

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

Srikanth Gadicherla

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

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

DRAFT! — September 21, 2018 — DRAFT!

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

Author:	Srikanth Gadicherla		
Title:	Data Analysis and Memory Based Methods for RSS Bluetooth Low Energy Indoor Positioning		
Date:	!FIXME July 26, 2017 FIXME!	Pages:	xi + 93
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		
<p>The thesis aims at getting a proof of concept for solutions to the hybrid bluetooth low energy indoor positioning (BLE-IP) along with automatic node location identification (ANOLI). ANOLI is mainly aimed during the calibration phase which gives out the locations of the bluetooth low energy (BLE) beacons. The BLE beacons are present inside the luminaires. The luminaires might either be installed in the ceiling or tethered to it. The location of BLE beacons determined using ANOLI and received signal strength indicator (RSSI) together are used in the BLE-IP task. The factors influencing the RSSI values is also studied and based on the target environment and accuracy the model is decided. Simple applications like asset tracking is also implemented.</p>			
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've always stood like a wall. I'm in debt to my brother and friend *Badri Kakulavarapu* for his constant words of motivation throughout.

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

I thank Helvar for giving me this opportunity to work on my thesis, and especially my bosses *Henri Júlsen*, *Max Björkgren* and my daily advisor *Heikki Pulkkinen* for giving me enough freedom in my research and explore different avenues in my thesis. Toni Ajo and Guo Haipeng

!FIXME lab mates: Juho, Eero, Marko, Markus, Olli-Pekka, Akash Kumar Dhaka, Michael Riis Andersen !FIXME!

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

Otaniemi, !FIXME July 26, 2017 !FIXME!

Srikanth Gadicherla

Abbreviations

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

Contents

Acknowledgements	iii
Preface	ix
Symbols and Abbreviations	xi
1 Introduction	1
1.1 Motivation	2
1.2 Indoor Positioning	3
1.3 Bluetooth Low Energy	5
1.4 Challenges Plaguing the RSS based methods	5
1.5 Goal of the thesis	6
1.6 Contribution of the thesis	6
1.7 Outline of the thesis	6
2 Location Fingerprinting	7
2.1 Location Fingerprinting	7
2.1.1 Radiomap	8
2.1.1.1 Reference Table	9
2.1.2 Raw data	10
2.2 RSS based methods	10
2.3 Positioning Algorithms	12
2.3.1 Non-memory based methods	12
2.3.1.1 k-Nearest Neighbour	12
2.3.2 Memory based methods	13
2.4 Related Work	13
2.4.1 Signals of Opportunity (SO _{Op})	14
2.4.2 Cellular Networks	14
2.4.2.1 Enabling Technologies	14
2.5 Performance metrics	15

3 Gaussian Process Model for Bayesian Filtering	17
3.1 Motivation for using the GPs	18
3.2 Advantages of GPs	18
3.3 GP modeling for indoor positioning	19
4 Memory Based Filtering Methods	23
4.1 Introduction	23
4.2 Sequential Monte Carlo	23
4.3 Bayesian Filtering	23
4.3.1 State Space Modeling	24
4.3.2 Advantages of Bayesian Filtering	26
4.4 Filtering	27
4.5 Particle Filters	27
4.5.1 Importance Sampling	27
4.5.2 Filtering	29
4.6 Unscented Kalman Filters	33
4.6.1 Unscented Transform	33
4.6.2 Filtering	34
5 Measurement Setup & Data Collection	37
5.1 Experimental Testbed	37
5.2 Measurement Configuration	37
5.3 Bluetooth Low Energy	38
5.3.1 Physical Layer	39
5.3.2 Advertisement channels	39
5.3.3 Data channels	40
5.3.4 Different BLE protocols	40
5.4 Assumptions on the RSSI values	41
5.5 Radio Analyzer	41
5.6 Measurement Application	41
5.6.1 Smartphones	42
5.7 Data Collection	44
5.7.1 For Data Analysis	44
5.7.1.1 Different User Orientations	44
5.7.1.2 Different Phone Orientations	44
5.7.1.3 Outdoor Measurements	45
5.7.2 For Positional Algorithms	46
5.7.2.1 Setup for Calibration Phase and Test Phase .	46
5.7.2.2 Calibration Data	46
5.7.2.3 Test data	46

6 Data Analysis of RSSI Values	49
6.1 RSSI as a measure of distance	50
6.2 Statistical Hypothesis Testing	50
6.2.1 Two Sample Kolmogorov-Smirnov test	52
6.2.2 Two Sample Z-test	52
6.2.3 Levene's test	53
6.2.4 Wilcoxon rank-sum test	54
6.2.5 Experiments	55
6.2.5.1 Bias due to user's presence	56
6.2.6 Bias due to different smartphone	56
6.2.6.1 Orientation of Smart phone	56
6.2.6.2 Material of the Luminaires	56
6.2.6.3 Comparisons of Smartphone	63
6.2.6.4 Comparison with Radio Analyzer	63
6.3 Received Signal Strength Indication: Revisited	63
6.3.1 Measurement Outdoors with Radio analyzer	66
7 Experiments and Results	73
7.1 Effect of parameters in calibration phase on positioning	73
7.1.1 Separate parameter evaluation	73
7.1.1.1 Calibration points	73
7.1.1.2 Calibration time	74
7.1.2 Combined parameter evaluation	75
7.2 Filtering location estimation algorithms	75
7.2.1 Summary of the location estimation algorithms	78
8 Discussion & Future Work	81
8.1 Discussion	81
8.1.1 Data Analysis	82
8.1.2 Radio map	82
8.1.3 Indoor Positioning	82
8.2 Future Work	82
References	85
A Appendix	93

Preface

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

Uncertainty is endemic in the problems related to decision analysis. It can range from taking a personal choice to making policy decision for multi million dollar company. In that sense, this concept is extremely innate to the field of Machine Learning which involves prediction having grave consequences.

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

One significant example how powerful this methodology is finding the wreckage of *Air France Flight AF447* (Stone et al., 2014), only a week after undersea search was deployed. The analysis not only included the flight data but also included the failures of the previous methods to correctly estimate the location of the wreckage. Subsequently, it was also used to locate the *Malaysian Airliner MH370* (Davey et al., 2016).

It is evident that *Bayesian Inference* is a powerful tool to combine knowledge from multiple sources. The above analysis particularly involved an iterative method called *Particle Filters*.

Symbols and Abbreviations

Symbols

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

Operators

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

Abbreviations

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

Chapter 1

Introduction

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

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

New age, new learnings.

Come twenty-first century, there are successful accomplished efforts to connect mankind in the outdoor space. The outdoor navigation systems includes the ubiquitous Global Positioning System (GPS) and regional systems like Russian Global Navigation Satellite System (GLONASS), DORIS, Galileo, BeiDou (now BedDou-2 or COMPASS), NAVIC (or IRNSS) and QZSS. With the explosion of usage of smart-phones helped the cause. !FIXME
rephrase this FIXME!

!FIXME move the discussion to IP and move to motivation. FIXME!

The peregrination for such a kind of connectivity has moved indoors. In the pursuit for cheap and reliable indoor positioning system, the research community has tried solving the problem using Wi-Fi, Zigbee, GSM mobile network signal, RADAR, etc., but now it is lingering upon Bluetooth Low Energy technology and it has the capability to become the de facto IoT communication device !FIXME **cite Umar Ahmad thesis** FIXME!.

1.1 Motivation

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

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

A scalable, seamless solution which equally performs well both indoors and outdoors is need of the hour.

The IPS opens up a great avenues for different Location based services (LBS) or Location Aware Services (LAS) like assets tracking, product flow optimization, product recommendation etc. which is would have market value of 10 billion US dollars by the year 2020 (He and Chan, 2016).

As most of the positioning solution are built on smartphones, it calls for effective usage of the sensor suite in the smartphones (Langlois et al., 2017).

Unlike technologies like Ultra Wide Band solutions, which need heavy investments on infrastructural installations, *Wireless Local Area Networks* and *Bluetooth* are existential and cheap.

There is an implicit requirement for reliable IPS (or LBS), as the error in location estimation can cause either locating wrong room at offices or wrong product in a super-store.

!FIXME **check Au Anthea thesis** FIXME!

As the technology is becoming more context and location aware for the world of Internet of Things, for sensor (device) connectivity the location is imperative for intelligent solutions. !FIXME **change this** FIXME!

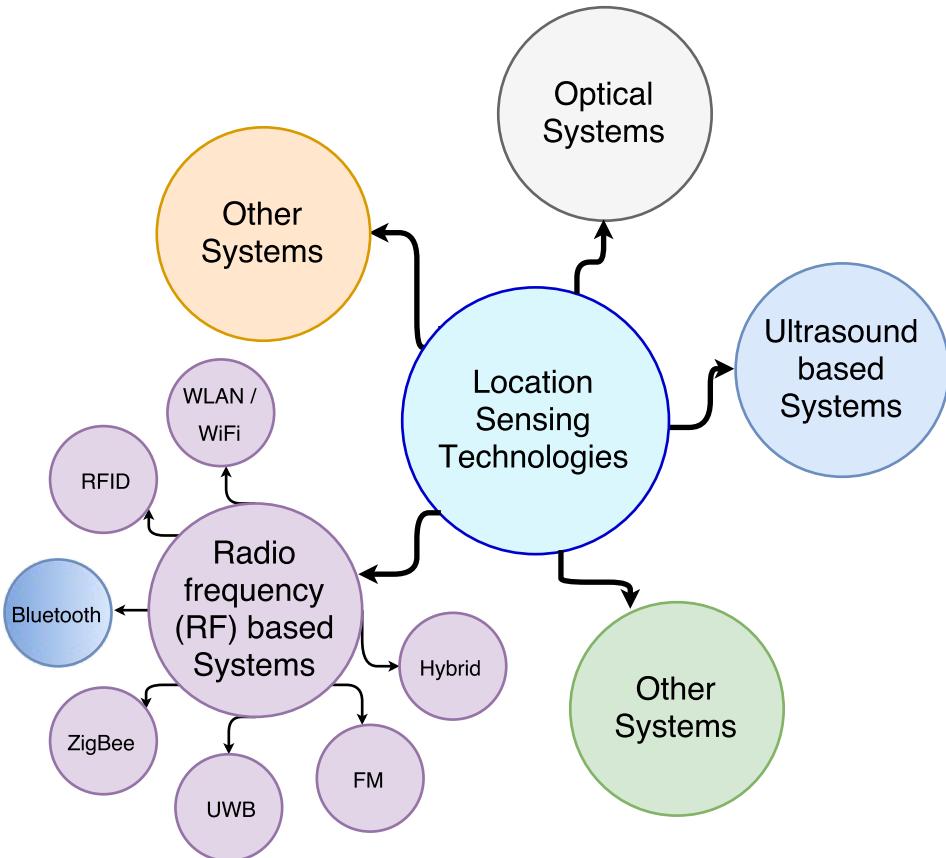


Figure 1.1: Technologies used in the Indoor Localization systems

FIXME use Umar's thesis. FIXME! The IPS opens up a great avenues for different Location based services (LBS) or Location Aware Services (LAS) like assets tracking, product flow optimization, product recommendation etc. which would have market value of 10 billion US dollars by the year 2020 (He and Chan, 2016).

1.2 Indoor Positioning

Over two decades now, the research in the field *Indoor Positioning (IP; or Indoor Localization)* has seen both infrastructure based and infrastructure free solutions (Hazas et al., 2004). Indoor positioning system (IPS) is either device aided (Liu, 2008a) or device-free (Patwari and Wilson, 2010) location sensing system which can accurately estimate the physical location of object (or person) indoors. These solutions are used for both research and

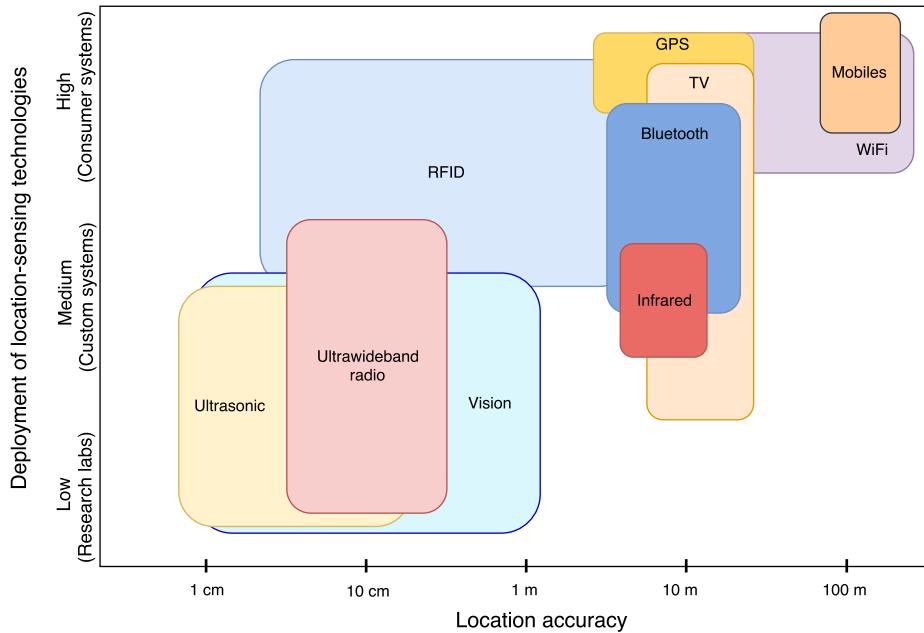


Figure 1.2: Technologies used in the Indoor Localization systems

commercial applications and to aid other location based services.

Indoor positioning systems in vague terms can be referred to as *indoor GPS*. Indoor environment can be office, hospital, shopping mall, school, airport, metro station etc and positioning makes much sense for larger arena. Research and development in IP would proliferate the location aware and location-based applications (Hazaras et al., 2004) making them a cornerstone in the world of IoT.

The Signals of opportunity: magnetic field, pressure, light, sound intensity, global system of mobiles (GSM) network signals, global network of satellite systems, etc. !FIXME **citation?** FIXME!

other signals: Wireless Fidelity, FM radio, radio-frequency identification (RFID), ultrasound or sound, light, magnetic field !FIXME **cite sources from he_chan_2016** FIXME!

Aim was not to add any extra hardware and get the IP and asset tracking applications possible.

!FIXME **copied from (De Luca et al., 2006) active vs passive devices** FIXME!

The generic localization procedure based on RF power measurements can be divided into two phases:

(1)the terminal measures the received power(s) of the signals transmitted by some devices used for localization;

(2) power samples are processed (somewhere) to estimate the position of

the terminal.

IoT and 5G.

!FIXME from my IP slides, get the source. and combine with above FIXME!
The different ways how different signal's are received and analyzed:

Active: Here the device generates signal's. For example, Asset Tracking.

Passive: Here the device receives the signal. The signals could also be categorized based on embedded information

1.3 Bluetooth Low Energy

Also called Bluetooth smart, BTLE, Bluetooth LE, or simply BLE Most the solutions require expensive infrastructure installations, and with the advent of beacon technology it has dragged all the attention towards itself mainly due to its low installation time and for being economical. (cite that report; print out taken)

Estimote's *nearables* is a new term for BLE plus the sensors.

1.4 Challenges Plaguing the RSS based methods

Going about solving this problem can be challenging as it is tagged by high non line-of-sight occurrences, effects of multiple obstacles, density and movement of human beings. This is exacerbated by signal fluctuations, channel interference and reflection leading to multipath, signal attenuation (Kaemarungsi and Krishnamurthy (2012), He and Chan (2016)). The problem of indoor positioning is inherently a challenging (Roos et al., 2002) and its 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 as it suffers from large-scale fading due to multipath, reflection, refraction (Hashemi, 1993) and small-scale fading due to dynamic nature of environment (Kaemarungsi and Krishnamurthy (2012), Luo et al. (2011a)). The problem exacerbated by co-channel interference given BLE radio signals follows 2.4 GHz ISM band (Hashemi, 1993). As the radio waves are readily absorbed by water, a single human can attenuate the signal by -3.5 dBm (Bahl et al., 2000).

The sweeping fluctuations in the RSSI measurements could be attributed to the low bandwidth and low transmission power of BLE protocol in contrast to WiFi. This makes BLE signals vulnerable to fast fading Faragher and Harle (2014).

1.5 Goal of the thesis

The goal of the thesis is to find an indoor positioning solution without adding extra hardware to the existing infrastructure. Using the already available Bluetooth modules¹ as the part of the luminaires for positioning. This study focuses on the methodology called location fingerprinting which capitalizes on the unique relationship between the received signal strength and the location.

1.6 Contribution of the thesis

The main contributions of this thesis is:

- Comprehensive data analysis for the signal strength pertaining to Bluetooth low energy beacons.
- An overview of Gaussian processes in the field of indoor positioning.
- Understanding of the flaws of other non-linear filtering methods in the aspect of BLE-IP.
- Rule of thumb for calibration point to luminaires in the calibration phase.

This thesis aims at creating a solution - which doesn't require necessary attention in terms of user or phone orientation.

1.7 Outline of the thesis

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

Chapter 2

Location Fingerprinting

!FIXME add the structure of the chapter FIXME!

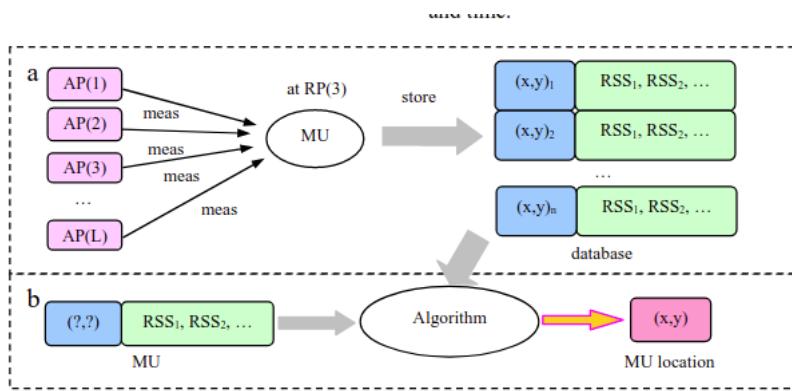
In this chapter, we delineate the background for the thesis with the relevant previous work. We start with detailed introduction of location fingerprinting and then move to preliminary concepts building on to the indoor positioning methods.

2.1 Location Fingerprinting

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

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

Being a laborious task, one of the few proposed reduced fingerprinting techniques like compressive sensing based *sparsity rank singular value decomposition (SRSVD)* in conjunction with k-Nearest Neighbour to compensate for the missing values and mitigate the redundancies caused due to multi-path and interference. Yiu et al. (2017) enumerates the effects of outdated radiomaps shows that the positioning accuracies are overestimated. They



show it by using realistic and unrealistic scenarios.

