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

Srikanth Gadicherla

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

Master's Thesis
Espoo, November 19, 2018

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

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

ABSTRACT OF
MASTER'S THESIS

Author: Srikanth Gadicherla

Title:
Data Analysis and Memory Methods for RSS Bluetooth Low Energy Indoor Positioning

Date: November 19, 2018 **Pages:** ix + 88

Major: Machine Learning and Data Mining **Code:** SCI3044

Supervisor: Professor Aki Vehtari

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

The thesis aims at finding a feasible solution to Bluetooth low energy indoor positioning (BLE-IP) including comprehensive data analysis of the received signal strength indication (RSSI) values. The data analysis of RSSI values was done to understand different factors influencing the RSSI values so as to gain better understanding of data generating process and to improve the data model. The positioning task is accomplished using a methodology called *fingerprinting*. The fingerprinting based positioning involves two phases namely *calibration phase* and *localization phase*. The localization phase utilises the memory methods for positioning. In this thesis, we have used *Gaussian process* for generation of radio maps and for localization we focus on memory methods: *particle filters* and *unscented Kalman filters*. The Gaussian process radio map is used as the measurement model in the Bayesian filtering context. The optimal fingerprinting phase parameters were determined and the filtering methods were evaluated in terms root mean square error.

Keywords: Indoor Positioning, Bayesian Filtering, Gaussian Processes, Fingerprinting

Language: 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 have 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 have worked with them. I am indebted to my advisor and guide *Roland Hostettler* for many wonderful discussions both offline and online. A special thanks to Tommi Gröndahl for helping me with "the very complicated *p-values*" and hypothesis testing, without which the thesis would be incomplete.

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 me regarding floorplans and measurement application.

I would like to thank my fellow lab-mates *Tuomas Sivula*, *Juho Piironen*, *Eero Siivola*, *Marko Järvenpää*, *Markus Paasiniemi*, *Olli-Pekka Koistinen*, *Akash Kumar Dhaka* and *Michael Riis Andersen* in the PML group for being accommodating and being readily available for discussions in the times when I was struck and out of ideas.

Last but not the least, I thank my friends and life gurus *Nonappa Bilinele*, *Jussi Gillberg*, *Kalyan G.V.S*, *Bhanu Prakash Reddy*, *Karthik Upadhya*, *Purendar Venkateshan*, *Srikkanth Balasubramanian*, *Afaque Hussian*, *Elīza Barkāne*, *Pradeep Kumar Eranti*, *Sriharsha Kuchimanchi*, *Manoj Kumar Tyagi*, *Mayank #nevermind Khandelwal*, *Sunny Vijay*, *Kunal Ghosh*, *Shishir Bhattarai*, *Maria Valdimirovna Pochtar*, *Abheeshta Putta*, *Ashish Sultania* for everything.

Otaniemi, November 19, 2018

Srikanth Gadicherla

Contents

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. Given the context, the concept of uncertainty is exceptionally relevant to the field of Machine Learning which involves making predictions. To formalise the concept of uncertainty and to generate valid measures for making decisions in terms of probability distributions, we use a methodology called *Bayesian Inference*. In this methodology, we update our beliefs 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. Initially, there were multiple search operations deployed to find the wreckage of the ill-fated flight, but all in vain. Stone et al. (2014) used the methodology mentioned above to recover the flight debris successfully only one week into search operation. The analysis incorporated not only the flight navigation data but also the failure points of the previous search operations. This prior information helped to estimate the exact location of the wreckage. Subsequently, this methodology was used in the search operations for the *Malaysian Airliner MH370* (Davey et al., 2016) in 2014.

From the above examples, it is evident that *Bayesian Inference* is a powerful tool to combine knowledge from multiple sources. The original notion with the current model evidence gives an opportunity to the modeller to criticise and update one's model. The above analysis mainly involved an iterative algorithm called *Particle Filter*. This thesis also implements Bayesian filters for solving the positioning problem in the indoor environment.

Chapter 1

Introduction

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

The quest to accurately navigate through unknown terrains has troubled humanity since time unknown. The sixteenth-century approach to solve this problem moved from being deterministic to one being deduced. The deterministic methods involved the rigorous application of astronomy and mathematics with a spate of tools like celestial globe, astrolabe, quadrant, cross-staff (Ahmad, 2015). 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 the magnetic compass, which was pivotal and it completely changed how humans navigated. Barring the fact that the dead reckoning accumulated errors, under certain constraints, it was quite accurate.

New age, new learnings.

Today in the twenty-first century, there are successful efforts for outdoor navigation systems which include the ubiquitous Global Positioning System (GPS) and regional systems, like, for example, Russia's Global Navigation Satellite System (GLONASS) and India's Indian Regional Navigation Satellite System (IRNSS). The explosion in the usage of smartphones, widespread cellular connectivity and GPS availability has made it convenient to navigate in the outdoor environment.

Now, the quest for such a kind of connectivity has moved indoors. In the pursuit for cheap and reliable indoor positioning system (IPS), the research

community has examined the problem mainly using technology like wireless local area network (WLAN), ZigBee, cellular mobile network, Bluetooth Low Energy (BLE) technology. Standing out, the BLE could become the de facto IoT connectivity device according to Ahmad (2015).

1.1 Motivation

It is well founded that we spend 90% of our time indoors (EU (2003), USEPA (2013)) and given the advancement in radio-based communications, there is ample opportunity for infrastructure based positioning. The indoor setting is complex, dynamic, and modelling such data makes it a hard problem to solve.

It is also a well-known fact that GPS operations (like Assisted GPS (Djuknic and Richton, 2001)) are limited indoors, and there is a void for a scalable solution which can seamlessly work both indoors and outdoors. There have been efforts to increase the GPS accuracy in the indoor environment using GPS-repeaters (Jardak and Samama, 2009) but this solution has higher initial costs and longer start-up times. In short, there is no widespread positioning system which works seamlessly both indoors and outdoors.

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

As technology is becoming more context and location-aware in the world of the Internet of Things (IoT), the location of the device is imperative for intelligent solutions. Research and development in indoor positioning would drastically proliferate the location-aware and location-based applications (Hazas et al., 2004) like, for example, assets tracking, product flow optimisation and product recommendation, making them a cornerstone in the world of IoT. The IPS is estimated to have a market value of 10 billion US dollars by the year 2020 (He and Chan, 2016).

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

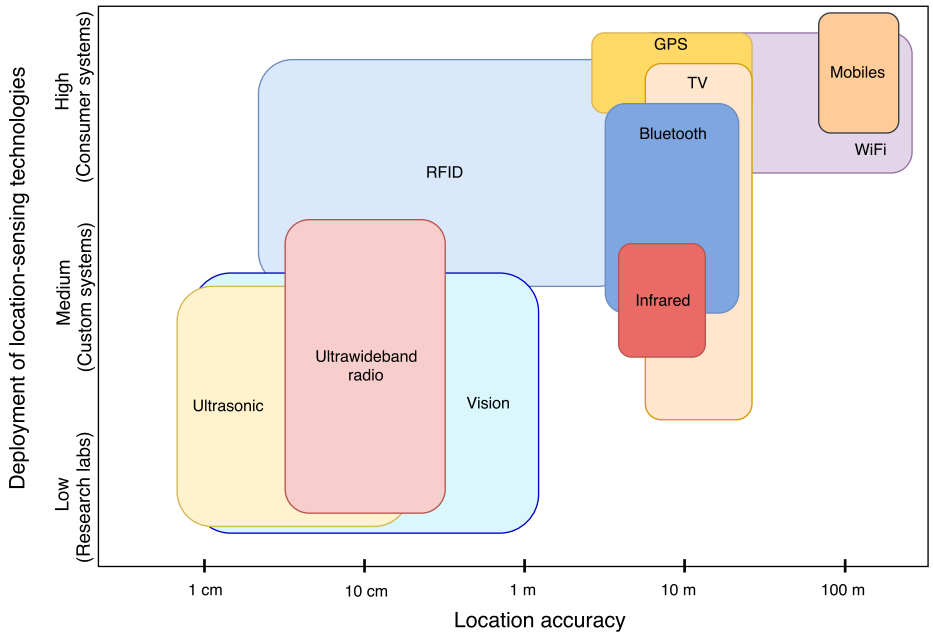


Figure 1.1: Technologies used in the Indoor Positioning systems with their location accuracies (Hazas et al., 2004).

1.2 Indoor Positioning

Indoor positioning systems in vague terms can be referred to as *indoor GPS*. Indoor environment can be office, hospital, shopping mall, school, airport, or a metro station, and positioning system using BLE makes sense for a larger arena as the accuracy of such a system falls within 10 metres accuracy (refer to Figure 1.1).

The indoor positioning problem pose different challenges from the aspect of the indoor environment, technology leveraged and methodology used for solving the problem. Specifically the BLE technology, which utilises received signal strength indicator (RSSI) from the signal advertised for positioning, even the factors like the orientation of the BLE chip installed can affect the accuracy of the positioning system in addition to signal attenuation. The RSSI based indoor positioning methods directly or indirectly learns the path loss component of the propagation of the signal in an indoor setting. For better understanding the signal characteristics, the data analysis of RSSI from BLE is imperative and moreover, this aspect is lacking in the previous literature. One of the methods, *fingerprinting* involves inferring about the location based on observations from the scene. This is accomplished using the features recorded for a set of locations in the form of a prerecorded dataset and the

fingerprinting method aims at creating a unique feature map for all locations. This methodology is well studied and performs well for indoor positioning, and we use it to generate radiomaps with a non-parametric method called Gaussian processes. A class of methods which can utilise the radiomap in fingerprinting method are *memory methods*. Memory methods are the probabilistic methods that keep a memory of both previous (predicted) locations and measurements. These methods are flexible to incorporate additional information like the physical model of the process. Bayesian filters are a classic case of memory methods.

Fine tuning the parameters in both the fingerprinting and memory methods in addition to the attenuation factors in RSSI makes indoor positioning an interesting problem to solve, hence, a detailed study of Bluetooth low energy based indoor positioning system (BLE-IP) is inevitable.

1.3 Goal & Contribution of the thesis

The goal of the thesis is to develop a feasible BLE-IP solution, given that the beacons[1] are embedded inside the luminaires. The thesis aims at using memory method in conjunction with the fingerprinting methodology. We also aim to record the RSSI values using a smartphone device utilising an in-house developed android measurement application. We plan to implement and evaluate the memory methods in MATLAB.

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 fingerprinting parameters and evaluating the memory methods.

1.4 Outline of the thesis

This thesis is structured as follows. Chapter 2 gives the background to various topics dealt in the thesis. Chapter 3 covers the concept of fingerprinting and method for generating the radiomaps for memory methods. It also covers a review into non-memory based algorithms and related technologies

[1] We will interchangeably use *module* or *beacon* in the thesis.

concerning IP. The theoretical distinction between memory and non-memory methods is also given. In Chapter 4, a detailed background on Gaussian processes for IP is given. In Chapter 5, we introduce and explain the memory methods – the non-linear Bayesian filtering algorithms. The measurement setup, including the measurement application and BLE modules for the experiments, are described in the Chapter 6. The Chapter 7 discusses in detail the statistical analysis of the RSSI data from the BLE and quantify the different selected bias factors that affect the signal indoors. In Chapter 8, we evaluate the memory methods using the different optimal fingerprinting parameters discussed therein, and results are stated. We discuss the results in Chapter 9 and also state the future course of action.

Chapter 2

Background

In this chapter, we give background to the concepts used in the thesis. First, in Section 2.1, we introduce Bluetooth low energy. Next, in Section 2.2 and 2.3, we discuss about the received signal strength indication and the challenges involved in using it. Next, in Section 2.4, we introduce the memory and non-memory methods for localization. Last, in Section 2.5, we discuss about the hypothesis testing and different tests used in the thesis.

2.1 Bluetooth Low Energy

Bluetooth Low Energy (BLE) is the latest low-powered wireless technology operating 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). The BLE uses 0.50 W power when compared to classic Bluetooth which uses 1 W and has maximum peak current of 15 mA in contrast to classic Bluetooth's 30 mA. 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.

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 either be directly triangulated to find the location or radiomaps could be learnt to estimate the location based on that. We discuss in detail about BLE's in Chapter 6.

2.2 Received Signal Strength Indication

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 widely used

parameters from the radio signals for indoor positioning. One of the characteristics of RSSI is that the values are quantised, that is, the values are measured only in integers. It is 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 RSSI is inversely proportional to the distance and successfully used it for positioning. Generally, the units of RSSI is irrelevant as we are interested only on its variations[1].

2.3 Challenges in 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, movement of human beings. 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 and multipath, reflection, refraction (Hashemi, 1993) and small-scale fading due to dynamic nature of environment (Kaemarungsi and Krishnamurthy (2012), Luo et al. (2011)). The problem is exacerbated by co-channel interference given the fact that BLE radio signals follow 2.4 GHz ISM band (Hashemi, 1993). As the radio waves are readily absorbed by water, a single human can attenuate the signal upto -3.5 dBm (Bahl et al., 2000). It is also seen that BLE RSSI values are troubled with 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). Moreover, the malfunctioning BLE modules can lead to not registering of the beacon's RSSI either during the calibration phase or in the localisation phase.

The primary challenge which arises from the issues mentioned above for BLE-IP is getting a right data model. The probability distribution of the recorded RSSI values can be either be left skewed, right skewed or symmetric based on the indoor environment. The time of the day is also a factor for the variations in the RSSI values, for example, in an office space, there are more people in the day when compared to other times. Hence, the RSSI is a function not only of location but also of time. Log-normal, Gaussian and exponential distributions were different models which were used previously (Honkavirta, 2008). The variety of smartphone hardware and its internal

[1]RSSI is defined in units of signal power, that is, mW or dBm.

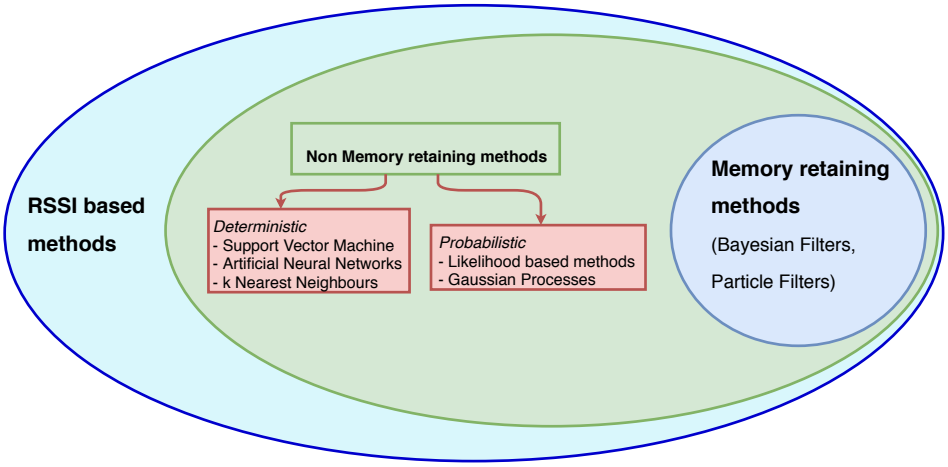


Figure 2.1: Categorization of RSS based methods for indoor positioning. The memory methods are special case of non memory methods.

preprocessing of RSSI for quantisation also adds bias exacerbating the issues for data modelling. Honkavirta (2008) and Kaemarungsi and Krishnamurthy (2012) also explore cell based positioning for overcoming the quatisation issue. Next, we discuss different algorithms used for localisation.

2.4 RSS based Localization Algorithms

The RSS based methods can be categorized as *non-memory methods* and *memory methods*. The memory methods are a special case of non-memory methods (refer Figure 2.1). Additionally, the non-memory methods can be used in conjunction with memory methods. For example, the *k-Nearest Neighbours* likelihood model can be used in conjunction with Particle filters. We next introduce the non-memory and memory methods.

The non-memory methods are range based methods (Aravecchia and Messelodi, 2014). The range based methods basically transform the RSSI to a distance measure, that is, the distance between the mobile unit (MU) and the beacon. Hence, the non-memory methods are *one-shot positioning* methods and they do not keep memory of either previous locations or previous measurements. As seen from Figure 2.1, the non-memory methods can either be deterministic or probabilistic. We describe the probabilistic non-memory method *Gaussian processes* in Chapter 4.

The memory based methods are the range-free methods (Aravecchia and Messelodi, 2014). These methods utilise the previous predicted locations and measurement to predict the current location and, hence, can track tra-

jectories. Tracking trajectories can minimize the prediction fluctuations in comparison to on-shot predictions. These are probabilistic methods and use a observation model to generate likelihood for the current RSSI measurements. For example, the observation model can either utilise the path loss propagation or utilise the stored knowledge base in the form of either the reference table or the radiomap from the fingerprinting methodology for getting the predictions. The accuracy of location predictions depends on the precision of the observation model. Additionally, the probabilistic nature of the memory methods allow seamless integration of inertial sensor data. The widely used memory methods are Bayesian filters – Grid Filter, Kalman Filter, Extended Kalman Filter, Unscented Kalman Filter, Sequential Monte Carlo methods like Particle Filters (see Figure 2.1). Refer to Chapter 5 for detailed discussion on these methods.

2.5 Statistical Hypothesis Testing

For the empirical (sample) data collected for experiments, we need to draw inferences in order to understand the data generating process. This allows us to hypothesise about the data population and utilise them to model our problem. Hence, the hypothesis testing is also called *confirmatory data analysis*.

The hypothesis testing procedure is as follows:

1. A *null hypothesis* H_0 and an *alternative hypothesis* H_1 are formulated.
2. A sample statistics for the given hypothesis is considered.
3. Samples from the population are selected.
4. The test statistic is determined and the results are inferred.

Generally, the null hypothesis states the status quo and the alternative hypothesis states otherwise. For example, the 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. If the computed test statistic falls in the region of acceptance, which is a range of sample statistics, then we have no evidence for rejecting the null hypothesis. The critical values are the threshold from the region of acceptance to the region of rejection. These are also sometimes referred to as *critical region*. In this thesis, we

investigate the statistical significance for the parameter mean and median from the different samples recorded. The different tests used are described below.

2.5.1 Shapiro-Wilk Normality Test

The *Shapiro-Wilk* test uses the W statistic to confirm if a sample comes from normal distribution. This test calculates the W statistic. For a given random sample, $x_1, x_2, ..., x_n$, the W statistic is given by

$$W = \frac{(\sum_{i=1}^{n} a_i x_{(i)})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \tag{2.1}$$

where the $x_{(i)}$ are the ordered (or sorted) sample values and a_i are the mean, variance and covariance generated constants using the order statistic of sample from a normal distribution. If the *p-value* computed is less than the critical value (here 0.05), we reject the null hypothesis and we have no evidence to state that given sample follows normal distribution. For comprehensive discussion refer to the pioneering paper by Shapiro and Wilk (1965).

2.5.2 Wilcoxon Rank-Sum Test

The *Wilcoxon rank-sum test* or *Mann-Whitney U test* for median is a non-parametric statistical test (Daniel, 1999). The non-parametric nature of the test is due to two assumption about the underlying distribution, one, that it is continuous, and two, its probability density function is symmetric. The Wilcoxon signed-rank test requires equal samples, but the rank-sum test applies to unequal samples (Daniel, 1999).

The null hypothesis H_0 states that the two medians scores M_1, M_2 from the sample data are equal while the alternative hypothesis H_1 states otherwise.

$$\begin{aligned} H_0 &: M_1 = M_2 \\ H_1 &: M_1 \neq M_2 \end{aligned} \tag{2.2}$$

The assumptions about the sample data are

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

Table 2.1: A 2×2 Contingency table

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

3. the parameter (median) being inferred is continuous.

The Wilcoxon rank-sum test uses the *chi square statistic* for making inferences. The chi square statistic X^2 is computed using the equation

$$X^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}, \tag{2.3}$$

where a, b, c, d and n are defined in the Table 2.1. It is important to note that in the construction of the 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 χ^2 distributed with degree of freedom 1. It implies that computed X^2 should be less than 3.841 given that α is defined as *0.05*.

Chapter 3

Location Fingerprinting

In this chapter, we describe in detail about location fingerprinting. In Section 3.1, we introduce the concept of location fingerprinting technique. In Section 3.2, we discuss about the reference maps and give the mathematical formulation for it. In Section 3.3, we describe about the radiomaps.

3.1 Introduction

Fingerprinting is a method for signal pattern recognition (Aguilar-Garcia et al. (2015), Honkavirta (2008)). It exploits the relationship between the signal based characteristics to the location for positioning (Kaemarungsi and Krishnamurthy, 2012). Bahl and Padmanabhan (2000) first introduced fingerprinting for wireless local area network based positioning. A typical fingerprinting-based localisation technique juxtaposes the RSSI to the already ones present in the form of a reference table or a radiomap (Gu et al., 2016). Reference table is the database of locations with their corresponding signal strength values from all the available beacons in the positioning space whereas the radiomap is a continuous interpolation of signal values in the concerned positioning space for a beacon. Yiu et al. (2017) calls usage of the reference table for localisation as traditional fingerprinting.

