TECHNISCHE
UNIVERSITÄT
WIEN

VIENNA
UNIVERSITY OF
TECHNOLOGY

DISSERTATION

Mobile Handset Localization by Received Signal Level Pattern Matching

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften

eingereicht an der
Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

durch

Dipl.-Ing. Harald Kunczier
König und Bauer-Strasse 23, 2344 Ma. Enzersdorf
geboren am 22. Mai 1976 in Mödling
Matr. Nr. 9525539

Wien, im Juni 2006 _____

Supervisor:

Prof. Dr. Ernst Bonek

Technische Universität Wien, Austria

Examiner:

Prof. Dr. Klaus Felsenstein

Technische Universität Wien, Austria

Kurzfassung

In dieser Arbeit untersuchen wir Positionierungsverfahren, die auf Mustererkennung basieren. Signalstärkemessungen werden dabei zur Charakterisierung der Position herangezogen und mit vorher aufgezeichneten oder simulierten Daten verglichen.

Durch eine zeitlich veränderliche Umgebung, durch Mehrwegeausbreitung und durch Unterbrechungen der Sichtverbindung zwischen Sender und Empfänger entstehen allerdings mehrdeutige Charakterisierungen der Position. Um diesem prinzipiellen Problem entgegenzuwirken, modellieren wir jede Position als stochastisches Modell. Dabei verwenden wir Bayessche Netze, um die Abhängigkeiten zwischen den Signalen der empfangenen Basisstationen zu beschreiben.

Wir optimieren die Bayessche Netzstruktur solcherart, dass sie Positionen im städtischen Raum bestmöglich repräsentiert. Als Metrik wird dabei die Verbundwahrscheinlichkeit zwischen Modellstruktur und gemessenen Daten an den jeweiligen Positionen verwendet. Dies erlaubt eine objektive Evaluierung der Qualität der Modelle.

Wir zeigen, dass a-priori Wissen über die Verteilung der Signalstärkemuster einer Position aus benachbarten Positionen gewonnen werden kann.

Weiters wird ein Algorithmus zur Positionsbestimmung, basierend auf den Bayesschen Netzen, mit geringer Latenzzeit vorgestellt. Wir führen dessen Komplexität auf eine einzige Maximierungsaufgabe zurück. Die Berechnung der zugrunde liegenden Daten in Echtzeit ist dabei nicht notwendig.

Der Algorithmus wird schließlich mit gängigen Mustererkennungsmethoden und mit satellitenbasierter Positionierung verglichen. Die Ergebnisse zeigen dabei

Fehler von 13.75m in 67% und 93.45m in 95% aller Fälle, welche den in den USA vorgeschriebenen Erfordernissen für mobile Notfalldienste entsprechen. Diese schreiben in 67% aller Fälle Fehler kleiner 100m und in 95% aller Fälle Fehler kleiner 300m für netzbasierte Methoden vor.

Wir zeigen weiters durch Langzeituntersuchungen von aufgezeichenten Signalstärkemesswerte die Notwendigkeit der regelmäßigen Aktualisierung der Signalstärkemusterdatenbank. Sinnvolle Aktualisierungsintervalle liegen bei etwa 42 Stunden.

Abstract

In this thesis we investigate pattern-matching-based positioning. Based on signal-power measurements, the position of the mobile is determined by either matching the measured data to pre-stored data in a database or by matching the measured data to simulated data.

To allow for the uncertainty about the environment, including time variant scatterers, multipath and non-line-of-sight situations, we apply a stochastic perspective. Propagation effects, which create the ambiguity of the radio frequency patterns, are modelled by the capability of Bayesian networks to encode uncertainty. The Bayesian networks are used to describe the dependencies between the different measured cell identities (serving cells and list of the neighboring cells ranked according to the received power levels) for each position.

We optimize the Bayesian model structure to represent positions in urban environment by maximizing the joint probability of the data and the model's structure over all models. This forms not only a view of the relative quality of the candidate models, but it also gives an assessment of the adequacy as an explanation of the data.

To include environment-caused physical facts we introduce an adapted prior distribution. This includes no multiple occurrence of a certain cell ID within a single pattern or less received power of more distant base stations.

We develop a positioning algorithm whose complexity is reduced to a single maximization of a database data vector, containing one element per node and position of the model. The effort-consuming computations are done off-line, allowing low latency of the positioning.

The algorithm is tested and compared against several state-of-the-art positioning methods, including other pattern matching algorithms and satellite based positioning. We achieve an error of 13.75m in 67% of all cases and an error of 93.45m in 95% of all cases in an urban environment. This is fully compliant with the requirements for mobile emergency services in the USA (E911), specifying an error of 100m in 67% of all cases and 300m in 95% of all cases for network-based methods.

We further evaluated long-term signal-level measurements revealing a validity of the signal level database of about 42 hours. During that period, it is still possible to determine 50% of all received fingerprints at the measurement site in 90% of all cases.

Acknowledgements

First I would like to thank Professor Vincent Hetherington, my uncle, who strongly influenced my first decision to add five additional years of school to write a PhD-thesis.

I would like to extend my gratitude to my advisor and first examiner Professor Ernst Bonek, who somehow always finds a time-slot for me in his busy calendar, whenever I ask for it. An equal amount of gratitude goes to my second supervisor, Professor Klaus Felsenstein, whose "door is always open". Thanks also to Dr. Markus Kommenda and Mag. Horst Rode, for hiring me at ftw. in the first place and, thus, for providing the opportunity to conduct the research for this thesis. Already in my first week I realized that ftw. is a great place with many friendly colleagues around. In particular I would like to mention Dr. Günther Pospischil, who first introduced me to the topic of this thesis and also Hermann Anegg for many nice and fruitful discussions.

I want to acknowledge mobilkom austria for the great support of this thesis in particular with the measurements.

A big hug goes to my parents Wilhelm and Christine, to my brother Stefan and of course to my friends.

Finally, my gratitude goes to my girlfriend Dominga, who does not put up with me working all the time, but lets me know when it is time to quit and come home.

Harald Kunczier

Table of Contents

Chapter 1

Introduction

Increasing interest in opportunities of mobile services has led to extended effort in research and development of localization technologies as a basis for location based services. The reason for this development is manifold but two main contributors can be identified.

First, in 1994, the United States Federal Communication Commission (FCC) issued a Notice of Proposed Rulemaking [1] in which it suggested regulations, requiring mobile operators to provide enhanced 911 emergency services including, among other, emergency call positioning. While there was general support for the goals set forth in this notice, the wireless industry and public safety groups differed with regard to the schedule for E911 deployment and the need for Federal regulation. Early in 1996, a Consensus Agreement between all groups was signed proposing a two-step implementation. This resulted in an adoption of the wireless rules in the same year [3]. Under step 1, named Phase I, which started April 1, 1998, wireless carriers had to provide Public Safety Answering Points (PSAPs) with a callback number and the location of the cell site/sector receiving a 911 call. Under Phase II, still ongoing and revised several times, the wireless carriers are required to provide to PSAPs the location of all 911 calls by longitude and latitude with an accuracy of 150 meters for 95 percent of calls. Table 1.1 shows a summary of the current requirements. The strong pressure from FCC has, on the one hand, led wireless operators to invest money in poor working localization solutions, but has, on the other hand, also pushed research and development. Europe has acted less regulative. The European Commission (EC) issued a recommendation (under Universal Service Directive [4]) instructing fixed and mobile network operators to provide caller location information to emergency service centers "to the extend technically feasible". So far, however, only a rough time line is scheduled, where the Cell-ID method is considered to be the only realistic minimum accuracy requirement. High accuracy services, which are equivalent to the US E911 services, are not realistic before 2007 [45].

Solution	1996 E911 requirements	1999 E911 requirements	
	67% of calls	67% of calls	95% of calls
Handset based	125 meters	50 meters	150 meters
Network based	125 meters	100 meters	300 meters

Table 1.1: FCC E911 Requirements.

Second, industry analysts predict annual location related revenue of several billions of euros by 2008. Although these values are less than predicted back in 2000, market is the significant driving force for new research and development effort. This is also proven by the current effort of the European satellite based localization system Galileo - opposed by the US - which will become fully operational in 2008.

During the work on my thesis, attention has several times turned towards different technologies and solutions. While, at the beginning of my work, strong attention was put on terrestrial time-of-arrival based solutions, some of these solutions failed important requirements like feasibility or accuracy (e.g. Enhanced Observed Time Difference of Arrival, refer to Section 2.3). Today, the Global Positioning System (GPS) and assisted GPS in particular, has become more interesting, since production costs and size for GPS receivers have decreased significantly. An additional push is coming from Galileo, which will double the number of space vehicles in orbit and hence improve accuracy and availability for users. The required line-of-sight to the satellites, however, is still a knock-out criteria for urban areas with narrow street canyons and indoor environments. In these areas, patter matching has proven a reliable alternative.

1.1 Field of Research and Objectives

Various positioning technologies exist, whereas most methods rely on communication between known reference points (e.g. base stations) and the target (e.g. mobile station) which are aimed to be localized. The known geometric configuration of the reference points enables the determination of the target's position. These methods, however, lack of efficiency in urban and heavy urban areas, where most of the mobile users are located.

This thesis proposes another approach which we call Network-Measurement-Report (NMR) based handset localization[1]. Based on signal power measurements

[1]This thesis was written in close cooperation with two research projects - C1: UMTS Applications Development, `http://lola.ftw.at/homepage` and C5: Network Management Report

already available in the network (due to handover reasons), the position of the mobile is evaluated either by matching the measured data to pre-stored data in a database or by matching the measured data to simulated data. The difference to known fingerprint or pattern matching methods is the usage of Bayesian models, offering the ability to describe the localization problem in a stochastic manner.

The objective of the research presented in this thesis is to study possibilities to improve and optimize localization methods for urban areas. The main emphasis was put on the NMR based handset localization, since it seems to offer the most potential solution for personal positioning in terms of accuracy in urban areas.

1.2 Thesis Organization and Contributions

The thesis is organized in the following chapters and the author's contributions are as follows:

Chapter 2 - Mobile Positioning: Chapter 2 introduces the reader to the topic of mobile positioning. We discuss the various localization technologies including satellite-based, terrestrial-based and signal-strength-based positioning. We illustrate advantages and disadvantages of all methods and conclude this chapter with a comparison of all methods in terms of accuracy, reliability, applicability, and latency. For a more thorough discussion of these topics we refer to reader to [34].

Chapter 3 - NMR based Positioning: Chapter 3 develops the physical framework for the NMR based handset localization introduced in this thesis. We describe the relevant propagation effects being responsible for the location dependence of the RF signals and show the complexity of the inverse problem, which is to derive the position of ambiguous signals. Based on the ideas of [32, 46] we develop a pattern matching method [35, 68], which is suitable for urban environments. We introduce the required pattern recognition approaches, the feature selection and the decision making. The chapter closes with an summary of all steps required for the NMR based handset localization method.

Chapter 4 - Probabilistic Networks: In Chapter 4 we concentrate on the construction of the Bayesian framework and introduce probabilistic models. Probabilistic models have the ability to learn about causal relationships and to encode the correlation between input variables. Requiring an optimal

Based Handset Localization, `www.ftw.at/ftw/research/projects/ProjekteFolder/C5` - at the Telecommunications Research Center Vienna.

representation of a certain position in terms of the received signal for the NMR based localization method, this mathematical framework provides the necessary tools and allows us to develop the positioning algorithms.

Chapter 5 - The Bayesian Model: In this chapter we developed the system model [37]. It is based on multiple received signals of different cells, which are characterized by different nodes of the model. The dependencies between the different nodes are optimized using measurements to achieve the best representation of a certain position. To allow for an easy computation of the posterior probability of the Bayesian network, we construct a prior probability distribution from expert knowledge and measurements from neighboring positions [36]. We apply the developed model on test measurements from urban environment and compare its performance to reference models utilizing the non-informative Jeffrey's prior.

Chapter 6.2 - The Positioning Algorithm: In this chapter we develop the positioning algorithm [39]. We show that the actual positioning can be reduced to a single maximization of a data vector, the elements of which can be computed off-line. This allows a fast and scalable localization with only low latency. We further evaluate the performance of the NMR based localization method in the city of Vienna (urban environment) [38] and compare it with reference pattern-matching-methods and with the satellite-positioning system GPS. We further discuss the weak point of pattern-matching methods in general, the durability of the required database and give indications of how long values in the database remain valid.

Chapter 7- Conclusions: This chapter summarizes the contributions of this thesis and contains a short statement about future possibilities for personal positioning.

Chapter 2

Mobile Positioning

2.1 Background and History

For over six thousand years humans have been developing ways of navigation to find remote destinations. Leaving stones, marking trees or referencing mountains where among the first concepts before more sophisticated techniques and instruments evolved. During star age, when men started to explore the oceans, visible objects like the Sun, the Moon or stars became "points of reference". When ancient Polynesians developed angular measurements celestial navigation was born. It was around the middle of the last century when scientist discovered radio to measure distance. That included radio beacons, long range radio navigation (LORAN [2]), and OMEGA [5], similar to LORAN on very low frequency (10.2 to 13.4 kHz). In the 1960s the first satellite based navigation system, the U.S. Navy's Navigation Satellite System (known as TRANSIT [17]), opened a new era of navigation technology and capability, before the first global navigation satellite system (GNSS), NAVSTAR Global Positioning System (GPS) was introduced in 1973.

2.2 Satellite Based Localization Methods

Global navigation satellite systems offer a world-wide, highly accurate and practically free-of-charge positioning service. Out of today's two existing GNSS alternatives, GPS is nowadays the only fully operational GNSS. The second alternative, namely the Russian global navigation satellite system (GLONASS), has suffered economical difficulties, which have led to the degradation of the space fleet and, in turn, also to the degradation of the positioning service. Mostly due to this reason, GPS has become the most widely used GNSS [63]. A European equivalent to GPS was decided by the European Commission in 2002 and is currently carried out by the European Space Agency (ESA). First space vehicles (SVs)

were launched in 2005 and the fully operational system will be ready by 2008.

2.2.1 GPS

The concept of GPS is primarily based on measuring the relative time of arrivals of signals which are sent simultaneously from different synchronized satellites (space vehicles or also called "NAVSTARs"). Due to the known position of the satellites, the position of the receiver (e.g. mobile equipment) can be estimated. A very slow data rate of 50 bits per second, however, requires at least 30 seconds to obtain satellite specific data for the exact position calculation, if the available satellites are already known. Furthermore, GPS positioning is completely paralyzed if one of the three required elements, GPS time, navigation data containing satellite orbital parameters or distance measurements are blocked. Unfortunately, these situations occur in areas with high density of expected service users - in urban areas where buildings and other obstacles attenuate GPS signals. Assisted GPS (AGPS) tries to resolve these problems by delivering missing elements (except distance measurements) through cellular networks[1] as shown in Fig. 2.1(a). While satellite orbital parameters can be measured by a reference receiver located somewhere in the network, the exact time is more difficult to achieve. In weak signal conditions, where navigation data is unobservable due to high bit error rate (BER), it is still possible to measure the C/A-code[2] phase. This allows a very accurate identification of the position within the C/A code sequence (period of 1 ms), but an identification of the absolute position is prevented. As a consequence, assistance is needed. There are several solutions existing, but the most straight forward one is to place Location Measurement Units (LMUs) in the network, which deliver the missing rough time information[3] while the accurate one is computed from the C/A-code phase.

In Tab. 2.1 we show a summary of AGPS quality of service parameters. Accuracy in defined as given in [66]: "Accuracy of the geolocation technology is a measure that defines how close the location measurements are to the actual location of the mobile station being located." Reliability is the ratio of successful positioning attempts out of all attempts made, while availability characterizes the conditions where or when the positioning method can be used. This definition inherently includes concepts of coverage and capacity. Latency measures the time passed from the time of power-up to the time the first position estimate is

[1] AGPS is not restricted to cellular networks; Any other wireless technology is also imaginable.

[2] GPS signalling is based an a code division multiple access (CDMA) principle, meaning that all SVs are transmitting at the same carrier frequencies, but separated by orthogonal codes, the C/A-codes for public use and the P-code for military use.

[3] This method, is very cost intensive. For other methods refer to [63].

obtained. Applicability finally summarizes all physical limitations and requirements associated with the implementation and use of a certain technology in terms of technical and financial issues [63]. We note that we give a more precise definition of accuracy in Chapter 6.2. For further information about GPS refer e.g. to [53, 54].

Criteria	Autonomous GPS	Assisted GPS (outdoor)	Assisted GPS (indoor)
Accuracy	<10m	1-10m	<30m
Reliability	high	medium	medium
Availability	global (outdoor)	global (within network coverage)	less than network coverage
Latency	<35s	1-10s	>10s
Applicability	high	medium	medium

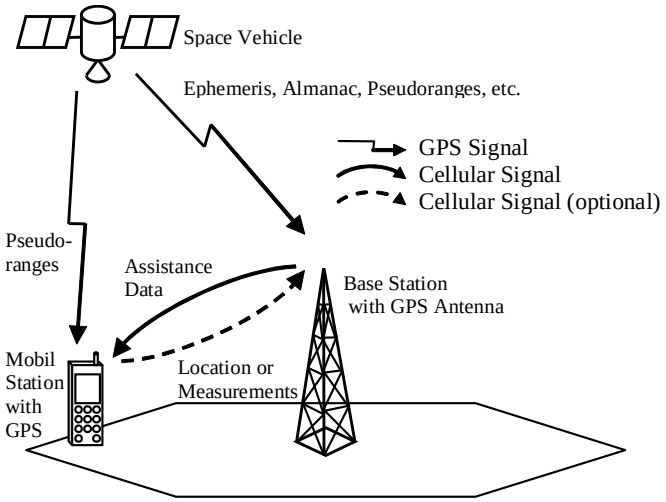
Table 2.1: AGPS Evaluation [63].

2.2.2 Galileo

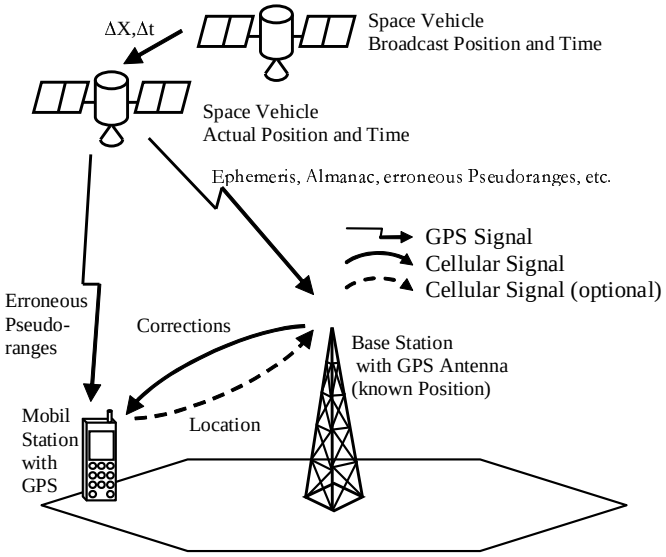
Galileo is the European counterpart to the American GPS satellite system. Completely driven by user needs and operated for civil purposes it is seen as an important step towards a higher number of location based services. Technically Galileo is to a great extend identical to its American counterpart using slightly different carrier frequencies and a minimal higher transmit power of 5 dB. The main distinction to GPS is a better coverage in regions of high latitude (like northern Europe) due to different satellite orbits and a service guarantee, stating the proper reaction in case of interruption or error and the availability of 99.7% [47].

2.2.3 Differential Positioning Techniques

Differential positioning techniques significantly improve both, the accuracy and the integrity of positioning systems. The most common version is Differential GPS (DGPS) as depicted in Fig. 2.1(b). DGPS requires high-quality GPS "reference receivers" at known, surveyed locations. The reference stations estimate the slowly varying error components of each satellite range measurement and form a correction for each GPS satellite in view. This correction is broadcast to all DGPS users on a convenient communications link. Typical ranges for a local area differential GPS station are up to 150km. Within this operating range,



(a) Assisted GPS.



(b) Differential GPS.

Figure 2.1: Assisted GPS as standardized for UMTS and local area differential GPS.

the differential correction improves accuracy to a large extent (e.g. mm range in geodesy) for all users.[4] This improvement arises because the largest GPS errors vary slowly with time and are strongly correlated over distance. By the same token, DGPS significantly improves the "integrity"[5], or truthfulness, of GPS for all classes of users, because it reduces the probability of an unacceptable position error attributable to an undetected system fault [54].

2.2.4 Other Satellite Based Navigation Systems

Beside GPS and Galileo, several other satellite based positioning systems are currently, though not globally, available:

EUTELTRACS is Europe's first Land Mobile Satellite Service similar to America's OMNITRACS system. It is in commercial operation since January 1991. Position determination is obtained by the transmission of a signal to the mobile via two separate satellites in geostationary orbits. The time difference between the two signals is measured by a Mobile Communications Terminal (MCT) and transmitted back to a hub station. During the data exchange the hub station measures the Round Trip Delay (RTD) from the hub to the MCT and back for every MCT within the network. This information is used (in conjunction with an altitude model database) to compute the exact position of the vehicle by multilateration.

OMNITRACS is a two-way satellite data communications system developed and operated by Qualcomm. It was introduced in 1988. The system allows for the receiving and sending of position reports in remote areas with satellite coverage, either GPS or Qualcomms proprietary positioning system Qualcomm Automatic Satellite Position Reporting (QASPR). QASPR is widely identical with EUTELTRACS's positioning method.

BEIDOU or BNS is China's satellite based positioning system and in operation since the end of 2000. This system uses two geostationary satellites and allows 3D positioning in China. Since 2003 China is considering an enhancement of the system with additional satellites in order to achieve complete independence of other positioning systems.

[4]This was also true before the U.S. Department of Defense (DOD) turned down selective availability (SA) in May 2000.

[5]Integrity is the probability that the displayed position is within the specified or expected error boundaries.

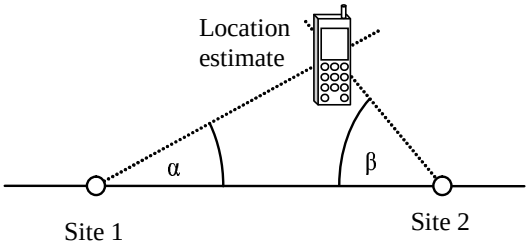


Figure 2.2: Location estimation using angle of arrival.

2.3 Terrestrial Radio Based Positioning Methods

Terrestrial radio based positioning methods typically use a network set-up consisting of fixed base stations and mobile transceivers, with one emitting the radio signals and the other one receiving them, in order to determine the position. (The role, who is the transmitter and who is the receiver and who actually calculates the position is in general the manufacturer's choice and varies in many cases.) For performance improvement various combinations of the techniques discussed in the following sections - hybrid solutions - are possible as well. Commonly studied techniques can be further classified into direction finding, range based, and signal based positioning systems.

2.3.1 Angle of Arrival

Direction finding systems identify the position of the mobile equipment, by measuring the Angle of Arrival (AOA) of the mobile's signal using antenna arrays[6]. The intersection of two directional lines define the unique position of the source which has to be located. For this method a minimum of two georeferenced base stations are required for a 2D localization (refer to Fig. 2.2).

Three major sources of error can be distinguished [73]:

- The bias in the antenna placement, which can be mitigated by calibration and the maintenance of the antenna systems.
- Noise background in the angular measurement due to multipath signals.
- Non line-of-sight measurement.

[6]In literature this method is sometimes also referred to as Direction of Arrival (DOA) method.