FIXME change it. create your own. FIXME!

Fingerprinting method has two phases: offline calibration phase and online localization phase.

1. *Calibration Phase:* It is an offline training phase where the *received signal strength indication (RSSI)* values of the radio-frequency signal from the *access points (APs)* at a particular *calibration point* for a certain time called *calibration time*. These measurements are called *fingerprints* (or signatures; Yiu et al. (2017)) and are location specific. This phase involves creating a *reference map* and in turn generating a *radiomap* using it.
2. *Localization Phase:* It is a online testing phase where the real-time signal strengths are recorded and location is estimated using an algorithm. This phase makes use of the radiomap generated in the calibration phase.

2.1.1 Radiomap

FIXME add radio-map 1 and 2 descriptions. FIXME! As discussed, the fingerprinting method involves learning the spatial properties of the signal in the calibration phase.

Radiomaps are the continuous interpolation of the signal strength values over the positioning space for a particular access point. There are other notable characteristics like *Signal to Noise ratio (SNR)* but the signal strength has stronger correlation to the distance. In fingerprinting technique, we are generally interested in the variation of RSSI values over the localization space,

hence, the units of RSSI is irrelevant¹. We discuss the relation of the signal strength and distance in detail in ??.

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 like a warehouse would need accuracy less than few centimeters but locating a room inside an office would allow us for error to few meters.

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

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

2.1.1.1 Reference Table

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

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

where N is number of calibration points, M is number of access points, P is location and access point dependent RSSI values and r_{min} is empirical minimum of all RSSI values in case an access point is not heard. It should be noted that due to indoor stochasticity the number of signal strength values vary for different access points and may vary in time even from the same access point. Yiu et al. (2016) suggests to use the minimum power device sensitivity level in case of unheard access point, but in practice the mobile unit can record a RSSI value below minimum sensitivity level.

It is also a common practice to use the mean estimate of all the RSSI values collected for a particular access point at a certain calibration point, but there is the downside of losing variance in the data. This is called the *mean reference table*.

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

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

Figure 2.1: Reference table generated as part of the fingerprinting technique. The first two columns represent the location in the form of *x coordinate* and *y coordinate* from a specific reference point. Depending on the number of access point the rest of columns are filled. !FIXME replace this with two columnned table. first columns showed mean estimated reference table. second non mean estimated reference table. FIXME!

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

!FIXME add the table for reference table. move rssi discussiob here FIXME!

2.1.2 Raw data

!FIXME check honkavirta thesis for more FIXME! Received Signal Strength Indication (RSSI) is a measurement of power present in the radio signal at the point of reception. RSSI is one of the mostly used parameters from the radio waves / signals for indoor positioning. It is widely used as a part of 802.11 standards. It is a pseudo location sensor and is the de facto measurement in the applications related to indoor positioning. As shown by Bahl and Padmanabhan (2000) that the signal strength shows proportional decrease with distance.

2.2 RSS based methods

The RSS based methods can be categorized as *Non-memory based* and *Memory based*. The non-memory based methods are a special case of memory based methods (refer figure ??). The various non-memory based methods

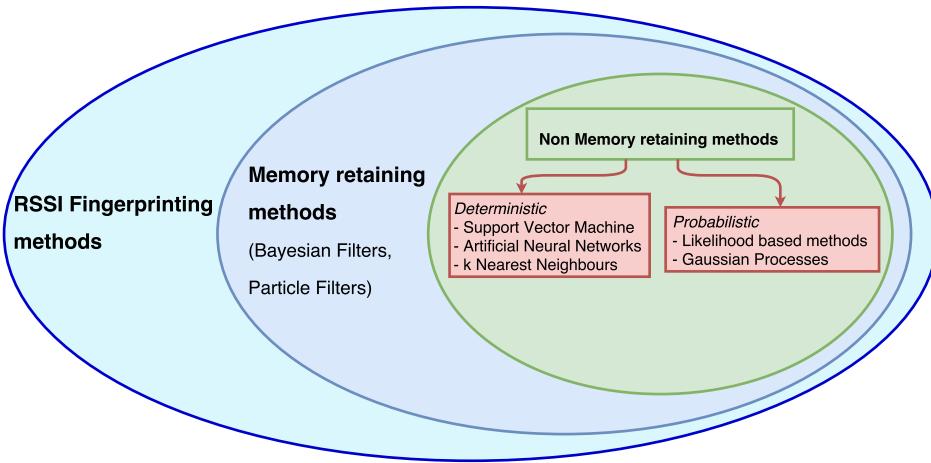


Figure 2.2: Illustration of received signal strength indication value decreases with increase in the distance of the luminaire

could be used as the measurement models which can efficiently capture the data generating process.

The non-memory based methods are range based (Aravecchia and Messelodi, 2014), parametric, deterministic methods where the RSSI is converted to a distance measure between the mobile unit and the access points. This measure is used for localizing the mobile unit. Various approaches have been employed for eg. like using polynomial curve to learn non-linearities in the RSSI (Feldmann et al., 2003), or by using space grids method with the mean of RSSI values (Elnahrawy et al. (2004), Luo et al. (2011b)), or by deploying the path loss model (Zanca et al. (2008), Pivato et al. (2011)). Bahl and Padmanabhan (2000) used *Cohen-Sutherland line clipping algorithm* (Foley, 1990) to obtain the accurate building layout information to estimate the radio propagation model.

The memory based methods are the range-free based (Aravecchia and Messelodi, 2014), stochastic methods where a likelihood model is used to compare the current RSSI measurements to predicted measurements from the previously built knowledge base in the training phase. The knowledge base can be either the reference table or radio maps, and the precision of measurements depends on the type of measurement model used. The predictions are computed over the parameter space, which could be a grid of points, or dispersed location particles.

The Bluetooth module present in the phone reads the signal and calculates the signal strength value. One noticeable feature of the RSSI values is that the values are quantized / discrete.

RSSI is designed for wireless communication and designing receiver an-

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

2.3 Positioning Algorithms

2.3.1 Non-memory based methods

The various non-memory based methods are described next.

2.3.1.1 k-Nearest Neighbour

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

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

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

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

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

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

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

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

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

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

The weights denote the reliability of the measurements as shorter distance gives accurate measurements. Any criteria for calculating the weights can be used, for example, Rizos et al. (2007) uses the inverse of the distance as weights.

2.3.2 Memory based methods

The probabilistic nature of our positioning solution would allow seamless integration of inertial sensor data for increased performance. Refer to Chapter 4 for detailed discussion on these methods.

2.4 Related Work

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

Accurate positioning systems required extensive auxiliary infrastructural services with sensors and time expensive calibration. Subsequently, the research community focused on radio based methods including WiFi, Blue-

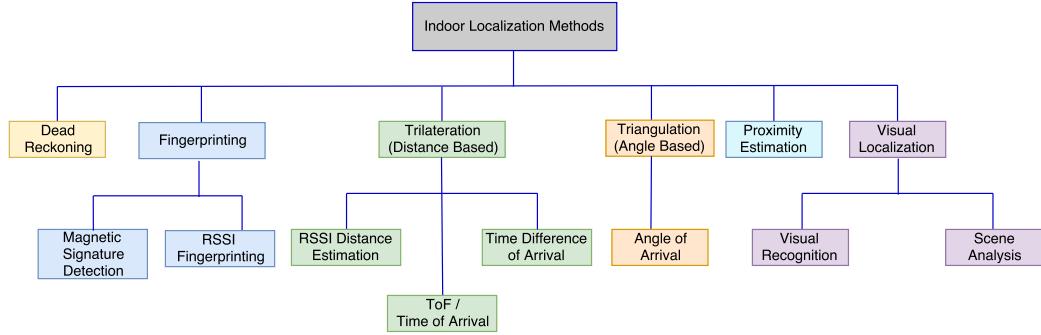


Figure 2.3: Taxonomy of Indoor Positioning

tooth, radio frequency identification tags and making the short range technologies solutions fine-grained for accurate (less than 30 cms) location estimation. But these technologies couldn't compete with wide coverage technologies like GPS, cellular networks, TV broadcasts.

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

2.4.1 Signals of Opportunity (SO_{Op})

!FIXME add the reference for SO_{Op} FIXME! Signals of opportunity (Curran et al., 2011) represent ubiquitous signals inherent in our working lives. They could be either due natural phenomenon or ubiquitous context aware technologies. These include magnetic field, pressure, light, sound intensity, cellular signals, GNSS, FM signals etc. Indoor positioning using SO_{Op} would be discussed here:

TV signals. (Hazaras et al., 2004)

2.4.2 Cellular Networks

2.4.2.1 Enabling Technologies

check Mainetti et al. (2014). get Graphical overview of the technologies enabling the indoor localization in dependence of accuracy and coverage; also in Umar Ahmad. Also the comparison between enabling technologies. Also Table II.

Categorize in terms of Signals of Opportunity.

2.5 Performance metrics

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

Chapter 3

Gaussian Process Model for Bayesian Filtering

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

- Radford M. Neal, 1996

Gaussian Processes (GPs) are non-parametric¹, stochastic process for describing distributions over functions (Rasmussen and Williams, 2005). Hence, we model directly over the functions and their statistical properties are inferred (Solin, 2016). The term 'process' originates from signal processing while 'random fields' is used in spatial statistics (Solin, 2016).

Definition 3.0.1. A Gaussian process is a collection of random variables, any finite (GPs) are non-parametric², stochastic process for describing distributions over functions (Rasmussen and Williams, 2005). Hence, we model directly over the functions and their statistical properties are inferred (Solin, 2016). The term 'process' originates from signal processing while 'random fields' is used in spatial statistics (Solin, 2016).e number of which have a joint Gaussian distribution (Rasmussen and Williams, 2005).

¹going by the definition of basic Gaussian process, mean of Gaussian is non-parametric, but the conditional distribution is Gaussian, i.e, parametric. Hence it can also be called semi-parametric methods.

²going by the definition of basic Gaussian process, mean of Gaussian is non-parametric, but the conditional distribution is Gaussian, i.e, parametric. Hence it can also be called semi-parametric methods.

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

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

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

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

GPs are generalization of multivariate Gaussian distribution (Rasmussen and Williams, 2005). Contrary to sampling from distributions which yields finite dimensional vectors, Gaussian processes yields infinite dimensional vectors of which any finite section would follow a multivariate Gaussian distribution. GPs can also be seen as Bayesian neural networks with neural network model as the prior over the unknown non-linear function and learning as the posterior distribution (Neal, 2012).

3.1 Motivation for using the GPs

Due to complexity of the indoor environment coupled with interference, multipath propagation of radio-frequency signals, obstacles, leading to distorted spatial distributions of the signal strength values. The simple parametric distributions are inadequate in modeling the complex RSS distributions (Seco et al., 2010). Hence, we need flexible models for tackling this problem and Gaussian processes are perfect solution for that.

3.2 Advantages of GPs

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

1. *Continuous Locations:* Traditionally, GPs were known as *kriging*, which is an regression task, hence they can predict the signal strength values (with the uncertainty estimates). GPs have excellent capabilities of interpolating over other test locations. They are flexible as they don't need any designated training points for accomplishing this task.

2. *Arbitrary likelihood models*: A wide variety of complex data models can be approximated given the non-parametric nature of GPs: multiple kernels could be used in conjunction with each other Rasmussen and Williams (2005). Hence, GPs are can model highly non-linear signals such as RSSI (Aravecchia and Messelodi, 2014).
3. *Correct uncertainty handling*: As the GPs come with a Bayesian flavor, along with the mean estimates they also spit out the uncertainty estimates for each value in the state space. This is mainly dependent on amount of data and the estimated noise around the test points ?.
4. *Consistent parameter estimation*: The model selection problem in GPs helps solving the obtaining the optimal (hyper-) parameters. This is done via the maximizing the marginal likelihood ?. These point to spatial correlation between measurements and learn the measurement noise?.

3.3 GP modeling for indoor positioning

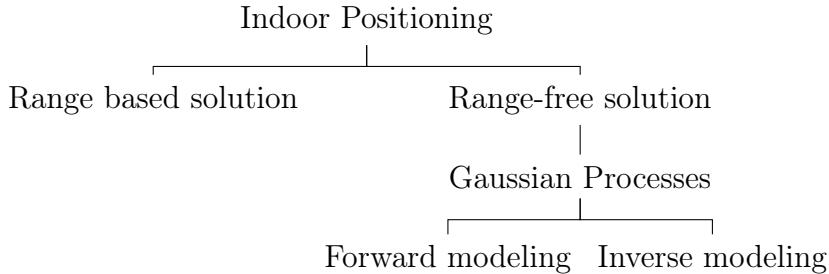


Figure 3.1

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

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

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

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

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

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

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* Schwaighofer et al. (2004) modeling by Schwaighofer et al is quite misleading.

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

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

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 could be modeled directly from signal space to metric space.

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

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

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

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

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

!FIXME change it; for reference only. FIXME!

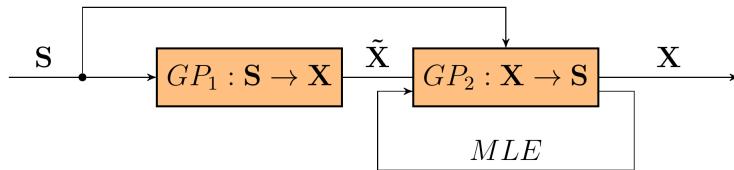


Figure 3.2: Hybrid model. courtesy Aravecchia and Messelodi (2014).

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

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

!FIXME send me to chapter 2. FIXME! GPs were first used by Schwaighofer et al. (2004) and later on by Li et al. (2005) as error correction map for the non-line-of-sight problem for positioning in the cellular networks.

Image based localization using GP's.

Schüssel and Pregizer (2015) talks about hybrid method i.e., log path loss model as the mean function in the GP's **!FIXME similar to Nokia paper** FIXME!.

!FIXME copied; rephrase FIXME!

The authors of Atia proposed a GPS-Like zero-configuration indoor positioning system based on received signal strength (RSS) of the popular WiFi network as shown in Figure 9. The proposed system does not require a time-consuming offline radio survey prior knowledge about the area or new hardware unlike current RSS-based indoor systems. Similar to GPS, the proposed system consists of three sections as shown in Figure 9: network segment(WiFi), control segment, and user segment. Between network segment and control segment, RSS observations are exchanged periodically. The control segment uses a novel hybrid propagation modeling (PM) technique using logarithmic decay model augmented by a nonlinear Gaussian process regression (GPR) that models RSS residuals that cannot be modeled by the traditional logarithmic decay models indoors. The proposed system provides 2-3 m accuracy in indoor environments.

Schwaighofer et al. (2004) solved the localization using GP's and kNN algorithm in DETC cellular networks. The authors showed that GP's gave an accuracy of 7.5 m, bit worse than kNN's 7 m. Matérn covariance function was used for GP modeling of individual stations and maximum likelihood was used for estimation of location.

by the IMU process model in (2), is a linear Gaussian process Li et al. (2013)

Bisio et al. (2017) latest GP paper.

!FIXME copied from Indoor Location with Wi-Fi Fingerprinting Noah
Pritt FIXME!

WiFi-SLAM employs Gaussian process latent variable models to build an indoor map as the user moves throughout a building [20]. Unlike RADAR and Horus, it does not require a calibration stage. However, it only works with very simple, rectangular floor plans and makes restrictive assumptions about user movements. It also requires integration with additional sensors to determine the direction and distance of movement.

Chapter 4

Memory Based Filtering Methods

!FIXME change it. create your own. FIXME!

4.1 Introduction

? has justification for k-means. the orientation if the error vector is different for each beacon. Hence, averaging leads us closer to the true location.

The memory based methods in this thesis are the non-linear Bayesian filtering methods: particle filter and unscented Kalman filter.

4.2 Sequential Monte Carlo

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

Bayesian Filtering forms a case of SMC method used in the field of state estimation.

Particle filters are the SMC methods

4.3 Bayesian Filtering

!FIXME may be move this to introduction FIXME!

Most of the real-world problems circumscribe evaluating an unknown quantity given some measurements. Often, sufficient prior information about the data generating and dynamic process is available. This a classic setup for

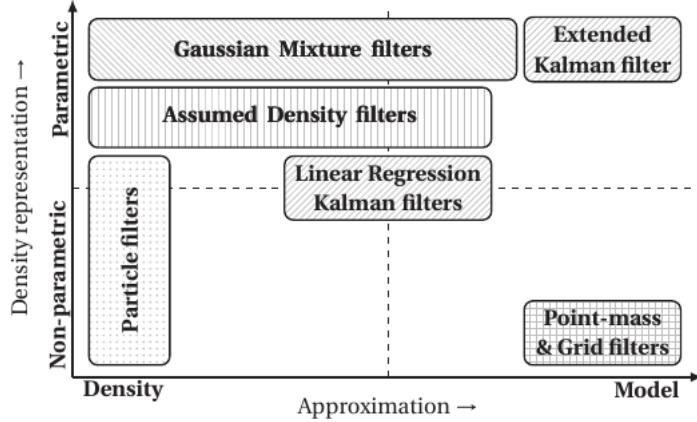


Figure 1.2: Popular approximate nonlinear filtering approaches.

using Bayesian Inference and forms another view for optimal filtering (refer Särkkä (2013)). It involves incorporating the initial notion of the unknown quantity giving rise to updated belief about the quantity. Precisely, the prior distribution is updated in the light of new evidence to give rise to posterior distribution which allows for inference on the quantity of interest (for eg. location of person). For comprehensive overview on Bayesian inference, refer (Gelman et al., 2014).

In the field of *sensor informatics* and *time varying systems*, the data arrives sequentially and the marginal posterior needs to be updated simultaneously. This Bayesian recursive state estimation can be termed as *Bayesian Filtering* (Särkkä, 2013). Bayesian Filtering resembles the statistical inversion problem (Särkkä, 2013) with states as latent variables and measurements as observed variables, figure ???. 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 of navigation, telecommunication, economics, etc.

4.3.1 State Space Modeling

We consider the following state space model

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

Restricting ourselves to Markovian, non-linear and non-Gaussian scenario, (Doucet et al., 2001), the unobserved states (hidden or latent vari-

ables) are $\{\mathbf{x}_t; t \in \mathbb{N}\}$, $\mathbf{x} \in \mathbb{R}^n$ are modeled as a *Markov process*. Using the initial distribution $p(\mathbf{x}_0)$, transition equation $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ and marginal distribution $p(\mathbf{y}_t|\mathbf{x}_t)$ given the noise riddled observed measurements $\mathbf{y} \in \mathbb{R}^m$, we write our model as

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

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

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

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

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

where

- $p(\mathbf{x}_{0:t})$, is the dynamic model which forms the prior distribution,
- $p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})$, is the measurement model¹ which forms the likelihood function,
- $p(\mathbf{y}_{1:t})$, is the evidence and is a normalizing constant.