A dense network of beacons helps the fingerprints to be unique corresponding to a location (Huang et al., 2011), in turn, improving the accuracy of the positioning system. Fingerprinting is a widely researched positioning methodology because of its ease in implementation and economical in the context of usage of existing infrastructure (Yiu et al., 2017). However, the collection of fingerprints is especially tiresome and time-consuming task, and recovering the missing data involves additional manual intervention. To overcome the shortcomings mentioned earlier, Zhang et al. (2013) proposed

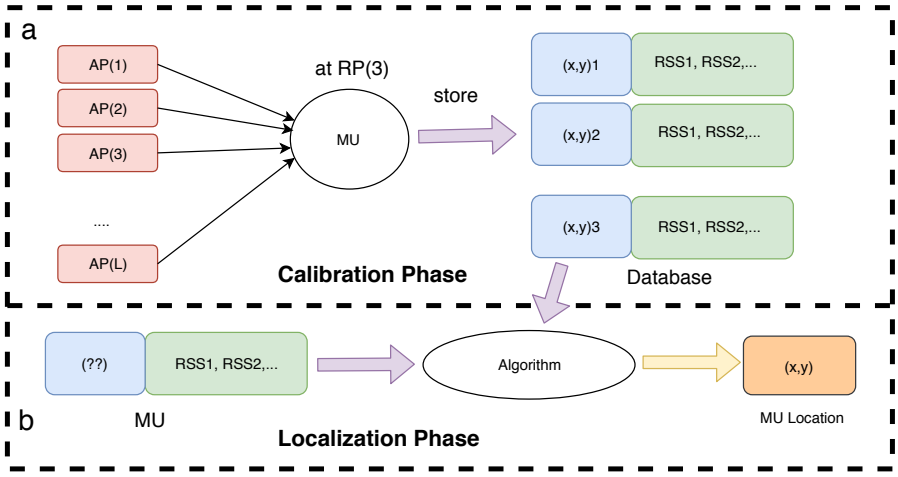


Figure 3.1: An overview of the fingerprinting technique. (a) Calibration phase, (b) Localisation Phase.

a compressive sensing based *sparsity rank singular value decomposition* technique used with *k-Nearest Neighbours*. This technique compensates for the missing values in the data and mitigates the redundancies caused due to multipath and interference. In addition, there is also the problem of redundant radiomaps. This is caused due to a change in the indoor environment. Yiu et al. (2017) shows that the positioning accuracies are overestimated due to this and explains its ill effects.

As illustrated in Figure 3.1, the fingerprinting based positioning method is categorized into two phases: one *offline calibration phase* and next, a *online localisation phase*. The offline calibration phase is the training phase. In this phase, we measure the RSSI of the radio-frequency signal from all the detectable beacons. These measurements are recorded using a mobile unit (MU) at pre-determined locations called *calibration points* and these measurements are recorded for a certain time called *calibration time*. These location-specific measurements are called *fingerprints* (or signatures; Yiu et al. (2017)). This phase involves building a database of fingerprints called *reference table* and is illustrated in 3.1(b). This database is later used for generating *radiomaps*. Next, the online localisation phase involves recording the real-time RSSI, and the locations are predicted using an algorithm. This phase makes use of the radiomap generated in the offline calibration phase. Figure 3.1(b) illustrates the same.

3.2 Reference Table

The first step in the creation of the radiomap is to generate a *reference table* (refer to figure (3.2)). A reference table is a collection of calibration points and RSSI from all the beacons. Mathematically, it is

$$\mathcal{R} = \bigcup_{n=1}^{N} \bigcup_{m=1}^{M} \{(x_n, y_n), \mathbf{r}_{nm}\}, \quad if\ \mathbf{r}_{nm} \in \emptyset,\ \ \mathbf{r}_{nm} = r_{min}, \tag{3.1}$$

where N is the number of calibration points, M is the number of beacons and r_{min} is the empirical minimum for the RSSI. In case the device is not heard during the calibration phase, we use the empirical minimum of a beacon. Note that due to indoor signal stochasticity the number of RSSI records varies for different beacons at different locations, and may even vary in time.

To investigate the effects of signal stochasticity on positioning accuracy, we introduce two reference tables. *Reference table 1* ($\mathcal{R}_1$) has a single entry of RSSI for a beacon at each calibration point (Honkavirta, 2008). It is a common practice to use the mean estimate of all the RSSI collected for a particular beacon at a certain calibration point, but there is the downside of losing variance in the data. Hence, this is also called the *mean reference table*. Refer to Figure 3.2(a) for the Matlab data structure of $\mathcal{R}_1$.

$$\mathcal{R}_1 = \bigcup_{i=1}^{N} \bigcup_{j=1}^{M} \{(x_i, y_i), r_j^{mean}\}, \quad if\ r_j^{mean} \in \emptyset,\ \ r_j^{mean} = r_{min}. \tag{3.2}$$

Next, we consider the *reference table 2* ($\mathcal{R}_2$) and the equation of reference table remains as is as defined in Equation (3.1). Therefore, there are multiple entries of RSSI values for a single calibration point and this number could be variable. Refer to Figure 3.2(b). It is hypothesised that if the variance of the RSSI for a particular beacon is intact, it would add to the accuracy of the constructed radiomap and in turn add to the accuracy of the positioning algorithm.

Yiu et al. (2016) suggests to use the *minimum power device sensitivity* of the beacon as an empirical minimum in case of an unheard access point. In practice, a mobile unit can record an RSSI value below the minimum sensitivity level. In this thesis, if a particular beacon is unheard or is below device sensitivity level, we update the RSSI to the device sensitivity level which is $-110\,\text{dBm}$. So the same is used in the Equation (3.1).

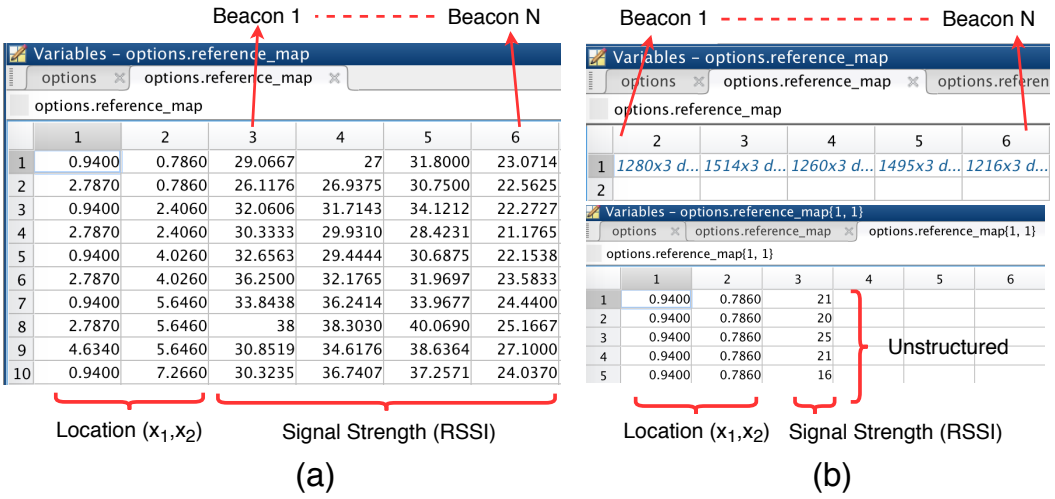


Figure 3.2: Illustration of the reference tables used in this thesis. (a) It shows the conventional mean reference table $\mathcal{R}_1$. (b) It shows the $\mathcal{R}_2$ reference table. It depicts multiple RSSI measurements for the same location.

3.3 Radiomap

As mentioned earlier, fingerprinting involves learning the spatial properties of the signal in the calibration phase. Radiomap is a learnt outcome of the calibration phase and is a continuous interpolation of the signal strength values over the positioning area for a particular access point. Hence, each beacon has its own radiomap mimicking its signal path propagation. In the fingerprinting technique, we are generally interested in capturing the variation of signal strength values over the localisation space. The radiomap can be obtained 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). Wang et al. (2011) uses the difference of the RSSI values in the radiomap generation process, and it was done primarily to mitigate the receiver gain difference in the devices. Kriz et al. (2016) proposes using distributed systems for creating and updating the radiomap. We can also utilise characteristics like *Signal to Noise ratio* (Bahl and Padmanabhan, 2000) in the generation of radiomaps but RSSI usually suffices as it has a stronger correlation to distance. We describe in detail about RSSI – distance relationship in Chapter 7.

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 warehouse might need more accurate position prediction, that is,

less than few centimetres but location of a room inside an office would allow us for errors up to a few metres. Hence, tuning the calibration parameters for a particular indoor environment is crucial, we discuss about this in Chapter 8.

Chapter 4

Gaussian Process Model for Indoor Positioning

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

In this chapter, we introduce Gaussian processes in Section 4.1. Next we give the motivation for using Gaussian processes in Section 4.2. Finally, in Section 4.4, we explain in detail the different modelling techniques for IP.

4.1 Gaussian Processes

Gaussian Processes (GPs) are a non-memory methods which are used for one-shot positioning (Schwaighofer et al., 2004). GPs are random processes for describing distributions over functions (Rasmussen and Williams, 2005) and are non-parametric. The term 'process' originates from signal processing while 'random field' is used in spatial statistics (Solin, 2016). Traditionally, GPs were known as *kriging*, which is a regression task but they can also be used for classification task (Rasmussen and Williams, 2005). The spatial properties of the target function are inferred using the learnt hyperparameters of the kernel function (Solin, 2016) and we use them for generating the radiomaps, basically forming the observation model.

Gaussian process is a collection of random variables, of which any finite collection of variables follows a joint Gaussian distribution (Rasmussen and

Williams, 2005). For a latent[1] stochastic process $f(\mathbf{x})$, we define mean function $m(\mathbf{x})$ and covariance function $k(\mathbf{x}, \mathbf{x}')$ as

$$\begin{aligned} m(\mathbf{x}) &= \mathbb{E}_p\big[f(\mathbf{x})\big] \\ k(\mathbf{x}, \mathbf{x}') &= \mathbb{E}_p\big[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))\big] \end{aligned} \tag{4.1}$$

and then we can draw inferences from the function $f(\mathbf{x})$ by putting a GP prior

$$f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{4.2}$$

GPs are generalization of multivariate Gaussian distribution (Rasmussen and Williams, 2005) and contrary to sampling from distributions which yields finite-dimensional vectors, sampling here yields infinite dimensional vectors. In practise, as computation on infinite dimensions is infeasible, we have a finite grid space on which we draw the predictions.

4.2 Motivation for using the GPs

Due to the complexity of the indoor environment, the simple parametric distributions are inadequate to model such complex RSSI data (Seco et al., 2010). Hence, we need flexible models to accurately built the data model, and Gaussian processes are the perfect solution for that. The motivation for using GPs for RSS based indoor positioning are (Ferris et al., 2007b),

1. *Continuous Locations*: GPs have excellent capabilities of interpolating over other test locations, and hence, can overcome the quantisation issue in the RSSI. They are flexible as they do not need any designated training points for accomplishing the positioning task.

2. *Arbitrary likelihood models*: A wide variety of complex data models can be approximated given the non-parametric nature of GPs: multiple kernels can be used in conjunction with each other (Rasmussen and Williams, 2005). Hence, GPs 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 give the uncertainty estimates for the predicted location value. This is mainly dependent on

[1] because it is an unobserved function.

the amount of data and the estimated noise around the training 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 by maximizing the marginal likelihood (Rasmussen and Williams, 2005). For example, given the knowledge of signal propagation indoors, informative priors can be used for the length scale parameters in the kernel functions. In addition, these also point to spatial correlation between the recorded measurements and, hence, can learn the measurement noise based on it (Ferris et al., 2007b).

4.3 Application in Indoor Positioning

GPs were first used by Schwaighofer et al. (2004) and later by Li et al. (2005) as error correction map to overcome the non-line-of-sight problem. The localisation was performed using GPs and k-nearest neighbour algorithm on a cellular network. They showed that GP's gave an accuracy of 7.5 metres. They used Matérn covariance function for modelling the individual stations and used the maximum likelihood method for prediction of location. Bisio et al. (2017) proposes a hybrid method to solve the problem with GP measurement model aiming to reduce the complexity of the calibration phase. Schüssel and Pregizer (2015) uses a log path loss model as the mean function in the GP's which is in similar lines with Yiu et al. (2016).

Atia et al. (2013) proposed a solution similar to the GPS but replaced the GPS signals to the RSSI values of the WiFi network. The solution neither had any overhead in the form of fingerprinting nor had any additional infrastructural needs. They used a hybrid propagation model using GPs and showed that it models RSS residuals better than path loss models. Their proposed solution gave $2 - 3$ metres accuracy.

WiFi-SLAM (Ferris et al., 2007a) uses GP latent variable models to build an indoor map as the user moves through a building. This methodology does not require fingerprinting, hence, has no calibration phase. The caveats are that the method has only been experimented with simple, rectangular floor plans and additionally makes restrictive assumptions about the user movements. This methodology requires integration with additional sensors to determine the direction and distance of movement.

4.4 GP modeling for Indoor Positioning

In a broader sense, the GPs can be used in the following ways:

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

$$s_j = f(x) + \epsilon \tag{4.3}$$

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

$$\begin{aligned} GP :& \mathbf{R}^d \rightarrow \mathbf{R} \\ & x \ \mapsto \ s \end{aligned} \tag{4.4}$$

This might look unremarkable but works for most of the problems and can 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 *radiomaps*. Radio maps are discussed in Chapter 3 and see Figure 4.1 for a sample radiomap. GP here can be exploited in the form of measurement model using the learnt radio maps, as in (Ferris et al., 2007b).

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 get the location estimate directly from the RSSI values, that is, from signal space to metric space

$$x = f(\mathbf{s}) + \epsilon, \tag{4.5}$$

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

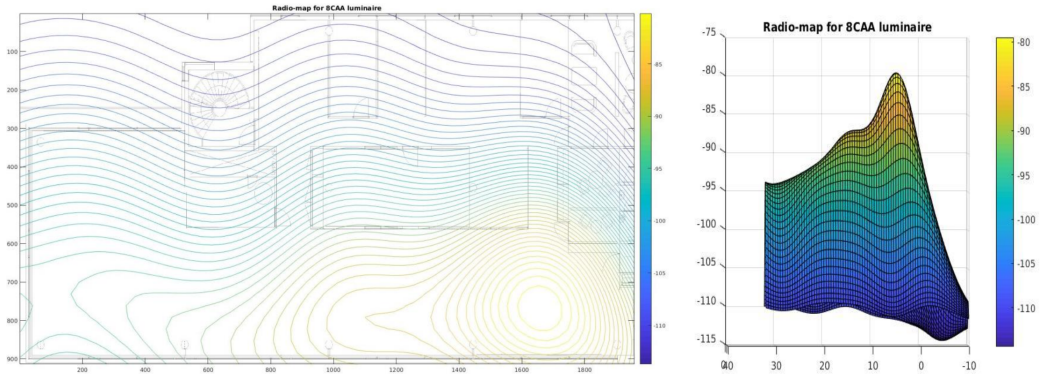


Figure 4.1: Illustration of GP radiomap generated from beacon *8CAA* at Helvar R&D.

$$\begin{aligned} GP : \mathbf{R}^q &\rightarrow \mathbf{R}^d \\ \mathbf{s} &\mapsto x \end{aligned} \tag{4.6}$$

This is achieved through *Maximum Likelihood estimation (MLE)*. MLE suffers from initialization problem and also depends on the type and convexity of likelihood function. 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 conter intuitive. One observation from Equation (4.5), evidently a vice, is that it assumes that the input RSSI values are noise free, that is, we tend to ignore the stochasticity of the signal propagation.

3. *Hybrid* modeling: 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 (Aravecchia and Messelodi, 2014). This forms the first GP fold. The second fold uses vague location estimates and passes 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.

$$\begin{aligned} \tilde{x} &= f_{GP_1}(\mathbf{s}) \\ x &= g_{GP_2}(\tilde{x}) \end{aligned} \tag{4.7}$$

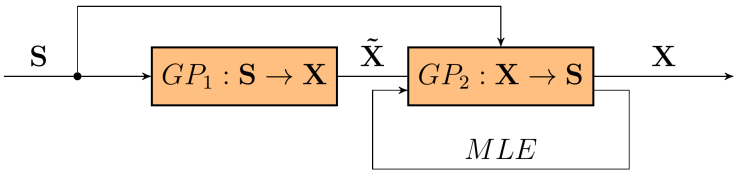


Figure 4.2: Hybrid model. courtesy (Aravecchia and Messelodi, 2014).

where f is function which follows GP_1 from the Equation (4.6) while g follows GP_2 from the Equation (4.4), $\tilde{x}$ is the predicted estimate of location from signal strengths $\mathbf{s}$ from luminaires whereas x is the updated location estimate.

Chapter 5

Memory Based Filtering Methods

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

Memory methods are the probabilistic methods which implicitly keep a memory of the previous states and measurements to predict the current state (position) parameter. The memory methods typically include the Bayesian filtering methods which make use of *Hidden Markov models*. Bayesian filters following the Markov property performs the statistical inversion to predict the hidden (or latent) state. The Markov property helps relax computational complexity, and we only compute the marginal posterior for the present state. In this chapter, in Section 5.1, we introduce the Bayesian filtering. Next, in Section 5.1.1, we formulate the filtering problem and derive a simple prediction, update equations for predicting the state. Then, in Section 5.2 we explain the filtering methods – Particle and Unscented Kalman filters.

5.1 Bayesian Filtering

Most of the real-world problems involve evaluating an unknown quantity given some external measurements and 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 world-view for optimal filtering (Särkkä, 2013). Optimal filtering historically refers to statistical methods for predicting state for time-varying systems. These methods use statistical guarantees for optimality. The Bayesian inference methods involve incorporating the initial belief about 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 the posterior

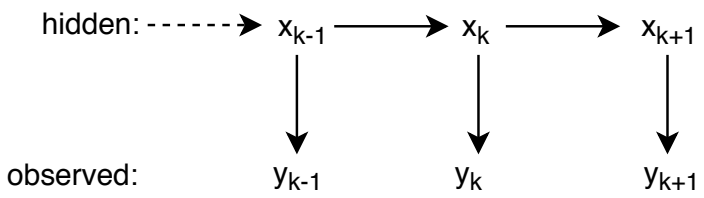


Figure 5.1: The Hidden Markov model illustrating the states $\mathbf{x}$ being indirectly estimated from the measurements $\mathbf{y}$.

distribution which allows for inference on the quantity of interest (for example, the location of person). For a comprehensive overview of 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 (see Figure 5.1). This finite state-space Markov chain with data partially observed can be termed as hidden Markov model filter (Doucet et al., 2001). Applications can be found in the field like navigation, telecommunication and economics.

5.1.1 Problem Formulation

We consider the following state space model

$$\begin{aligned} \mathbf{x}_t &= f(\mathbf{x}_{t-1}, \boldsymbol{\theta}) + \mathbf{q}_{t-1}, \\ \mathbf{y}_t &= h(\mathbf{x}_t, \boldsymbol{\theta}) + \mathbf{r}_t, \end{aligned} \tag{5.1}$$

where f is the state function for the dynamic model, h is the measurement function for data model, $\mathbf{q_{t-1}}$ is the process noise, $\mathbf{r_t}$ is the measurement noise. The state and the measurement transition are the function of state $\mathbf{x}$ and parameter $\boldsymbol{\theta}$.

Restricting ourselves to Markovian, non-linear and non-Gaussian scenario (Doucet et al., 2001), the unobserved states (hidden or latent variables) are $\{\mathbf{x}_t; t \in \mathbb{N}\}$, $\mathbf{x}_t \in \mathbb{R}^n$ 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 likelihood function $p(\mathbf{y}_t|\mathbf{x}_t)$ given the noisy observations $\{\mathbf{y}_t; t \in \mathbb{N}\}$, $\mathbf{y}_t \in \mathbb{R}^m$, we write our probabilistic model as

$$\begin{aligned} p(\mathbf{x}_0) & \\ p(\mathbf{x}_t|\mathbf{x}_{t-1}) & \ for\ \ t \geq 1 \\ p(\mathbf{y}_t|\mathbf{x}_t) & \ for\ \ t \geq 1. \end{aligned} \tag{5.2}$$

We aim at computing the *joint posterior distribution* $p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$ at any time t and using the Bayes' theorem

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})\ p(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t})}. \tag{5.3}$$

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[1] which forms the likelihood function,
- $p(\mathbf{y}_{1:t})$, is the evidence and is a normalizing constant

$$p(\mathbf{y}_{1:t}) = \int p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})\ \ p(\mathbf{x}_{0:t}) d\mathbf{x}_{0:t}. \tag{5.4}$$

We are principally interested in the expectation of the marginal posterior distribution called the *filtering distribution* $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ and given an arbitrary function $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$, the expectation is

$$I(\mathbf{g}_t) = \mathbb{E}_{p(\mathbf{x}_t|\mathbf{y}_{1:t})}[g(\mathbf{x}_t)] \triangleq \int g(\mathbf{x}_t)\ p(\mathbf{x}_t|\mathbf{y}_{1:t})\ d\mathbf{x}_t \tag{5.5}$$

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

$$p(\mathbf{x}_{0:t+1}|\mathbf{y}_{1:t+1}) = p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})\ p(\mathbf{x}_{t+1}|\mathbf{x}_t)}{p(\mathbf{y}_{t+1}|\mathbf{y}_{1:t})}. \tag{5.6}$$

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

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

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)\ p(\mathbf{x}_t|\mathbf{y}_{1:t})\ d\mathbf{x}_t \tag{5.7}$$

[1] also called data model or observation model. Here, we use the terms interchangeably.