While the first source of error can be omitted by carefully placing the antenna, the remaining sources of error require a high number of antenna elements to resolve a large number of different propagation paths. Accurate angle estimation becomes therefore challenging and cost intensive. For methods to improve the positioning accuracy see e.g. [73].

2.3.2 Time Based Positioning Systems

Time based positioning systems can be classified as ranging, rang sum or range difference positioning systems [50]. In all cases a unique geometry configuration is defined, being the base for the calculation of the position estimate. Time based positioning systems measure the distance between the mobile equipment and a set of base stations by time-of-arrival (TOA) measurements. Since the propagation time of the radio wave is directly proportional to the traversed range, a multiplication with the speed of light c results in the distance between the base station and the mobile equipment. For a single path additive white Gaussian noise (AWGN) channel [55] and [11] show that the best achievable accuracy of a distance estimate $\hat{d}$ derived from TOA estimation satisfies the following inequality:

$$\sqrt{var(\hat{d})} \geq \frac{c}{2\sqrt{2}\pi\sqrt{SNR}\beta}, \tag{2.1}$$

where SNR denotes the signal-to-noise ratio and, β is the effective (or root mean square) signal bandwidth defined by

$$\beta = \sqrt{\left[\frac{\int_{-\infty}^{\infty} f^2|S(f)|^2 df}{\int_{-\infty}^{\infty}|S(f)|^2 df}\right]} \tag{2.2}$$

with $S(f)$ being the Fourier transform of the transmitted signal. Equation 2.2 shows that by increasing the SNR or the effective signal bandwidth, the accuracy of time based positioning can be improved. Ultra wide band (UWB) signals have, for instance, very large bandwidths and as a result extremely accurate location estimates are possible. An UWB pulse of $1.5GHz$ bandwidth at SNR = 0 dB allows an theoretical accuracy of less than 30 millimeters.

Range Based Positioning

Ranging positioning systems measure the absolute distance between the mobile equipment and a set of base stations by time-of-arrival measurements. The intersection of the thereby defined circles indicate the position estimate. This principle is also useable for 3D positioning, where the circles become spheres.

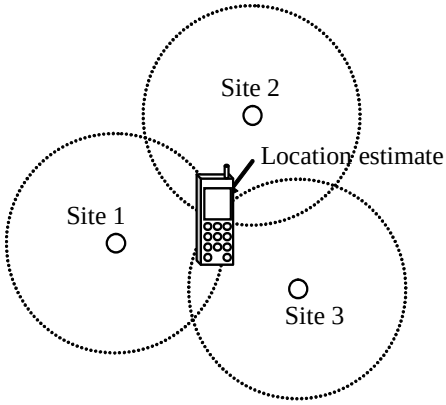


Figure 2.3: Location determination using time of arrival (TOA).

If the circles intersect in more than one single point (e.g. because of non-line-of sight conditions), an ambiguous solution to the position estimate results. This can be reduced by redundant range measurements and the application of e.g. the least squares (LS) technique.

To illustrate this method consider a 3D range positioning system using 4 base stations. The TOA is measured at each receiver and the ranges are

$$R_i = c * d_i = \sqrt{(X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2} \quad ; \quad i = 1 \ldots 4, \qquad (2.3)$$

where c is the propagation speed of the signal, X_i, Y_i, Z_i are the co-ordinates of the base stations, x, y, z are the co-ordinates of the source to be estimated and d_i is the TOA of the i-th receiver.

Due to the overdetermination of the system, LS techniques are used to calculate the position estimate.

Strict clock synchronization between the mobile unit and the base stations is required. This can be accomplished, for instance, by the use of stable clocks, such as rubidium or cesium standard clocks, at both the mobile equipment and the base stations. This, however, would require additional hardware in the mobile units, resulting in additional size, power consumption, weight and costs.

Another disadvantage is the sensitivity for system geometry. Highest accuracy is only attainted when all ranging circles intersect at 90 degrees, which is, due to moving mobile stations, not very likely.

Elliptical Positioning Systems

The mobile equipment is located at the intersection of ellipsoids with $R_i + R_j = const = rangesum$ measured between source and base station i and j ($i = 1$ and $j = 2$ in the 2D example shown in Fig. 2.4). The range sum is determined from

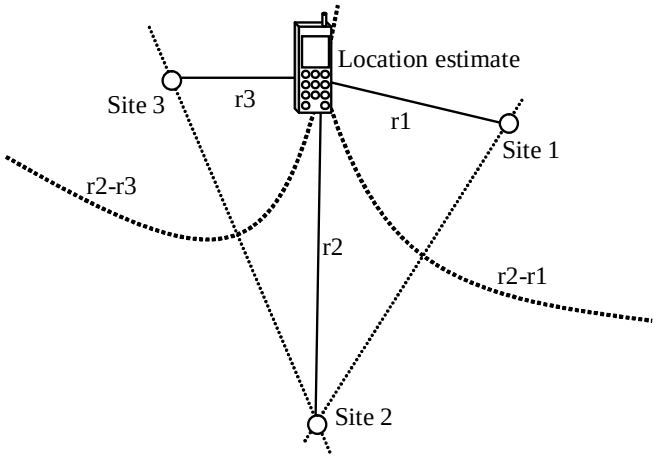


Figure 2.4: 2D elliptical positioning system.

the sum of the signal's TOAs at multiple receivers. The relationship between the range sum $R_{i,j}$ and the TOAs is given by

$$R_{i,j} = cd_{i,j} = R_i + R_j, \tag{2.4}$$

where $d_{i,j}$ is the sum of the TOAs from receiver i and receiver j, and c is the propagation speed of the signal.

The geometrical relation defines the equation system to calculate the location estimate. Referring to the 2D example in Fig. 2.4 we find non-linear equations

$$\begin{aligned} R_{1,2} &= \sqrt{(X - X_1)^2 + (Y - Y_1)^2} + \sqrt{(X - X_2)^2 + (Y_2 - Y)^2} \\ R_{1,3} &= \sqrt{(X - X_1)^2 + (Y - Y_1)^2} + \sqrt{(X_3 - X)^2 + (Y - Y_3)^2}, \end{aligned} \tag{2.5}$$

which can be solved by LS techniques. Other eligible solutions are presented in [10] and [21].

Hyperbolic Positioning Systems

The mobile equipment is located based on trilateration. Time difference measurements between two base stations are interpreted as constant distances to define a hyperbolic curve. In the 2D case, the intersection of at least two hyperbolas determines the estimated position of the mobile equipment. The range difference $R_{i,j}$ between base station i and base station j is given by

$$R_{i,j} = cd_{i,j} = R_i - R_j, \tag{2.6}$$

where c expresses the propagation speed of the signal and $d_{i,j}$ is the TDOA between base station i and base station j.

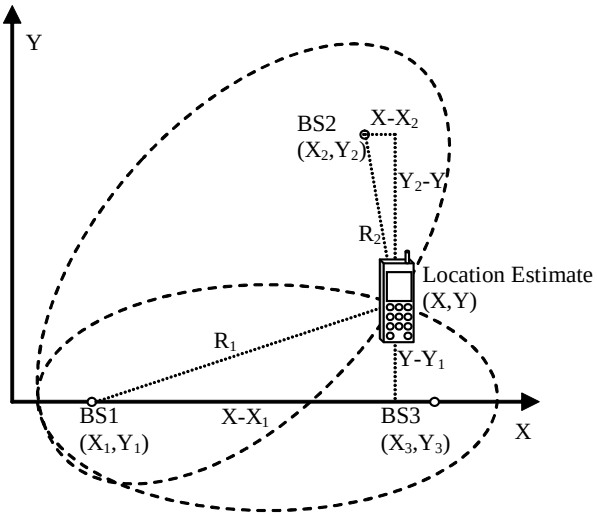


Figure 2.5: 2D hyperbolic positioning system.

In the 3D case at least 3 pairs of base stations (hence, 4 independent base stations) have to be available for positioning. The hyperbolas are given by

$$R_{i,j} = \sqrt{(X_i - x)^2 + (Y_i - y)^2 + (Z_i - z)^2} - \sqrt{(X_j - x)^2 + (Y_j - y)^2 + (Z_j - z)^2}, \tag{2.7}$$

where (X_i, Y_i, Z_i) and (X_j, Y_j, Z_j) define the base stations i and j, and (x, y, z) are the unknown coordinates of the mobile equipment. No strict clock synchronization between the mobile equipment and the base station is required. Any timing error or any common error experienced at all base stations caused by the channel can be reduced due to the difference of the range measurements. The network, however, still has to be synchronized, which can be done by additional LMUs.

Time difference of arrival methods have been and are still being standardized for the Universal Mobile Telecommunications System UMTS by the 3^{rd} Generation Partnership Project Agreement (3GPP). The main representatives are the enhanced observed time difference method (EOTD) and the advanced forward link trilateration method (AFLT).

EOTD: EOTD standardization has been finalized by the European Telecommunications Standards Institute (ETSI) already in R98 and R99. For UMTS the standardization is still in progress. Three basic timing quantities are important for the localization of the mobile equipment:

1. Observed Time Difference (OTD): This is the time which is measured by the mobile station between the reception of a burst from a base station 1

(say at time t_1) and the reception of a burst from a base station 2 (say at time t_2). OTD is then given by

$$OTD = t_2 - t_1. \tag{2.8}$$

OTD is zero if the two bursts arrive at exactly the same time. (Note: This means that the mobile station is in equal distance to each of the two base stations, only if the base stations are synchronized.)

2. Real Time Difference (RTD): Is the time emerging when the base stations in the network are not synchronized. Say t_3 is the moment when base station 1 sends a burst and t_4 the time when base station 2 sends a burst. The RTD is then calculated by

$$RTD = t_4 - t_3. \tag{2.9}$$

Only if perfect synchronization is available in the network RTD equals zero.

3. Geometric Time Difference (GTD): Is the time difference defined by the geometry, and the propagation speed of the signals:

$$GTD = (d_2 - d_1)/c, \tag{2.10}$$

where d_2 and d_1 are the lengths of the propagation paths between the base stations and the mobile station and c is the propagation speed. GTD is zero, if the distances from the base stations to the mobile station are the same.

Important for the final calculation, by intersecting the hyperbolas, is only GTD. Since the mobile equipment knows the OTD, and RTD can be measured by an additional LMU in the network, GTD can be calculated according to

$$GTD = OTD - RTD. \tag{2.11}$$

Advanced Forward Link Trilateration (AFLT): This method has been standardized in the U.S. for IS-95. Because IS-95 is a time synchronized system, time difference measurement are easier to accomplish. The basic idea is to compare the pilot signal of the serving cell with the pilot signals of the neighboring cells. The time difference is then converted into range information. Since this method is, first, not really different to the above described TDOA method and, second, not standardized in UMTS, it will not be further described in this thesis.

2.3.3 Main Sources of Error in Time Based Positioning

Multipath Propagation

An optimal estimate of the arrival time is obtained by using matched filters or a bank of correlation receivers [67], which is mathematically equivalent and, hence, usually a choice of design and implementation costs. Both methods produce a peak in the output signal which provides the arrival time estimate. In narrowband systems, however, multiple replicas of the transmitted signal, due to multipath propagation, overlap and thereby shift the position of the correlation peak. The multipath channel creates a mismatch between the received signal of interest and the template the receiver is using. Instead of auto-correlation, one obtains cross-correlation which does not necessarily result in a peak at the correct timing. High resolution delay estimation techniques, such as in [52], have been proposed to resolve this effect, but are highly complex. Less complex algorithms to detect the first arriving signal path are e.g. discussed in [44].

Non Line of Sight Propagation

If the direct path between the transmitter and the receiver is obstructed, only reflections of the transmitted signal from scatterers reach the receiver. The delay of the first arriving peak does not correspond with the true TOA and thus neither with the distance. The resulting additive distance is called non-line-of-sight (NLOS) error. In such cases, LS techniques are inaccurate and without further information this problem is not resolvable. M. Wylie and J. Holtzman [72] observed that the variance of the TOA measurements in the NLOS case is usually much larger than the one of the line-of-sight (LOS) case. They observed that the information contained in the time history of the range measurements can be used to distinguish between LOS and NLOS situations via a simple hypothesis test. They also show that after the recognition of an error an correction is possible, using prior statistical knowledge of the characteristics of the system's standard measurement noise.

In case of tracking a mobile device Kalman filters [23,38,43] or particle filters [18] can be applied for an even more precise positioning.

Synchronization and Power Control

Due to the unsynchronized nature of most mobile networks, the accurate measurement of the TDOA is problematic. A 1μs time error can lead to a 300m positioning error. Precise correction time must thus be provided by additional network components, LMUs as described in the previous section.

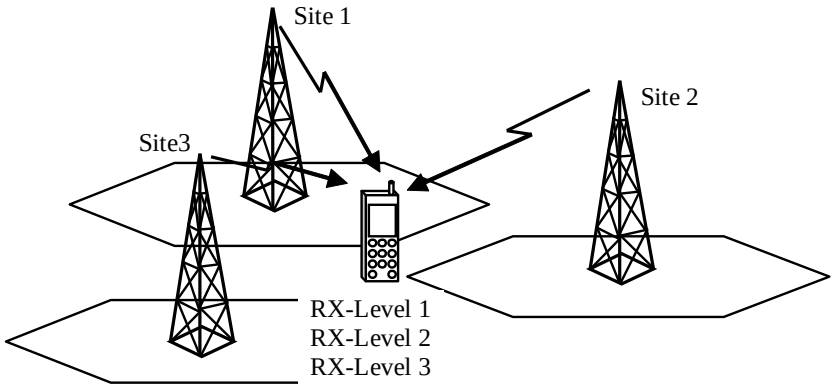


Figure 2.6: Received power level based positioning.

Another main challenge is the power control, which avoids unnecessary interference between different mobile stations. But time based positioning can only deliver results, when at least three BS are "hearable" (if best with good geometry). One solution proposed is called Idle Period Downlink (IPDL). The transmission of the serving cell is paused for a very short time, so that mobile equipment next to the serving cell can listen to the neighboring base stations.

2.4 Signal Strength Based Positioning

Signal strength (SS) based positioning (Fig. 2.6) relies on the measured energy of the received signal at the mobile equipment. Applying path loss models [60], the received signal strength corresponds to the distance between the mobile station and the base station, allowing the utilization of the same triangulation techniques as described earlier for TOA based methods. To determine the distance from SS measurements accurately, however, the characteristics of the channel must be known. This makes these methods very sensitive to environmental changes. The lack of precise models for the signal strength attenuation considering all environment obstacles like houses, vegetation, hills etc. prevents high accuracy. Results can be expected within a few hundred meters. Controversially, Rx-levels are inherent measurements (e.g. for handover reasons) in cellular systems and are therefore available with great probability also within the core network.

2.5 Pattern Matching Methods

Pattern matching methods utilize distinct radio frequency (RF) patterns (multipath phase and amplitude characteristics) of the radio signals arriving at a receiver antenna from a single caller. The unique characteristic of the signal,

including its multipath pattern is analyzed and a "fingerprint" is determined for a defined area. The fingerprint is then compared with a database of previously "fingerprinted" locations, and a match is made. The metrics which are applied range from taking the serving antenna's position directly as location estimate, to path loss dependent combinations of the position from the serving and neighboring antennas [75]. These methods require only the position of the base stations to be stored in the database. The achieved accuracy is about 300 meters in 95% of all measured cases. Other authors [18–20, 41] use correlation algorithms to find the best match between the measured fingerprint and the fingerprints in the database. Widely known is VTTs Database Correlation Method (DCM) [41] showing an accuracy of about $90m$ in 90% of all measured case for urban environment. To omit the necessity of a database with pre-measured values, in [74] the COST 231 Walfisch/Ikegami propagation model is used to predict the fingerprints. The accuracies achieved are in the range between 190 and 280 meters in 95% of all cases.

2.6 Short Comparison of Positioning Methods

All methods discussed in the previous sections show advantages and disadvantages. No single method outperforms its competitors. Satellite based localization methods (e.g. GPS, AGPS, DGPS, Galileo) have a very wide availability with very good accuracy, except for street canyons, indoor environment or underground areas. TOA based methods show their strength in indoor areas (e.g. UWB) or in underground areas and in outdoor environment if multipath propagation and NLOS is unlikely. Pattern matching methods, achieve good results in areas where multipath propagation is likely because different sites show different received power levels. But they require static environment. Figure 2.7 qualitatively compares the different methods in terms of availability and accuracy. Published accuracy results are given in Tab. 2.2. The results from the two suburban EOTD trials differ because of their strong NLOS dependence. The pattern matching methods (DCM and Wireless Location Signatures WLS) perform best in urban environment and show less accurate results in suburban areas. The Cell-ID method, which is currently widely used in Europe, shows the worst results, but does not require any additional hardware.

Table 2.3 compares different positioning technologies in terms of reliability, latency and applicability.

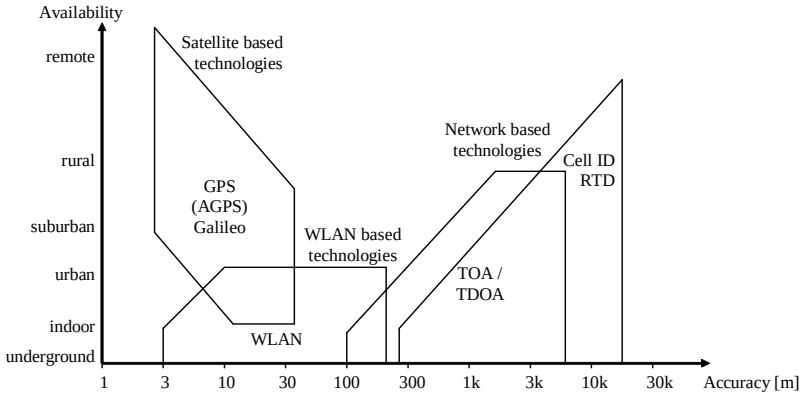


Figure 2.7: Qualitative comparison of satellite-, network- and WLAN-based positioning technologies based on [63].

Urban Environment			Suburban Environment		
Trial	67%	90%	Trial	67%	90%
Cell-ID [40]	328m	519m	Cell-ID [40]	639m	1085m
Cell-ID + TA [40]	283m	475m	Cell-ID + TA [40]	415m	666m
DCM [41]	44m	90m	DCM [41]	74m	190m
WLS [20]	44m	90m	WLS [19]	81m	220m
			AOA [62]	45m	89m
			OTD [26]	167m	323m
E-OTD [42]	125m		E-OTD [42]	75m	

Table 2.2: Accuracy of different positioning technologies.

Type	Positioning Technology	Reliability	Latency	Applicability
Satellite based	GPS	High	< 35s	High
	AGPS	Medium	1-10s	Medium
Cellular Network based	AOA	Medium	≈ 10s	Low
	Signal Strength	High	< 5s	High
	Pattern Matching	Low	< 10s	Low
	Cell-ID + TA	High	< 5s	High
	TOA/TDOA	Medium	< 10s	Low
	EOTD/OTDOA-IPDL	Medium	< 10s	Medium
WLAN based	Zone Positioning	High	< 5s	Low
	TDOA	High	< 5s	Low

Table 2.3: Reliability, latency and applicability of satellite-, network- and WLAN-based positioning technologies based on [63].

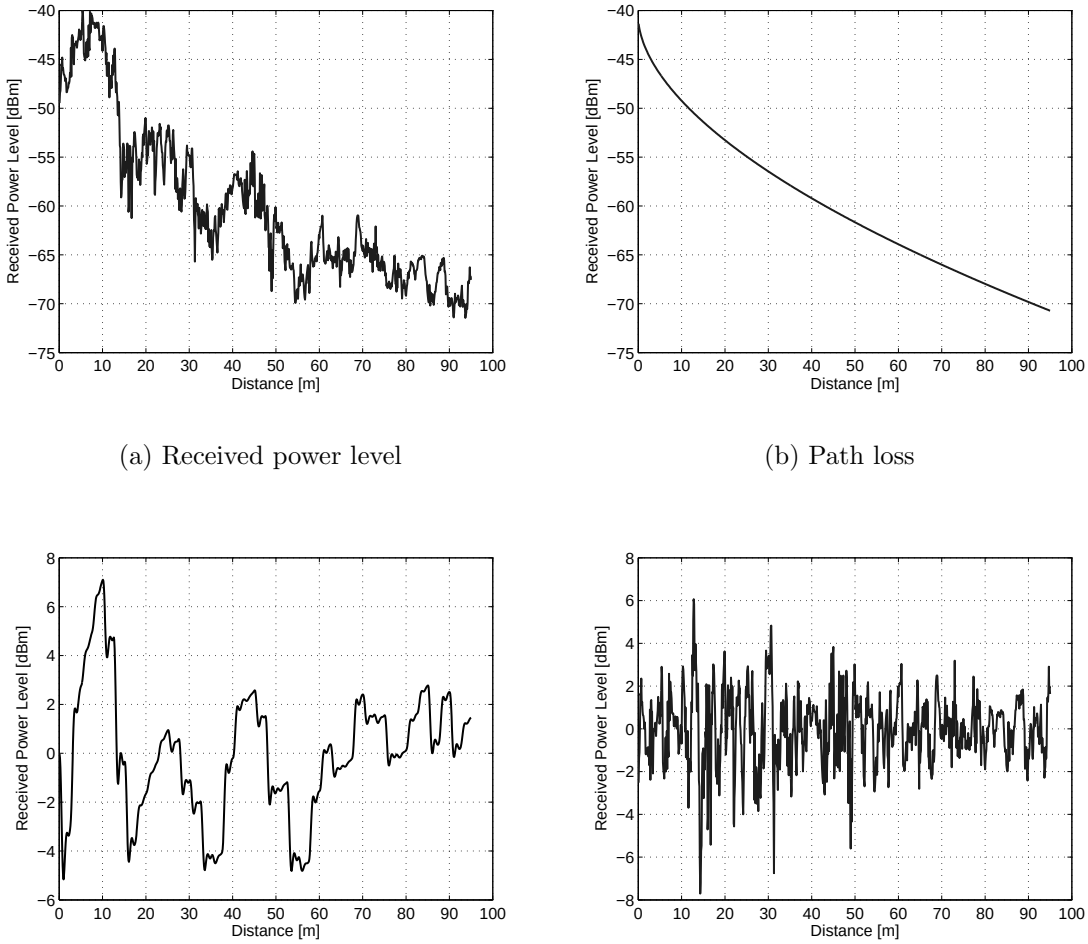
Chapter 3

NMR based Positioning

This chapter gives an introductory overview of the pattern matching positioning method developed in this thesis. We begin with an review of the basic RF signal propagation phenomena being responsible for the location dependence of the RF signals. We briefly explain the computational effort to simulate accurate RF signals and discuss the inverse problem of deriving the position from the signal. Impractical computational effort and the lack of accurate environmental knowledge will then motivate us to develop flexible models to characterize a position, containing only as much "reasonable" information as necessary.

3.1 Real Positioning Environment

For all wireless communication, the mobile communication channel is the transmission medium. The understanding of the wireless channel is essential for the analysis of any wireless system, in particular if we want to derive the position from the signal. Theoretically, the channel is determined exactly. The existence of propagating electromagnetic waves can be predicted as direct consequence of Maxwell's equations [49] and it should be possible to compute the electric field $\boldsymbol{E}$ at any position at any time. This, however, is theory. Practically, limitations in computational power and especially the lack of accurate side conditions prevent exact results. Most of the time we will miss knowledge about the environment, about obstacles like cars, buildings, etc. as well as their material coefficients with sufficient accuracy. More practically we have to approximate the electric field strength by applying the basic physical phenomena. Three fundamental phenomena (Fig. 3.1) can be identified that are key to develop a reliable localization method.



