TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Jukka Talvitie

Algorithms and Methods for Received Signal Strength Based Wireless Localization

Julkaisu 1365 • Publication 1365

Tampere 2016

Tampereen teknillinen yliopisto. Julkaisu 1365
Tampere University of Technology. Publication 1365

Jukka Talvitie

Algorithms and Methods for Received Signal Strength Based Wireless Localization

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Sähkötalo Building, Auditorium S2, at Tampere University of Technology, on the 15th of January 2016, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2016

Supervisor

Markku Renfors, Prof.
Department of Electronics and Communications Engineering
Tampere University of Technology
Tampere, Finland

Pre-examiners

Pau Closas, Dr.
Centre Tecnològic de Telecomunicacions de Catalunya
Barcelona, Spain

Traian Abrudan, Dr.
Department of Computer Science
University of Oxford
Oxford, United Kingdom

Opponent

Mérouane Debbah, Prof.
CentraleSupélec
Gif-Sur-Yvette, France

ISBN 978-952-15-3672-4 (printed)
ISBN 978-952-15-3679-3 (PDF)
ISSN 1459-2045

Abstract

In the era of wireless communications, the demand for localization and localization-based services has been continuously growing, as increasingly smarter wireless devices have emerged to the market. Besides the already available satellite-based localization systems, such as the GPS and GLONASS, also other localization approaches are needed to complement the existing solutions. Finding different types of low-cost localization methods, especially for indoors, has become one of the most important research topics in recent years.

One of the most used approaches in localization is based on Received Signal Strength (RSS) information. Specific fingerprints about RSS are collected and stored and positioning can be done through pattern or feature matching algorithms or through statistical inference. A great and immediate advantage of the RSS-based localization is its ability to exploit the already existing infrastructure of different communications networks without the need to install additional system hardware. Furthermore, due to the evident connection between the RSS level and the quality of a communications signal, the RSS is usually inherently included in the network measurements. This favors the availability of the RSS measurements in the current and future wireless communications systems.

In this thesis, we study the suitability of RSS for localization in various communications systems including cellular networks, wireless local area networks, personal area networks, such as WiFi, Bluetooth and Radio Frequency Identification (RFID) tags. Based on substantial real-life measurement campaigns, we study different characteristics of RSS measurements and propose several Path Loss (PL) models to capture the essential behavior of the RSS levels in 2D outdoor and 3D indoor environments. By using the PL models, we show that it is possible to attain similar performance to fingerprinting with a database size of only 1-2% of the database size needed in fingerprinting. In addition, we study the effect of different error sources, such as database calibration errors, on the localization accuracy. Moreover, we propose a novel method for studying how coverage gaps in the fingerprint database affect the localization performance. Here, by using various interpolation and extrapolation methods, we improve the localization accuracy with imperfect fingerprint databases, such as those including substantial coverage gaps due to inaccessible parts of the buildings.

Preface

This thesis is based on the research work carried out during years 2008-2015 at the Department of Electronics and Communications Engineering, Tampere University of Technology, Tampere, Finland. I greatly appreciate the financial support provided by Tampere Doctoral Programme in Information Science and Engineering (TISE), Jenny and Antti Wihuri Foundation (2010), Tuula and Yrjö Neuvo Foundation (2011), and Doctoral training network in Electronics, Telecommunications and Automation (DELTA).

Firstly, I would like to thank my supervisor Prof. Markku Renfors for his invaluable guidance and personal example, which has shaped me into the researcher I am today. Besides his widely recognized scientific contributions, he has also highly impacted on elevating the team spirit prevailing among the colleagues at the department. Secondly, I would like to express my gratitude to Dr. Elena-Simona Lohan for her enormous contribution to our mutual publications and research activities. I am also extremely grateful to my superior Prof. Mikko Valkama for providing me all the possible support that I could ask for during my whole time at the department.

I would like to thank the pre-examiners of this thesis, Dr. Pau Closas and Dr. Traian Abrudan, for agreeing to review the thesis, and for providing constructive comments, which definitely enhanced the overall quality of the thesis. Furthermore, I am highly grateful to Prof. Mérouane Debbah for agreeing to act as the opponent in the public examination of this thesis.

Numerous good ideas are often aroused out of interaction between fellow workers. Hence, I would like to express my gratitude to all my colleagues and co-authors, especially to Toni Levanen, Markus Allén, Jaakko Marttila, Elina Laitinen, Shweta Shrestha, Pedro Figueiredo e Silva, Orod Raeesi, Ville Syrjälä, Janis Werner, Tero Isotalo, Lauri Anttila, Vesa Lehtinen, Jussi Turkka, Aki Hakkarainen, Jukka Rinne, Yaning Zou, Ondrej Daniel, Ari Asp, Dani Korpi, and Joonas Säe. During my years at the department I have been very fortunate to work with such amazing people, and moreover, I have been privileged to form true lifelong friendships. I am also grateful to Tarja Erälaukko, Sari Kinnari, Kirsi Viitanen, Marianna Jokila and Soile Lönnqvist for their assistance in the practical matters during my doctoral studies. In addition, I have great-

ly appreciated the possibility to work with true professionals in several industrial-funded projects. Thus, I would like to thank Jari Syrjärinne and Lauri Wirola from HERE and Juho Pirskanen from former Renesas Mobile for their open minded, but rigorous project management.

I have always liked the phrase 'Work hard, play hard'. As I have already recognized people from the first part of the phrase, I also wish to consider the second part. Hence, I am tremendously thankful to all my friends who have stood by me during all these years. With you guys I have enjoyed our long conversations, holiday trips, parties with good food and drinks, and music, of course.

Finally, I would like to thank my parents Matti and Irma Talvitie for their love and fundamental support during my whole life. In addition, I would like to thank Kaj Jonaeson and my brother Jani Talvitie including his beautiful family, and my grandparents Laura and Reino Vainio and Signe Talvitie, for the lifelong support. Most of all, I thank my beloved Heidi for her unconditional love, as in uphill she pushes me forward and in downhill she breaks my speed. Last but definitely not least, I thank Heidi's family Seppo, Tarja and Marjo Mikkonen for warmly welcoming me to join their wonderful family.

Tampere, Finland, December 2015

Jukka Talvitie.

Contents

List of Symbols

$A^{(r)}$	PL constant of the r^{th} TX
$\mathbf{C}_W^{(r)}$	Shadowing covariance matrix of r^{th} TX
$\mathbf{C}_\theta^{(r)}$	Covariance matrix of the prior distribution of the PL parameter vector of the r^{th} TX
d_{block}	Radius of the chunk of removed fingerprints in the fingerprint removal process
$d_{BP,m}$	Breakpoint distance (a boundary between two PL slopes) between the m^{th} and the $(m+1)^{\text{th}}$ slope
$d_{EXTRAP,j}$	Distance between the extrapolated grid point coordinate and the j^{th} measurement coordinate
$d_j^{(r)}$	Distance between the estimated TX location and the location of the j^{th} measurement for the r^{th} TX
$\hat{d}_{RANGE}^{(r)}$	The estimated distance (or range) to the r^{th} heard TX in trilateration
d_{REF}	Reference distance in the logarithms used with the PL models
D_{decor}	De-correlation distance defining the magnitude of the shadowing correlation
f	Floor index
$f_{carr}^{(r)}$	Carrier frequency of the r^{th} TX
$F_L^{(r)}$	Floor loss of the r^{th} TX
g	Fingerprint grid interval
$\mathbf{h}_j^{(r)}$	j^{th} row in the matrix $\mathbf{H}_{multis}^{(r)}$ of the r^{th} TX
$\mathbf{H}_{floorL}^{(r)}$	System matrix used to present single-slope PL models including the floor losses (and possibly the carrier frequency)
$\mathbf{H}_{logdist}^{(r)}$	System matrix used to present single-slope log-distance P models
$\mathbf{H}_{method}^{(r)}$	System matrix for the desirable PL modeling method (*method* equals either *logdist* or *multis*) of the r^{th} TX

$\mathbf{H}_{multis}^{(r)}$	System matrix used to present multi-slope log-distance PL models
i	Fingerprint index in the learning database
$\mathrm{I}_m(\cdot)$	Indicator function used for determining whether the argument is larger than the breakpoint distance $d_{BP,m}$ or not.
j	Index for the TX-wise organized RSS values
k	Running sum index used in the estimation algorithms
K_{NN}	Number of nearest neighbors included in the location estimate in KNN algorithm
l	Running sum index used in the estimation algorithms
m	PL model slope index
$n^{(r)}$	PL exponent of the r^{th} TX
$n_m^{(r)}$	PL exponent for the m^{th} slope of the r^{th} TX
$N_{counter}$	Number of available RSS measurements for a certain TX in one location
$N_{floors,j}^{(r)}$	Number of floors between the j^{th} measurement and the location of the r^{th} TX
N_{FP}	Total number of fingerprints in the database
$N_{FP}^{(r)}$	Number of RSS measurements for the r^{th} TX
N_{high}	Number of used TX location trials (with the smallest error norm) in the final TX location estimate with the deconvolution method
$N_{RSS,i,r}$	Number of observed RSS measurements in the i^{th} fingerprint and r^{th} TX
N_{slopes}	Number of slopes in the multi-slope PL model
$N_{TXheard}$	Number of heard TXs in the user measurement
$p(\cdot)$	Probability density function
$p_{RSS}(\cdot)$	Probability density function of the RSS distribution for one TX in a single location
P_{EXTRAP}	Extrapolated RSS value
$P_{i,r}$	RSS value of the r^{th} TX in the i^{th} fingerprint stored in the learning database

$\tilde{P}_{i,r,q}$	q^{th} RSS measurement of the r^{th} TX in the i^{th} fingerprint (before taking the average RSS)
$P_{i,r,updated}$	Updated RSS value in the recursive RSS mean calculation in the i^{th} fingerprint and r^{th} TX
$P_j^{(r)}$	RSS value of the j^{th} measurement of the r^{th} TX
$P_{USER,r}$	RSS value for the r^{th} heard TX in the user measurement
$\hat{P}_{\mathbf{x}_g}^{(r)}$	Estimated RSS value for the r^{th} TX at any coordinate $\mathbf{x}_g$
$\mathbf{P}^{(r)}$	Vector of the RSS values of the r^{th} TX
$\hat{\mathbf{P}}^{(r)}$	Vector of the estimated RSS values of the r^{th} TX
q	Index for the RSS measurements for a certain TX in one location (before taking the average RSS)
$\mathbf{Q}^{(r)}$	Weighting matrix for the r^{th} TX used with WLS
r	TX index
$R_{shadow}(\cdot)$	Shadowing autocorrelation function
s	Arbitrary fingerprint index used in the random fingerprint removal method
u	Design parameter for the IDW method
$W_j^{(r)}$	Noise term for the j^{th} measurement of the r^{th} TX in the PL modeling process
$\mathbf{W}^{(r)}$	Noise vector r^{th} TX in the PL modeling process
x_{EXTRAP}	x coordinate of the extrapolated RSS value
x_i	x coordinate of the i^{th} fingerprint
$x_j^{(r)}$	x coordinate of the j^{th} measurement of the r^{th} TX
$\hat{x}_{TX}^{(r)}$	x coordinate of the estimated location of the r^{th} TX
x_{USER}	x coordinate of the true user location
$\mathbf{x}_g$	Coordinate vector of a grid point at an arbitrary location (not necessarily within the coverage of the database)
$\mathbf{x}_i$	Coordinate vector of the i^{th} grid point

$\hat{\mathbf{x}}_{TX}^{(r)}$	Coordinate vector of the estimated location of the r^{th} TX
$\mathbf{x}_{USER}$	Coordinate vector of the true user location
$\hat{\mathbf{x}}_{USER,KNN}$	Coordinate vector of the estimated user location by using KNN method
$\hat{\mathbf{x}}_{USER,MAP}$	Coordinate vector of the estimated user location by using MAP method
$\hat{\mathbf{x}}_{USER,MMSE}$	Coordinate vector of the estimated user location by using MMSE method
$\hat{\mathbf{x}}_{USER,NN}$	Coordinate vector of the estimated user location by using NN method
$\hat{\mathbf{x}}_{USER,trilater}$	Coordinate vector of the estimated user location by using trilateration
$\hat{\mathbf{x}}_{USER,WKNN}$	Coordinate vector of the estimated user location by using WKNN method
y_{EXTRAP}	y coordinate of the extrapolated RSS value
y_i	y coordinate of the i^{th} fingerprint
$y_j^{(r)}$	y coordinate of the j^{th} measurement of the r^{th} TX
$\hat{y}_{TX}^{(r)}$	y coordinate of the estimated location of the r^{th} TX
y_{USER}	y coordinate of the true user location
z_i	z coordinate of the i^{th} fingerprint
$z_j^{(r)}$	z coordinate of the j^{th} measurement of the r^{th} TX
$\hat{z}_{TX}^{(r)}$	z coordinate of the estimated location of the r^{th} TX
z_{USER}	z coordinate of the true user location
Δd	Distance difference in the shadowing autocorrelation function
$\boldsymbol{\theta}_{floorL}^{(r)}$	PL parameter vector of the r^{th} TX for the PL model including the floor losses
$\hat{\boldsymbol{\theta}}_{floorL}^{(r)}$	NNLS estimate of the PL parameter vector of the r^{th} TX for the PL model including the floor losses
$\boldsymbol{\theta}_{logdist}^{(r)}$	PL parameter vector of the r^{th} TX for the log-distance PL model
$\hat{\boldsymbol{\theta}}_{logdist,LS}^{(r)}$	LS estimate of the PL parameter vector of the r^{th} TX for the log-distance PL model

$\boldsymbol{\theta}_{method}^{(r)}$	PL parameter vector for a desirable method (*method* equals either *logdist* or *multis*) of the r^{th} TX
$\hat{\boldsymbol{\theta}}_{method,LS}^{(r)}$	LS estimate of the PL parameter vector for a desirable method (*method* equals either *logdist* or *multis*) of the r^{th} TX
$\hat{\boldsymbol{\theta}}_{method,MMSE}^{(r)}$	MMSE estimate of the PL parameter vector for a desirable method (*method* equals either *logdist* or *multis*) of the r^{th} TX
$\hat{\boldsymbol{\theta}}_{method,WLS}^{(r)}$	WLS estimate of the PL parameter vector for a desirable method (*method* equals either *logdist* or *multis*) of the r^{th} TX
$\boldsymbol{\theta}_{multis}^{(r)}$	PL parameter vector of the r^{th} TX for the multi-slope PL model
$\hat{\boldsymbol{\theta}}_{multis,LS}^{(r)}$	LS estimate of the PL parameter vector of the r^{th} TX for the multi-slope PL model
μ	Removal percentage (the percent of removed fingerprints in the random fingerprint removal process)
$\boldsymbol{\mu}_{\theta}^{(r)}$	Mean of the prior distribution of the PL parameter vector of the r^{th} TX
v	Function argument for the $p_{RSS}(\cdot)$
σ_{RSS}^2	Standard deviation of the RSS distribution used in the Gaussian shaped $p_{RSS}(\cdot)$
$\upsilon_j^{(r)}$	Coordinate weight for the j^{th} measurement and the r^{th} TX used with the weighted centroid method
ψ_i	Value of the cost function in fingerprinting at the i^{th} fingerprint
$\chi_i(\mathbf{x}_{USER})$	Indicator function for determining whether the argument coordinate is mapped to the i^{th} fingerprint of not
$\Omega_f^{(r)}$	Set of measurement indices of the r^{th} TX found in the f^{th} floor
Ω_{FULL}	Set of all fingerprint indices in the original learning database (the starting point in the random fingerprint removal process)
Ω_{KNN}	Set of fingerprint indices i pointing out the K_{NN} smallest values of the cost function ψ_i.

$\Omega_{PARTIAL}$	Set of fingerprint indices in the partial learning database after some fingerprints have been artificially removed from the database in the random fingerprint removal process
$\Omega_{RSS,i,r}$	Set of RSS measurements taken from the r^{th} TX in the i^{th} fingerprint (before a taking the average RSS)
$\Omega_{RSS,USER}$	Set of RSS measurements observed by the user
Ω_{TX}	Set of all TX indices in the learning database
$\Omega_{TXheard}$	Set of TX indices heard in the user measurement

List of Abbreviations

2G	2nd Generation
3G	3rd Generation
4G	4th Generation
AOA	Angle-of-Arrival
AP	Access Point
API	Application Programming Interface
ARFCN	Absolute Radio-Frequency Channel Number
ASIC	Application Specific Integrated Circuit
BCCH	Broadcast Control Channel
BLE	Bluetooth Low Energy
BS	Base Station transmitter
BSIC	Base Station Identity Code
CPICH	Common Pilot Channel
EM	Expectation Maximization
EPC	Electronic Product Code
FSPL	Free-Space Path Loss
GNSS	Global Navigation Satellite System
GSM	Global System for Mobile Communications
IDW	Inverse Distance Weighting
ISM	Industrial, Scientific and Medical
KNN	K-Nearest Neighbor
LOS	Line-Of-Sight
LS	Least Squares
LTE-A	Long Term Evolution – Advanced
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MVU	Minimum Variance Unbiased estimator

NLOS	Non-Line-Of-Sight
NN	Nearest Neighbor
NNLS	Non-Negative Least Squares
OFDMA	Orthogonal Frequency-Division Multiple Access
PL	Path Loss
RFID	Radio Frequency Identification
RSCP	Received Signal Code Power
RSRP	Reference Symbol Received Power
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RXLEV	Received signal level (GSM notation)
SLAM	Simultaneous Localization and Mapping
SNR	Signal-to-noise-power ratio
TDOA	Time-Difference-Of-Arrival
TOA	Time-Of-Arrival
TX	Radio transmitter
UARFCN	UMTS Terrestrial Radio Access ARFCN
UHF	Ultra-High Frequency
WCDMA	Wideband Code Division Multiple Access
WKNN	Weighted K-Nearest Neighbor
WLAN	Wireless Local Area Network
WLS	Weighted Least squares

List of Publications

This thesis is a compilation of the following eight publications:

[P1] J. Talvitie, M. Renfors and E.S. Lohan, "Distance-based Interpolation and Extrapolation Methods for RSS-Based Localization with Indoor Wireless Signals," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1340-1353, Apr. 2015.

[P2] J. Talvitie, E.S. Lohan and M. Renfors, "The Effect of Coverage Gaps and Measurement Inaccuracies in Fingerprinting based Indoor Localization," in *Proc. Int. Conf. Localization and GNSS*, Helsinki, Finland, pp. 1-6, 2014.

[P3] J. Talvitie, M. Renfors and E.S. Lohan, “A Comparison of Received Signal Strength Statistics between 2.4 GHz and 5 GHz bands for WLAN-based Indoor Positioning,” accepted in *Proc. IEEE Globecom 2015 Workshop Localization for Indoors, Outdoors, and Emerging Networks*, San Diego, CA, USA, Dec. 2015.

[P4] S. Shrestha, J. Talvitie and E.S. Lohan, "Deconvolution-based Indoor Localization with WLAN Signals and Unknown Access Point Locations," in *Proc. Int. Conf. Localization and GNSS*, Turin, Italy, pp.1-6, 2013.

[P5] J. Talvitie and E.S. Lohan "Modeling Received Signal Strength Measurements for Cellular Network based Positioning," in *Proc. Int. Conf. Localization and GNSS*, Turin, Italy, pp. 1-6, 2013.

[P6] S. Shrestha, J. Talvitie and E.S. Lohan, "On the Fingerprints Dynamics in WLAN Indoor Localization," in *Proc. Int. Conf. ITS Telecommun*., Tampere, Finland, pp. 122-126, 2013.

[P7] E. S. Lohan, J. Talvitie, P. Figueiredo e Silva, H. Nurminen, S. Ali-Loytty and R. Piche, "Received Signal Strength Models for WLAN and BLE-based Indoor Positioning in Multi-floor Buildings," in *Proc. Int. Conf. Localization and GNSS*, Gothenburg, Sweden, pp. 1-6, 2015.

[P8] E.S. Lohan, K. Koski, J. Talvitie and L. Ukkonen, "WLAN and RFID Propagation Channels for Hybrid Indoor Positioning," in *Proc. Int. Conf. Localization and GNSS*, Helsinki, Finland, pp. 1-6, 2014.

1 Introduction

1.1 Background and motivation

During the last two decades, wireless communications have become a significant part of our everyday life. The new continuously developing wireless technologies have brought about new opportunities for utilizing the available radio signals. One of the most attractive use cases is the radio-network-based localization, where the available wireless signals are exploited for the localization purposes. Although satellite-based localization with Global Navigation Satellite Systems (GNSS) can provide high localization accuracy in outdoor environments at the global scale, the challenge of the development of a global scale indoor localization system remains unsolved. Moreover, the performance of the localization system should be founded not only on the actual localization accuracy, but also on the overall cost of the localization system. The overall cost is composed of economic factors, such as the implementation and maintenance costs of the system, and operating factors, such as the energy-efficiency factor. The latter one is a crucial factor for the user experience due to the possibility of enhancing the battery durability in the wireless devices. For example, GNSS enables accurate navigation for vehicles, but it is unnecessarily accurate and energy-inefficient for location-aware marketing that can run in the mobile device as a background process informing the user whenever something interesting emerges nearby. Hence, as pointed out in [77], a single localization system is unable to meet all the requirements set by the industry, and therefore, continuing studies over a variety of positioning technologies is necessary.

The market of the indoor localization has been increasing rapidly over the recent years and the growth is still accelerating [116],[129]. This has produced an increasing demand for cost-efficient wireless indoor localization systems operating at the global scale. As a result, the use of the already available wireless networks has been proposed in the literature for the localization purposes [10],[37],[49],[59],[85],[147]. Here, the fundamental advantage is that no additional hardware is required for the new localization system, and moreover, the availability of the wireless systems,

such as Wireless Local Area Networks (WLAN) [64] and Bluetooth Low Energy (BLE) [27], is yet expected to be increased in the future.

Most of the wireless communications networks provide a straightforward access to the Received Signal Strength (RSS) values, which has made the RSS-based localization one of the most attractive network-based localization approaches. Probably the most often mentioned RSS-based localization approach is the fingerprinting studied, for example, in [16],[21],[76],[80],[81],[92],[94],[109],[117],[138],[141]. Here, the fundamental idea is to first collect learning data from the target area and then use it later in the user localization, by comparing the user measurements with the learning data. However, if the target area, the number of the observable wireless networks, and the number of radio transmitters (TX) grow, the size of the learning database might become intolerably large. Even if there would be enough storage capacity in the database, the data traffic handled by the database server might become a bottleneck of the system. To tackle this problem, by using Path Loss (PL) models, such as in, for example, [1],[100],[101],[112],[122], the size of the database can be significantly reduced. However, due to harsh and unpredictable radio propagation environment, the accuracy of the PL models might sometimes be inadequate, which directly results in decreased localization performance. There is thus a fundamental tradeoff between the accuracy of a localization algorithm and its required database size and storage capacity, a tradeoff that is addressed all through this thesis.

We use the above-mentioned TX acronym for any transmitting radio device, whose signal can be utilized for the RSS-based localization. These are, for example, Access Points (AP) in WLANs and Base Station transmitter (BS) in cellular networks.

1.2 The scope and objectives of research

The main focus in this thesis is to study the RSS-based localization from the fingerprinting and PL modeling point of view and analyze the performance of the localization systems under the influence of different error sources and coverage gaps in the learning data. We focus on the two-step localization approach, which is divided into the learning data collection phase and the user localization phase, often referred to as the offline phase and the online phase [37]. However, there are also localization approaches available in the literature, which perform the localization while simultaneously learning about the environment. These types of approaches are generally referred to as Simultaneous Localization and Mapping (SLAM), which is described in a tutorial format in [22] and [44]. Moreover, in [95] a computationally efficient fastSLAM-algorithm was introduced. The Bayesian-based fastSLAM-algorithm exploits specific tree structures for providing considerable reduction

in the computation complexity compared to traditional Kalman-filter-based approaches. Furthermore, in [133] the full SLAM problem (or offline SLAM problem), where the localization and mapping are done only after all the data has been collected, was solved by introducing a novel GraphSLAM-algorithm. However, on top of this, in this thesis we are also interested in the learning data from the RSS statistics point of view, since such knowledge provides valuable information for the SLAM approach and for the RSS-based localization algorithms in general. Having a clear understanding of the properties of the radio propagation environment and the RSS statistics is in vital role when developing new localization algorithms and PL models and creating computer-based simulations for the localization purposes.

The main focus of the thesis regarding the localization systems is on the 3D indoor localization with multi-storey buildings, but also outdoor scenarios for suburban and urban environments are considered. In this thesis, the vertical coordinates of the 3D systems are always discretized to the known floor heights, which is adequate for the considered use cases. For the indoor case, we study WLANs at both 2.4 GHz and 5 GHz carrier frequencies, BLE beacon signals at 2.4 GHz carrier, and passive Radio Frequency Identification (RFID) tags at Ultra-High Frequency (UHF). For the outdoor case we consider 2nd generation (2G) and 3rd generation (3G) cellular networks, namely the Global System for Mobile Communications (GSM) [2] and the Wideband Code Division Multiple Access (WCDMA) [5]. For all the considered cases we are interested in the observed RSS statistics and the PL model parameters, and especially, in how the parameters differ between separate localization systems. Besides these, one of the objectives is to compare the localization accuracy between the considered localization systems and include the aspect of the database size in the performance comparison. Thus, we consider the two-step localization approach and study the advantages and disadvantages of the traditional fingerprinting and PL-model-based approaches. Both of these approaches require collection of the training data, is first processed and stored in the learning database. The fundamental comparison between the performances of the two approaches is performed based on the localization accuracy and the required size of the learning database. Furthermore, since RSS-based localization introduces various error sources in the localization estimation, we are also highly interested in studying the effects of different error sources on the localization accuracy.

The essential description of the considered localization system in this thesis is illustrated in Fig. 1-1. One example of a similar type of a system is the Horus, presented earlier in [147] and [148]. The Horus is a WLAN-based indoor localization system, which uses the two-step fingerprinting approach including the offline-phase and online-phase (i.e., learning data collection phase and the user localization phase). In addition, the Horus system incorporates several algorithms and meth-

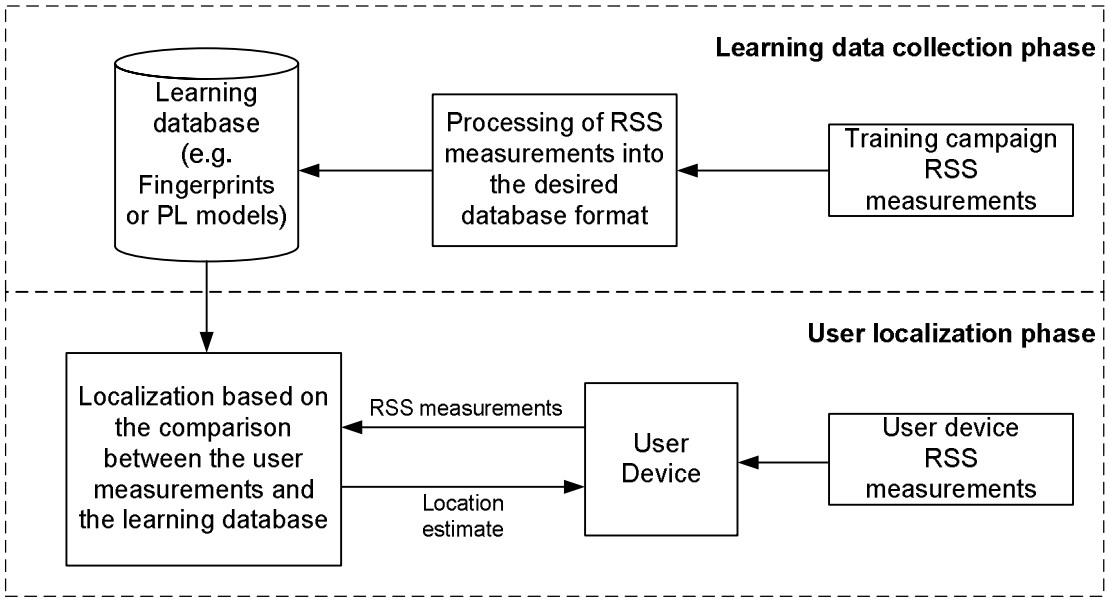

Fig. 1-1 *An example of the considered two-step localization system including the training phase and the user localization phase. In this thesis, we concentrate on the mobile-centric approach where the localization is performed by the user device (i.e., the localization block is physically located inside the user device).*

ods for handling different statistical properties of the RSS measurements, such as the temporal and statistical correlation of the RSS measurements taken from the same AP.