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

$$p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t+1}) = \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})\, p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})}{\int p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})\, p(\mathbf{x}_{t+1}|\mathbf{y}_{1:t})\, d\mathbf{x}_{t+1}} \tag{5.8}$$

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

If assumed a linear Gaussian 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 does not 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, Gauss-Hermite Kalman filter, Cubature Kalman filter and Particle filters (Särkkä, 2013). Particle filters are sequential Monte Carlo methods of which Bayesian filtering is a special case (Särkkä, 2013).

5.1.2 Advantages of Bayesian Filtering

In the filtering framework, it can accommodate adding additional prior information. The information like the (angular) speed of the person can be incorporated through the dynamic model. The presence of the walls and explicit information can be used manipulate the filtering algorithm for additional accuracy. For example, in case of the particle filters, if the sampled particle exceeds or hits the wall, they can be replaced by particles in valid positions. This framework also provides the opportunity to track the trajectory of the state. In the case of the malfunctioning of the sensor or noisy measurements, the solution degrades gracefully providing a chance for a contingency plan.

5.2 Filtering

In this section we will describe the probabilistic filtering methods. Based on the literature study, we select the promising filters – Particle filters and Unscented Kalman filters in this thesis.

5.2.1 Particle Filters

Sequential Monte Carlo (SMC) methods are simulation-based methods for generating draws from target distribution through sequentially generating weighted particles from an intermediate sampling distributions. These are a set of flexible methods for approximating any sequence of probability distributions. Sequential Monte Carlo methods with hidden Markov model are called *Particle Filters*. The weights of the particles are corrective measures for bias reduction with respect to corresponding *marginal posterior distribution* (or filtering distribution; (Liu (2008a), Doucet et al. (2001)).

Importance Sampling (Gelman et al., 2014) is an efficient Monte Carlo integration method applied when sampling from the target posterior distribution is implausible. It was named "importance" so as to underscore the important regions which become critical in the high dimensional posterior space (Liu, 2008a). 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 expectation

$$\begin{aligned} I(\mathbf{g}_t) &= E_{p(\mathbf{x}_t|\mathbf{y}_{1:t})}[g(\mathbf{x}_t)] \\ &= \int g(\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{y}_{1:t})\, d\mathbf{x}_t \\ &= \int \left[g(\mathbf{x}_t) \frac{p(\mathbf{x}_t|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_t|\mathbf{y}_{1:t})} \right] \pi(\mathbf{x}_t|\mathbf{y}_{1:t})\, d\mathbf{x}_t \,. \end{aligned} \tag{5.9}$$

The choice of importance distribution is critical and has to be non-zero in the important posterior regions. For a comprehensive discussion on Monte Carlo error due to importance sampling, refer to Vehtari et al. (2015). The Monte Carlo approximation of N samples drawn from $\pi(\mathbf{x}_t|\mathbf{y}_{1:t})$ is

$$\begin{aligned} E[g(\mathbf{x}_t)|\mathbf{y}_{1:t}] &\approx \frac{1}{N} \sum_{i=1}^{N} \frac{p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})} g(\mathbf{x}_t^{(i)}) \\ &= \sum_{i=1}^{N} w_t^{(i)}\, g(\mathbf{x}_t^{(i)}) \end{aligned} \tag{5.10}$$

where the weights w_t are defined as:

$$w_t^{(i)} = \frac{1}{N} \frac{p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})} \tag{5.11}$$

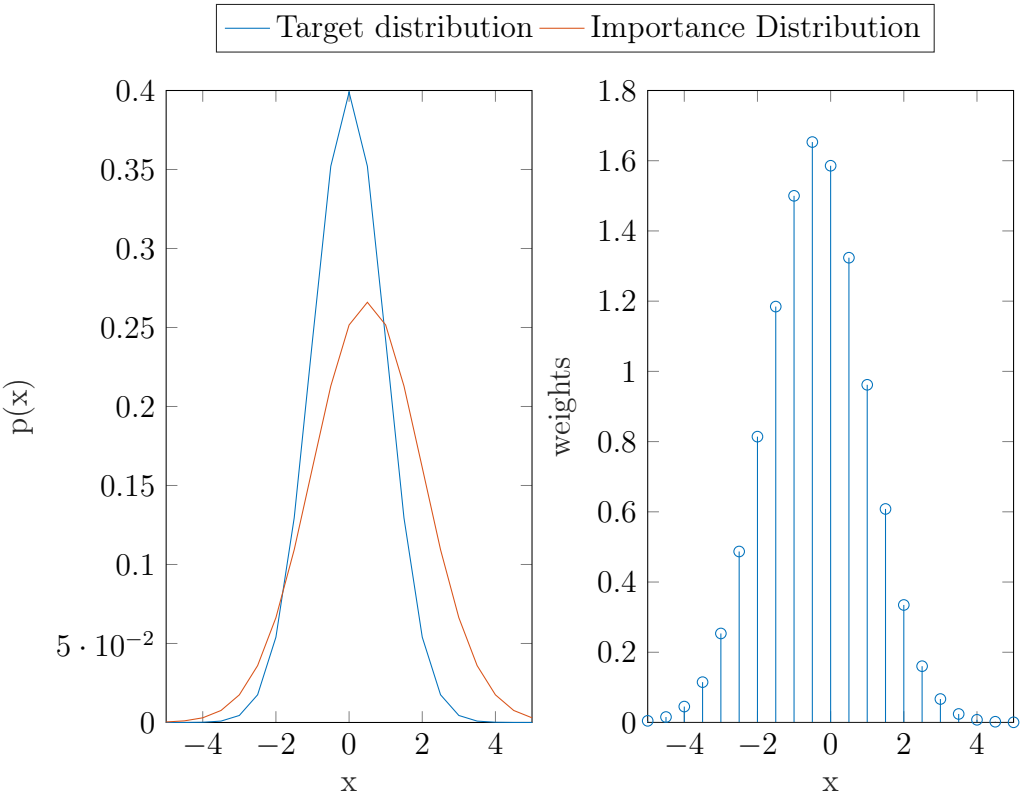


Figure 5.2: (left) The importance distribution and target distribution. (right) Approximate posterior distribution in the form of weights.

Now, the approximate filtering distribution can be written as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \approx \sum_{i=1}^{N} w_t^{(i)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}) \tag{5.12}$$

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

As we do not have the filtering distribution readily available, we compute it using the Bayes' rule. Hence,

$$p(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)})}{\int p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)})} \tag{5.13}$$

Using equation (5.13) in equation (5.9), we arrive at

$$
\begin{aligned}
E[g(\mathbf{x}_t)|\mathbf{y}_{1:t}] &= \sum_{i=1}^{N} \left[\frac{\frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{y}_{1:t})}}{\sum_{j=1}^{N} \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t^{(j)})p(\mathbf{x}_t^{(j)})}{\pi(\mathbf{x}_t^{(j)}|y_{1:t})}} \right] g(\mathbf{x}_t^{(i)}) \\
&= \sum_{i=1}^{N} \widetilde{w}^{(i)} g(\mathbf{x}_t^{(i)}).
\end{aligned} \tag{5.14}
$$

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 eventually will have infinite variance due to the presence of heavy right tail (Vehtari et al., 2015), we need sophisticated methods to solve this issue. We summarise the importance sampling algorithm in Algorithm 1.

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 move to the next filtering sequence. Resampling is performed using the criterion *effective sample size* (S_{eff}; Liu (2008b)). The effective sample size is found using

$$
S_{eff} = \frac{1}{\sum_{i=1}^{N} (w_t^{(i)})^2}. \tag{5.15}
$$

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.

The particle filter algorithm forms weighted set of particles from the importance distribution at every time step t, that is, $\{(\mathbf{w}_t^{(i)}, \mathbf{x}_t^{(i)}) : i = 1, ..., N\}$, which approximates the filtering distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. To derive the algorithm, we consider the full posterior distribution consisting of all previous

Algorithm 1: Importance Sampling

1 Sample from the importance distribution:

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

2 Calculate the unnormalized weights w_t:

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

3 Normalize the weights:

$$\widetilde{w}^{(i)} = \frac{w^{(i)}}{\sum_{j=1}^{N} w^{(j)}}, \qquad i = 1, ..., N$$

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

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

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

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

.

Algorithm 2: Resampling

1 Sample from the current state particles $\{\mathbf{x}_t^{(i)}, i = 1, ..., N\}$ with probability equal to $\{w_t^{(i)}, i = 1, ..., N\}$.

2 Substitute the old particle set with the newly drawn particles.

3 Reweight the new particles as $w_t^{(i)} = 1/N$

states and measurements. The recursion goes as

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_{0:t}, \mathbf{y}_{1:t-1})p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t-1}) \tag{5.16a}$$
$$= p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1})p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}) \tag{5.16b}$$
$$= p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}). \tag{5.16c}$$

Equations (5.16) use the Markov property that probability distribution of current measurement only depends on current state and probability distribution of current state depends only on previous state.

Using equations (5.13) and (5.14), we can similarly write here:

$$w_t^{(i)} \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})p(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})}{\pi(\mathbf{x}_{0:t}^{(i)}|\mathbf{y}_{1:t})} \tag{5.17}$$

The importance distribution can be conveniently factorised as:

$$\pi(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \propto \pi(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})\pi(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}). \tag{5.18}$$

Using equation (5.18) in the equation (5.17), we get

$$w_t^{(i)} \propto \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} \frac{p(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})}{\pi(\mathbf{x}_{0:t-1}^{(i)}|\mathbf{y}_{1:t-1})} \tag{5.19a}$$
$$= \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(i)})p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{\pi(\mathbf{x}_t^{(i)}|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_{1:t})} w_{t-1}^{(i)} \tag{5.19b}$$

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

Algorithm 3: Particle Filter (Sequential Importance Resampling)

Draw samples from the prior distribution and set all the weights $w_0^{(i)} = 1/N$. Set a threshold $\mathcal{N}$ for resampling.

$$\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0), \qquad i = 1, \ldots, N.$$

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

Sample particles from the importance distribution using previous state particles and all the measurements.

$$\mathbf{x}_t^{(i)} \sim \pi(\mathbf{x}_t | \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}), \qquad i = 1, \ldots, N.$$

Update and normalize the weights using

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

Calculate the state estimate.

Calculate the effective sample size S_{eff}

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

If the effective sample S_{eff} is less than threshold $\mathcal{N}$ perform resampling and set all the weights to $1/N$.

end

5.2.2 Unscented Kalman Filters

Unscented Kalman filter performs unscented transform for performing the filtering task. 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

$$\begin{aligned} \mathbf{x} &\sim N(\mathbf{m}, \mathbf{P}) \\ \mathbf{y} &= g(\mathbf{x}). \end{aligned} \tag{5.20}$$

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 transformed 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 the non-linear transformation. For convenience, the transformed variable is approximated as a Gaussian distribution.

The *unscented Kalman filter* (UKF) or *sigma point filter* is a non-optimal Bayesian filter which utilizes unscented transform for approximating the filtering distribution (Särkkä, 2013). It is a better alternative to *extended Kalman filter* as it is Jacobian free. We aim at evaluating the joint filtering distribution

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \simeq N(\mathbf{x}_t|\mathbf{m}_t, \mathbf{P}_t), \tag{5.21}$$

where $\mathbf{m}_t$ and $\mathbf{P}_t$ are the estimated mean and covariance. We follow the following procedure for deriving the mean and covariances of the filtering distribution.

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

$$\begin{aligned} \mathcal{X}^{(0)} &= \mathbf{m}, \\ \mathcal{X}^{(i)} &= \mathbf{m} + \sqrt{n+\lambda}[\sqrt{\mathbf{P}}]_i, \\ \mathcal{X}^{(i+n)} &= \mathbf{m} - \sqrt{n+\lambda}[\sqrt{\mathbf{P}}]_i, \quad i = 1, \ldots, n \end{aligned} \tag{5.22}$$

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

$$\lambda \triangleq \alpha^2(n + \kappa) - n. \tag{5.23}$$

  α and κ are the user set parameters which dictate the spread of the sigma points around the mean.

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

$$\mathcal{Y}^{(i)} = g(\mathcal{X}^{(i)}), \quad i = 0, \ldots, 2n. \tag{5.24}$$

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

$$\begin{aligned} E[g(\mathbf{x})] \simeq \boldsymbol{\mu}_{ut} &= \sum_{i=0}^{2n} W_i^{(m)} \mathcal{Y}^{(i)}, \\ Cov[g(\mathbf{x})] \simeq \mathbf{S}_{ut} &= \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{Y}^{(i)} - \boldsymbol{\mu}_{ut})(\mathcal{Y}^{(i)} - \boldsymbol{\mu}_{ut})^T, \end{aligned} \tag{5.25}$$

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

$$\begin{aligned} W_0^{(m)} &= \frac{\lambda}{n+\lambda}, \\ W_0^{(c)} &= \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta), \\ W_i^{(m)} &= \frac{1}{2(n+\lambda)}, \quad i = 1, \ldots, 2n, \\ W_i^{(c)} &= \frac{1}{2(n+\lambda)}, \quad i = 1, \ldots, 2n, \end{aligned} \tag{5.26}$$

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

We summarise the algorithm in Algorithm 4. Refer to Chapter 5 of Särkkä (2013) for a comprehensive introduction.

Algorithm 4: Unscented Kalman Filter

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

Prediction:

- Get the sigma points:

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

  where n is state dimension and λ is defined in (5.23).

- Transform the sigma points using the dynamic model:

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

- 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 (5.26)

Update:

- Get the sigma points:

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

3

- Transform the sigma points using the measurement model:

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

- 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 6

Measurement Setup & Data Collection

In this chapter, in Section 6.1, we describe in detail about the experimental area and beacons used for measurements. Next, in Section 6.2, we describe in detail the BLE with its architecture and different channels. Next, in Section 6.3, we state various assumption used in this thesis. Next, in Sections 6.4 and 6.5, we give the details about the radio analyser and the RSSI measurement application. Last, in Sections 6.6 and 6.7, we describe the procedures followed for collecting the data for data analysis experiments and position evaluations.

6.1 Experimental Testbed

Our experiment area is an office space located on the fourth floor at Helvar R&D office (see Figure 6.1). The area is a vacant space with luminaires with beacons embedded. The stem part of L-shaped floored space has the dimensions 32×5.5 metres ($length \times width$), and the leg part is 10×7 metres in dimension. For simplicity and accurate results, the movement of personnel during the experiments was avoided. The low right corner in Figure 6.1 is considered the origin and the locations mentioned in this thesis are with respect to this.

The beacons in the testbed are inside the luminaires. The luminaires are at the height of 1.37 metres 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 cannot be estimated inside the luminaire. As shown in Figure 6.1, there are 28 beacons, and the measurements are recorded from all the beacons for the data analysis experiments and for the positioning evaluation in this thesis. In addition to the Helvar beacons, we also record measurements from the Kontakt beacons (`https://store.kontakt.io/`) for data analysis.

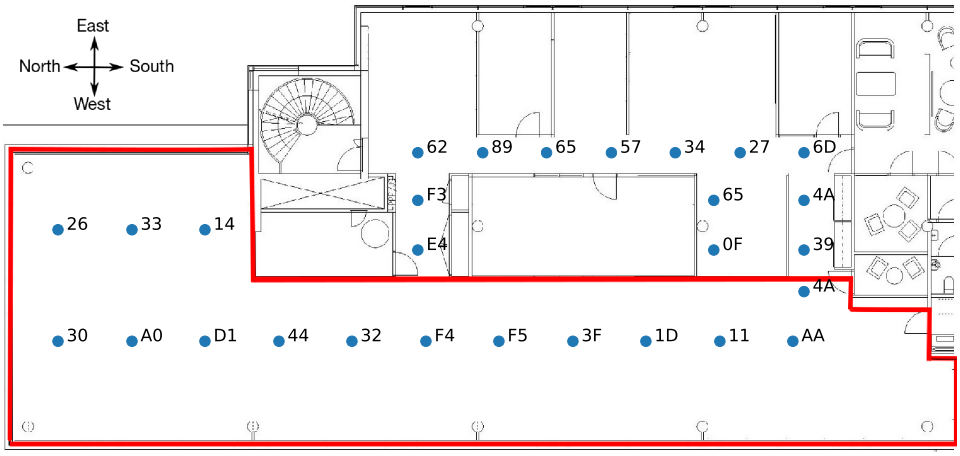


Figure 6.1: Floor plan of Helvar's R&D section in Keilaniemi office with the location of the beacons.

6.2 Bluetooth Low Energy

Bluetooth Low Energy (BLE) is a 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 a 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 the context of positioning techniques, it does not require any connection between devices and only a mobile unit (MU) is required to read the beacon signal packets from the *Wireless Access Points (WAPs)* at regular intervals. The BLE has a physical layer with advertisement channels and data channels.

BLE has 40 physical radio channels with each radio channel spaced out of 2 MHz in between them (refer to Figure 6.2). The BLE has the data rate of 1 Mbps and like the *classic Bluetooth*, uses the *Gaussian Frequency Shift Keying* modulation. However, both the technologies have different spacing of radio channels, hence are incompatible and therefore, can not communicate.

As mentioned before, the physical channels are categorised based on the type of data they transmit, and we describe them next.

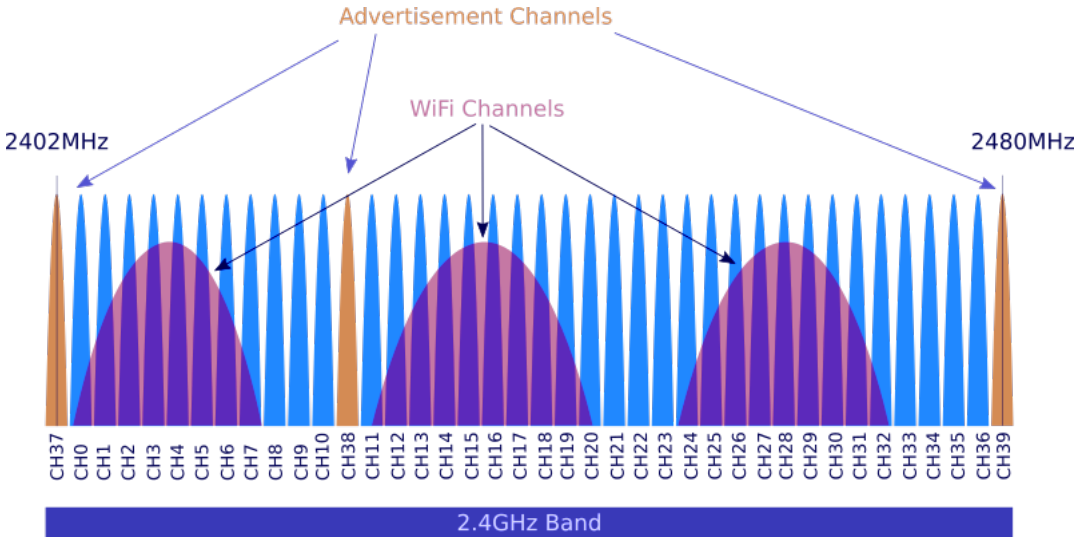


Figure 6.2: Channels in Bluetooth low energy (Argenox, 2015).

6.2.1 Advertisement channels

BLE beacons are used to advertise and communicate the data. The advertisement and data transmission window is 1 second. The beacons advertise at the start of the second which is approximately 1 ms. The channels 37, 38 and 39 are the advertisement channels. These radio channels are strategically placed to avoid interference from the WiFi (refer to Figure 6.2). The advertisement channels are critical as they are responsible for making a connection to an MU (like smartphones, smartwatches, and so on) and three channels are used to increase the probability of central device reading the advertisement packets. This mode serves for uni-directional communication and for the positioning applications we do not require any connection between devices. In case of access to the hardware, the BLE technology provides an option for masking the advertisement channels.