(a) Received power level

(b) Path loss

(c) Shadowing

(d) Fast scale fading

Figure 3.1: Attenuation effects: 3.1(a) shows the received power level measured at different distances to the transmitter. It is the superposition of path loss 3.1(b), shadowing 3.1(c) and fast scale fading 3.1(d).

3.1.1 Path Loss

The path loss is an overall decrease in field strength, as the distance between the transmitter and the receiver increases. This phenomenon is a physical process, caused by the spreading of waves in free space and the obstructing effects of buildings, trees or other obstacles. The following three propagation mechanisms are responsible:

Attenuation

In free space, an (hypothetical) isotropic antenna radiates the power uniformly and without loss in all directions. As the propagation distance increases, the radiated energy is spread over the surface of a sphere of radius r according to

$$S = \frac{P_t}{4\pi r^2}, \tag{3.1}$$

where P_t is the transmission power. The power received at the receiver decreases proportional to r^2. Note that this is only valid for the far-field where the propagation distances r is much larger than two times the square of the antenna size divided by the wavelength (rayleigh distance). Here, the wavefronts appear very closely as spherical waves and only the power radiated in radial direction is of importance. For localization purpose the far-field approximation is valid, since the rayleigh distance for typical base stations is between several meters (macro cells) to sub-meters (micro and pico cells).

Reflection and Transmission

If an electromagnetic wave incidences onto a planar boundary between two different media, parts of the power is reflected and parts is transmitted. If Maxwell's equations are solved, new waves with the same frequency are produced, one propagating away from the boundary, obeying Snell's law of reflection, the other travelling into the new medium behind the boundary according to Snell's law of refraction[1]. If the surface of the boundary is made progressively rougher, the reflected wave becomes scattered. This reduces the energy in specular direction and increases the energy radiated in other directions. For the computation two different theories are applied: 1) the Kirchhoff-Theory, assuming knowledge about the probability density function of the surface height and 2) perturbation theory using additional information about the correlation of the surface height. While for the localization method, introduced in this thesis, reflection is one of

[1]Note that in case of lossy media, Snell's law of refraction is no longer valid. For mobile communications, however, adapted reflection and transmission coefficients can be used as approximation. See [7] for more details.

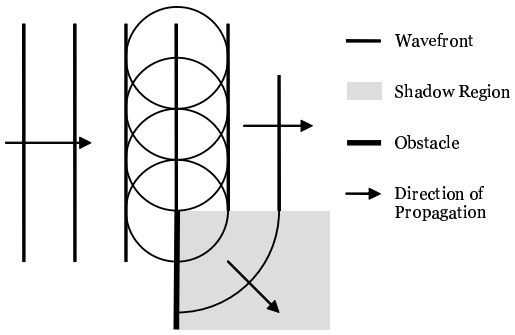


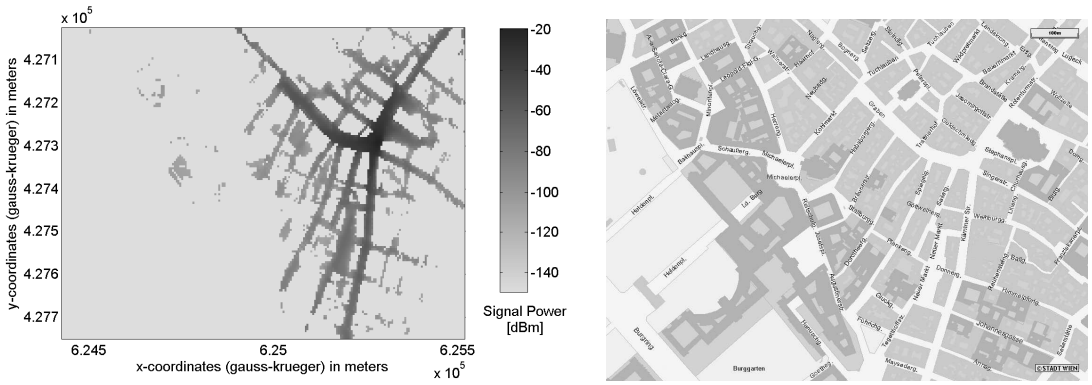
Figure 3.2: Huygens' principle for knife-edge diffraction (source: [58]).

the important effects, the computation is of less importance, since we will rely on measurements as we will describe in Chapter 5.

Diffraction

Diffraction is responsible for electromagnetic fields in shadow regions behind obstructions. While geometric optic simply denies the occurrence of waves in the shadow region, this is incorrect in reality, since shadows are never sharp and some energy does propagates in the shadow region. This fact is important for mobile communications, in particular for predictions in sub-urban, urban and indoor areas. Diffraction is explained by *Huygens'* principle: Each single point of a wavefront can be regarded as the center of a spherical wavelet. The superposition of all spherical wavelets results in the actual wavefront. This yields for a plane wavefront, impinging on a wide screen not allowing energy to pass through (single knife-edge), waves to propagate in the shadow region. For deep inside the geometrical shadow region, rays appear to emerge from a point close the the edge as shown in Fig. 3.2. To compute, however, the propagation of waves over rooftops, the single knife-edge has to be generalized to several knife-edges. This is computational extensive but several approximations with advantages and disadvantages exist. For more information refer to [16].

By combining these three basic propagation effects we receive a path loss as shown in Fig. 3.1(a). In direct vicinity of the transmit antenna the received power decreases proportional to r^2. Attenuation (line of sight is assumed) is dominating the path loss. For positions in greater distance to the transmit antenna, the power decreases with higher degree. In these areas the path loss is found empirically to be proportional to r^n with $1.5 < n < 5.5$. The actual value of n is strongly dependent on the environment. To achieve reasonable positioning accuracy we thus have to look for very accurate models, requiring good knowledge about the environment (geometric data, surface data) and high computational effort. Such



(a) Predicted received power levels for urban environment

(b) Geometry of the environment

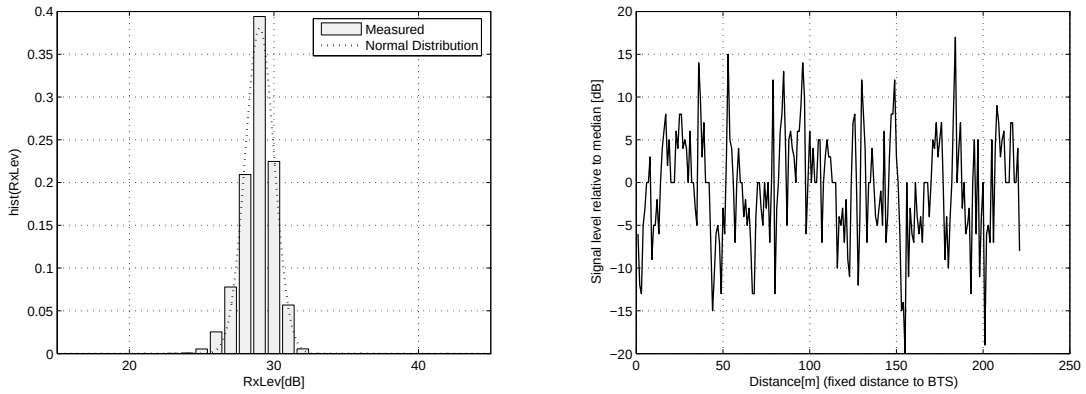
Figure 3.3: Ray based prediction of received power levels for one base station in urban environment. The discrete simulation grid is 5 meters.

accuracy is only achievable by geometry based models, like ray models. Here, the prediction of the received power is based on the theory of diffraction (and extensions) rather than by statistical means. The lack of environmental data and the high computational effort, however, limits their practicability. An example for the prediction of the received power levels in the city of Vienna is depicted in Fig. 3.3.

3.1.2 Large Scale Fading

Large scale fading (slow fading, shadowing) is the variation of the field strength caused by different obstruction for different propagation paths. When moving the receiver at a given constant distance to the transmitter, the changing clutter (buildings, trees, etc.) along the path will cause variations in the field strength as depicted in Fig. 3.4(b). These variations around the 50% percentile are empirically found normal distributed on a logarithmical scale (Fig. 3.4(a)) and are called location variation.

Theoretically this normal distribution is explained by the central limit theorem. It is assumed that between the transmitter and the receiver several independent reflection-, scattering- or diffraction-processes are responsible for the attenuation. Because of the central limit theorem and the independence of all individual losses, the attenuation in dB is normal distributed. This explanation, however, shows limitations when only a low number of (or even one) reflection or diffraction process is located between the transmitter and the receiver.



(a) Probability mass function of received power levels

(b) Variation around the median of the received power levels

Figure 3.4: Probability mass function and the variation around the median of the received power levels at an suburban site near Vienna. The mobile station was moved along a path with fixed distance to the base station.

3.1.3 Correlated Shadowing

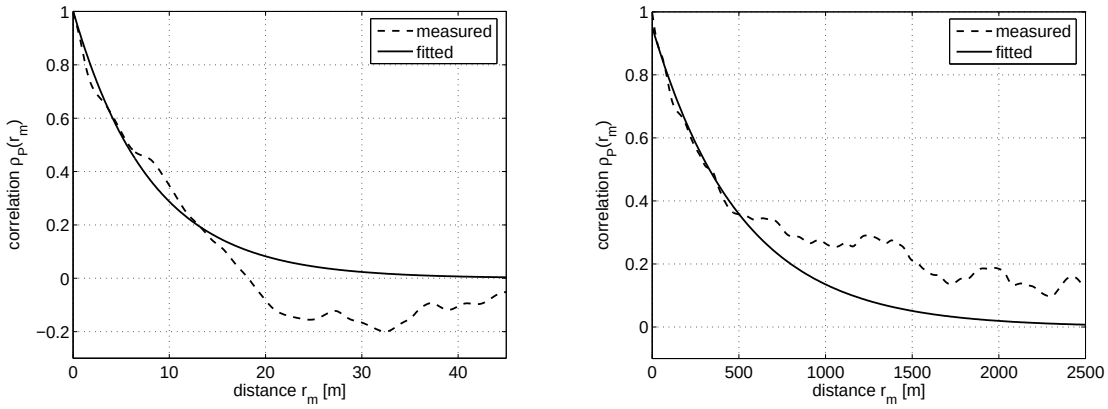
So far, we have assumed each propagation path from the base station to the mobile station to be independent. Positions, however, separated only by a small distance, might share several obstructions in their path profiles. This results in a correlation of the received signal strength between adjacent positions. Measurements [25, 48] of the correlation between received power levels at different positions suggest a first-order exponential correlation function

$$\rho_P(r_m) = \sigma^2 e^{-\alpha r_m}, \tag{3.2}$$

where σ denotes the location variability. The variable r_m denotes the distances between the positions i (shadowing correlation distance). The coefficient alpha was estimated from data to be between $\alpha = -\frac{1}{500}$ for suburban areas and $\alpha = -\frac{1}{8}$ for urban areas.

In [25] typical values for r_m range between a few tens (urban) to several hundreds (suburban to rural) meters for $\rho_P = 0.37$.

While this correlation effect limits the effort of power control - if the shadowing correlation decreases very rapidly, the estimate of the received power of the base station will be inaccurate by the time the (moving) mobile reacts on the command - localization methods can benefit. Due the low correlation, adjacent positions will carry different received power level patterns and will, hence, be easier to distinguish based on these patterns.



(a) Urban environment

(b) Suburban environment

Figure 3.5: Normalized autocorrelation (measured and fitted) in different environments (Source: [25]).

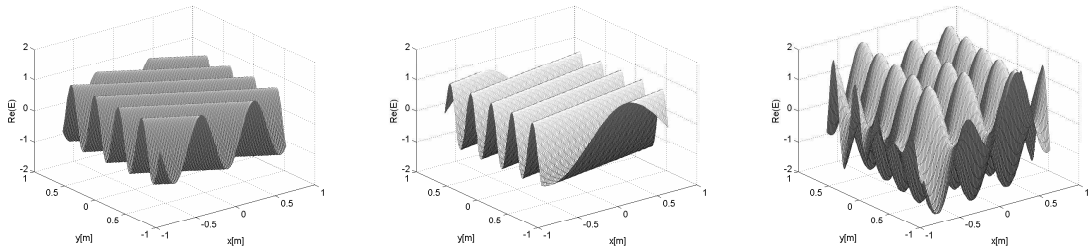
In Fig. 3.5 the normalized autocorrelation (measured and fitted) is shown for different environments. In urban environment a low correlation occurs with r_m being between several meters to tens of meters. In suburban environment the same correlation is shown for distances of over 1000 meters. This already indicates that pattern matching methods are preferably developed for urban areas.

3.1.4 Small Scale Fading

After a careful prediction of path loss and shadowing there are still significant variations in the received signal as the mobile moves over short distances up to one wavelength. This phenomenon is called small scale fading (fast fading) with variations so rapid that only statistical description is useful. Imaging a transmitter being surrounded by different obstacles causing reflection- and diffraction-effects. At the receiver, located in the far-field of these scatterers, plane waves arrive and interfere. The result is a location dependent fading as shown in Fig. 3.6 for the simplest case of only two waves[2].

In general, a higher number of paths and the lack of very exact knowledge of the positions and the electromagnetic characteristics of all scatterers prevent an accurate modelling of the field strength. Hence, the field strength does not seem to be a good indicator for the location. Its missing reversibility does not allow a unique prediction. In real world situations however, static environment is seldom. Scatterers will move and the paths will change, even for fixed transmitters and

[2]A constant amplitude for both waves was assumed due to the close vicinity to the receiver.



(a) Wave 1 (b) Wave 2 (c) Superposition of wave 1 and wave 2

Figure 3.6: Interference between two plane waves with amplitude $A_1 = A_2 = 1$ and $\angle(k_1, k_2) = 40$.

receivers. It can be shown that in cases of no dominant path (NLOS condition), the small scale fading is normal distributed with zero mean. The mean signal power at the receiver is then

$$P_{NLOS} = \sigma^2. \tag{3.3}$$

In case of a dominant path (LOS condition) the signal is composed of a random multipath component plus a coherent line-of-sight component essentially with constant power[3]. The mean signal power at the receiver is then

$$P_{LOS} = \sigma^2 + \frac{A^2}{2} \tag{3.4}$$

with A being the amplitude of the dominating component. For positioning purposes it is thus useful for both situations, LOS and NLOS, to average over time to reduce the small scale fading effect[4].

3.2 Positioning with Bayesian Networks

We have seen in the last section that path loss, shadowing and scattering effects cause considerable variations of the signal strength measured by the mobile. These variations of the signal strength can be used as a characteristic feature for a reliable recognition of the location within the mobile environment.

[3]Path loss and shadowing are not considered here.

[4]This is actually automatically performed within the mobile.

3.2.1 Pattern Matching

The task of pattern matching is a generic operation in pattern recognition and is used to determine similarity between two entities. In general, there are several different approaches; two of them can be adapted for the localization problem. *Template matching* denotes a method where the pattern to be recognized is matched against a stored template. The *statistical approach*, represents each pattern in terms of d-features and the pattern is viewed as a point in a d-dimensional space. The goal is to choose those features, which allow patterns belonging to different categories, to occupy compact and disjoint regions in the d-dimensional feature space [29]. The effectiveness of the feature set is determined by how well patterns belong to different classes. For the NMR based localization the received power levels of the different base stations themselves are used as features. The localization system combines the pattern matching and the statistical approach in order to find the position of a mobile. First, for each position a template is constructed represented by a Bayesian network. Second, the statistical approach is used to find the probability of belonging to a certain class of patterns. We only assume two classes, either "pattern of this position" or "no pattern of this position" and can then identify the most likely Bayesian network. Since the Bayesian network is directly linked to a certain position, the localization is finished.

3.2.2 Features Selection

Statistical pattern recognition has been applied in various fields during the last decades. However, the best pattern recognizers only perform well, if the pattern contains information which can be classified well. The so called features, should thus, first, be easy to access, second, be low of dimension for computational reasons and, third be easy to distinguish to allow easy classification. However, a reduction in the number of features may lead to a loss in the discrimination power and thereby may lower the accuracy of the resulting recognition system. On the contrary, a reduction can as well increase the discriminating power by omitting ambiguous characteristics. An illustrative example supporting the need for a careful choice of the features is given with Watanabe's "ugly duckling theorem" [69]:

Suppose there are only three things in the world, two 1's and one 0. One is usually tempted to say that there are two "kinds" of things, the two 1's and the one 0. This, however, is not all. If we line the digits up, the position is another property we have not mentioned so far. The two 1's are now different in the spatial property. Hence, we cannot say that some things are more similar than others, unless we specify some properties as representatives.

We have already mentioned before that we use the received power levels as

features characterizing a position. This characterization will be ambiguous, since in variegated and heterogeneous environments, the received power is a very complex function of the distance (see Section 3.1) and will result in equal values at multiple positions due to propagation effects. In particular the complex urban geometry with lots of scatters will produce ambiguity. Controversially, the high number of scatters and the dense number of base stations will produce recognizable shadowing, a low shadowing correlation distance and hence, will allow a better distinction between adjacent positions.

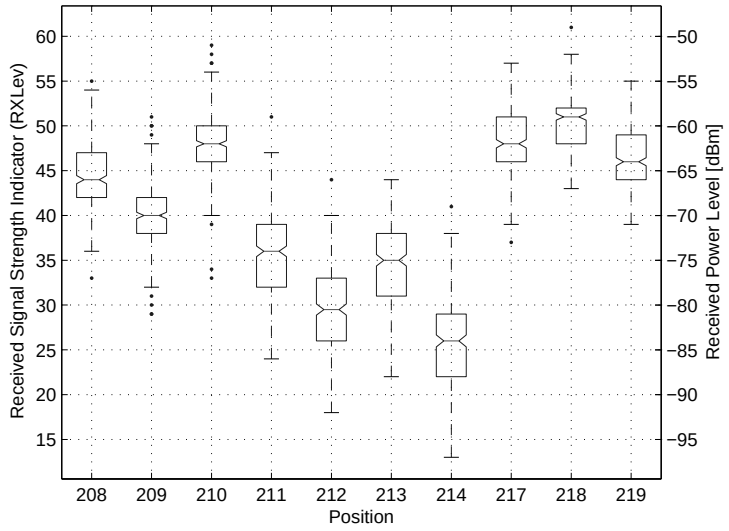
Figure 3.7(a) shows an example of the received power levels of the serving cell for adjacent positions (5 meter spacing). The end of the whiskers indicate the minimum and maximum data values, unless outliers (denoted by dots) are present in which case the whiskers extend to a maximum of 1 times the interquartile range. The median is shown as a horizontal line between the notches. Figure 3.7(b) shows the corresponding map. The positions are denoted by dots. In the area of the intersection (Positions 211 to 214) the power level decreases because of shadowing and allows a spacial distinction. On the contrary, positions 217, 218 and 219, located in a street canyon, show an overlapping plot, even though the medians differ with 95% confidence indicated by not overlapping notches. Outside or inside the street canyon - which is the general case - the difference of the received power levels will be too small for a reliable distinction based on simple comparison. Defining the received power levels to be the only feature will thus limit the accuracy of the system significantly.

More important are the dependencies between the different measured cell IDs (serving cell and neighboring cells) for every position. Because of the geometry of the environment a set of different base stations will be received with different power levels at the mobile station. Ordered according to their power levels with the serving cell always first, a location dependent pattern is defined.

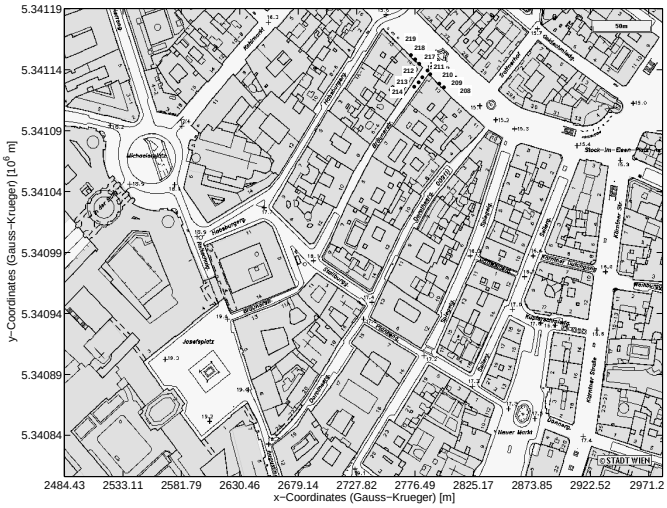
This has two additional advantages: First, different receivers and antennas would not allow a direct comparison of the absolute power levels. The ordering of the received cell IDs according to their power levels, however, is robust against different receivers and antennas. Second, the effect of power control is also compensated. Power control prevents the usage of the absolute power level of the serving cell without additional knowledge of the network[5].

During an active call, the power control limits the transmit power of the base station to reduce interference, as illustrated in Fig. 3.8. At $t = 0$ a call is established and the received power level of the serving cell (BSIC = 42) equals 61 dB. At time $t > 0$ the power level decreases. Hence, different power levels are measured at the same position, depending on the active call duration. An

[5]RF power control in the BSS is optional in GSM. The broadcast channel is never power controlled. In UMTS power control is mandatory.



(a) Boxplot of received power levels



(b) Corresponding map. The positions shown in the boxplot are denoted by dots.

Figure 3.7: Received power levels of the serving cell at several closely co-located positions.

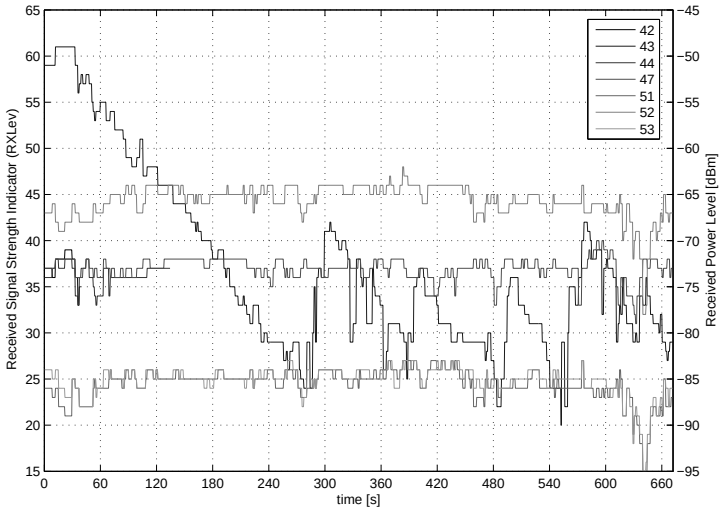


Figure 3.8: Time variation of the received power levels received from different cells and measured at a single site.

estimation of the position based on the absolute power level is thus impossible, if the power control cannot be compensated.

3.2.3 Decision Making

To model the complex dependencies between the measured cell IDs, we use probabilistic models with sufficient large numbers of parameters and automated learning techniques. Bayesian networks are an effective solution. Their capability of encoding uncertainty allows to describe the ambiguity of the measured cell IDs caused by the propagation effects. Their capabilities further allows to incorporate prior information extracted from neighboring positions. We will describe this in Chapter 5.

Once defined, the Bayesian networks are trained with pre-measured data, optimizing all probability distributions within the model in such a way that the model could reproduce the training data sets with highest probability. The search for the position is then straight forward. The mobiles position is estimated by searching for the model which produces a given sequence of snapshots (taken from the mobile currently to be localized) with highest probability. The so found models position equals the position estimate.