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

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

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

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

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

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

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

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

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

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

If assumed Gaussian state-space model, the filtering solution gives rise to the optimal *Kalman filter* which has a neat closed form analytical expression (Särkkä (2013), Doucet et al. (2001)). This assumption at times doesn't hold as the empirical data suffer the recursive equation lets consider a new observation y_{t+1} , hence, the updated joint posterior isrs from non-Gaussianity, high dimensionality and non-linearity (Doucet et al., 2001) which requires filtering techniques with statistical workarounds and posterior approximations. These methods include extended and unscented Kalman filters, Gaussian filter, Gauss-Hermite Kalman filter, Cubature Kalman filter and Particle filters (Särkkä, 2013).

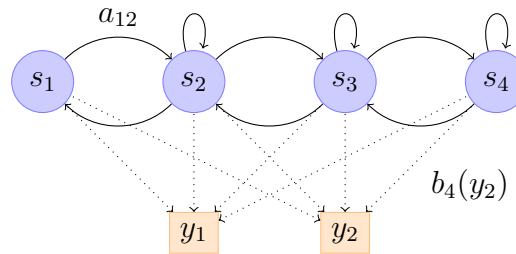


Figure 4.1: edit it similar to Sarkka figure 1.5

Based on the previous literature, we select only promising filtering

4.3.2 Advantages of Bayesian Filtering

One major advantage is that the solution degrades gracefully.

4.4 Filtering

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

In this chapter we will describe the probabilistic methods used in the thesis. We will start with dynamic models used for the filtering models.

4.5 Particle Filters

4.5.1 Importance Sampling

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

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

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

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

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

where the weights w_t are defined as:

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

Now, the approximate filtering distribution can be written as:

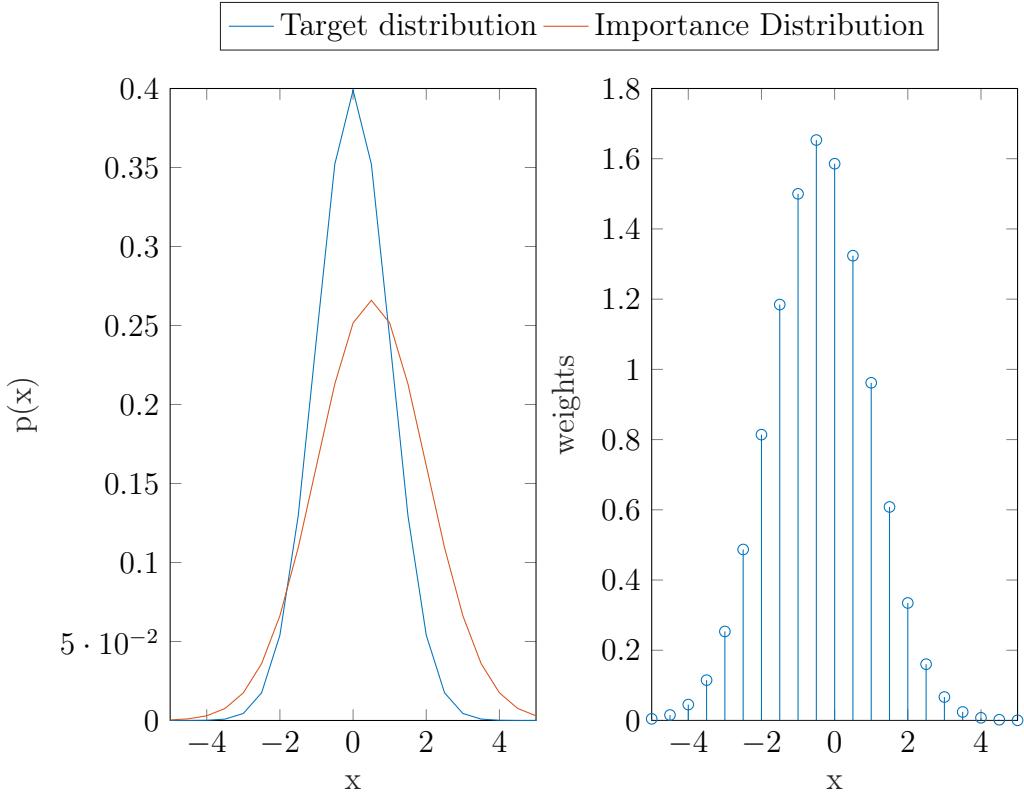


Figure 4.2: (left) The importance distribution and target distribution. (right) Approximate posterior distribution in the form of weights.

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

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

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

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

Using equation 4.13 in equation 4.9, we arrive at

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

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

4.5.2 Filtering

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

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

Effective sample size is a way of determining the exact samples from the target distribution. It interprets the number of particles which effectively contributes to the estimation of the state and shows the efficiency of the estimation (Martino et al., 2017). Mathematically, it is a ratio estimate (Kong, 1992) but it can also be explained as a discrepancy measure i.e., the euclidean distance between the probability mass function of the normalized weights to its discrete uniform probability mass function (Martino et al., 2016). A detailed discussion on alternative S_{eff} criteria are discussed in Martino et al. (2017).

We have summarized the resampling algorithm in algorithm 2.

Algorithm 1: Importance Sampling

- 1 Sample from the importance distribution:

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

- 2 Calculate the unnormalized weights w_t :

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

- 3 Normalize the weights:

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

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

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

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

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

.

Algorithm 2: Resampling

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

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

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

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

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

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

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

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

The importance distribution can be conveniently split as:

Algorithm 3: Particle Filter (Sequential Importance Resampling)

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

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

Using equation 4.18 in the equation 4.17, we get

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

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

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

4.6 Unscented Kalman Filters

4.6.1 Unscented Transform

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

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

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

We follow the following procedure for forming the approximation:

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

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

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

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

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 function $g(\cdot)$:

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

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

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

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

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

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

4.6.2 Filtering

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

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

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

Algorithm 4: Unscented Kalman Filter

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

2 *Prediction:*

- Get the sigma points:

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

where n is state dimension and λ is defined in (4.22).

- Transform the sigma points using the dynamic model:

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

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

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

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

3 *Update:*

- Get the sigma points:

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

3

- Transform the sigma points using the measurement model:

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

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

$$\boldsymbol{\mu}_t = \sum_{i=0}^{2n} W_i^{(m)} \hat{\mathcal{Y}}_t^{(i)},$$

$$\mathbf{S}_t = \sum_{i=0}^{2n} W_i^{(c)} (\hat{\mathcal{Y}}_t^{(i)} - \boldsymbol{\mu}_t) (\hat{\mathcal{Y}}_t^{(i)} - \boldsymbol{\mu}_t)^T + \mathbf{R}_t,$$

$$\mathbf{C}_t = \sum_{i=0}^{2n} W_i^{(c)} (\mathcal{X}_t^{(i)} - \boldsymbol{\mu}_t^-) (\hat{\mathcal{Y}}_t^{(i)} - \boldsymbol{\mu}_t)^T.$$

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 :

$$\mathbf{K}_t = \mathbf{C}_t \mathbf{S}_t^{-1},$$

$$\mathbf{m}_t = \mathbf{m}_t^- + \mathbf{K}_t [\mathbf{y}_t - \boldsymbol{\mu}_t],$$

$$\mathbf{P}_t = \mathbf{P}_t^- - \mathbf{K}_t \mathbf{S}_t \mathbf{K}_t^T.$$

5 end

Chapter 5

Measurement Setup & Data Collection

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

The minimum data was thresholded to -110 dBm.

FIXME check the dimensions of the floor and add beacons location. FIXME!

5.1 Experimental Testbed

Our experiment area is an office space located on fourth floor at Helvar office. The area is a vacant space with luminaires fitted with the beacons. The stem part of L-shaped floored space has the dimensions is 32 meters by 5 meters and the leg part is 6 meters by 3 meters in dimension, refer the Figures 5.1 and 5.6a. For simplicity and accurate results, the movement of personnels during the experiments were from minimal to zero. The low right corner in the Figure 5.6a is considered the origin and the locations mentioned are with respect to this.

5.2 Measurement Configuration

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

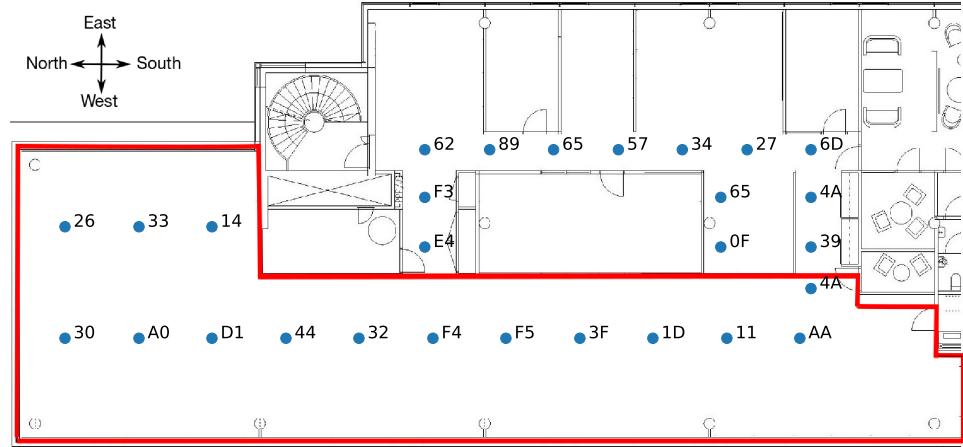


Figure 5.1: Floor plan of Helvar's R&D section in Keilaniemi office with the location of the beacons.

5.3 Bluetooth Low Energy

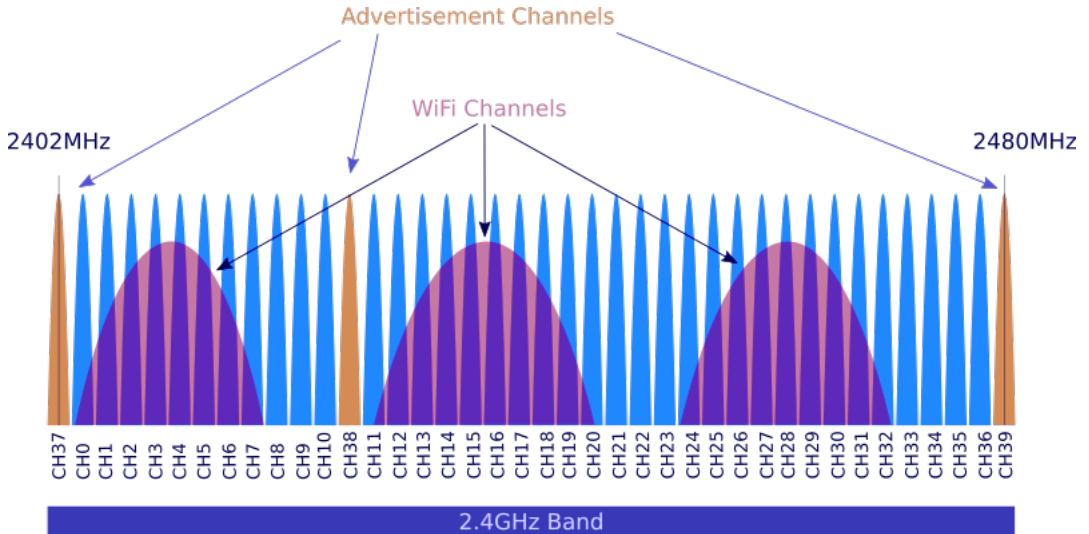


Figure 5.2: Channels in Bluetooth low energy.

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

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

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

The BLE devices are available in market with the names *beacons*,

!FIXME what does the beacon packet have? FIXME!

5.3.1 Physical Layer

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

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

5.3.2 Advertisement channels

The channels 37, 38 and 39 are the advertisement channels. These radio channels are strategically placed to avoid interference from the WiFi (see figure 7.6). The advertisement channels are critical as they are responsible for making a connection to a mobile devices (like smartphones, smartwatches, etc.) and three channels are used to increase the probability of central device reading the advertisement packets. This mode serves for uni-directional communication. The BLE technology also gives the option of masking any channel.

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

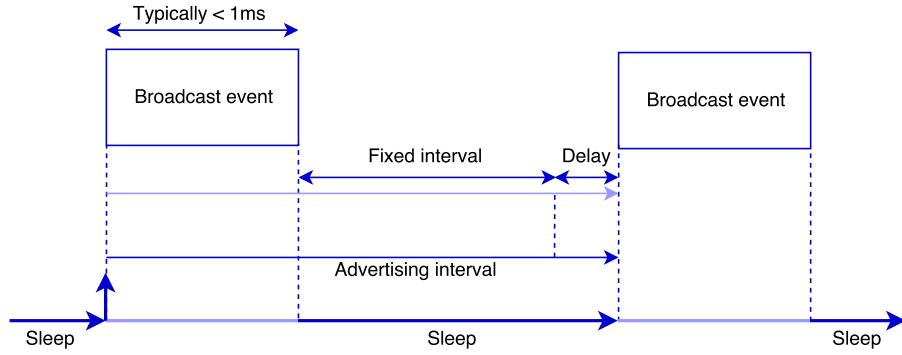


Figure 5.3: Advertisement in BLE

5.3.3 Data channels

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

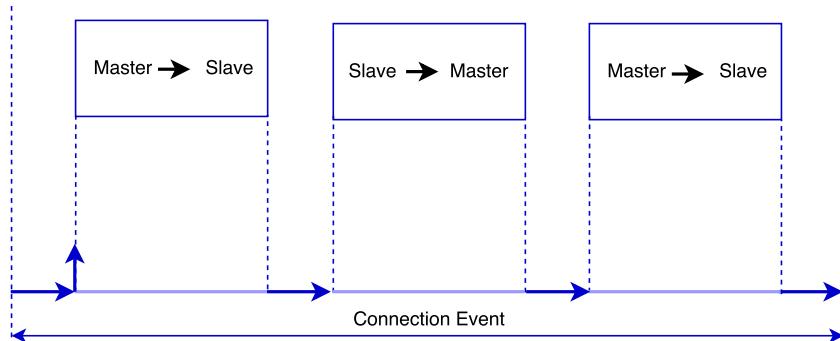


Figure 5.4: Data communication in BLE

5.3.4 Different BLE protocols

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

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

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

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

¹

Kriz et al. (2016) check references 18–21.

5.4 Assumptions on the RSSI values

We assume that in-luminaire beacons are omni-directional with the luminaires always being located over the head of the user. The interference of in-luminaire working and other radio signal is consider negligible. The localization algorithm aims at point estimates rather than area estimates.

5.5 Radio Analyzer

The Frontline Sodera LE !FIXME **cite it** !FIXME! radio-analyzer was used for reading the RSSI values. The radio-analyzer accurately reads the signal strength values and gives additional data like channel information, radio channel information !FIXME **correct it** !FIXME!, universally unique identifier (UUID) and data type.

Kaemarungsi and Krishnamurthy (2012) discussion on different devices to record RSSI values.

Discussion on the various ways to measure RSSI values. like in Kaemarungsi and Krishnamurthy (2012)

5.6 Measurement Application

The RSSI signal strength measurement application was built using the AltBeacon Android library (Networks, 2017). This library incorporates extra features like foreground and background interval. The in-l beacons we used didn't follow the either the iBeacon (Newman, 2014) or Eddystone (Eddy-stone, 2016) protocol and the library allows reading the radio signals from

¹Technically the hardware is wireless access points which can adverstise, but we also use the term *beacons*.

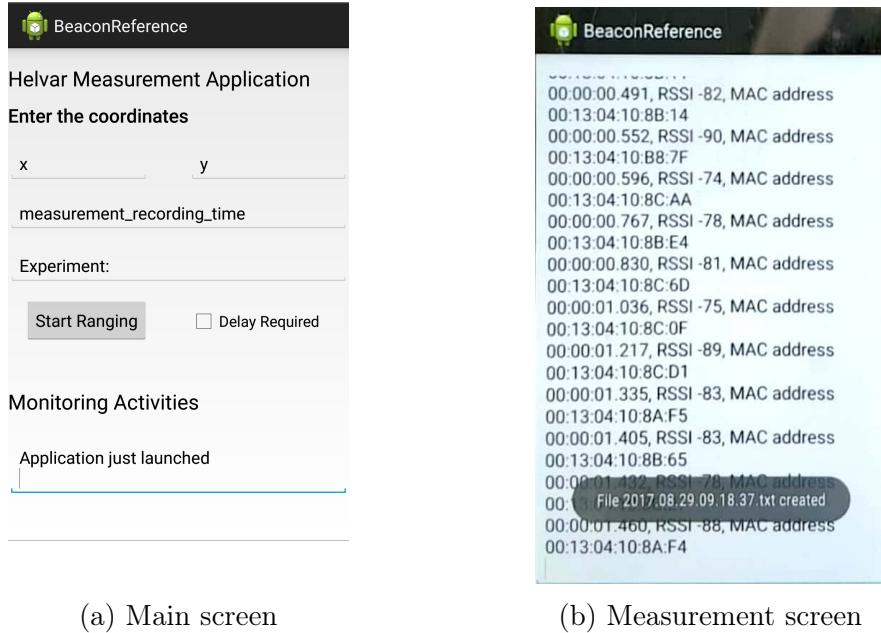


Figure 5.5: Caption place holder

non-standard advertising beacons. But, these standards could be incorporated in to the in-luminaire beacons.

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

5.6.1 Smartphones

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



(a) measurement arena



(b) Tripod with mobile holder for user free measurements at one of the measurement landmarks.

Figure 5.6: measurement setup

5.7 Data Collection

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

5.7.1 For Data Analysis

For better understanding the data generating process we collect the data from different smartphone and human orientation to better construct the measurement model in the Bayesian filtering context. The different user orientations will be discussed in the following section.

5.7.1.1 Different User Orientations

The main idea here is to cover different aspects of the the user directions while recording the measurements from the WAP's. In this thesis, we recorded the user measurements in the directions north, east, west and south (for directions refer Figure 5.1). For the user measurements, the user held the smartphone device at a 45°angle, with head parallel to the screen of the phone (Hansraj, 2014). We refer this as *standard usage configuration*. By observation, this most common way of usage when the user is walking. With the same configuration, we recorded the shadow and rotate measurements. In the shadow mode, the user stands in between the signal generating beacon and the smartphone shadowing the measurements cutting the line of sight to the beacon. In the rotation mode (Honkavirta, 2008), the user rotates while recording the measurements.