1.3 Main contributions of the thesis

This compound thesis consists of 8 publications, none of which has been used as a part of any other PhD thesis. The main contributions of the thesis can be concisely described as follows:

- Different RSS interpolation and extrapolation algorithms are studied and derived to estimate the RSS values in areas, where no learning data has been collected.
- A randomized model for simulating RSS coverage gaps in the learning database is derived. Based on the model, the effect of coverage gaps on the localization performance is studied and the above-mentioned interpolation and extrapolation methods are proposed to alleviate the reduction in the localization performance with coverage gaps.
- Novel deconvolution approach based on estimating the PL models without the knowledge of TX locations is presented. In addition, characteristics of traditional PL models are compared between different localization systems, including both indoor and outdoor localization, and different carrier frequencies in case of WLAN (2.4 GHz and 5 GHz).

- The characteristics of the RSS measurements obtained during the learning phase are discussed. These include the format of the RSS distribution in a fixed location, and the format of the RSS shadowing distribution with spatial correlation.
- The effect of the learning database calibration error and the bias error on the performance of the RSS-based localization system is evaluated. These types of errors occur when there are not enough repetitive RSS measurements collected for the learning data, or there is a bias between the RSS values of the user device and the device used for collecting the learning data.

In [P1] and [P2] we studied the effect of coverage gaps on the localization accuracy and on the floor detection probability in a multi-storey building with fingerprinting approach. For this we developed a randomized process to simulate incomplete fingerprint database with realistic coverage gaps caused by an inadequate learning data collection. To reduce the negative effect of the coverage gaps, in [P1] we proposed several different interpolation and extrapolation methods to retrieve the missing fingerprint data and showed that with proper interpolation and extrapolation the average localization error could be decreased by up to 12%. In addition, in [P2] we studied the effect of different error sources on the localization performance, such as the database calibration errors and RSS bias errors.

In [P3]-[P8] we studied and introduced new models regarding the estimation of PL model parameters and the RSS statistics in various localization systems. In [P3] we compared the RSS characteristics between the 2.4 GHz and 5 GHz carriers for the indoor WLAN with different building types, including office buildings and shopping malls. With PL models we either have to assume the TX positions known, or we have to estimate them with the training data. Hence, in [P4] we proposed a method for estimating the TX location and the PL parameters jointly by using a specific deconvolution principle. In [P5] we studied the modeling of the RSS measurements in cellular networks, and in [P6], we studied RSS measurement distributions and tested how the localization accuracy was changed after a re-configuration of a WLAN system. Furthermore, a comparison of the PL model characteristics between WLAN and RFID was conducted in [P8].

In [P7] a comparison of the localization performance between WLAN and BLE was done by considering multiple different localization approaches. Moreover, in [P4] and [P7] also the aspect of the database size was taken into account in the performance comparison.

1.4 Outline of the thesis

This thesis is organized in three conceptual parts, which can be briefly described as the learning phase part, the PL modeling part, and the user localization part.

In Chapter 2 we first discuss the suitability and motivation for using RSS measurements for localization purposes and provide methods on how to access the RSS data in all the considered localization systems. After this, the RSS measurements are used to build-up the learning database similar to the learning phase step in the fingerprinting approach. Besides the description of the database, we also discuss different aspects of the learning data collection and analyze distributions of the RSS values.

In Chapter 3, we introduce several feasible PL models for the localization purposes in both indoor and outdoor environments. In addition, we analyze the distribution of shadowing, which describes the local variations of the RSS values around the PL model, and we use the obtained results to generate TX-wise RSS models based on computer-simulations. After this, we introduce various approaches, such as the Least Squares (LS) method and the Minimum Mean Square Error (MMSE) method, to estimate the PL model parameters. In the end, we compare the PL model characteristics between all the considered localization systems, including both indoor and outdoor cases.

In Chapter 4 we focus on various different deterministic and probabilistic user localization algorithms by considering the fingerprinting and PL-model-based approaches. Then, by using the presented algorithms, the localization accuracy is studied under influence of different error sources in the learning database. In addition to this, we study the effect of database coverage gaps and introduce sufficient interpolation and extrapolation methods to alleviate the effect of the incomplete learning data. Finally, the localization performance between the considered localization systems is compared and analyzed.

Finally, in Chapter 5, we draw conclusions by unifying the most important results presented in the thesis. Furthermore, we address some remarks on the possible future studies for the RSS-based localization.

1.5 Author's contribution to the publications

For all the publications [P1]-[P8], the basis of the computer analysis software, used to process, analyze and visualize the RSS measurements and the PL models, was written by the author. In

[P1],[P2],[P3],[P5], the author derived all the main results and performed the majority of the writing work. In publications [P4],[P6],[P7],[P8], the contribution of the author regarding the derivation of the main results is considered to be equal with the corresponding first authors. However, in these papers, most of the writing effort was contributed by the corresponding first authors. In addition, the author was the sole collector of the outdoor RSS measurement data and participated in the collection of the indoor measurement data which were used in the results of the publications. Dr. Elena Simona Lohan ([P1]-[P8]) and Prof. Markku Renfors ([P1]-[P3],[P5]) provided valuable comments and suggestions for the publications.

2 RSS Measurements and Learning Phase: Generation and Calibration of a Learning Database

Before concentrating on the technical details of the RSS-based localization approaches, it is important to justify the use of RSS in wireless localization systems. Therefore, in this chapter we first discuss the applicability of RSS in localization systems by stating several reasons favoring the RSS approach. Moreover, we describe how the RSS can be accessed in various communications systems considered later in the thesis. After this, we describe the generation of the learning phase used in the considered RSS localization approaches and study the statistics of the RSS measurements.

There are generally two fundamental approaches in the RSS-based localization: the approach with assumed prior information on the environment, such as the SLAM methods [22],[44], and the two-step approach without such prior information. In the two-step approach, there are two separate phases in the localization. Firstly, there is the learning phase, where data from the target area is collected and stored in the learning database. Secondly, there is the user localization phase, where the localization is performed based on the data obtained from the learning phase. In this thesis we consider the two-step approach, because the availability of learning data enables studies regarding PL models and RSS measurement distributions, and thus, it provides more insight into the properties of the radio propagation environment. Moreover, understanding the RSS measurement behavior in various radio propagation environments is a vital issue in developing new SLAM-based localization approaches.

There are many different ways to collect the learning data and to construct the actual learning database. Besides the amount of collected and stored learning data, also many practical issues regarding the data collection affect the eventual user localization performance. In this chapter, we discuss about some practical considerations during the data collection process and describe the structure of the database used later in this thesis. In addition, we study the distribution of the RSS

measurements taken from the same TX in one location. The characteristics of the RSS distribution are important in understanding the underlying error sources in the localization systems.

2.1 Suitability of RSS in localization systems

RSS is one of the most important quantities in modern localization systems. This is because together with the noise power, the RSS defines the signal-to-noise-power ratio (SNR) of a communications signal. Moreover, according to the famous Shannon's law (also Shannon-Hartley theorem) [123], the SNR defines the achievable capacity of the communications system. In several present communications systems, including cellular networks and WLAN, the available bit rate of the network user is dynamically adjusted by modifying the modulation and coding scheme of the used communications waveform, according to the observed signal quality. Therefore, constant monitoring of the RSS is typically incorporated into common functionalities of a communications system, as it is extremely advantageous to approaching theoretically achievable system bit rates. In addition, the RSS measurements are also playing an important role in radio resource management as they can be used in monitoring the signal levels of neighboring BS in cellular networks or AP in WLANs [78]. Monitoring the RSS of neighboring cells can be used, for example, in advanced interference management and in making decisions of handovers by which the user device switches the serving cell [78]. Due to the above facts, the RSS measurements are expected to be maintained also in future communications systems, which only increases the motivation of studying and exploiting RSS in the localization-based services.

From the localization point of view, RSS provides information on the distance between a radio transmitter and a radio receiver, such as between the AP and the user device in WLANs. Assuming the knowledge of the location of the radio transmitter and its transmission power, it is possible to approximate the physical distance between the communicating radios based on the observed RSS value by using proper PL models, which are further discussed in Chapter 3. Moreover, with multiple distance approximations from multiple radio transmitters, it is possible to estimate the location of the radio receiver geometrically by trilateration [48],[90],[128],[145]. However, due to a vastly dynamic and heterogeneous radio propagation environment [112],[122], the accuracy of PL models is often relatively poor, especially when most of the system parameters, such as the transmitter locations, are unknown. In addition, the majority of the PL models available in the literature are not fully adequate to be exploited in localization systems, since the required radio environment parameters are unavailable [28],[82]. Nevertheless, PL-based localization has certain

considerable advantages, such as a small learning database size [P4],[P7], over the traditional fingerprinting methods, which makes it a noteworthy option for many localization systems.

Another of the most appealing and practical reasons to use the RSS in localization is its availability in Application Programming Interfaces (API) of the most common operating systems. For example, Android (Google Inc.) and Windows Phone (Microsoft Inc.) provide a direct access to the RSS measurements of hearable WLAN APs via their operating system APIs [53],[93]. This makes the usage of RSS measurements in localization services attractive, since it enables the development of localization-based applications by only creating new software updates and without the need to install or access into specific hardware components. In many other well-known localization methods, such as Angle-Of-Arrival (AOA), Time-Of-Arrival (TOA) and Time-Difference-Of-Arrival (TDOA), dedicated measurement hardware and accurate synchronization algorithms [9] are required to be used in the devices, which can increase the cost of the localization system. Hence, this particular economical aspect of the RSS-based localization favors its status in the future market of localization-based services [116],[129].

Although GPS is able to provide accurate localization outdoors, there is still motivation for the RSS-based approaches. Whereas GPS outperforms the RSS-based approaches in localization accuracy and localization reliability in outdoor scenarios, RSS-based approaches support outstanding energy-efficiency in many use cases. For example, in cellular-based localization, the RSS measurements are continuously monitored during the normal network operations [6]. Thus, the RSS measurements can be considered to be obtained as free of charge from the energy-efficiency and localization points of view, because no additional signaling is required. Moreover, if the learning data is stored beforehand in the user device, the localization can be performed without any help from an online localization server. In this case, the only additional effort for acquiring the user location estimate is in the computational burden of the desired localization estimation algorithm. It means that the RSS-based localization can also run as a background process and provide constant localization awareness for the user device, which can be further exploited in many location-based services. In [96], the energy-efficiency of a localization system was optimized by using game-theoretical algorithms.

2.2 Accessing RSS indicators in localization systems

The RSS measurements can be accessed via the API of the operating system found in the measurement device. Depending on the operating system in the device and on the measured communications network, the extent of radio measurement reports might differ, but at least the required

RSS measurement and the corresponding radio transmitted identities can be observed in the majority of the communications networks. There are several RSS-like indicators in the radio measurement reports, such as Received Signal Code Power (RSCP), Received Signal Strength Indicator (RSSI) and E_c/N_0 (Energy per Chip to Noise power spectral density ratio) in WCDMA cellular networks [5]. It is very important to understand which of the available indicators represents the essential RSS value to be considered in the localization systems.

In 2G cellular networks, namely the GSM, the RSS measurements, generally referred to as the received signal level (RXLEV) in the standard [2], are reported together with the Absolute Radio-Frequency Channel Numbers (ARFCN) on the active cell and each neighbor cell. The measurements are done based on the Broadcast Control Channel (BCCH), which is a logical channel working under the ARFCN, regardless of the device being in idle or connected mode [2]. The BS identification is done based on the reported ARFCN and on the Base Station Identity Code (BSIC) of each BS. If the radio network planning has been appropriately conducted, there should always be a unique pair of ARFCN and BSIC for each BS heard in the same area. Typically the actual cell identity indicator is reported only for the serving cell.

In the GSM system the RSS measurements from different BSs are taken from orthogonal channels in time and frequency, and thus, the measurements from different BSs do not practically interfere with each other. However, in 3G networks, here referred to as WCDMA networks, the Node Bs, referred to for simplicity also as BSs, can operate in the same frequency simultaneously as the signal separation is managed in the code domain [4]. This implies that the orthogonality between the signals transmitted by the BSs is achieved only in the code domain and the traditional signal power measurements do not reveal the BS-wise RSS levels. Now, since there are different RSS-like indicators found in the 3G measurement report, it is important to understand which of the measurements are relevant in the localization context. The RSS measurements are generally obtained via the Common Pilot Channel (CPICH), which provides two types of signal level measurements: RSSI and RSCP. Here, the RSSI measures the total signal power over the measured physical channel, which can contain signals from multiple BSs due to the used CDMA approach. As a result, the RSSI is not an appropriate RSS measure for the RSS-based localization, since separate BSs cannot be properly distinguished from each other and RSSI measures a joint effect of all BSs. Instead, RSCP is a proper RSS measurement, since it is obtained after processing the signal in the code domain by de-scrambling the received signal with the BS-wise scrambling code. Thus, by using the RSCP, the RSS levels from separate BSs can be appropriately distinguished and exploited for localization purposes. Similar to the ARFCN and BSIC in GSM, a unique combination of the used channel frequency, described by the UMTS Terrestrial Radio Access ARFCN (UARFCN)

in the WCDMA [3], and the scrambling code should provide a unique BS identity for each region. In addition, similar to the GSM, the global cell identity is typically reported only for the serving cell.

In 4th generation (4G) cellular networks, namely the Long Term Evolution – Advanced (LTE-A), the RSS measurement set is very close to the one used in the WCDMA. In LTE-A, there are also the RSSI measurements, which include the power from the whole signal band at the used carrier frequency. Since the LTE-A is based on the Orthogonal Frequency-Division Multiple Access (OFDMA) scheme, the RSSI measurement includes also the interference from neighboring cells using the same carrier frequency. Thus, RSSI is again not a useful RSS measure in LTE. Instead of the RSSI, the signal power from a certain BS can be obtained from the Reference Symbol Received Power (RSRP) measurement [6], which offers the corresponding RSS measurements for the desired localization purposes. The cell identification can be obtained from the primary and secondary synchronization sequences used by the BS, but also a global cell identity can be found in the broadcasted system information block.

In WLANs and BLE networks, the RSS value of the heard AP or the BLE beacon is defined based on the signal strength measurement from a specific preamble or a beacon included in the received signal. Unlike in the case of cellular networks, which have their own dedicated frequency bands, WLANs and BLE networks operate in a contested frequency band, namely as the Industrial, Scientific and Medical (ISM) band. For this reason, the RSS measurements from WLANs and BLE networks may contain interference from other TXs. In addition, one considerable issue with the RSS measurements in WLANs [64] is the interpretation of the RSS measurements, which are often given differently by each chip-set vendor. This should be taken into account in the design of the localization system by carrying out a separate calibration phase for handling different chip-set vendors. Otherwise, the localization accuracy might drop drastically. The identity of WLAN APs or BLE beacons can be found by globally unique MAC addresses found in the control fields of the received signal frame.

Compared to all above-mentioned communications networks, passive RFID tags introduce a noticeably distinctive approach for accessing and exploiting the RSS information. The localization with RFID tags have been earlier studied, for example, in [34],[121],[151]. Since passive RFID tags are not transmitting any communications signals of their own, the RSS measurements are based on backscattered signal powers. When the tag is read with the tag reader device, the Application Specific Integrated Circuit (ASIC) attached to the RFID tag modulates and emits the signal back to the reader device. This signal returning from the RFID tag to the tag reader is called the backscattered signal. The modulation performed in the ASIC conveys a specific Electronic Product Code (EPC), which can be used to identify the tag. Thus, the backscattered signal power of passive

RFID tags can be directly used in RSS-based localization purposes in a similar manner as the reference signals in cellular networks, WLANs, and Bluetooth beacons, and an example of this is shown in [P8].

Table 1. Methods for *accessing the RSS indicators and the corresponding TX identity information in the considered communications systems. Here, GSM, WCDMA and LTE-A are able to provide also a separate global cell-identity information, but only for the serving cell.*

System	RSS indicator name	RSS measurement channel/origin	TX identity
GSM	RXLEV	BCCH	BSIC/ARFCN
WCDMA	RSCP	CPICH	scrambling code /UARFCN
LTE-A	RSRP	Reference signals across the bandwidth	Primary/secondary synchronization sequences
WLAN	RSS (or RSSI)	Preamble/beacon	MAC address
BLE	RSS (or RSSI)	Preamble/beacon	MAC address
RFID	RSS (or RSSI)	Backscattered signal power	EPC

2.3 Collection of RSS measurements

2.3.1 Determination of the measurement coordinates

Besides measuring the actual RSS values from different TXs, determining the correct measurement coordinates is vital in the learning phase. Depending on the considered signals and communication system type, the determination of the measurement coordinates can be a fairly straightforward or a very cumbersome task. For example, in outdoor environments when cellular data is collected, there is typically an access to GNSS-based coordinate estimates, whereas indoors there are no globally valid localization systems providing adequate coordinate estimates.

When collecting the measurements from cellular networks, the exploitation of GNSS-based coordinate estimates is extremely advantageous. To manually insert each coordinate at each measurement location for large areas would be an exhausting process. Mostly, the GNSS-based coordinates are adequately accurate for the RSS-based localization purposes. If the measurements are mapped into a synthetic grid, as described in Section 2.4, the coordinate errors will be roughly less

than half of the used grid interval. Conversely, in some areas, for example in urban canyons, the GNSS-based coordinates can be rather inaccurate, which will automatically reduce the quality of the learning phase data. In addition, if cellular measurements are taken indoors, typically the GNSS coordinates become inaccurate which results in inconsistency between the nearby indoor and outdoor RSS measurements.

Due to the absence of reliable GNSS-based coordinate estimates in indoor environment, the measurement campaigns are often much more complicated indoors. Furthermore, since the target localization accuracy for indoors is typically below 2-3 meters, the tolerable errors in the learning data coordinates should be less than half of this, i.e. below 1 meter. It is clear that there are yet no global localization methods to achieve this level of accuracy for indoors. Thus, since the coordinates cannot be obtained with any existing localization system, determining the coordinates manually is one considerable option. In this case, the measurement coordinates have to be manually inserted by the measurer at each location where the RSS measurements are obtained. Although here the chance of causing substantial coordinate errors due to the human factor is evident, by carefully conducted measurement campaign and with good building maps it is still possible to have very accurate and trustworthy coordinate estimates.

Nowadays most of the smartphones have inbuilt GNSS capability, which makes different crowdsourcing-based data collection approaches very cost-effective for localization service providers [71],[120],[142],[149]. In the crowdsourcing approach, the collection of learning data is conveniently outsourced to common mobile users, which allows a straightforward access to the GNSS-coordinates and the corresponding RSS measurements in a large scale system. However, since there is no guarantee of the measurement quality, the crowdsourcing methods require specific signal processing methods for handling the measurement outliers and for monitoring the consistency of the data.

Crowdsourcing methods are also possible in indoor environment, as studied in [47],[52],[62],[110],[140],[143],[152], but in this case the complexity increases rapidly due to lack of globally available coordinate estimates. For example, by exploiting different sensors included in the mobiles, such as accelerometers, gyroscopes, magnetometers, barometers and pedometers, it is possible to generate the learning database based on advanced machine-learning algorithms. Nonetheless, for research purposes the manually determined coordinates are a safe approach, since it is always clear in which way and in which coordinates the measurements were truly taken.

2.3.2 Error sources and practical consideration of the RSS measurement campaign

The manual collection of fingerprints, including measurement coordinates and the corresponding RSS measurements, can be organized in many different ways and can lead to various outcomes of the system performance. For example, in indoors data collection, the measurement device can be attached into a specifically designed platform, where the orientation and movement of the device is extremely steady, or the measurement device can be held in hand. In [39] the performance of the localization system is compared between two cases, where in the first case the device is on the hand of the user, and in the second case the device is on a flat-surface table. In addition, the measurements can be taken during a time period when nobody else remains in the building, which reduces the influence of the radio propagation environment on the measured RSS values. These kinds of measurement arrangements are desirable for studying certain radio propagation characteristics and new localization algorithms, but often they give too optimistic results for real-life localization accuracy. Conversely, by taking the measurements as randomly as possible during different times of a day with arbitrary device orientation and with random levels of crowd, the localization results should be more realistic. On the other hand, it might be very difficult to study the underlying system models, since abrupt errors from unfamiliar error sources might occur.

Since the radio environment is not stationary, it is generally not enough to gather learning data by taking only one set of RSS measurements per each location. Especially indoors the difference of RSS levels between Line-Of-Sight (LOS) and Non-Line-Of-Sight (NLOS) signals can be significant. The LOS signal can be easily interrupted with any obscuring object including walls, doors, furniture, people and the body of the device holder [15],[102]. Although some of the obscuring objects might be stationary with respect to the building, they still move with respect to the movement of the measurement device and might any time emerge between the device and the TX. For example, in [72] it has been reported up to 20dB to 30dB signal variations due to obscured furniture and people presence in the 2.4 GHz ISM band. Thus, in order to study the characteristics of the RSS behavior in a fixed location, numerous measurements are required to reveal the distribution shape. The shape of the distribution has been further discussed in Section 2.5 and has also been briefly tackled in publications [P2],[P5],[P6].

In some localization algorithms, such as in [117], it is desired to acquire the complete distribution of RSS values from all locations, whereas in some algorithms, as in [109], only the mean of the RSS values is desired. For both of the cases, the more measurements are obtained, the more accurate distribution parameter estimates can be achieved. This procedure is often referred to as calibration of the RSS mean and its effect on the positioning performance is further studied in [P2] and in Section 4.2.

Because of the apparent uncertainties in the learning data collection, the performance of the localization system depends greatly on the variety of the conducted measurement campaign. In addition, the TX density, the building type, the area size and the number of floors are all affecting the localization performance. Therefore, in the literature it is very difficult to find a fair comparison between different localization approaches. For example, in our own studies the average indoor localization error without advanced tracking or filtering methods varies roughly between 3m and 25m depending on the considered building. The only way to have a fair comparison between different localization methods would be to use exactly the same data set in all studied cases. For this reason we have also distributed some of our own indoor measurement data publicly in [132], which allows researchers to compare their algorithms with each other by using the same reference dataset.

2.4 Database structure

It is common to map the measurements obtained from the learning data measurement campaign into a synthetic grid with some predefined discrete coordinate values, as done in [61]. In this procedure, based on the measurement coordinates, the measurements are mapped to the closest coordinates found in a predefined synthetic grid. Thus, the database size can be considerably reduced and the nearby RSS measurements can be efficiently combined together. After the grid mapping process, at each fingerprint (i.e. grid point) there are RSS measurements taken from one or multiple TXs, where each observed TX might have one or multiple RSS measurements. Thus, the set of RSS measurements taken from the r^{th} TX in the i^{th} fingerprint $\Omega_{RSS,i,r}$ is given as

$$\Omega_{RSS,i,r} = \left\{ \tilde{P}_{i,r,q} : q = 0,1,..., N_{RSS,i,r} - 1 \right\}, \tag{2.4.1}$$

where $\tilde{P}_{i,r,q}$ and $N_{RSS,i,r}$ are the q^{th} RSS measurement (in dBm), and the number of RSS measurements in the i^{th} fingerprint and taken from the r^{th} TX, respectively. Throughout the thesis, mathematical sets are always denoted with the letter Ω (omega) with appropriate subscripts and superscripts. Now by including the coordinates of the fingerprints, the measurement set can be described as

$$\left\{ x_i, y_i, z_i, \left\{ \Omega_{RSS,i,r} : r \in \Omega_{TX} \right\} : i = 0,1,..., N_{FP} - 1 \right\}, \tag{2.4.2}$$

where x_i, y_i and z_i are the x-coordinate, y-coordinate and z-coordinate of the i^{th} fingerprint, N_{FP} is the total number of fingerprints in the database, and $\Omega_{TX} = \{0,1,2,..., N_{TX} - 1\}$ is the set of TX indi-

ces used to identify the TXs, where N_{TX} is the total number of TXs in the database. For example, if TXs with indices r=3, r=7, r=14 and r=23 are heard in the fingerprint index i=0, the fingerprint can be given as $\{x_0, y_0, z_0, \Omega_{RSS,0,3}, \Omega_{RSS,0,7}, \Omega_{RSS,0,14}, \Omega_{RSS,0,23}\}$. Moreover, it should be noticed that both the number of heard TXs (i.e., the number of sets $\Omega_{RSS,i,r}$) and the number of RSS measurements for each observed TX (i.e., the number of elements in each $\Omega_{RSS,i,r}$) varies between separate fingerprints. If the altitude or the floor-index of a building is not desired to be considered in the localization system, the z-coordinate can be simply neglected.

In case multiple RSS measurements are heard from the same TX in the same grid point (i.e., there is more than one element in the set $\Omega_{RSS,i,r}$), it is possible to store either the complete histogram of the RSS values, as done in [117], or only one or several RSS distribution parameters, such as the mean of the RSS values as proposed in [109], and the standard deviation. In our studies we have consistently used the latter approach and stored only the arithmetic mean of the RSS values in case multiple RSS values from the same TX have been observed in the same grid point. Now, by considering the RSS measurements from the r^{th} TX in the i^{th} fingerprint, given in (2.4.1), we can calculate the mean RSS value $P_{i,r}$ as

$$P_{i,r} = \frac{1}{N_{RSS,i,r}} \sum_{q=0}^{N_{RSS,i,r}-1} \tilde{P}_{i,r,q} \,, \tag{2.4.3}$$

where $P_{i,r}$ is the mean RSS value stored in the learning database. Finally, the elements in a complete learning database can be described as

$$\left\{x_i, y_i, z_i, \left\{P_{i,r} : r \in \Omega_{TX}\right\} : i = 0,1,...,N_{FP}-1\right\}. \tag{2.4.4}$$

Of course, by storing only the mean RSS value, a part of the information is lost, but the required database size is much smaller compared to the case when we would store the whole RSS histogram. Moreover, using the arithmetic mean, it is possible to update the database incrementally by only keeping a count of the total number of RSS values as new measurements become available. In this case the updated RSS value in the i^{th} fingerprint and r^{th} TX in the database can be defined as

$$P_{i,r,updated} = \frac{1}{N_{counter}} \left(N_{counter} P_{i,r} + \tilde{P}_{i,r,N_{RSS,i,r}} \right), \tag{2.4.5}$$

where $P_{i,r,updated}$ is the updated RSS value, $P_{i,r}$ is the original RSS value, $\tilde{P}_{i,r,N_{RSS\,i,r}}$ is the new RSS measurement, and $N_{counter}$ is the number of measurements used to calculate the $P_{i,r}$. The incremental updating can be a great advantage in large-scale systems, since there is no need to store and process all the measurement data whenever new measurements are desired to be updated in the database. It should also be noticed that the use of the RSS mean values can be seen as the Maximum Likelihood (ML) estimation of the expected value of a Gaussian distributed random variable.

The grid coordinates can be organized either in a uniform rectangular grid, as done in [70] and [74], or in a non-uniform grid, as studied in [13] and [61]. The non-uniform grid can be especially useful for indoor localization as the grid point locations can be adapted according a specific building floor plan. By this way the grid points are always found in the middle of the corridors and in the most vital areas, whereas in the uniform grid the grid points might locate inside walls and other obstacles, which are unreasonable from the localization point of view. For this reason, the non-uniform grid might provide better localization accuracy in practice compared to the uniform grid. However, building floor plans are often unavailable, which makes the efficient utilization of non-uniform grids very difficult. Therefore, we have decided to use the uniform grid, which offers simplified implementations for certain localization algorithms and an efficient design of the database structure due to the regularity of the grid.

The density of the grid points affects the database calibration accuracy. If the grid interval is chosen to be very large, the average number of RSS measurements per TX at one grid point increases and the RSS distribution estimates improve. Additionally, a large grid interval results in a sparse fingerprint structure which saves space in the database, but it might decrease the localization accuracy. On the other hand, with a very small grid interval, the database size increases and the dense fingerprint structure enables a high resolution for the localization algorithms. However, at the same time the number of RSS measurements per TX at one grid point decreases, which automatically leads to lower quality RSS distribution estimates. Because of this, the RSS measurements from the same AP in adjacent grid points might have unrealistically large differences which cause instability in certain localization algorithms. An example of using a uniform grid interval of 1m and 5m in one floor of a University building in Tampere, Finland, is shown in Fig. 2-1.