During the advertisement interval, the three advertisement channels transmit the packets sequentially in under 1 ms then followed by sleep period (refer to Figure 6.3). The sleep period consists of a fixed interval and a pseudo-random delay. Based on the target application, the fixed interval generally ranges from 20 ms to 10.24 seconds and the pseudo-random delay from 0 ms to 10 ms. The pseudo-random delay helps to avoid central device missing the beacon advertisement packets. The overlap might happen in the case of overlap of advertisement interval and central device's (or MU's) *scanning interval*. In this phase, the device is in sleep mode most of the time and transmission power in between -20 dBm to 10 dBm, which guarantees low

power consumption. For more technical details refer to Lindh (2015).

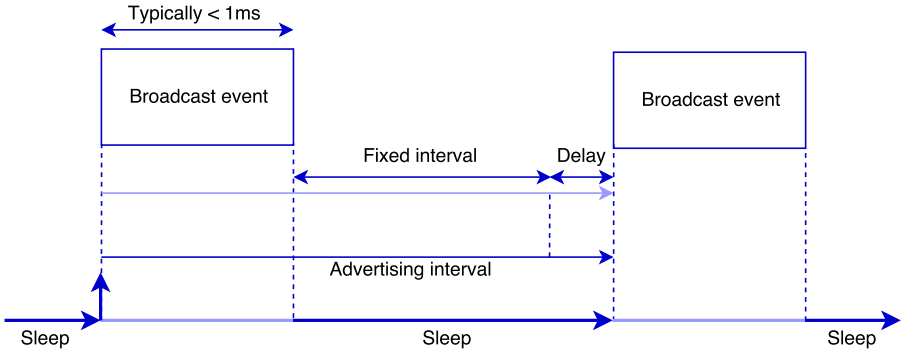


Figure 6.3: Advertisement phase in BLE.

6.2.2 Data Channels

The rest of the radio channels, that is, channel 0 to channel 36, are dedicated data channels. These channels are in use once a device is discovered and connection with a master device is established. The data transmission phase is a bi-directional mode as shown in Figure 6.4 where we can see two-way communication between master and slave. We do not use this mode for the positioning applications.

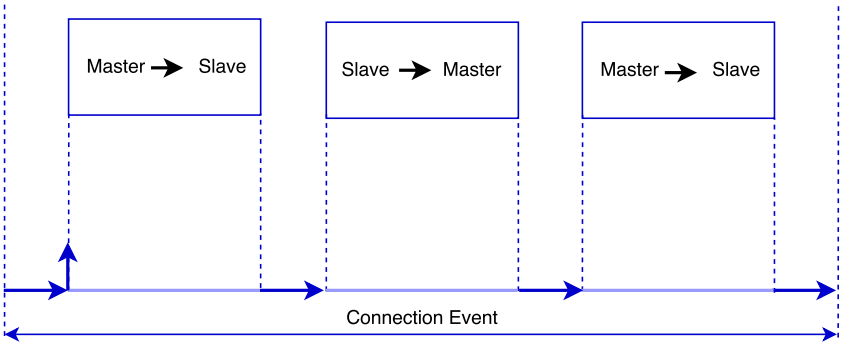


Figure 6.4: Data communication in BLE

6.2.3 Different BLE protocols

The BLE beacons are configured into two types of communication protocols. The communication protocols have a *universally unique identifier* which are used for reading the signal packets in the smartphone application.

- **iBeacon**: The iBeacon (Newman, 2014) is a 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.

6.3 Assumptions on the RSSI values

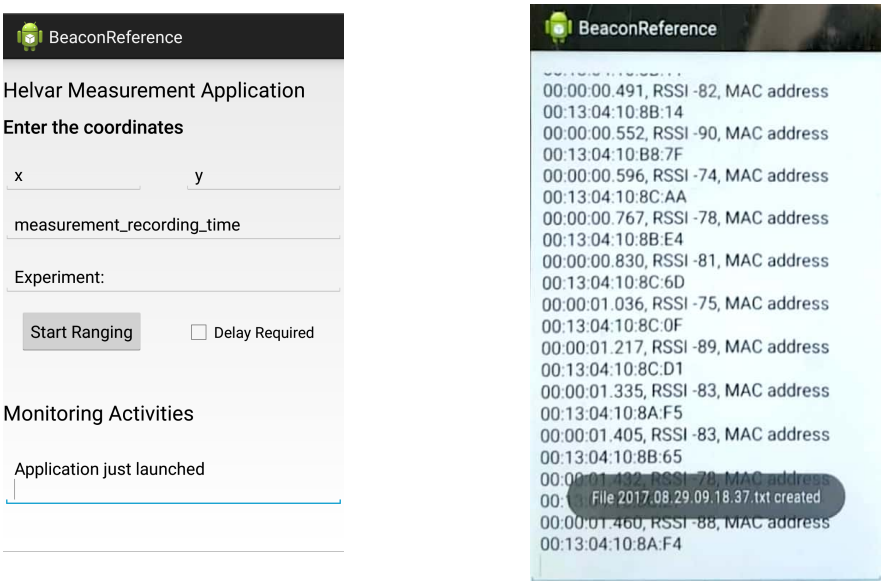
RSSI is designed for wireless communication and designing receiver antenna but not for positioning applications. The RSSI values are quantised, that is, the MU outputs the values in the steps of 1 dBm, so in theory, the quantised value of RSSI represents an area rather than a point (Kaemarungsi and Krishnamurthy, 2012), and for data analysis, we consider the data distribution to be continuous.

We assume that in-luminaire BLE beacons are omnidirectional with the luminaires always located over the head of the user. The variation of RSSI in case of a change in the location of the beacon inside the luminaires was assumed negligible.

The interference of in-luminaire working and other radio signal is considered negligible. The bias due to the direction of the phones when taking user-free measurements was also assumed negligible. We study the caveats of RSSI for positioning thoroughly in Chapter 7.

6.4 Radio Analyzer

In addition to the smartphones, we use the Frontline Sodera LE (Teledyne Lecroy, 2016) radio analyser for recording the RSSI values. The radio analyser accurately reads the signal strength values and also provides additional information like physical channel, radio channel, universally unique identifier and data type. A comprehensive discussion on the various ways to measure RSSI is discussed in Kaemarungsi and Krishnamurthy (2012).



(a) Main screen

(b) Measurement screen

Figure 6.5: The screnshots of the measurement application used in the thesis.

6.5 Measurement Application

The RSSI measurement application was built using the AltBeacon Android library (Networks, 2017). This library incorporates extra features like application foreground and background signal scan interval. The background scan interval is useful for tracking when the MU's screen is turned off. As the in-luminaire beacons did not follow either the iBeacon (Newman, 2014) or Eddystone (Eddystone, 2016) protocol, we used this library for reading the radio signals from a non-standard advertising beacon.

The application allows for entering in the current measurement location, i.e., x and y coordinates, measurement recording time, and an experiment description. It also provides an option to delay recording the measurements in the case of user-free measurements. The application buzzes after the completion of measurement time. As shown in Figure 6.5, the application displays relative time from the start, the beacon's media access control (MAC) address and RSSI value while recording the measurements. The same data is logged in a text file with the addition of absolute time-stamp for each measurement. The file is saved with the absolute time-stamp when opened as the file name.

(a) (b)

Figure 6.6: (a) Fingerprinting setup at Helvar R&D. (b) Tripod with mobile holder for user free measurements.

6.6 Data for Data Analysis Experiments

For gaining better insight about the data generating process, we collect data from different smartphones and different human orientations. This was done to construct the measurement model in the Bayesian filtering context better. We discuss the different user orientations in the following section. We selected two locations for conducting the following experiments – location 1 at (2.77 metres, 9.78 metres) and location 2 at (8.54 metres, 25.24 metres), which we, from now on will refer to as L1 and L2 respectively.

6.6.1 Different Smartphones

All the measurements were recorded with a smartphone, unlike the other studies where the experiments were conducted using a laptop (Kaemarungsi and Krishnamurthy (2004); Bahl and Padmanabhan (2000)). The smartphones used during the project were Samsung S7, Samsung S4 and Samsung S4 mini. Due to the instability of the measurement application, the Sony and LG devices were not used. The reason is beyond the scope of this thesis.

6.6.2 Different User Orientations

The main idea here is to cover different user directions while recording the measurements from the beacons. In this thesis, we record the user mea-

surements in all the cardinal directions north, east, west and south (Richter and Toledano-Ayala (2015); for directions refer to Figure 6.1). For the user measurements, the user held the smartphone device at a 45° angle, with the head tilted forward making it parallel to the screen of the phone (Hansraj, 2014). We refer to this as *standard usage configuration*. By observation, this is the most common way of usage when a user is walking. With the same configuration, we record the shadow and the rotation measurements.

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

6.6.3 Different Phone Orientations

In addition to the different user orientations, we also experimented with different phone orientation, as described in Hansraj (2014). The different configuration of phone angles used were 0°, 45° and 90° with respect to the floor. The signals strengths recorded in this configurations were user free. For recording the user-free measurements, the smartphone device was clamped and mounted on the tripod as shown in Figure 6.6. Additionally, we took the with user measurements with a smartphone in the side pocket of the trousers for juxtaposing against the standard usage configuration.

6.6.4 Outdoor Measurements

To avoid the possible attenuation of the signal due to reflection from walls indoors, we recorded measurements in an outdoor setting. At a fairly empty car parking area behind Jämeräntaival 1, Espoo, we arranged the setup as seen in Figure 6.7. We mounted a luminaire with beacon on a tripod at the height of 0.6 metres and connected it to a power source. At the same height and 1 metre away, we placed the radio analyser (refer to Section 6.4) on polystyrene thermocol boxes, and a tripod clamped smartphone. The radio analyser was powered up by connecting it to a power source and was later connected to a laptop for running the radio analyser's measurement application. The radio analyser's and smartphone's respective applications were set up to record measurements simultaneously. In case of a mismatch, the absolute time-stamps were used to sync the measurements for further data analysis.

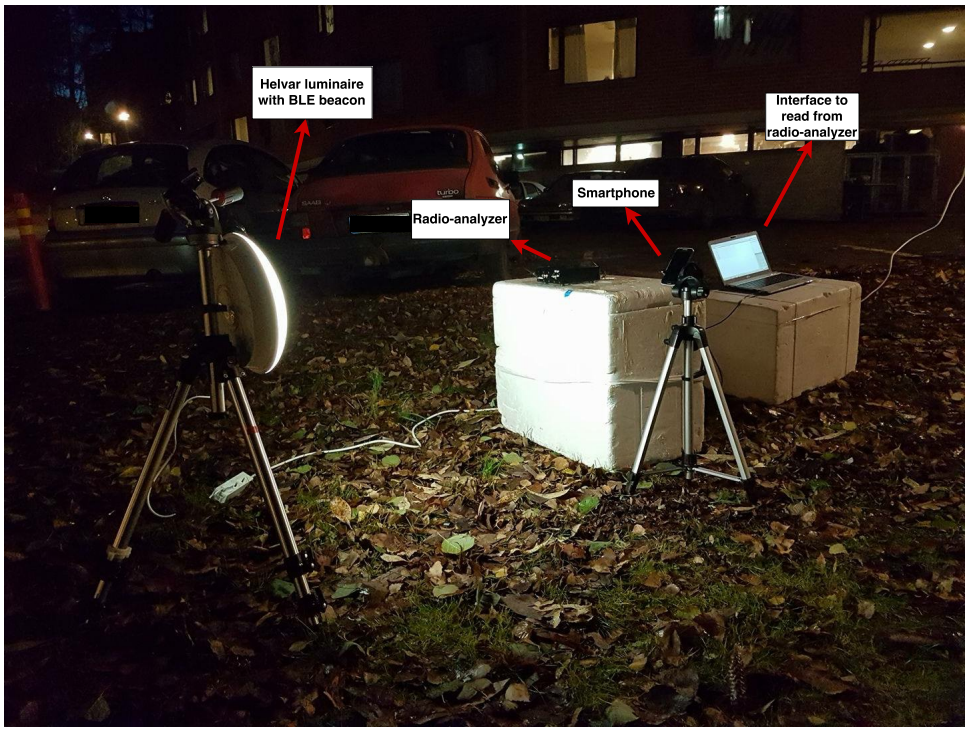


Figure 6.7: Measuring the RSSI values outdoors using the smartphone and radio analyzer.

6.7 Data for Evaluation of Positioning Algorithms

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

6.7.1 Setup for Calibration Phase and Test Phase

As said, the calibration phase involves the fingerprinting process. In this phase, we identify particular locations in the experiment test bed and collect the signal measurements. As shown in Figure 6.6, we landmark the locations with respect to the origin as described in Section 6.1. In this thesis, we selected a total of 63 calibration points and recorded measurements for 50 seconds.

For the test phase, we landmark the exact locations on the floor, and we recorded the measurements by following the track (see Figure 6.8) using the smartphone application. The measurement application was used for time flagging whenever the landmarks were reached for recording the ground truth

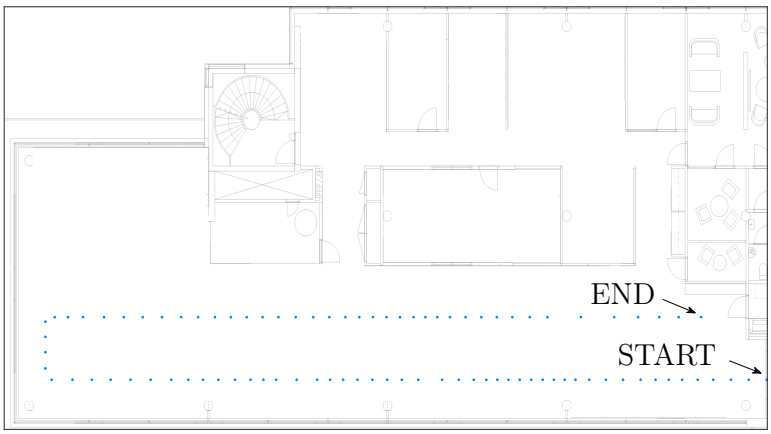


Figure 6.8: Track for collecting of test data from Helvar R&D arena.

timestamp and later use it for getting the accuracy.

6.7.2 Calibration Data

The calibration data is used for generating the reference table (both $\mathcal{R}_1$ and $\mathcal{R}_2$) which is in turn used for creating the radiomaps. The radiomaps are used as the measurement model in the filtering process as described in Chapter 5. For collecting the data, the user measures the signal strength values at a particular calibration point at a random direction. This was done keeping in mind the final indoor positioning product, where the user is not be forced to choose a direction but rather take measurements as is.

6.7.3 Test data

Obtaining the data was done using the mobile unit (MU) walking at a constant speed in the test setup. The measurements were collected using the smartphone device Samsung S7. The data was ported to MATLAB where it was converted to an appropriate data structure for evaluating the filtering methods. The test device Samsung S7 was the same device as was used for fingerprinting and generating the radiomaps. The track where the test data was collected can be seen in Figure 6.8.

Chapter 7

Data Analysis of RSSI

You attract the right things when you have a sense of who you are.
- Amy Poehler

Before we start evaluating the methods described in Chapter 5 for the indoor positioning problem, the initial challenge lies in getting the right measurement (data) model as the RSSI varies due to factors like signal attenuation due to obstacles, human beings, signal interference, type of smartphone hardware, orientation of the phone, algorithmic related factors and other factors like malfunctioning BLE modules. The challenges have been discussed in detail in Chapter 2.

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 by studying the statistical properties of the RSSI values. This knowledge about the data generating process and RSSI's can help in the efficient modelling of the measurement model which could lead to a better indoor positioning system (IPS).

The *initial data analysis*[1] is the primary task in statistical analysis and modelling which yields critical statistical properties about data generating process. It veers us towards finding the right solution, that is, finding the right data model for our problem and learn the peculiarities in the data. It helps to design and analyse an IPS efficiently. A comprehensive data analysis of RSSI data from the beacons has not been done (for WiFi, check Kae-

[1]not to be confused with exploratory data analysis. Check Chatfield (2006) for more details.

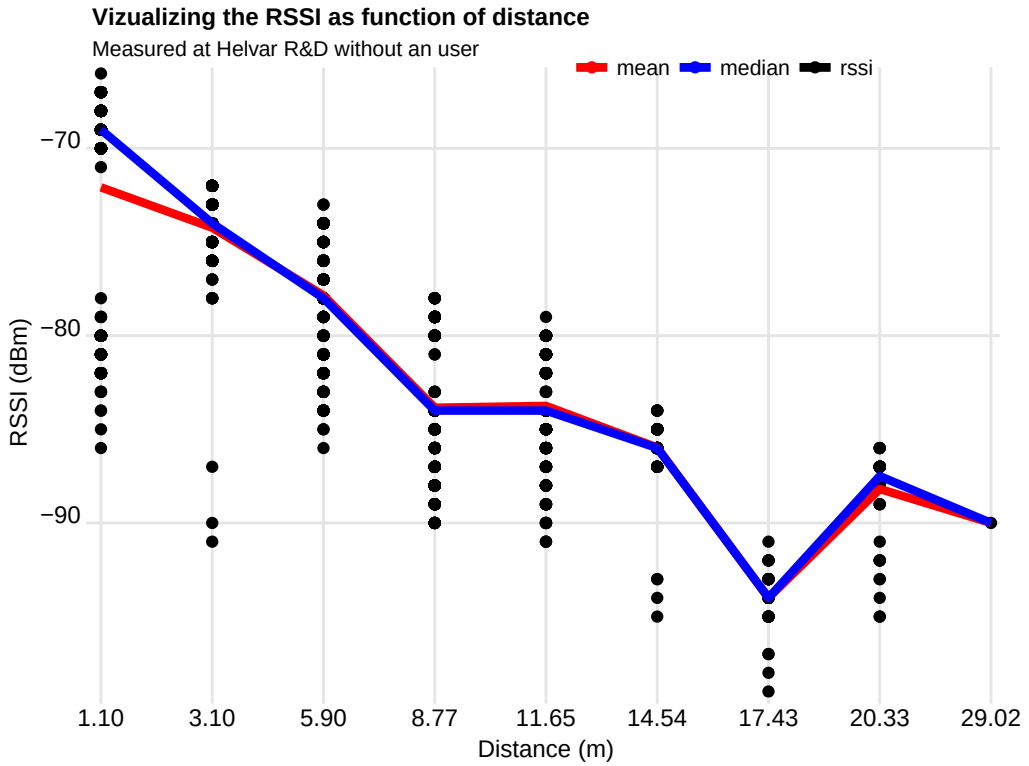


Figure 7.1: Illustration of received signal strength indication value decrease with increase in the distance to the luminaire.

marungsi and Krishnamurthy (2012)) in previous literature with most of the researchers just focusing on the algorithms. Based on our review, there is no in-depth analysis of statistical properties of RSSI values of the BLE signals, so in this thesis, we take a closer look at the RSSI and investigate different factors which affect its variation. Due to the paucity of time, the analysis of the long-term behaviour of RSSI values is not studied, and it was taken care that all the measurements were recorded in close successions in time. We take into consideration the factors such as the user's presence, smartphones and orientation of smart-phones and material of the luminaire. First, in Section 7.1, we show that the RSSI data is a suitable distance criterion. Next, in Section 7.2, we summarise the results for different experiments conducted to understand the factors biasing the RSSI values. Last, in Section 7.3, we go deeper in understanding the RSSI with the perspective of BLE's hardware architecture.

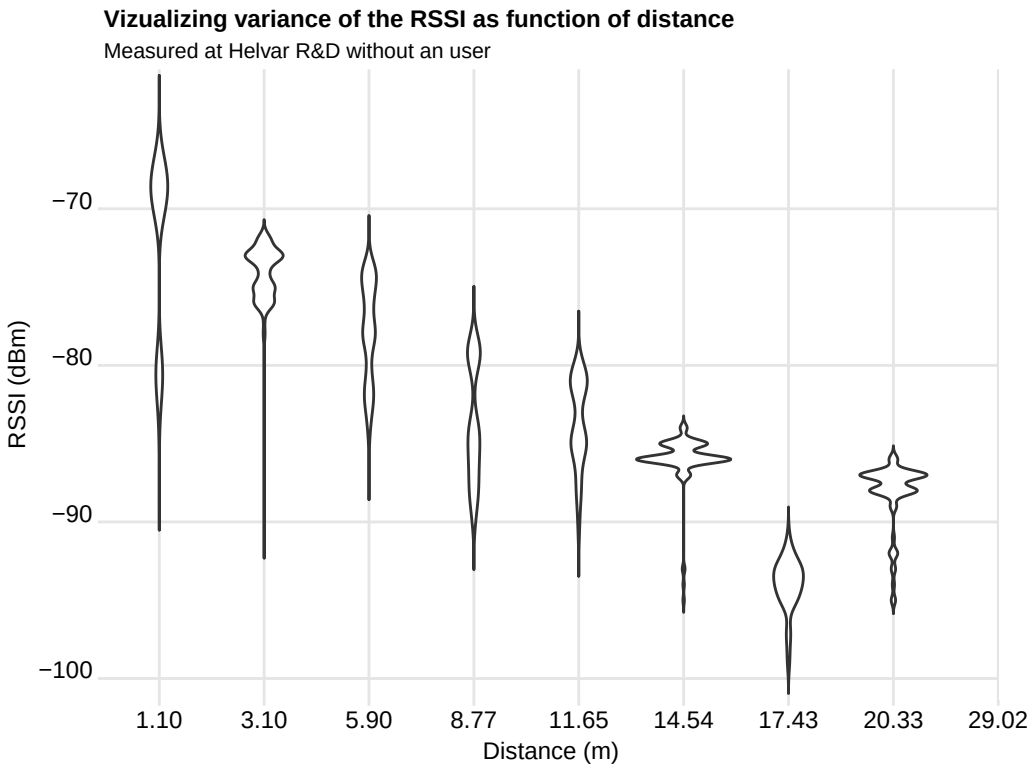


Figure 7.2: Illustration of variance of signal strength with the luminaire at increasing distance.