5.7.1.2 Different Phone Orientations

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

5.7.1.3 Outdoor Measurements

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

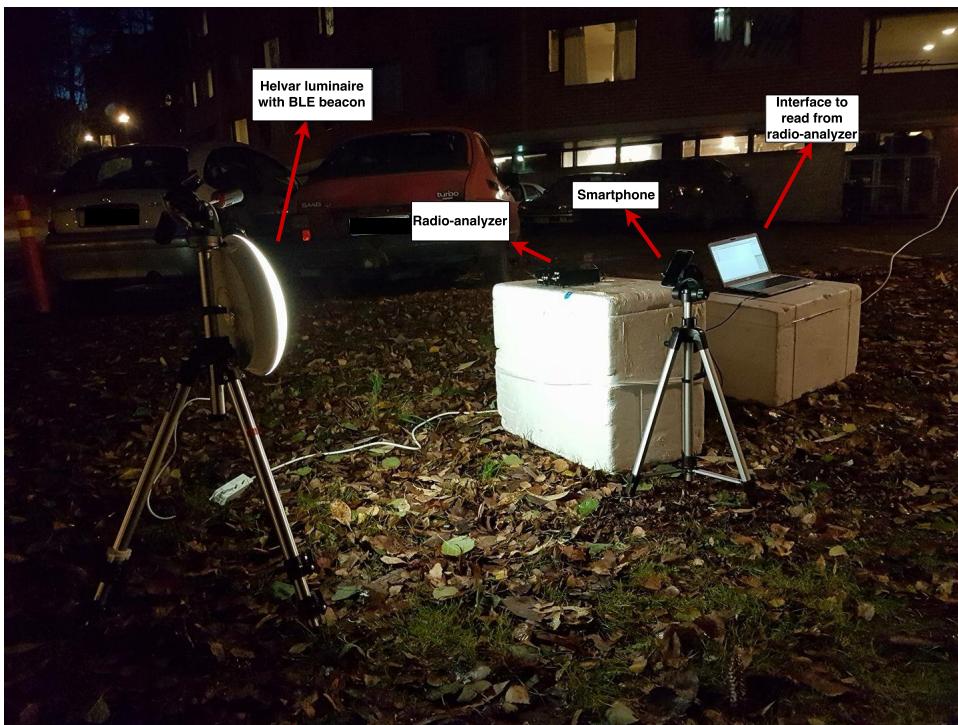


Figure 5.7: Measuring the RSSI values outdoors using the smartphone and radio analyzer.

FIXME update this section... FIXME!

5.7.2 For Positional Algorithms

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

5.7.2.1 Setup for Calibration Phase and Test Phase

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

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

5.7.2.2 Calibration Data

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

5.7.2.3 Test data

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

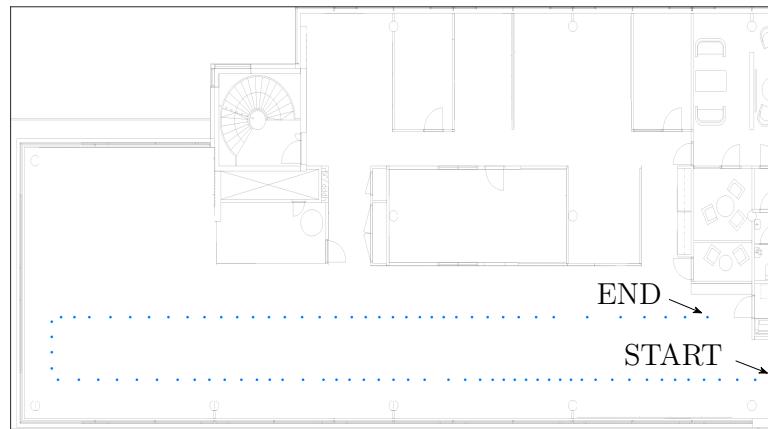


Figure 5.8: Test data from Helvar R&D

Chapter 6

Data Analysis of RSSI Values

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

- Amy Poehler

Before we start exploring the different methods for solving the indoor positioning problem, the initial challenge lies in getting the right measurement (data) model as the RSSI values vary due to various factors like human beings, type of smart-phone, orientation of the phone and algorithm related factors like calibration time, calibration points and other factors like malfunctioning bluetooth modules. [FixMe: add factors here or refer to previous sections].

The Figure !FIXME get the 4.5 (a) of Honkavirta FIXME! shows the number of AP's heard per second. !FIXME write about the figure introduced before FIXME!. !FIXME move to chapter 6 data analysis FIXME!

The data analysis of RSSI values is vital to the understanding of the location dependent features in fingerprinting methodology and hence designing an accurate indoor positioning algorithm. The knowledge of the statistical properties of the RSSI's can later help in the efficient modeling of the measurement model which would in-turn lead to better localization performance.

RSS analysis based on time not done here. !FIXME add me FIXME!

The use of RSS indication was proposed in the pioneering work by ? where the wireless local network (WLAN) was used.

Based on our review, there was no in depth analysis of statistical properties of RSSI values from the bluetooth beacons.

The *initial data analysis*¹ is the primary task in statistical analysis and modeling which yields critical statistical properties about data generating process. It veers us towards finding the right solution i.e., finding the right

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

data model for our problem and learn the peculiarities in the data. It helps to efficiently design and analyze an IPS. A comprehensive data analysis of RSSI data from the beacons is hardly done² in previous literature with most of the researchers focusing on the algorithms. In this chapter, we take closer look at the data and try to find different factors which affect its variation. We take into consideration the factors such as user's presence, smart-phones and orientation of smart-phones and material of the luminaire. First, in 6.3, we look at the raw RSSI data. Next, ... !FIXME **add the details about the chapter** FIXME! A detailed description about RSSI is given in chapter ??.

It was taken care that the measurements were taken in close successions.

6.1 RSSI as a measure of distance

The vital proposition to the thesis is that RSSI is a valid measure convenient for solving the indoor positioning problem. Similar to Bahl and Padmanabhan (2000), in figure 6.1 we show that the RSSI is a realistic criteria with the measure inversely correlating with the distance. The experiment included the tripod held mobile unit which recorded the signal strength for a particular period of time. Evidently, conforming to the common knowledge the signal strength at the beacon closest to the mobile unit has the highest signal strength and reduces as the beacon is farther away.

!FIXME change the x label to "Distance of luminaires" for rssi vs distance plot or add the luminaire numbers?; add from other phones as well. FIXME!

6.2 Statistical Hypothesis Testing

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

The hypothesis testing procedure goes as follows:

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

²for WiFi, check Kaemarungsi and Krishnamurthy (2012)

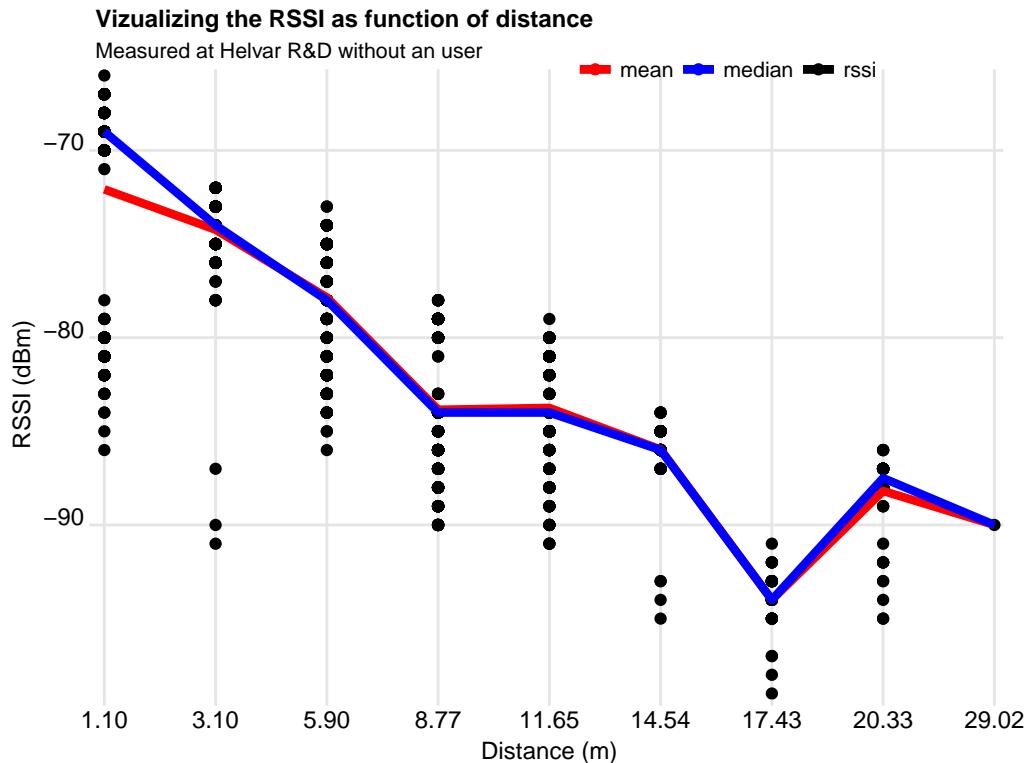


Figure 6.1: Illustration of received signal strength indication value decreases with increase in the distance of the luminaire

4. Determine the test statistic and infer from the results.

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

The critical values are the threshold from region of acceptance to region of rejection, also sometimes referred to as *critical region*.

6.2.1 Two Sample Kolmogorov-Smirnov test

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

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

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

$$D = \sup_x |\mathcal{F}_1(x) - \mathcal{F}_2(x)| \quad (6.1)$$

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

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

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

6.2.2 Two Sample Z-test

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

1. $H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2,$
2. $H_0 : \mu_1 \geq \mu_2 \quad H_1 : \mu_1 < \mu_2,$
3. $H_0 : \mu_1 \leq \mu_2 \quad H_1 : \mu_1 > \mu_2.$

As the data is greater than 30, we use the Z test in opposed to t test and *central limit theorem* allows us to use the equation 6.2 and adds robustness. We assume that the samples are drawn from normal distribution with equal but unknown variances and null hypothesis being that the means are equal. Consider two samples with sample means \bar{x}_1 and \bar{x}_2 , the test statistic z is given by

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (6.2)$$

where the μ_1, μ_2 with subscript zero are hypothesized parameters, σ_1^2, σ_2^2 are known variances, if unknown sample variance are calculated and n_1, n_2 are number of data points in each corresponding samples. With the *level of significance* α as 0.05, and assumed variance of 10 and equal known means (i.e., $\mu_1 = \mu_2$) for the samples, the region of acceptance lies in between ± 6.1980 .

6.2.3 Levene's test

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

$$\begin{aligned} H_0 : \sigma_1^2 &= \sigma_2^2 \dots = \sigma_n^2 \\ H_1 : \sigma_i^2 &\neq \sigma_j^2, \quad \exists(i, j). \end{aligned}$$

The test statistic is defined as

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

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

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

$Z_{i.}$ are the group means of the Z_{ij} and $\bar{Z}_{..}$ is the overall mean of Z_{ij} . !FIXME
change the above line FIXME!

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

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

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

One way to infer about the variances of two distribution is to use its ratio estimate i.e., σ_1^2/σ_2^2 (Daniel, 1999). The *F distribution*, given its definition, conveniently allows us to conclude if the two distributions have the same variance or not. Given two independent samples, assumed to be drawn from different normal distributions, then a random variable X ,

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

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

To find the confidence interval for the interval estimate σ_1^2/σ_2^2 given a significance level α , we use the following expression

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

which can be re-written as

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

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

6.2.4 Wilcoxon rank-sum test

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

Table 6.1: A 2×2 Contingency table

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

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

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

The assumptions in the test given the ordinal, unequal size 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
3. the parameter being inferred is continuous.

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

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

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

For the null hypothesis to be true X^2 needs to be approximately χ^2 distributed with 1 degree of freedom. It implies that computed X^2 should be less than 3.841 (for $\alpha = 0.05$).

6.2.5 Experiments

The main aim of these experiments was to find the bias of different factors, hence, it was taken care to change only single factor keeping the other factors

unchanged. Measurements were taken with one phone at a time to avoid any kind of unknown interference. We accomplish the statistical hypothesis testing methods discussed in section 6.2.

6.2.5.1 Bias due to user's presence

In this experiment, we investigated the effect of user's presence on the recorded RSSI values. The smartphone Samsung S7 phone was selected and experiments were conducted with that. It was assumed that when collecting the data without the user, the direction of the phone had minimal bias on the RSSI value. Hence, a random direction was chosen and this direction was consistent over all the experiments with different smartphones. In addition, we also checked for the shadowing effect, that is, when the bias when the user completely blocks the signals from a certain beacon.

From the location of measurement, we selected seven beacons with increasing distance to observe the effects.

6.2.6 Bias due to different smartphone

In this experiment, the bias due to difference in the hardware of the smartphones. It is a user free experiment where we recorded the measurements over the phones at 45°(the normal human holding position). A random direction in relation to the room was chosen and the it was kept consistent over the different smartphone measurements.

6.2.6.1 Orientation of Smart phone

!FIXME add the inferences from the hypothesis testing results. FIXME! We recorded the RSSI values for the smartphones over different orientation of the phones. We chose different orientations like for example, 0°, 45°and 90°using the S4 mini smartphone.

In the experiment with 45°

In the second experiment 90°orientation, the smartphone wasn't able to capture the properly beyond 8.53 meters

From the figures, the *hand vs pocket* experiment doesn't show significant difference in the RSSI values.

6.2.6.2 Material of the Luminaires

In this experiment, we investigate if the material of the luminaire adds any bias to the signal strength values from the BLE access point. In order to eliminate any other factor affecting the measurements, a single smartphone

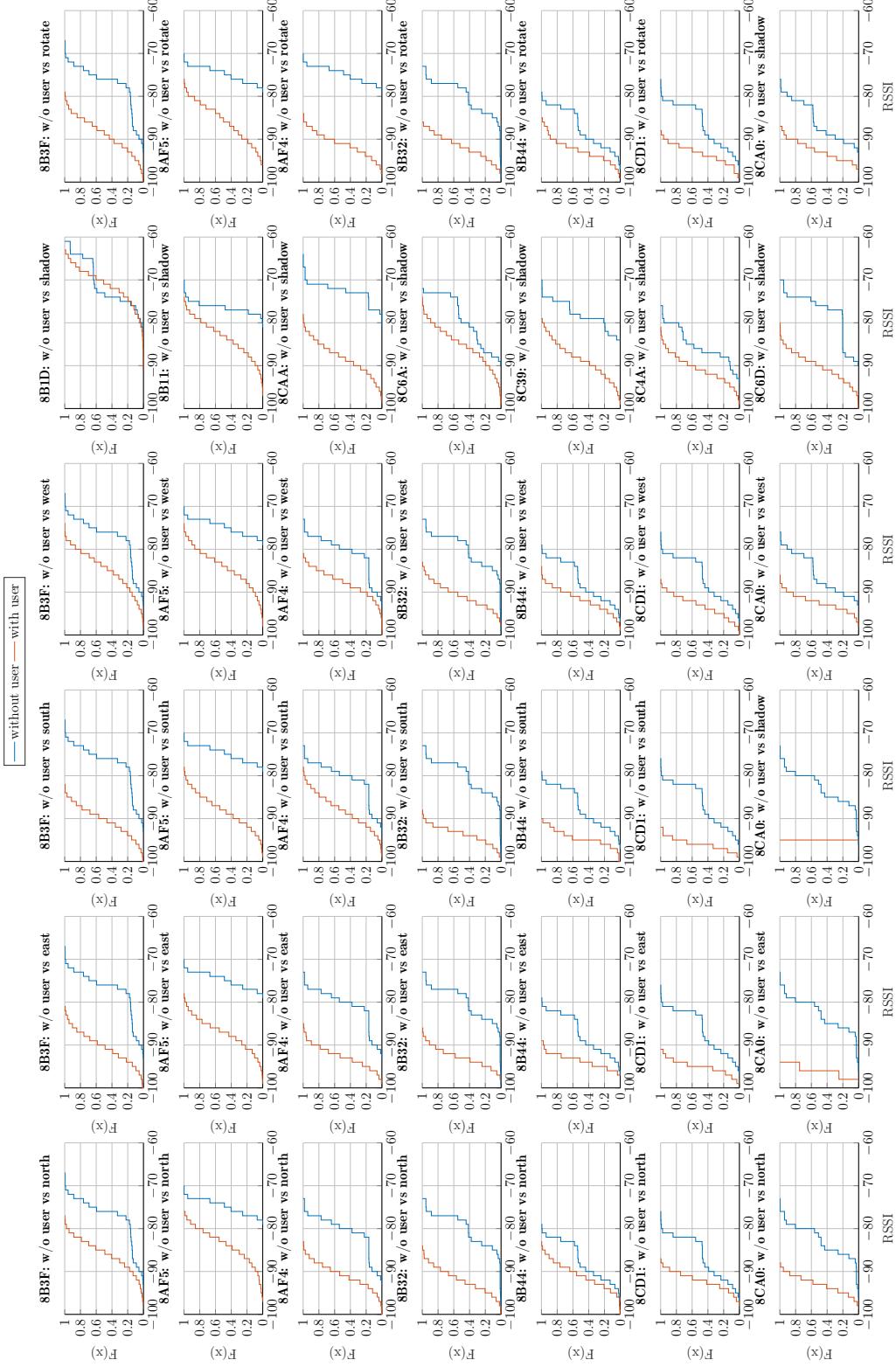


Figure 6.2: Experiment 1: Helvar user presence

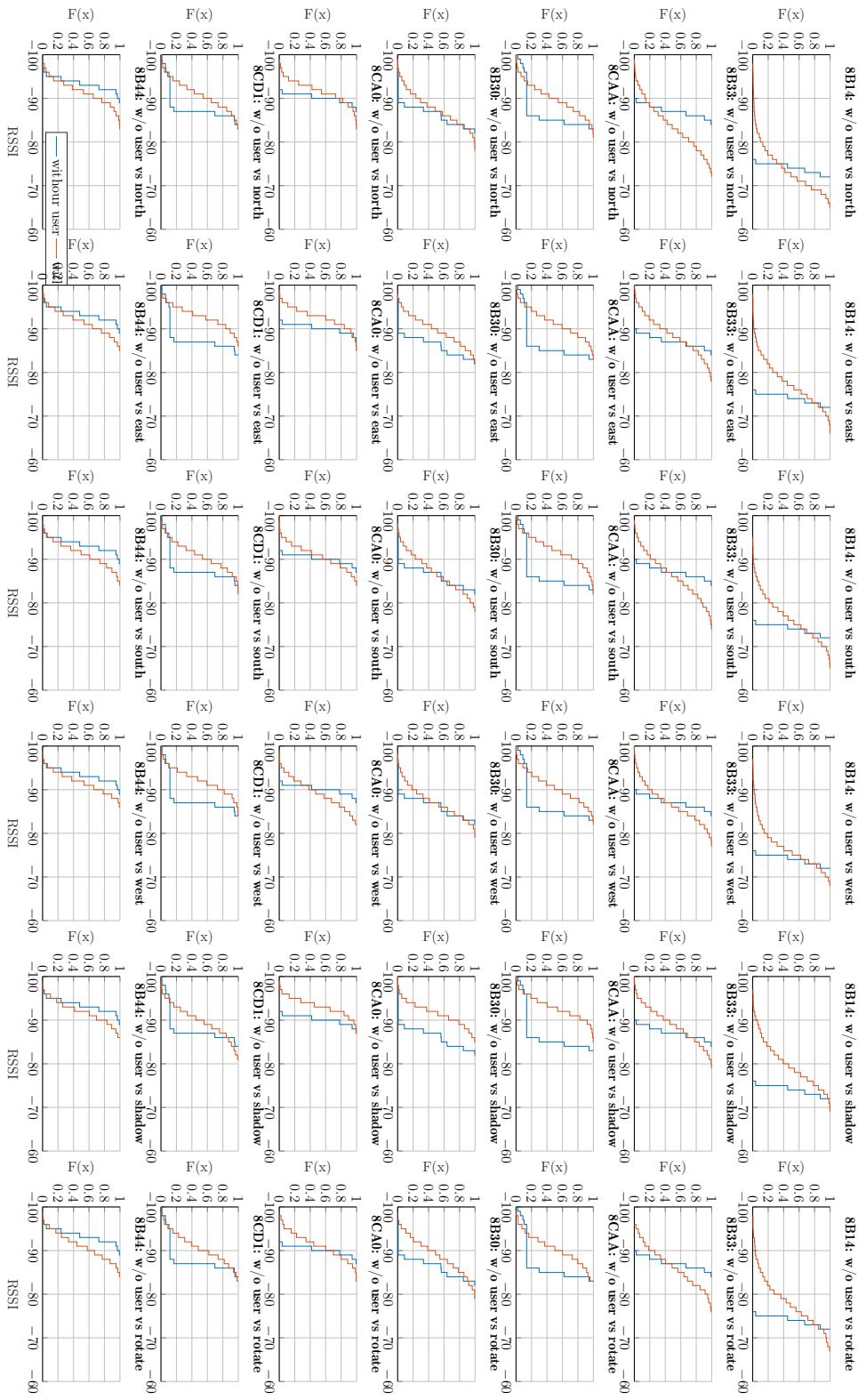


Figure 6.3: Helvar User presence experiment 2.