3.3 Summary of the Network Measurement based Handset Localization Method

In the previous section we have discussed several propagation effects which are responsible for the location dependence of the received power levels. We have explained how we can take advantage of probabilistic models to find a positioning solution for mobile handsets. The main characteristics of the introduced NMR method are summarized as follows:

- Propagation effects are responsible for ambiguous received power levels which we use as location dependent parameters.
- The received power levels of the serving cell do not carry enough information to locate a mobile handset accurately.
- Urban areas, however, show significant shadowing which is advantageous to distinguish different positions.
- Received power levels and cell IDs of neighboring cells and their dependencies are used to increase the discriminating power of the resulting cell ID patterns.
- We avoid considering absolute power levels, because we do not know receiver characteristics and power control effects.
- We stick to measured reference data, because of limited environmental data available.
- Probabilistic models are chosen to represent the dependencies between the received cell IDs and relative power levels.
- One model represents a certain position and is trained with the reference data measured at that position.
- Localization is performed by comparing the current cell ID pattern with all models in the area of interest (the current serving cell area).
- The corresponding position of the most likely model is chosen to be the position estimate.

Chapter 4

Probabilistic Networks

A Bayesian network is defined as a graphical model containing variables, represented by nodes, and relationships between the variables indicated by edges. A simple Bayesian network containing only three nodes is shown in Fig. 4.1, where the node X_1 has some (currently still undefined) influence on its children X_2 and X_3. The strength of Bayesian networks is the capability of mathematically encoding uncertain expert knowledge, by introducing modularity - for instance, a given realization of $X1$ separates (d-separates) the nodes X_2 and X_3 - and hence increases the explanatory power of a model and at the same time reducing its complexity to a feasible number. In other words, it is a way to represent a situation, to structure it and to allow a user to draw a decision even under uncertainty.

Four important answers are given to the question of what Bayesian networks and methods have to offer [27]:

- Bayesian networks can handle incomplete data sets, since they naturally encode the correlation between input variables. If we would only know the state of X_1 we could infer the probabilities of interest (e.g. a certain state of X_2).
- Bayesian networks allow to learn about causal relationships. This means we can gain understanding of a problem during data analysis and also allows us to make predictions in the presence of interventions.

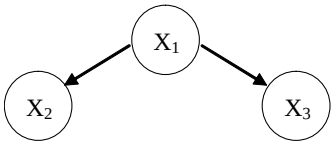


Figure 4.1: Simple Bayesian network with nodes represented by random variables $X_1, \ldots, X_3$. The directed edges indicate relationships between the variables.

- Bayesian networks in conjunction with statistical techniques allow to naturally combine prior knowledge with observed data. This is especially important if data is scarce or expensive. If there is some believe about the probability density distribution of X_1 the model do not have to start from complete ignorance.

- Bayesian methods in conjunction with Bayesian networks allow to avoid overfitting of data.

In this thesis we will not make use of all of the Bayesian networks's strengths, and hence we limit this chapter to the topics we need throughout the thesis. For a more comprehensive discussion we recommend to read for instance [14, 27, 31, 51].

4.1 Reasoning under Uncertainty

To see the basic idea of reasoning under uncertainty we assume only one single discrete node. Imagine a Bernoulli process and suppose to observe a sequence $\mathbf{D} = (X(1), X(2), \ldots, X(M))$ of M identically and independent (iid) Bernoulli trials. Each trail is represented by a random variable $X(m) : m = 1, \ldots, M$ being either in the state 0 or 1. The question of interest is: How probable is it to receive 1 again at the $(M+1)^{th}$ observation? To solve this problem the probability of receiving 1 is defined as $p(X(m) = 1) = \theta$ and hence $p(X(m) = 0) = 1-\theta$ represents the probability of receiving 0. In general, having multinomial random variables, a parameter $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_s)$ will be defined with values $\theta_1, \theta_2, \ldots, \theta_s$ corresponding to the possible true values of the physical probability for the different states s of the random variable. Now, the goal is to find θ in such a way that the data $\mathbf{D}$ is best represented. Starting with the likelihood function for the Bernoulli process

$$L(\theta) = p(\boldsymbol{X}|\theta) = \prod_{m=1}^{M} p(X(m) \mid \theta), \tag{4.1}$$

it can be found for the example if assuming a certain realization $\mathbf{d} = (1, 0, 0, 1, 1)$ for $\boldsymbol{D}$:

$$\begin{aligned} L(\theta) = p(\boldsymbol{d}|\theta) = \prod_{m=1}^{5} p(x(m)|\theta) &= p(1|\theta)p(0|\theta)p(0|\theta)p(1|\theta)p(1|\theta) \\ &= \theta(1-\theta)(1-\theta)\theta\theta \\ &= \theta^{\eta_1}(1-\theta)^{\eta_2}, \end{aligned} \tag{4.2}$$

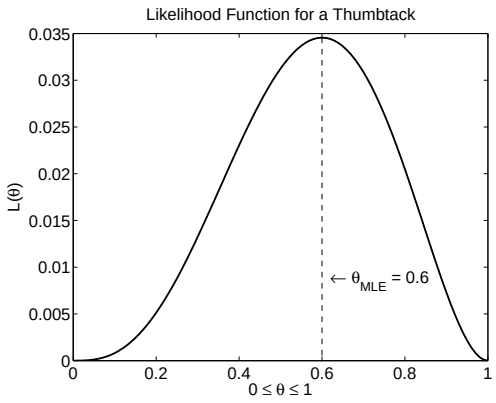


Figure 4.2: Likelihood function of the Bernoulli example for a given observation sequence of $\mathbf{d} = (1, 0, 0, 1, 1)$.

where the variable η_1 denotes the number of occurring $1s$ and η_2 denotes the number of occurring $0s$.

Staying in the framework of the classical statistic, the likelihood function could now being maximized with regard to θ. This would result in the desired value of θ where $\mathbf{d}$ would most likely occur according to

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \theta^{\eta_1}(1-\theta)^{\eta_2} = \frac{\eta_1}{\eta_1 + \eta_2}. \tag{4.3}$$

This, however, gives a single scalar for the parameter θ and does not represent the uncertainty about the parameter anymore.

For the sample observation sequence $\mathbf{d} = (1, 0, 0, 1, 1)$ the likelihood function is shown in Fig. 4.2. The most likely value of the parameter θ given the observation sequence $\boldsymbol{d}$ is thus 0.6. Hence, the probability of finding the $(M+1)^{th}$ observation being 1 is also 0.6.

The Bayesian approach however utilizes also prior information about the parameter to model uncertainty about the parameters. For instance, if assuming that the Bernoulli process represents a coin being thrown, this could be the believe of using a loaded coin, favoring heads or tails.

To compute the probability $p(x_{M+1}|\boldsymbol{D})$ for receiving 1 at the next observation in the Bayesian framework we have to apply Bayes' rule according to

$$p(\theta|\boldsymbol{D}) = \frac{p(\theta)p(\boldsymbol{D}|\theta)}{p(\boldsymbol{D})} \tag{4.4}$$

with

$$p(\boldsymbol{D}) = \int p(\boldsymbol{D}|\theta)p(\theta)d\theta. \tag{4.5}$$

Combining (4.2) and (4.4) results in

$$p(\theta|\boldsymbol{D}) = \frac{p(\theta)\theta^{\eta_1}(1-\theta)^{\eta_2}}{p(\boldsymbol{D})}, \tag{4.6}$$

with the probability density distribution $p(\theta)$ being called the prior and $p(\theta|\boldsymbol{D})$ being called the posterior. With (4.6) one has now an equation representing the knowledge about the parameter θ after observing the data $\boldsymbol{D}$.

For the determination of the probability that the $(M+1)^{th}$ observation in the example is 1, the joint probability distribution of $X(M+1) = 1$ and the parameter θ given the observed data $\boldsymbol{D}$ is computed. Afterwards the parameter θ is marginalized out:

$$\begin{aligned} p(X(M+1) = 1|\boldsymbol{D}) &= \int p(X(M+1) = 1|\theta)p(\theta|\boldsymbol{D})d\theta \\ &= \int \theta p(\theta|\boldsymbol{D})d\theta. \end{aligned} \tag{4.7}$$

$$\tag{4.8}$$

To compute (4.7) it is necessary, however, to know $p(\theta|\boldsymbol{D})$ given by (4.6) and thus to know the prior distribution $p(\theta)$.

A common approach is to use the beta-distribution $Beta(\alpha, \beta)$ for the prior according to

$$p(\theta) = Beta(\theta|\alpha_1, \alpha_2) = c\theta^{\alpha_1 - 1}(1-\theta)^{\alpha_2 - 1} \tag{4.9}$$

with $c = \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$ being a normalization constant and $\Gamma(.)$ denoting the Gamma function.

The parameters $\alpha_1 > 0$ and $\alpha_2 > 0$ of the distribution are called hyperparameters to distinguish them from the actual parameter θ. Both hyperparameters must be greater zero otherwise a division by zero would prevent normalization.

The beta distribution has the same form as the likelihood function (4.2) of the Bernoulli process and hence the posterior (4.6) also results in a beta distribution according to

$$p(\theta|\boldsymbol{D}) \propto \theta^{\alpha_1+\eta_1-1}(1-\theta)^{\alpha_2+\eta_2-1} \propto Beta(\theta|\alpha_1+\eta_1, \alpha_2+\eta_2). \tag{4.10}$$

Since the prior, as well as the posterior is in the family of Beta distributions the Beta family is conjugate for sampling from binomial process.

To obtain the probability of a 1 in the $(M+1)^{th}$ observation according to (4.7) we receive, using $\Gamma(x+1) = x\Gamma(x)$ and $\int_0^1 x^\alpha(1-x)^\beta dx = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}$:

$$p(X(M+1)=1|\boldsymbol{D}) = \frac{\alpha_1 + \eta_1}{\alpha_1 + \alpha_2 + \eta_1 + \eta_2}. \tag{4.11}$$

To finally access the prior's hyperparameters α_1 and α_2 a method known as equivalent samples can be used. Other methods are available, but will not be discussed here. See for instance [71] for further information. In this method the hyperparameters are considered as pseudo counts. Suppose first that one is completely ignorant about the distribution of θ. It is hence desirable to pick a prior that is non-informative. We will see a non-informative prior constructed based on Jeffreys' rule in section 4.3; for now we will use the natural choice of $Beta(\theta|0,0)$. In this case, the posterior distribution according to (4.10) yields $Beta(\theta|\eta_1,\eta_2)$ resulting in a posterior mean of

$$\hat{\theta}_{bayes} = \frac{\eta_1}{\eta_1 + \eta_2}. \tag{4.12}$$

This, however, is equivalent to the maximum likelihood given in (4.3), which is entirely determined by the observed outcomes. $Beta(\theta|0,0)$ can thus be assigned to be the least influential prior[1]. When this prior is now used to compute the posterior distribution according to (4.10) the current state of information (after the observation of the data) equals the prior knowledge for future observations. In other words, the hyperparameters are interpreted as the number of observations one would have had to see to achieve the current state of information, if he would have started from complete ignorance about the distribution of θ.

In Fig. 4.3 the probability of 1 in the $(M+1)^{th}$ observation is given for different hyperparameters. The state without any knowledge about the prior ($Beta(\theta|0,0)$) achieves the same results of 0.6 as in the classical framework. Note however that $Beta(\theta|1,1)$ equivalent to a uniform distribution differs from 0.6, since here the assumed uniform prior distribution is already knowledge about the problem.

4.2 The Bayesian Network

Let us assume discrete Bayesian networks $\lambda = (\mathcal{G}, \mathcal{P})$ consisting of the directed acyclic network structure $\mathcal{G}$ and a probability set $\mathcal{P}$ as shown in Fig. 4.4. The network structure $\mathcal{G} = (\mathcal{V}, E)$ consists of the node set $\mathcal{V} = (1, \ldots, V)$ represented

[1] $Beta(\theta|0,0)$ however is not defined. To be precise we would have to consider $Beta(\theta|\epsilon,\epsilon)$ for arbitrarily small $\epsilon > 0$. This, however, would result in a distribution, which approaches 2 point masses at $\theta = 0$ and $\theta = 1$. Such a prior belief seems extremely strong, since it says θ is essentially either 0 or 1. Intuitively we would consider this unreasonable, but exactly this distribution yields a posterior mean as close as possible to the maximum likelihood estimator.

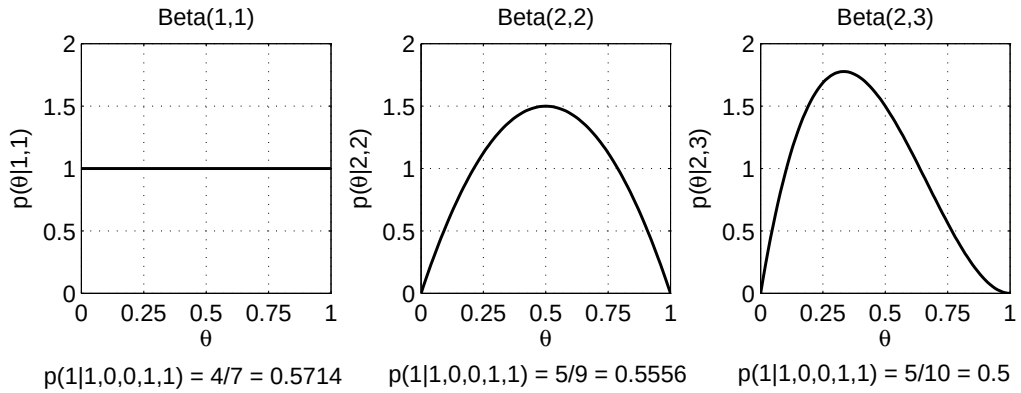


Figure 4.3: Different Beta distributions and associated results for $p(X(M+1) = 1|\boldsymbol{D} = (1,0,0,1,1))$.

by a set of multinomial random variables $X = (X_v : v \in \mathcal{V})$ and encodes the conditional independence assertions about the variables in X indicated by the edges E. The edges are a subset of the set $\mathcal{V} \times \mathcal{V}$ of ordered pairs of nodes. They are called undirected if two ordered pairs (α, β) and (β, α) of nodes belong to E. If $(\alpha, \beta) \in E$ but $(\alpha, \beta) \notin E$ the edge is directed and α is a *parent* of β. The set of parents of a node β is denoted by $pa(\beta)$, and the set of children of node α by $ch(\alpha)$. The alphabet of each discrete variable X_v is defined to be $\mathcal{S} = \{x^s : s = 1, \ldots, S\}$, where S represents the number of different states of the random variable X_v. Note that we have limited the alphabet to be the same for every variable X_v . A more general approach is not required in this thesis.

For the joint distribution $p(\boldsymbol{X})$, given a particular structure $\mathcal{G}$, can now be written:

$$\begin{aligned} p(\boldsymbol{X}) &= p(X_1, \ldots, X_V) \\ &= \prod_{v=1}^{V} p(X_v | X_{v-1}, \ldots, X_1) \quad , \quad X_0 \doteq 1 \\ &= \prod_{v=1}^{V} p(X_v | pa(X_v)), \end{aligned} \tag{4.13}$$

where the term $pa(X_v)$ denotes the parent set of the variable X_v. For the Bayesian network shown in Fig. 4.4 (4.13) yields

$$p(\boldsymbol{X}) = p(X_4|X_3)p(X_3|X_2, X_1)p(X_2)p(X_1). \tag{4.14}$$

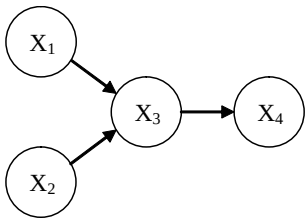


Figure 4.4: Simple Bayesian network with nodes represented by random variables $X_1, \ldots, X_4$ and directed edges.

Conditional independencies
$p(X_1)$
$p(X_2)$
$p(X_3\|X_1, X_2)$
$p(X_4\|X_3)$

Table 4.1: Conditional independence assertion encoded by the edges in Fig. 4.4

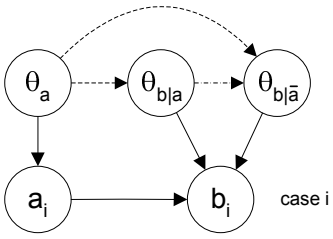


Figure 4.5: A Bayesian network of two binary variables $A \in \{a, \bar{a}\}$ and $B \in \{b, \bar{b}\}$ showing the assumptions about parameter independence. The dashed link has to be removed for global parameter independence, the dash-dotted link for local parameter independence.

Since the Bayesian network probability set $\mathcal{P}$ should rather be updated with the data than to be specified precisely, we consider the joint distribution as belonging to some specified family of possible distributions, parameterized by $\boldsymbol{\theta}$. In the Bernoulli process example, for instance, the probability of $1s$ were not learned; instead the posterior distribution $p(\theta|\boldsymbol{D})$ for the variable θ representing the probabilities of $1s$ has been updated.

The joint probability of a certain realization $\boldsymbol{x}$ of $\boldsymbol{X}$ for a Bayesian network given the parameters is then given by

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{v=1}^{V} p(x_v|pa(X_v), \boldsymbol{\theta}). \tag{4.15}$$

Note that the parameter vector $\boldsymbol{\theta}$ still contains all parameters and thus each factor of the product cannot be considered independent of the others. An update of the parameters would thus only be possible jointly. To allow further simplifications a certain parent configuration is defined as $\boldsymbol{\rho}_{vp} = pa(X_v)$. All different parent configurations are enumerated from $p = 1$ to $p = p_v$ for every node X_v. Consider for instance the node X_3 in Fig. 4.4 having two binary parent nodes X_1 and X_2 and an possible set of instantiations according to Tab. 4.2. The first parent configuration is then $\boldsymbol{\rho}_{31} = (X_1 = 0, X_2 = 0)$. Accordingly, the second

Case Number	X_1	X_2	X_3
1	0	0	0
2	0	0	1
3	0	1	0
4	0	1	1
5	1	0	0
6	1	0	1
7	1	1	0
8	1	1	1

Table 4.2: A example of possible instantiations for the Bayesian network given in Fig. 4.4.

parent configuration is $\boldsymbol{\rho}_{32} = (X_1 = 0, X_2 = 1)$.

The parameter vector $\boldsymbol{\theta}$ is then defined according to

$$\begin{aligned} \boldsymbol{\theta} &= (\boldsymbol{\theta}_v : v = 1, \ldots, V) \\ \boldsymbol{\theta}_v &= (\boldsymbol{\theta}_{vp} : p = 1, \ldots, p_v) \\ \boldsymbol{\theta}_{vp} &= (\theta_{vps} : s = 1, \ldots, S) \end{aligned} \tag{4.16}$$

with $\sum_{s=1}^{S} \theta_{vps} = 1$. A parameter θ_{vps} represents the probability $p(X_v = x^s | pa(X_v) = \boldsymbol{\rho}_{vp})$. Thus, one parameter is defined for every realization x_v for all nodes and all parent configurations.

The assumption of global and local parameter independence (see also Fig. 4.5) defined as

$$p(\boldsymbol{\theta}) = \prod_{v=1}^{V} p(\boldsymbol{\theta}_v) \tag{4.17}$$

$$p(\boldsymbol{\theta}_v) = \prod_{p=1}^{p_v} p(\boldsymbol{\theta}_{vp}) \tag{4.18}$$

allows to re-write (4.15) according to

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{v=1}^{V} p(x_v | pa(X_v), \boldsymbol{\theta}_v) \tag{4.19}$$

containing independent factors $p(x_v|pa(X_v), \boldsymbol{\theta}_v)$.

Further factorization because of local parameter independence yields

$$
\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{\theta}) &= \prod_{v=1}^{V} p(x_v|pa(X_v), \boldsymbol{\theta}_v) \\
&= \prod_{v=1}^{V}\prod_{p=1}^{p_v} p(x_v|pa(X_v) = \boldsymbol{\rho}_{vp}, \boldsymbol{\theta}_{vp}). \qquad (4.20)
\end{aligned}
$$

Now suppose a database $\boldsymbol{D} = (\boldsymbol{x}(1), \ldots, \boldsymbol{x}(m))$ of m independent and identically distributed (iid) observations with $\boldsymbol{x} = (x_1, \ldots, x_V)$ being a realization of $\boldsymbol{X}$. Assume further that η^{vps} is the multiplicity of $(X_v = x^s|pa(X_v) = \boldsymbol{\rho}_{vp})$. We then write

$$
p(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_{v=1}^{V}\prod_{p=1}^{p_v}\prod_{s=1}^{S} p(x_v = x^s|pa(X_v) = \boldsymbol{\rho}_{vp}, \boldsymbol{\theta}_{vp})^{\eta_{vps}}. \qquad (4.21)
$$

Since $\eta_{vps} = 0$ for all x^s not equal a certain node realization x_v and not element of the assigned parent configuration $\boldsymbol{\rho}_{vp}$ in the database, they do not contribute to the product. We can hence factorize over all x^s in the alphabet.

Similar to the Bernoulli case where the parameter θ was equal to the probability of observing a 1, the parameters are now set to $\theta_{vps} = p(x_v = x^s|pa(X_v) = \boldsymbol{\rho}_{vp})$. Equation 4.21 is now rewritten according to

$$
p(\boldsymbol{D}|\boldsymbol{\theta}) = \prod_{v=1}^{V}\prod_{p=1}^{p_v}\prod_{s=1}^{S} (\theta_{vps})^{\eta_{vps}}. \qquad (4.22)
$$

Hence, the global and local parameter independence allows to estimate all parameters $\boldsymbol{\theta}_{vp}$ for a given node v and parent configuration $\boldsymbol{\rho}_{vp}$ separately.

The uncertainty about the parameters θ_{vps} can be expressed with a prior distribution $p(\boldsymbol{\theta}_{vp})$ which is commonly assumed for the multinomial case to be Dirichlet[2] defined as:

$$
p(\boldsymbol{\theta}_{vp}) = \mathcal{D}(\boldsymbol{\theta}_{vp}|\alpha_{vp1}, \ldots, \alpha_{vpS}) = \frac{\Gamma(\alpha^+)}{\prod_{s=1}^{S}\Gamma(\alpha_{vps})}\prod_{s=1}^{S}(\theta_{vps})^{\alpha_{vps}-1} \qquad (4.23)
$$

with $\alpha^+ = \sum_{s=1}^{S}\alpha_{vps}$, and $\alpha_{vps} > 0, s = 1, \ldots, S$. The Dirichlet distribution is a generalization of the Beta-distribution used in the binomial case of the last section. Note that the prior distribution could be any distribution in general, however [22, 57] proves that the prior has to be Dirichlet if the global and local parameter independence assumption should hold.

[2]For a short introduction about the multinomial process see Appendix B.