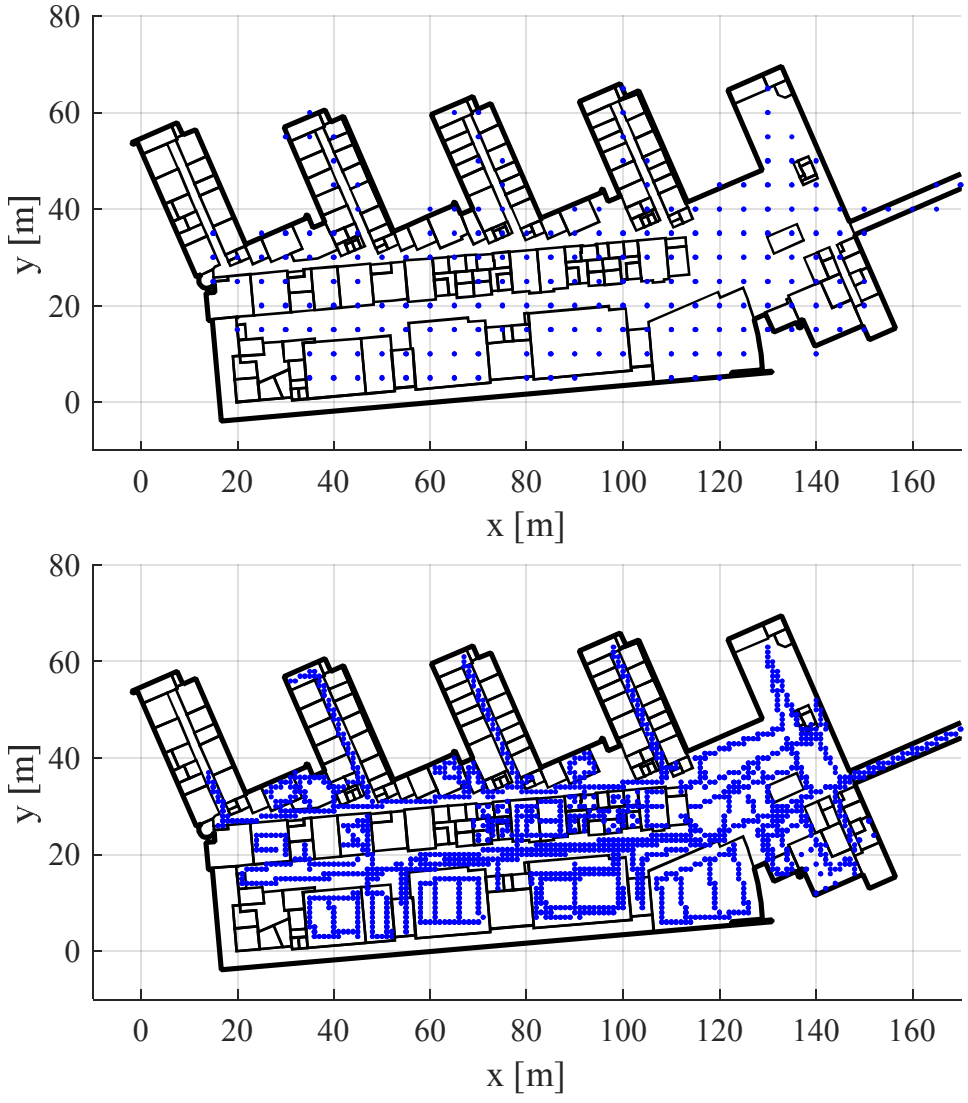

Fig. 2-1 *An example of a uniform fingerprint grid with grid interval of 5m (upper) and 1m (lower) in a University building in Tampere, Finland.*

2.5 Analysis of RSS measurement distribution

For each grid coordinate and TX there is a distinct RSS measurement distribution. More specifically, we refer to the set of RSS measurements $\Omega_{RSS,i,r}$, given by (2.4.1), where the measurements have not been processed, excepting the mapping to the grid coordinates. These local TX-wise RSS fluctuations can be caused by the changes in the radio propagation channel, the measurement error and the data traffic in a communications network. The magnitude of the fluctuation is essential to the localization performance, since it also defines the variance of the used RSS mean estimate, given in (2.4.3), in the learning phase. In addition, assuming a stationary system in time, the same RSS fluctuation is conveyed to the localization phase and it causes a similar RSS distribution as the user takes measurements from the corresponding location and TX. The RSS distribution shape and the distribution parameters have been studied in [31],[32],[68],[69] and [84]. In [P6]

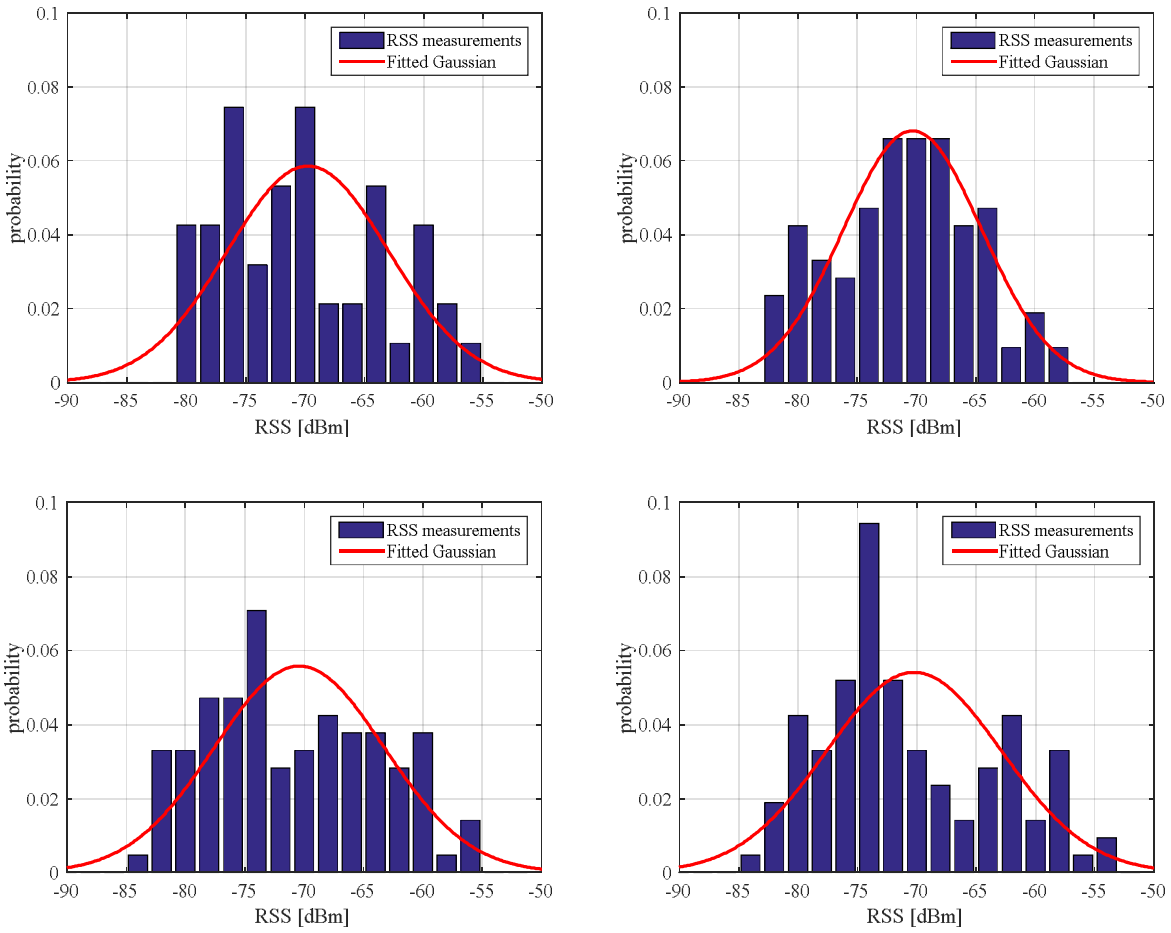

Fig. 2-2 *An example of RSS measurement histograms and the corresponding fitted Gaussian distributions observed in a single grid location by a single AP in 2.4GHz WLAN networks. Each sub-figure represents the RSS distribution for a separate AP in one specific location. All the distributions have the same mean RSS value as -70dBm.*

we studied the shape and the parameters of the RSS distribution at different distances from the TX. The results in [P6] suggested that the best fit was with the lognormal distribution (i.e., Gaussian in dB-scale) among several tested ones including the Weibull distribution. A few examples of the RSS measurement distributions and the fitted Gaussian distributions are illustrated in Fig. 2-2 for an indoor 2.4 GHz WLAN systems where all the cases have the same average RSS as -70 dBm. It can be seen that the Gaussian distribution does not fit nicely into the RSS measurements due to the heavy skewness and kurtosis visible in the RSS measurements. Nevertheless, since the RSS values are often observed in a discrete format, also the probability distribution could be defined as discrete, such as the multinomial distribution.

In many cases the distribution fitting process gets very challenging due to heavy skewness of the observed distributions. In addition, there might be multiple clearly distinguishable peaks in the distribution as it would be a mixture of different distributions. For example, in an indoor environment multiple peaks might occur, if a door between the measurement device and the TX is either open or closed. This creates a mixture of two distributions separated by a RSS bias subject to the attenuation of the door. Thus, a better fit compared to the single Gaussian fitting case, can be achieved by using a mixture of Gaussians as shown in Fig. 2-3. Here, the mixture of two Gaussians is fitted into the RSS measurements by using the Expectation Maximization (EM) algorithm [40]. The skewness of the RSS distribution has been studied in [68],[70],[126] in more detail. In addition, the distribution kurtosis was considered in [P7] and [32], and in [84], where the skewness and kurtosis of the Gaussian distribution was adjusted to get a better fit with the RSS distribution.

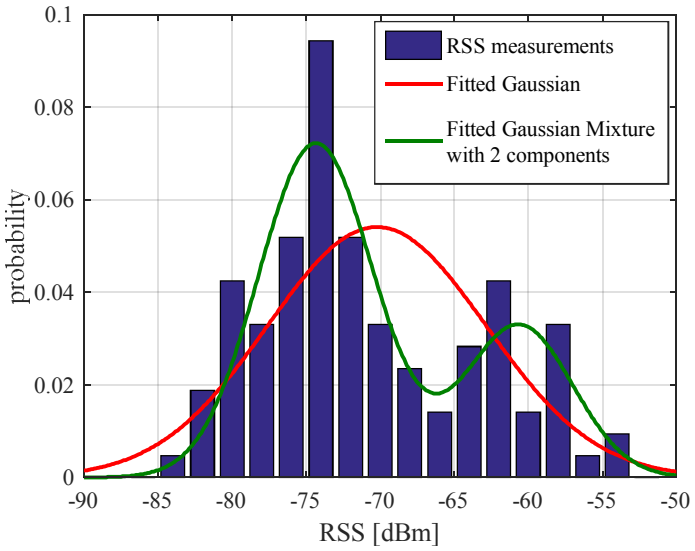

Fig. 2-3 *An example of RSS measurement histogram (the same found in the bottom-right corner in Fig. 2-2) and the corresponding fitted single Gaussian distribution and the fitted Gaussian mixture distribution with two Gaussian components.*

3 Path Loss Models for RSS-based Localization

As the radio wave propagates in the transmission medium, it is always attenuated due to the travelled distance and encountered obstacles in the propagation path. In addition, the radio propagation medium introduces multiple other phenomena which affect the RSS values, such as reflection, absorption, diffraction, and refraction of the radio waves. The dependency between the RSS measurements and the radio propagation distance is modeled with PL models [7],[8],[45],[103],[112],[122],[130]. The PL models can be classified into different categories such as physical models, deterministic models, empirical models, and hybrid models. Probably the most famous model based on a physical law is the Free-Space Path Loss (FSPL) based on the general inverse-square law. However, the FSPL provides often far too optimistic propagation losses, for which reason empirical and deterministic models have been introduced [112]. Empirical models are based on massive measurement campaigns, and typically include several tuning parameters for the models in order to cover several types of propagation environments. Instead, deterministic models, such as ray tracing [112] and dominant path modeling [139], aim to create an accurate model of the propagation environment including all the obstacles and their electromagnetic properties.

Because the PL models and the observed RSS values together are able to provide a distance estimate between the TX and the user device, PL models are extremely useful for localization purposes. Nevertheless, the selection of feasible PL model for the localization purposes is limited by the fact that the parameters of the propagation environment are not typically known beforehand. For example, in the well-known empirical-based Okumura-Hata model [112],[122], it is required to include parameters concerning the type of the propagation environment (rural, suburban or urban (small or large city)), the used carrier frequency, and the antenna heights of the TX and the user device. Therefore, the deterministic and empirical models are not directly feasible for the RSS-

based localization, because information about antenna heights and sometimes also the user carrier frequencies are not easily accessible or not available at all.

When considering the learning-data-based localization systems, PL models are particularly advantageous for reducing the size of the learning database. This is because PL models can compactly predict the RSS values in any coordinates by using only a few model parameters per each TX. Due to PL modeling inaccuracies, the localization accuracy is often somewhat reduced compared to the conventional fingerprinting approach. The comparison between the localization accuracy and the database size considering the PL-model-based approach and the fingerprinting-based approach is further conducted in Section 4.4.

PL models are also in vital role when simulating communications networks for localization-based studies as done in [P5] and [115]. Although only the real-life network measurements can truly validate the performance of the studied localization algorithms, the simulated network models are able to verify the algorithm consistency for various use cases and reveal severe algorithm issues which have not been found with the real-life measurements. Therefore, with appropriate simulation models, it is possible to diminish the workload introduced by the real-life measurement campaigns.

3.1 Feasible PL models for localization purposes

Since in the localization systems the radio propagation environment parameters are not typically known beforehand, the number of PL model parameters must be minimized. One of the most used and practical model is the log-distance PL model [19],[67],[98],[112],[119], whose different variants can be used in both outdoor and indoor environments. A common requirement for all the considered PL models is the availability of the TX locations. Of course, these TX locations are not usually known in advance, and therefore, the TX location estimation is part of the actual PL model estimation problem and it is further considered in Section 3.3.1.

The PL models are always considered as TX-wise, and hence, it is convenient to also reorganize the RSS measurements in the database to be TX-wise. Thus, we denote the set of measurements including the coordinates and the corresponding RSS values for each observed TX as

$$\left\{ x_j^{(r)}, y_j^{(r)}, z_j^{(r)}, P_j^{(r)} : j = 0,..., N_{FP}^{(r)} \right\}, \tag{3.1.1}$$

where $x_j^{(r)}$, $y_j^{(r)}$ and $z_j^{(r)}$ are the j^{th} x-coordinate, y-coordinate and z-coordinate of the r^{th} TX and $N_{FP}^{(r)}$ is the number of measurements in the r^{th} TX. These TX-wise coordinates and RSS values are

extracted from the complete fingerprint data set given in (2.4.4) so that only the values with the correct TX index *r* are considered, and hence, $x_j^{(r)} \in \{x_i\}$, $y_j^{(r)} \in \{y_i\}$, $z_j^{(r)} \in \{z_i\}$ and $P_j^{(r)} \in \{P_{i,r}\}$ for $i = 0,1,...,N_{FP}-1$.

3.1.1 Log-distance path loss model

One of the most used PL model in the RSS-based localization systems is the log-distance PL model, which contains only two unknown PL parameters in its simplest format. Now, with the log-distance PL model the observed RSS at distance $d_j^{(r)}$ can be presented as

$$P_j^{(r)} = A^{(r)} - 10n^{(r)} \log_{10}\left(d_j^{(r)}\right) + W_j^{(r)}, \tag{3.1.2}$$

where $P_j^{(r)}$, $d_j^{(r)}$ and $W_j^{(r)}$ are the RSS value, the measurement distance and the noise term of the j^{th} measurement of the r^{th} TX, respectively. Besides the actual measurement noise, $W_j^{(r)}$ includes shadowing and other effects of random RSS fluctuations within the grid point. Furthermore, $A^{(r)}$ and $n^{(r)}$ are the path loss constant and the path loss exponent of the r^{th} TX. The measurement distance $d_j^{(r)}$ is the distance between the estimated TX location and the location of the j^{th} measurement defined as

$$d_j^{(r)} = \sqrt{\left(\hat{x}_{TX}^{(r)} - x_j^{(r)}\right)^2 + \left(\hat{y}_{TX}^{(r)} - y_j^{(r)}\right)^2 + \left(\hat{z}_{TX}^{(r)} - z_j^{(r)}\right)^2} \tag{3.1.3}$$

where $\hat{x}_{TX}^{(r)}$, $\hat{y}_{TX}^{(r)}$ and $\hat{z}_{TX}^{(r)}$ are the estimates of the x-coordinate, the y-coordinate and the z-coordinate of the position of the r^{th} TX.

The path loss constant $A^{(r)}$ represents the RSS value at the 1m reference distance from the TX, but in general the reference distance is not restricted to this specific value. By introducing an additional reference distance d_{REF} inside the logarithm $\log_{10}\left(d_j^{(r)}/d_{REF}\right)$ given in (3.1.2), the path loss constant $A^{(r)}$ becomes the RSS value at the d_{REF} distance. However, throughout this thesis we define d_{REF}=1m, and thus, consider the $A^{(r)}$ as the RSS value at the 1m distance.

In addition, compared to various empirical PL models, the $A^{(r)}$ aggregates all the additive distance-independent terms into a single parameter. For example, considering the Okumura-Hata model, the $A^{(r)}$ can be conceived as the summation of terms depending on the propagation environment, which are the used carrier frequency and the antenna heights of the TX and the user device.

Moreover, since we define the PL model to represent the observed RSS value, not the actual path loss value, the parameter $A^{(r)}$ embodies also the transmission power of the TX.

The path loss exponent $n^{(r)}$ describes the steepness of the RSS attenuation as a function of distance. The attenuation rate of the RSS can be determined as $10n^{(r)}$ dB per a decade of the distance. In a free space environment the path loss exponent is physically defined as $n^{(r)}$=2, but the value is typically increased as obstacles appear in the radio path. However, in certain propagation environments, such as in narrow corridors, it is not uncommon to observe smaller values than the free space exponent (i.e., $n^{(r)}$<2) [113]. Furthermore, as seen later in Section 3.4, the correlation between $A^{(r)}$ and $n^{(r)}$ can also affect the observed path loss exponent values. Consequently, if the parameter $A^{(r)}$ gets a very high value, the overall PL model is typically compensated with a relatively small value of $n^{(r)}$.

Since the PL models in general are functions of the propagation distance, the modeling can be naturally adopted in both 2D and 3D environments. In this thesis we consider the 3D model to include the 2D horizontal coordinates and a discretized vertical coordinate for each floor. In other words, the location of the user device and the TXs are always restricted in the known floor levels, and thus, cannot float in between the floors. In case of the 2D PL models, the z-coordinate will simply be neglected. However, due to asymmetric properties of the radio propagation parameters in the vertical and horizontal direction inside buildings, the indoor 3D models can be inaccurate if floor losses are not included in the modeling process. Floor losses in 3D models are discussed in Section 3.1.3.

3.1.2 Multi-slope path loss models

Occasionally, the single slope log-distance PL model cannot describe the RSS behavior with required accuracy. Particularly in the indoor environment due to walls and other obstacles, the radio propagation parameters might change as a function of distance. To incorporate this into the propagation modeling process, multi-slope PL models have been proposed in [P4] and in the literature in [17],[30],[46],[75],[150]. One well-known indoor multi-slope PL model is the Ericsson model [17],[112],[122], which is essentially a multi-slope log-distance model with empirically tuned PL parameter values. Other PL models for the indoor environment have been proposed in [33] and [65].

By diving the fundamental log-distance model, given in (3.1.2), into two separate PL slopes we end up with a dual-slope PL model, where the j^{th} RSS value of the r^{th} TX can be given as

$$P_j^{(r)} = \begin{cases} A^{(r)} - 10n_0^{(r)} \log_{10}\left(d_j^{(r)}\right) + W_j^{(r)}, & d_j^{(r)} \le d_{BP,0} \\ A^{(r)} - 10n_0^{(r)} \log_{10}\left(d_{BP,0}\right) - 10n_1^{(r)} \log_{10}\left(\dfrac{d_j^{(r)}}{d_{BP,0}}\right) + W_j^{(r)}, & d_j^{(r)} > d_{BP,0} \end{cases} \quad (3.1.4)$$

where $n_0^{(r)}$ and $n_1^{(r)}$ are the path loss exponents of the r^{th} TX for the first slope and the second slope, and $d_{BP,0}$ is a breakpoint distance, defining the exact distance where the slope exponents change. We consider the breakpoint distance $d_{BP,0}$ to be maintained constant throughout the TXs, but it is also possible to optimize its value as TX-wise. However, this would directly mean an additional PL parameter for each TX to be stored in the database.

For considering PL models with more than two slopes, it is convenient to switch to matrix notations both in the PL model presentation and further in the PL model estimation discussed in Section 3.3. Firstly, the TX-wise RSS measurements are included into a single column vector as $\mathbf{P}^{(r)} = \begin{bmatrix} P_0^{(r)} & P_1^{(r)} & \cdots & P_{N_{FP}^{(r)}-1}^{(r)} \end{bmatrix}^T$, and secondly the PL parameters are grouped into one parameter vector defined as $\boldsymbol{\theta}_{multis}^{(r)} = \begin{bmatrix} A^{(r)} & n_0^{(r)} & n_1^{(r)} & \cdots & n_{N_{slopes}-1}^{(r)} \end{bmatrix}^T$. With multi-slope PL models it is often desired that the overall PL function is continuous at the breakpoint distances. Fundamentally, this means that the next slope always begins from the same RSS value where the previous slope ended. Now, by taking this into account, after a few arithmetic operations, the multi-slope log-distance PL model including N_{slopes} slopes in total for the r^{th} TX can be written as

$$\mathbf{P}^{(r)} = \mathbf{H}_{multis}^{(r)} \boldsymbol{\theta}_{multis}^{(r)} + \mathbf{W}^{(r)}, \quad (3.1.5)$$

where $\mathbf{W}^{(r)}$ is a noise vector for the r^{th} TX and $\mathbf{H}_{multis}^{(r)}$ is the system matrix given as

$$\mathbf{H}_{multis}^{(r)} = \begin{bmatrix} \mathbf{h}_0^{(r)} \\ \mathbf{h}_1^{(r)} \\ \vdots \\ \mathbf{h}_{N_{FP}^{(r)}-1}^{(r)} \end{bmatrix}, \quad (3.1.6)$$

in which the j^{th} row is further defined as

$$\mathbf{h}_j^{(r)} = 10 \begin{bmatrix} \frac{1}{10} \\ \mathrm{I}_0\left(d_j^{(r)}\right)\log_{10}\left(\frac{d_j^{(r)}}{d_{BP,0}}\right) - \log_{10}\left(d_j^{(r)}\right) \\ \mathrm{I}_1\left(d_j^{(r)}\right)\log_{10}\left(\frac{d_j^{(r)}}{d_{BP,1}}\right) - \mathrm{I}_0\left(d_j^{(r)}\right)\log_{10}\left(\frac{d_j^{(r)}}{d_{BP,0}}\right) \\ \vdots \\ \mathrm{I}_{N_{slopes}-1}\left(d_j^{(r)}\right)\log_{10}\left(\frac{d_j^{(r)}}{d_{BP,N_{slopes}-1}}\right) - \mathrm{I}_{N_{slopes}-2}\left(d_j^{(r)}\right)\log_{10}\left(\frac{d_j^{(r)}}{d_{BP,N_{slopes}-2}}\right) \\ -\mathrm{I}_{N_{slopes}-1}\left(d_j^{(r)}\right)\log_{10}\left(\frac{d_j^{(r)}}{d_{BP,N_{slopes}-1}}\right) \end{bmatrix}^T . \quad (3.1.7)$$

Moreover, $d_{BP,m}$ is the breakpoint distance (i.e., a boundary between two slopes) between the m^{th} and the $(m+1)^{\text{th}}$ slope and $\mathrm{I}_m\left(d_j^{(r)}\right)$ is an indicator function described as

$$\mathrm{I}_m\left(d_j^{(r)}\right) = \begin{cases} 1, & \text{if } d_j^{(r)} > d_{BP,m} \\ 0, & \text{otherwise} \end{cases} . \quad (3.1.8)$$

Here, the indicator function is used to define whether the distance $d_j^{(r)}$ of the j^{th} measurement of the r^{th} TX is larger than the breakpoint distance $d_{BP,m}$.

3.1.3 Indoor log-distance model featuring floor losses and frequency-dependency

One of the main differences between the radio wave propagation in outdoor and indoor environments is the floor loss in multi-storey buildings. The floor losses introduce asymmetric PL models, in which the radio signal attenuation differs in horizontal and vertical direction. Typically the floor losses are much larger than wall losses, and therefore, RSS measurements which have the equal 3D distances to the TX, but are located in separate floors, might differ significantly.

Besides the floor losses, also the carrier frequency can be included in the PL model. In most cases the used carrier frequency is available directly or indirectly (e.g., a channel identification number etc.) in the API of the measurement device. Of course, the usage of frequency-dependent PL models is not only restricted to indoors, but they can be sometimes very useful for outdoor models as well.

Motivated by to the above discussion, in [P7] we have also considered the single-slope log-distance-based PL model including floor losses and frequency-dependency as

$$P_j^{(r)} = A^{(r)} - 10n^{(r)} \log_{10}\left(d_j^{(r)}\right) - 10n^{(r)} \log_{10}\left(f_{carr}^{(r)}\right) - F_L^{(r)} N_{floors,j}^{(r)} + W_j^{(r)} \quad (3.1.9)$$

where $f_{carr}^{(r)}$ and $F_L^{(r)}$ are the carrier frequency and the floor loss of the r^{th} TX, and $N_{floors,j}^{(r)}$ is the number of floors between the j^{th} measurement and the TX location. Here it is assumed that for each TX the floor loss is constant for all floors. Furthermore, the floor loss parameters $F_L^{(r)}$ are assumed to be unknown and they are part of the PL parameter estimation problem discussed later in Section 3.3.

3.2 Shadowing analysis and simulation of RSS measurements

The noise variable $W_j^{(r)}$ in the PL models includes the effect of shadowing, RSS fluctuations within the measurement location and other type of measurement errors. The term shadowing has been used in variety of different contexts, but we refer by it to the random fluctuations of the RSS values around the PL model. If we assume noise free measurements and a perfect RSS calibration, which captures the true average RSS values for each location and TX, only shadowing is left from the PL noise parameter $W_j^{(r)}$. As the PL models describe a linear dependency between the logarithmic propagation distance and the observed RSS value, physically the shadowing values present the local variations of the RSS values due to heterogeneous radio propagation environment. It should be noticed that without any changes in the propagation environment, the pure shadowing value is fixed per each coordinate location and TX.

A large variety of different shadowing distributions has been proposed in the literature, such as the log-normal distribution, Nakagami-lognormal distribution, Weibull-lognormal distribution, Gamma distribution, and Weibull-gamma distribution [25],[31],[114]. In this context, it is quite common to refer to linear scale RSS values with the distributions. Consequently, with linear scale RSS values the lognormal distribution results in a Gaussian distribution with logarithmic scale RSS values.

Since the shadowing is caused by fixed obstacles in the radio propagation path, the effect of each obstacle on the RSS values is comparable in neighboring coordinates. This distance dependent correlation can be modeled with an environment specific autocorrelation function, as discussed in [12],[54],[58],[105],[107],[127],[147],[148]. The actual shape of the autocorrelation function de-

pends on the radio propagation environment and based on the Gudmundson model [56], the autocorrelation function for the shadowing can be modeled as

$$R_{shadow}(\Delta d) = e^{-\Delta d / D_{decor}}, \tag{3.2.1}$$

where Δd is the distance difference and D_{decor} is de-correlation distance, which depends on the radio propagation environment. In [P5] we have further studied the shadowing distributions and the shadowing autocorrelation function by using estimated PL models to remove the dependency of the propagation distance from the RSS measurements.

From the localization point of view, shadowing can be understood as both a negative and a positive phenomenon. In the PL-model-based localization, shadowing introduces local variations of the RSS values with respect to the average RSS value provided by the PL models, and thus, it reduces the modeling accuracy. On the other hand, in conventional fingerprinting-based approach, the local RSS variations due to the shadowing can sometimes be useful in the localization, since they induce differences between nearby fingerprints making their comparison feasible.

A proper understating and modeling of the shadowing phenomenon is also very important in computer-based simulations of the RSS measurements. The simulations can be very useful in validating localization systems and in finding problematic cases in the considered approach without the need of extensive measurement campaigns. In [P5] we have studied the simulation of RSS measurements for cellular-based localization system. Although the simulations of the RSS measurements considering a complete cellular network are quite complicated, modeling RSS measurements of a single TX is rather straightforward. Here we can directly utilize the pre-discussed PL models and the correlated shadowing model, given in (3.2.1). The basic steps for creating simulated RSS measurements for a certain grid coordinates from a single TX can be described as follows:

1. Create tentative RSS measurements based on the desired PL model and the TX location
2. Create uncorrelated shadowing values from the desired distribution (such as Gaussian)
3. Filter the uncorrelated shadowing values with a proper filter based on (3.2.1) in order to achieve a correlated shadowing values with the desired autocorrelation function
4. Add the correlated shadowing values to the tentative RSS measurements created in step 1 in order to get the final simulated RSS measurements.