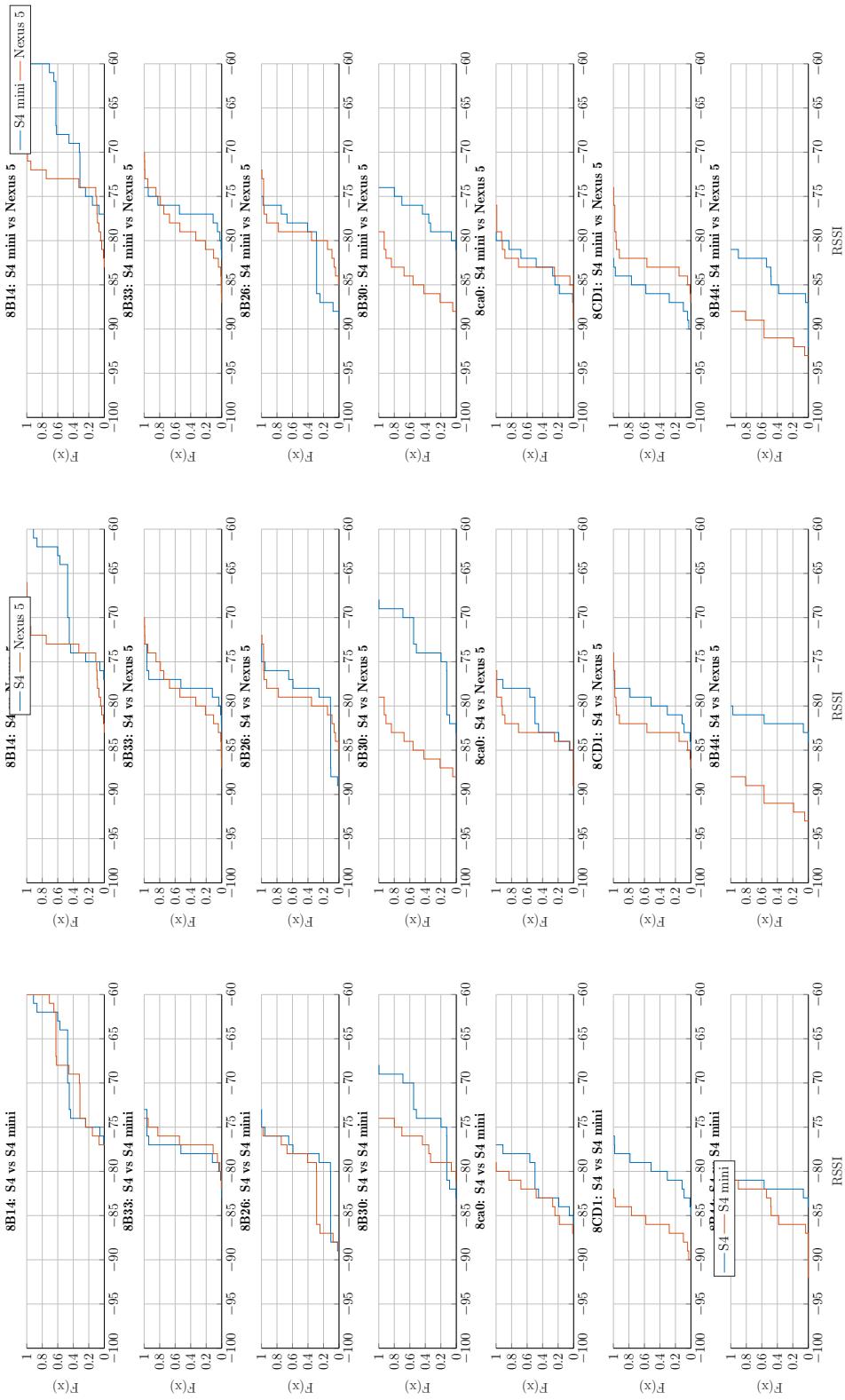


Figure 6.4: Smartphone helvar location 1

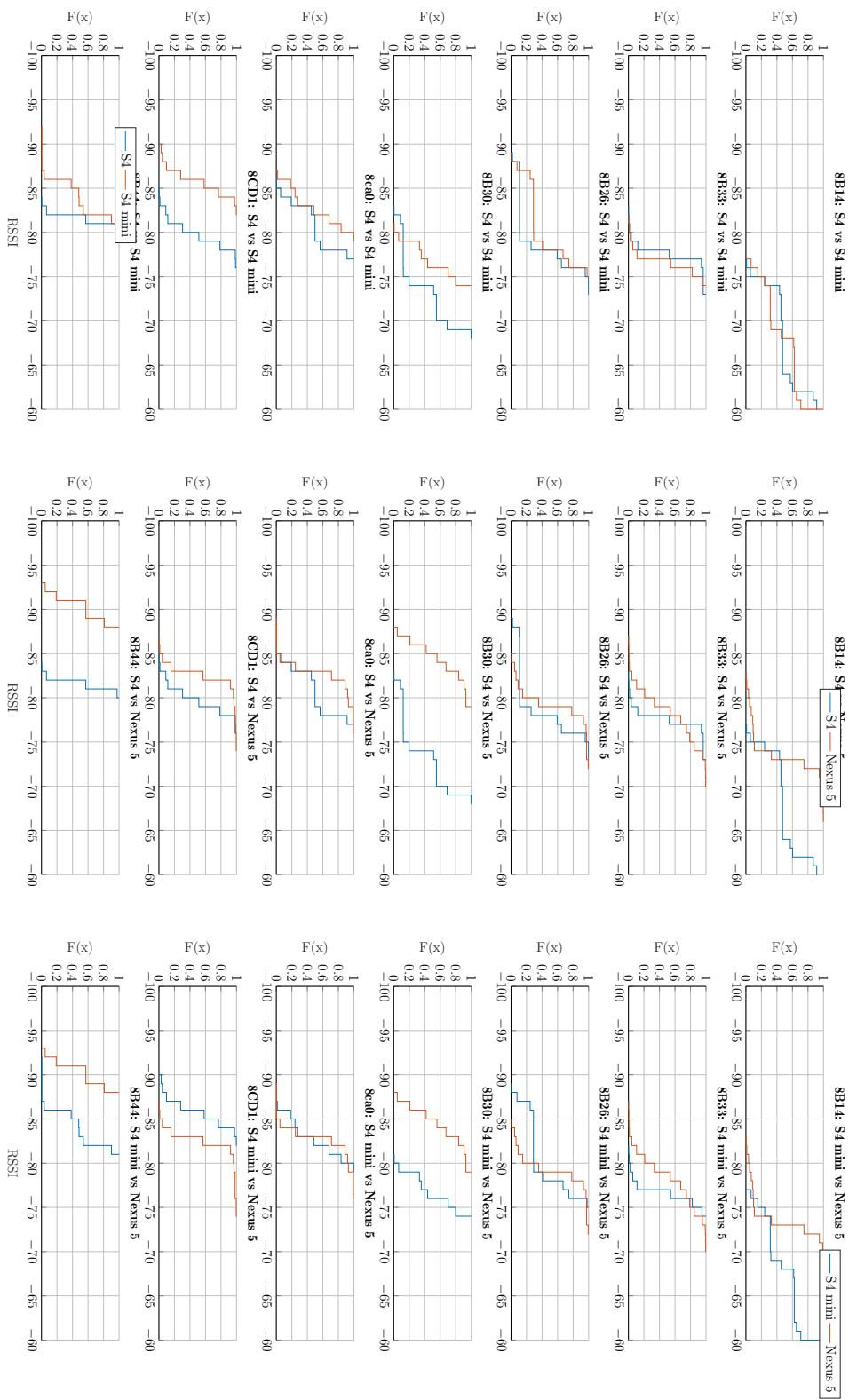


Figure 6.5: Smartphone helvar location 2.

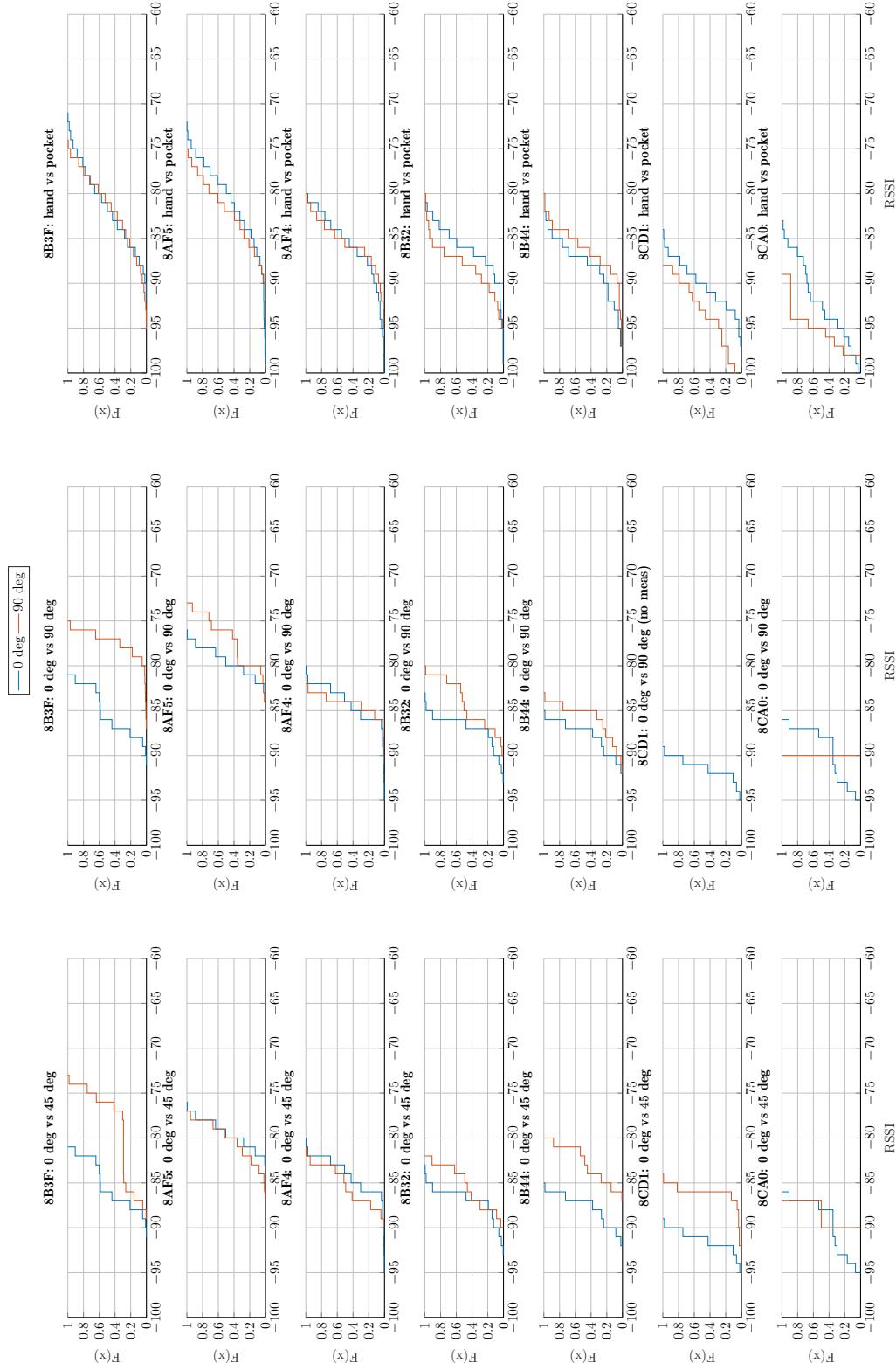


Figure 6.6: Orientation Helvar location 1

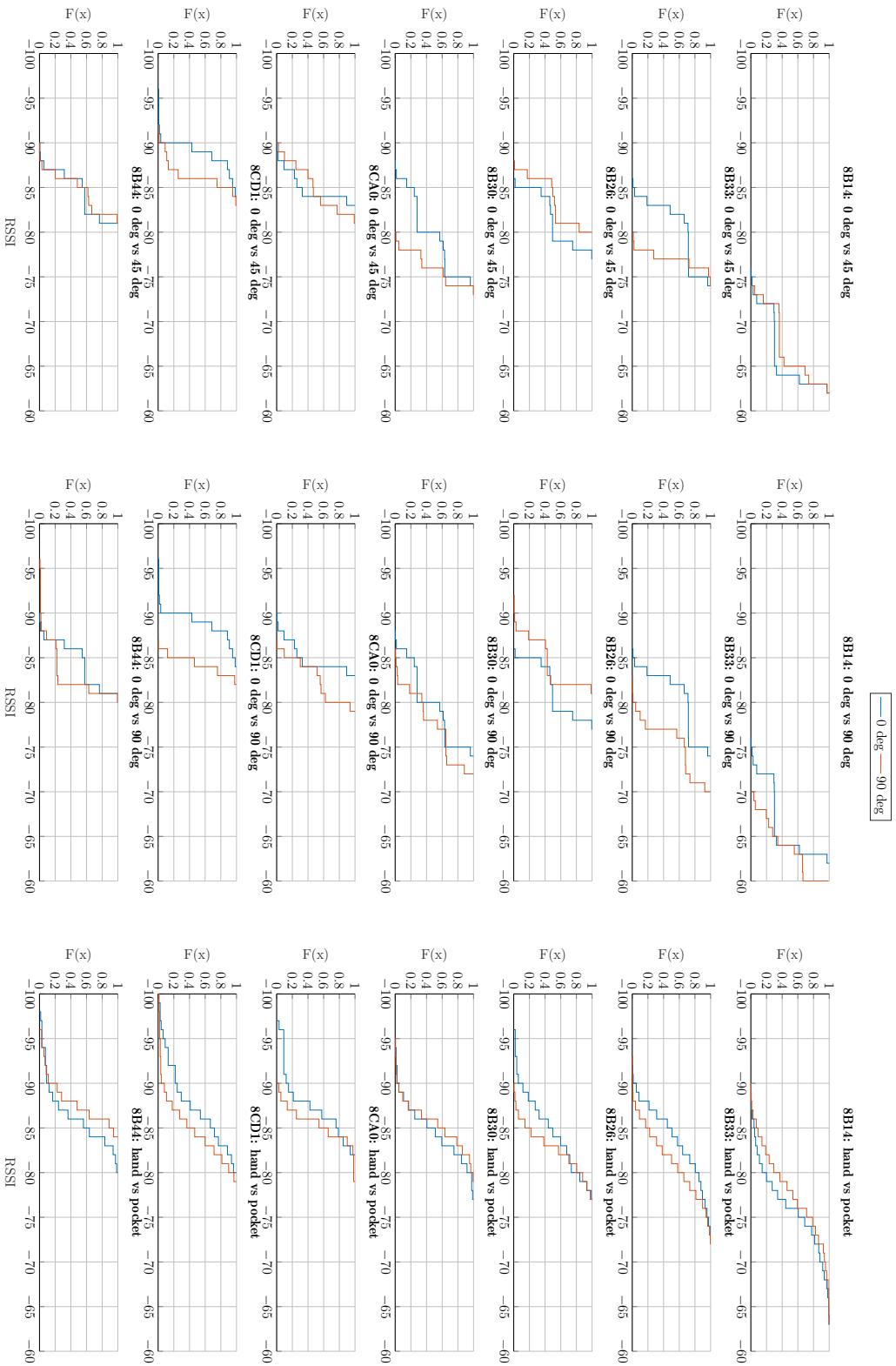


Figure 6.7: Helvar orientation location 2

namely Samsung S4 mini was chosen and measurements were recorded without any user. The phone was present at 90 degrees angle at both the locations right below the access point at a height of 2.898 meters for *metal-plastic* luminaire and 2.707 meters for plastic luminaire. It was assumed that the orientation of the Bluetooth access point or mobile unit have minimal effect on the signal strength values.

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

We conducted two sets of experiments, with one luminaire made of plastic and metal, and other one made only of plastic with the access point inside the ceiling. The results were inconclusive based on the tests. The KS test and median test confirm for both the experiments that the underlying distribution is different but the t-test and f-test infer conflicting results. Refer table ?? for the results.