7.1 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 7.1, we show that the RSSI is a realistic distance criterion. It conforms to the common knowledge that the signal strength for the beacon closest to the mobile unit has the highest signal strength and it reduces for the beacon is farther away. The experiment included the tripod clamped mobile unit (MU) which recorded the signal strength for a particular period of time. Hence, it was a user-free experiment to avoid additional biases to the RSSI values. The measurements were recorded with the Samsung S7 smartphone device.

Additionally, we observe from Figure 7.2 an interesting phenomenon, the variance of the RSSI values is more for the luminaires close-by when compared

to the luminaires far away. For instance, for the luminaire at 1.10 metres distance, we see the variance is over 20 dBm. We discuss more on the phenomenon in Section 7.3.

7.2 Experiments

The main aim of these experiments was to quantify the bias of different factors in the RSSI; hence, it was taken care to change only a single factor keeping the other factors unchanged. Measurements were taken with one phone at a time to avoid any unknown interference. We discuss the results for the statistical hypothesis testing methods discussed in Chapter **??**.

7.2.1 Test for Normality

In this experiment, we investigated the test for normality for the RSSI values. We used the smartphones Samsung S4, Samsung S4 mini, Samsung S7 and the radio analyser. The avoid additional factor adding bias to the RSSI, we decided to take the user-free measurements. The tripod and clamp (see Figure 6.6) with a random direction were set up. The direction was kept consistent for the measurements with different devices.

The Table 7.1 summarises the results for the test. We considered seven beacons with increasing distance for the experiment. The results show low p-values consistently for the beacons 8B3F and 8AF5 for all the devices, we can reject the null hypothesis as we have no evidence to assert that the sample comes from a population which has normal distribution. However, we see contrasting results for the beacons farther away (8AF4 – 8CA0) from the MU. Similar phenomenon can be observed in Figure Figure 7.3. Overall, the test was inconclusive.

Table 7.1: Summary of normality tests for different smartphones and radio analyzer (p-values).

Access Point	S4	S4 mini	S7	Radio analyzer
8B3F	8.62E-09	3.29E-05	3.59E-05	0
8AF5	7.15E-13	5.71E-07	9.64E-06	0
8AF4	1.39E-05	3.09E-06	0.5195	0
8B32	1.35E-05	0.1653	0.0377	0
8B44	5.82E-04	NA	0.0797	0
8CD1	0.0021	0.0013	NA	0.0115
8CA0	0.0117	NA	NA	3.96E-14

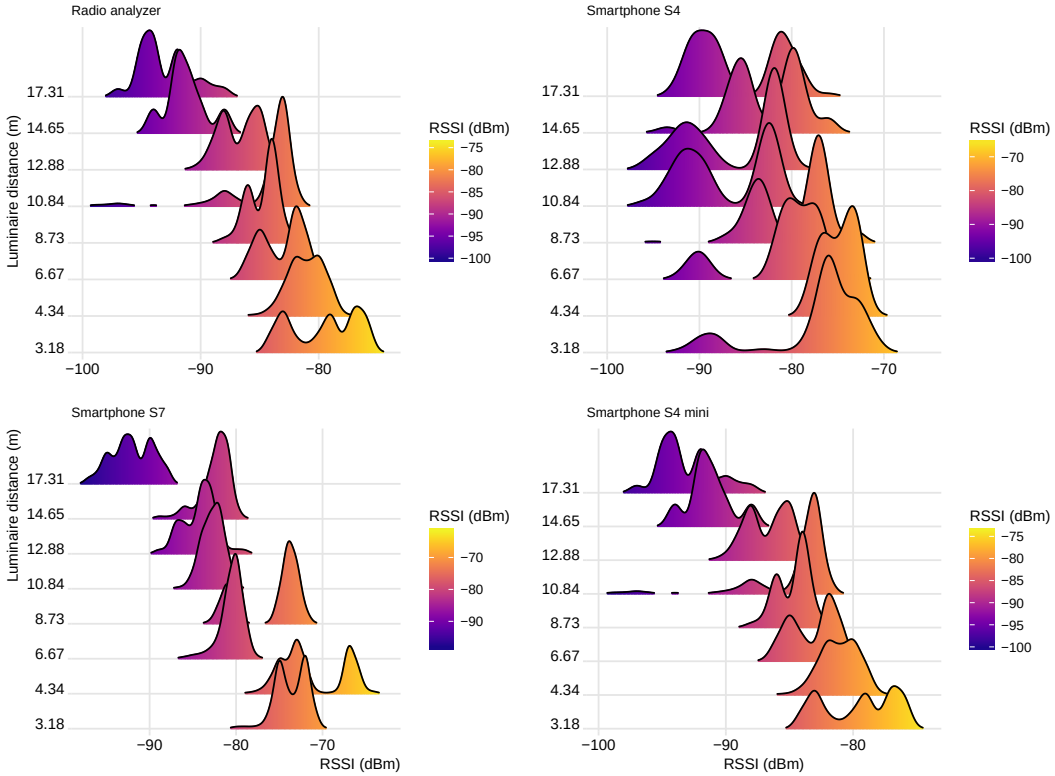


Figure 7.3: Illustration of the density of RSSI with increase in the distance.

7.2.2 Bias due to User's Presence

In this experiment, we investigated the effect of the user's presence on the recorded RSSI values. We used the smartphone Samsung S7 for the experiments. It was assumed that the direction of the phone has a minimal bias on the RSSI value when taking user-free measurements. Hence, a random direction was chosen. In addition, we also checked 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 (see the Table 7.2 and 7.3), the null hypothesis was rejected for most of the experiments including shadowing and rotation experiments and we had no evidence to say that the RSSI recorded with and without user have the same median. We can say the bias due

Table 7.2: Summary of user presence bias to the RSSI measurements for L1.

Without User vs User facing north direction				
Access Point	w/o user RSSI (count)	User RSSI (count)	MAD	Wilcoxon test
8B3F	-77.1349 (645)	-85.8808 (453)	8.7459	1.50E-103
8AF5	-74.8347 (629)	-83.7002 (487)	8.8655	3.58E-175
8AF4	-80.8054 (596)	-90.7882 (321)	9.9828	3.53E-142
8B32	-79.9181 (415)	-91.9278 (291)	12.0098	1.65E-108
8B44	-87.1852 (351)	-90.8154 (298)	3.6303	2.88E-20
8CD1	-86.3827 (405)	-92.1137 (387)	5.731	8.15E-48
8CA0	-82.6421 (461)	-93.0089 (224)	10.3668	2.81E-94
Mean	*-81.2719*	*-89.7479*	*8.476*	
Without User vs User facing east direction				
Access Point	w/o user RSSI (count)	User east (count)	MAD	Wilcoxon test
8B3F	-77.1349 (645)	-90.4377 (345)	13.3028	2.02E-118
8AF5	-74.8347 (629)	-87.2577 (388)	12.4231	6.92E-160
8AF4	-80.8054 (596)	-92.6728 (162)	11.8675	1.26E-87
8B32	-79.9181 (415)	-92.8242 (165)	12.9062	8.85E-77
8B44	-87.1852 (351)	-93.7024 (84)	6.5172	2.96E-31
8CD1	-86.3827 (405)	-95.0484 (62)	8.6657	1.18E-30
8CA0	-82.6421 (461)	-96 (4)	13.3579	5.28E-04
Mean	*-81.2719*	*-92.5633*	*11.2914*	
Without User vs User facing south direction				
Access Point	w/o user RSSI (count)	User south (count)	MAD	Wilcoxon test
8B3F	-77.1349 (645)	-90.4205 (440)	13.2856	6.97E-137
8AF5	-74.8347 (629)	-88.0415 (434)	13.2068	1.02E-170
8AF4	-80.8054 (596)	-92.8765 (170)	12.0711	6.51E-91
8B32	-79.9181 (415)	-93.6351 (222)	13.7171	1.12E-93
8B44	-87.1852 (351)	-94.4821 (56)	7.297	1.83E-25
8CD1	-86.3827 (405)	-95.9697 (33)	9.587	3.23E-20
8CA0	-82.6421 (461)	-95 (2)	12.3579	1.40E-02
Mean	*-81.2719*	*-92.9179*	*11.646*	
Without User vs User facing west direction				
Access Point	w/o user RSSI (count)	User west (count)	MAD	Wilcoxon test
8B3F	-77.1349 (645)	-84.6492 (553)	7.5143	8.46E-109
8AF5	-74.8347 (629)	-84.1693 (561)	9.3347	9.40E-185
8AF4	-80.8054 (596)	-87.7208 (480)	6.9155	6.97E-181
8B32	-79.9181 (415)	-90.7911 (426)	10.873	2.64E-129
8B44	-87.1852 (351)	-91.6395 (319)	4.4543	1.88E-29
8CD1	-86.3827 (405)	-93.1667 (252)	6.784	9.79E-51
8CA0	-82.6421 (461)	-92.4114 (299)	9.7693	1.49E-91
Mean	*-81.2719*	*-89.2211*	*7.9492*	
Without User vs User facing shadowing direction				
Access Point	w/o user RSSI (count)	User shadow (count)	MAD	Wilcoxon test
8B1D	-70.8425 (546)	-71.1298 (524)	0.2873	**0.2479**
8B11	-76.4113 (586)	-83.3536 (509)	6.9424	3.60E-149
8CAA	-72.7231 (585)	-88.1457 (405)	15.4226	7.55E-160
8C6A	-79.1505 (651)	-85.2004 (489)	6.0499	1.15E-49
8C39	-78.0507 (670)	-88.5194 (387)	10.4686	3.68E-148
8C4A	-85.2737 (464)	-90.7217 (357)	5.448	9.97E-64
8C6D	-77.581 (568)	-90.1753 (405)	12.5943	1.11E-121
Mean	*-77.1475*	*-85.3208*	*8.1733*	
Without User vs User with rotation				
Access Point	w/o user RSSI (count)	User rotate (count)	MAD	Wilcoxon test
8B3F	-77.1349 (645)	-88.8276 (203)	11.6927	3.81E-74
8AF5	-74.8347 (629)	-85.8068 (208)	10.9721	6.67E-104
8AF4	-80.8054 (596)	-91.1449 (138)	10.3396	2.38E-77
8B32	-79.9181 (415)	-91.7045 (132)	11.7865	3.72E-65
8B44	-87.1852 (351)	-92.5375 (80)	5.3523	1.33E-19
8CD1	-86.3827 (405)	-93.3084 (107)	6.9257	4.34E-31
8CA0	-82.6421 (461)	-92.6267 (75)	9.9846	5.67E-31
Mean	*-81.2719*	*-90.8509*	*9.579*	

Table 7.3: 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	Wilcoxon test
8B14	-74.034 (235)	-73.6262 (650)	-0.4079	3.47E-06
8B33	-87.1358 (243)	-83.7008 (625)	-3.435	6.05E-25
8B26	-86.5161 (124)	-89.9404 (554)	3.4243	1.86E-26
8B30	-86.0913 (252)	-87.6348 (564)	1.5435	7.78E-08
8CA0	-90.1701 (241)	-91.7809 (324)	1.6107	5.22E-22
8CD1	-87.8323 (155)	-90.9436 (461)	3.1113	9.53E-29
8B44	-93.4215 (121)	-91.6318 (402)	-1.7896	2.54E-12
Mean	*-86.4573*	*-87.0369*	*0.5796*	
Without User vs User facing east direction				
Access Point	w/o user RSSI (count)	User east (count)	MAD	Wilcoxon test
8B14	-74.034 (235)	-77.6172 (674)	3.5832	1.75E-26
8B33	-87.1358 (243)	-88.5488 (605)	1.413	5.91E-08
8B26	-86.5161 (124)	-91.4196 (491)	4.9034	8.91E-34
8B30	-86.0913 (252)	-89.7476 (519)	3.6563	1.22E-44
8CA0	-90.1701 (241)	-92.7592 (245)	2.5891	8.85E-42
8CD1	-87.8323 (155)	-92.8976 (332)	5.0653	9.28E-43
8B44	-93.4215 (121)	-91.6407 (334)	-1.7808	1.76E-11
Mean	*-86.4573*	*-89.2329*	*2.7756*	
Without User vs User facing south direction				
Access Point	w/o user RSSI (count)	User south (count)	MAD	Wilcoxon test
8B14	-74.034 (235)	-76.5601 (666)	2.526	1.15E-12
8B33	-87.1358 (243)	-86.2746 (590)	-0.8612	0.0065
8B26	-86.5161 (124)	-92.1648 (449)	5.6487	2.56E-33
8B30	-86.0913 (252)	-86.9623 (610)	0.871	0.013
8CA0	-90.1701 (241)	-90.7882 (439)	0.618	1.37E-04
8CD1	-87.8323 (155)	-90.9836 (488)	3.1513	3.41E-31
8B44	-93.4215 (121)	-91.3266 (346)	-2.0949	1.98E-14
Mean	*-86.4573*	*-87.8657*	*1.4084*	
Without User vs User facing west direction				
Access Point	w/o user RSSI (count)	User west (count)	MAD	Wilcoxon test
8B14	-74.034 (235)	-76.3689 (656)	2.3349	6.42E-14
8B33	-87.1358 (243)	-86.6918 (623)	-0.444	8.38E-04
8B26	-86.5161 (124)	-90.5759 (514)	4.0597	1.65E-30
8B30	-86.0913 (252)	-87.1509 (603)	1.0596	5.60E-04
8CA0	-90.1701 (241)	-89.2592 (517)	-0.9109	6.43E-05
8CD1	-87.8323 (155)	-92.1986 (423)	4.3663	3.40E-41
8B44	-93.4215 (121)	-91.6941 (340)	-1.7274	**7.49E-13**
Mean	*-86.4573*	*-87.7056*	*1.2483*	
Without User vs User facing shadowing direction				
Access Point	w/o user RSSI (count)	User shadow (count)	MAD	Wilcoxon test
8B14	-74.034 (235)	-79.8323 (650)	5.7983	2.78E-65
8B33	-87.1358 (243)	-88.931 (536)	1.7952	2.76E-11
8B26	-86.5161 (124)	-92.9302 (344)	6.4141	2.84E-34
8B30	-86.0913 (252)	-91.508 (374)	5.4168	7.96E-78
8CA0	-90.1701 (241)	-93.25 (172)	3.0799	2.53E-47
8CD1	-87.8323 (155)	-89.9053 (433)	2.0731	3.55E-17
8B44	-93.4215 (121)	-91.9349 (215)	-1.4866	1.62E-08
Mean	*-86.4573*	*-89.756*	*3.2987*	
Without User vs User with rotation				
Access Point	w/o user RSSI (count)	User rotate (count)	MAD	Wilcoxon test
8B14	-74.034 (235)	-77.9962 (261)	3.9621	9.13E-21
8B33	-87.1358 (243)	-86.3686 (236)	0.7672	1.84E-02
8B26	-86.5161 (124)	-90.3661 (183)	3.85	1.78E-19
8B30	-86.0913 (252)	-89.0408 (196)	2.9495	5.38E-17
8CA0	-90.1701 (241)	-91.2 (130)	1.0299	9.03E-05
8CD1	-87.8323 (155)	-90.6932 (176)	2.8609	5.10E-18
8B44	-93.4215 (121)	-91.1071 (140)	-2.3143	3.75E-11
Mean	*-86.4573*	*-88.1103*	*1.653*	

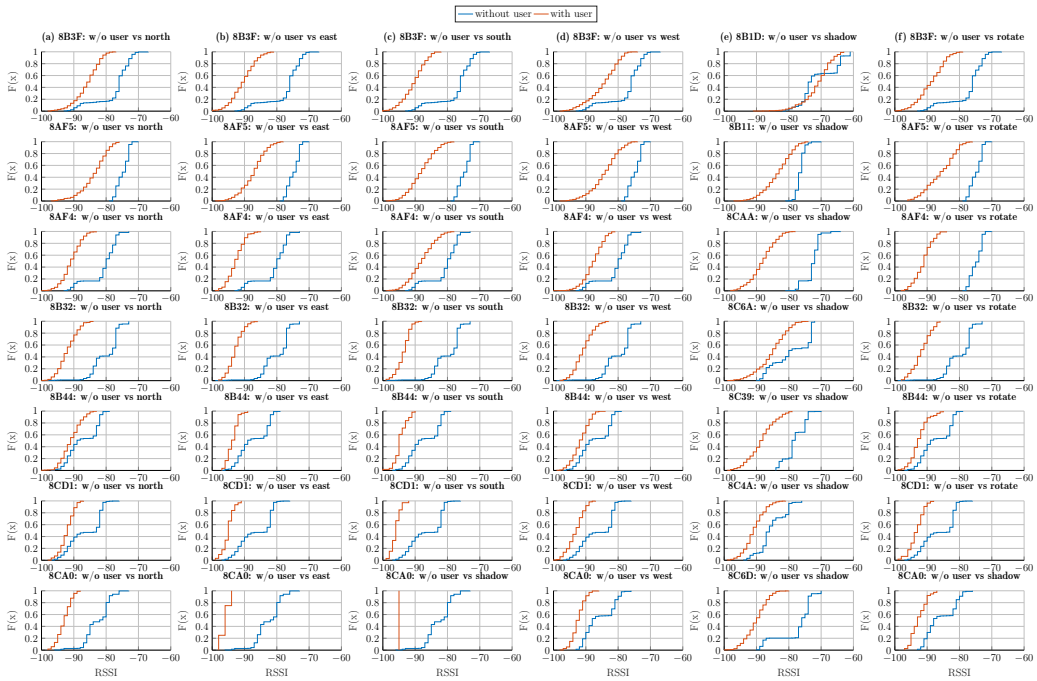


Figure 7.4: Illustration of bias of RSSI values for with and without user scenario for location 1 (L1). The cumulative distribution function for the recorded RSSI shows a higher negative bias when recorded with a user.

to user presence can have a maximum mean average deviation (MAD) of 11.646 dBm. We can see the bias in Figures 7.4 and 7.5.

On a granular level, the observations are:

- **Shadowing**: As mentioned above, the shadowing experiment is performed to quantify the decrease in signal strength due to the presence of the human body in between the mobile device and BLE beacon. It can be seen from the Tables 7.2 and 7.3 that the MAD for the shadowing experiment is not significantly different from the other experiments.

- **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 7.2 and 7.3, the results suggest that rotation does not significantly improve the RSSI measurement in terms of signal strength or signal measurement count. This can be observed from Figures 7.4 and 7.5

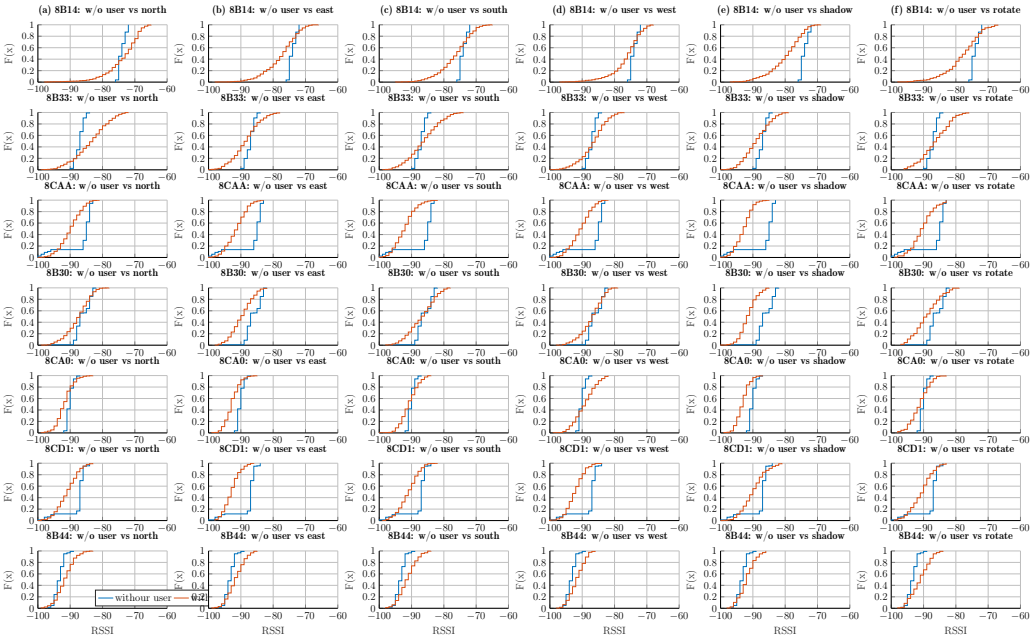


Figure 7.5: Illustration of bias of RSSI values for with and without user scenario for location 2 (L2). In contrast to experiment at L1, the cumulative distribution function for the recorded RSSI shows less significant negative bias when recorded with a user.