In Fig. 3-1, the different phases of the process are illustrated together with a real-life BS measurement example. First, to get the simulated RSS values we have estimated the PL model parameters from a real-life BS using the log-distance PL model, as given further in (3.3.4). Then, based on the estimated PL model we have estimated the distribution and the correlation of the shadowing values

and generated random shadowing with similar statistics with the real-life RSS measurements. In the end, the correlated shadowing was added to the PL model to get the final simulated RSS model. Hence, when comparing the RSS plots of the simulated BS with the real-life BS, the main difference can be found from the different realizations of the shadowing fluctuations on top of the PL

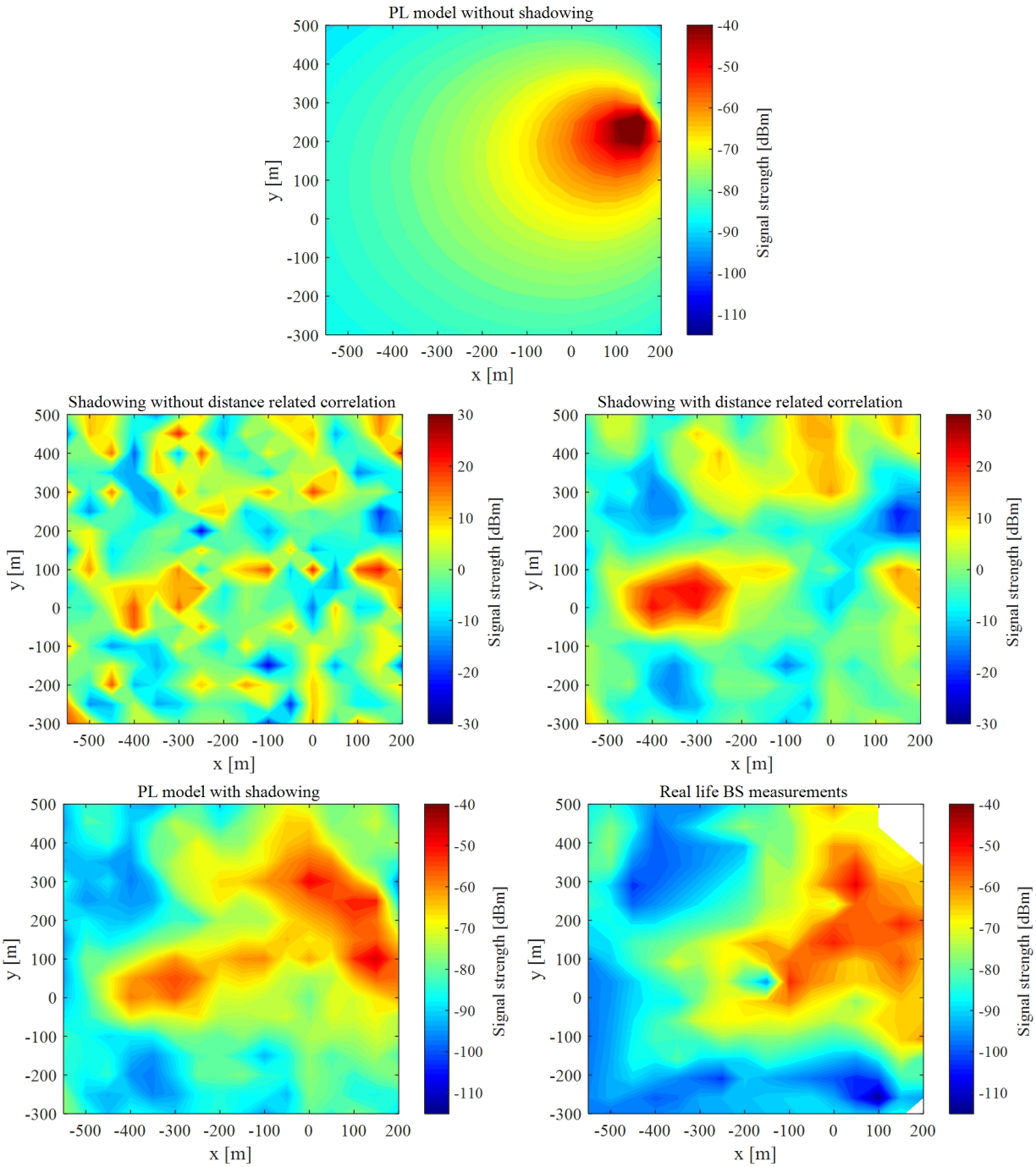

Fig. 3-1 *An example of the RSS Modeling process using the estimated PL model from real-life BS measurements: PL model (top), uncorrelated shadowing (middle-left), correlated shadowing (middle-right), the final RSS model as the sum of the PL model and correlated shadowing (bottom-left), and the real-life BS measurements (bottom-right).*

models, which are the same for both of the cases.

3.3 Estimation of the PL model parameters

3.3.1 Estimation of TX location

In order to use the PL models, the estimates of the TX locations are required to determine the distances between the measurements and TX location. Since we know that the radio signal begins to attenuate as soon as it departs from the transmission antenna, it would be logical to estimate the TX location based on the maximum observed RSS level. However, due to measurement errors, obstacles and different types of non-idealities in the radio path, using only the coordinates of the maximum heard RSS value as the TX location estimate, might result in large estimation errors. Nevertheless, since the TX is expected to be located around the area where most of the maximum RSS values are observed, the weighted centroid method has been proposed and used in [24], and in [P3],[P7],[P8]. Here the basic principle is to estimate the TX location as a weighted mean over the measurement coordinates as

$$\begin{aligned} \hat{x}_{TX}^{(r)} &= \sum_{j=0}^{N_{FP}^{(r)}-1} \upsilon_j^{(r)} x_j^{(r)} \\ \hat{y}_{TX}^{(r)} &= \sum_{j=0}^{N_{FP}^{(r)}-1} \upsilon_j^{(r)} y_j^{(r)}, \\ \hat{z}_{TX}^{(r)} &= \sum_{j=0}^{N_{FP}^{(r)}-1} \upsilon_j^{(r)} z_j^{(r)} \end{aligned} \tag{3.3.1}$$

where the coordinate weights for the j^{th} measurement and the r^{th} TX is given as

$$\upsilon_j^{(r)} = \frac{10^{\frac{P_j^{(r)}}{10}}}{\sum_{k=0}^{N_{FP}^{(r)}-1} 10^{\frac{P_k^{(r)}}{10}}}. \tag{3.3.2}$$

Hence, the weights are simply the linear scale RSS values of each measurement point. With such linear scale approach the weights for the large RSS values are relatively higher compared the low RSS values.

Another approach for the TX location estimation is the deconvolution-based approach introduced in [P4], where the AP location estimates are obtained jointly with the PL model parameter estimates. This method is discussed later in Section 3.3.4 in more detail.

3.3.2 Parameter estimation of the log-distance PL models

In the PL parameter estimation process it is convenient to use the matrix notations. Thus, using the same approach as shown in (3.1.5), the RSS value vector for the single-slope log-distance model can be modeled as $\mathbf{P}^{(r)} = \mathbf{H}_{logdist}^{(r)}\boldsymbol{\theta}_{logdist}^{(r)} + \mathbf{W}^{(r)}$, where the PL parameter vector is given as $\boldsymbol{\theta}_{logdist}^{(r)} = \begin{bmatrix} A^{(r)} & n^{(r)} \end{bmatrix}^T$ and the system matrix $\mathbf{H}_{logdist}^{(r)}$ is described as

$$\mathbf{H}_{logdist}^{(r)} = \begin{bmatrix} 1 & -10\log_{10}(d_0^{(r)}) \\ 1 & -10\log_{10}(d_1^{(r)}) \\ \vdots & \vdots \\ 1 & -10\log_{10}(d_{N_{FP}^{(r)}-1}^{(r)}) \end{bmatrix}, \tag{3.3.3}$$

Now, the above estimation problem can be solved by using well-known linear regression methods. Depending on the availability and usage of information regarding the statistics of the estimated parameters and the noise term, there are numerous approaches for obtaining the parameter estimates as shown in [26] and [111]. Another excellent source for the analysis and estimation algorithms regarding the linear regression is presented by Kay in [73].

Assuming no prior knowledge of the estimation problem, a straightforward approach is the famous LS solution, where the estimated parameter vector for the r^{th} TX is given as

$$\hat{\boldsymbol{\theta}}_{method,LS}^{(r)} = \left(\mathbf{H}_{method}^{(r)T}\mathbf{H}_{method}^{(r)}\right)^{-1}\mathbf{H}_{method}^{(r)T}\mathbf{P}^{(r)}, \tag{3.3.4}$$

where $\hat{\boldsymbol{\theta}}_{method,LS}^{(r)}$ and $\mathbf{H}_{method}^{(r)}$ are the parameter vector estimate and the system matrix of the r^{th} TX for either the single-slope ($\hat{\boldsymbol{\theta}}_{logdist,LS}^{(r)}$ and $\mathbf{H}_{logdist}^{(r)}$) PL model or the multi-slope ($\hat{\boldsymbol{\theta}}_{multis,LS}^{(r)}$ and $\mathbf{H}_{multis}^{(r)}$) PL model. If the noise term $\mathbf{W}^{(r)}$ is white, the LS approach can be considered as Best Linear Unbiased Estimator (BLUE), and moreover, if additionally the noise is Gaussian, the LS approach is actually the Minimum Variance Unbiased estimator (MVU) [73].

If the quality of the measurements in $\mathbf{P}^{(r)}$ are considered to be varying, one option is to utilize the Weighted Least squares (WLS) approach, where the estimated parameter vector for the r^{th} TX is given as

$$\hat{\boldsymbol{\theta}}^{(r)}_{method,WLS} = \left(\mathbf{H}^{(r)^T}_{method}\mathbf{Q}^{(r)}\mathbf{H}^{(r)}_{method}\right)^{-1}\mathbf{H}^{(r)^T}_{method}\mathbf{Q}^{(r)}\mathbf{P}^{(r)} \tag{3.3.5}$$

where $\mathbf{Q}^{(r)}$ is a specific diagonal weighting matrix for the r^{th} TX. Considering RSS-based localization systems, high RSS values are often considered as important from the localization accuracy point of view. Thus, in [P4] have we used the linear scale RSS values as the weights of the measurements as $\mathbf{Q}^{(r)} = \mathrm{diag}(10^{\frac{\mathbf{P}^{(r)}}{10}})$. However, with this weighting allocation the localization accuracy with the WLS-based PL model estimates was not generally improved compared to the LS-based approach.

Occasionally there might be a priori information regarding the statistics of the PL parameters and the noise term. In this case the linear system, given in (3.1.5), can be considered as a Bayesian linear system. In [30], the Bayesian approach was considered for multi-slope PL models including estimation of the breakpoint distance. Nevertheless, assuming that both the PL parameter vector $\boldsymbol{\theta}^{(r)}_{method} \sim \mathrm{N}(\boldsymbol{\mu}^{(r)}_{\theta}, \mathbf{C}^{(r)}_{\theta})$ and the noise term $\mathbf{W}^{(r)} \sim \mathrm{N}(\mathbf{0}, \mathbf{C}^{(r)}_{W})$ are Gaussian distributed and independent of each other, we can find the MMSE estimate [73] of the estimated parameter vector as

$$\hat{\boldsymbol{\theta}}^{(r)}_{method,MMSE} = \left(\mathbf{C}^{(r)^{-1}}_{\theta} + \mathbf{H}^{(r)^T}_{method}\mathbf{C}^{(r)^{-1}}_{W}\mathbf{H}^{(r)}_{method}\right)^{-1}\left(\mathbf{H}^{(r)^T}_{method}\mathbf{C}^{(r)^{-1}}_{W}\mathbf{P}^{(r)} + \mathbf{C}^{(r)^{-1}}_{\theta}\boldsymbol{\mu}^{(r)}_{\theta}\right), \tag{3.3.6}$$

where $\boldsymbol{\mu}^{(r)}_{\theta}$ and $\mathbf{C}^{(r)}_{\theta}$ are the mean vector and the covariance matrix of the prior distribution of the PL parameter vector, and $\mathbf{C}^{(r)}_{W}$ is the covariance matrix of the noise term. Under the given conditions, the estimate is the mean of the posterior function resulted from the Bayesian inference. In addition, it is also the MVU for the considered estimation problem. If the prior information is not available, then the inverse of the prior covariance matrix approaches to zero (i.e., $\mathbf{C}^{(r)^{-1}}_{\theta} \to 0$) and the estimate becomes as $\left(\mathbf{H}^{(r)^T}_{method}\mathbf{C}^{(r)^{-1}}_{W}\mathbf{H}^{(r)}_{method}\right)^{-1}\left(\mathbf{H}^{(r)^T}_{method}\mathbf{C}^{(r)^{-1}}_{W}\mathbf{P}^{(r)}\right)$, which resembles the WLS estimate, given earlier in (3.3.5). With white noise assumption $\mathbf{C}^{(r)^{-1}}_{W}$ becomes diagonal, and based on the variance of the noise components, it can be considered as the weighing matrix $\mathbf{Q}^{(r)}$ in the WLS. In Fig. 3-2, we have illustrated the PL model estimation using the single-slope and dual-slope log-distance PL models with measurements from a real-life cellular BS. As expected, the dual-slope PL model appears to give a better fit with the measurements than the single-slope PL model. However, with the dual-slope PL models it is obvious that the quality of the LS fitting depends greatly on the used breakpoint distance $d_{BP,0}$.

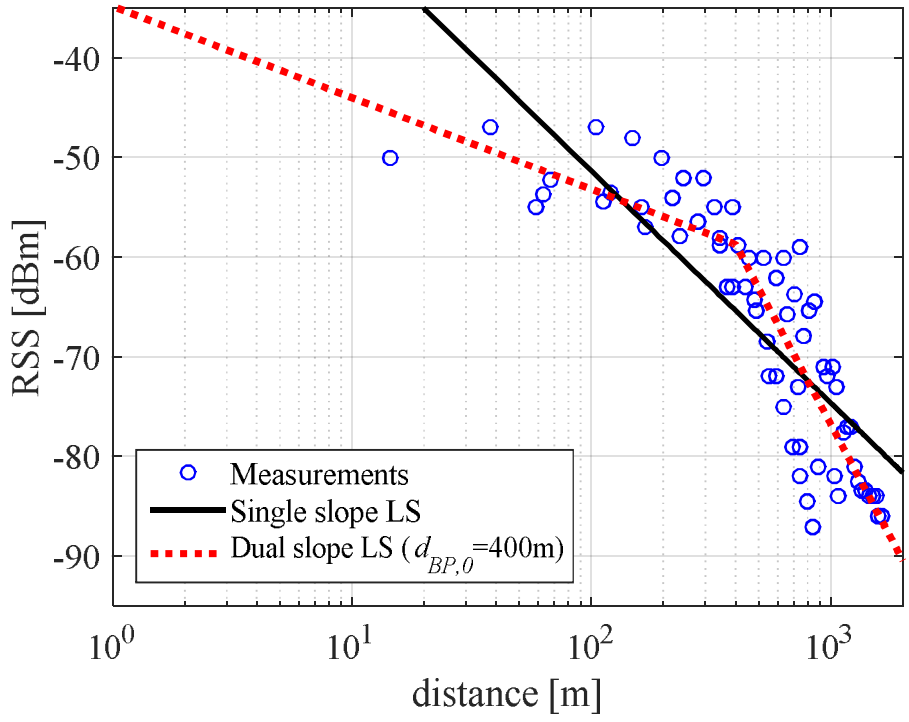

Fig. 3-2 *An example of the LS estimation of the PL model for measurements from a real-life cellular BS using the single-slope and the dual-slope log-distance PL models. The breakpoint distance of the dual-slope model is set to* $d_{BP,0}$ *=400m.*

3.3.3 Parameter estimation of the log-distance PL model with floor losses and frequency-dependency

Due to the differences between horizontal and vertical radio propagation in indoor environment, introducing floor losses in the estimation problem is beneficial. Here we have included also the frequency-dependency on the same PL model, but if preferred, it can be easily neglected from the equations. By using the matrix notation, the PL model given in (3.1.9) can be presented as

$$\mathbf{P}^{(r)} = \mathbf{H}_{floorL}^{(r)} \boldsymbol{\theta}_{floorL}^{(r)} + \mathbf{W}^{(r)}, \tag{3.3.7}$$

where the system matrix $\mathbf{H}_{floorL}^{(r)}$ is given as

$$\mathbf{H}_{floorL}^{(r)} = \begin{bmatrix} 1 & -10\log_{10}(d_0^{(r)} f_{carr}^{(r)}) & -N_{floors,0}^{(r)} \\ 1 & -10\log_{10}(d_1^{(r)} f_{carr}^{(r)}) & -N_{floors,1}^{(r)} \\ \vdots & \vdots & \vdots \\ 1 & -10\log_{10}(d_{N_{FP}^{(r)}-1}^{(r)} f_{carr}^{(r)}) & -N_{floors,N_{FP}^{(r)}-1}^{(r)} \end{bmatrix}, \tag{3.3.8}$$

And where the PL parameter vector is defined as $\boldsymbol{\theta}_{floorL}^{(r)} = \begin{bmatrix} A^{(r)} & n^{(r)} & F_L^{(r)} \end{bmatrix}^T$. Hence, besides the traditional log-distance PL model parameters $A^{(r)}$ and $n^{(r)}$, the floor loss parameter $F_L^{(r)}$ is also included in the estimated parameters. In addition, it is worth of noticing that the carrier frequency found from the second column of $\mathbf{H}_{floorL}^{(r)}$ could be introduced in similar way also with system matrix of the basic log-distance PL model given in (3.3.3).

Although the above system model could be solved using the LS, WLS and MMSE methods as discussed earlier, the physical interpretation of the floor loss parameter $F_L^{(r)}$ constrains its value to be larger than zero. In other words, the possible gain from the floor is desired to be avoided. This results in a Non-Negative Least Squares (NNLS) problem in which the estimated PL parameter vector can be found as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{floorL}^{(r)} &= \underset{\boldsymbol{\theta}_{floorL}^{(r)}}{\arg\min} \left\| \mathbf{H}_{floorL}^{(r)} \boldsymbol{\theta}_{floorL}^{(r)} - \mathbf{P}^{(r)} \right\|^2 \\ \text{subject to } & \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \boldsymbol{\theta}_{floorL}^{(r)} \geq \mathbf{0}. \end{aligned} \tag{3.3.9}$$

Here in fact, all the PL parameters are constrained as $A^{(r)} \leq 0$, $n^{(r)} \geq 0$ and $F_L^{(r)} \geq 0$. At least in most of the communications networks, these are very reasonable assumptions about the PL models. To solve the given PL parameter estimate $\hat{\boldsymbol{\theta}}_{floorL}^{(r)}$ an iterative algorithm described in [79] can be exploited. In many practical cases the PL parameter vector $\hat{\boldsymbol{\theta}}_{floorL}^{(r)}$ can also be solved by using the MMSE approach given in (3.3.6) and using appropriately chosen prior values $\boldsymbol{\mu}_{\theta}^{(r)}$ and $\mathbf{C}_{\theta}^{(r)}$ to avoid negative floor losses. However, since the prior distribution is assumed to be Gaussian, the process does not strictly limit the estimated vector and so negative floor losses might occasionally occur. In Fig. 3-3, an illustration of the 3D PL modeling is shown with the PL model with floor losses and without floor losses.

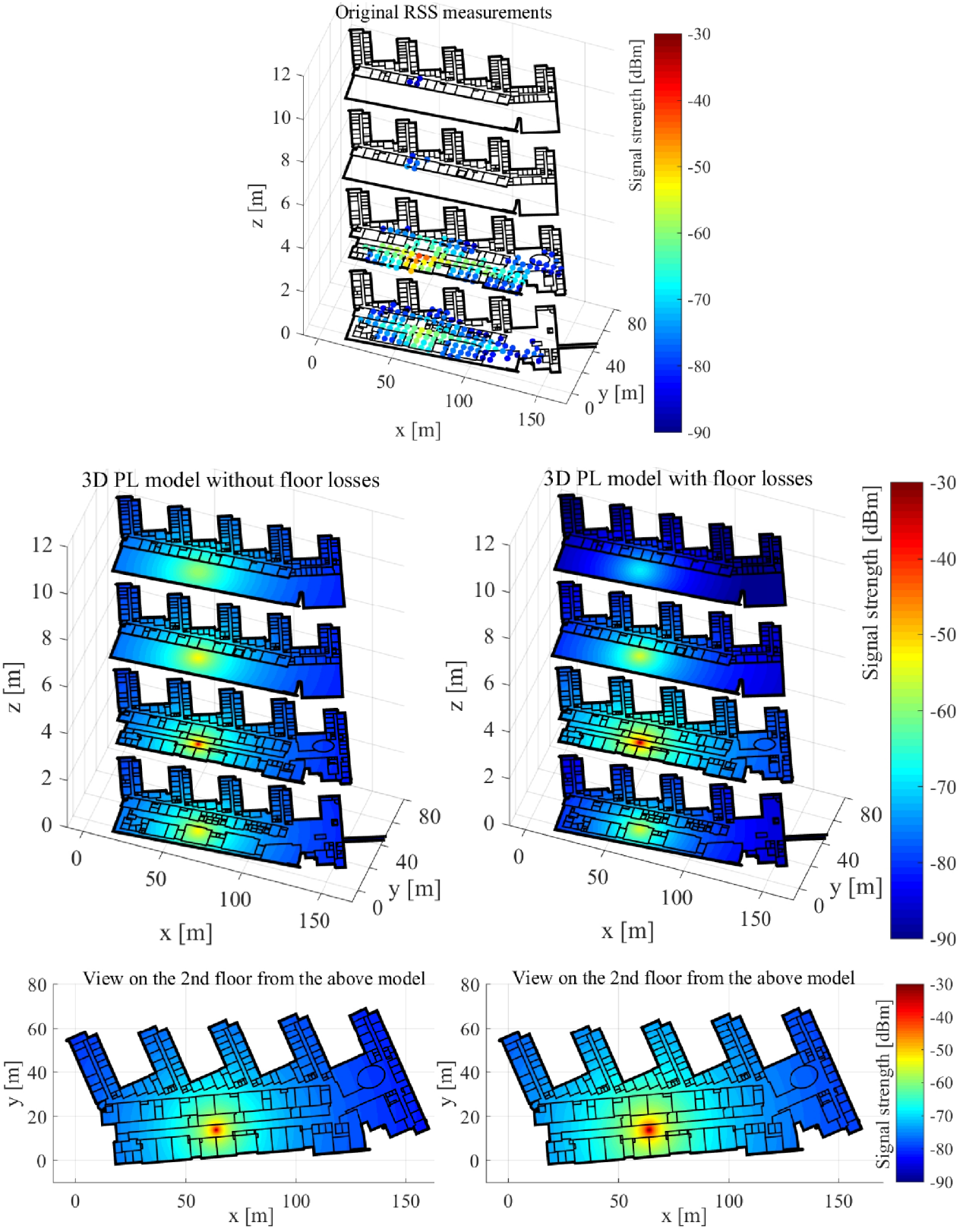

Fig. 3-3 *An example of the 3D PL model for indoor environment for one AP in 2.4GHz WLAN network: the original RSS measurements (top), the log-distance PL model without floor losses (left), and with floor losses (right), where the estimated floor loss was 6.8dB.*

3.3.4 Joint estimation of the PL parameters and the TX location by deconvolution

In all previously discussed PL model estimation approaches, the location the on TX was assumed to be known based on the TX location estimation by the weighted centroid given in (3.3.1). In [P4], we have studied a different approach for obtaining the estimates of the PL parameters and the TX location by deconvolution. Here the fundamental idea is to trial the assumed PL model over all the available grid coordinates and to assume the AP location to be in the grid coordinate at each trial. After this, the TX location can be estimated based on the residual errors of the PL model fitting in each grid point. The process for the r^{th} TX can be described as follows:

1. For each grid point in which the TX has been heard $\left\{x_j^{(r)}, y_j^{(r)}, z_j^{(r)} : j = 0, ..., N_{FP}^{(r)}\right\}$, set the AP location as the grid point coordinates $\left(\hat{x}_{TX}^{(r)} = x_j^{(r)}, \hat{y}_{TX}^{(r)} = y_j^{(r)}, \hat{z}_{TX}^{(r)} = z_j^{(r)}\right)$ and calculate the corresponding system matrices $\mathbf{H}_{method}^{(r)}$ for each trialed grid point.
2. Compute the PL parameter estimate vector $\hat{\boldsymbol{\theta}}_{method}^{(r)}$ for each trialed grid point by using the desired estimation method.
3. Compute the estimated RSS values based on the PL parameter estimates as $\hat{\mathbf{P}}^{(r)} = \mathbf{H}_{method}^{(r)} \hat{\boldsymbol{\theta}}_{method}^{(r)}$ for each trialed grid point.
4. Compute the norm of the error vector for each trialed grid point as $\left\|\hat{\mathbf{P}}^{(r)} - \mathbf{P}^{(r)}\right\|$ and define the coordinates of the TX as the coordinates of the grid point with the smallest error norm. Alternatively, estimate the coordinates of the TX as the average of the coordinates of the N_{high} grid points having the smallest error norms.
5. If $N_{high} > 1$ (i.e., average over multiple coordinates is taken), re-compute the PL parameter estimate $\hat{\boldsymbol{\theta}}_{method}^{(r)}$ for the estimated TX location by using the desired PL parameter estimation method.

The above deconvolution approach utilizes brute-force ideology when trialing the TX locations over the known grid coordinates and choosing the location where the PL model has the best fit with the observed measurements. An alternative approach for the joint estimation of the TX location and the PL parameters is to use advanced iterative methods, such as the non-linear LS, as studied in [100] and [101]. However, with these algorithms one must deal with several design parameters, such as the values of the initial parameter estimates, the iteration step size, and the convergence criterion, which are further discussed in [26] and [118]. In addition, the convergence of the algorithm to the global optimum is not generally guaranteed, but multiple runs with different parameter

initializations are often required. For example, the well-known Gauss-Newton for solving the non-linear LS problem can be very sensitive to poor parameter initializations [100].

3.4 Comparison of PL model parameter statistics for the considered localization systems

The results shown in this section are achieved using the single-slope log-distance PL model with the LS principle. Although other approaches, such as the WLS and MMSE, could benefit in some modeling scenarios, the LS is a well-justified approach in the general case, where no information from the modeled system is available. For example, the LS approach always finds a better fit between the PL model and the measurements compared to the MMSE method, since the LS is purely based on the measurement data. This is because the MMSE introduces a priori information, which drives the estimates into the direction determined by a priori distribution values. Of course, if the a priori information is correct, the estimate accuracy will increase. Nonetheless, in practice it is very difficult, even impossible, to define correct a priori distributions. Moreover, as seen later on, the PL parameter values might vary considerably between different communications systems and radio propagation environments.

We have studied the estimation of PL parameters in different communication systems in [P3]-[P8]. The average values of the estimated path loss constant $A^{(r)}$, path loss exponent $n^{(r)}$, and the shadowing standard deviation are shown for different communications systems in Table 2.

Table 2. The average estimated PL parameter values for the considered communications systems

	PL Parameter		
System	**Average of the PL constant $A^{(r)}$ [dBm]**	**Average of the PL exponent $n^{(r)}$ [-]**	**Average of the shadowing standard deviation [dB]**
WLAN 2.4 GHz	-52	1.5	5.9
WLAN 5 GHz	-45	1.6	5.3
BLE	-71	1.3	6.2
RFID	-7	2.2	2.8
Cellular GSM (Suburban)	4	2.9	6.2
Cellular WCDMA (Urban)	-18	2.6	10.8

Here the results for the WLANs, BLE and the RFID are acquired from one university building in Tampere, Finland, and the cellular network results are based on outdoor measurements taken also in Tampere. To maintain a reasonable comparability between different systems, for all cases the PL model parameters have been estimated using the single-slope log-distance model without taking floor losses into account. The differences between the average parameter values are fairly visible between different communication systems. This indicates that a system specific optimization of the PL modeling approach is advantageous for increasing the modeling accuracy.