6.2.6.3 Comparisons of Smartphone

S4 has better reading capability than S4 mini. Nexus 5 has bad reception.

6.2.6.4 Comparison with Radio Analyzer

6.3 Received Signal Strength Indication: Revisited

As mentioned in chapter 5, we record the signal strength along with their MAC address with absolute and relative time of recording using the android measurement application. For simplicity, we investigate only three closest luminaires with user-free setting.

We look at few nearest beacons and plot them against time. user free measurements

FIXME rss first look helvar beacon and kontakt as subplots. FIXME!

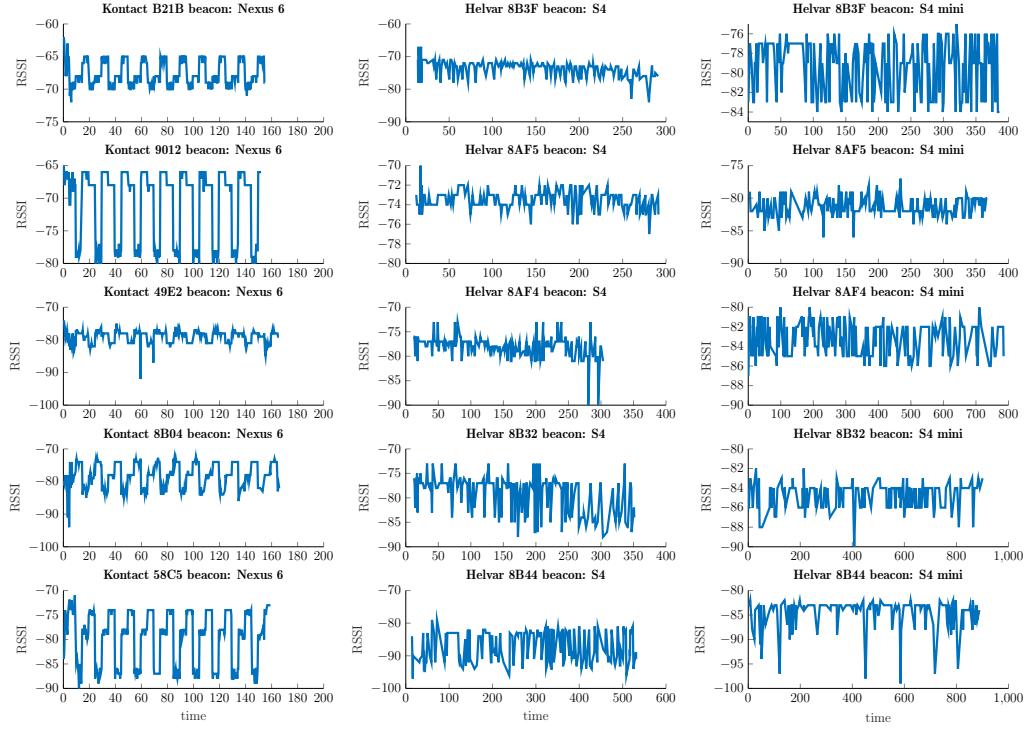


Figure 6.12: The figure illustrates the RSSI versus the time from Helvar and Kontakt beacons. It shows a systematic periodic pattern.

The figure ?? shows a peculiar but periodic characteristic of the RSSI signal for Helvar and Kontakt beacons. It is too systematic to be attributed to the signal fluctuation. This calls for a detailed investigation of the RSSI values. In order to avoid all the different environmental factors which could add bias to RSSI values we take the measurements of the beacons in a outdoor setting with the android phone and radio-analyzer.

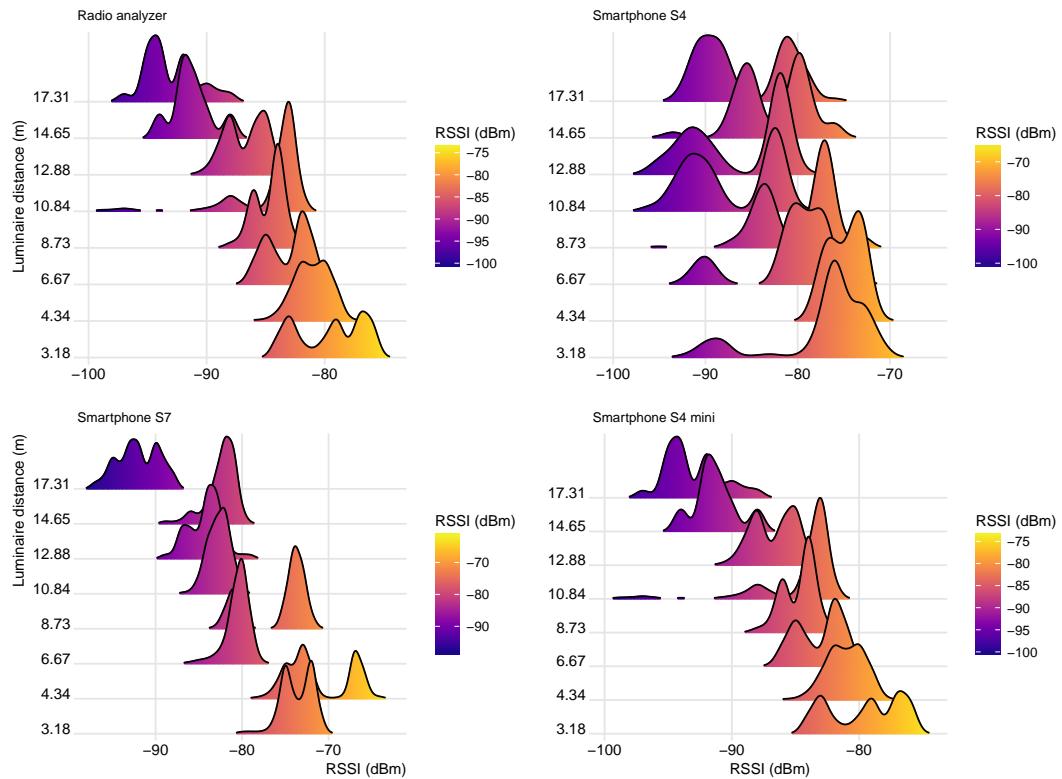


Figure 6.13: The figure shows the density of signal strength as distance increases.

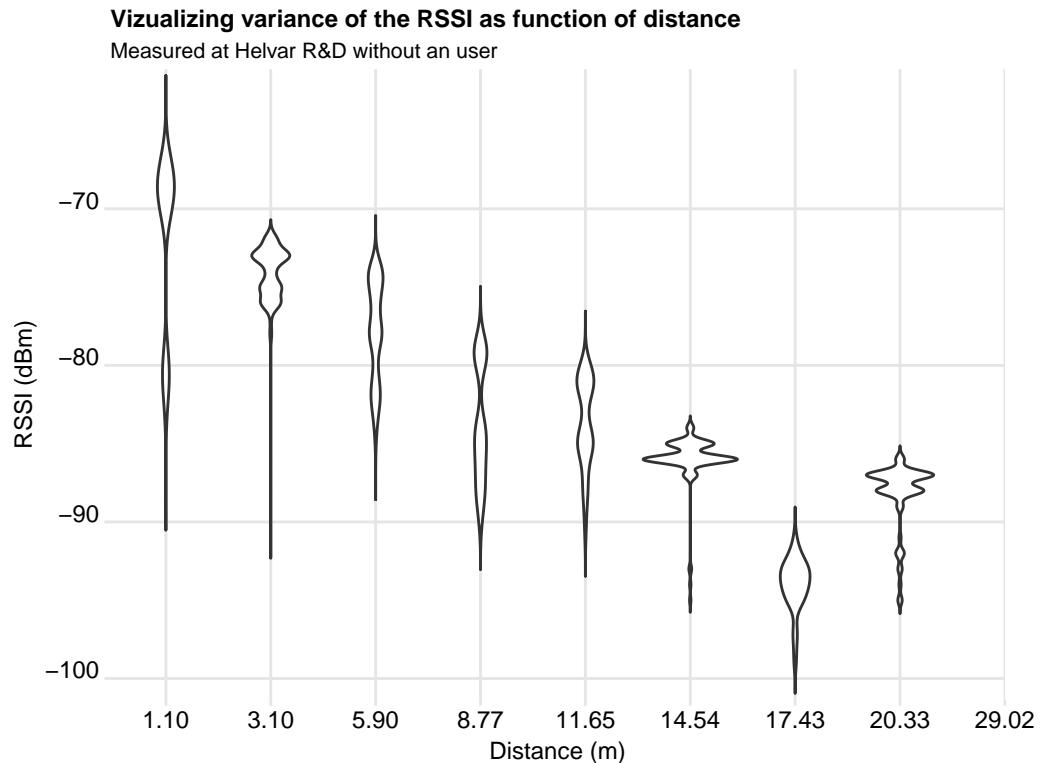


Figure 6.14: The figure shows the variation of signal strength as distance increases.

6.3.1 Measurement Outdoors with Radio analyzer

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

From the figure ??, it is evident that the periodic trend in the RSSI values is due to the difference in the performance of the channels of the BLE beacon. Due to the mismatch in the performance of the channels the variation in the RSSI values were over 20 dBm. Testing with the different beacons showed that the channels which performed the worse always wasn't same.

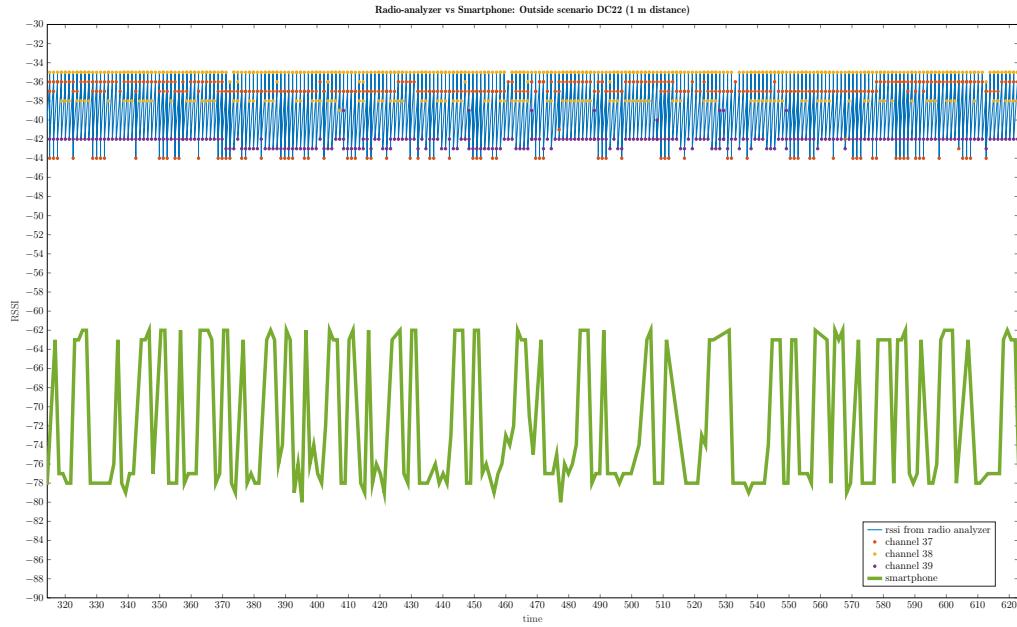


Figure 6.15: Helvar beacon RSSI measurement outside using radio analyzer and smartphone.

Kaemarungsi and Krishnamurthy (2012) discussion on the quantization of RSSI values in the smartphones.

Kaemarungsi and Krishnamurthy (2012) averaging of the RSSI values in effect mitigates the effect of quantization. is also reduces the small-scale fading effects.

Kaemarungsi and Krishnamurthy (2012) indoor - multipath environment and propagation effects (reflection, diffraction, scattering). multipath fading effect is caused due to either constructive or destructive interference of signal causing the mean of the signal strength to fluctuate. large scale fading vs small scale fading Sklar (1997): check Kaemarungsi and Krishnamurthy (2012) for summary.

NLOS - Rayleigh distribution ; LOS - Rician distribution.; mean of RSSI is log-normal distribution Sklar (1997)

the mean of the distribution would follow one of the standard path loss models. from Kaemarungsi and Krishnamurthy (2012) cite this [32].

Kaemarungsi and Krishnamurthy (2012) the sd and stationarity of signal strength values are often understood poorly

Kaemarungsi and Krishnamurthy (2012) multimodal distribution of RSSI [33].

According to Kaemarungsi and Krishnamurthy (2012), the deviation of the measurements is no greater than 10 dBm. The received signal strength

is no greater than maximum transmitted power and lowest value is bounded receiver device sensitivity.

Kaemarungsi and Krishnamurthy (2012): if average at particular location is higher (say -80 dBm or above), then the most probable distribution is left skewed or the distribution would either be symmetric or log-normal. For points farther away the distribution is symmetric.

Kaemarungsi and Krishnamurthy (2012) add discussion of the histograms observed from various locations.

Kaemarungsi and Krishnamurthy (2012) how to find left skewed? [35].

Discussion on skewness in relation with the distance.

Left skewed are prominent in either short duration measurements or very long duration measurements.

Standard deviation vs rss values with reference to the LOS or NLOS beacons.

Stationarity vs RSSI

Temporal received signal strength outage. According to (Kaemarungsi and Krishnamurthy, 2012) our test bed should not have any issues.

Kaemarungsi and Krishnamurthy (2012) shows how the RSSI values are time dependent.

Kaemarungsi and Krishnamurthy (2012) human factor dominates the RSSI values so we add it to the model. the noise part. or something.

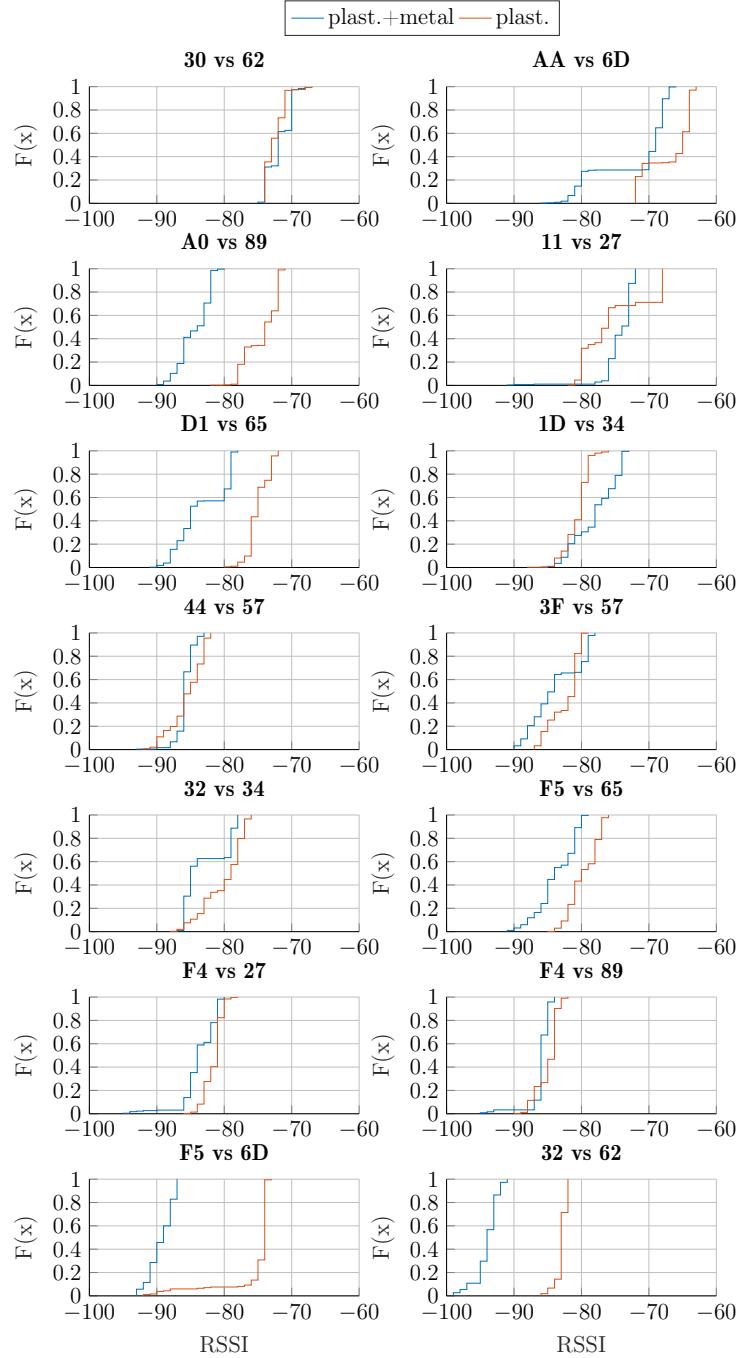


Figure 6.8: Comparing the RSSI values from luminaires with different material at different distances (check the table 6.3).