The posterior of the parameters $\boldsymbol{\theta}$ becomes by applying Bayes' rule, (4.22), the assumption of global and local parameter independence and a Dirichlet distribution $\mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\alpha})$ for the prior:

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{\alpha}) &= \frac{p(\boldsymbol{D}|\boldsymbol{\theta})\prod_{v=1}^{V}\prod_{p=1}^{p_v}\mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\alpha}_{vp})}{p(\boldsymbol{D})} \\ &\propto \prod_{v=1}^{V}\prod_{p=1}^{p_v}\left[\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}}\prod_{s=1}^{S}(\theta_{vps})^{\alpha_{vps}-1}\right] \\ &= \prod_{v=1}^{V}\prod_{p=1}^{p_v}\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}+\alpha_{vps}-1} \\ &\propto \mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\eta}_{vp}+\boldsymbol{\alpha}_{vp}). \end{aligned} \tag{4.24}$$

Equation (4.24) shows that $p(\boldsymbol{\theta}_{vp}|\boldsymbol{D},\boldsymbol{\alpha}_{vp})$ has the same mathematical form as a Dirichlet distribution $\mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\eta}_{vp}+\boldsymbol{\alpha}_{vp})$. Since both distributions, $p(\boldsymbol{\theta}_{vp})$ and the Dirichlet distribution $\mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\eta}_{vp}+\boldsymbol{\alpha}_{vp})$ are normalized with respect to $\boldsymbol{\theta}_{vp}$ the equality

$$P(\boldsymbol{\theta}_{vp}|\boldsymbol{D},\boldsymbol{\alpha}_{vp}) = \mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\eta}_{vp}+\boldsymbol{\alpha}_{vp}) \tag{4.25}$$

is received. Parameter updating is thus reduced to only adopting the hyperparameters of the Dirichlet distribution, rather than computing the complete posterior distribution numerically. The posteriori hyperparameters are therefore

$$\tilde{\boldsymbol{\alpha}}_{vp} = \boldsymbol{\eta}_{vp}+\boldsymbol{\alpha}_{vp}. \tag{4.26}$$

Equivalent to the pseudo counts in the binomial case the hyperparameters $\boldsymbol{\alpha}_{vp}$ can also be interpreted as pseudo counts in the multinomial case. The approach is exactly the same and α_{vps} hence represents the number of "prior observations" of $(X_v = x^s|pa(X_v) = \boldsymbol{\rho}_{vp})$.

4.3 The Non-Informative Prior

In section 4.1 we argued that the $Beta(\theta|0,0)$ distribution is a correct choice to encode a complete ignorance about the distribution of θ. We proved its correctness by showing the equality to the classical maximum likelihood estimator. There exist, however, several other approaches for the creation of a suitable prior if neither expert knowledge, nor experimental insight is available. A widely used one, the Jeffreys prior [30], is introduced in this section. It is, however, even

though existing for Bayesian networks in general, NP hard for arbitrary network structures [12].

The Jeffreys prior is motivated by an invariance argument: Suppose $\pi_p(p)$ as an prior for p. In order for $\pi_p(p)$ to be non-informative it is required that a parameter transformation $\theta = p(\theta)$ results in a non-informative prior $\pi_\theta(\theta)$.

To meet this requirement Jeffreys suggests to pick $\pi_p(p)$, given p, according to

$$\pi_p(p) \propto I(p)^{\frac{1}{2}} \quad , \quad I(p) = -\mathbb{E}(\frac{\partial^2 \log f(X|p)}{\partial p^2}) \tag{4.27}$$

where $I(p)$ denotes the *Fisher information* and $f(X|p)$ the likelihood function.

To prove the invariance with respect to the parameter transformation we apply $\theta = p(\theta)$ in (4.27), which yields

$$\begin{aligned} \pi_\theta &\propto \sqrt{-\mathbb{E}_X\left(\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2}\right)} = \\ &= \sqrt{-\mathbb{E}_X\left(\frac{\partial^2 \log f(X|p)}{\partial p^2}\left(\frac{\partial p}{\partial \theta}\right)^2\right)} = \\ &= \sqrt{I(p)\left(\frac{\partial p}{\partial \theta}\right)^2} \propto \pi_p(p)\left|\frac{\partial p}{\partial \theta}\right|. \end{aligned} \tag{4.28}$$

$$\tag{4.29}$$

This satisfies the invariance argument.

The motivation for Jeffreys' method is that the Fisher Information $I(\theta)$ is an indicator of the amount of information brought by the observations about θ. By finding a parameter θ, which results in large $I(\theta)$, is thus equivalent to minimizing the influence of the prior [64].

The Jeffreys prior can be generalized to a parameter vector $\boldsymbol{\theta}$ according to

$$\pi(\boldsymbol{\theta}) \approx |\boldsymbol{I}(\boldsymbol{\theta})|^{\frac{1}{2}} \tag{4.30}$$

where a single entry of the *Fisher Information Matrix* $\boldsymbol{I}(\boldsymbol{\theta})$ is defined by

$$[\boldsymbol{I}(\boldsymbol{p})]_{l,k} = -\mathbb{E}\left(\frac{\partial^2 \log f(\boldsymbol{X}|\boldsymbol{p})}{\partial p_l \partial p_k}\right) \tag{4.31}$$

with l, k line and column indices respectively.

4.3.1 The Jeffreys prior for Bayesian Networks

The following section will review the derivation of the Jeffreys prior for Bayesian Networks with arbitrary structure. For a more detailed version we refer the reader to [33].

We consider only a single data vector $\boldsymbol{x}$. The log-likelihood of a data vector is then written according to

$$logP(\boldsymbol{x}|\boldsymbol{\theta}) = \sum_{v=1}^{V}\sum_{p=1}^{p_v} z_{vp}\left(\sum_{s=1}^{S}(z_{vs}\log\theta_{vps})\right), \tag{4.32}$$

where z_{vp} and z_{vs} are two indicator variables defined as follows:

For $\boldsymbol{x} \in \boldsymbol{D}$, let x_v denote the value of X_v in $\boldsymbol{x}$ and $\boldsymbol{x}_{pa(X_v)}$ the value of $pa(X_v)$ in $\boldsymbol{x}$. The indicator variables are then given by

$$z_{vs} = \begin{cases} 1, & \text{if } x_v = x^s, \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad z_{vp} = \begin{cases} 1, & \text{if } \boldsymbol{x}_{pa(X_v)} = \rho_{vp}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.33}$$

Recalling that $\theta_{vpS} = 1 - \sum_{s=1}^{S-1}\theta_v ps$ we find the first derivative according to

$$\frac{\partial \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_{vpl}} = \frac{z_{vp}z_{vl}}{(\theta_{vpl})^2} - \frac{z_{vp}z_{vS}}{\theta_{vpS}}. \tag{4.34}$$

For the second derivative $\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_{vpl}\partial\theta_{vpk}}$ we consider two different cases $l \neq k$ and $l = k$ with $l, k = 1, \ldots, S-1$ which yield

$$\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_{vpl}\partial\theta_{vpk}} = \begin{cases} -z_{vp}z_{vS}\theta_{vpS}^{-2} = -\frac{z_{vp}zvS}{\theta_{vpS}^2} & , l \neq k, \\ -\frac{z_{vp}z_{vl}}{\theta_{vpl}^2} - \frac{z_{vp}z_{vS}}{\theta_{vpS}^2} & , l = k \end{cases} \tag{4.35}$$

The elements of the Fisher information matrix are now the expectations of (4.35) over the set of all possible data vectors $\boldsymbol{x} \in \boldsymbol{D}$. For cases $l \neq k$ this is

$$\begin{aligned} -\mathbb{E}_X\left[\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial\theta_{vpl}\partial\theta_{vpk}}\right] &= \sum_{\boldsymbol{x}\in\boldsymbol{D}} p(\boldsymbol{x}|\boldsymbol{\theta})\frac{z_{vp}z_{vS}}{(\theta_{vpS})^2} \\ &= \frac{1}{(\theta_{vpS})^2}\sum_{\boldsymbol{x}\in\boldsymbol{D}|pa(X_v)=\rho_{vp},X_v=x^S} p(\boldsymbol{x}|\boldsymbol{\theta}) \\ &= \frac{p(pa(X_v)=\rho_{vp}, X_v = x^S|\boldsymbol{\theta})}{(\theta_{vpS})^2} \\ &= \frac{p(pa(X_v)=\rho_{vp}|\boldsymbol{\theta})}{(\theta_{vpS})}. \end{aligned} \tag{4.36}$$

For case $l = k$ we get similar to (4.36)

$$-\mathbb{E}_X\left[\frac{\partial^2 \log p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial(\theta_{vpl})^2}\right] = \frac{p(pa(X_v) = \rho_{vp}|\boldsymbol{\theta})}{(\theta_{vpl})} + \frac{p(pa(X_v) = \rho_{vp}|\boldsymbol{\theta})}{(\theta_{vpS})}. \tag{4.37}$$

To obtain the Fisher Information Matrix for m data vectors we have to multiply all elements of the matrix by m [33]. The sub-matrix $\boldsymbol{I}_{vp}(\boldsymbol{\theta})$ is then given by

$$\boldsymbol{I}_{vp}(\boldsymbol{\theta}) = \begin{pmatrix} m(\frac{p_{vp}}{\theta_{vp1}} + \frac{p_{vp}}{\theta_{vpS}}) & m(\frac{p_{vp}}{\theta_{vpS}}) & \cdots & m(\frac{p_{vp}}{\theta_{vpS}}) \\ m(\frac{p_{vp}}{\theta_{vpS}}) & m(\frac{p_{vp}}{\theta_{vp2}} + \frac{p_{vp}}{\theta_{vpS}}) & \cdots & m(\frac{p_{vp}}{\theta_{vpS}}) \\ \vdots & \vdots & \ddots & \vdots \\ m(\frac{p_{vp}}{\theta_{vpS}}) & m(\frac{p_{vp}}{\theta_{vpS}}) & \cdots & m(\frac{p_{vp}}{\theta_{vp,S-1}} + \frac{p_{vp}}{\theta_{vpS}}) \end{pmatrix} \tag{4.38}$$

where $p_{vp} = p(pa(X_v) = \rho_{vp}|\boldsymbol{\theta})$. The sub-matrix $\boldsymbol{I}_{vp}(\boldsymbol{\theta})$ is an $(S-1)\times(S-1)$ matrix, since $\boldsymbol{\theta_{vp}}$ is completely determined by $(\theta_{vp1}, \ldots, \theta_{vp,S-1})$ according to $\sum_{s=1}^{S} \theta_{vps} = 1$.

The determinant of (4.38) is given by

$$|\boldsymbol{I}_{vp}(\boldsymbol{\theta})| = \frac{(m \cdot p_{vp})^{S-1}}{\prod_{l=1}^{S} \theta_{vpl}}, \tag{4.39}$$

which is shown e.g. in [6]. The Fisher information matrix $\boldsymbol{I(\theta)}$ is a block diagonal matrix, where the blocks are defined by $\boldsymbol{I}_{vp}(\boldsymbol{\theta})$. The determinant of a block diagonal matrix is the product of the determinants of the blocks. Hence, we finally receive for the Jeffreys prior

$$\begin{aligned} \pi(\boldsymbol{\theta}) &\propto |\boldsymbol{I}(\boldsymbol{\theta})|^{\frac{1}{2}} \\ &\propto \prod_{v=1}^{V}\prod_{p=1}^{p_v}\left(\frac{(m \cdot p_{vp})^{S-1}}{\prod_{l=1}^{S} \theta_{vpl}}\right)^{\frac{1}{2}} \\ &\propto \prod_{v=1}^{V}\prod_{p=1}^{p_v}\left((m \cdot p_{vp})^{\frac{S-1}{2}} \prod_{l=1}^{S} (\theta_{vpl})^{-\frac{1}{2}}\right) \\ &\propto \prod_{v=1}^{V}\prod_{p=1}^{p_v}\left((p_{vp})^{\frac{S-1}{2}} \prod_{l=1}^{S} (\theta_{vpl})^{-\frac{1}{2}}\right). \end{aligned} \tag{4.40}$$

Chapter 5

The Bayesian Model

In Chapter 3 we introduced the general idea of localizing a mobile handset via pattern matching. We have seen that models are required to represent the received power signals at a specific position. The question is how to get a suitable model for a specific position? In this chapter we will show how to find an optimal model, based on statistical evidence and, by the same token, trying to avoid vague terms like "significant" or "not significant", which we have used in the previous chapters.

5.1 Model Selection

The basic problem of finding a "true" model to a given set of data is to identify all relevant variables, parameters and all the causal relationships between them. Since we consider the "reality" as being very complex, e.g. of infinite dimension with lots of effects and interactions, it is hardly possible to describe the "reality" with a finite set of data. Hence, it is for sure impossible to describe reality with a single model. All models we can find are therefore in the very best case only approximations.

We can use the Bayesian framework, however, to define an appropriate model for the given data and to objectively and efficiently assess its usefulness to the localization problem.

Recommended by [14] we split the construction of the model into three distinct components:

1. The qualitative stage, which considers only general relationships between the variables of interest.

2. The probabilistic stage resulting in a graph allowing an probabilistic interpretation.

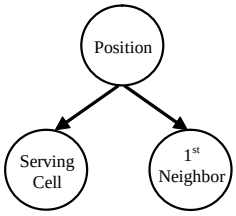


Figure 5.1: Simple localization model: Directed acyclic graph representing dependencies between a position and cell IDs.

3. The quantitative step specifying the numerical conditional probability distributions.

5.1.1 The Qualitative Stage

One strength of probabilistic networks is their pictorial representation. It is possible to get a feeling about the existing relationships between the variables of a problem. Let us consider the serving cell and the first neighbor cell of our localization method. It is obvious that once a cell is declared as serving cell, it can never appear to be the first neighboring cell at the same time. A dependency exists between the two cells. Hence, we can define a simple model for this case as shown in Fig. 5.1. The vertices or nodes in the graph represent variables and the edges between the nodes encode dependencies. The directed edges represent causal relationships. The directed edge between *Serving Cell* and *1st Neighbor* models the assumption that we receive a certain first neighbor given the serving cell. If we reversed the directed edge, this would model the first neighbor to be the causal reason for the serving cell. This case can be ruled out, since the model should focus on representing the reality and should not try to introduce implausible hypotheses. Burnham and Anderson [9] for example states: "If a particular model (parametrization) does not make biological sense, this is reason to exclude it from the set of candidate models ...".

5.1.2 The Probabilistic Stage

The second step is to turn the qualitative representation into some algebraic properties. For a probabilistic network this is the computation of the joint probability $p(X)$, which may be difficult especially in the case of undirected graphs. For Bayesian networks $p(X)$ simplifies according to (4.13). For the example shown in Fig. 5.1 we get

$$P(\mathit{ServingCell,1stNeighbor}) = P(\mathit{1stNeighbor}|\mathit{ServingCell})P(\mathit{ServingCell}). \quad (5.1)$$

5.1.3 The Quantitative Stage

The quantitative stage plugs numbers into (4.13). Our localization example represents the simplest case of a directed acyclic graph with only discrete nodes. The numerical inputs are thus only the conditional distributions of each node given the parent configuration as already indicated in (5.1). In our case, the numerical probabilities will most likely be drawn from measurements, since we can physically stop at a specific position and take some cell ID measurements. For other networks, however, it might be a better choice to ask experts to provide sufficient conditional probabilities to finally specify the joint distribution or to combine both approaches.

5.2 Model Optimization

As illustrated in the sections before, we are able to construct a model for the positioning problem. We still do not know, however, if the model represents an optimal solution to the problem. It is necessary to find a metric which allows to objectively judge if the model is "good" or "not good". We formulate the problem as maximizing the joint probability of the data $\boldsymbol{D}$ and the model's structure $\mathcal{G}$ according to

$$\hat{\mathcal{G}} = \arg\max_{\mathcal{G}}(p(\boldsymbol{D}, \mathcal{G})). \tag{5.2}$$

We assume the same settings for the graphical network as in Chapter 4. The discrete Bayesian networks $\lambda = (\mathcal{G}, \mathcal{P})$ consists of the network structure $\mathcal{G}$ and a probability set $\mathcal{P}$. The network structure consists of the node set $\mathcal{V} = (1, \ldots, V)$ represented by a set of discrete random variables $X = (X_v : v \in \mathcal{V})$. The alphabet of each discrete variable X_v is defined to be $\mathcal{S} = \{x^s : s = 1, \ldots, S\}$, where S represents the number of different states of all random variable X_v. The parameter $\boldsymbol{\theta}$ is defined according to (4.16). θ_{vps} denote the multinomial parameter corresponding to the probability $p(X_v = x^s | pa(X_v) = \boldsymbol{\rho}_{vp})(\theta_{vps} > 0, \sum_{s=1}^{S} \theta_{vps} = 1)$.

That is, the parameters $\boldsymbol{\theta}$ correspond to the probability set $\mathcal{P}$ for a single-case Bayesian network.

To find the most probable network structure $\hat{\mathcal{G}}$ given a database we have to compute

$$\hat{\mathcal{G}} = \arg\max_{\mathcal{G}} p(\mathcal{G}|\boldsymbol{D}) \tag{5.3}$$

over all possible network structures.

To find a computational traceable equation five additional assumptions [27,28] are introduced:

1. Assumption 1: There exist a database $\boldsymbol{D}$ containing only discrete multinomial variables.

2. Assumption 2: All parameters associated with each variable in a network structure are independent (global parameter independence) and all the parameters associated with each instance of the parents of a variable are independent (local parameter independence) according to

$$p(\boldsymbol{\theta}) = \prod_{v=1}^{V} p(\boldsymbol{\theta}_v) \tag{5.4}$$

$$p(\boldsymbol{\theta}_v) = \prod_{p=1}^{p_v} p(\boldsymbol{\theta}_{vp}). \tag{5.5}$$

   See also (4.17) of Chapter 4.

3. Assumption 3: The densities for parameters $\boldsymbol{\theta}_{vp}$ depend only on the structure of the network that is local to variable X_v - namely, $\boldsymbol{\theta}_{vp}$ only depends on X_v and its' parents according to

$$p(\boldsymbol{\theta}_{vp}|\mathcal{G}_1) = p(\boldsymbol{\theta}_{vp}|\mathcal{G}_2) \quad p = 1, \ldots, p_v \tag{5.6}$$

   with $\mathcal{G}_1, \mathcal{G}_2$ being two possible network structures and X_v having the same parents in $\mathcal{G}_1$ and $\mathcal{G}_2$.

4. Assumption 4: All parameters $\boldsymbol{\theta}_{vp}$ are assumed to be Dirichlet distributed. That is, there exists exponents α_{vps} that satisfy

$$p(\boldsymbol{\theta}_{vp}) = c \cdot \prod_{s=1}^{S} (\theta_{vps})^{\alpha_{vps}-1}, \tag{5.7}$$

   where c is a normalization constant. See also (4.23) of Chapter 4.

5. Assumption 5: The database is complete. That is, there are no missing values in the database $\boldsymbol{D}$.

Exploring the consequences of these assumptions, the application of the discrete multinomial variable assumption (assumption 1) and the complete database assumption (assumption 5) yields

$$p(\boldsymbol{D}|\boldsymbol{\theta},\mathcal{G}) = \prod_{v=1}^{V}\prod_{p=1}^{p_v}\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}}, \tag{5.8}$$

if η_{vps} denotes the multiplicity of cases in database $\boldsymbol{D}$ in which $X_v = x^s$ and $pa(X_v) = \boldsymbol{\rho}_{vp}$. Combining (5.8) with the conditional probability

$$p(\boldsymbol{D}|\mathcal{G}) = \int p(\boldsymbol{D}|\boldsymbol{\theta},\mathcal{G})p(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta} \tag{5.9}$$

results in

$$p(\boldsymbol{D}|\mathcal{G}) = \int \left[\prod_{v=1}^{V}\prod_{p=1}^{p_v}\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}}\right] p(\boldsymbol{\theta}|\mathcal{G})d\boldsymbol{\theta}. \tag{5.10}$$

From the parameter independence assumption (assumption 2) $p(\boldsymbol{\theta}|\mathcal{G})$ can be expressed according to

$$p(\boldsymbol{\theta}|\mathcal{G}) = \prod_{v=1}^{V}\prod_{p=1}^{p_v} p(\boldsymbol{\theta}_{vp}|\mathcal{G}) \tag{5.11}$$

and substituting (5.11) into (5.10) yields

$$p(\boldsymbol{D}|\mathcal{G}) = \int \left[\prod_{v=1}^{V}\prod_{p=1}^{p_v}\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}}\right]\left[\prod_{v=1}^{V}\prod_{p=1}^{p_v} p(\boldsymbol{\theta}_{vp}|\mathcal{G})\right] d\boldsymbol{\theta}. \tag{5.12}$$

After interchanging the integral with the products one receives

$$p(\boldsymbol{D}|\mathcal{G}) = \prod_{v=1}^{V}\prod_{p=1}^{p_v}\int \left[\prod_{s=1}^{S}{\theta_{vps}}^{\eta_{vps}}\right] p(\boldsymbol{\theta}_{vp}|\mathcal{G})d\boldsymbol{\theta}. \tag{5.13}$$

Assuming further, the prior parameter distributions $p(\boldsymbol{\theta}_{vp}|\mathcal{G})$ to be uniform[1] it follows that $p(\boldsymbol{\theta}_{vp}|\mathcal{G}) = \boldsymbol{C}_{vp}$ for some constants $\boldsymbol{C}_{vp}$. Since $p(\boldsymbol{\theta}_{vp}|\mathcal{G})$ is a probability density function it follows that

$$\int \boldsymbol{C}_{vp}d\boldsymbol{\theta}_{vp} = 1. \tag{5.14}$$

Since $p(\boldsymbol{\theta}_{vp}|\mathcal{G})$ is also a special case of the Dirichlet distribution we can use Dirichlet's integral with the following solution [70]

[1]This assumption was first introduced by Cooper and Herskovits [13] for their derivation of the K2 algorithm.

$$\int \prod_{s=1}^{S} (\theta_{vps})^{\eta_{vps}} d\boldsymbol{\theta}_{vp} = \frac{\prod_{s=1}^{S} \eta_{vps}!}{(\eta_{vp} + S - 1)!} \tag{5.15}$$

with $\eta_{vp} = \sum_{s=1}^{S} \eta_{vps}$. Setting all $\eta_{vps} = 0$ and therefore $\eta_{vp} = 0$, $\boldsymbol{C}_{vp}$ can be found to be

$$\boldsymbol{C}_{vp} = (S-1)! = p(\boldsymbol{\theta}_{vp}|\mathcal{G}). \tag{5.16}$$

Substituting (5.16) into (5.13) and taking into account that $(S-1)!$ is constant one obtains

$$p(\boldsymbol{D}|\mathcal{G}) = \prod_{v=1}^{V} \prod_{p=1}^{p_v} (S-1)! \int \prod_{s=1}^{S} (\theta_{vps})^{\eta_{vps}} d\boldsymbol{\theta}. \tag{5.17}$$

This, however, is again Dirichlet's integral and (5.15) can be used as solution according to

$$p(\boldsymbol{D}|\mathcal{G}) = \prod_{v=1}^{V} \prod_{p=1}^{p_v} \frac{(S-1)!}{(\eta_{vp} + S - 1)!} \prod_{s=1}^{S} \eta_{vps}!. \tag{5.18}$$

By using Bayes's rule and by being indifferent about the prior model structure, (5.18) can finally be rewritten as

$$p(\mathcal{G}|\boldsymbol{D}) = c \prod_{v=1}^{V} \prod_{p=1}^{p_v} \frac{(S-1)!}{(\eta_{vp} + S - 1)!} \prod_{s=1}^{S} \eta_{vps}! \tag{5.19}$$

with c being a constant prior probability $p(\mathcal{G})$ for each model structure.

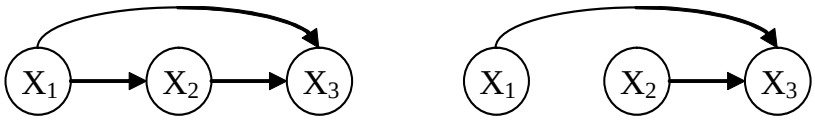
To find the best model, we use the algorithm proposed by Cooper and Herskovits [13]. The algorithm starts with a node v assuming no parents and computes $p(\mathcal{G}, \boldsymbol{D})$, which we can do using (5.19). Single parents are added as long as this addition increases the probability of the resulting network structure. Adding parents to a node is stopped if $p(\mathcal{G}, \boldsymbol{D})$ does not increase further. The same procedure is then repeated with all other nodes according to a initial node ordering.