Since the RSS-based localization is still mostly associated with the WLANs and cellular networks, we turn the focus of more detailed analysis on these systems. For the following studies we consid-

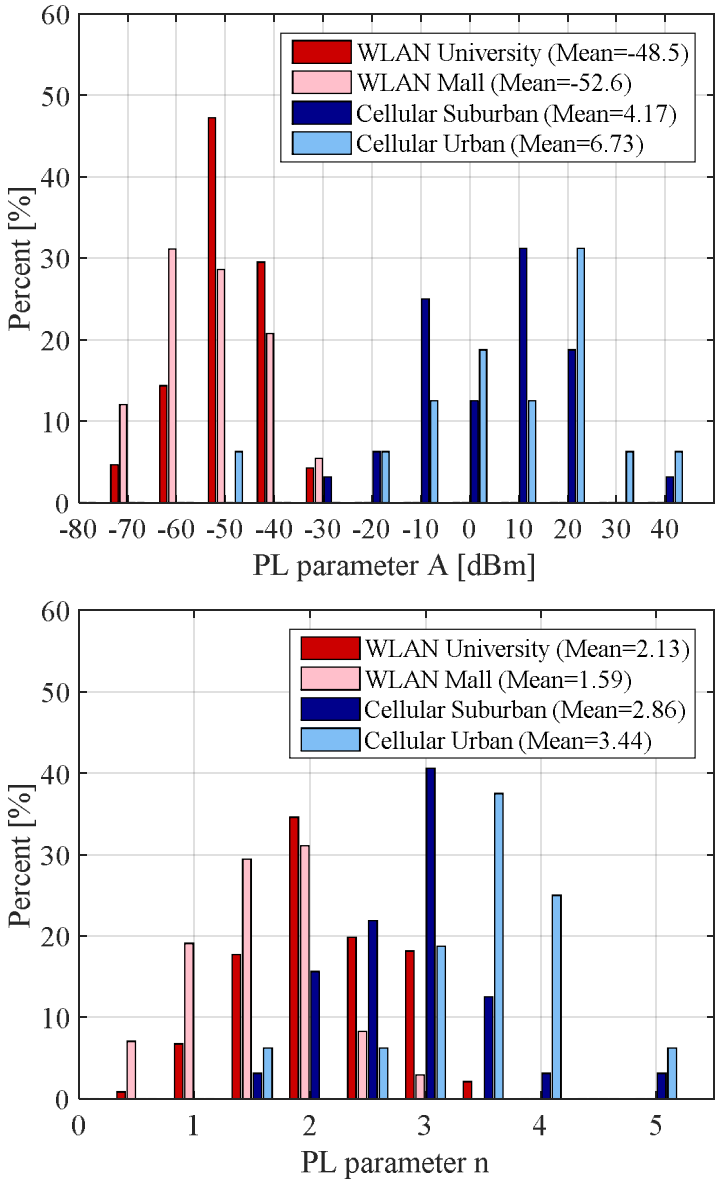

Fig. 3-4 *Histograms of the estimated PL parameter values for 2.4GHz WLAN networks in a University building and Mall building, and for cellular networks in suburban and urban environments.*

er the previously mentioned University building (*WLAN 2.4 GHz*) and the two cellular network cases (*Cellular GSM (Suburban)* and *Cellular WCDMA (Urban)*), and one shopping mall with 2.4 GHz WLAN from Berlin, Germany. To get a better impression of the distribution of the PL parameter estimates, in Fig. 3-4 we show the histograms of the estimated PL parameter values for the 4 different scenarios.

Here, to get reliable parameter estimates we have only considered TXs with more than 30 measurements in the histograms. It should be noticed that this is not beneficial from the localization accuracy point of view, since many TXs are dropped from the database due to this procedure. From Fig. 3-4 it can be seen that the PL parameters of WLANs can be clearly distinguished from the cellular network parameters, especially regarding the path loss constant $A^{(r)}$. Nevertheless, the differences between the two buildings as well as the differences between the two cellular cases are relatively small. However, the path loss exponent $n^{(r)}$ in the University building is generally larger compared to the shopping mall, since in the shopping mall there are more open spaces than in the University building, which reduces the value of the path loss exponent. A similar observation can be made also with the cellular networks between the urban and sub-urban case. Here, in the urban case, where there are more obstacles in the radio path compared to the sub-urban case, the val-

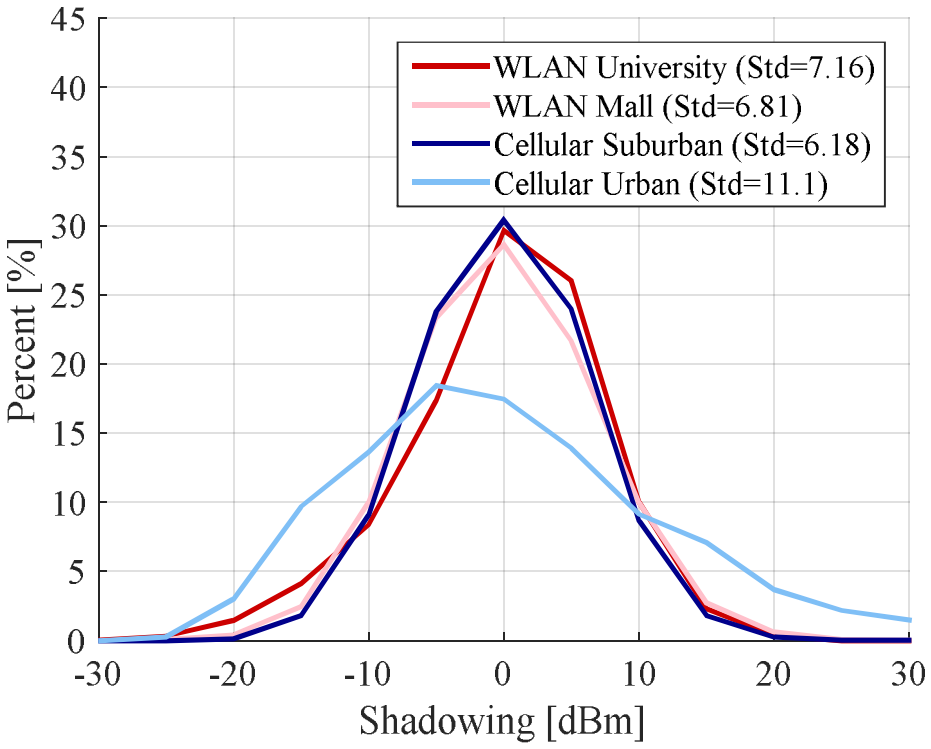

Fig. 3-5 *Distributions of the shadowing values for 2.4GHz WLAN networks in a University building and Mall building, and for cellular networks in suburban and urban environments.*

ues of the path loss exponents are generally larger than in the sub-urban case.

In Fig. 3-5, we illustrate the distribution of the shadowing values for the above scenarios. With the exception of the urban WCDMA network, all other scenarios share a rather similar shadowing distribution. The reason for the unique and slightly skewed distribution of the urban cellular case might be originated either from the propagation environment effects or from the WCDMA access method. Unlike in the GSM, the RSS indicator in the WCDMA system is the RSCP, which can be affected by the data traffic load of the system [5].

One important notice, which has not been often mentioned in the literature is a positive correlation between the path loss parameters $A^{(r)}$ and $n^{(r)}$. On average, whenever the parameter $A^{(r)}$ increases, also $n^{(r)}$ increases. Therefore, the previously discussed comparison of the PL model parameters between different communications systems should always be performed based on both of the parameters. As we look back at the Table 2, it is possible to see a pattern in the PL parameter pairs. For example, by sorting the table rows based on the values of the parameter $A^{(r)}$ from the smallest to the largest, also the parameter $n^{(r)}$ values become sorted apart from the WCDMA

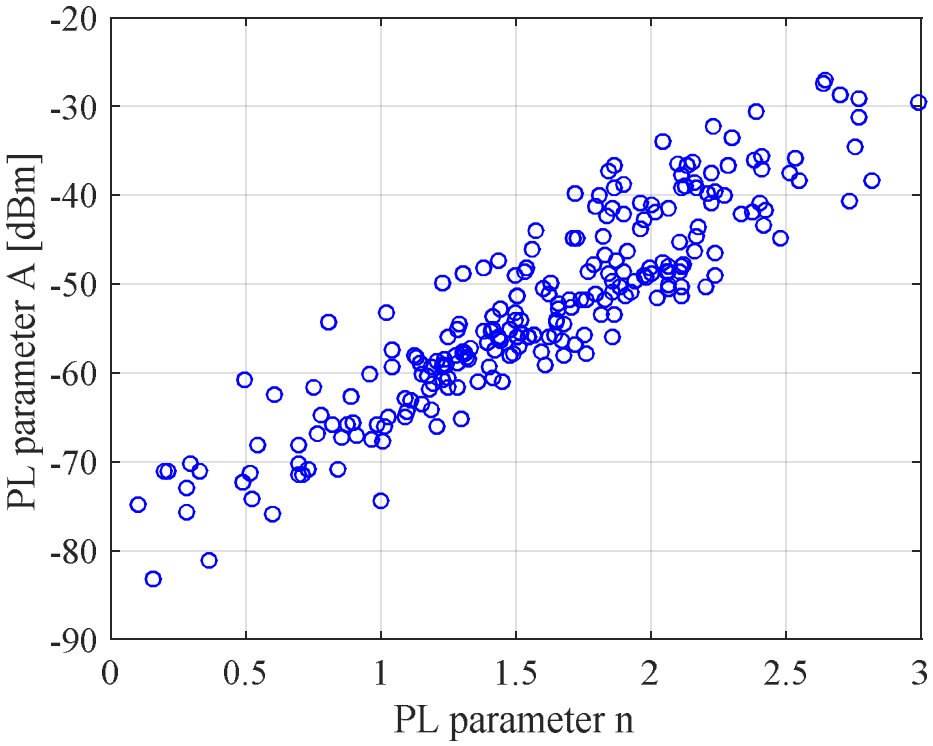

Fig. 3-6 *Illustration of the correlation between the PL parameters $A^{(r)}$ and $n^{(r)}$ for all the estimated 2.4GHz WLAN APs in the shopping mall in Berlin. Each circle in the plot describes the estimated $A^{(r)}$ and $n^{(r)}$ for one AP.*

case. The correlation of the PL parameters is further illustrated in Fig. 3-6, where the $A^{(r)}$ and $n^{(r)}$ are jointly given for all the APs heard in the Berlin shopping mall case.

Based on the Fig. 3-6, there is a clear linear dependency between the two parameter values and the Pearson product-moment correlation coefficient can be computed as high as 0.91. This information can be used, for example, in defining the prior covariance matrix $\mathbf{C}_{\theta}^{(r)}$ in the MMSE estimation approach, given in (3.3.6). These correlation characteristics are mostly a property of single-slope models. With dual-slope models the correlation between the parameter $A^{(r)}$ and the first slope exponent $n_0^{(r)}$ is still considerable, but with the second slope exponent $n_1^{(r)}$ the correlation with the $A^{(r)}$ is greatly decreased.

4 Localization Phase with User RSS Measurements

Regarding the localization phase, in which the user is localized based on the data collected in the learning phase, there are numerous different approaches for estimating the location discussed in the literature [16],[21],[76],[80],[81],[92],[94],[109],[117],[137],[138],[141],[147],[148]. Noteworthy surveys between different localization algorithms are found in [10],[49],[59] and [61].

One important study topic within the field of RSS-based localization is Bayesian-based location filtering and tracking methods [13],[14],[23],[35],[41],[63],[83],[89],[97],[125],[144]. Here, the fundamental idea is to estimate the user location recursively via the Markov process based on the assumptions on the user movement by means of the state-transition model and on the RSS measurements by means of the measurement (or observation) model. One of the most famous Bayesian-based tracking algorithms is the Kalman filter, studied in [13],[35],[83],[89], which provides the MMSE location estimates in the case of linear system model with Gaussian distributed random variables. Other Bayesian filtering methods without necessarily having the assumptions on the linearity or on the Gaussian distributed variables are discussed in [14],[23],[29],[41],[63],[97],[125], [144]. Although the filtering algorithms should always be included in the practical localization systems, we have left them out in our studies, since they can be considered as a completely new field of studies. Nonetheless, we fairly assume that the quality of the static estimates, provided in our studies, reflect the quality of the filtered location estimates.

In this Section we first present different considered localization algorithms for both the fingerprint and PL model based localization. Then, we analyze the effect of different types of errors, such as the database calibration errors and bias errors, on the localization performance. After this, the effect of the coverage gaps in the learning database is modeled in our simulator and analyzed. The degradation of the localization performance due to the coverage gaps is reduced by introducing different interpolation and extrapolation methods for the lost RSS values. We conclude the section

by comparing the localization performance in different communications systems by considering both fingerprinting and PL-model-based localization approaches.

4.1 Localization algorithms

4.1.1 Conventional fingerprinting

In the conventional fingerprinting the RSS measurements taken by the user are compared with the measurements available in the learning database. The user position estimate is determined based on coordinates of one or multiple fingerprints with the best match for the user measurements. Hence, eventually the fingerprinting-based localization is a classification problem, where the multi-dimensional user measurement is attempted to be categorized into the correct fingerprint. The fingerprinting algorithms can be roughly divided into deterministic approaches [21],[80],[109] and probabilistic approaches [16],[80],[81],[92],[94],[117]. In the deterministic approach the location of the user is assumed to be a non-random variable whereas in the probabilistic approach the location is assumed to be a random variable with a specific probability distribution.

Usually, the deterministic approaches are based on computing a certain predefined cost function in each grid point. Now, we define the set of RSS measurements taken by the user from $N_{TXheard}$ TXs as

$$\Omega_{RSS,USER} = \left\{ P_{USER,r} : r \in \Omega_{TXheard} \right\}, \tag{4.1.1}$$

where $P_{USER,r}$ is the user RSS measurement from the r^{th} TX and $\Omega_{TXheard} \subset \Omega_{TX}$ is a set of indices r of those TXs, which are heard in the current measurement, and thus, the number of elements in $\Omega_{TXheard}$ is given as $\left|\Omega_{TXheard}\right| = N_{TXheard}$. A widely used cost function is based on the squared Euclidian distance $\|\cdot\|^2$ between the user measurements and the database fingerprints as

$$\psi_i = \sum_{r \in \Omega_{TXheard}} \left(P_{USER,r} - P_{i,r} \right)^2, \tag{4.1.2}$$

where ψ_i is the value of the cost function at the i^{th} fingerprint and $P_{i,r}$ is the RSS value of the r^{th} TX stored in the i^{th} fingerprint. Besides the Euclidian norm, other norms such as the p-norm and the maximum norm might also be suitable. Moreover, if information on the standard deviation of the RSS distribution in each fingerprint is available, also the Mahalanobis-norm can be exploited [61].

In case that one or more TXs are not found in the studied fingerprint, a small heuristic bogus value can be used instead. Typically the level of the bogus value should be equal or less than the smallest observed RSS value in the database [117].

For the sake of notational convenience we denote the fingerprint location coordinates and the user location coordinates by using vectors as

$$\begin{aligned} \mathbf{x}_i &= \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T \\ \mathbf{x}_{USER} &= \begin{bmatrix} x_{USER} & y_{USER} & z_{USER} \end{bmatrix}^T \end{aligned}, \tag{4.1.3}$$

where $\mathbf{x}_i$ is the coordinate vector of the i^{th} fingerprint and $\mathbf{x}_{USER}$ is the coordinate vector of the user locations in which x_{USER}, y_{USER} and z_{USER} are the x-coordinate, y-coordinate and the z-coordinate of the user location. The most straightforward method to estimate the user location based on the cost-function-based approach is the Nearest Neighbor (NN) method [20],[80], where the estimated user location is determined at the fingerprint where the cost function is minimized:

$$\hat{\mathbf{x}}_{USER,NN} = \mathbf{x}_k, \text{ where } k = \arg\min_i \psi_i . \tag{4.1.4}$$

Another popular method is the K-Nearest Neighbor (KNN) method [20],[38],[57],[80],[108],[134], where the user location is obtained by taking an average of the K_{NN} fingerprint coordinates with the smallest cost function as

$$\hat{\mathbf{x}}_{USER,KNN} = \frac{1}{K_{NN}} \sum_{k \in \Omega_{KNN}} \mathbf{x}_k , \tag{4.1.5}$$

where Ω_{KNN} is the set of fingerprint indices *i* pointing out the $|\Omega_{KNN}| = K_{NN}$ smallest values of the cost function ψ_i.

The cost-function can also be used directly as a weighting function, which can be used to indicate the relative categorization precision between the user measurements and each fingerprint. Thus, by exploiting the KNN principle, the user location estimate can also be calculated by using a weighted average as

$$\hat{\mathbf{x}}_{USER,WKNN} = \sum_{k \in \Omega_{KNN}} \frac{\psi_k^{-1}}{\sum_{l \in \Omega_{KNN}} \psi_l^{-1}} \mathbf{x}_k , \tag{4.1.6}$$

in which case the approach is referred as the Weighted K-Nearest Neighbor (WKNN) [20],[80]. Here the inverse of the cost function is required, since the weight value should always increase as the cost-function decreases.

It is shown in [61], that the nearest-neighbor-based methods give relatively good performance compared to more advanced localization algorithms, such as the probabilistic Bayesian approach discussed later on. With a tight fingerprint grid interval, the NN is able to provide roughly the same localization accuracy as the KNN and WKNN. However, as the grid interval is increased, KNN and WKNN can improve their accuracy compared to the NN, since they are able to estimate the user in the middle of the fingerprint coordinates, whereas the estimates of the NN are always found exactly at the fingerprint locations. Conversely, a general problem in the KNN method is that although there would be a perfect match with the user measurement and the correct fingerprint, the estimate is always biased due to the effect of other fingerprints, with possibly much worse match with the location estimate. This issue can be somewhat alleviated with the WKNN assuming that the weights are scaled properly.

In addition to the above discussed localization algorithms, also other interesting approaches have been proposed in the literature. One of these is the rank-based approach, studied in [86],[87],[88],[91],[146], where the RSS measurements are not used in the traditional manner for calculating the cost function or the likelihood. In the rank-based method the basic idea is to organize the RSS measurements from different TXs into a descending order based on the RSS values. The resulted ordered TX index vector is called as the ranking vector. In the user localization phase, instead of using the actual RSS values, ranking vector of the heard TXs is compared with ranking vector in each fingerprint. Hence, the cost function is based on similarity measures, such as the Spearman distance, Spearman's footrule, Jaccard coefficient, Hamming distance, or Canberra distance, as discussed in [86] in more detail. One great advantage of the rank-based method is its tolerance against biased RSS measurement errors. For example, if one device is used to collect the learning data and another device to perform the localization, in rank-based approach the mutual RSS bias between the RSS measurements of separate devices does not affect the localization performance as long as the order of the RSS values remains the same [86].

A probabilistic localization approach using Bayesian methods can be very advantageous in fingerprinting. Firstly, the probabilistic approach offers widely-known tools for describing the theoretical framework of the estimation problem, and thus, it provides an access to several useful concepts in the estimation process, such as the evaluation of the localization error quantiles. Secondly, the Bayesian methods act as the backbone of Bayesian filters, such as the famous Kalman filter [13], which are used for efficient tracking of the user location. In the Bayesian framework, one of the

most important concepts is the likelihood function, which binds the user RSS measurements and the system coordinates together. Assuming independent measurements from each TX, the probability of observing the RSS measurements $\Omega_{RSS,USER}$ at the fingerprint coordinates $\mathbf{x}_i$ can be written as

$$p(\Omega_{RSS,USER} \mid \mathbf{x}_i) = \prod_{r \in \Omega_{TXheard}} p_{RSS}\left(P_{USER,r} - P_{i,r}\right), \tag{4.1.7}$$

where $p_{RSS}(\cdot)$ is the probability density function of the measured RSS values defining how RSS values vary in one fingerprint when multiple measurement from the same TX are considered. Based on the given definition of the likelihood, we assume that $p_{RSS}(\cdot)$ is identical for all fingerprints and TXs. Although this is not exactly true in practice, we assume that the database includes only one RSS value per each fingerprint and AP, and so, any information regarding the parameters of individual RSS probability density functions is inaccessible. The shape of the probability densify function $p_{RSS}(\cdot)$ can be freely chosen by design. Two distributions which are often mentioned in the literature are the Gaussian distribution as $p_{RSS}(v) = 1/\sqrt{2\pi\sigma_{RSS}^2}\exp\left(-v^2/2\sigma_{RSS}^2\right)$ with the variance σ_{RSS}^2 as a design parameter, and the exponential distribution as $p_{RSS}(v) = 1/2\exp\left(-\|v\|\right)$, where v is the function argument [61].

Using the fact that the fingerprint grid is uniform and assuming that there is no available a priori information on the true user position $\mathbf{x}_{USER}$ (i.e., $p(\mathbf{x}_{USER})$ is not known), based on the Bayes' rule [73], the posterior probability density function for the user position is directly proportional to the likelihood function as

$$p(\mathbf{x}_{USER} \mid \Omega_{RSS,USER}) \propto \sum_{i=0}^{N_{FP}-1} p(\Omega_{RSS,USER} \mid \mathbf{x}_i)\chi_i(\mathbf{x}_{USER}), \tag{4.1.8}$$

where

$$\chi_i(\mathbf{x}_{USER}) = \begin{cases} 1, & \text{if } \mathbf{x}_{USER} \text{ is mapped to the } i^{\text{th}} \text{ fingerprint} \\ 0, & \text{otherwise} \end{cases} \tag{4.1.9}$$

is an indicator function. Here the indicator function is used to apply the posterior function over the whole coordinate space so that the posterior function would not be restricted only to the known fingerprint coordinates. Thus, the posterior function is determined to be piecewise defined around

each fingerprint. As an example, the normalized posterior (i.e. likelihood) function of a real-life WLAN scenario in the multi-storey University building is given in Fig. 4-1.

The actual location estimate of the user can be obtained in several ways from the posterior function. One of the most famous approaches is the MAP estimation [86], where user estimate is the coordinates of the maximum value posterior function value as

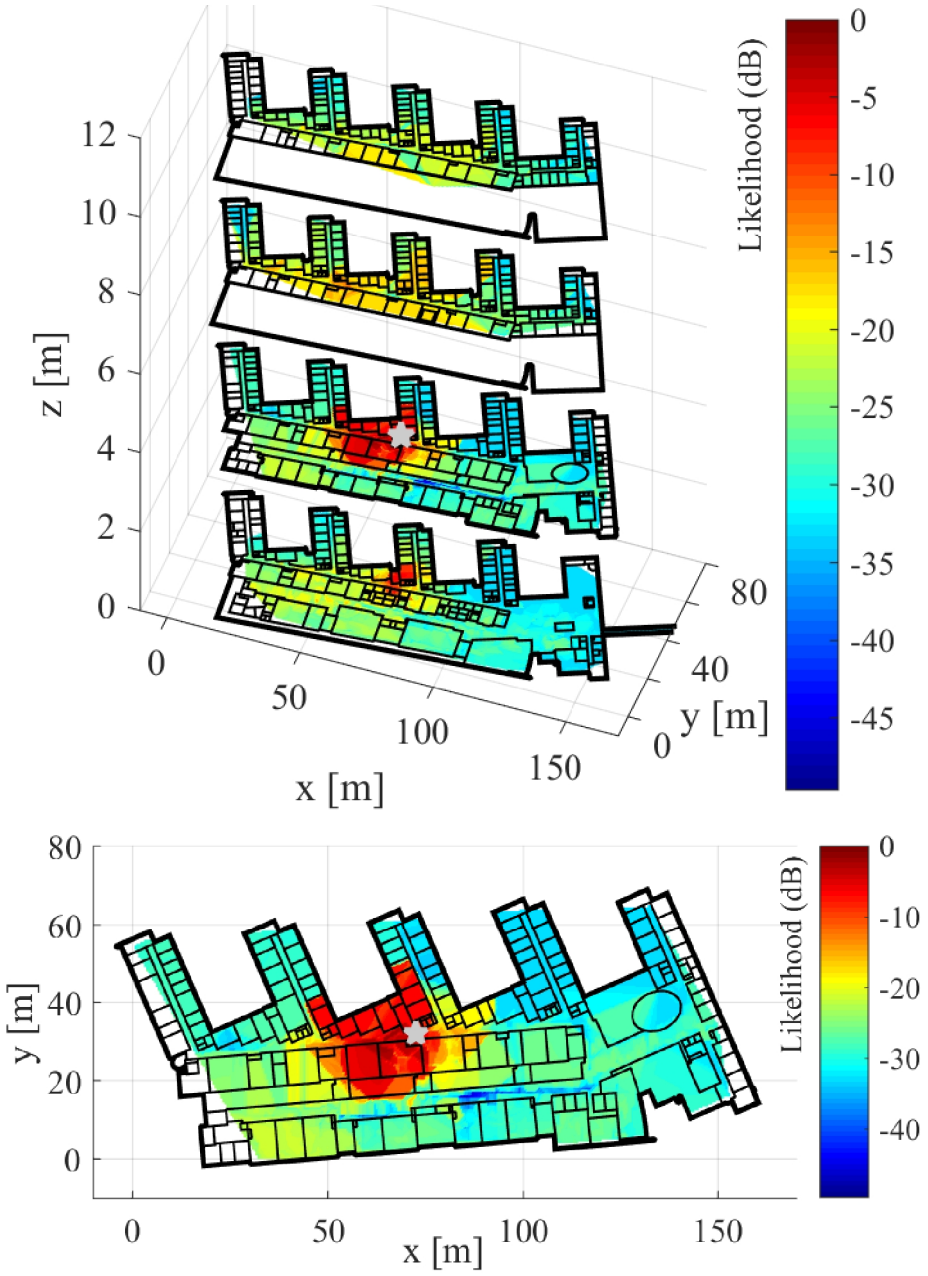

Fig. 4-1 *A normalized Bayesian fingerprinting-based likelihood (max. value is set to 0dB) of the user location for one set of user RSS measurements $\Omega_{RSS,USER}$ in the University building in 2.4GHz WLAN network: all floors (top) and the view on the 2nd floor only (bottom). The grey star around the maximum value of the likelihood is the true user location.*

$$\hat{\mathbf{x}}_{USER,MAP} = \arg\max_{\mathbf{x}_{USER}} p(\mathbf{x}_{USER} \mid \Omega_{RSS,USER}) . \tag{4.1.10}$$

In this case, since the a priori distribution is assumed to be uniform, the MAP estimate is identical with the ML estimate. Nevertheless, based on our studies and the studies in [61], a better average performance compared to the MAP estimate is obtained by taking the mean of the posterior function as

$$\hat{\mathbf{x}}_{USER,MMSE} = \sum_{i=0}^{N_{FP}-1} \frac{p\left(\Omega_{RSS,USER} \mid \mathbf{x}_i\right)}{\sum_{k=0}^{N_{FP}-1} p\left(\Omega_{RSS,USER} \mid \mathbf{x}_k\right)} \mathbf{x}_i , \tag{4.1.11}$$

which can be considered as the MMSE estimate of the user location.

4.1.2 Probabilistic localization using the PL models

There are several reasons for exploiting PL-model-based localization in present and future communications networks. Firstly, as the number of available TXs increase the fingerprinting-based approaches begin to suffer from the excessive amount of data needed to store and process from multiple different communications networks and TXs. In addition, because of the growth of the indoor localization systems, the data is also expanding vertically as multi-floor buildings have to be taken into account. Secondly, the PL models are naturally interpolating and extrapolating the learning data in all desired coordinates, whereas the in the fingerprinting, the fingerprints are restricted to provide information from only their own coordinates.

One of the traditional approaches for the PL-model-based localization is the trilateration principle [48],[90],[128],[145]. In trilateration, the fundamental idea is to determine the user location based on the range estimates to a set of known TX locations. Here, for simplicity, we focus on using only the single-slope log-distance model to estimate the ranges, but any other PL model is also applicable. Now, assuming the user measurement set $\Omega_{RSS,USER}$ given in (4.1.1), there range to the r^{th} heard TX can be obtained by solving the measurement distance parameter from (3.1.2) as

$$\hat{d}_{RANGE}^{(r)} = 10^{\frac{\hat{A}^{(r)} - P_{USER,r}}{10\hat{n}^{(r)}}} , \tag{4.1.12}$$

where $\hat{A}^{(r)}$ and $\hat{n}^{(r)}$ are the PL parameter estimates of the r^{th} TX. In addition, by assuming the estimated TX locations to be known, the distance between the user measurement location and each TX location is obtained as $\left\|\mathbf{x}_{USER} - \hat{\mathbf{x}}_{TX}^{(r)}\right\|$, where $\hat{\mathbf{x}}_{TX}^{(r)} = \begin{bmatrix} \hat{x}_{TX}^{(r)} & \hat{y}_{TX}^{(r)} & \hat{z}_{TX}^{(r)} \end{bmatrix}^T$ is the TX location esti-

mate given earlier in (3.3.1). Thus, the user location can be estimated by exploiting the LS principle as

$$\hat{\mathbf{x}}_{USER,trilater} = \arg\min_{\mathbf{x}_{USER}} \sum_{r \in \Omega_{TXheard}} \left(\left\| \mathbf{x}_{USER} - \hat{\mathbf{x}}_{TX}^{(r)} \right\| - \hat{d}_{RANGE}^{(r)} \right)^2 . \quad (4.1.13)$$

This equation can be solved, for example, by using iterative non-linear LS methods. An illustration of the trilateration principle is given is Fig. 4-2, where an artificial test set-up is created. In the set-up there are 3 TXs with known locations and noisy range estimates. The non-linear LS estimation is done by using the Newton-Raphson method [55] with two separate initializations. The figure also reveals one of the negative sides of using iterative methods as the first initialization convergences into an incorrect local maximum far away from the true user location.