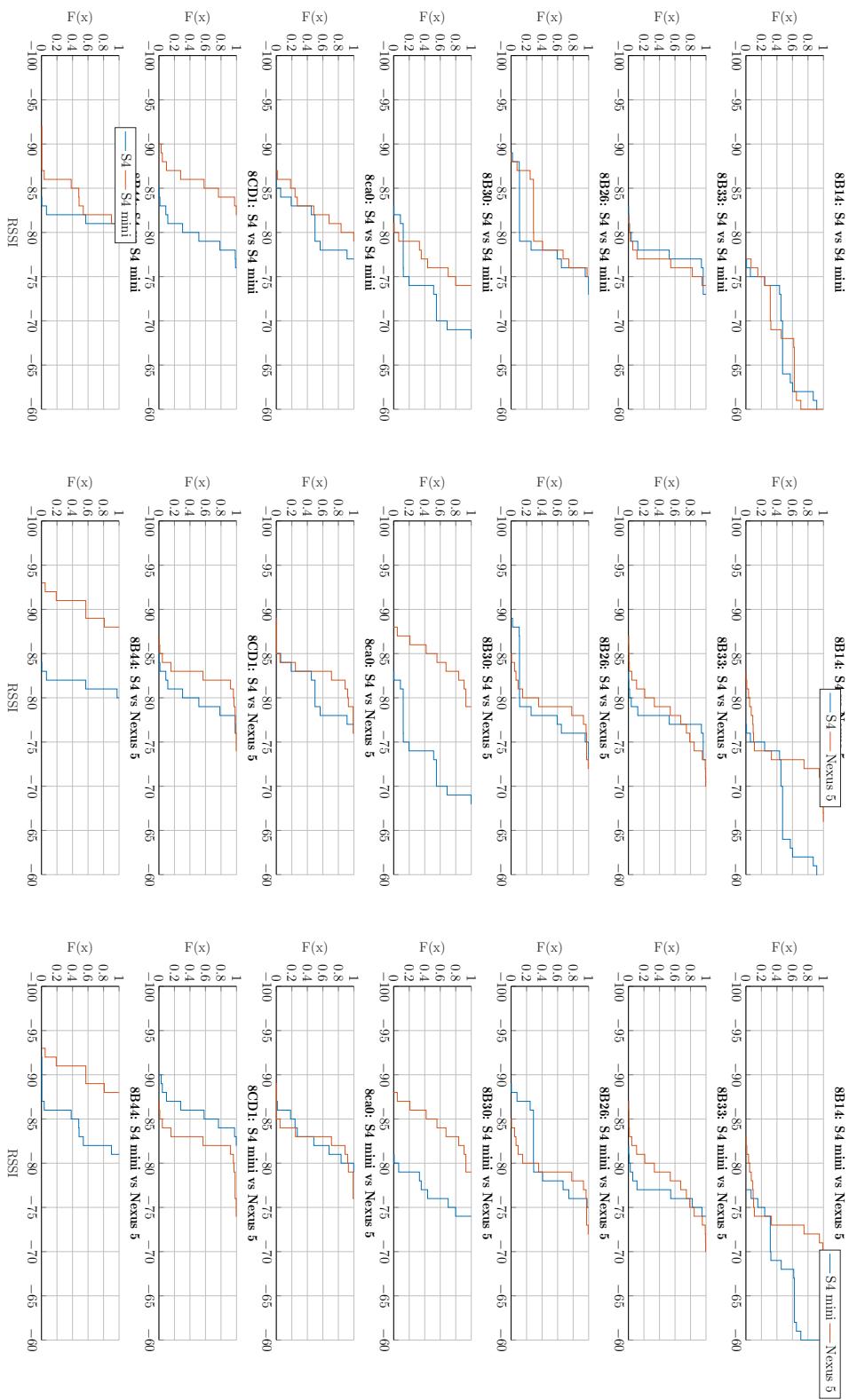


Figure 6.9: Smartphone helvar location 1

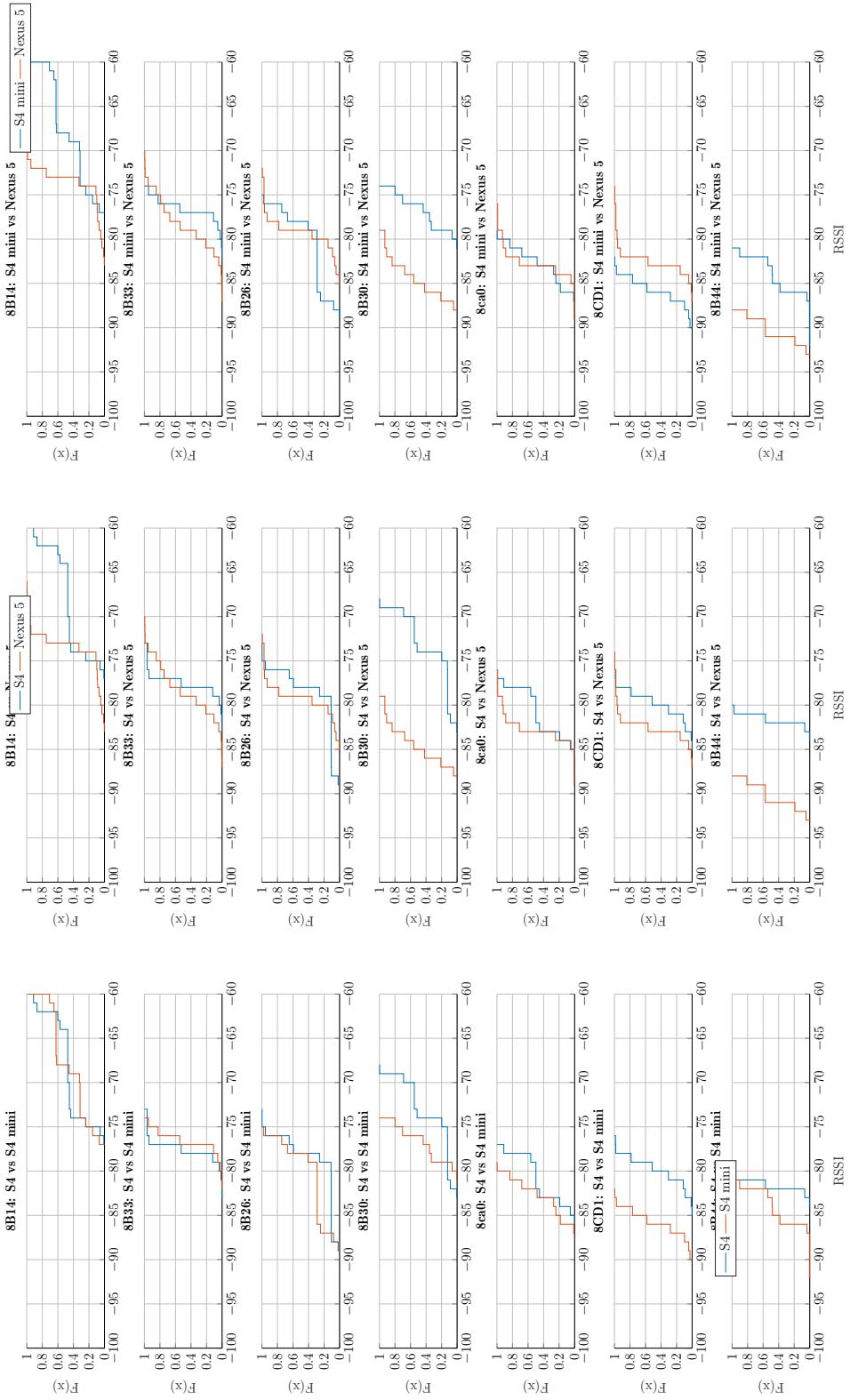


Figure 6.10: Smartphone helvar location 2.

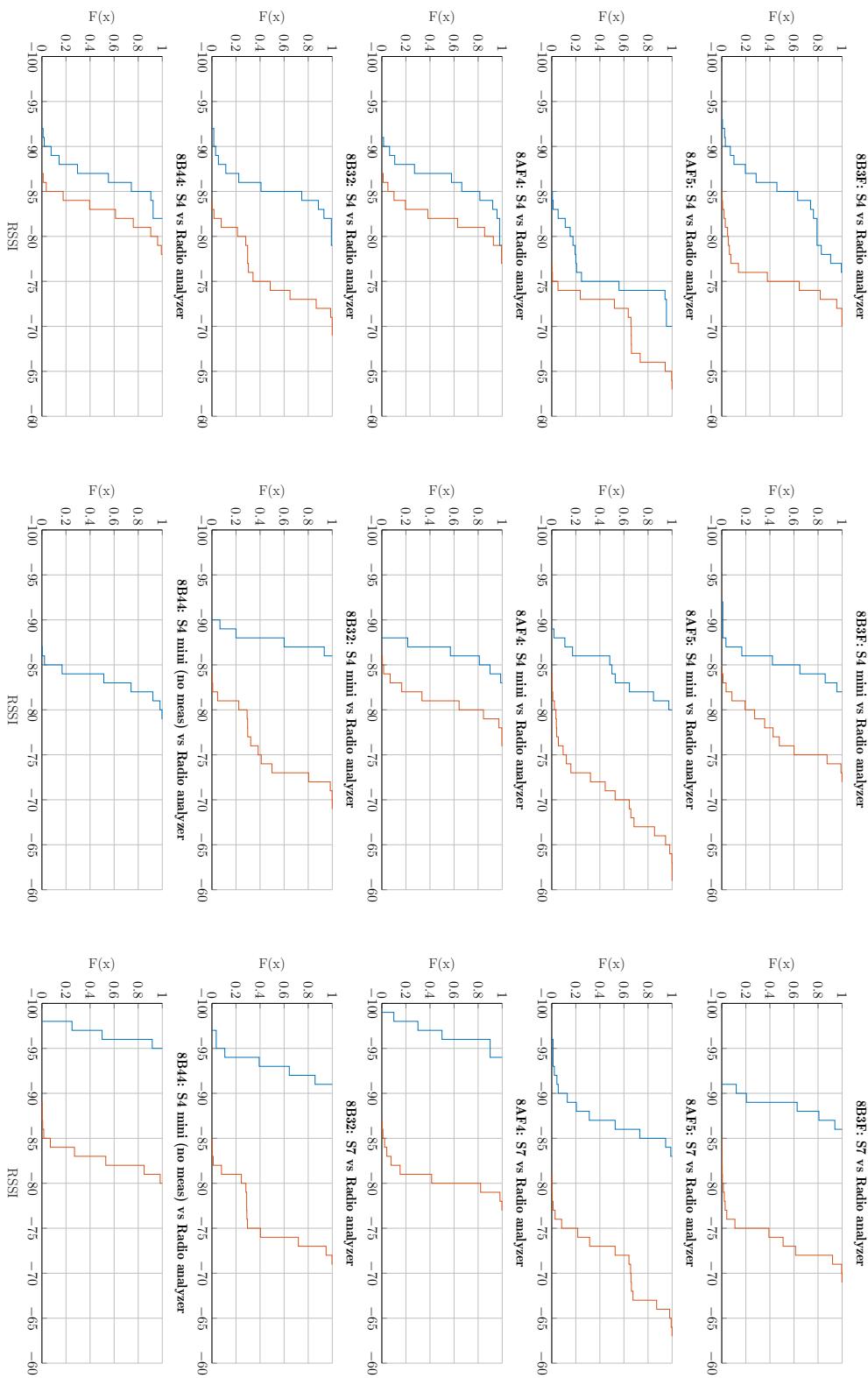


Figure 6.11: Radio analyzer bias from smartphones location 1

Chapter 7

Experiments and Results

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

7.1 Effect of parameters in calibration phase on positioning

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

7.1.1 Separate parameter evaluation

7.1.1.1 Calibration points

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

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

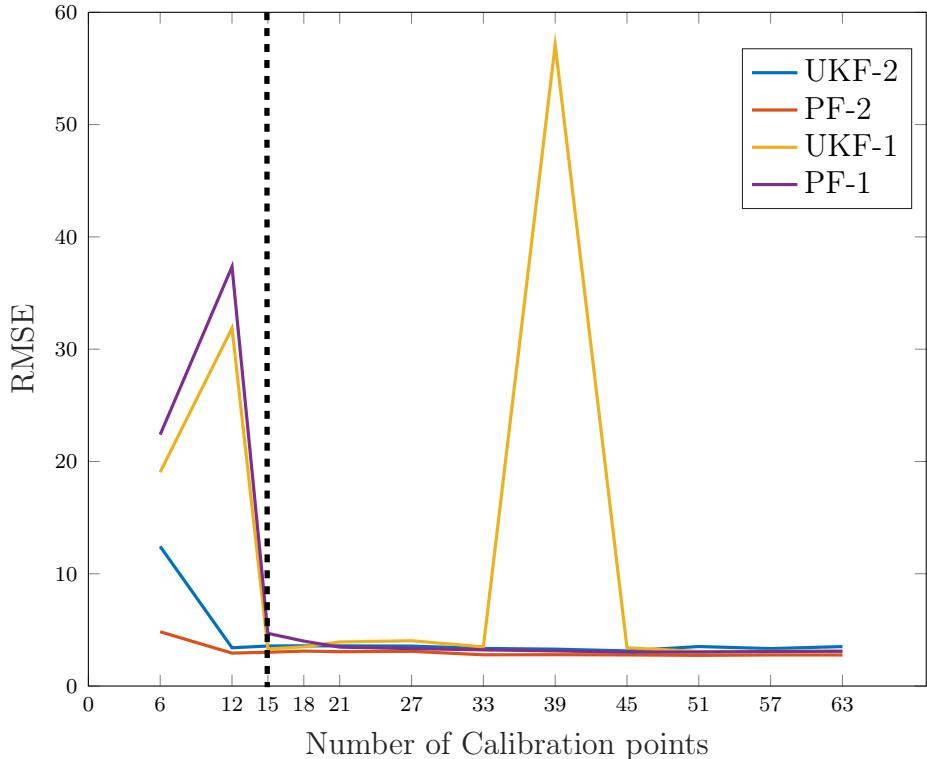


Figure 7.1: Experiment with increasing calibration points for GP-PF and GP-UKF

7.1.1.2 Calibration time

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

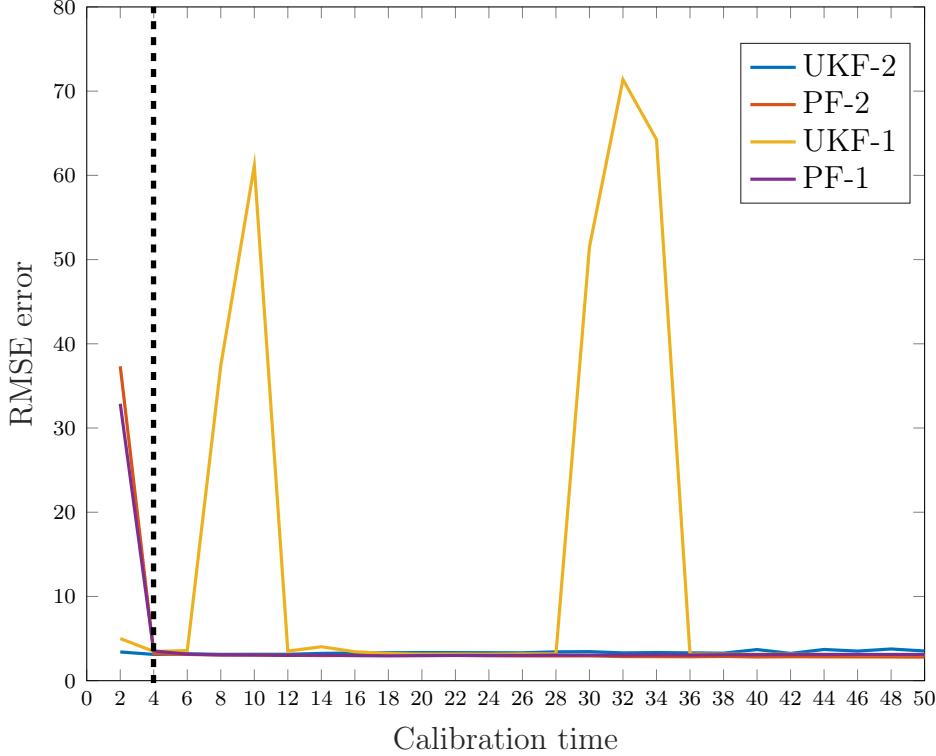


Figure 7.2: Experiment with increasing calibration time for GP-PF and GP-UKF.