To show an example, let us consider only the serving, the first and the second neighboring cells. Table 5.2 shows an excerpt of the measured database which will be $\boldsymbol{D}$ for this example.

Applying (5.18) to compute $p(\mathcal{G}_1, \boldsymbol{D})$ for the Bayesian network structure shown in Fig. 5.2(a) results in

Case	Serving Cell X_1	X_2	X_3
1	131	56	41
2	131	56	41
3	56	131	41
4	131	56	41
5	56	41	131
6	56	131	41
7	131	56	41
8	56	41	131
9	131	56	41
10	56	131	41

Table 5.1: Excerpt of database $\boldsymbol{D}$ showing 10 cases used for the Bayesian networks presented in Fig. 5.2.



(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$

Figure 5.2: Examples of Bayesian network structures.

$$\begin{aligned} p(\mathcal{G}_1, \boldsymbol{D}) &= p(\mathcal{G}_1)\frac{(3-1)!0!5!5!}{(10+3-1)!}\frac{(3-1)!0!5!0!}{(5+3-1)!}\frac{(3-1)!2!0!3!}{(5+3-1)!} \\ &\quad \frac{(3-1)!5!0!0!}{(5+3-1)!}\frac{(3-1)!3!0!0!}{(3+3-1)!}\frac{(3-1)!0!0!2!}{(2+3-1)!} \\ &= 1.08204810^{-11}. \end{aligned} \tag{5.20}$$

See table 5.2 for corresponding node and parent configurations which have been used in this example.

The same computation is done for all different possible network structures. For the probability $p(\mathcal{G}_2, \boldsymbol{D})$, given the network structure shown in Fig. 5.2(b), we obtain $p(\mathcal{G}_2, \boldsymbol{D}) = 2.86906610^{-13}$. If we have no prior assumption about one of the two Bayesian network structures and, hence, set $p(\mathcal{G}_1) = p(\mathcal{G}_2)$ the comparison of the two network structures results in

$$\frac{p(\mathcal{G}_1, \boldsymbol{D})}{p(\mathcal{G}_2, \boldsymbol{D})} = 37.7. \tag{5.21}$$

The data implies that the network structure $\mathcal{G}_1$ shown in Fig. 5.2(a) is 37.7 times more likely than $\mathcal{G}_2$. This result is not surprising, since omitting the edge

Variable	Value
V	3
$\mathcal{S}$	$\{131, 56, 41\}$
S	3
$\boldsymbol{\rho}^{11}$	[]
$\boldsymbol{\rho}^{21}$	[131]
$\boldsymbol{\rho}^{22}$	[56]
$\boldsymbol{\rho}^{31}$	[131, 56]
$\boldsymbol{\rho}^{32}$	[56, 131]
$\boldsymbol{\rho}^{33}$	[56, 41]

Table 5.2: Variables used for the computation of (5.20). All values are based on the data shown in table 5.2.

between node X_0 and X_1 would mean to deny any dependence between the serving cell and the strongest neighboring cell. Looking at the data $\boldsymbol{D}$ in Tab. 5.2 reveals, however, that whenever $X_0 = 131$, X_1 equals 56 and whenever $X_0 = 56$, X_1 equals either 131 or 41.

5.3 The NMR Model

We now resume to the original problem of finding the optimal Bayesian network to characterize a position.

5.3.1 Assumptions and Restrictions

We assume complete data in our observations $\mathbf{X} = (X_v : v = 1, \ldots, 7)$, with X_v representing the measured serving cell ID. Complete data is likely since we aim for urban or suburban environment. If no cell ID value of any neighboring cell can be measured, we assign the symbol 'not available' or $x^s = 0$ to this state. That is, we rather assume that we cannot receive an Rx-level at a specific position, instead of assuming that any neighbor cell could be available there, but we just cannot measure it. In such a case we would have to define the cell ID value as being 'missing'. In the measured test database of $m > 250000$ samples only $\approx 1.8\%$ of all samples contain missing values. This indicates that our assumption of complete data is legitimate.

5.3.2 Scenarios

We consider the following two situations. First, we construct models from databases $\boldsymbol{D}$ containing different numbers of measured samples and second, we

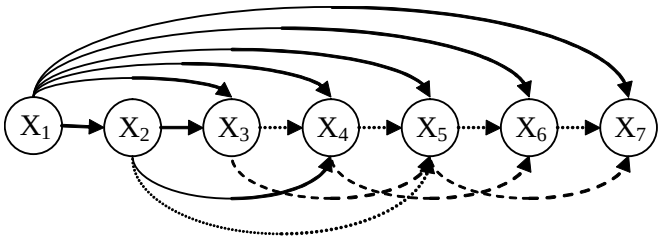


Figure 5.3: Constructed Bayesian network structure based on a different number of measured samples $\boldsymbol{D}$. The solid lines, the dotted lines and the dashed lines denote the model structure constructed from $m_1 = 5000$, $m_2 = 30000$ and $m_3 = 60000$ samples respectively.

construct models based on different types of areas, namely *street*, *place*, *crossing* and *all*. *All* denotes the model constructed from all measured samples regardless of the actual position.

Different Sample Numbers

To evaluate the impact of the training sample number on the model's structure, we apply a different number of training samples $m \in \boldsymbol{D}$ and create the optimized structures (Fig. 5.3). In the first situation we find the optimized network structure using $m_1 = 250000$ samples, which is still computational traceable. Arbitrarily chosen lower training sample numbers $m_2 = 60000$, $m_3 = 30000$ and $m_4 = 5000$ are used to give an indication which arcs are still present with a lower number of training samples. This represents a strong relationship between nodes. All found structures are similar but show additional arcs when constructed from a larger number of samples (except for 250000 samples). The Bayesian network structure based on $m_4 = 5000$ samples shows dependencies mainly from the serving cell (node 1) with additional dependencies of node 3 and 4 from the strongest neighboring cell (node 2). The network based on $m_3 = 30000$ samples has an edge added from node 2 to node 5 and all nodes are now dependent from their direct predecessors. If we, once more, increase the relevant sample number to 60000, the dependencies of all nodes v from the nodes $v - 2$ emerge. If m is set to all 250000 measured samples no additional arcs are added compared to $m_3 = 60000$. For an easier comparison of each arc's contribution to $p(\mathcal{G}, \boldsymbol{D})$ we use a measure similar to (5.21) defined as

$$\psi_{ij} = \log p(\mathcal{G}_{opt}|\boldsymbol{D}) - \log p(\mathcal{G}_{ij}|\boldsymbol{D}), \tag{5.22}$$

where $p(\mathcal{G}_{opt}|\boldsymbol{D})$ is the probability for the optimal model structure $\mathcal{G}_{opt}$ according to (5.19), $\boldsymbol{D}$ is the data and $p(\mathcal{G}_{ij}|\boldsymbol{D})$ represents the probability resulting from the model omitting the arc between nodes i and j. This is a measure of how much less

i-j	ψ_{ij}						
	m=5k	m=30k	m=60k	street	place	crossing	m=250k (≡ all)
1-2	*0.36*	*2.44*	*5.03*	*1.65*	*1.17*	*1.35*	*7.26*
1-3	*0.23*	*1.89*	*4.05*	*1.38*	*0.70*	*1.04*	*6.76*
1-4	*0.20*	*0.44*	*1.03*	*0.57*	*0.19*	*0.25*	*2.52*
1-5	0.37	0.36	0.03	0.02	0.01	0.01	0.02
1-6	0.35	1.06	0.80	0.09	0.50	0.54	0.58
1-7	0.32	1.09	0.57	0.09	0.29	0.48	0.68
2-3	0.15	1.46	3.23	1.36	0.67	0.86	5.25
2-4	0.02	0.50	1.34	0.34	0.11	0.11	1.67
2-5		0.03	0.17	0.20	0.09	0.03	1.05
3-4		0.59	1.63	0.52	0.24	0.19	2.32
3-5			0.33	0.36	0.21	0.17	2.30
4-5		0.16	0.62	1.01	0.43	0.35	3.80
4-6			0.28	0.20	0.22	0.27	0.32
5-6		0.39	0.73	0.33	0.19	0.32	0.80
5-7			0.38	0.19	0.41	0.37	0.54
6-7		0.40	0.68	0.39	0.20	0.37	1.06

Table 5.3: ψ_{ij} for different databases $\boldsymbol{D}$. ψ_{ij} is assigned to the arc from node i to node j. "Important" arcs indicating a strong impact on the model's suitability to the trainings data are emphasized. All numbers shown have to be multiplied by 10^4.

likely the model would be, if an arc, and therefore also the dependency, would be omitted. The logarithm is used for more comfortable numbers, since (5.22) grows fast with the number of samples, nodes, parent combinations and the cardinality of the alphabet of the nodes. In Tab. 5.3 all values for ψ_{ij} are shown for different tested databases $\boldsymbol{D}$.

The most "important" dependencies (*italic* in Tab. 5.3), indicated by high numbers of ψ_{ij}, are the ones from the serving cell. This is plausible and caused by the same reason we have already seen in the simple example at the beginning of this section. A certain serving cell is always found jointly with a limited number of different neighboring cell IDs. Note also that $\psi_{45} = 3.80$, which indicates a "higher importance" than the arcs starting in the serving cell's node. Such cases occur if e.g. a certain pattern of cell IDs are measured in a high number of training samples. In our case a pair of cell IDs occurring as neighbor 3 and neighbor 4[2].

[2]Note, that ψ_{45} represents the arc between node 4 and 5 which is equal to a arc between neighbor 3 and neighbor 4, since node 1 represents the serving cell

Different Area Types

For the second situation, where models are constructed for different types of areas, we refer to Fig. 5.4, where all measurement positions are indicated. The symbols $\times, \square, \circ$ denote the measurement positions of the type *street*, *place* and *crossing* respectively. *Street* only contains positions located in a distance to an intersection or a place of at least 10 meters. *Place* addresses positions that are located at open areas with a diameter of at least 30 meters. *Crossing* contains positions in the vicinity (10 meters) of street corners. The measure of how much less likely a model would be, if the arc was omitted is also shown in Tab. 5.3.2.

We find the same model structure $\mathcal{G}_{NMR}$ for all area types as depicted in Fig. 5.5. No arcs are omitted or added for any area type, even though the "importance" of the arcs is different. It is also the same model structure as found for $m_3 = 60000$ and $m_4 = 250000$. Hence, for all further investigations we use $\mathcal{G}_{NMR}$.

5.4 Parameter Learning

In the last sections we have derived a method to construct a graphical model, which represents the most optimal structure given the measured observations. In this section we will find the optimal parameters. The Bayesian approach for this estimation problem is to express uncertainty about the model's probabilities in terms of a prior distribution for the parameters, determining those probabilities and then computing the wanted posterior distribution from the prior and the given data.

We assume discrete Bayesian networks $\lambda^i = (\mathcal{G}, \mathcal{P}^i)$, consisting of the structure $\mathcal{G}_{BNT}$ as shown in Fig. 5.5 for all positions $i \in$ target area $\mathcal{I}$. We denote X_1 to be the serving cell ID and $X_v : v = 2, \ldots, 7$ to be the neighboring cell IDs ordered by decreasing Rx-levels. Note that we do not use the Rx-level of the serving cell, which might be power controlled. The probability set $\mathcal{P}^i$ is defined as collection of distributions with densities $p^i(X_v|pa(X_v), \boldsymbol{\theta^i})$, where the random parameters $\boldsymbol{\theta}^i$ determine all possible distributions p^i can belong to. The parameters $\boldsymbol{\theta}^i$ we want to estimate.

We think of the pre-measured reference data for all positions $i \in \mathcal{I}$ as a sequence of M independent and identical distributed observations[3] and call it database $\boldsymbol{D}$. The subset measured at a certain position $i \in \mathcal{I}$ is labelled $\boldsymbol{D}^i = (\boldsymbol{x}(m) : m = 1, \ldots, M)$ containing the same number of samples M for

[3]In general we can use all location dependent parameters, but for this work we concentrate on the cell IDs and the downlink power level of the neighboring cells reported e.g. by the network measurement report in GSM.

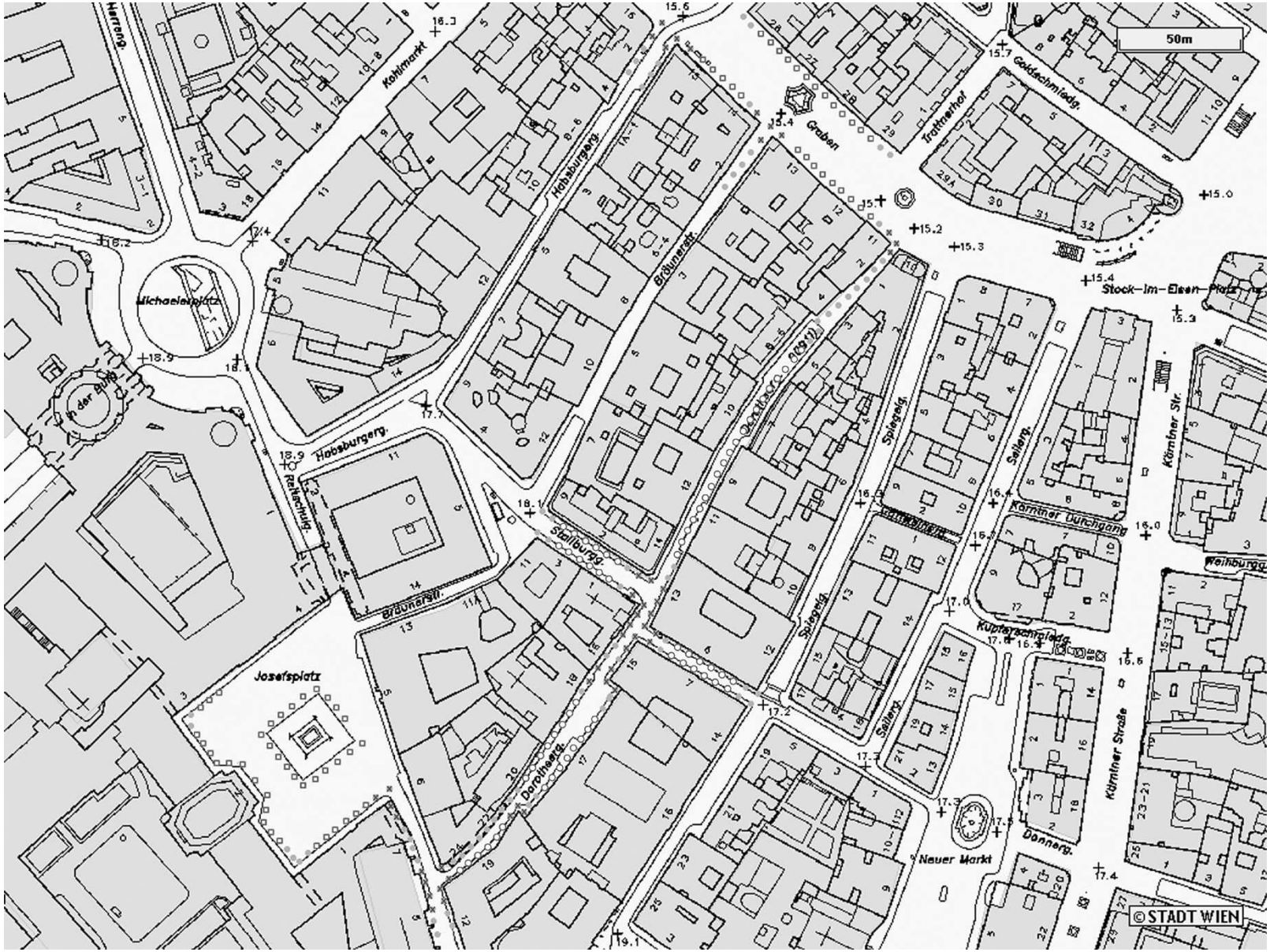


Figure 5.4: Map of the Rx-level measurements. $\times, \square, \circ$ denote measurement positions of type *street*, *place* and *crossing* respectively. Type *all* denotes all positions shown in the map.

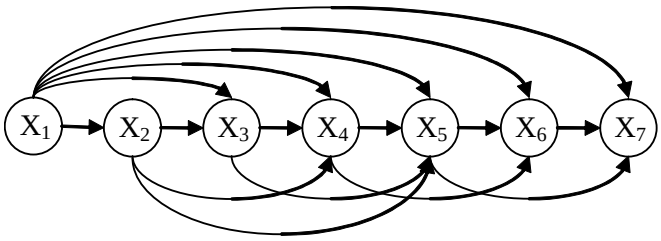


Figure 5.5: NMR based Bayesian network structure $\mathcal{G}_{NMR}$ independent of area type. The model is constructed from samples taken at positions according to Fig. 5.4.

all positions. The alphabet $\mathcal{S} = \{x^s : s = 1, \ldots, S\}$ for all nodes $d_1, \ldots, d_7$ is taken from the operator's network planning tools, extended by the symbol $x^0 =$ "not available", which is used if no cell ID can be measured by the mobile. We note that the sample $\boldsymbol{X}(m)$ is still dependent on the mobile's receiver, because one receiver might still be able to receive the Rx-levels while another receiver might fail. We assume, however, the same sensitivity of the receiver (including antenna) over the entire frequency spectrum of all available cell IDs at its current position. The same receiver characteristics therefore apply for all nodes of the Bayesian network ensuring the same situation for the dependencies between the nodes.

For the parameters $\boldsymbol{\theta}$ we have assumed global and local parameter independence in the last section.

We note that the *global parameter independence* is challenged by the fact that changing the distribution of a certain node can have an impact on the distribution of the remaining nodes (see Fig. 5.6). Imagine for example that the power level of the current serving cell is raised (Fig. 5.6 - Cell ID 5, $\theta_1 = b$). In such a case it will become more unlikely that the current serving cell's power level drops below the strongest neighbor's power level. Thus, a change of the assigned nodes in the Bayesian network for the serving cell and the first neighboring cell will also become less likely. In Fig. 5.6 this means that the probability of being the strongest neighbor is increased for Cell ID 4. Similar to [24], however, we will stick here to the assumption of global parameter independence for computational reasons.

Local parameter independence means that observing a certain node, given a certain realization of parents will not alter our beliefs about the node's distribution if another realization of parents is observed. Let us illustrate this again with an example. We consider only node X_3 and it's parents $pa(X_3) = (X_1, X_2)$, all having a certain distribution $p(X_v|pa(X_v), \boldsymbol{\theta})$ with $v = 1, 2, 3$. Now, if the power level for a certain base station, let us say $X_2 =$ Cell ID b , being currently the

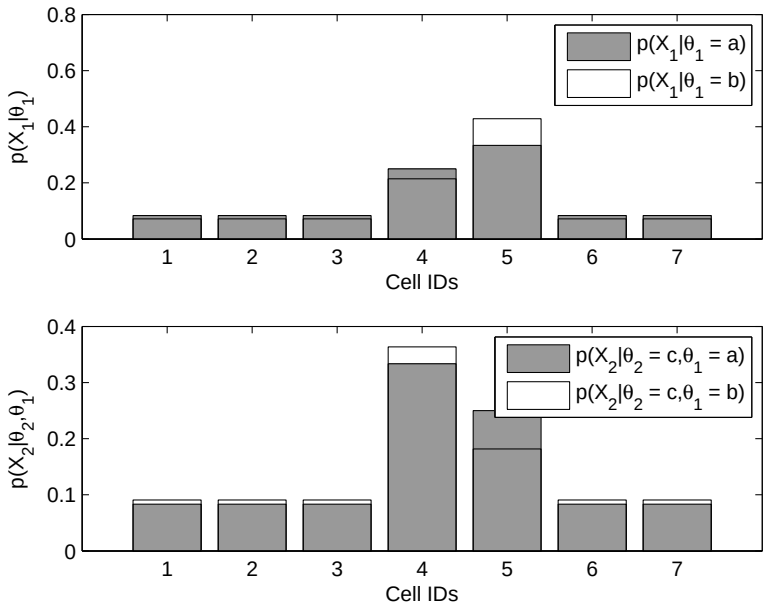


Figure 5.6: Global parameter independence is challenged, since the distribution of a certain node (here X_1) can have an impact on the distribution of the remaining nodes (here X_2). If the probability of $p(X_1 = 5)$ is increased, e.g. due to an power level increase of this Cell ID, the probability of $p(X_2 = 5)$ is decreased.

strongest neighbor, is raised, it has an impact on a certain number of parameters $\boldsymbol{\theta}_{3p} = p(X_3 = x^s | pa(X_3) = \boldsymbol{\rho}_{3p})$. If Cell ID b passes the power level of the current serving base station, e.g. Cell ID a, it might suddenly become the serving cell. Hence, the realization $pa(X_3) = (X_1 = b, X_2 = a)$ is now more likely than $pa(X_3) = (X_1 = a, X_2 = b)$.

The two base stations Cell ID a and Cell ID b have already changed two parameters $\boldsymbol{\theta}_{31}$ and $\boldsymbol{\theta}_{32}$ enumerating $pa(X_3) = (X_1 = b, X_2 = a)$ and $pa(X_3) = (X_1 = a, X_2 = b)$ respectively. Hence, the two parameters $\boldsymbol{\theta}_{31}$ and $\boldsymbol{\theta}_{32}$ are not completely independent. The quite strong assumption about local parameter independence is thus too restrictive in reality. However, we will again stick to this assumption for computational reasons and refer to [24] where the relaxation of the local independence assumption is discussed. The results indicate that the relaxation requires more computational effort but performs better only marginally.

5.5 Informative Priors

For a computation of the posterior probability[4] of a Bayesian network it is necessary to assess the parameter priors $p(\boldsymbol{\theta})$. In general its construction is difficult. We will show, however, that a prior can be constructed for our cases and under certain assumptions, with only a small number of observations.

Since all nodes of the Bayesian network are modelled as multinomial random variables, a Dirichlet distribution $Dir(\boldsymbol{\theta}_{vp} | \boldsymbol{\alpha})$ for all nodes and parent configurations is assumed. While we are in general free to assess any probability density for the parameters $\boldsymbol{\theta}$, the global and local independence assumption limits the prior to be Dirichlet [22, 57]. In such a case we have seen in Section 4.2 that the variable α_{vps}, denoting the multiplicity, can be thought of as pseudo counts: The number of observations one would have had to see in order to achieve the current state of information, if one would have started from complete ignorance.

However, we do not necessarily have to use past samples to obtain information about our Bayesian network's parameters. Since we want to reduce the number of samples in our database as much as possible, we have to extract the information from some other source. The standard approach is to use "expert knowledge".

We assume a similar power level situation in closely co-located positions. The situation is shown in Fig. 5.7 for two positions $i = 44$ and $i = 45$ spatially separated by 5 meters. The serving cell and the first neighbor remain (almost) the same for both positions, while the second neighbor is rather $x_1 = 740$ at position $i = 44$ and $x_1 = 52$ at position $i = 45$ for the tested cases. All other

[4] If we use *approximations* for the posterior probability like the Bayesian information criterion (BIC) [59], or the Minimum Description Length (MDL) [56] the prior is not necessary.