Besides the trilateration, PL models can be used to recreate the original fingerprint data by approximating the RSS levels of each TX in each fingerprint This idea can be directly used for the localization purposes in similar way as in the Bayesian-based fingerprinting. Hence, again for simplicity

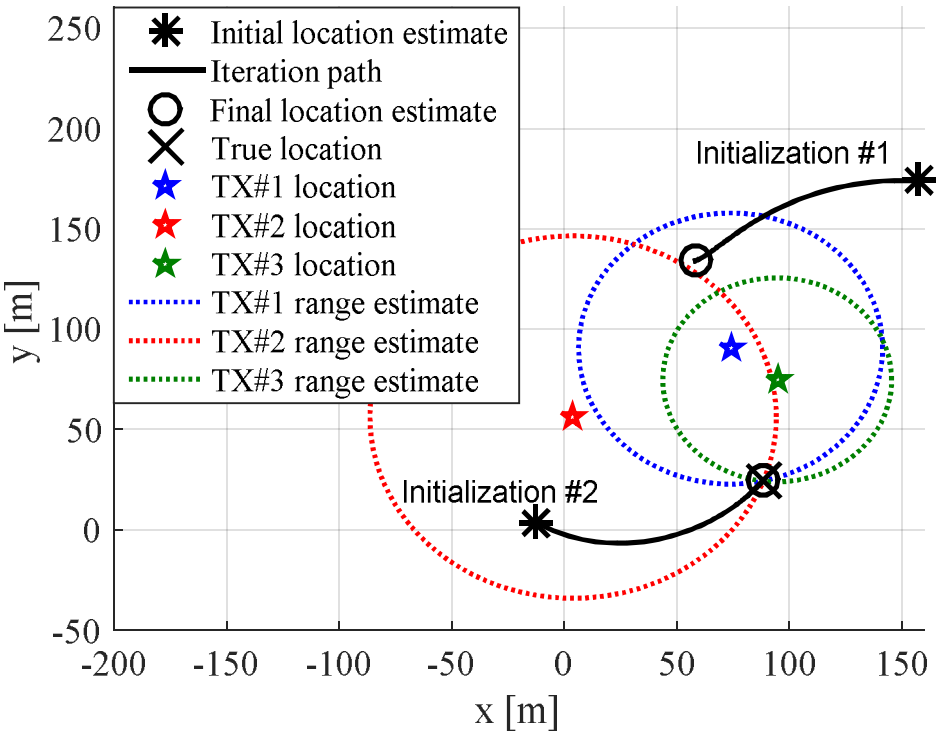

Fig. 4-2 *A simulated example of the trilateration principle, where 3 noisy range estimates are obtained from separate TXs with known TX locations. The final location estimate is achieved by solving a non-linear LS problem iteratively with the Newton-Raphson algorithm. The two tested parameter initializations in the algorithm result in different final location estimates.*

we restrict our analysis to the single-slope log-distance models without excluding the possibility to use other PL models. Based on the estimated PL model parameters, the estimated RSS level from the r^{th} TX at any coordinate $\mathbf{x}_g$ can be given as

$$\hat{P}_{\mathbf{x}_g}^{(r)} = A^{(r)} - 10n^{(r)} \log_{10}\left(\left\|\hat{\mathbf{x}}_{RE}^{(r)} - \mathbf{x}_g\right\|\right). \tag{4.1.14}$$

Now, by considering the estimated RSS level $\hat{P}_{\mathbf{x}_g}^{(r)}$ as the estimate of the fingerprint measurement $P_{i,r}$, the NN, KNN, and WKNN along with the Bayesian method are all directly applicable with the PL models. However, on contrary to the fingerprinting methods, the RSS values of each TX can be defined in arbitrary coordinates in the system, and thus, no bogus values have to be used with the PL models. An example of the likelihood function based on the Bayesian-based approach using the PL models is shown in Fig. 4-3. The presented likelihood is given for exactly the same case used earlier with the Bayesian fingerprinting in Fig. 4-1. Here, we have used the re-created fingerprint grids based on the LS-based estimates of the single-slope log-distance PL model including the floor losses. The likelihood is calculated with the posterior function given in (4.1.8) by substituting $\mathbf{x}_i \rightarrow \mathbf{x}_g$ and by replacing the fingerprint RSS value $P_{i,r}$ found in the probability density function $p_{RSS}\left(P_{USER,r} - P_{i,r}\right)$ with the estimated value $\hat{P}_{\mathbf{x}_g}^{(r)}$. By comparing the PL-model-based likelihood with the earlier fingerprinting-based likelihood (i.e., Fig. 4-1 with Fig. 4-3), the PL-model-based likelihood seems smoother and it covers the whole building area including those areas where no learning data were original taken. However, in the PL-model-based approach, the likelihood is widely spread in multiple floors, which might affect the floor detection probability. Moreover, if the floor losses would not be included in the PL model, the likelihood spread would be even more severe due to asymmetric propagation losses in horizontal and vertical directions.

Similar to the fingerprinting algorithms, there are also other approaches for the user location estimation originating from the PL model estimation. One of these is the weighted-centroid-based method studied in [P7] and [20], in which the user location is determined as the weighted centroid of the location estimates of the heard TXs, similar to the TX location estimation given in (3.3.1). This intuitive method is based on the well-justified assumption that the user is located near the AP with the highest observed RSS level. As shown in the localization performance results shown in Section 4.4, the weighted centroid method provides comparable localization performance with PL-modeling-based estimation methods.

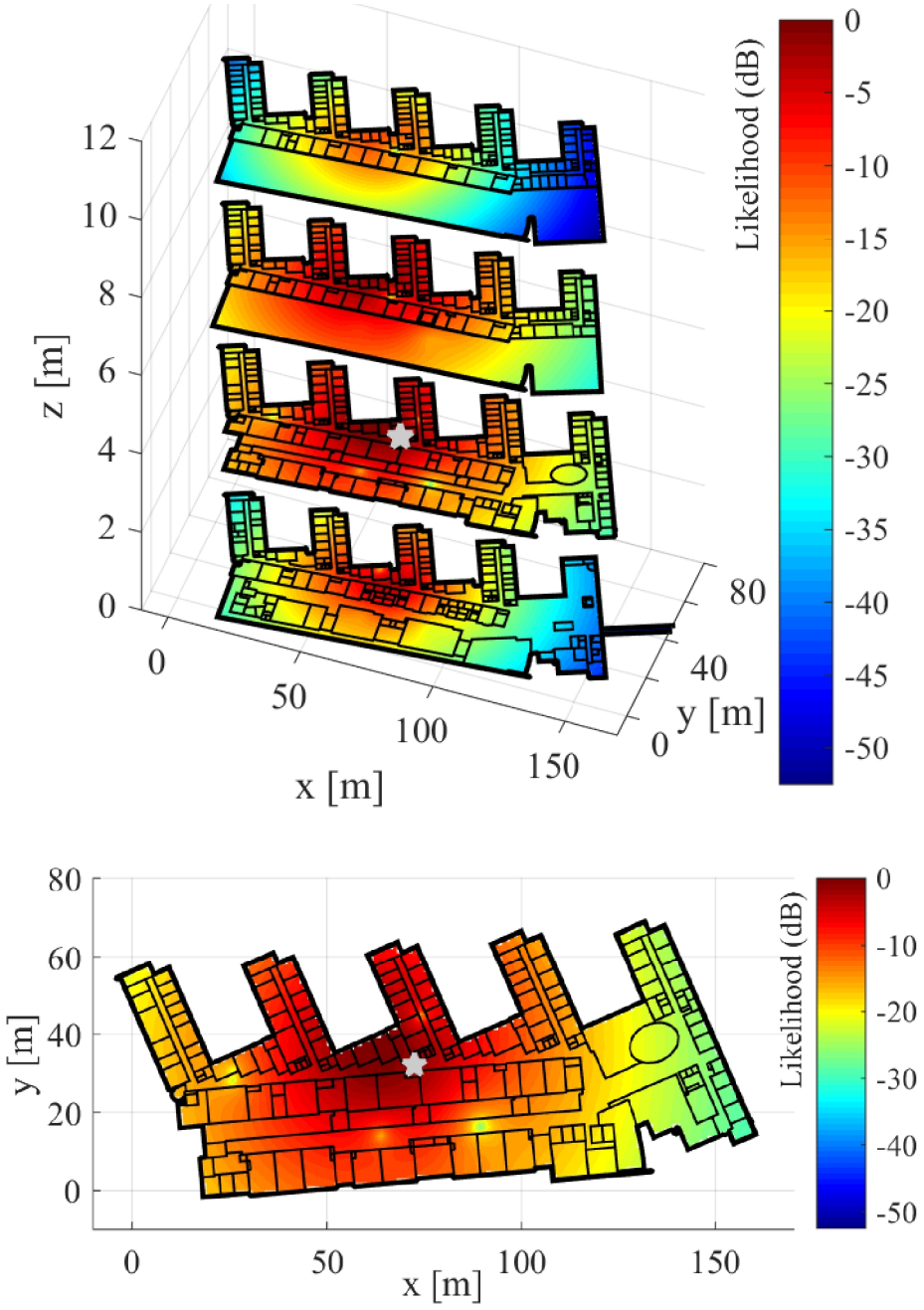

Fig. 4-3 *A normalized Bayesian PL-modeling-based likelihood (max. value is set to 0dB) of the user location for one set of user RSS measurements $\Omega_{RSS,USER}$ in the University building in 2.4GHz WLAN network: all floors (top) and the view on the 2nd floor only (bottom). The grey star around the maximum value of the likelihood is the true user location.*

4.2 Error sources in RSS-based user localization

In learning-based RSS data localization system, there are multiple sources of error affecting the localization accuracy. Since the localization is based on the pre-collected learning data, all changes in the learning data between the data collection time and the user localization time affect the localization performance. In [P6] we studied the effect of WLAN 2.4 GHz AP configuration changes in a University building by comparing two data sets taken from two years apart. Based on the given results, the effect of re-configuration was not significant as long as the learning data was kept updated. This implies that the building structure and the propagation environment is more influential for the localization accuracy than the AP topology.

Besides the outdated learning data induces error in the localization, also the quality of the data is an essential error source in localization systems. For example, if only one RSS measurement per TX is collected from each fingerprint, the mean of the local RSS distribution might be estimated very poorly. In fact, with only one RSS measurement, the variance of the estimated RSS mean is equal to the variance of the true RSS distribution in the fingerprint. This induces a calibration error in the learning data, where the error distribution is inherited from the RSS distribution, as discussed in Section 2.5. In addition to the calibration error due to low number of measurements, the RSS bias between the device used to collect the learning data and the device used to perform the localization can decrease the localization performance considerably. The problem with the RSS bias is that the error is affected in the same direction in each fingerprint, whereas with the typical calibration error the error distribution is symmetric. We have studied the effect of database calibration error in [P2], where we have considered a Gaussian distributed calibration error and the RSS bias error for both positive and negative bias values. The effect of the different error sources are depicted in Fig. 4-4 considering the KNN-based fingerprinting localization in the 2.4 GHz WLAN network in the University building. From the results, it is evident that the Gaussian distributed calibration error has the smallest effect on the localization error. Moreover, even with 20dB standard deviation, the localization performance is dropped by only less than 3 meters. Conversely, significant performance degradation can be witnessed with the positive RSS bias error. It is widely recognized the high level RSS measurements are very important from localization accuracy point of view, as they always point out the user location to be in very close proximity to the location of the observed TX. However, due to practical RSS boundaries, a large positive RSS bias results in RSS clipping and widens the area where the high RSS levels can be heard.

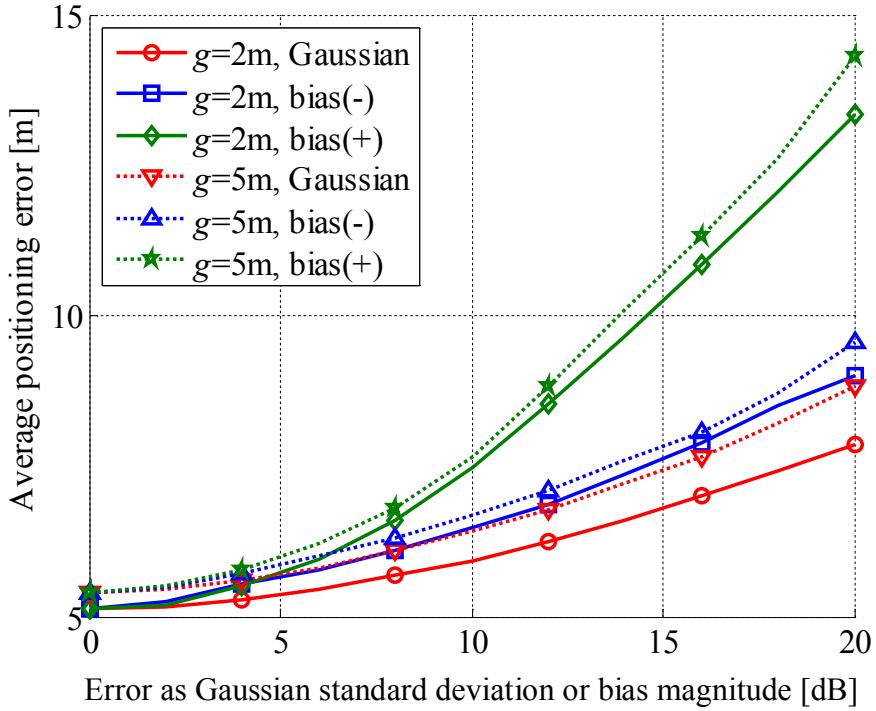

Fig. 4-4 *The effect of the RSS calibration error and the RSS bias error on average localization error using KNN-algorithm in the University building with 2.4GHz WLAN network. The results are given separately for fingerprint grid intervals of g=2m and g=5m.*

Depending on the considered communications network there might be some issues which cannot be anyway affected by the localization system design. For example, the number of heard TXs is often assumed to be depending on how many of the TXs are within the range (i.e., above some certain RSS level) of the measurement device. In WLANs this is roughly true, as all hearable APs are reported during each scan of the APs up to the some practical limit of maximum number of simultaneously scanned APs. However, in cellular networks the measurement process is strictly managed by the network and all the observable BSs are not necessarily measured during the process. Although this cannot be considered exactly as an error source of the localization system, it surely affects the localization performance. In Fig. 4-5, we show the Gaussian fitted distributions of the measured RSS values for different numbers of heard BSs per measurement in a GSM cellular network. It can be clearly seen that, if high level RSS values are available in the measurements, the number of measured BSs is smaller than with the low RSS values, where the number of measured BSs increase. This happens in the GSM network, because without any BSs with a high RSS level available, the measurement device is probably around the cell edge and several BSs are monitored in order to prepare for an upcoming handover between cells. On the other hand, if high

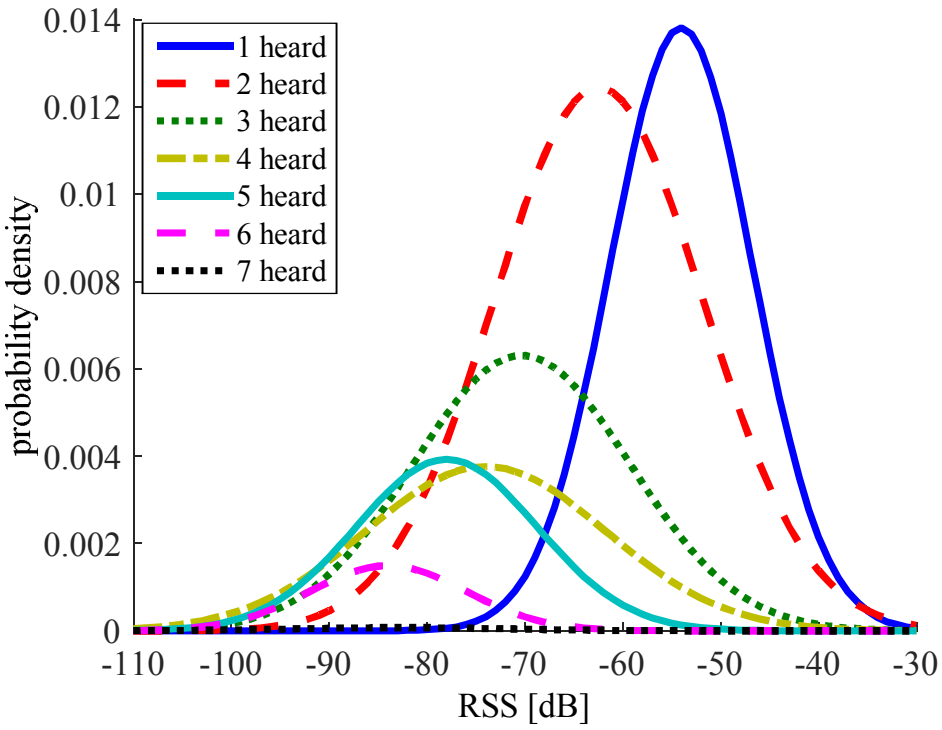

Fig. 4-5 Gaussian fitted distributions of the measured RSS values for different numbers of heard BSs per measurement set in a GSM cellular network.

RSS levels are observed, the measurement device is probably near the BS and imminent handovers are not required, and thus, there is no need to monitor other BSs.

4.3 Effect of coverage gaps on the positioning accuracy

Occasionally, the learning data does not necessarily cover the whole target area of the localization system. From the localization point of view this is problematic, since some of the areas are left in a total unawareness of the surrounding radio environment. These coverage gaps can be originated simply from poorly covered collection of learning data or due to the fact that some of the areas are restricted or inaccessible by the persons collecting the data. In addition to these, parts of the learning data might get outdated or lose their integrity, which might lead to exclusion of the data.

The study of the coverage gaps is not generally straightforward. To reveal the effect of coverage gaps on the localization performance, we must understand the performance of the original localization system without the coverage gaps. Therefore, in [P1] and [P2] we have proposed a randomized method to synthetically remove fingerprints from the original full coverage fingerprint database. By this way, it is possible to study how the missing fingerprints specifically affect the system per-

formance, since the original fingerprint database is fully known as a reference. The removal of the fingerprints cannot be done based on the uniform probability distribution, where each fingerprint has an equal probability of being removed. Instead, the coverage gaps are assumed to include large chunks of nearby measurements, which results in visible holes in the fingerprint grid. For this purpose we introduce a design parameter d_{block}, which describes the radius of the chunk (in meters) of removed fingerprints. Another required design parameter is μ, which determines the percent of removed fingerprints with respect to the initial number of fingerprints. Furthermore, the coverage gaps are assumed to be formulated independently in a floor-wise manner. Now, by denoting the set of original fingerprint indices as $\Omega_{FULL} = \{0,1,2,...,N_{FP}-1\}$, and by initializing the set of partial fingerprint indices as $\Omega_{PARTIAL} = \Omega_{FULL}$, the fingerprint removal method can be described as follows:

1. Select randomly the coordinates (x_s, y_s, z_s) of one fingerprint from $s \in \Omega_{PARTIAL}$ by using the uniform probability distribution.
2. Remove all fingerprints whose Euclidian distance in the horizontal plane (the xy-plane) to the randomly selected fingerprint (x_s, y_s, z_s) is smaller or equal to the block radius parameter d_{block}. The preserved fingerprints are now defining the new partial database $\Omega_{PARTIAL}$.
3. Check if enough fingerprints have been removed and the desired removal percentage μ satisfies the inequality $|\Omega_{PARTIAL}| \leq (1-\mu)|\Omega_{FULL}|$. If yes, continue to the part 4. Otherwise, go back to the part 1 and continue removing fingerprints.
4. In case $|\Omega_{PARTIAL}| = (1-\mu)|\Omega_{FULL}|$ is not satisfied with one fingerprint accuracy, retrieve a required number of the fingerprints from the last removed block starting from the fingerprints with largest distance to (x_s, y_s, z_s).

In Fig. 4-6, the removal process has been illustrated with the original fingerprints, the partial fingerprints obtained by the above described removal process, and the partial fingerprint set obtained by the uniform removal process. The percent of removed fingerprints is μ=30% for both of the removal methods. It can be clearly seen that the uniform removal method does not create coverage gaps, but more likely, it decreases the average density of the original fingerprints. Because of the randomized nature of the above described removal process, the process can be repeated over several times with different random number realizations. Consequently, the results of the effects of the coverage gaps in the following analysis are based on the average of 100 different realizations of the removal process.

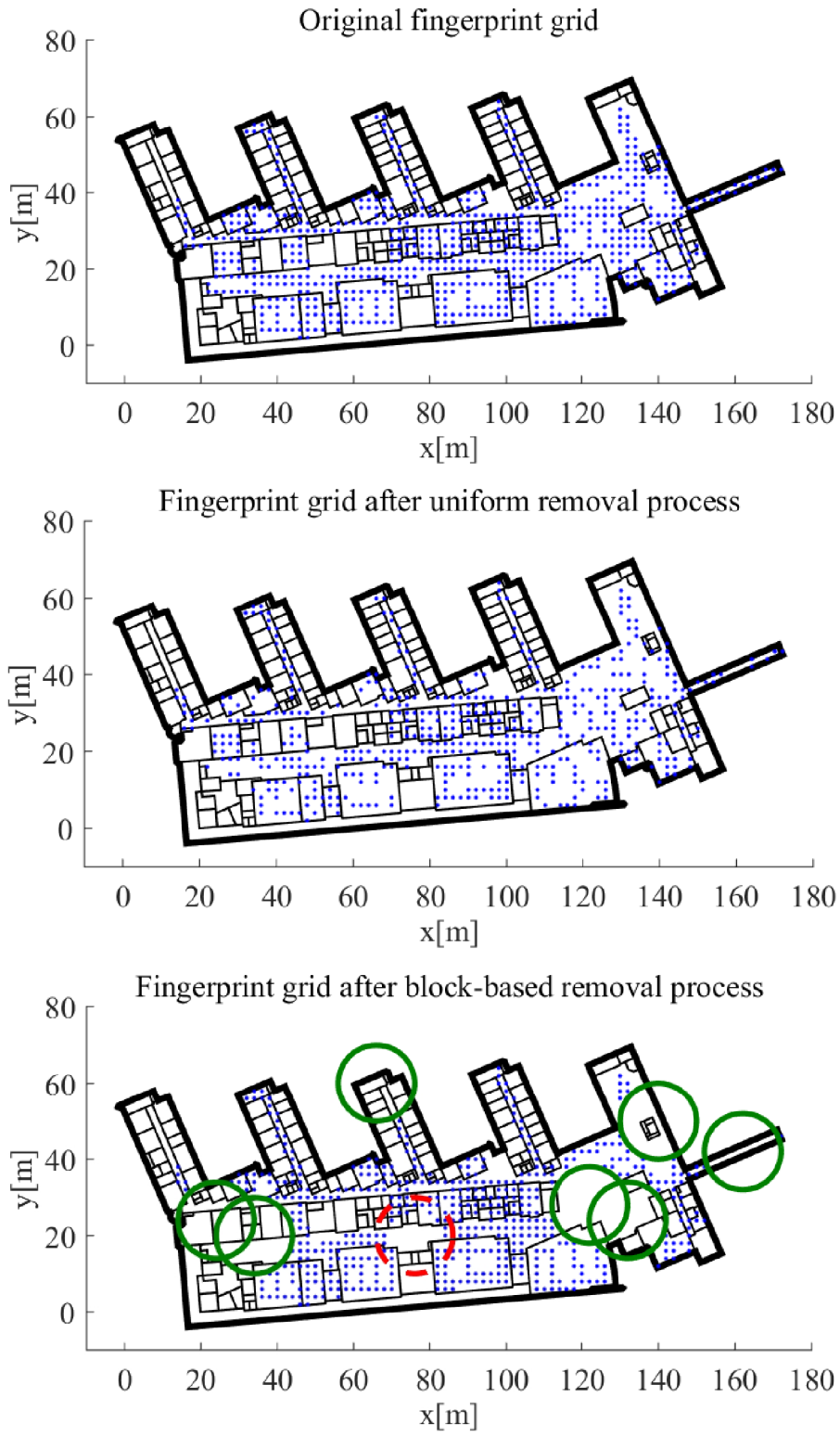

Fig. 4-6 *An illustration of the original fingerprint grid (top) and the partial fingerprint grids with 30% of the fingerprints removed (i.e., μ = 30%) by using the uniform removal (middle) and the block-based removal with d_{block} =10m (bottom). In the block removal case the circles indicate the removed areas of the removal process. The red and dashed circle is the last removed area, where part of the fingerprints has been retrieved in order to satisfy the desired removal percentage μ.*

In Fig. 4-7 we have studied the effect of coverage gaps on the average localization error and the floor detection probability by using 2 different grid intervals and 3 different fingerprint removal methods. The used localization algorithm was the KNN-based fingerprinting. It can be seen that

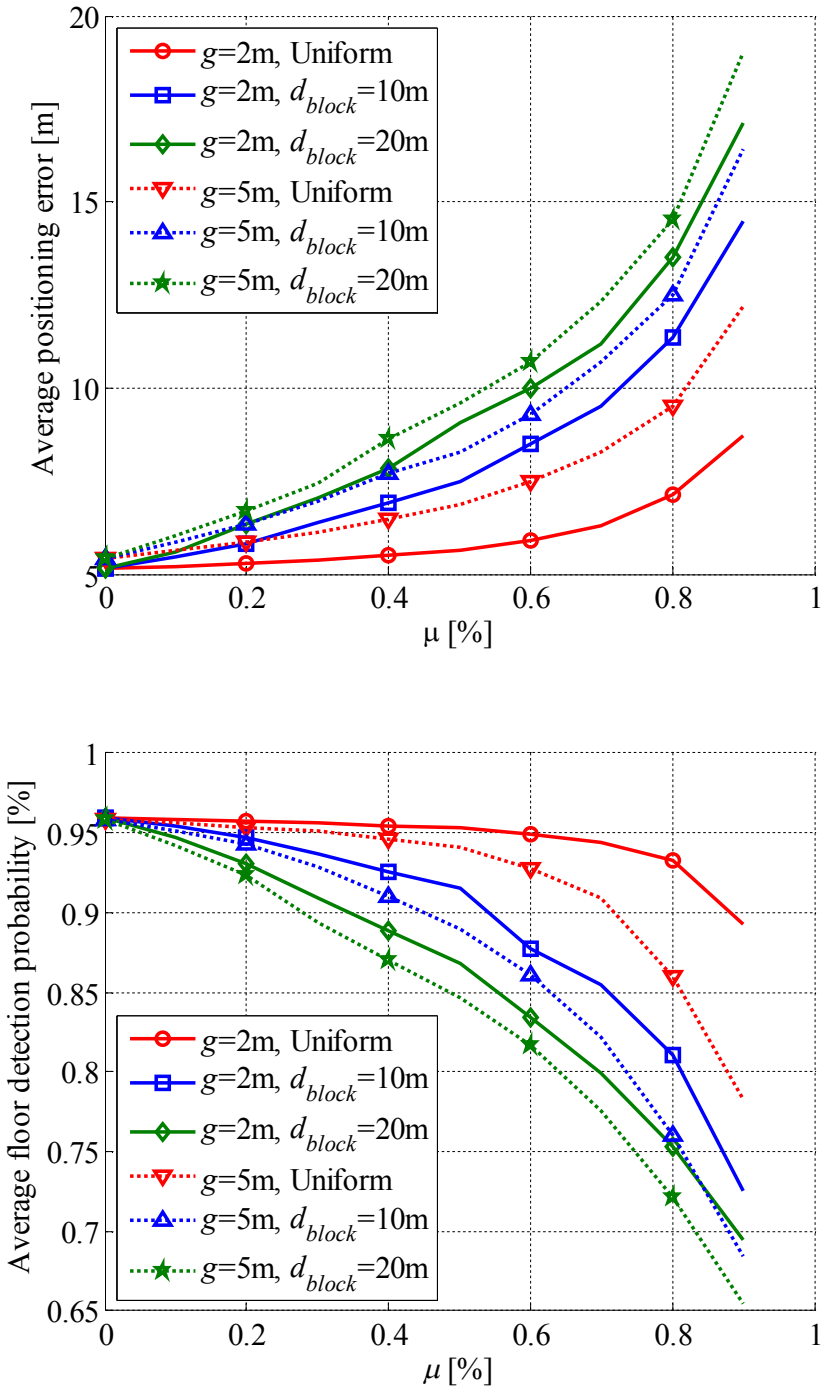

Fig. 4-7 *The effect of the coverage gaps on the average localization error and floor detection probability as a function of removal percentage μ with different fingerprint removal methods using the KNN-algorithm in the University building with 2.4GHz WLAN network. The results are given separately for fingerprint grid intervals of g=2m and g=5m.*

the block-based removal methods have considerably lower localization accuracy compared to the uniform removal method. Here, it should also be noticed that the selection of the localization algorithm has a crucial role in achieving reasonable localization accuracy. If coverage gaps are present in the dataset, the algorithms whose estimates are based on taking an average over multiple fingerprint coordinates, such as the KNN, WKNN and MMSE, are superior against the NN. This is because, if the user is in the middle of a coverage gap, the NN is forced to place the estimates somewhere in the gap boundaries, whereas KNN, WKNN, and MMSE are able to find the estimate inside the gap.