- **Increased variance**: From Figures 7.4 and 7.5, additionally we observe increased variance for user measurements at both L1 and L2 locations.

7.2.3 Comparison of Smartphones

In this experiment, we investigated the RSSI bias due to the difference in the hardware of the smartphones. This is a user-free experiment where we recorded the measurements using the tripod with smartphone clamped to it using the standard usage configuration (refer to Chapter 6). A random direction was chosen, and it was kept consistent for different smartphone measurements. We used three smartphones, namely, Samsung S4, Samsung S4 mini and Google Nexus 5 in this experiment.

Overall for locations L1 and L2 (see the Table 7.4 and 7.5), the null hypothesis was rejected for most of the experiments, and we had no evidence to say that the RSSI recorded for different smartphones have the same median. We infer that difference in hardware can cause a maximum bias of 4.63 dBm. The bias in the RSSI distributions can be seen from Figures 7.6 and 7.7.

Table 7.4: 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	Wilcoxon test
8B3F	-77.1349 (645)	-79.5686 (503)	2.4337	2.42E-55
8AF5	-74.8347 (629)	-81.0689 (450)	6.2342	2.43E-174
8AF4	-80.8054 (596)	-83.1289 (225)	2.3235	9.94E-40
8B32	-79.9181 (415)	-84.7287 (188)	4.8107	7.12E-47
8B44	-87.1852 (351)	-84.4841 (157)	2.7011	1.11E-04
8CD1	-86.3827 (405)	-86.6853 (197)	0.3026	0.0653
8CA0	-82.6421 (461)	-91.25 (8)	8.6079	3.41E-06
Mean	*-81.2719*	*-84.4163*	*3.1444*	
Smartphone: S4 vs Nexus 5				
Access Point	S4 RSSI (count)	Nexus 5 (count)	MAD	Wilcoxon test
8B3F	-77.1349 (645)	-77.7444 (90)	0.6096	0.3761
8AF5	-74.8347 (629)	-76.1156 (467)	1.281	9.87E-19
8AF4	-80.8054 (596)	-83.7955 (44)	2.9901	2.07E-09
8B32	-79.9181 (415)	-83.75 (20)	3.8319	3.32E-05
8B44	-87.1852 (351)	-84.5 (10)	2.6852	**0.194**
8CD1	-86.3827 (405)	-79.7979 (94)	6.5848	4.24E-42
8CA0	-82.6421 (461)	-81.1889 (90)	1.4532	0.0163
Mean	*-81.2719*	*-80.9846*	*0.2873*	
Smartphone Orientation: S4 mini vs Nexus 5				
Access Point	S4 mini (count)	Nexus 5 (count)	MAD	Wilcoxon test
8B3F	-76.176 (375)	-77.7444 (90)	1.8241	3.21E-06
8AF5	-83.8701 (154)	-76.1156 (467)	4.9533	1.89E-136
8AF4	-84.5294 (102)	-83.7955 (44)	0.6666	**0.0535**
8B32	-84.6071 (84)	-83.75 (20)	0.9787	0.0025
8B44	-87.303 (33)	-84.5 (10)	0.0159	**0.1428**
8CD1	-87.359 (78)	-79.7979 (94)	6.8874	8.32E-42
8CA0	-86.2576 (66)	-81.1889 (90)	10.0611	3.99E-06
Mean	*-84.3003*	*-80.9846*	*3.6267*	

Table 7.5: 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	Wilcoxon test
8B14	-67.7138 (622)	-67.5916 (404)	0.1222	**0.1851**
8B33	-77.5363 (606)	-76.5158 (380)	1.0205	7.22E-53
8B26	-78.4482 (589)	-80.1535 (443)	1.7053	1.88E-06
8B30	-72.8802 (668)	-76.6959 (467)	3.8157	1.07E-80
8CA0	-80.6734 (646)	-82.7034 (381)	2.03	5.42E-21
8CD1	-79.7937 (630)	-85.7704 (318)	5.9768	3.81E-140
8B44	-81.6128 (545)	-83.8438 (256)	2.2309	4.04E-51
Mean	*-76.9512*	*-79.0392*	*2.4144*	
Smartphone: S4 vs Nexus 5				
Access Point	S4 RSSI (count)	Nexus 5 (count)	MAD	Wilcoxon test
8B14	-67.7138 (622)	-73.5583 (609)	5.8445	1.93E-14
8B33	-77.5363 (606)	-78.2458 (419)	0.7095	1.45E-14
8B26	-78.4482 (589)	-79.303 (330)	0.8548	1.93E-42
8B30	-72.8802 (668)	-84.5116 (43)	11.6314	4.71E-28
8CA0	-80.6734 (646)	-82.7299 (211)	2.0565	1.56E-15
8CD1	-79.7937 (630)	-82.5678 (273)	2.7741	7.25E-85
8B44	-81.6128 (545)	-90.1905 (21)	8.5776	1.07E-17
Mean	*-76.9512*	*-81.5867*	*4.6355*	
Smartphone Orientation: S4 mini vs Nexus 5				
Access Point	S4 mini (count)	Nexus 5 (count)	MAD	Wilcoxon test
8B14	-67.5916 (404)	-73.5583 (609)	5.9667	1.95E-34
8B33	-76.5158 (380)	-78.2458 (419)	1.73	1.76E-28
8B26	-80.1535 (443)	-79.303 (330)	0.8505	8.88E-08
8B30	-76.6959 (467)	-84.5116 (43)	7.8157	5.29E-27
8CA0	-82.7034 (381)	-82.7299 (211)	0.0264	0.0164
8CD1	-85.7704 (318)	-82.5678 (273)	3.2027	41.51E-85
8B44	-83.8438 (256)	-90.1905 (21)	6.3467	4.04E-15
Mean	*-79.0392*	*-81.5867*	*2.5475*	

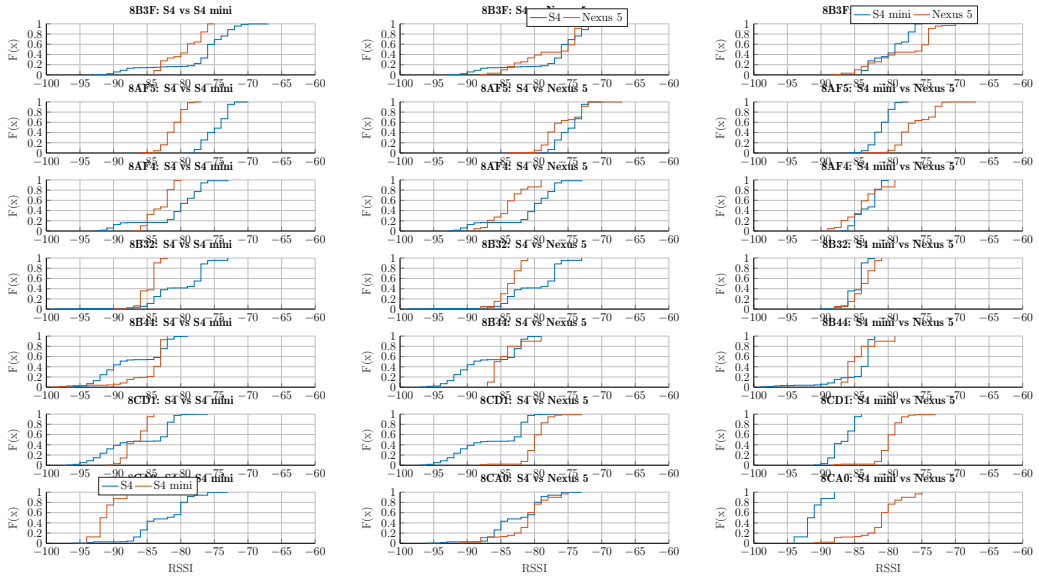


Figure 7.6: Illustration of RSSI bias due to smartphone hardware differences at L1.

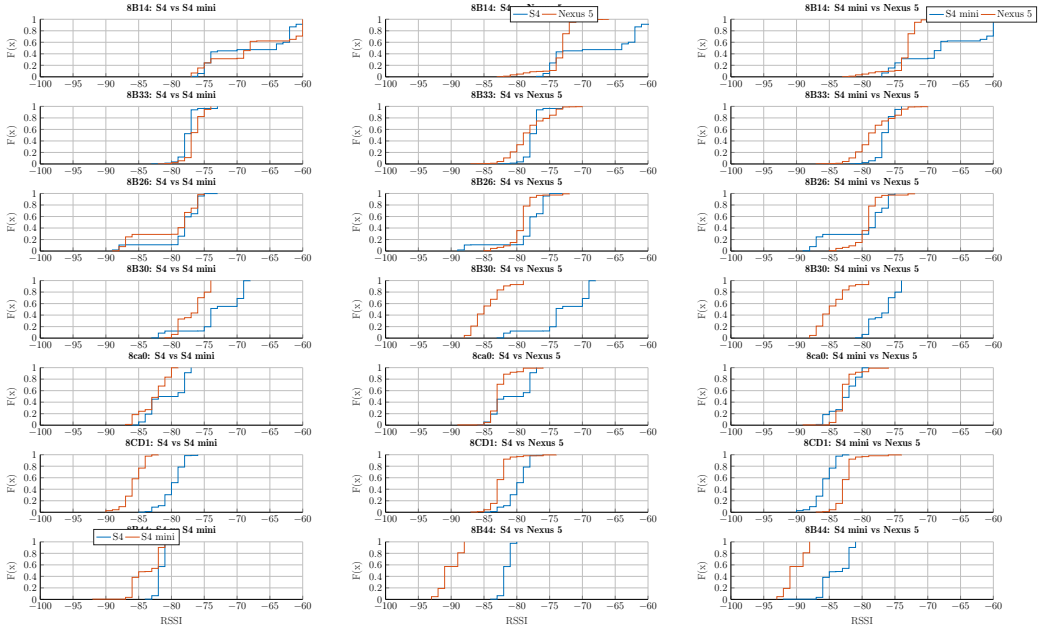


Figure 7.7: Illustration of RSSI bias due to smartphone hardware differences at L2.

7.2.4 Orientation of Smartphone

In this experiment, we investigated the effect of orientation of smartphone on the RSSI values. Similar to user's presence experiment, this was a user-free experiment with a random direction. As discussed in Chapter 6, we chose different orientations 0°, 45°and 90° using the S4 mini smartphone.

Overall for locations L1 and L2 (see Table 7.6 and 7.7), for the user-free experiments, that is, 0° versus 45°and 0° versus 90°, the null hypothesis was rejected for most of the experiments, and we had no evidence to say that the RSSI recorded over different orientation have the same median. We see higher bias for 0° versus 90°, with a maximum bias of 3.20 dBm. For the hand vs pocket experiment, the results are contrasting for the location L1 and L2. This experiment showed a maximum bias of 2.22 dBm. The bias in the RSSI distributions can be seen from Figures 7.8 and 7.9.

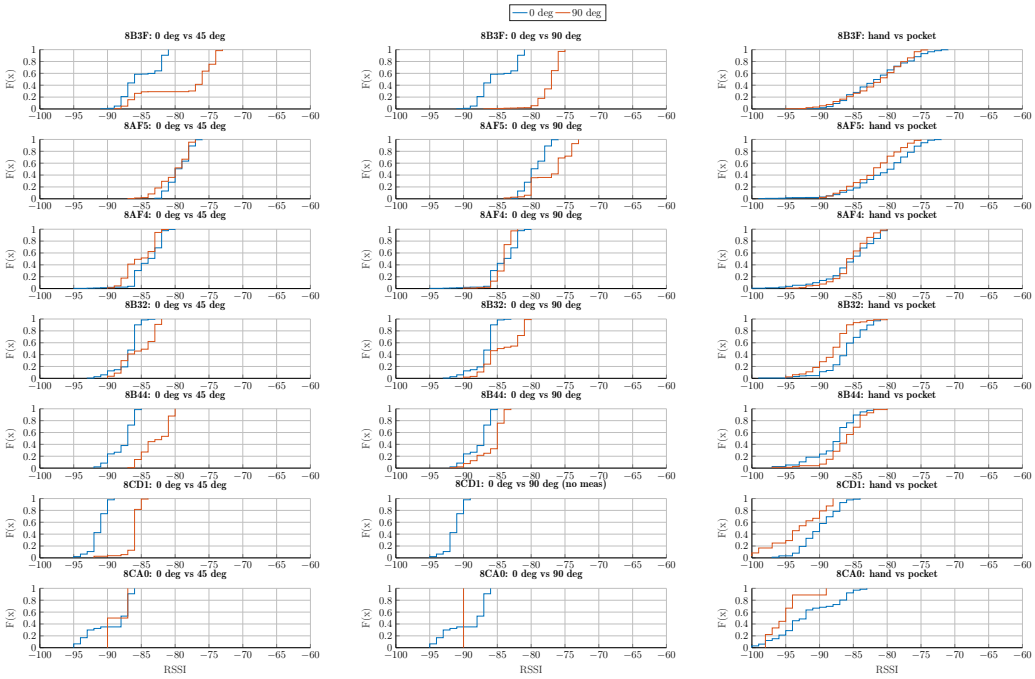


Figure 7.8: Illustration of RSSI bias due to orientation of the smartphone at L1.

Table 7.6: Summary of smartphone orientation bias to the RSSI measurements at L1.

Smartphone Orientation: 0° vs 45°				
Access Point	0° RSSI (count)	45° RSSI (count)	MAD	Wilcoxon test
8B3F	-85.0298 (369)	-78.6019 (422)	6.4279	1.88E-55
8AF5	-79.4463 (419)	-80.1002 (499)	0.6539	2.13E-04
8AF4	-84.0418 (287)	-85.2217 (221)	1.1799	2.97E-12
8B32	-86.9134 (254)	-85.3123 (285)	1.6011	3.76E-10
8B44	-87.7042 (144)	-82.7621 (269)	4.9421	5.61E-58
8CD1	-91.3404 (47)	-86.1193 (109)	5.2211	7.59E-24
8CA0	-89.3506 (77)	-88.5 (2)	**0.8506**	0.8718
Mean	*-86.2609*	*-83.8025*	*2.4584*	
Smartphone Orientation: 0°vs 90°				
Access Point	0° RSSI (count)	90° (count)	MAD	Wilcoxon test
8B3F	-85.0298 (369)	-77.2751 (458)	7.7547	7.48E-133
8AF5	-79.4463 (419)	-76.9649 (484)	2.4814	8.73E-43
8AF4	-84.0418 (287)	-84.189 (328)	0.1472	0.027
8B32	-86.9134 (254)	-84.1456 (309)	2.7678	3.05E-25
8B44	-87.7042 (144)	-85.7639 (144)	1.9403	1.06E-19
8CD1	-91.3404 (47)	NaN (0)	NaN	NaN
8CA0	-89.3506 (77)	-90 (2)	0.6494	0.4691
Mean	*-86.2609*	*-83.0564*	*3.2045*	
Smartphone Orientation: Hand vs Pocket				
Access Point	Hand (count)	Pocket (count)	MAD	Wilcoxon test
8B3F	-81.4135 (445)	-81.5547 (375)	0.1412	**0.8251**
8AF5	-80.5614 (399)	-81.9297 (256)	1.3683	2.06E-06
8AF4	-85.5031 (161)	-85.4275 (131)	0.0756	**0.5429**
8B32	-86.0152 (197)	-88.1098 (82)	2.0945	1.65E-09
8B44	-87.8684 (38)	-86.1622 (74)	1.7063	0.0066
8CD1	-90.0795 (88)	-93.2083 (24)	3.1288	5.38E-04
8CA0	-91.9394 (66)	-95.1111 (9)	3.1717	0.0258
Mean	*-86.1972*	*-87.3576*	*1.1604*	

Table 7.7: 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	Wilcoxon test
8B14	-66.1402 (428)	-67.209 (421)	1.0688	1.58E-09
8B33	-80.6326 (430)	-77.0043 (460)	3.6282	1.41E-30
8B26	-81.531 (226)	-83.5867 (225)	2.0557	3.29E-20
8B30	-79.5641 (390)	-75.9783 (415)	3.5858	1.07E-36
8CA0	-84.8239 (159)	-85.0227 (220)	0.1988	**0.2264**
8CD1	-88.9256 (121)	-86.2967 (246)	2.6289	8.14E-31
8B44	-84.4265 (272)	-84.6272 (228)	0.2007	**0.5018**
Mean	*-80.8634*	*-79.9607*	*0.9027*	
Smartphone Orientation: 0° vs 90°				
Access Point	0° RSSI (count)	90° (count)	MAD	Wilcoxon test
8B14	-66.1402 (428)	-63.6649 (370)	2.4753	1.84E-10
8B33	-80.6326 (430)	-75.259 (417)	5.3736	3.54E-55
8B26	-81.531 (226)	-84.3929 (168)	2.8619	9.97E-15
8B30	-79.5641 (390)	-77.3325 (382)	2.2316	6.88E-11
8CA0	-84.8239 (159)	-82.5994 (362)	2.2245	3.00E-13
8CD1	-88.9256 (121)	-84.32 (90)	4.6056	2.22E-57
8B44	-84.4265 (272)	-82.9281 (167)	1.4983	2.55E-06
Mean	*-80.8634*	*-78.6424*	*3.0386*	
Smartphone Orientation: Hand vs Pocket				
Access Point	Hand (count)	Pocket (count)	MAD	Wilcoxon test
8B14	-76.176 (375)	-77.9639 (360)	1.7879	2.81E-07
8B33	-83.8701 (154)	-81.3208 (293)	2.5493	1.80E-10
8B26	-84.5294 (102)	-82.728 (250)	1.8014	1.40E-04
8B30	-84.6071 (84)	-85.4365 (197)	0.8294	**0.0181**
8CA0	-87.303 (33)	-85.4865 (111)	1.8165	0.0034
8CD1	-87.359 (78)	-84.7133 (150)	2.6456	7.89E-06
8B44	-86.2576 (66)	-87.831 (71)	1.5734	8.23E-05
Mean	*-84.3003*	*-83.64*	*2.221*	

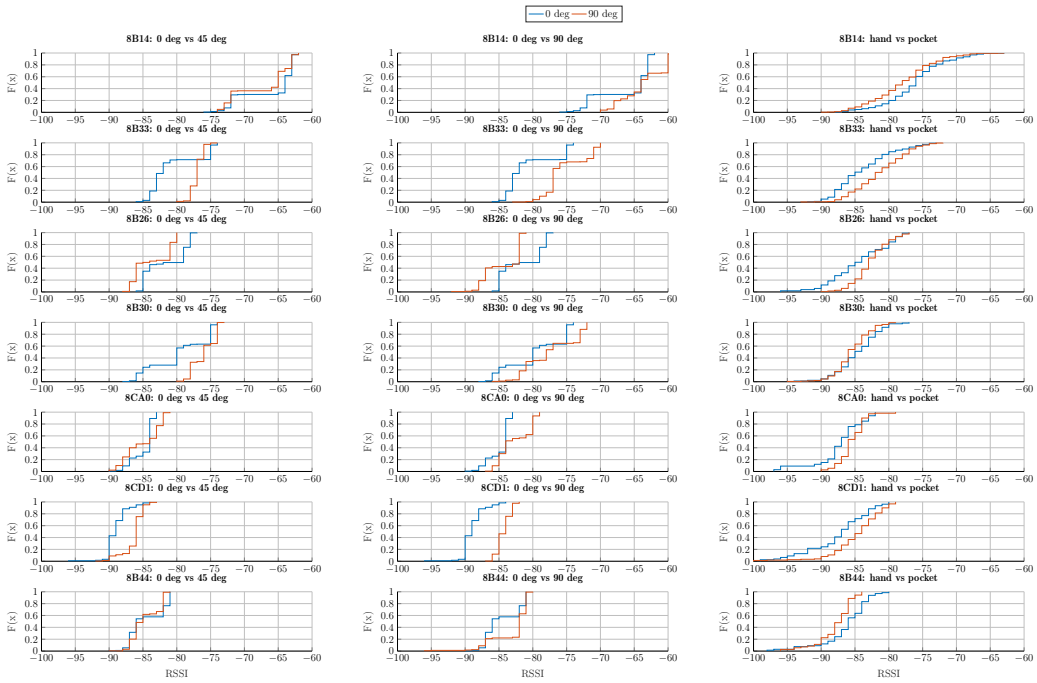


Figure 7.9: Illustration of RSSI bias due to orientation of the smartphone at L2.