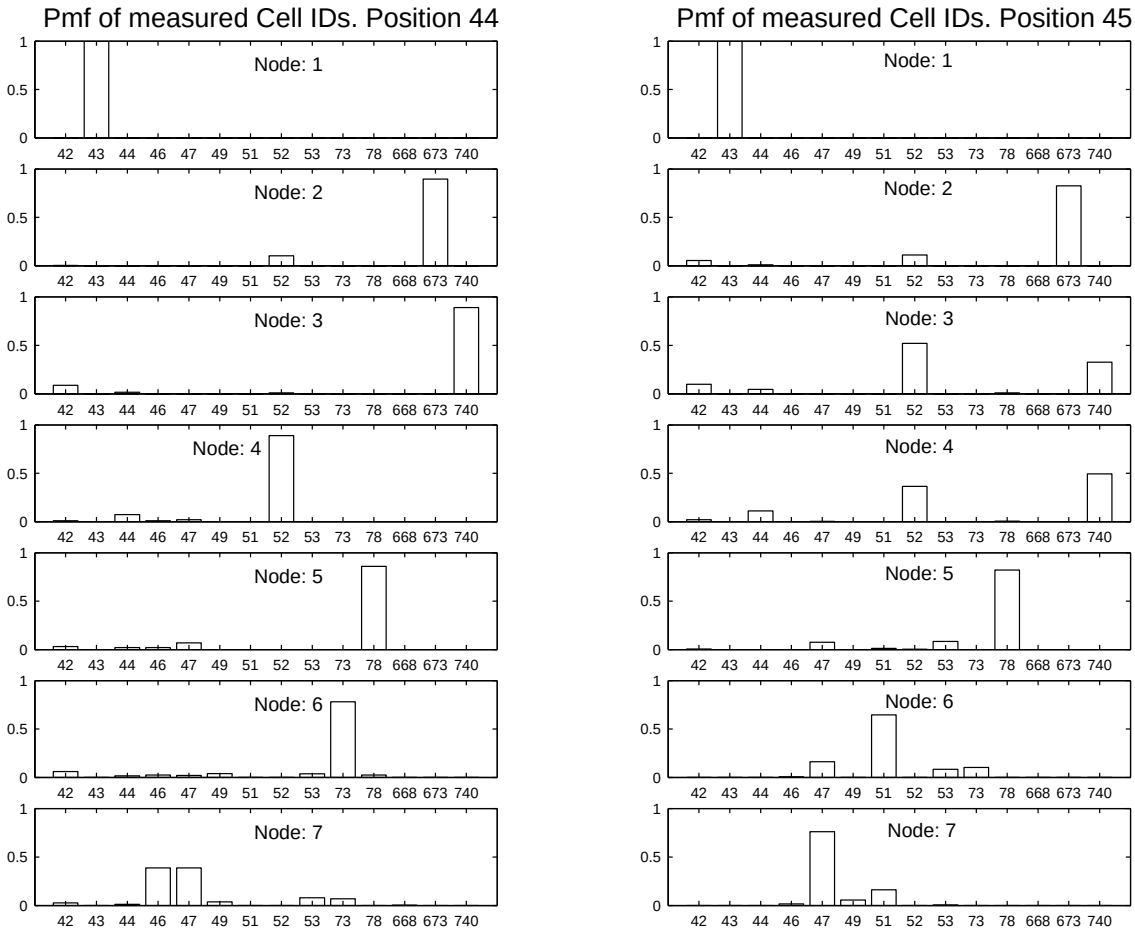


Figure 5.7: Measured cell IDs of two closely co-located positions.

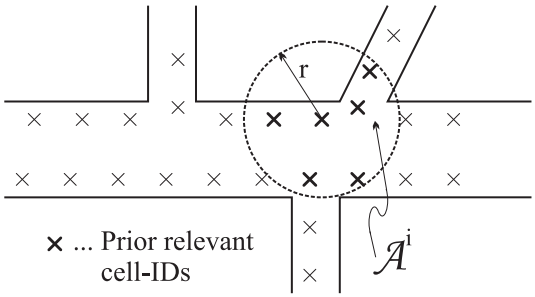


Figure 5.8: Concept of NMR prior.

combinations can also occur, but are less likely.

We thus try to create the prior by spanning a certain area $\mathcal{A}^i$ with radius r around the current position and count the number of equal realization for each node v and parent configuration $\boldsymbol{\rho}_{vp}$ in this area (Fig. 5.8), instead of counting past samples. Informally spoken, we deduct the prior hyperparameters of the current position from knowledge about neighboring positions.

We argue here by induction. That is from the occurrence of an effect - here the similarity between adjacent positions - to the presumed cause - the realization of a certain cell ID pattern at our position of interest. It would be more desirable to reference a relevant series from which the case under discussion could have been drawn from. This has, however, already been rejected for practical reasons, since we do not have a series of earlier measurements of patterns for the exact position of interest.

To determine the radius r it would be desirable to have full knowledge about the distributions $p^i(X_v|pa(X_v), \boldsymbol{\theta}_{vp})$ at every position i. But this is in general not achievable [8, 65]. The description of $p(X_v|pa(X_v), \boldsymbol{\theta}_{vp})$ is thus always based on a limited number of observations, which omits a exact description of the true reality.

For all further investigations we thus define long-term measurements, measured during extensive trails of several weeks and containing about 100000 samples per position (see Appendix D), to represent "true" reality.

This huge number of observations, however, does still not allow a traceable determination of the radius r. A further reduction of the number of samples is required. Studying the long-term measurements reveals only about 100 different Rx-level patterns per position in and around Vienna. This is an indicator of achieving a good description of the "true" reality with only a significantly limited number of observations.

To find good approximations, we use the *Kullback-Leibler* (KL) *information* as a distance measure. The KL distance is defined by

$$K(f,g) = \sum_k f_k log \frac{f_k}{g_k}, \tag{5.23}$$

where f and g are two discrete distributions and f_k and g_k are the probabilities of certain realizations of the distributions. We define a measured set of samples $\boldsymbol{D}_m^i = (x(1), \ldots, x(m))$ for every position i to be the basis of the "true" probability mass function at that position. Note that the sample set $\boldsymbol{D}_m^i$ is different to $\boldsymbol{D}^i$. It is only used to determine the radius r and it is not necessary for the implementation of the method in particular.

Figure 5.9(a) shows the pmf of the suburban site Aspern as bar graph. The overlayed line plots show the approximated pmfs using only $m = 30$ (line) and $m = 300$ (dashed) samples. While $m = 30$ samples are not sufficient to approximate the true pmf , $m = 300$ samples provide a good match and denotes a practically reasonable number of measurements per position[5]. Fig. 5.9(b) shows the corresponding KL distance in a logarithmic scale between the true distribution and the approximated one, computed from 5000 random sample sets. The KL-distance is greater 0 for the entire range of 0 to 300 samples which quantifies the difference between the "true reality" and the approximated distribution ($I(f,g) = 0$ would denote an exact match). Minimizing the KL-distance would yield the optimal sample size, which is in general the sample size of the "true" distribution. A good match, however, is already achieved with $m = 300$ samples, which would require a high number of additional samples to be reasonable increased.

Figure 5.10 shows the same plots of the urban site Hoher Markt in Vienna, which is surrounded by streets with heavy traffic volume. The pmf plot using only $m = 30$ samples is omitted, because of the high number of different measured Rx-level patterns.

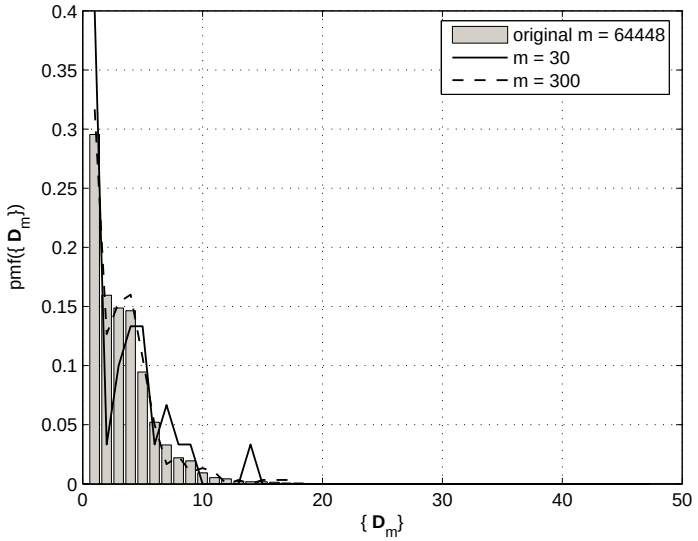
By setting $f(\zeta) = p^i(X_v)$, $g(\xi) = p^{\mathcal{A}^i}(X_v)$ and omitting all dependencies between any node and it's parents, the KL distance for a certain position and node is given by

$$K_v^i(p^i(X_v), p^{\mathcal{A}^i}(X_v)) = \sum_{x_v \in \mathcal{S}} p^i(x_v) \log \frac{p^i(x_v)}{p^{\mathcal{A}^i}(x_v)}, \tag{5.24}$$

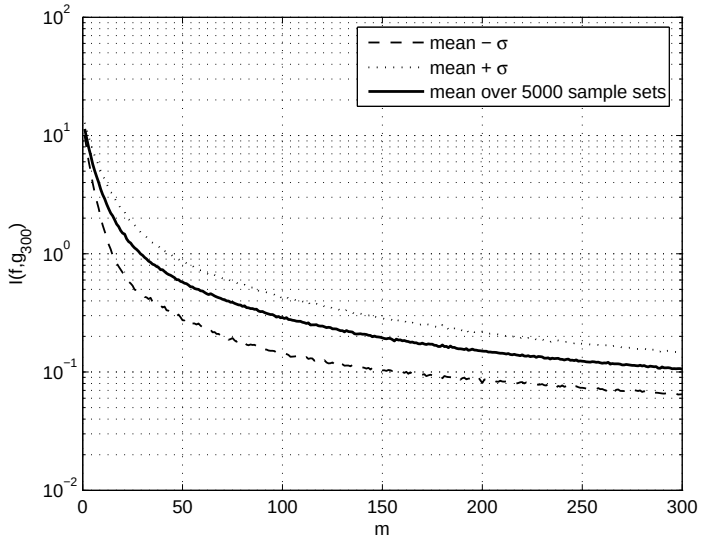
where $p^{\mathcal{A}^i}(X_v)$ represents the probability distribution function of all samples x_v^i contained in the area $\mathcal{A}^i$.

The separate discussion on a per node basis results from evidence given by long-term measurements (see Section 6.3). We expect, for instance, the serving

[5]The utilization of a cost function to search for the optimal number of samples was omitted here. A cost function would be position dependent, e.g. because of streets or pedestrian areas. A number of 300 samples represents a compromise chosen during the measurement campaign.

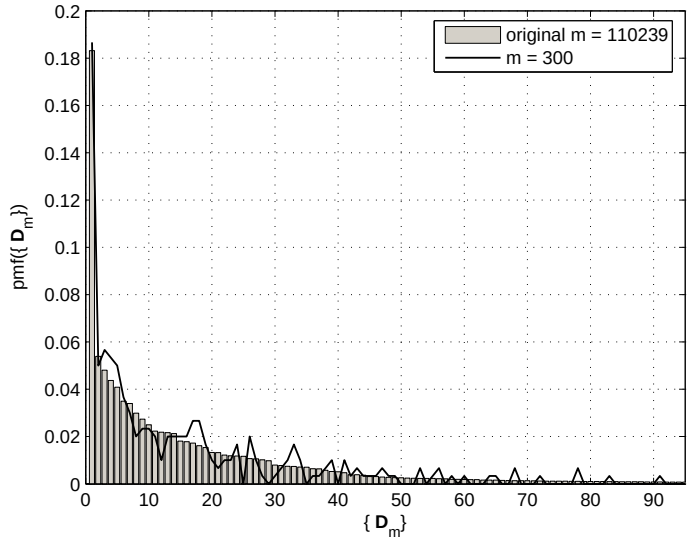


(a) True and approximated pmf of the enumerated measured Rx-level patterns.

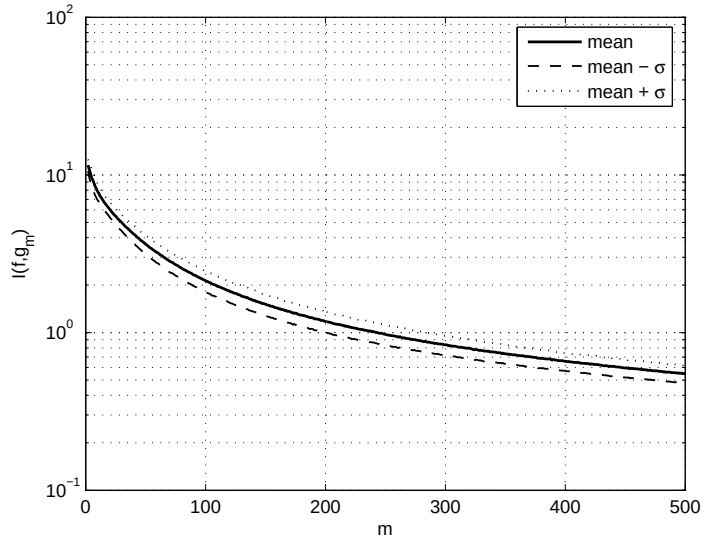


(b) Kullback-Leibler distance of true and approximated pmf with $m = 300$.

Figure 5.9: Comparison of true and approximated probability mass function at the suburban site *Aspern*. For further information about the site refer to Appendix D.



(a) True and approximated pmf of the enumerated measured Rx-level patterns.



(b) Kullback-Leibler distance of true and approximated pmf with $m = 300$.

Figure 5.10: Comparison of true and approximated probability mass function at the urban site *Hoher Markt*. For further information about the site refer to Appendix D.

cell to change its values less likely than the neighboring cells. This corresponds with reality, where the serving base station is statistically located in a shorter distance to the mobile user than neighboring base stations. Less environmental reflections and hence, fewer multipath effects, explain a more stable received power level. The same effects can be found when different neighbor cells are compared.

The assumption $p^i(X_v|pa(x_v)) = p^i(X_v)$ results from the lack of environmental knowledge. A causal relationship between a cell ID and its neighbors cannot be proposed as long as only statistical information about the environment is considered.

Going back to the search for an approximation of the distributions, representing the position i, we aim to loose as little information as possible. Hence, the KL distance $K_v^i(p^i(X_v), p^{\mathcal{A}^i}(X_v))$ should be minimal. We compute the radius to be the mean of the KL distance over all positions according to

$$\hat{r_v} = \arg\min_{r_v} \bar{K}_v \quad \text{with} \tag{5.25}$$

$$\begin{aligned}
\bar{K}_v &= \mathbb{E}_i \left[K_v^i(p^i(X_v), p^{\mathcal{A}^i}(X_v)) \right] && (5.26)\\
&= \frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \sum_{x_v\in\mathcal{S}} \left\{ p^i(x_v) \left[\log p^i(x_v) - \log p^{\mathcal{A}^i(r)}(x_v) \right] \right\} \\
&= \frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \sum_{x_v\in\mathcal{S}} \left\{ p^i(x_v) \log p^i(x_v) \right\} - \\
&- \frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \sum_{x_v\in\mathcal{S}} \left\{ p^i(x_v) \log p^{\mathcal{A}^i(r)}(x_v) \right\}, && (5.27)
\end{aligned}$$

where the first term in (5.27) is independent of the radius r and can be omitted for the minimization. The radius is then given by

$$\hat{r}_v = \arg\max_{r_v} \frac{1}{|\mathcal{I}|} \sum_{i\in\mathcal{I}} \sum_{x_v\in\mathcal{S}} \left\{ p^i(x_v) \log p^{\mathcal{A}^i(r)}(x_v) \right\}. \tag{5.28}$$

Figure 5.11 shows the resulting mean KL distance $\bar{K}_v$ for all nodes v. The x-axis denotes the radius r_v in meters. For illustration a fitted curve (dash-dotted) is shown. For each node, the best approximation (the circle denotes the minimum; the values in parenthesis are given in the format "radius, KL distance") of the distributions $p^i(x_v)$ can be found by using different radii ranging from 6 to 26 meters. The optimal radius for the serving cell has been found to be 6 meters in the tested area. This is, as expected, much lower than the radii of the neighboring cell priors, which were found to be e.g. 26 meters for node 7.

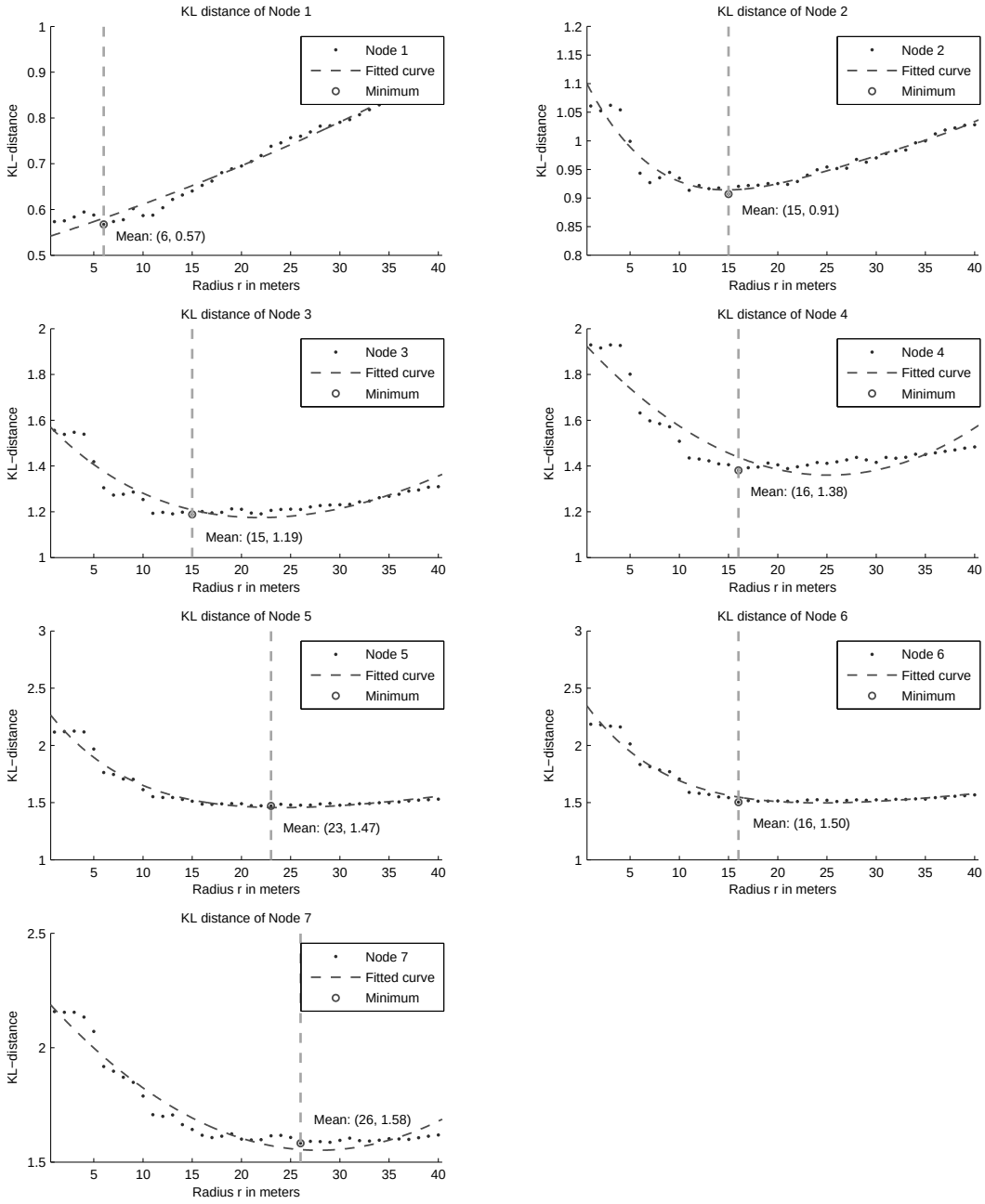


Figure 5.11: Mean KL distance $\bar{K}_v$ for all nodes.

Now, for the sake of simplicity and since it is clear from the context, we will omit the position index i in the further explanations.

We apply the radius $\hat{r}_v$ to characterize the area $\mathcal{A}$ and create an initial guess for the hyperparameters $\boldsymbol{\alpha}$ of the Dirichlet distribution with density $\mathcal{D}$. Since this prior will not represent the true density distribution of the current position and might even assign zero (or close to zero) values to certain probabilities, we add a small offset to the counted values within the circled area $\mathcal{A}$.

Let $\boldsymbol{\eta} = (\eta_{vps} : v = 1, \ldots, V; p = 1, \ldots, p_v; s = 1, \ldots, S)$ be the multiplicity of the observed cases where $(x_v = x^s | pa(X_v) = \boldsymbol{\rho}_{vp})$. Let further be $\boldsymbol{n} = (n_v : v = 1, \ldots, V)$ a parameter vector with its elements $\in [0, 1]$ controlling the level of the uniform offset. The hyperparameters $\boldsymbol{\alpha} = (\alpha_{vps} : v = 1, \ldots, V; p = 1, \ldots, p_v; s = 1, \ldots, S)$ can then be written as

$$\alpha_{vps} = w_v \left[\eta_{vps} n_v + \frac{1}{S}(1 - n_v) \right]. \tag{5.29}$$

The weighting factor $\boldsymbol{w} = (w_v : v = 1, \ldots, V)$ with $w_v \in [0, \infty[$ is used to control the level of believe in the prior compared to the observed data. Zero would indicate that we are completely ignorant about the prior.

Additionally we set[6]

$$\alpha_{vps} = \begin{cases} \alpha_{vps} & \text{if } x_v = x^s \notin \boldsymbol{\rho}_{vp} \\ \varepsilon & \text{otherwise} \end{cases} \tag{5.30}$$

with $\varepsilon = 0.001$. This describes the assumption of no multiple occurrence of a certain cell ID in one fingerprint. Equation (5.30) forces α_{vps} to be close to zero if the same symbol x^s for the node v is also contained in its parents configuration. For instance, a received pattern $(x_2 = 12, x_3 = 10, x_4 = 12)$ would be illegal. The realization x_2 would be equal to x_4, that is, the first and the third neighbor are represented by the same cell ID. This is illegal. However, (5.30) does not exclude the occurrence of $x_5 \ldots x_7$ to be 12. In reality these cases are not permitted either. This is a limitation of the used Bayesian network which only models the dependencies $p(X_v | pa(X_v))$ instead of modelling $p(X_v | X_1, \ldots, X_{v-1}, X_{v+1}, \ldots, X_V)$.

By combining (5.29) and (5.30) we can re-write the equation for the hyperparameters of the Dirichlet prior to

$$\alpha_{vps} = \begin{cases} w_v \left[\eta_{vps} n_v + \frac{1}{S}(1 - n_v) \right] & \text{if } x_v = x^s \notin \boldsymbol{\rho}_{vp} \\ 0 + \varepsilon & \text{otherwise.} \end{cases} \tag{5.31}$$

[6] We use ε to ensure that the hyperparameters remain greater zero to allow the normalization of the Dirichlet prior.