In [P1] we have studied different TX-wise and floor-wise interpolation and extrapolation methods for recovering the RSS values inside the coverage gaps. We considered several well-known interpolation and extrapolation methods and the target was to improve the localization performance decreased by the coverage gaps. It was shown that with the interpolation alone, there was no significant effect on the localization performance, since the average-based localization algorithms were able to find the location estimates inside the coverage gaps anyway. However, if also extrapolation of the RSS values was considered, it was possible to improve the localization performance

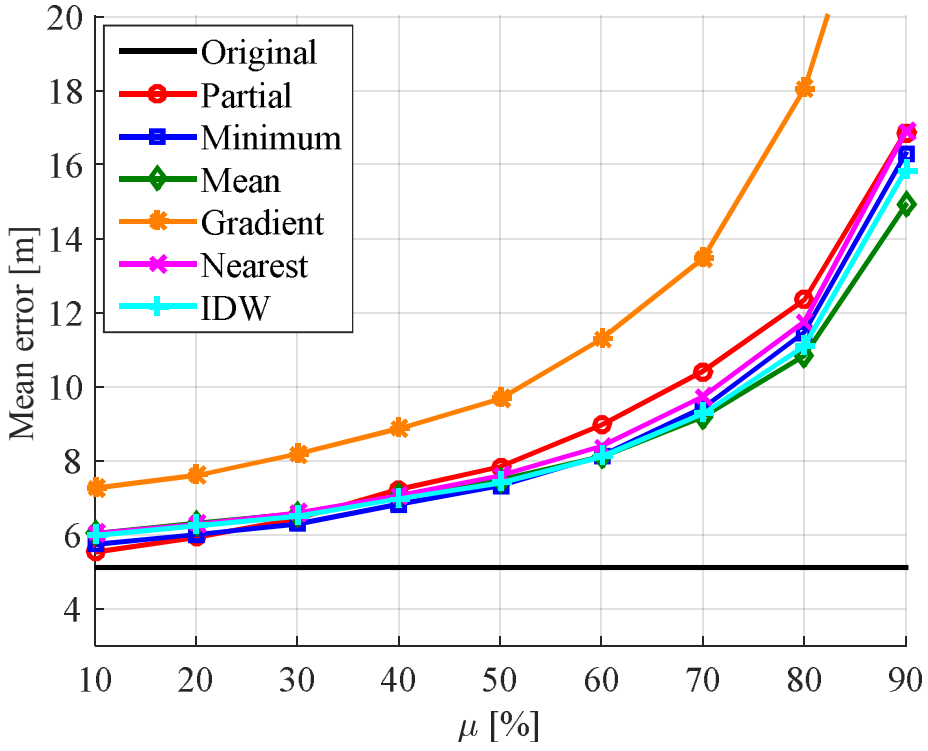

Fig. 4-8 *Mean positioning error as a function of the removal percentage μ considering original fingerprints, partial fingerprints (block-based removal with d_{block} =10m) and for different extrapolation methods after interpolation in the University building with 2.4GHz WLAN network (fingerprint grid interval of 5m).*

by certain extrapolation approaches. Fig. 4-8 shows the results regarding the mean localization error as a function of the removal percent μ for different studied extrapolation approaches and using the block removal method with d_{block} =10m. Here the minimum and mean methods refer to constant-based extrapolation approaches, where the extrapolated value is always either the minimum or the mean value over all available RSS values. In the gradient method the extrapolated values are based on the gradient of the closest available RSS values. In the nearest-method the extrapolated values were determined as the RSS value in the closest known location. Finally, the most consistent extrapolation method was determined to be the Inverse Distance Weighting (IDW), also referred as the Shepard's algorithm [124]. Here the extrapolated RSS value P_{EXTRAP} at the location (x_{EXTRAP}, y_{EXTRAP}) in the f^{th} floor is defined based on distance-weighted RSS levels as

$$P_{EXTRAP} = \sum_{j \in \Omega_f^{(r)}} \frac{d_{EXTRAP,j}^{-u}}{\sum_{k \in \Omega_f^{(r)}} d_{EXTRAP,k}^{-u}} P_j^{(r)}, \text{ where}$$
$$d_{EXTRAP,j} = \sqrt{\left(x_{EXTRAP} - x_j^{(r)}\right)^2 + \left(y_{EXTRAP} - y_j^{(r)}\right)^2} \quad (4.3.1)$$

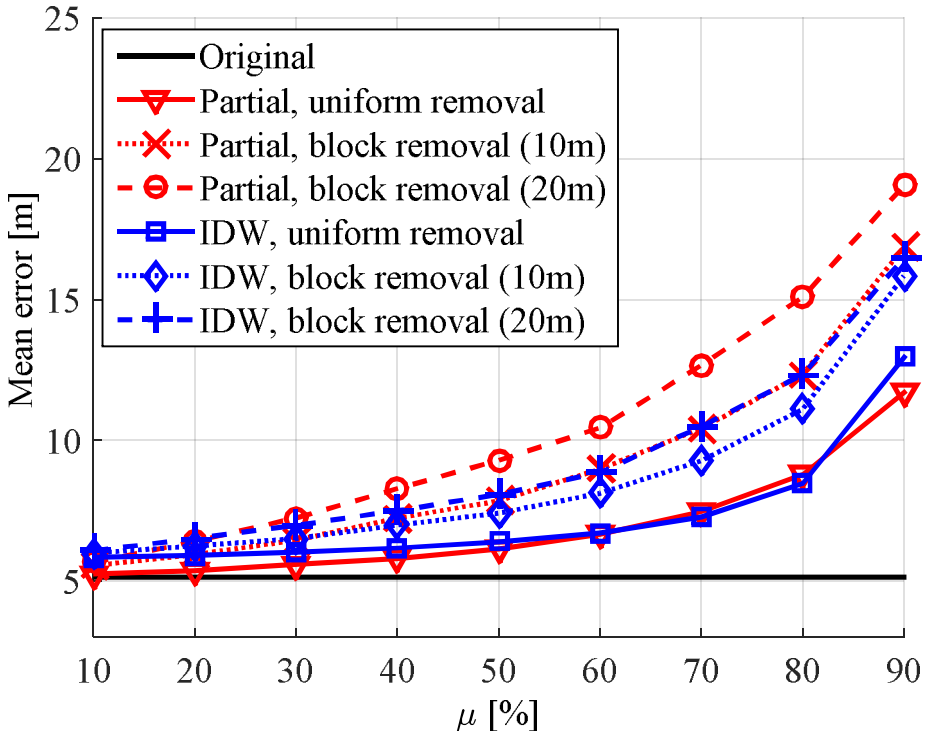

Fig. 4-9 *Mean positioning error as function of removal percentage μ considering original fingerprints, partial fingerprints and IDW interpolated/extrapolated fingerprints for uniform fingerprint removal and block removal with d_{block} =10m and d_{block} =20m in the University building with 2.4GHz WLAN network (fingerprint grid interval of 5m).*

where $\Omega_f^{(r)}$ is the set of measurement indices of the r^{th} TX found in the f^{th} floor, and thus a subset of the TX measurement set given earlier in (3.1.1). Here, the design parameter u can be used to control the relative weight between different distances. Based on the results, apart from the gradient-based approach the extrapolation seems to reduce the mean localization error up to 12% compared to the case without the extrapolation. The problem with the gradient-based method is that sometimes increasing gradient in the edge of the area of known RSS values results in very unrealistic extrapolation results. In Fig. 4-9, the IDW method is separately studied with different fingerprint removal methods. It seems that the more severe is the effect of the coverage gaps on the localization accuracy the more gain can be achieved via the extrapolation.

4.4 Comparison of localization performance for the considered localization systems

We have studied the localization performance using the discussed RSS-based localization algorithms in various different communications systems, including the WLANs at both 2.4 GHz and 5 GHz carriers, BLE, and cellular networks with GSM and WCDMA. For each system we have compared the performance of 3 different localization approaches: the conventional fingerprinting discussed in Section 4.1.1, the PL-model-based localization and the heuristic weighted centroid method, discussed in Section 4.1.2. In case of the PL-model-based approach we have considered the log-distance model, and moreover, for the indoor environment we study the models with and without the floor losses including the frequency-dependency.

In Fig. 4-10 the cumulative localization error of the indoor localization is showed for WLANs at the 2.4 GHz and 5 GHz carriers and the BLE at the 2.4 GHz carrier. All results are obtained from the same University building used earlier in this thesis to maintain a fair comparison between different systems. Based on the results, the 2.4 GHz WLAN and the BLE outperform the 5 GHz WLAN. However, this is mainly due to lack of measurements at 5 GHz carrier, since most of the WLAN APs are still operating at the 2.4G Hz carrier. In the WLANs the traditional fingerprinting is the most accurate localization approach, but in the BLE the weighted centroid method seems to be the most accurate one. In all cases the PL model where the floor losses are included provides better results compared to the model without the floor losses. Especially the floor detection probability is significantly improved by including the floor losses into the PL model.

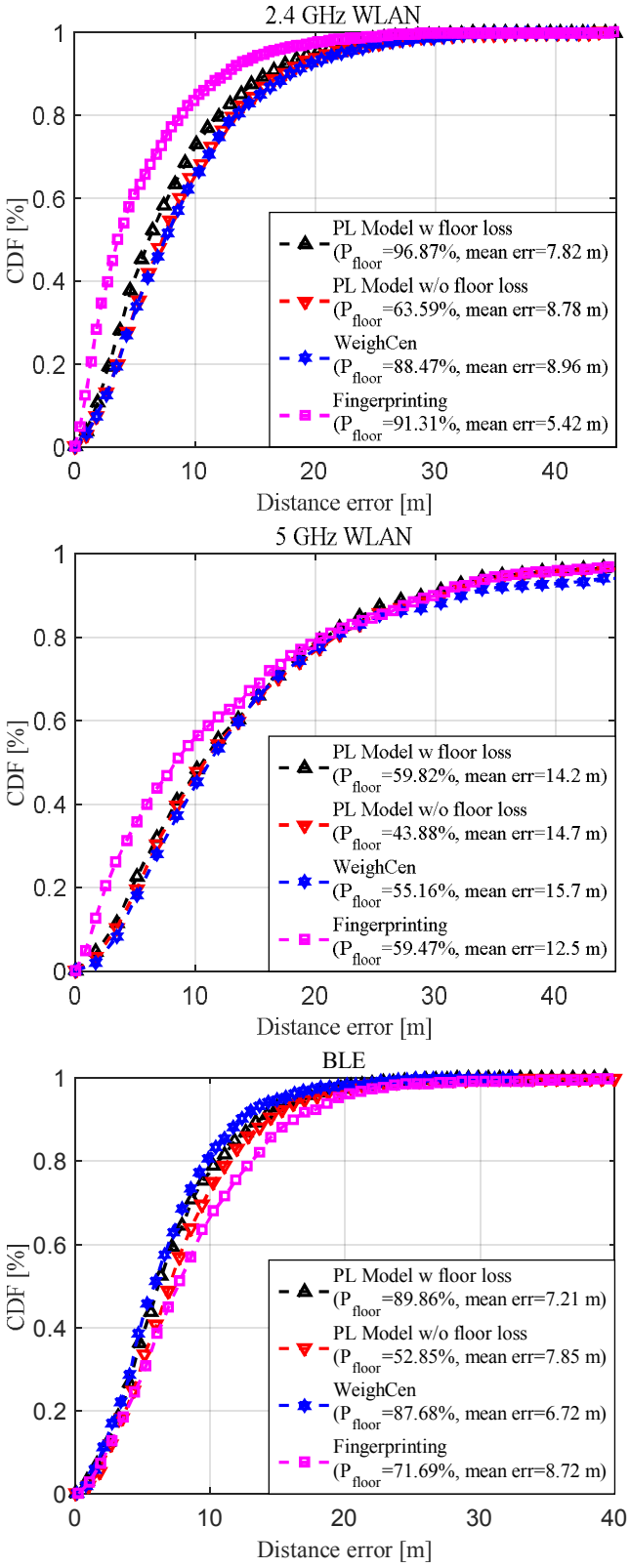

Fig. 4-10 *Comparison between cumulative localization error between different localization approaches in 2.4 GHz and 5 GHz WLANs and BLE in the University building. The results include the Bayesian PL model approach with and without floor losses, the weighted centroid approach, and the Bayesian fingerprinting approach. The mean errors and the floor detection probabilities P_{floor} are given in the legend for each localization approach.*

In Fig. 4-11 similar localization accuracy results are shown for the cellular networks, including the sub-urban GSM network and an urban WCDMA network. In both cases, the fingerprinting achieves the best average localization accuracy and the PL-model-based localization has the worst average accuracy among the considered approaches. However, in the GSM case, the PL models provide consistently better accuracy up to the 90% quantile, where the PL error curve saturates and the weighted centroid method begins to reach the fingerprinting curve. Hence, it seems that handling certain large errors in the PL modeling approach would increase the average estimation accuracy considerably.

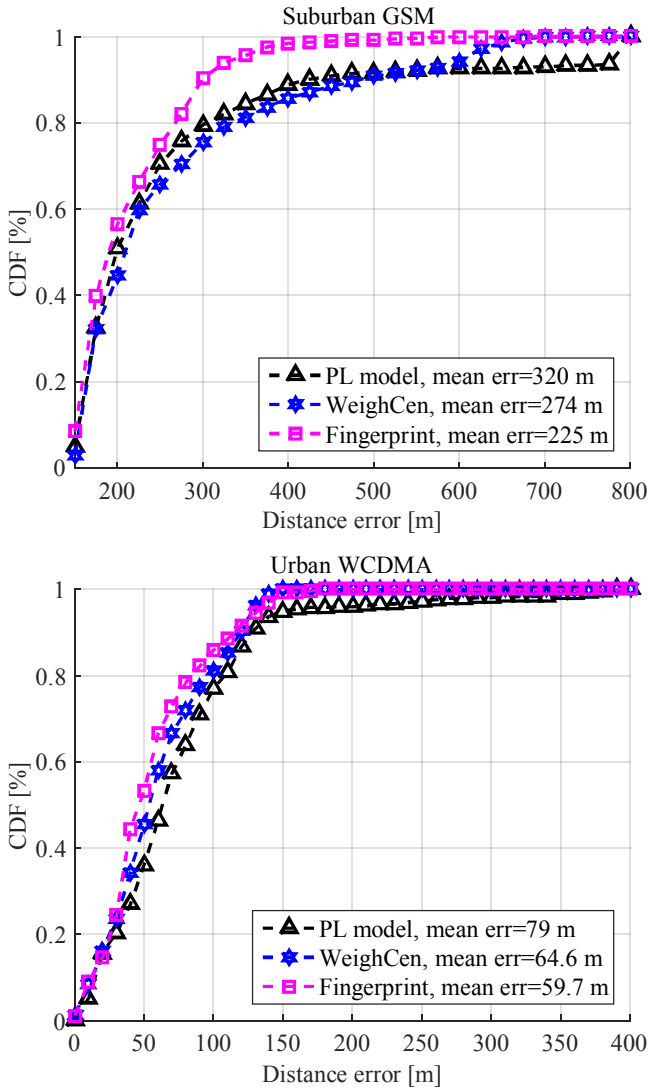

Fig. 4-11 *Comparison between cumulative localization error between different localization approaches in the suburban GSM and the urban WCDMA cellular networks. The results include the Bayesian PL model approach, the weighted centroid approach, and the Bayesian fingerprinting approach. The mean errors are given in the legend for each localization approach.*

The overall performance of the localization system is not only about the actual localization accuracy, but only about the algorithm complexity and the size of the learning database. Thus the number of parameters required to be stored in the learning database are shown in Table 3. Here, in case of the outdoor systems, namely the GSM and WCDMA, there are no values for the PL model with floor losses, since the outdoor models are considered only in 2D.

Table 3. *Number of parameters required to be stored in the database for the considered localization approaches in different localization systems.*

	System				
Localization approach	**WLAN 2.4GHz**	**WLAN 5GHz**	**BLE**	**GSM**	**WCDMA**
Weighted centroid	1764	460	312	96	78
PL model with floor losses	3087	805	546	-	-
PL model without floor losses	2646	690	468	160	130
Fingerprinting	139302	22525	41715	5481	1246

By comparing the numbers of stored database elements in the fingerprinting with the corresponding number in the weighted centroid or the PL-model-based approaches, it is evident that the fingerprinting approach is not able to compete with the other approaches in terms of database size. For example, in the 2.4 GHz WLAN, the fingerprinting database includes 45 times more database elements than the database of PL models with floor losses. Moreover, it is a fair assumption that the database size correlates with the computational complexity of the localization algorithm. Because of these reasons, the applicability of the fingerprinting methods at a global scale, especially as new exploitable communications network emerge in the future, is not straightforward. Although the problem regarding the actual database size could be solved, the data traffic between the database server and the user device might become as a bottleneck of the system. In addition, the data traffic handling in the server must also be designed to take into account the traffic generated from the database updates and general database maintenance duties. Nevertheless, the decision of which localization approach to use should depend on the scale of the localization system and on the availability of the database and computational resources.

5 Conclusions and Future Considerations

In this thesis we have studied different approaches for RSS-based localization in indoor and outdoor wireless localization systems. In the beginning, we justified the utilization of the RSS measurements for the localization purposes based on the fact that the RSS measurements are widely available in various wireless communications networks. In addition, the RSS measurements can be easily accessed via the APIs of different operating systems, which enables the use of RSS-based localization in most of the user devices available in the market. The number of the available networks and the TXs is expected to be increased in the future, which enables improved availability and accuracy for the upcoming localization systems.

The research focus in this thesis has been in the two-step localization approach, where in the first step the learning data is collected from the target area, and in the second step, the user localization is performed by exploiting the learning phase data. We addressed different aspects of the learning data collection and indicated that the statistics of the RSS measurements are heavily affected by the manner how the data collection is performed. In any case, due to the unpredictable radio propagation environment, simplified models for the RSS measurement statistics can be very difficult to obtain.

The RSS value can be considered as a function of the propagation distance between the measurement device and the TX. This dependency between the propagation distance and the RSS value can be modeled using the PL models. Thus, by assuming known PL model parameters, the user device can be located based on the triangulation principle and the localization can be performed without the collection of learning data. However, at the global scale it is very challenging, or even impossible, to find PL models with a reasonable number of model parameters so that it suits in all possible radio propagation environments. Moreover, certain PL model parameters in the localization systems, such as the effective antenna heights of the user device and the TX, are often unavailable, which makes certain PL models unfeasible for the localization purposes.

We have studied several PL models for both the 3D indoor and the 2D outdoor environments and used them to reduce the database size compared with the traditional fingerprinting, since only a few parameters per PL model have to be stored per each TX. We discussed several PL parameter estimation methods, such as the LS, WLS, and MMSE, from which the used method can be selected based on the availability of the prior information on the shadowing standard deviation and the a priori distribution of the PL parameters. We compared the distribution of the estimated PL model parameters between different indoor and outdoor systems. Based on the results, we observed that, as the number of obstacles in the radio path increases, also the PL exponent increases. This observation is well-aligned with the intuition and the earlier results indicated in the literature. Nevertheless, we also pointed out the significant correlation between the PL constant and the PL exponent parameters in the single-slope log-distance model. Besides the single-slope models, we provided methodology for modeling and estimating multi-slope PL models, which can improve the modeling accuracy for certain radio propagation environments.

We have studied different localization methods for the deterministic and probabilistic fingerprinting approaches and for the probabilistic PL-model-based approach. The comparison of the localization performance between the probabilistic fingerprinting approach and the probabilistic PL-model-based approach showed that the fingerprinting approach provides better localization accuracy and floor detection probability compared to the PL-model-based approach. However, in terms of the required database size, the PL-model-based approach rises clearly above the fingerprinting approach with up to 50 times smaller database size compared to the fingerprinting in case of the BLE system. This sort of difference in the database sizes cannot be simply ignored in the localization performance comparison, and thus, the exploitation of the PL-model-based approach should be seriously considered as a reasonable option for the future localization systems. Another type of a method for reducing the size of the learning database is proposed in [131]. Here the learning database is compressed based on spectral analysis, which exploits the spatial correlation of the RSS measurements taken from the same TX. Here, up to a certain database compression ratio, the localization accuracy with the compressed database is surprisingly slightly better compared to the original uncompressed database. This is because with relatively small compression ratios, the proposed method basically removes only the noise from the fingerprints and leaves the essential RSS information untouched.

In addition to the localization accuracy comparisons, we have studied also the effect of RSS mean calibration error and RSS bias error in the learning database on the localization performance. It was shown that the Gaussian distributed calibration error was not as severe as the bias errors. Moreover, the direction of the bias error was shown to be important regarding the degradation of the localization performance. Besides the RSS errors in the database, we also studied the effect of

incomplete database on the localization performance by introducing a randomized iterative method to create coverage gaps in the fingerprint grid. To recover the lost fingerprints we proposed multiple interpolation and extrapolation methods and reduced the localization error with the recovered database by up to 12% compared to the incomplete database.

As already indicated earlier, the number of the available networks and the corresponding TXs is expected to be increased in the future. Due to the excessive amount of data included in all localization systems, fingerprinting-based systems might become very challenging from the data handling point of view. Thus, building-up and maintaining extensive databases might become an overwhelming task in the future. To tackle this problem, compressive sensing methods [50],[99], and advanced machine learning methods [11],[18],[22],[42],[43],[44],[51],[60],[66],[95],[133],[135],[136] have been proposed in the literature. These latter ones are generally referred to as the SLAM methods, as they perform the learning phase and the user localization phase simultaneously. Of course, here the estimation problem becomes more challenging as the number of unknown variables increase. In [36] it has been shown that several traditional signal processing methods are inconsistent when the system dimensions are large. To tackle this, the authors in [36] use random matrix theory jointly with complex analysis methods for an improved consistency in large-dimensional systems. Nevertheless, by efficiently exploiting the studies regarding the RSS-related radio propagation characteristics in the literature, such as the ones discussed in this thesis, the SLAM approach could be the next big step in the future RSS-based localization systems.

Another important future study topic is the hybridization of the RSS with other types of localization information, such as the AOA, TOA and TDOA, as studied, for example, in [104] and [106]. In addition, hybridization of the RSS with possible inertial sensor measurements is especially interesting, as the inertial sensor data is often available by the API of mobile operation systems. Moreover, inertial sensors have also an important role in the self-learning process of the above-mentioned SLAM methods. Another exciting localization enabling sensor is the magnetometer, which measures the direction of the earth's magnetic field. This can be used for the indoor localization, since the local distortions in the magnetic field caused by the structural properties of the building can be considered as location-based fingerprints. Since each localization approach have unique advantages and disadvantages, a proper combination of different approaches should improve the overall localization performance.

Bibliography

[1] *Technical Specification Group Services and System Aspects;Evaluation of path-loss technologies for Location Services(LCS)(Release 9)*, 3GPP TR 25.907, 2010.

[2] *Technical Specification Group GSM/EDGE Radio Access Network;Radio subsystem link control(Release 1999)*, 3GPP TS 05.08, 2005.

[3] *Technical Specification Group Radio Access Network;Base Station (BS) radio transmission and reception(FDD)(Release 12)*, 3GPP TS 25.104, 2014.

[4] *Technical Specification Group Radio Access Network;Spreading and modulation(FDD)(Release 12)*, 3GPP TS 25.213, 2014.

[5] *Technical Specification Group Radio Access Network;Physical layer;Measurements(FDD)(Release 12)*, 3GPP TS 25.215, 2015.

[6] *Technical Specification Group Radio Access Network;Evolved Universal Terrestrial Radio Access (E-UTRA);Physical layer;Measurements(Release 12)*, 3GPP TS 36.214, 2015.

[7] M.N. Abdallah, W. Dyab, T.K. Sarkar, M.V.S.N. Prasad, C.S. Misra, A. Lamparez, M. Salazar-Palma and S.W. Ting, "Further Validation of an Electromagnetic Macro Model for Analysis of Propagation Path Loss in Cellular Networks Using Measured Driving-Test Data," *IEEE Antennas Propag. Mag.*, vol. 56, no. 4, pp. 108-129, Aug., 2014.

[8] M.N. Abdallah, W. Dyab, T.K. Sarkar, C.S. Misra, M.V.S.N. Prasad, A. Lamparez and M. Salazar-Palma, "Electromagnetic macro model for analysis of propagation path loss in cellular networks," in *Proc. 2014 IEEE Antennas and Propagation Soc. Int. Symp.*, 2014, pp. 947-948.

[9] T.E. Abrudan, A. Haghparast and V. Koivunen, "Time Synchronization and Ranging in OFDM Systems Using Time-Reversal," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 12, pp. 3276-3290, Dec., 2013.

[10] M.A. Al-Ammar, S. Alhadhrami, A. Al-Salman, A. Alarifi, H.S. Al-Khalifa, A. Alnafessah and M. Alsaleh, "Comparative Survey of Indoor Positioning Technologies, Techniques, and Algorithms," in *Proc. 2014 Int. Conf. Cyberworlds*, 2014, pp. 245-252.

[11] P.F. Alcantarilla, J.J. Yebes, J. Almazan and L.M. Bergasa, "On combining visual SLAM and dense scene flow to increase the robustness of localization and mapping in dynamic environments," in *Proc. 2012 IEEE Int. Conf. Robotics and Automation*, 2012, pp. 1290-1297.

[12] A. Algans, K.I. Pedersen and P.E. Mogensen, "Experimental analysis of the joint statistical properties of azimuth spread, delay spread, and shadow fading," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 523-531, Apr., 2002.

[13] S. Ali-Loytty, T. Perala, V. Honkavirta and R. Piche, "Fingerprint Kalman Filter in indoor positioning applications," in *Proc. 2009 IEEE Control Appl. Intelligent Control*, 2009, pp. 1678-1683.

[14] A. Alin, J. Fritsch and M.V. Butz, "Improved tracking and behavior anticipation by combining street map information with Bayesian-filtering," in *Proc. 2013 16th Int. IEEE Conf. Intelligent Transportation Syst.*, 2013, pp. 2235-2242.

[15] N. Alsindi, Z. Chaloupka and J. Aweya, "Entropy-based non-line of sight identification for wireless positioning systems," in *Proc. 2014 Ubiquitous Positioning, Indoor Navigation and Location Based Service*, 2014, pp. 185-194.

[16] N. Alsindi, Z. Chaloupka, N. AlKhanbashi and J. Aweya, "An Empirical Evaluation of a Probabilistic RF Signature for WLAN Location Fingerprinting," *IEEE Trans. Wireless Commun.*, vol. 13, no. 6, pp. 3257-3268, June, 2014.

[17] V.J. Arokiamary, *Mobile Communications*, 1st ed. Pune, India: Technical Publications Pune, 2009.

[18] C. Arth, C. Pirchheim, J. Ventura, D. Schmalstieg and V. Lepetit, "Instant Outdoor Localization and SLAM Initialization," *IEEE Trans. Vis. Comput. Graphics*, vol. PP, no. 99, pp. 1-1, 2015.

[19] M. Babalou, S. Alirezaee, A. Soheili, A. Ahmadi, M. Ahmadi and S. Erfani, "Microcell path loss estimation using Log-Normal model in GSM cellular network," in *Proc. 2015 Int. Symp. Signals Circuits and Syst.*, 2015, pp. 1-4.

[20] T. Bagosi and Z. Baruch, "Indoor localization by WiFi," in *Proc. 2011 IEEE Int. Conf. Intelligent Computer Communication and Process.*, 2011, pp. 449-452.

[21] P. Bahl and V.N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proc. Nineteenth Annu. Joint Conf. IEEE Computer and Commun. Societies INFOCOM 2000*, 2000, pp. 775-784 vol.2.

[22] T. Bailey and H. Durrant-Whyte, "Simultaneous localization and mapping (SLAM): part II," *IEEE Robot. Autom. Mag.*, vol. 13, no. 3, pp. 108-117, Sept., 2006.

[23] M.R. Basheer and S. Jagannathan, "Localization and Tracking of Objects Using Cross-Correlation of Shadow Fading Noise," *IEEE Trans. Mobile Comput.*, vol. 13, no. 10, pp. 2293-2305, Oct., 2014.

[24] M.W. Bern, D. Eppstein, L.J. Guibas, J. Hershberger, S. Suri and J. Wolter, "The Centroid of Points with Approximate Weights," in *Proc. Third Annu. Eur. Symp. Algorithms*, 1995, pp. 460-472.

[25] P.S. Bithas, "Weibull-gamma composite distribution: alternative multipath/shadowing fading model," *Electron. Lett.*, vol. 45, no. 14, pp. 749-751, July, 2009.

[26] A. Bjorck, *Numerical Methods for Least Squares Problems*, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1996.

[27] *Specification of the Bluetooth System: Covered Core Package version: 4.0*, Bluetooth SIG, 2010.

[28] L. Bruno, P. Addesso and R. Restaino, "Indoor Positioning in Wireless Local Area Networks with Online Path-Loss Parameter Estimation," *Scientific World J.*, vol. 2014, Article ID 986714, Aug., 2014.

[29] J.M. Castro-Arvizu, J. Vila-Valls, P. Closas and J.A. Fernandez-Rubio, "Simultaneous tracking and RSS model calibration by robust filtering," in *Proc. 2014 48th Asilomar Conf. Signals, Systems and Computers*, 2014, pp. 706-710.

[30] J.M. Castro-Arvizu, P. Closas and J.A. Fernandez-Rubio, "Cramer-Rao lower bound for breakpoint distance estimation in a path-loss model," in *Proc. 2014 IEEE Int. Conf. Commun. Workshops*, 2014, pp. 176-180.