7.2.5 Material of the Luminaires

In this experiment, we investigated the RSSI bias due to the material of the luminaire. In order to eliminate any other factors affecting the measurements, a single smartphone Samsung S4 mini was chosen and measurements were recorded without any user. In the setup, the phone was perpendicular (at 90°) to the floor. It was assumed that the orientation MU have minimal effect on the signal strength values.

We conducted two sets of experiments, with one luminaire enclosure made of plastic and metal, and other luminaire made only of plastic with the access point inside the ceiling. For inferring the bias and comparison, the different material luminaires selected were at same distance. Overall for both the experiments, the null hypothesis was rejected for all the experiment, and we had no evidence to say that the RSSI recorded over luminaires of different material have the same median. The results (see Table 7.8) shows the maximum bias 5.24 dBm between the two material of luminaires on the RSSI values. The bias in RSSI distributions can be seen from Figure 7.10.

Table 7.8: Summary of luminaire material bias to the RSSI measurements.

Experiment 1				
Distance (m)	Metal + plastic RSSI (count)	Plastic RSSI (count)	MAD	Wilcoxon test
1.84074	-71.8418 (392)	-72.5564 (532)	0.7146	1.09E-10
3.68148	-84.4118 (272)	-74.3834 (493)	10.0284	3.52E-118
5.52222	-83.25 (392)	-74.9864 (368)	8.2636	8.51E-126
7.36296	-85.8042 (240)	-85.529 (293)	0.2752	0.0013
9.2037	-82.9 (230)	-80.1208 (356)	2.7792	7.61E-24
11.12811	-83.6786 (224)	-81.5825 (424)	2.096	1.09E-34
13.05252	-89.3429 (35)	-75.3946 (185)	13.9483	2.45E-22
Mean	*-83.0327*	*-77.7933*	*5.2394*	
Experiment 2				
Distance (m)	Metal + plastic RSSI (count)	Plastic RSSI (count)	MAD	Wilcoxon test
1.84074	-72.0806 (434)	-66.9766 (471)	5.104	7.01E-40
3.68148	-74.233 (279)	-75.1373 (437)	0.9043	1.46E-11
5.52222	-77.8531 (422)	-80.5909 (506)	2.7378	1.74E-41
7.36296	-83.8486 (370)	-82.3724 (427)	1.4763	1.70E-05
9.2037	-83.7517 (286)	-79.6709 (471)	4.0808	3.51E-61
11.12811	-85.975 (120)	-84.9881 (420)	0.9869	1.89E-13
13.05252	-94 (37)	-82.9429 (105)	11.0571	3.70E-21
Mean	*-81.6774*	*-78.9542*	*2.7232*	

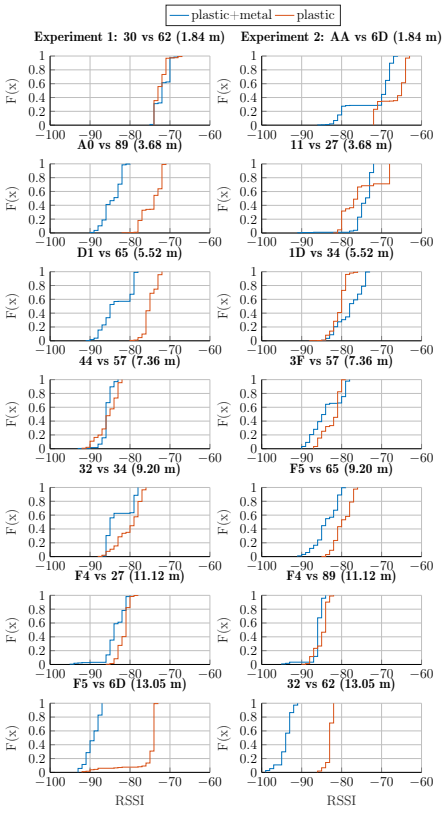


Figure 7.10: Illustration of RSSI bias due to material of the luminaire at L1 (Experiment 1) and at L2 (Experiment 2). Corresponding experiments in the Table 7.8 can be found using the distance and experiment number. Last two characters of MAC address (as in Figure 6.1) are used for denoting the experiment.

7.2.6 Comparison with Radio Analyzer

In this experiment, we investigate the bias induced by the smartphones in comparison to the radio analyser. We used smartphones Samsung S4 mini, Samsung S4 and Samsung S7 for this experiment. Using the same configuration as previous experiments, we assumed minimal bias due to direction of the smartphone and user-free measurements were recorded. One direction was chosen and it was consistent for all the experiments.

Overall for location L1 (see Table 7.9), the null hypothesis was rejected for all the experiments and we had no evidence to say that the RSSI recorded for the experiments with different smartphone and radio analyzer have the same median. The maximum bias inferred is 15.6 dBm. The bias in the RSSI distributions can be seen from Figure 7.11.

Table 7.9: 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	Wilcoxon test
8B3F	-75.2315 (1326)	-84.4048 (168)	9.1732	3.60E-94
8AF5	-70.7684 (1956)	-75.7176 (170)	4.9492	1.20E-69
8AF4	-82.1404 (178)	-86.3286 (140)	4.1881	1.92E-39
8B32	-75.8115 (1751)	-85.4 (130)	9.5885	1.48E-80
8B44	-82.8472 (386)	-86.5905 (105)	3.7433	2.19E-40
8CD1	-82.8471 (157)	-92.0886 (79)	9.2415	1.02E-36
8CA0	-82.8815 (540)	-88.0294 (34)	5.1479	1.49E-22
Mean	*-78.9325*	*-85.5085*	*6.576*	
RSSI bias in Smartphones: Radio analyzer vs S4 mini				
Access Point	Radio analyzer RSSI (count)	S4 mini RSSI (count)	MAD	Wilcoxon test
8B3F	-77.3256 (1689)	-85.1228 (114)	7.7972	1.79E-71
8AF5	-70.6342 (2452)	-84.2727 (110)	13.6386	2.58E-70
8AF4	-81.042 (405)	-86.481 (79)	5.439	5.60E-45
8B32	-75.5629 (2288)	-87.8 (15)	12.2371	8.90E-12
8B44	-83.3467 (424)	NA	NA	NA
8CD1	-84.619 (21)	-88.2143 (14)	3.5952	1.45E-04
8CA0	-82.975 (519)	-89 (1)	-6.025	8.13E-02
Mean	*-79.3579*	*-86.8151*	*7.4572*	
RSSI bias in Smartphones: Radio analyzer vs S7				
Access Point	Radio analyzer RSSI (count)	S7 RSSI (count)	MAD	Wilcoxon test
8B3F	-73.6596 (1416)	-88.6988 (83)	15.0392	1.48E-55
8AF5	-71.3179 (2120)	-86.9892 (93)	15.6713	2.82E-61
8AF4	-80.5172 (435)	-96.7 (10)	16.1828	1.60E-08
8B32	-75.8498 (1977)	-93.0714 (28)	17.2217	1.12E-20
8B44	-82.7293 (548)	-96.6667 (12)	13.9374	1.14E-09
8CD1	-84.1875 (64)	NA	NA	NA
8CA0	-82.3946 (299)	NA	NA	NA
Mean	*-76.8148*	*-92.4252 (till 8b44)*	*15.6104 (till 8b44)*	

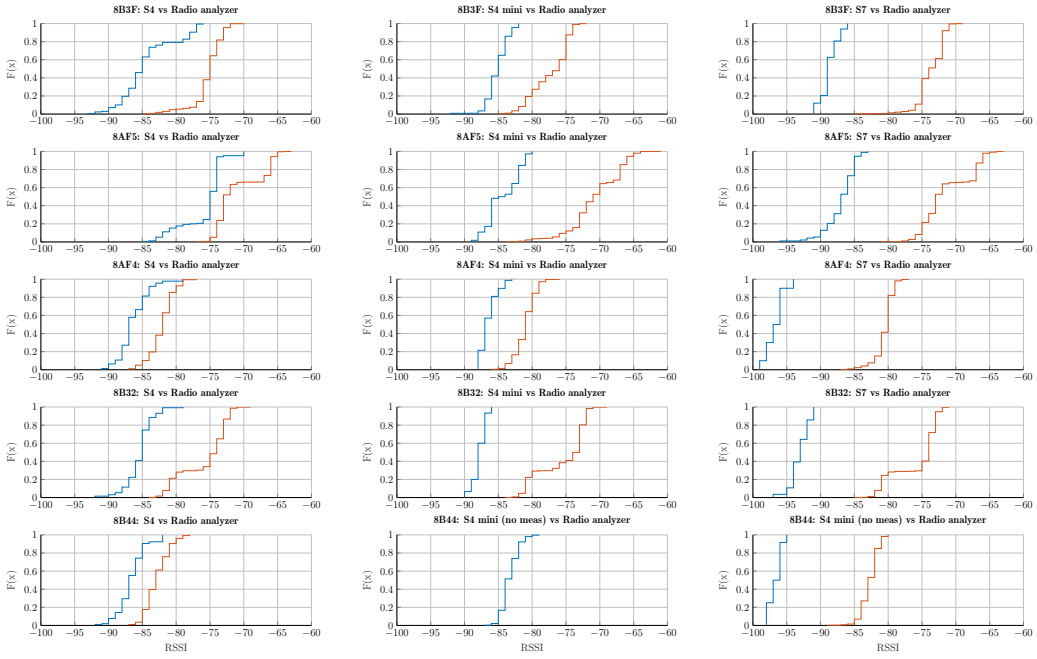


Figure 7.11: Illustration of RSSI bias in smartphones when compared to radio analyzer at L1.

7.3 Received Signal Strength Indication: Revisited

As mentioned in Chapter 6, we record the RSSI values along with their MAC address with absolute timestamp and relative time of recording using the android measurement application. In this experiment, we investigate the temporal behaviour of RSSI values. For simplicity, we selected three closest luminaires with the user-free setting. We used the smartphones Nexus 6, Samsung S4 and Samsung S4 mini for recording the RSSI values.

Figure 7.12 shows a peculiar but periodic characteristic of the RSSI signal from Helvar and Kontakt beacons. It shows a systematic pattern and, hence, can not be attributed to the signal fluctuations. The RSSI pattern calls for a detailed investigation into this phenomenon. In order to avoid all the different environmental factors which add bias to RSSI, we record the measurements of the beacons at an outdoor setting with the android phone and radio analyser. We describe the experiment next.

With a user-free setting in an outdoor environment, we recorded the RSSI measurements using the smartphones Nexus 6 and Samsung S7 (see Figure 6.7) in conjunction with the radio analyzer.

From Figure 7.13, it is evident that the periodic trend in the RSSI values

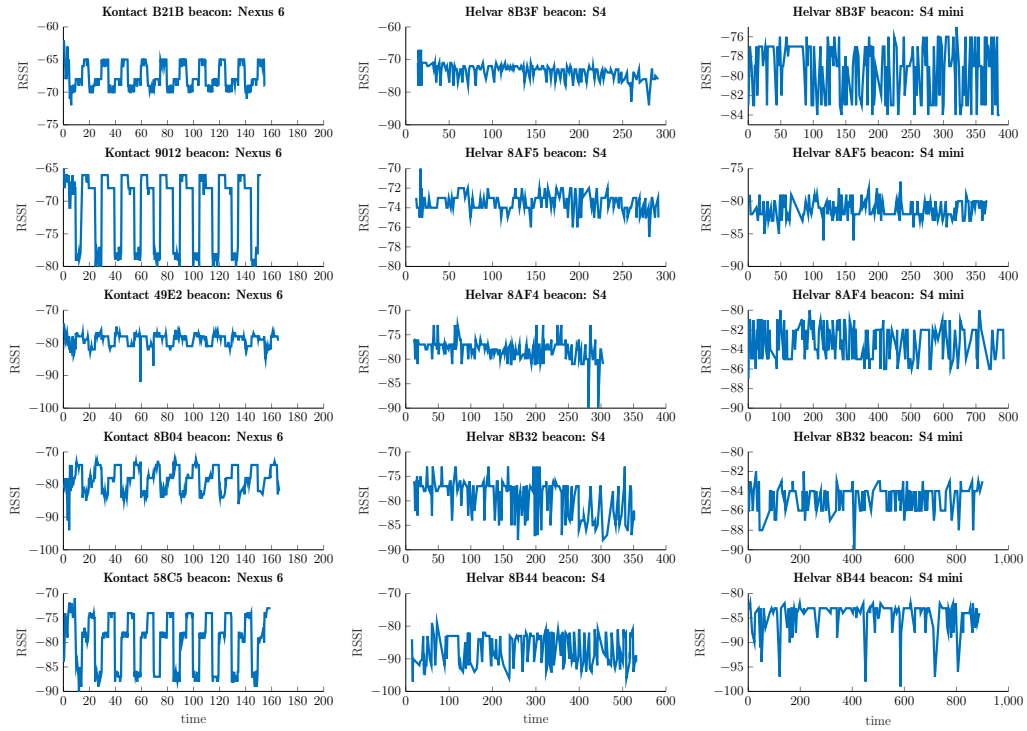
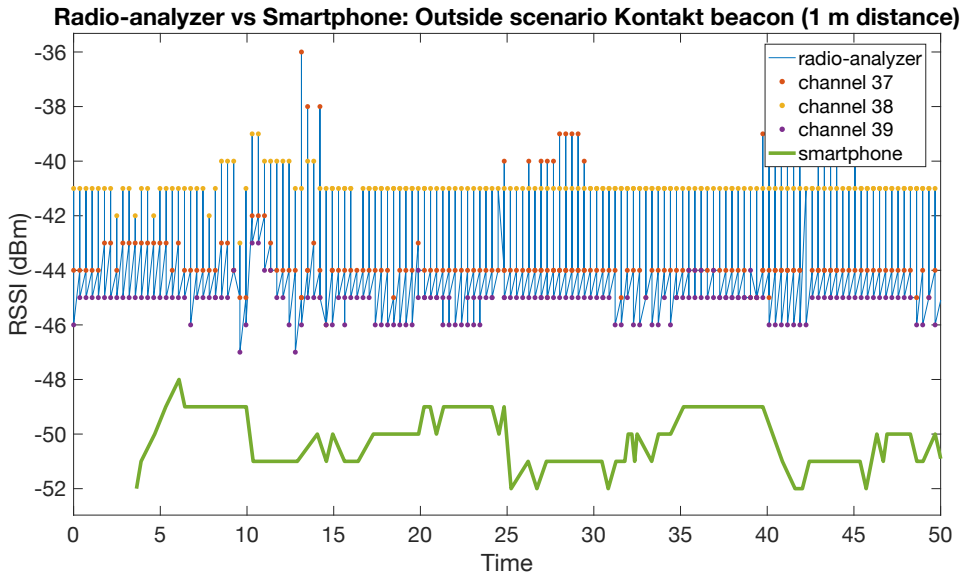
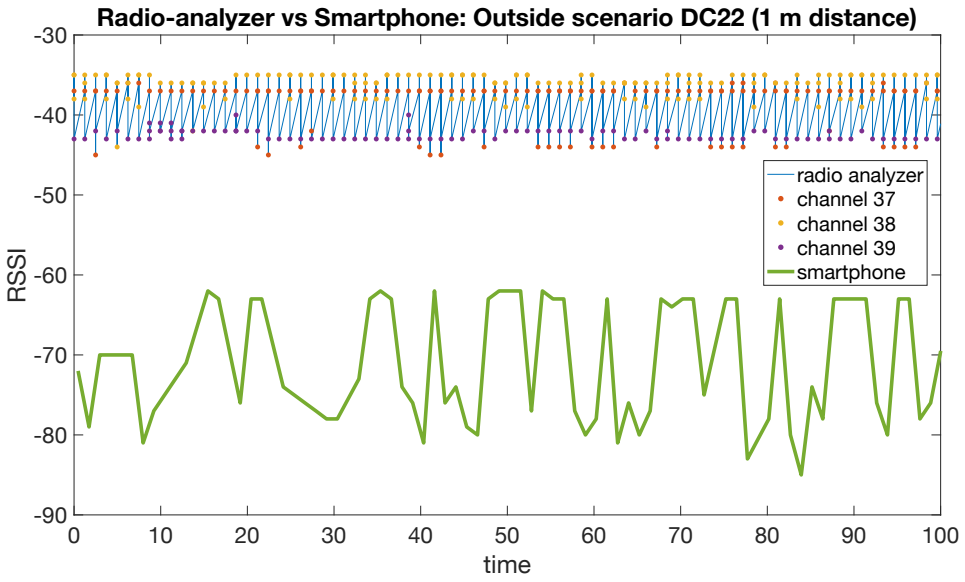


Figure 7.12: Illustration of the RSSI versus the time for Kontakt and Helvar beacons with measurements recorded using smartphones Nexus 6, Samsung S4 and Samsung S4 mini. It shows a systematic periodic pattern.

is due to the difference in the performance of the advertisement channels of the BLE beacon. Due to the mismatch in the performance of the channels, the power of the signal strength advertised by BLE varies, which directly affects the RSSI. The variation due to this phenomenon is over 20 dBm (see Figure 7.3). Testing with the different beacons showed that the channel which performed the worse (least in terms of RSSI) was not always the same. Now, we understand that the multimodality of the RSSI values is due to the different performance of the advertisement channels.



(a) Kontakt beacon



(b) Helvar beacon

Figure 7.13: Comparison of RSSI from radio analyzer and smartphones in a outdoor setting. The smartphone used were (a) Nexus 6 and (b) Samsung S7 respectively.

Chapter 8

Experiments and Results

In this chapter the algorithms explained in Chapter 5 are implemented and evaluated using the recorded RSSI data at Helvar R&D. In Section 8.1, we describe the performance metrics used for evaluation of the positioning algorithms. Next, in Section 8.2, we aim to find the optimum parameter for the parameters in fingerprinting, that is, the number of calibration points and calibration time. Last, in Section 8.3, we evaluate the results for the positioning algorithms.

8.1 Performance metrics

The performance of the positioning methods are evaluated in terms of the absolute errors of the prediction. We get the one-shot location predictions and the optimal criterion for evaluation is the root mean square error (RMSE). The other criteria used are the mean, 90^{th} percentile, maximum absolute error and variance of the error estimates.

8.2 Effect of Parameters in Calibration Phase

In this section, for the given Helvar R&D positioning arena, we optimize the number of calibration points and calibration time. The simulations are designed such that for a particular set of calibration parameters we generate the Gaussian process radiomap using the data structures $\mathcal{R}_1$ and $\mathcal{R}_2$. The algorithms particle filter and unscented Kalman filter are run using both the radiomap. In the experiment, if we run a simulation using particle filter with radiomap $\mathcal{R}_1$, we label it with PF-1 and other simulations are labelled accordingly. We compute the RMSE error for the test data averaged over

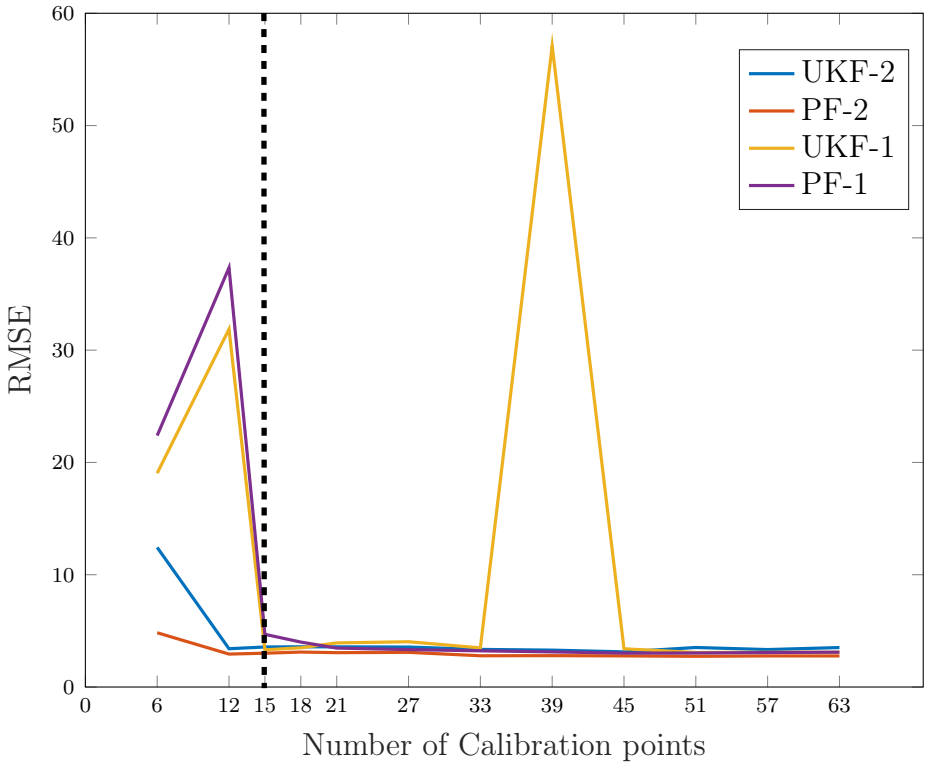


Figure 8.1: Optimisation experiment for the number of calibration points for PF and UKF using the radiomaps $\mathcal{R}_1$ and $\mathcal{R}_2$. PF-1 denotes particle filter algorithm with radiomap $\mathcal{R}_1$.

independent 100 Monte Carlo simulations and then infer the optimal parameters. We refer to parameters inferred from this experiment as *optimal parameters 1*.