7.1.2 Combined parameter evaluation

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

7.2 Filtering location estimation algorithms

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

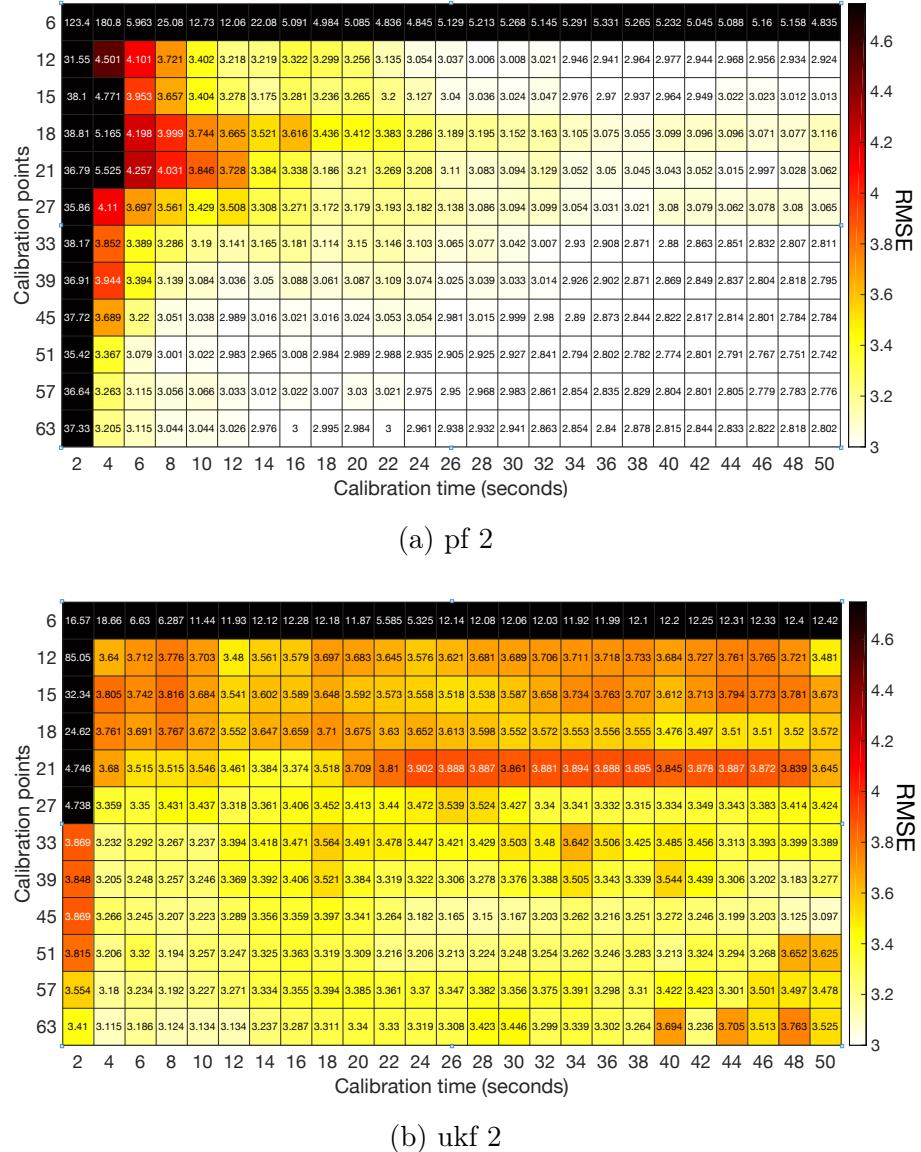


Figure 7.3: radio-map 1

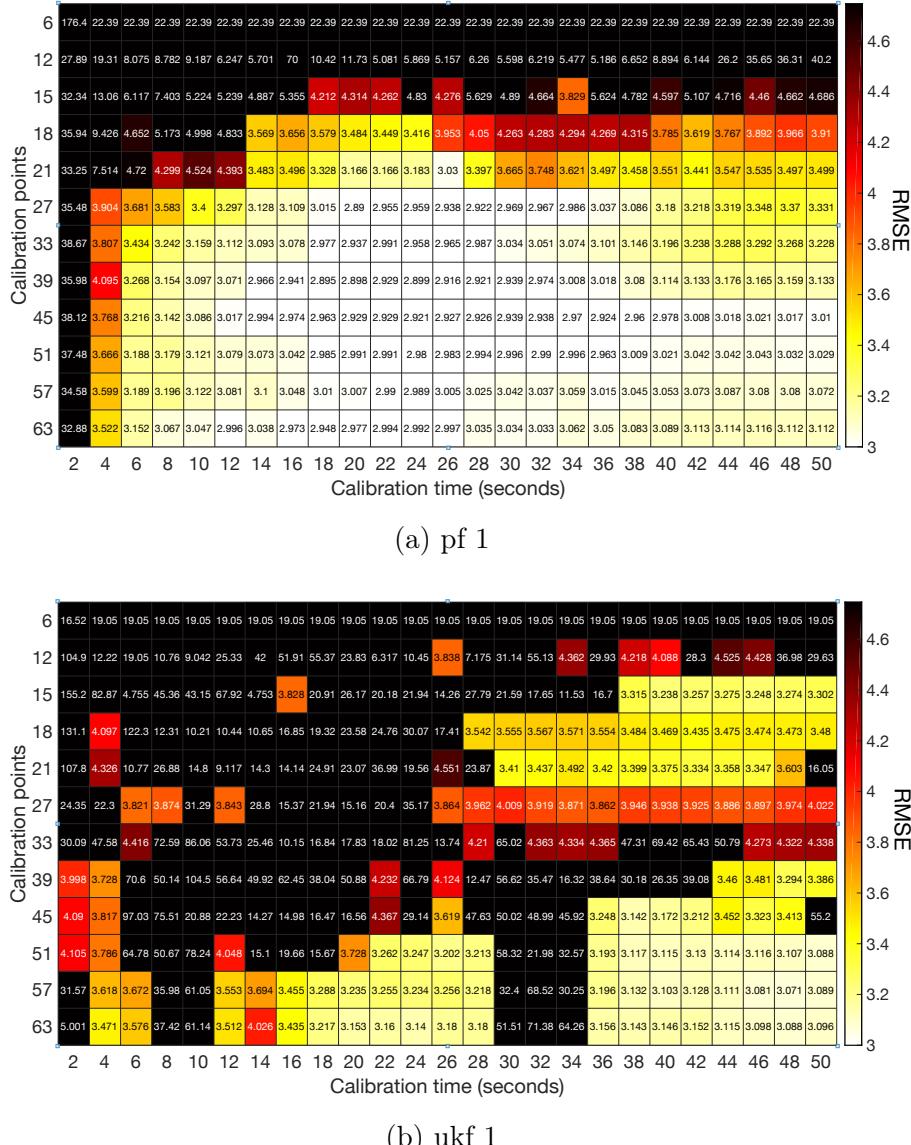


Figure 7.4: radio-map 2

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

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

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

7.2.1 Summary of the location estimation algorithms

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

In comparison to radio-map 1, the results in the different parameter zones shows significant improvement in terms of performance for both particle filter and unscented Kalman filter.

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

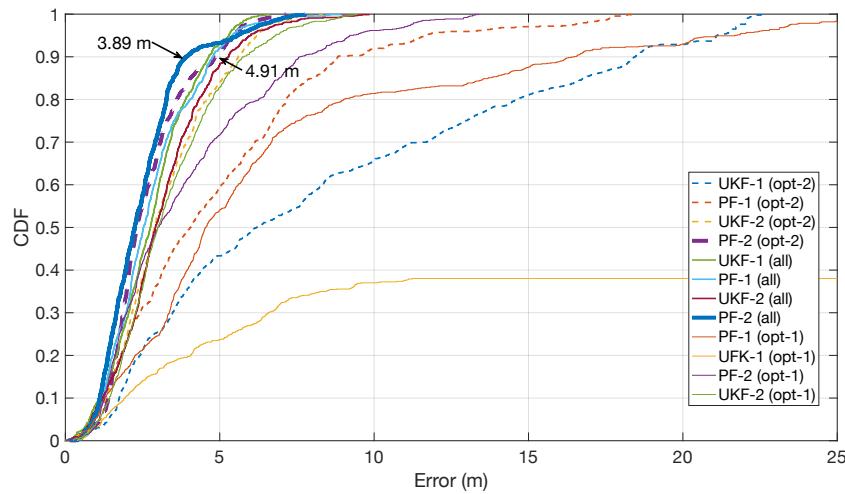


Figure 7.5: Indoor position cumulative error distribution evaluation for particle and unscented Kalman filter using radio-map 1 & 2

We have additionally plotted Figure 7.5, showing the cumulative error

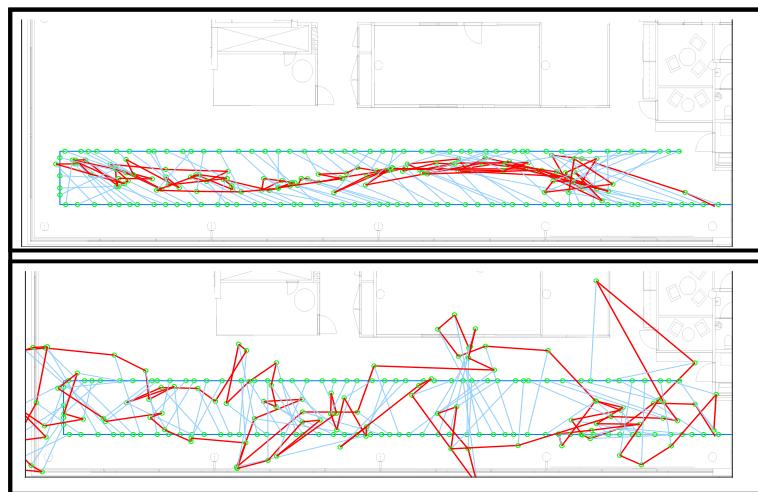


Figure 7.6: Positioning result for (top) Particle Filter using radio-map 2 (bottom) Unscented Kalman Filter using radio-map

Chapter 8

Discussion & Future Work

As learnt from the Chapter data analysis, the distance interpreting methods if not implemented with caution the prediction could be biased.

8.1 Discussion

This study takes a step backwards in understanding the data and in that process we understood that the variation in RSSI is inherent in its architecture and technology. The difference in the performance of the channel could also attributed to the reason it is low energy device. With this characteristic of the RSSI signal, we can say that averaging the signal strength values could lead to misleading results.

This thesis discovered a unique characteristic of the BLE devices which shows a distinction in working of the three different channels. As seen from the experiments in the data analysis the difference is quite huge when the device near with variance spreading 20 dBm. The bias due to this is two fold. One in the construction of the radio map. And next in the running of the filter, this effect is evident when we try to combine the signal information from various beacons in the calculation of weights in the particle. This bias confuses the filter into selecting an estimate outside the state space.

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

This study also looked at how the masking of the channel (or multiple channels) in BLE beacons would help in improvement of accuracy of the localization problem. We were in an advantageous position to implement this as our aim was to implement an IP solution which involved BLE beacons of same make !FIXME i.e., **nRF51822** !FIXME!. The case would be different if we

had to create a single standalone solution for beacons of different make.

These methods are capable of integrating multiple data information and can pave path for a unified solution.

8.1.1 Data Analysis

The BLE architecture has 40 channels and has three advertisement channels which are strategically placed to avoid interference with the WiFi channels
!FIXME add bluetooth specifications **FIXME!**.

The advertisement channels are spread over the frequency channels and the performance of the channels vary. The difference in the performance channel affects the signal strength value to vary over 10 dBms.

8.1.2 Radio map

Additionally, GPs also learnt the presence of the wall and it was evident in the radio-maps. The advantages of GP based regression: (1) more flexible spatial distribution. (2) ability to predict outside the specified state space region. (3) considerably good accuracy even with less calibration points.

The radio-map 1 is built on top of averaged RSSI values which leads loss of information in terms of variance. This is preserved in the radio-map 2.

8.1.3 Indoor Positioning

In this thesis we test with three different methods and experimented with the number of calibration points and time for taking each measurements. We found out that.. **!FIXME relation between calibration points and time for measurements.** **FIXME!**

8.2 Future Work

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

Use the truncated singular value decomposition (TSVD) to map the signal strength measure to spatial locations. TSVD encapsulates the dynamics of the indoor environment and our indoor setting fulfills the criteria of dense

APs (Lim et al., 2005). Lim et al. (2005) also discusses about zero configuration positioning.

Inertial Navigation systems (Solin et al., 2017)

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

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

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

Masking the channels

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

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

different dynamic models.

References

- Addlesee, M., Curwen, R., Hodges, S., Newman, J., Steggles, P., Ward, A., and Hopper, A. (2001). Implementing a sentient computing system. *Computer*, 34(8):50–56.
- Aguilar-Garcia, A., Fortes, S., Molina-García, M., Calle-Sánchez, J., Alonso, J. I., Garrido, A., Fernández-Durán, A., and Barco, R. (2015). Location-aware self-organizing methods in femtocell networks. *Computer Networks*, 93:125–140.
- Aittola, M., Ryhänen, T., and Ojala, T. (2003). Smartlibrary—location-aware mobile library service. In *International Conference on Mobile Human-Computer Interaction*, pages 411–416. Springer.
- Aravecchia, M. and Messelodi, S. (2014). Gaussian process for RSS-based localisation. In *2014 IEEE 10th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pages 654–659.
- Au, A. W. S. (2010). *Rss-based wlan indoor positioning and tracking system using compressive sensing and its implementation on mobile devices*. PhD thesis.
- Bahl, P. and Padmanabhan, V. N. (2000). Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. IEEE.
- Bahl, P., Padmanabhan, V. N., and Balachandran, A. (2000). A software system for locating mobile users: Design, evaluation, and lessons. *Microsoft Research, MSR-TR-2000-12*.
- 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*.
- Curran, K., Furey, E., Lunney, T., Santos, J., Woods, D., and McCaughey, A. (2011). An evaluation of indoor location determination technologies. *Journal of Location Based Services*, 5(2):61–78.
- 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.
- De Luca, D., Mazzenga, F., Monti, C., and Vari, M. (2006). Performance evaluation of indoor localization techniques based on rf power measurements from active or passive devices. *EURASIP Journal on Applied Signal Processing*, 2006:160–160.
- Djuknic, G. M. and Richton, R. E. (2001). Geolocation and assisted gps. *Computer*, 34(2):123–125.
- Doucet, A., De Freitas, N., and Gordon, N. (2001). Sequential monte carlo methods in practice. series statistics for engineering and information science.
- Eddystone, I. (2016). The new beacon format from google. URL <http://developer.radiusnetworks.com/2015/07/14/introducing-eddystone.html>.
- Elnahrawy, E., Li, X., and Martin, R. P. (2004). The limits of localization using signal strength: A comparative study. In *Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on*, pages 406–414. IEEE.
- Faragher, R. and Harle, R. (2014). An analysis of the accuracy of bluetooth low energy for indoor positioning applications. In *Proceedings of the 27th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS+’14)*, pages 201–210.

- Feldmann, S., Kyamakya, K., Zapater, A., and Lue, Z. (2003). An indoor bluetooth-based positioning system: Concept, implementation and experimental evaluation. In *International Conference on Wireless Networks*, pages 109–113.
- Ferris, B., Hahnel, D., and Fox, D. (2007). *Gaussian processes for signal strength-based location estimation*, volume 2, pages 303–310. MIT Press Journals.
- Foley, J. D. (1990). A. vandam, sk feiner, and jf hughes. *Computer Graphics: Principles and Practice*.
- Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., and Rubin, D. (2014). *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)*. Chapman and Hall/CRC, London, third edition.
- Gomez, C., Oller, J., and Paradells, J. (2012). Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology. *Sensors*, 12(9):11734–11753.
- Gu, Z., Chen, Z., Zhang, Y., Zhu, Y., Lu, M., and Chen, A. (2016). Reducing fingerprint collection for indoor localization. *Computer Communications*, 83:56–63.
- Hansraj, K. (2014). Assessment of stresses in the cervical spine caused by posture and position of the head. 25:277–9.
- Hashemi, H. (1993). The indoor radio propagation channel. *Proceedings of the IEEE*, 81(7):943–968.
- Hazas, M., Scott, J., and Krumm, J. (2004). Location-aware computing comes of age. *Computer*, 37(2):95–97.
- He, S. and Chan, S. H. G. (2016). Wi-fi fingerprint-based indoor positioning: Recent advances and comparisons. *IEEE Communications Surveys Tutorials*, 18(1):466–490.
- Honkavirta, V. (2008). Location Fingerprinting Methods In Wireless Local Area Networks. Master’s thesis, Tampere University of Technology, Finland.
- Huang, J., Millman, D., Quigley, M., Stavens, D., Thrun, S., and Aggarwal, A. (2011). Efficient, generalized indoor wifi graphslam. In *Robotics and*

- Automation (ICRA), 2011 IEEE International Conference on*, pages 1038–1043. IEEE.
- Jardak, N. and Samama, N. (2009). Indoor positioning based on gps-repeaters: performance enhancement using an open code loop architecture. *IEEE Transactions on Aerospace and electronic systems*, 45(1):347–359.
- Julier, S. J. and Uhlmann, J. K. (1997). New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–194. International Society for Optics and Photonics.
- Kaemarungsi, K. and Krishnamurthy, P. (2004). Properties of indoor received signal strength for wlan location fingerprinting. In *Mobile and Ubiquitous Systems: Networking and Services, 2004. MOBIQUITOUS 2004. The First Annual International Conference on*, pages 14–23. IEEE.
- Kaemarungsi, K. and Krishnamurthy, P. (2012). Analysis of wlan’s received signal strength indication for indoor location fingerprinting. *Pervasive and mobile computing*, 8(2):292–316.
- 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.
- Li, W. W.-L., Iltis, R. A., and Win, M. Z. (2013). Integrated imu and radio location-based navigation using a rao-blackwellized particle filter.
- Lim, H., Kung, L.-C., Hou, J. C., and Luo, H. (2005). Zero-configuration, robust indoor localization: Theory and experimentation. Technical report.
- Lindh, J. (2015). Bluetooth low energy beacons. *Texas Instruments*, page 15.
- Liu, J. S. (2008a). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

- Liu, J. S. (2008b). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.
- Luo, X., O'Brien, W. J., and Julien, C. L. (2011a). Comparative evaluation of received signal-strength index (rss) based indoor localization techniques for construction jobsites. *Advanced Engineering Informatics*, 25(2):355–363.
- Luo, X., O'Brien, W. J., and Julien, C. L. (2011b). Comparative evaluation of received signal-strength index (rss) based indoor localization techniques for construction jobsites. *Advanced Engineering Informatics*, 25(2):355–363.
- Mainetti, L., Patrono, L., and Sergi, I. (2014). A survey on indoor positioning systems. In *Software, Telecommunications and Computer Networks (Soft-COM), 2014 22nd International Conference on*, pages 111–120. IEEE.
- 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.
- Neal, R. M. (2012). *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media.
- Networks, R. (2017). Altbeacon: Beacon android library.
- Newman, N. (2014). Apple ibeacon technology briefing. *Journal of Direct, Data and Digital Marketing Practice*, 15(3):222–225.
- Patwari, N. and Wilson, J. (2010). Rf sensor networks for device-free localization: Measurements, models, and algorithms. *Proceedings of the IEEE*, 98(11):1961–1973.
- Pivato, P., Palopoli, L., and Petri, D. (2011). Accuracy of rss-based centroid localization algorithms in an indoor environment. *IEEE Transactions on Instrumentation and Measurement*, 60(10):3451–3460.
- Priyantha, N. B. (2005). *The cricket indoor location system*. PhD thesis, Massachusetts Institute of Technology.

- Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Rizos, C., Dempster, A. G., Li, B., and Salter, J. (2007). Indoor positioning techniques based on wireless lan.
- Roos, T., Myllymäki, P., Tirri, H., Misikangas, P., and Sievänen, J. (2002). A probabilistic approach to wlan user location estimation. *International Journal of Wireless Information Networks*, 9(3):155–164.
- Särkkä, S. (2013). *Bayesian filtering and smoothing*, volume 3. Cambridge University Press.
- Schüssel, M. and Pregizer, F. (2015). Coverage gaps in fingerprinting based indoor positioning: The use of hybrid gaussian processes. In *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, pages 1–9. IEEE.
- Schwaighofer, A., Grigoras, M., Tresp, V., and Hoffmann, C. (2004). Gpps: A gaussian process positioning system for cellular networks. In *Advances in Neural Information Processing Systems*, pages 579–586.
- Seco, F., Plagemann, C., Jiménez, A. R., and Burgard, W. (2010). Improving rfid-based indoor positioning accuracy using gaussian processes. In *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, pages 1–8. IEEE.
- Sklar, B. (1997). Rayleigh fading channels in mobile digital communication systems. i. characterization. *IEEE Communications magazine*, 35(9):136–146.
- Solin, A. (2016). *Stochastic Differential Equation Methods for Spatio-Temporal Gaussian Process Regression*. Phd dissertation, Aalto University.
- Solin, A., Cortes, S., Rahtu, E., and Kannala, J. (2017). Inertial odometry on handheld smartphones. *arXiv preprint arXiv:1703.00154*.
- Steggles, P. and Gschwind, S. (2005). The ubisense smart space platform.
- Stone, L. D., Keller, C. M., Kratzke, T. M., Strumpfer, J. P., et al. (2014). Search for the wreckage of air france flight af 447. *Statistical Science*, 29(1):69–80.

- Vehtari, A., Gelman, A., and Gabry, J. (2015). Pareto smoothed importance sampling. *arXiv preprint arXiv:1507.02646*.
- Wang, J., Gao, Q., Wang, H., Chen, H., and Jin, M. (2011). Differential radio map-based robust indoor localization. *EURASIP Journal on Wireless Communications and Networking*, 2011(1):17.
- Yiu, S., Dashti, M., Claussen, H., and Perez-Cruz, F. (2016). Locating user equipments and access points using rssi fingerprints: A gaussian process approach. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–6. IEEE.
- Yiu, S., Dashti, M., Claussen, H., and Perez-Cruz, F. (2017). Wireless rssi fingerprinting localization. *Signal Processing*, 131:235–244.
- Zanca, G., Zorzi, F., Zanella, A., and Zorzi, M. (2008). Experimental comparison of rssi-based localization algorithms for indoor wireless sensor networks. In *Proceedings of the workshop on Real-world wireless sensor networks*, pages 1–5. ACM.

Appendix A

Appendix