[31] J. Chan and E.T. Wong, "Empirical modelling of received signal strength in indoor localization," in *Proc. Int. Conf. Automat. Control and Artificial Intell.*, 2012, pp. 978-981.

[32] Y. Chapre, P. Mohapatra, S. Jha and A. Seneviratne, "Received signal strength indicator and its analysis in a typical WLAN system (short paper)," in *Proc. 2013 IEEE 38th Conf. Local Computer Networks*, 2013, pp. 304-307.

[33] K. Cheung, J.H.-M. Sau and R.D. Murch, "A new empirical model for indoor propagation prediction," *IEEE Trans. Veh. Technol.*, vol. 47, no. 3, pp. 996-1001, Aug., 1998.

[34] J.S. Choi, H. Lee, D.W. Engels and R. Elmasri, "Passive UHF RFID-Based Localization Using Detection of Tag Interference on Smart Shelf," *IEEE Trans. Syst., Man, Cybern.,Syst.*, vol. 42, no. 2, pp. 268-275, Mar., 2012.

[35] P. Closas, C. Fernandez-Prades and J. Vila-Valls, "Multiple Quadrature Kalman Filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6125-6137, Dec., 2012.

[36] R. Couillet and M. Debbah, "Signal Processing in Large Systems: A New Paradigm," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 24-39, Jan., 2013.

[37] D. Dardari, P. Closas and P.M. Djuric, "Indoor Tracking: Theory, Methods, and Technologies," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1263-1278, April., 2015.

[38] L.B. Del Mundo, R.L.D. Ansay, C.A.M. Festin and R.M. Ocampo, "A comparison of Wireless Fidelity (Wi-Fi) fingerprinting techniques," in *Proc. 2011 Int. Conf. ICT Convergence*, 2011, pp. 20-25.

[39] F. Della Rosa, T. Paakki, J. Nurmi, M. Pelosi and G. Della Rosa, "Hand-grip impact on range-based cooperative positioning," in *Proc. 2014 11th Int. Symp. Wireless Commun. Syst.*, 2014, pp. 728-732.

[40] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Statistical Soc.*, vol. 39, no. 1, pp. 1-38, 1977.

[41] A. Dhital, P. Closas and C. Fernandez-Prades, "Bayesian filters for indoor localization using wireless sensor networks," in *Proc. 2010 5th ESA Workshop Satellite Navigation Technologies and Eur. Workshop GNSS Signals and Signal Process.*, 2010, pp. 1-7.

[42] G. Dissanayake, H. Durrant-Whyte and T. Bailey, "A computationally efficient solution to the simultaneous localisation and map building (SLAM) problem," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2000, pp. 1009-1014 vol.2.

[43] M.W.M.G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Trans. Robot. Autom.*, vol. 17, no. 3, pp. 229-241, June, 2001.

[44] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping: part I," *IEEE Robot. Autom. Mag.*, vol. 13, no. 2, pp. 99-110, June, 2006.

[45] E. Ekiz and R. Sokullu, "Comparison of path loss prediction models and field measurements for cellular networks in Turkey," in *Proc. 2011 Int. Conf. Select. Topics in Mobile and Wireless Networking*, 2011, pp. 48-53.

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

[47] L.M. Eslim, H.S. Hassanein, W.M. Ibrahim and A. Alma'aitah, "A cooperative localization scheme using RFID crowdsourcing and time-shifted multilateration," in *Proc. 2014 IEEE 39th Conf. Local Computer Networks*, 2014, pp. 185-192.

[48] C. Evrendilek and H. Akcan, "On the Complexity of Trilateration with Noisy Range Measurements," *IEEE Commun. Lett.*, vol. 15, no. 10, pp. 1097-1099, Oct., 2011.

[49] Z. Farid, R. Nordin and M. Ismail, "Recent Advances in Wireless Indoor Localization Techniques and System," *J. Comput. Networks. Commun.*, vol. 2013, Article ID 185138, 2013.

[50] C. Feng, S. Valaee and Z. Tan, "Multiple Target Localization Using Compressive Sensing," in *Proc. 2009 IEEE GLOBECOM Global Telecommun. Conf.*, 2009, pp. 1-6.

[51] J. Folkesson and H.I. Christensen, "Closing the Loop With Graphical SLAM," *IEEE Trans. Robot.*, vol. 23, no. 4, pp. 731-741, Aug., 2007.

[52] T. Gallagher, B. Li, A.G. Dempster and C. Rizos, "Database updating through user feedback in fingerprint-based Wi-Fi location systems," in *Proc. 2010 Ubiquitous Positioning, Indoor Navigation and Location Based Service*, 2010, pp. 1-8.

[53] Google Inc., WifiInfo. [Online]. Available: http://developer.android.com/reference/android/net/wifi/WifiInfo.html. Accessed: Sept. 9, 2015.

[54] F. Graziosi and F. Santucci, "A general correlation model for shadow fading in mobile radio systems," *IEEE Commun. Lett.*, vol. 6, no. 3, pp. 102-104, Mar., 2002.

[55] D.V. Griffiths and I.M. Smith, *Numerical Methods for Engineers, Second Edition*, 2nd ed. Boca Raton, FL, USA: CRC Press (Taylor & Francis Group), 2006.

[56] M. Gudmundson, "Correlation model for shadow fading in mobile radio systems," *Electron. Lett.*, vol. 27, no. 23, pp. 2145-2146, Nov., 1991.

[57] Z. Guowei, X. Zhan and L. Dan, "Research and improvement on indoor localization based on RSSI fingerprint database and K-nearest neighbor points," in *Proc. 2013 Int. Conf. Communications, Circuits and Syst.*, 2013, pp. 68-71.

[58] R. He, Z. Zhong, B. Ai and B. Zhang, "Measurement-based auto-correlation model of shadow fading for the high-speed railways in urban, suburban, and rural environments," in *Proc. 2014 IEEE Antennas and Propagation Soc. Int. Symp.*, 2014, pp. 949-950.

[59] S. He and S.-H.G. Chan, "Wi-Fi Fingerprint-based Indoor Positioning: Recent Advances and Comparisons," *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1-1, 2015.

[60] D.C. Herath, S. Kodagoda and G. Dissanayake, "New framework for Simultaneous Localization and Mapping: Multi map SLAM," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2008, pp. 1892-1897.

[61] V. Honkavirta, T. Perala, S. Ali-Loytty and R. Piche, "A comparative survey of WLAN location fingerprinting methods," in *Proc. 6th Workshop Positioning, Navigation and Communication*, 2009, pp. 243-251.

[62] Z. Huang, J. Xia, H. Yu, Y. Guan and J. Chen, "Clustering combined indoor localization algorithms for crowdsourcing devices: Mining RSSI relative relationship," in *Proc. 2014 Sixth Int. Conf. Wireless Commun. and Signal Process.*, 2014, pp. 1-6.

[63] J.M. Huerta, J. Vidal, A. Giremus and J.-Y. Tourneret, "Joint Particle Filter and UKF Position Tracking in Severe Non-Line-of-Sight Situations," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 5, pp. 874-888, Oct., 2009.

[64] *IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, IEEE Standard 802.11-2012, 2012.

[65] B.R. Jadhavar and T.R. Sontakke, "2.4 GHz Propagation Prediction Models for Indoor Wireless Communications Within Building," *Int. J. Soft Computing Eng.*, vol. 2, no. 3, pp. 108-113, 2012.

[66] M. Jung, H. Myung, S. Hong, D. Park, H. Lee and S. Bang, "Structured light 2D range finder for simultaneous localization and map-building (SLAM) in home environments," in *Proc. Int. Symp. Micro-Nanomechatronics and Human Sci. and The Fourth Symp. Micro-Nanomechatronics for Information-Based Soc.*, 2004, pp. 371-376.

[67] S. Jung, C. Lee and D. Han, "Wi-Fi fingerprint-based approaches following log-distance path loss model for indoor positioning," in *Proc. 2011 IEEE MTT-S Int. Microwave Workshop Series Intelligent Radio for Future Personal Terminals*, 2011, pp. 1-2.

[68] K. Kaemarungsi and P. Krishnamurthy, "Properties of indoor received signal strength for WLAN location fingerprinting," in *Proc. The First Annu. Int. Conf. Mobile and Ubiquitous Systems: Networking and Services*, 2004, pp. 14-23.

[69] K. Kaemarungsi, "Distribution of WLAN received signal strength indication for indoor location determination," in *Proc. 2006 1st Int. Symp. Wireless Pervasive Computing*, 2006, pp. 1-6.

[70] K. Kaemarungsi and P. Krishnamurthy, "Modeling of indoor positioning systems based on location fingerprinting," in *Proc. INFOCOM 2004. Twenty-third Annu. Joint Conf. IEEE Computer and Commun. Societies*, 2004, pp. 1012-1022 vol.2.

[71] S.S. Kanhere, "Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces," in *Proc. 2011 12th IEEE Int. Conf. Mobile Data Manage.*, 2011, pp. 3-6.

[72] A. Kara and H.L. Bertoni, "Effect of people moving near short-range indoor propagation links at 2.45 GHz," *J. Commun. Networks*, vol. 8, no. 3, pp. 286-289, Sept., 2006.

[73] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Upper Saddle River, NJ, USA: Prentice Hall, 1993.

[74] S. Khodayari, M. Maleki and E. Hamedi, "A RSS-based fingerprinting method for positioning based on historical data," in *Proc. 2010 Int. Symp. Performance Evaluation Computer and Telecommunication Syst.*, 2010, pp. 306-310.

[75] L. Klozar, J. Prokopec, S. Hanus, M. Slanina and Z. Fedra, "Indoor channel modeling based on experience with outdoor urban measurement - Multislope modeling," in *Proc. 2011 IEEE Int. Conf. Microwaves, Communications, Antennas and Electronics Syst.*, 2011, pp. 1-4.

[76] C. Koweerawong, K. Wipusitwarakun and K. Kaemarungsi, "Indoor localization improvement via adaptive RSS fingerprinting database," in *Proc. 2013 Int. Conf. Inform. Networking*, 2013, pp. 412-416.

[77] W. Kurschl, W. Gottesheim, S. Mitsch, R. Prokop, J. Schonbock and W. Beer, "Large-Scale Industrial Positioning and Location Tracking Are We There Yet?," in *Proc. 7th Int. Conf. Mobile Business2008*, 2008, pp. 251-259.

[78] S.A. Kyriazakos and G.T. Karetsos, *Practical Radio Resource Management in Wireless Systems*, Norwood, MA, USA: Artech House, 2004.

[79] C.L. Lawson and R.J. Hanson, *Solving Least Squares Problems*, Englewood Cliffs, NJ, USA: Prentice-Hall, 1974.

[80] B. Li, J. Salter, A.G. Dempster and C. Rizos, "Indoor positioning techniques based on wireless lan," in *Proc. 1st Int. Conf. Wireless Broadband and Ultra Wideband Commun.*, 2006, pp. 13-16.

[81] K. Li, J. Bigham and L. Tokarchuk, "Validation of a Probabilistic Approach to Outdoor Localization," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 167-170, Apr., 2013.

[82] X. Li, "RSS-Based Location Estimation with Unknown Pathloss Model," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3626-3633, Dec., 2006.

[83] D. Liu, Y. Xiong and J. Ma, "Exploit Kalman filter to improve fingerprint-based indoor localization," in *Proc. 2011 Int. Conf. Computer Sci. and Network Technol.*, 2011, pp. 2290-2293.

[84] J. Luo and X. Zhan, "Characterization of Smart Phone Received Signal Strength Indication for WLAN Indoor Positioning Accuracy Improvement," *J. Networks.*, vol. 9, no. 3, Mar., 2014.

[85] D. Macagnano, G. Destino and G. Abreu, "Indoor positioning: A key enabling technology for IoT applications," in *Proc. 2014 IEEE World Forum Internet Things*, 2014, pp. 117-118.

[86] J. Machaj, P. Brida and R. Piche, "Rank based fingerprinting algorithm for indoor positioning," in *Proc. 2011 Int. Conf. Indoor Positioning and Indoor Navigation*, 2011, pp. 1-6.

[87] J. Machaj and P. Brida, "Optimization of rank based fingerprinting localization algorithm," in *Proc. 2012 Int. Conf. Indoor Positioning and Indoor Navigation*, 2012, pp. 1-7.

[88] J. Machaj and P. Brida, "Impact of Wi-Fi Access Points on performance of RBF localization algorithm," in *Proc. 2012 ELEKTRO*, 2012, pp. 70-74.

[89] J. Machaj, P. Brida and J. Benikovsky, "Optimization of the RBF localization algorithm using Kalman filter," in *Proc. 2013 36th Int. Conf. Telecommun. and Signal Process.*, 2013, pp. 1-5.

[90] D.E. Manolakis, "Efficient solution and performance analysis of 3-D position estimation by trilateration," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 4, pp. 1239-1248, Oct., 1996.

[91] N. Marques, F. Meneses and A. Moreira, "Combining similarity functions and majority rules for multi-building, multi-floor, WiFi positioning," in *Proc. 2012 Int. Conf. Indoor Positioning and Indoor Navigation*, 2012, pp. 1-9.

[92] R.K. Martin, A.S. King, J.R. Pennington, R.W. Thomas, R. Lenahan and C. Lawyer, "Modeling and Mitigating Noise and Nuisance Parameters in Received Signal Strength Positioning," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5451-5463, Oct., 2012.

[93] Microsoft Inc., WiFiAvailableNetwork class. [Online]. Available: https://msdn.microsoft.com/en-us/library/windows/apps/windows.devices.wifi.wifiavailablenetwork.aspx. Accessed: Sept. 9, 2015.

[94] P. Mirowski, D. Milioris, P. Whiting and T.K. Ho, "Probabilistic radio-frequency fingerprinting and localization on the run," *Bell Labs Tech. J.*, vol. 18, no. 4, pp. 111-133, Mar., 2014.

[95] M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, "FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem," in *Proc. Eighteenth Nat. Conf. Artificial Intell.*, 2002, pp. 593-598.

[96] A. Moragrega, P. Closas and C. Ibars, "Potential Game for Energy-Efficient RSS-Based Positioning in Wireless Sensor Networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 7, pp. 1394-1406, July, 2015.

[97] C. Morelli, M. Nicoli, V. Rampa and U. Spagnolini, "Hidden Markov Models for Radio Localization in Mixed LOS/NLOS Conditions," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1525-1542, Apr., 2007.

[98] G. Morrison, M. Fattouche and H. Zaghloul, "Statistical analysis and autoregressive modeling of the indoor radio propagation channel," in *Proc. 1st Int. Conf. Universal Personal Commun.*, 1992, pp. 04.03/1-04.03/5.

[99] S. Nikitaki and P. Tsakalides, "Localization in wireless networks via spatial sparsity," in *Conf. Rec. Forty Fourth Asilomar Conf. Signals, Systems and Computers*, 2010, pp. 236-239.

[100] H. Nurminen, J. Talvitie, S. Ali-Loytty, P. Muller, E. Lohan, R. Piche and M. Renfors, "Statistical path loss parameter estimation and positioning using RSS measurements," in *Proc. 2012 Ubiquitous Positioning, Indoor Navigation, and Location Based Service*, 2012, pp. 1-8.

[101] H. Nurminen, J. Talvitie, S. Ali-Loytty, P. Muller, E. Lohan, R. Piche and M. Renfors, "Statistical path loss parameter estimation and positioning using RSS measurements in indoor wireless networks," in *Proc. 2012 Int. Conf. Indoor Positioning and Indoor Navigation*, 2012, pp. 1-9.

[102] S. Obayashi and J. Zander, "A body-shadowing model for indoor radio communication environments," *IEEE Trans. Antennas Propag.*, vol. 46, no. 6, pp. 920-927, Jun., 1998.

[103] T. Ohira, T. Hirai, S. Tomisato and M. Hata, "A study of mobile path loss estimation models for a sloping terrain area in cellular systems," in *Proc. 2012 18th Asia-Pacific Conf. Commun.*, 2012, pp. 472-477.

[104] L. Oliveira, C. Di Franco, T.E. Abrudan and L. Almeida, "Fusing Time-of-Flight and Received Signal Strength for adaptive radio-frequency ranging," in *Proc. 16th Int. Conf. Advanced Robotics*, 2013, pp. 1-6.

[105] C. Osterkorn, G. Ostermayer and M. Huemer, "Comparison of one- and two-dimensional slow fading models in mobile radio system simulations," in *Proc. IEEE 16th Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2005, pp. 2650-2654 Vol. 4.

[106] K. Papakonstantinou, M. Debbah and D. Slock, "MIMO Mobile Terminal Tracking Using Bayesian Probability Estimation," in *Conf. Rec. Forty-First Asilomar Conf. Signals, Systems and Computers*, 2007, pp. 1539-1543.

[107] M. Patzold, N. Avazov and V.D. Nguyen, "Design of measurement-based correlation models for shadow fading," in *Proc. 2010 Int. Conf. Advanced Technologies for Commun.*, 2010, pp. 112-117.

[108] V. Pawar and M. Zaveri, "Graph based K-nearest neighbor minutiae clustering for fingerprint recognition," in *Proc. 2014 10th Int. Conf. Natural Computation*, 2014, pp. 675-680.

[109] P. Prasithsangaree, P. Krishnamurthy and P.K. Chrysanthis, "On indoor position location with wireless LANs," in *Proc. The 13th IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2002, pp. 720-724 vol.2.

[110] A. Rai, K.K. Chintalapudi, V.N. Padmanabhan and R. Sen, "Zee: Zero-effort Crowdsourcing for Indoor Localization," in *Proc. 18th Annu. Int. Conf. Mobile Computing and Networking*, 2012, pp. 293-304.

[111] C.R. Rao, Shalabh, T. Helge and H. Christian, *Linear Models and Generalizations : Least Squares and Alternatives*, Berlin, Germany: Springer-Verlag Berlin Heidelberg, 2008.

[112] T.S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Piscataway, NJ, USA: Prentice Hall, 2002.

[113] *Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz*, Recommendation ITU-R P.1238-8, 2015.

[114] J. Reig and L. Rubio, "Estimation of the Composite Fast Fading and Shadowing Distribution Using the Log-Moments in Wireless Communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3672-3681, Aug., 2013.

[115] B. Roberts and K. Pahlavan, "Site-Specific RSS Signature Modeling for WiFi Localization," in *Proc. IEEE Global Telecommun. Conf.*, 2009, pp. 1-6.

[116] Mr. Rohan, Indoor Location Market Worth $3.9 Billion by 2019, Article in Directions Mag., Aug. 15th, 2014. [Online]. Available: http://www.directionsmag.com/pressreleases/indoor-location-market-worth-3.9-billion-by-2019/413751. Accessed: Sept. 9, 2015.

[117] T. Roos, P. Myllymaki, H. Tirri, P. Misikangas and J. Sievanen, "A Probabilistic Approach to WLAN User Location Estimation," *Int. J. Wireless Inform. Networks*, vol. 9, no. 3, pp. 155-163, July, 2002.

[118] A. Ruhe, "Accelerated Gauss-Newton algorithms for nonlinear least squares problems," *BIT Numerical Mathematics*, vol. 19, no. 3, pp. 356-367, 1979.

[119] P.M. Santos, T.E. Abrudan, A. Aguiar and J. Barros, "Impact of Position Errors on Path Loss Model Estimation for Device-to-Device Channels," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2353-2361, May., 2014.

[120] D. Schall, *Service-Oriented Crowdsourcing: Architecture, Protocols and Algorithms*, New York, NY, USA: Springer, 2012.

[121] M. Scherhaufl, M. Pichler and A. Stelzer, "UHF RFID Localization Based on Phase Evaluation of Passive Tag Arrays," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 4, pp. 913-922, Apr., 2015.

[122] J.S. Seybold, *Introduction to RF Propagation*, Hoboken, NJ, USA: Wiley & Sons, 2005.

[123] C.E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379-423, July, 1948.

[124] D. Shepard, "A Two-dimensional Interpolation Function for Irregularly-spaced Data," in *Proc. 23rd ACM Nat. Conf.*, 1968, pp. 517-524.

[125] S.T. Smith, "The Jump Tracker: Nonlinear Bayesian Tracking with Adaptive Meshes and a Markov Jump Process Model," in *Proc. Fortieth Asilomar Conf. SignalsSystems and Computers*, 2006, pp. 2004-2008.

[126] Y. Song and H. Yu, "A RSS Based Indoor Tracking Algorithm via Particle Filter and Probability Distribution," in *Proc. 4th Int. Conf. Wireless Communications, Networking and Mobile Computing*, 2008, pp. 1-4.

[127] T.B. Sorensen, "Correlation model for slow fading in a small urban macro cell," in *Proc. The Ninth IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 1998, pp. 1161-1165 vol.3.

[128] H. Staras and S.N. Honickman, "The accuracy of vehicle location by trilateration in a dense urban environment," *IEEE Trans. Veh. Technol.*, vol. 21, no. 1, pp. 38-43, Feb., 1972.

[129] G. Sterling and D. Top, Mapping the Indoor Marketing Opportunity, Opus Res. Rep., Feb. 1st, 2014. [Online]. Available: http://opusresearch.net/wordpress/pdfreports/OpusIndoorReport_24_Jan2014.pdf. Accessed: Sept. 9, 2015.

[130] A.I. Sulyman, A.T. Nassar, M.K. Samimi, G.R. MacCartney, T.S. Rappaport and A. Alsanie, "Radio propagation path loss models for 5G cellular networks in the 28 GHZ and 38 GHZ millimeter-wave bands," *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 78-86, Sept., 2014.

[131] J. Talvitie, M. Renfors and E.S. Lohan, "Novel Indoor Positioning mechanism via Spectral Compression," *IEEE Commun. Lett.*, accepted on Nov. 2015.

[132] The Signal Processing for wireless positioning group, WLAN indoor measurements. [Online]. Available: http://www.cs.tut.fi/tlt/pos/Measurements.htm. Accessed: Sept. 9, 2015.

[133] S. Thrun and M. Montemerlo, "The Graph SLAM Algorithm with Applications to Large-Scale Mapping of Urban Structures," *Int. J. Robot. Res.*, vol. 25, no. 5-6, pp. 403-429, May., 2006.

[134] P. Torteeka and X. Chundi, "Indoor positioning based on Wi-Fi Fingerprint Technique using Fuzzy K-Nearest Neighbor," in *Proc. 2014 11th Int. Bhurban Conf. Appl. Sciences and Technol.*, 2014, pp. 461-465.

[135] G. Tuna, K. Gulez, V.C. Gungor and T. Veli Mumcu, "Evaluations of different Simultaneous Localization and Mapping (SLAM) algorithms," in *Proc. IECON 2012 - 38th Annu. Conf. IEEE Ind. Electronics Soc.*, 2012, pp. 2693-2698.

[136] A. Vlavianos, L.K. Law, I. Broustis, S.V. Krishnamurthy and M. Faloutsos, "Assessing link quality in IEEE 802.11 Wireless Networks: Which is the right metric?," in *Proc. IEEE 19th Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2008, pp. 1-6.

[137] L. Wang and W. Wong, "A Novel RSS model and Power-bias Mitigation Algorithm in fingerprinting-based Indoor Localization in Wireless Local Area Networks," in *Proc. 20th Eur. Wireless Conf. Eur. Wireless 2014*, 2014, pp. 1-7.

[138] Y. Wen, X. Tian, X. Wang and S. Lu, "Fundamental limits of RSS fingerprinting based indoor localization," in *Proc. 2015 IEEE Conf. Computer Commun.*, 2015, pp. 2479-2487.

[139] G. Wolfle, R. Wahl, P. Wertz, P. Wildbolz and F. Landstorfer, "Deterministic Propagation Model for the Planning of Hybrid Urban and Indoor Scenarios," in *Proc. IEEE 16th Int. Symp. Personal, Indoor and Mobile Radio Commun.*, 2005, pp. 659-663.

[140] C. Wu, Z. Yang and Y. Liu, "Smartphones Based Crowdsourcing for Indoor Localization," *IEEE Trans. Mobile Comput.*, vol. 14, no. 2, pp. 444-457, Feb., 2015.

[141] L. Xin-di, H. Wei and T. Zeng-shan, "The Improvement of RSS-based Location Fingerprint Technology for Cellular Networks," in *Proc. 2012 Int. Conf. Computer Sci. Service System*, 2012, pp. 1267-1270.

[142] D. Yang, G. Xue, G. Fang and J. Tang, "Incentive Mechanisms for Crowdsensing: Crowdsourcing With Smartphones," *IEEE/ACM Trans. Netw.*, vol. PP, no. 99, pp. 1-13, 2015.

[143] S. Yang, P. Dessai, M. Verma and M. Gerla, "FreeLoc: Calibration-free crowdsourced indoor localization," in *Proc. 2013 Proc. IEEE INFOCOM*, 2013, pp. 2481-2489.

[144] Y. Yang, Y. Zhao and M. Kyas, "Recursive Bayesian Estimation Using a Topological Map for Indoor Position Tracking," in *Proc. 2014 IEEE 79th Veh. Technol. Conf.*, 2014, pp. 1-5.

[145] Z. Yang and Y. Liu, "Quality of Trilateration: Confidence-Based Iterative Localization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 5, pp. 631-640, May., 2010.

[146] K. Yedavalli, B. Krishnamachari, S. Ravula and B. Srinivasan, "Ecolocation: a sequence based technique for RF localization in wireless sensor networks," in *Proc. Fourth Int. Symp. Inform. Process. in Sensor Networks*, 2005, pp. 285-292.

[147] M. Youssef and A. Agrawala, "The Horus WLAN Location Determination System," in *Proc. 3rd Int. Conf. Mobile Systems Applications and Services*, 2005, pp. 205-218.

[148] M. Youssef and A. Agrawala, "The Horus Location Determination System," *J. Wireless Networks*, vol. 14, no. 3, pp. 357-374, June, 2008.

[149] D. Zeinalipour-Yazti, C. Laoudias, C. Costa, M. Vlachos, M.I. Andreou and D. Gunopulos, "Crowdsourced Trace Similarity with Smartphones," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 6, pp. 1240-1253, June, 2013.

[150] X. Zhang and J.G. Andrews, "Downlink Cellular Network Analysis With Multi-Slope Path Loss Models," *IEEE Trans. Commun.*, vol. 63, no. 5, pp. 1881-1894, May., 2015.

[151] Z. Zhang, Z. Lu, V. Saakian, X. Qin, Q. Chen and L. Zheng, "Item-Level Indoor Localization With Passive UHF RFID Based on Tag Interaction Analysis," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 2122-2135, Apr., 2014.

[152] Y. Zhuang, B. Wright, Z. Syed, Z. Shen and N. El-Sheimy, "Fast WiFi access point localization and autonomous crowdsourcing," in *Proc. 2014 Ubiquitous Positioning Indoor Navigation and Location Based Service*, 2014, pp. 272-280.

ORIGINAL PAPERS

Publication 1

J. Talvitie, M. Renfors and E.S. Lohan, "Distance-Based Interpolation and Extrapolation Methods for RSS-Based Localization with Indoor Wireless Signals," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1340-1353, 2015.

Publication 2

J. Talvitie, E.S. Lohan and M. Renfors, "The effect of coverage gaps and measurement inaccuracies in fingerprinting based indoor localization," in *Proc. Int. Conf. Localization and GNSS*, Helsinki, Finland, pp. 1-6, 2014.

Publication 3

J. Talvitie, M. Renfors and E.S. Lohan, “A Comparison of Received Signal Strength Statistics between 2.4 GHz and 5 GHz bands for WLAN-based Indoor Positioning,” accepted in *Proc. IEEE Globecom 2015 Workshop Localization for Indoors, Outdoors, and Emerging Networks*, San Diego, CA, USA, Dec., 2015.

Publication 4

S. Shrestha, J. Talvitie and E.S. Lohan, "Deconvolution-based indoor localization with WLAN signals and unknown access point locations," in *Proc. Int. Conf. Localization and GNSS*, Turin, Italy, pp.1-6, 2013.

Publication 5

J. Talvitie and E.S. Lohan "Modeling Received Signal Strength measurements for cellular network based positioning," in *Proc. Int. Conf. Localization and GNSS*, Turin, Italy, pp. 1-6, 2013.

Publication 6

S. Shrestha, J. Talvitie and E.S. Lohan, "On the fingerprints dynamics in WLAN indoor localization," in *Proc. Int. Conf. ITS Telecommun*., Tampere, Finland, pp. 122-126, 2013.

Publication 7

E. S. Lohan, J. Talvitie, P. Figueiredo e Silva, H. Nurminen, S. Ali-Loytty and R. Piche, "Received signal strength models for WLAN and BLE-based indoor positioning in multi-floor buildings," in *Proc. Int. Conf. Localization and GNSS*, Gothenburg, Sweden, pp. 1-6, 2015.

Publication 8

E.S. Lohan, K. Koski, J. Talvitie and L. Ukkonen, "WLAN and RFID Propagation channels for hybrid indoor positioning," in *Proc. Int. Conf. Localization and GNSS*, Helsinki, Finland, pp. 1-6, 2014.

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-3672-4
ISSN 1459-2045