8.2.1 Separate Parameter Evaluation

First, we experiment with an increasing number of calibration points keeping the calibration time constant to the maximum, that is, 50 seconds. As discussed in the Chapter 6, we have 63 calibration points in total, hence, we start with 6 points and keep increasing the points depending on the location in our testbed. From Figure 8.1, we observe an elbow at 15 calibration points for all the algorithms. We also see an unexplained spike at 40 calibration points for UKF-1 using the *radiomap 1* which is not visible in the *radiomap 2*, showing the shortcomings of the *radiomap 1*.

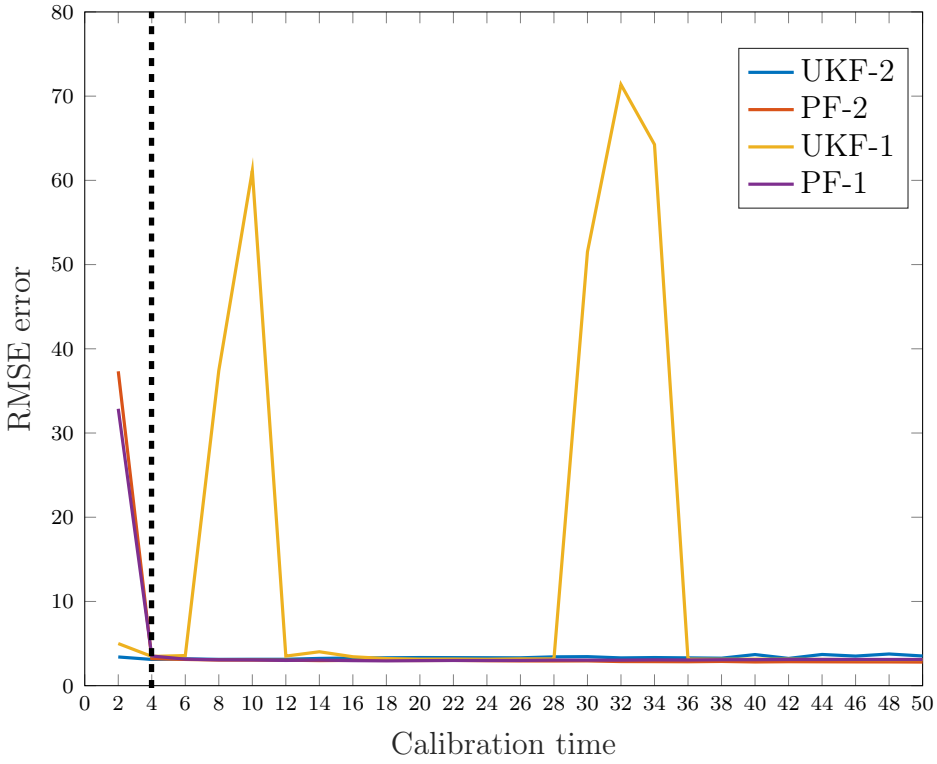


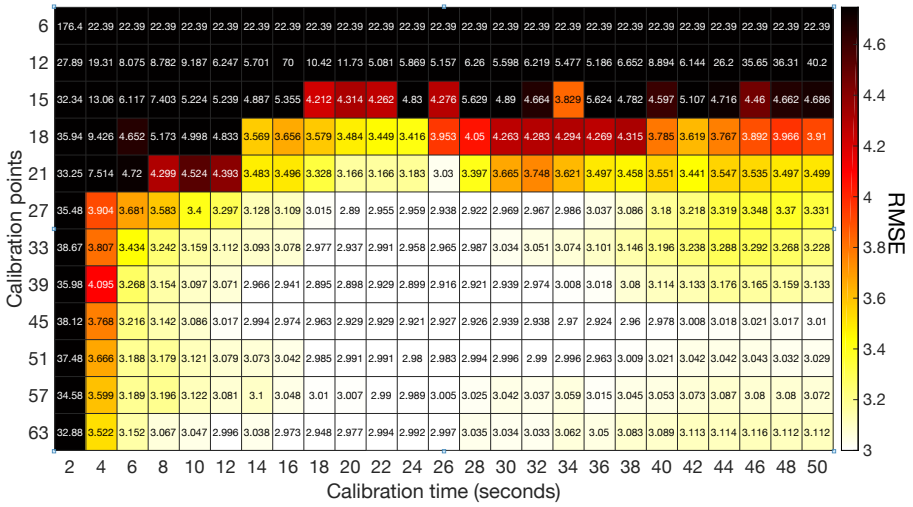
Figure 8.2: Optimisation experiment for the calibration time for PF and UKF using the radiomaps $\mathcal{R}_1$ and $\mathcal{R}_2$.

Second, we experiment with the increasing the calibration time keeping calibration points to maximum constant, that is, 63 points. As discussed in the Chapter 6, we recorded the calibration data for 50 seconds, so, in this experiment we run the simulations starting with 2 seconds and keep increasing the time at a step of 2 seconds, hence, we run 25 simulations in total. From Figure 8.2, we infer a clear elbow at 4 seconds. In addition, we also observe unusual spikes in the RMSE for UKF when used with the radiomap $\mathcal{R}_1$ (see Figures 8.1 and 8.2). We infer the *optimal parameters 1* as 15 calibration points and 4 seconds calibration time.

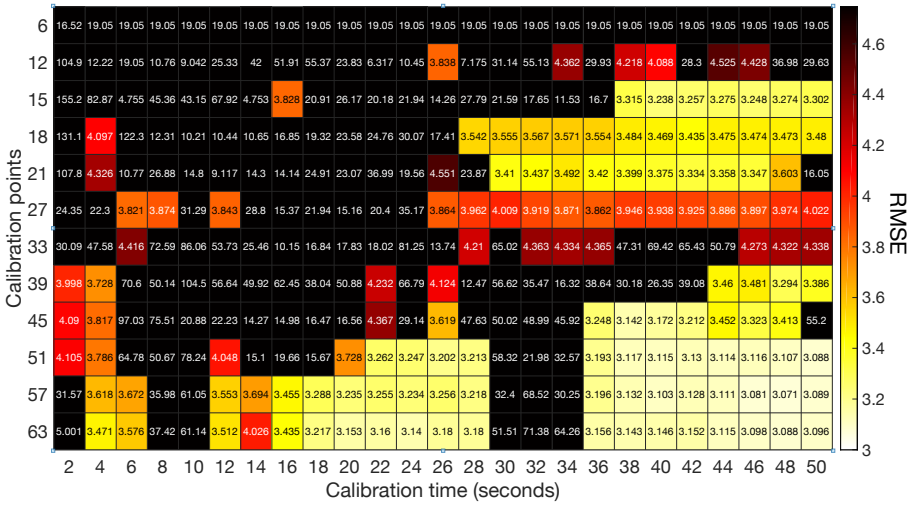
8.2.2 Combined parameter evaluation

Lastly, we vary both the parameters, that is, calibration points and calibration time combining the strategy from the previous experiments of calibration points and calibration time. We can see that the *radiomap 2*, gives stable

results than its counter part *radiomap 1*. The results have been displayed in Figures 8.3 and 8.4. We observe that Figure 8.4a gives best results and we infer that optimal parameters are 12 calibration points and 24 seconds as calibration time. We refer to these inferred parameters as *optimal parameters 2*. The spike in Figures 8.1 and 8.2 can also be seen in Figure 8.3b.

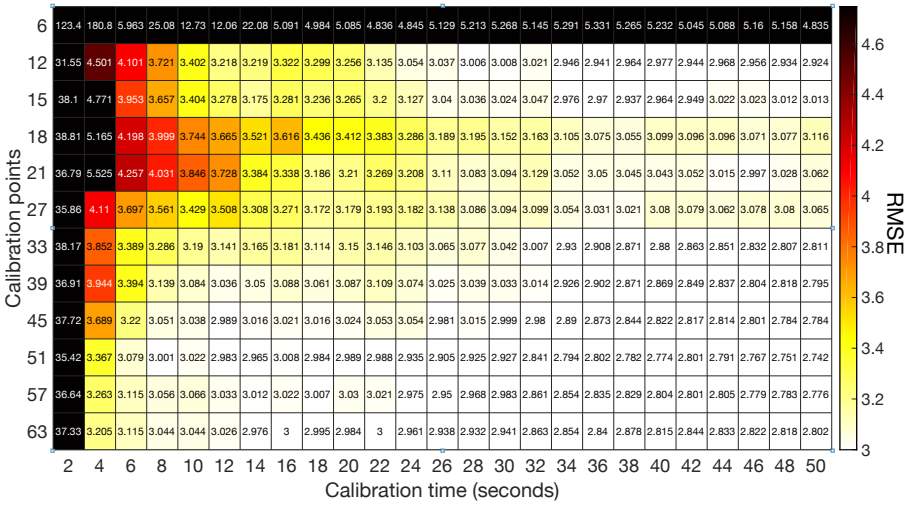


(a) PF-1

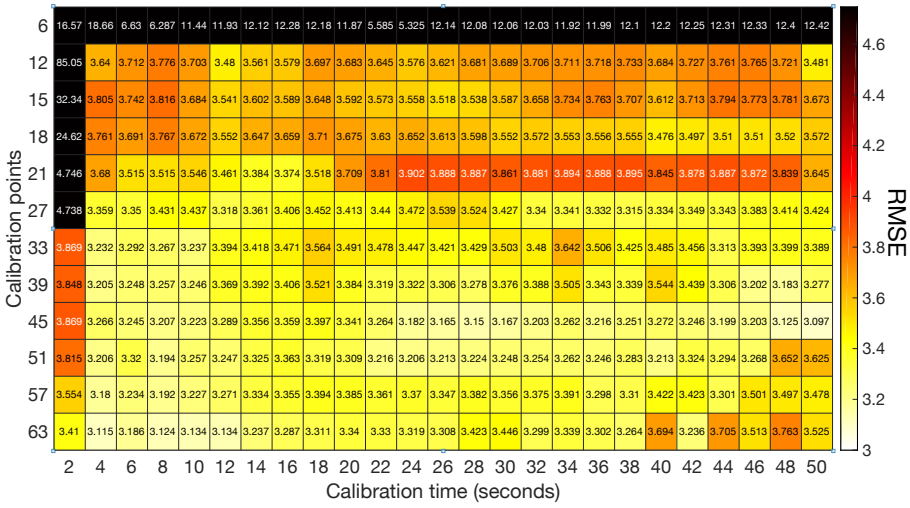


(b) UKF-1

Figure 8.3: Optimisation experiment varying both the calibration points and calibration time using the radiomap $\mathcal{R}_1$.



(a) PF-2



(b) UKF-2

Figure 8.4: Optimisation experiment varying both the calibration points and calibration time using the radiomap $\mathcal{R}_2$.

8.3 Filtering location estimation algorithms

The memory based Bayesian filtering methods described in the Chapter 5 were implemented in MATLAB. The test data and testbed, as discussed in Chapter 6, were used for evaluating these methods. The estimation of the location was done at every measurement step using all the 28 BLE beacons. These simulations produced the one-shot position estimate and the simulations were run in conjunction with both the radiomaps $\mathcal{R}_1$ and $\mathcal{R}_2$.

Table 8.1: Summary of performance of the memory based indoor positioning methods with radiomap $\mathcal{R}_1$ & $\mathcal{R}_2$

Method	**RMSE (m)**	**Mean (m)**	90^{th} **Percentile (m)**	**Max (m)**	**Variance**($\mathbf{m^2}$)
Optimal Calibration Parameters - 1					
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
Optimal Calibration Parameters - 2					
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
Full Calibration Parameters					
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

8.3.1 Summary of the location estimation algorithms

The *optimal parameters 1* and *2* gathered from the experiments in Section 8.2 were used and simulations were run with the algorithms particle filter and unscented Kalman filter. The results and performance metrics for the simulations with *optimal parameters* are summarised in the Table 8.1 (also see Figure 8.6). A summary for the simulations using the *full calibration parameters*, that is, 63 calibration points and 50 seconds calibration time is also given.

In comparison to radiomap $\mathcal{R}_1$, the results using the radiomap $\mathcal{R}_2$ for the different parameter shows significant improvement in terms of performance for both particle filter and unscented Kalman filter. The preserved variance in the reference table 2 translates into improved performance.

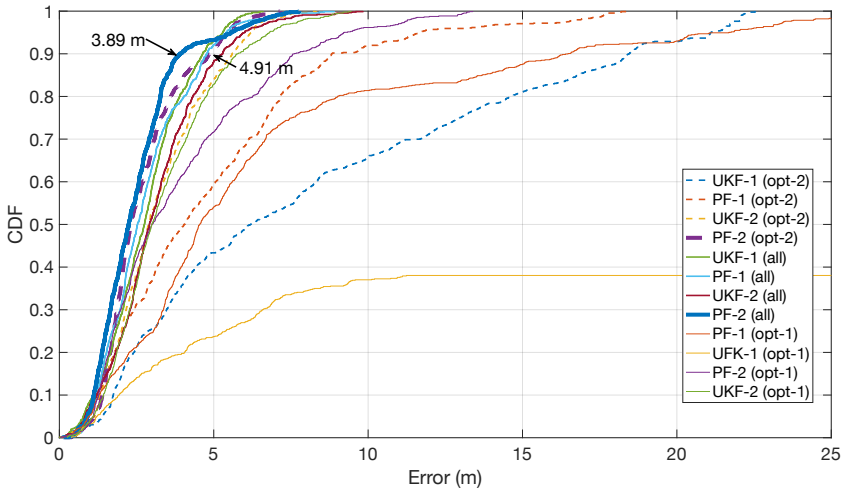


Figure 8.5: Performance evaluation for different positioning using the cumulative distribution of the errors. Apart from PF-2 using all the calibration parameters, the PF-2 using the optimal parameters 2 give the best results with 90% of errors falling under 4.91 metres.

We have additionally plotted Figure 8.5, showing the cumulative error for all the evaluated filtering methods with different radiomaps and different optimal 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 metres. This makes intuitive sense as more information is provided better the filter performs. With regards to optimal parameters for calibration phase, we see that *optimal parameters 2* give the best results for particle filters algorithm with 90% of the errors to be under 4.91 metres.

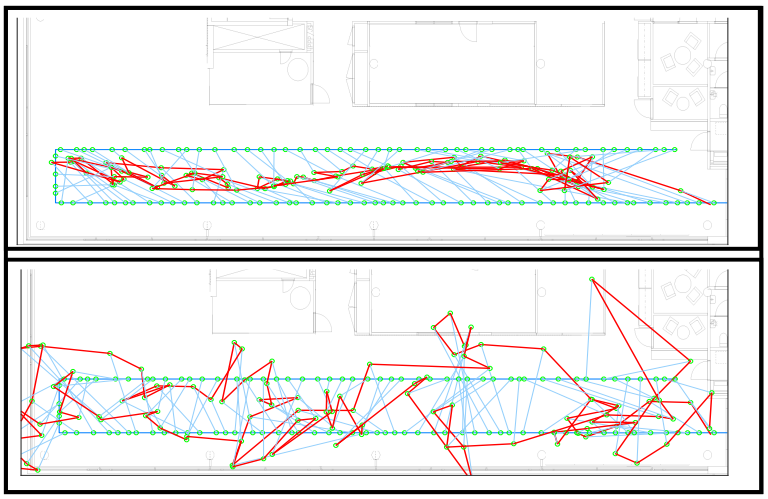


Figure 8.6: Positioning result for (top) Particle Filter using radiomap 2 (bottom) Unscented Kalman Filter using radiomap 2.

Chapter 9

Discussion & Future Work

9.1 Discussion

9.1.1 Data Analysis

This study takes a step backwards in understanding the data and in that process we understood that the variation in BLE RSSI is inherent in its architecture and its technology. This thesis made a **invention** in understanding the unique characteristic of the BLE devices which shows a distinction in performance of the three advertisement channels. The BLE architecture has 40 physical channels and has 3 advertisement channels which are strategically placed to avoid interference with the WiFi channels. The advertisement channels are spread over the frequencies and hence, due to BLE being a low powered device, the performance of the channels is distinct. The difference in the performance channel affects the signal strength to vary as much as 20 dBm and hence the data is tri-modal. This phenomenon could add two-fold bias. One, bias is added through the approximate construction of the radiomap (for example, radiomap $\mathcal{R}_1$). Next, bias is also added when a filter is run with sub-optimal data model.

Given the characteristic of the RSSI signal, we can say that averaging the signal strength values could lead to misleading results and the distance interpreting non-memory methods would perform poorly as the prediction would be be highly biased.

9.1.2 Radio map

This thesis looked at the generation of radiomaps using 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, in contrast to averaged RSSI in $\mathcal{R}_1$.

The $\mathcal{R}_2$ showed a improved performance in terms of RMSE for the filtering algorithms overcoming the effects of the unequal channel performance.
The advantages of GP based radiomaps are that it provides flexibility for learning the spatial characteristics, for example, GP based radiomap was able to learn the presence of the wall / obstruction just based on the RSSI values. In addition, the GP based radiomap with Bayesian filters gave a considerably good accuracy even with fewer calibration points and less calibration time. Moreover, the usage of GP makes the quantisation of the RSSI values insignificant, making the area / cell based location estimation methodology irrelevant.

9.1.3 Indoor Positioning

In this thesis, we test the memory methods in the form of Bayesian filtering, that is, particle filters and unscented Kalman filters. We experimented to find the optimal calibration parameters and found that particle filters performed best in the Helvar R&D office space setting. These methods add additional capabilities of integrating data from multiple sources, that is, for example, add inertial sensors data and hence, can pave a way for a unified indoor and outdoor positioning solution. This methodology also mitigates the problem of malfunction of BLE modules as the solution integrates information from multiple sources and as this methodology can track trajectories, usual fluctuation can be smoothed out.

Concerning the performance of particle filter and unscented Kalman filter for radiomap $\mathcal{R}_2$, the particle filter almost always performs better than unscented Kalman. Given the poor performance of radiomap $\mathcal{R}_1$, the results are considered sub-optimal and it gives usually high RSME when used in conjunction with UKF (see Figures 8.1 and 8.2).

9.2 Future Work

Using the additional information in terms of bias from the channel performance, smartphones, orientation of phones, material of the luminaire, we need to formulate a better data model for BLE RSSI. We can also look at masking of the advertisement channel (or multiple channels) in BLE beacons and investigate if it mitigates the fluctuation in RSSI. Aguilar-Garcia et al. (2015) talks about self-optimising and self-healing methods for Self-Organizing networks (SON) applied to cellular networks. These concepts are highly relevant to IPS with BLE and potentially could make finger-printing technique unsupervised. As proved in the thesis, the RSSI is highly non-

linear and tri-modal data, and we can use Deep Gaussian Processes as a measurement model to learn this complex representations.

References

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.

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.

Argenox (2015). A ble advertising primer. Accessed: 2018-11-11.

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.

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.

Bisio, I., Lavagetto, F., Sciarrone, A., and Yiu, S. (2017). A smart 2 gaussian process approach for indoor localization with rssi fingerprints. In *Communications (ICC), 2017 IEEE International Conference on*, pages 1–6. IEEE.

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*.

EU (2003). European commission press release (2003), indoor air pollution: new eu research reveals higher risks than previously thought. Accessed: 2018-11-11.

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.

Ferris, B., Fox, D., and Lawrence, N. D. (2007a). Wifi-slam using gaussian process latent variable models. In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 2480–2485.

Ferris, B., Hahnel, D., and Fox, D. (2007b). *Gaussian processes for signal strength-based location estimation*, volume 2, pages 303–310. MIT Press Journals.

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.

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. (2011). Comparative evaluation of received signal-strength index (rssi) 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.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Richter, P. and Toledano-Ayala, M. (2015). Revisiting gaussian process regression modeling for localization in wireless sensor networks. *Sensors*, 15(9):22587–22615.

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.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Solin, A. (2016). *Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression*. Phd dissertation, Aalto University.

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.

Teledyne Lecroy, I. (2016). Frontline sodera le wideband bluetooth protocol analyzer. Accessed: 2018-11-11.

USEPA (2013). *The Inside Story: A Guide to Indoor Air Quality*. BiblioGov.

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.

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.