To compute the unknown parameters $\boldsymbol{w}$ and $\boldsymbol{n}$ we use the marginal likelihood $\mathcal{L}$ as a score and compare all models λ given different $\boldsymbol{w}$ and $\boldsymbol{n}$, in the light of the position specific training data $\boldsymbol{D}$. This leads to

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}(\boldsymbol{w},\boldsymbol{n})) &= p(\boldsymbol{D}|\boldsymbol{\alpha}(\boldsymbol{w},\boldsymbol{n})) \\ &= \int p(\boldsymbol{D}|\boldsymbol{\theta})\mathcal{D}(\boldsymbol{\theta}|\boldsymbol{\alpha}(\boldsymbol{w},\boldsymbol{n}))d\boldsymbol{\theta}. \end{aligned} \tag{5.32}$$

Since the used Bayesian network has discrete variables, complete data and Dirichlet prior distributions, the marginal likelihood decomposes into factors for each parent-child combination [13, 28] that is given by

$$\mathcal{L}(\boldsymbol{\alpha}(\boldsymbol{w},\boldsymbol{n})) = \prod_{v=1}^{V}\prod_{p=1}^{p_v} \frac{\Gamma(\sum_{s=1}^{S}\alpha_{vps})}{\Gamma(\sum_{s=1}^{S}(\alpha_{vps}+\eta_{vps}))} \prod_{s=1}^{S} \frac{\Gamma(\alpha_{vps}+\eta_{vps})}{\Gamma(\alpha_{vps})}. \tag{5.33}$$

Maximization over all $\boldsymbol{w}$ and $\boldsymbol{n}$ according to

$$(\hat{\boldsymbol{w}},\hat{\boldsymbol{n}}) = \arg\max_{(\boldsymbol{w},\boldsymbol{n})} p(\boldsymbol{D}|\boldsymbol{\alpha}(\boldsymbol{w},\boldsymbol{n})) \tag{5.34}$$

results in the optimal parameters $\hat{\boldsymbol{w}}$ and $\hat{\boldsymbol{n}}$.

In 5.12 the performance of the constructed prior is compared with the non-informative Jeffreys prior for all nodes. The results shown are the mean results over all positions $i \in \mathcal{I}$ We use $m = 10$ training samples to update the Bayesian network and test the performance of the constructed prior using 60 trial measurements per position. The radii $\boldsymbol{r}$ of the area $\mathcal{A}$ are set to $\hat{\boldsymbol{r}} = [6, 15, 15, 16, 23, 16, 26]$ meters in accordance with the results shown in Fig. 5.11.

As a performance measure we use the marginal log likelihood L_v (or negative logarithmic penalty [14]) by applying batch *parent-child monitors* as derived in Appendix C. Compared with the non-informative Jeffreys prior, this is a relative measure of how adequate our assumptions about the conditional probability distribution of a node given its parents is.

Consider node $v = 1$, representing the serving cell ID. A maximum is shown at $w_1 = 9.5$ and $n_1 = 2$ indicated by the intersection of the two black curves in Fig. 5.12. This provides a mean Bayes factor of

$$L_1 - L_1(\hat{\boldsymbol{w}},\hat{\boldsymbol{n}}) = 1.02 \tag{5.35}$$

compared to the eight priors directly surrounding the maximum and an absolute value of $L_1 - L_1^{ninf} = 3.46$. The model, thus, performs $exp(3.46) = 31.76$ times

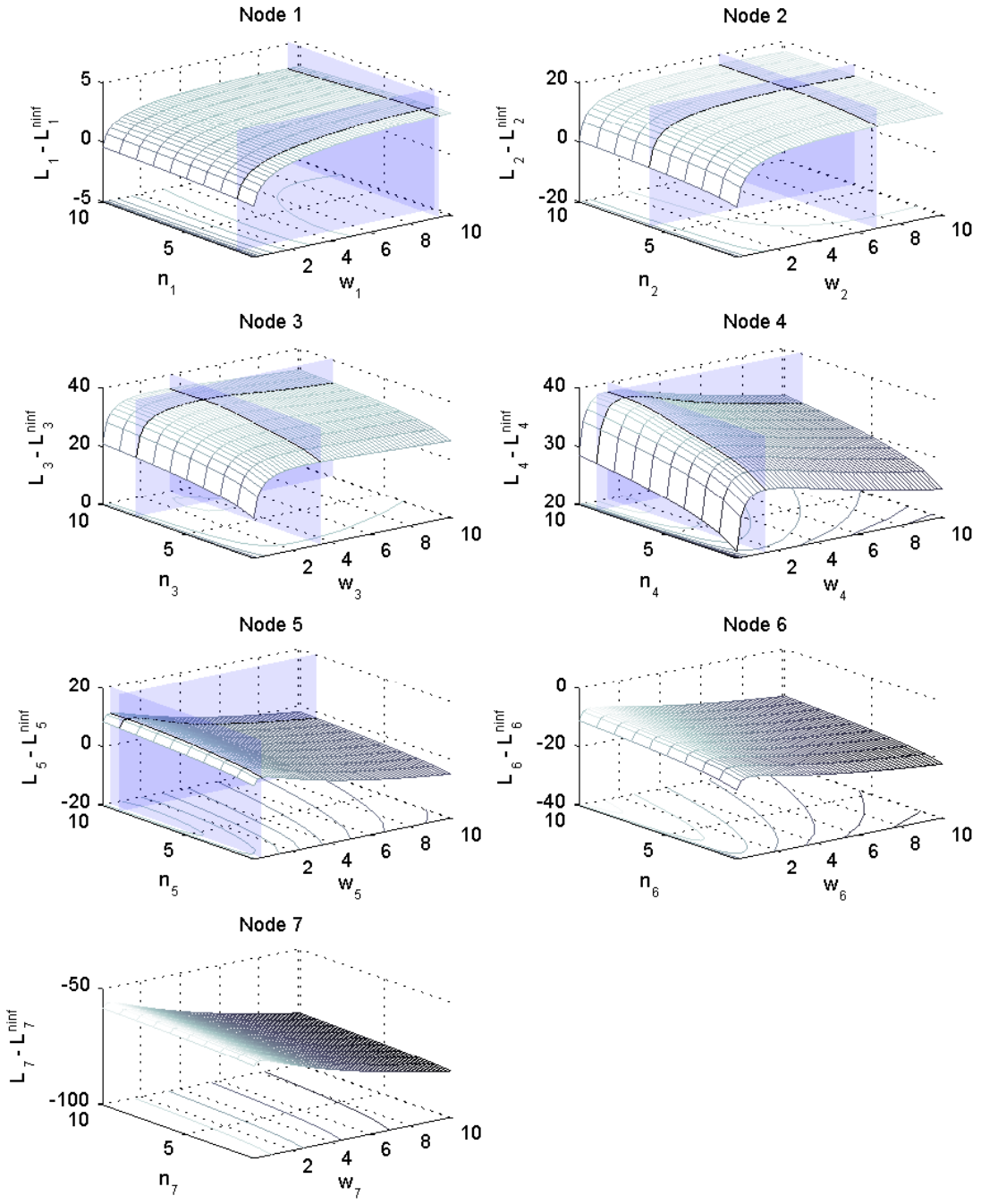


Figure 5.12: Normalized marginal log likelihood for different $\boldsymbol{w}, \boldsymbol{n}$.

Node	- Penalty	$\hat{w}^i$	$\hat{n}^i$
1	2.46	9.5	2
2	16.22	6.9	6
3	35.19	3.5	8
4	38.34	1.5	9
5	11.05	0.5	9

Table 5.4: Prior assessments for the nodes $1-5$ (batch).

Node v	$p(L_v - {L_v}^{ninf} < 0)$
1	0.045
2	0.055
3	0.041
4	0.045
5	0.026

Table 5.5: Prior assessments for the nodes $1-5$.

better than the model utilizing only the Jeffreys prior, if all other nodes remain unconsidered. For the other nodes similar results occur as given in Tab. 5.5.

Node 6 and node 7 show, both, negative results for the marginal log likelihood. That is, the non-informative prior performs better than the constructed prior. This is not surprising because of larger cell ID variations of the fifth and sixth neighboring cell, compared to the neighboring cells one to four (see Appendix D).

While in the previous discussion we observed a sequence of 60 fingerprints and tested whether the model is a good representation of a position, we now want to check whether the model gives a good prediction of a certain fingerprint. This is important since for the NMR based method (or pattern matching in general) only a single fingerprint is available for positioning. In Fig. 5.13 the cdf of the relative marginal likelihoods are shown for node 1 to node 5. The last two nodes are omitted, because we will use the non-informative prior. For node 1 only 4.5% of the tested fingerprints would have occurred with higher probability using the non-informative prior. In all other remaining cases the marginal likelihood is greater zero and, hence, in favor for the constructed prior. Table 5.5 shows the results for all nodes.

For node 3 and node 4 in Fig. 5.13 we find samples where the non-informative prior performs much better ($L_v - L_v^{ninf} \leq 0.2$) than the constructed prior. This explains the somehow "odd" cdf in this area. Unfortunately it was not possible to assign these samples to certain positions. This would have allowed a further improvement of the prior.

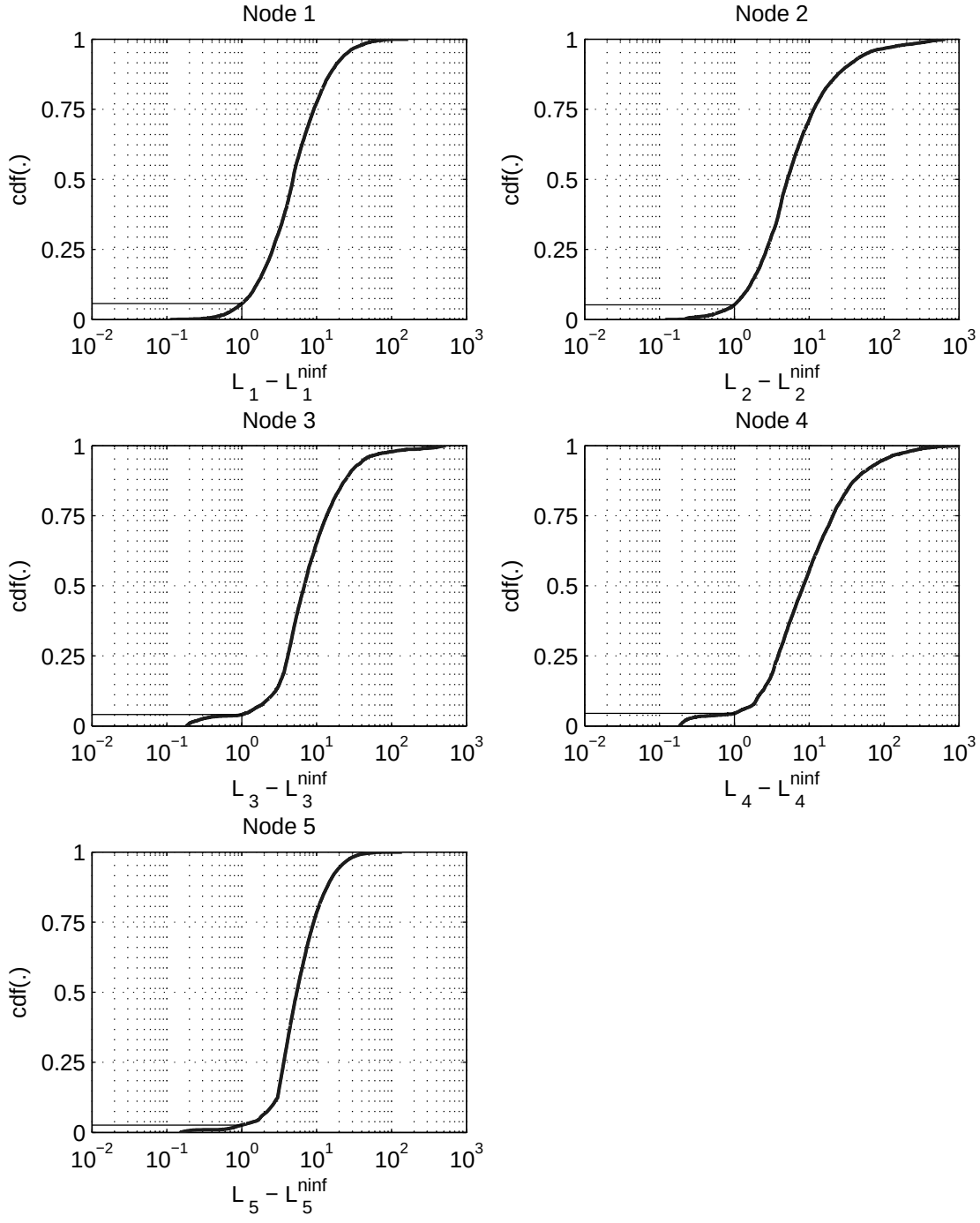


Figure 5.13: Cumulated distribution function of the relative marginal likelihood for node $1 - 5$.

Chapter 6

The Positioning Algorithm

In section 5.2 we have seen that in case of a *conjugated prior* the *marginal distribution* $p(\boldsymbol{D}|\mathcal{G})$ can be computed without solving an integral (see Equ. 5.18). We will use this feature to derive the localization algorithm in this section.

The method is based on *trained Bayesian networks* each representing one single position i. Once a mobile should be located, a snapshot of the received cell IDs is taken, containing several samples $\tilde{\boldsymbol{D}} = (\tilde{\boldsymbol{x}}(1), \ldots, \tilde{\boldsymbol{x}}(m))$, with $\tilde{\boldsymbol{x}} = (\tilde{x}_1, \ldots, \tilde{x}_V)$ ordered according to decreasing Rx levels. Here, $\tilde{x}_1$ represents the serving cell and $\tilde{x}_2$ to $\tilde{x}_7$ represent the neighboring cells. We ask all[1] models how likely each of them would yield the sample $\tilde{\boldsymbol{D}}$. This is equivalent with computing the marginal distribution $p(\tilde{\boldsymbol{D}}|\lambda^i)$, where λ^i represents the updated model. The maximization of the likelihood will finally give us the position estimate. A visual illustration of this algorithm is given at the end of this section. Before, however, we will proceed with the derivation.

Similar to the last chapter we omit the position index i for the further explanations for simplicity. All following equations refer to a single position i if not otherwise noted.

Recalling (4.21) we see that the likelihood can be split into a product of multivariate probabilities:

$$p(\tilde{\boldsymbol{D}}|\boldsymbol{\theta}) = L(\boldsymbol{\theta}) = \prod_{v=1}^{V} \prod_{p=1}^{p_v} \prod_{s=1}^{S} p(x_v = x^s | pa(X_v) = \boldsymbol{\rho}_{vp}, \boldsymbol{\theta}_{vp})^{\eta_{vps}}. \tag{6.1}$$

We can therefore compute the marginal distribution sequentially by updating only one node of the Bayesian network at a time and by finally constructing the product. We hence concentrate on the family of a single node v. Applying (5.9)

[1] To keep the explanation simple we currently refer to all models which have been trained. For the NMR based positioning method only a small subset based on the serving cell-ID and its sector is chosen.

for the marginal probability with $\boldsymbol{D} = \tilde{\boldsymbol{D}}$ for a single node and using the Dirichlet distribution as prior according to

$$p(\boldsymbol{\theta}_{vp}) = \mathcal{D}(\boldsymbol{\theta}_{vp}|\boldsymbol{\alpha}_{vps}) = \frac{\Gamma(\alpha_{vp}^{+})}{\prod_{s=1}^{S}\Gamma(\alpha_{vps})}\prod_{s=1}^{S}(\theta_{vps})^{\alpha_{vps}-1} \tag{6.2}$$

with $\alpha_{vp}^{+} = \sum_{s=1}^{S}\alpha_{vps}$ results in

$$p(\tilde{\boldsymbol{D}}_{\boldsymbol{v}}|\lambda) = \prod_{p=1}^{p_v}\int\left[\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}}\right]\frac{\Gamma(\alpha_{vp}^{+})}{\prod_{s=1}^{S}\Gamma(\alpha_{vps})}\prod_{s=1}^{S}(\theta_{vps})^{\alpha_{vps}-1}d\boldsymbol{\theta}. \tag{6.3}$$

Rearranging the equation yields

$$\begin{aligned} p(\tilde{\boldsymbol{D}}_{\boldsymbol{v}}|\lambda) &= \prod_{p=1}^{p_v}\frac{\Gamma(\alpha_{vp}^{+})}{\prod_{s=1}^{S}\Gamma(\alpha_{vps})}\int\prod_{s=1}^{S}(\theta_{vps})^{\eta_{vps}+\alpha_{vps}-1}d\boldsymbol{\theta} && (6.4)\\ &= \prod_{p=1}^{p_v}\left[\frac{\Gamma(\alpha_{vp}^{+})}{\prod_{s=1}^{S}\Gamma(\alpha_{vps})}\frac{\prod_{s=1}^{S}(\eta_{vps}+\alpha_{vps}-1)!}{\left[\sum_{s=1}^{S}(\eta_{vps}+\alpha_{vps}-1)+S-1\right]!}\right] && (6.5)\\ &= \prod_{p=1}^{p_v}\left[\frac{\Gamma(\alpha_{vp}^{+})}{\Gamma(\eta_{vp}^{+}+\alpha_{vp}^{+})}\prod_{s=1}^{S}\frac{\Gamma(\eta_{vps}+\alpha_{vps})}{\Gamma(\alpha_{vps})}\right] && (6.6) \end{aligned}$$

where we have used $\eta_{vp}^{+} = \sum_{s=1}^{S}\eta_{vps}$, $\alpha_{vp}^{+} = \sum_{s=1}^{S}\alpha_{vps}$ and $\Gamma(x) = (x-1)!$ for $x \in \mathbb{N}$.

Note that in (6.6) all information of the training-data is contained in the hyperparameters $\boldsymbol{\alpha}$ according to (4.26), where we found the updated posterior hyperparameters of the Bayesian network to be the sum of the prior hyperparameters and the multiplicity of $(X_v = x^s|pa(X_v) = \boldsymbol{\rho}_{vp})$ of the training data.

With

$$P(\tilde{\boldsymbol{D}}|\lambda) = \prod_{v=1}^{V}P(\tilde{\boldsymbol{D}}_{\boldsymbol{v}}|\lambda) \tag{6.7}$$

we have an equation for the marginal likelihood of a database $\tilde{\boldsymbol{D}}$ given the Bayesian network λ. By doing so for all networks λ^i and by taking the arg max according to

$$\hat{\lambda} = \arg\max_{\lambda^i}(p(\tilde{\boldsymbol{D}}, \lambda^i)), \tag{6.8}$$

we can find the network that approximates best the measured database $\tilde{\boldsymbol{D}}$. Since this Bayesian network is assigned to a certain position we have found an estimator for the position $i \in \mathcal{I}$.

6.1 Single Case Positioning

For the special case of a single measurement, $\tilde{\boldsymbol{D}} = \tilde{\boldsymbol{x}}(1)$ with $pa(\tilde{x}_v(1)) = \boldsymbol{\rho}_{v\tilde{p}}$, (6.6) can further be simplified. We use the fact that for every node v the multiplicity η_{vps} equals 0 except for the one single test case according to

$$\eta_{vps} = \begin{cases} 1 & \text{if } X_v = \tilde{x}_v; pa(X_v) = \boldsymbol{\rho}_{v\tilde{p}} \\ 0 & \text{otherwise.} \end{cases} \tag{6.9}$$

Equation (6.6) then yields

$$P(\tilde{\boldsymbol{D}}_{\boldsymbol{v}}|\lambda) = \prod_{p=1}^{p_v; p\neq\tilde{p}} \left[\underbrace{\frac{\Gamma(\alpha^+_{vp})}{\Gamma(0+\alpha^+_{vp})}}_{1} \prod_{s=1}^{S} \underbrace{\frac{\Gamma(0+\alpha_{vps})}{\Gamma(\alpha_{vps})}}_{1} \right] \cdot \tag{6.10}$$

$$\left[\frac{\Gamma(\alpha^+_{v\tilde{p}})}{\Gamma(1+\alpha^+_{v\tilde{p}})} \frac{\Gamma(1+\alpha_{v\tilde{p}\tilde{s}})}{\Gamma(\alpha_{v\tilde{p}\tilde{s}})} \right] \tag{6.11}$$

$$= \frac{\Gamma(\alpha^+_{v\tilde{p}})}{\Gamma(\alpha^+_{v\tilde{p}})\alpha^+_{v\tilde{p}}} \frac{\Gamma(\alpha_{v\tilde{p}\tilde{s}})\alpha_{v\tilde{p}\tilde{s}}}{\Gamma(\alpha_{v\tilde{p}\tilde{s}})} \tag{6.12}$$

$$= \frac{\alpha_{v\tilde{p}\tilde{s}}}{\alpha^+_{v\tilde{p}}} \tag{6.13}$$

$$= \alpha_{v\tilde{p}\tilde{s}}. \tag{6.14}$$

Since we interpreted the hyperparameters $\boldsymbol{\alpha}$ as pseudo counts, the computation of the marginal likelihood $P(\tilde{\boldsymbol{D}}_{\boldsymbol{v}}|\lambda)$ is simply a lookup in the corresponding conditional probability table storing the number of "prior observations". All computations required for the training of the Bayesian network are made off-line and the computation of the Gamma functions is completely omitted.

According to (6.7) we can now put the factors computed by (6.14) together and receive

$$\hat{i}_{NMR} = \arg\max_i \prod_{v=1}^{V} \alpha^i_{v\tilde{p}\tilde{s}} \tag{6.15}$$

for the final single case position estimator $\hat{i}_{NMR}$.

A summarization of the positioning algorithm is given in Fig. 6.1. During the off-line Phase I all Bayesian networks are trained with measured data. This phase ends with a rearrangement of the computed Bayesian model's data to efficiently search for fingerprints during the on-line Phase II. The serving cell ID and a single fingerprint serve as input data for the upcoming localization of a mobile.

We determine the marginal likelihood according to (6.14) using the serving cell ID to limit the search area and thus lower the search time. A maximization according to (6.15) finally delivers the position estimate.

6.2 Performance Evaluation

To evaluate the performance of the algorithm derived in the previous section we apply a multistage approach. In the following subsections we show a lower bound for the localization error resulting from the difference between the model and the true reality. We then discuss the relation between the number of training samples and the positioning error, which will allow the determination of an optimal training set size for positioning in urban areas. Given the training set, we will compare the discriminating power of $\boldsymbol{G}_{NMR}$ with several reference models given in literature. The final stage of the performance evaluation will show the positioning error of the method in relation to known pattern matching methods and also to GPS.

6.2.1 Model inadequacy

In section 5.3 we found the Bayesian model $\mathcal{G}_{NMR}$ based on a set of training data, assuming that the training data represents full information. In such a case the non informal prior is optimal and a fingerprint to be localized is contained in at least one training set. Consequently, at least one model returns a marginal log likelihood greater ϵ allowing for the positioning. For the database $\boldsymbol{D}$ containing $m = 300$ samples per position a lower bound for the positioning error is empirically found to be $e_{min} = 17.3m$. This error results from the multiplicity of the same fingerprints at various positions and thus denotes the "inadequacy" of the model to represent the true reality. A further improvement would only be possible if additional information would be added. This could for instance be a *sequence of fingerprints*. In this thesis, however, we will not further discuss this topic. For an approach using three consecutive fingerprints refer to [38] and [39].

6.2.2 Accuracy vs. training set size

To examine the performance of the NMR based positioning we apply the method to an urban environment, containing several streets in the city of Vienna. A map is shown in Fig. 5.4. For the training of the models we use a subset $\boldsymbol{D}_m$ of all measured fingerprints $\boldsymbol{D}$. For the performance test 6000 uniformly distributed test patterns $\tilde{\boldsymbol{X}} = (\tilde{X}_v : v = 1, \ldots, V)$ over the trial area are considered, equally

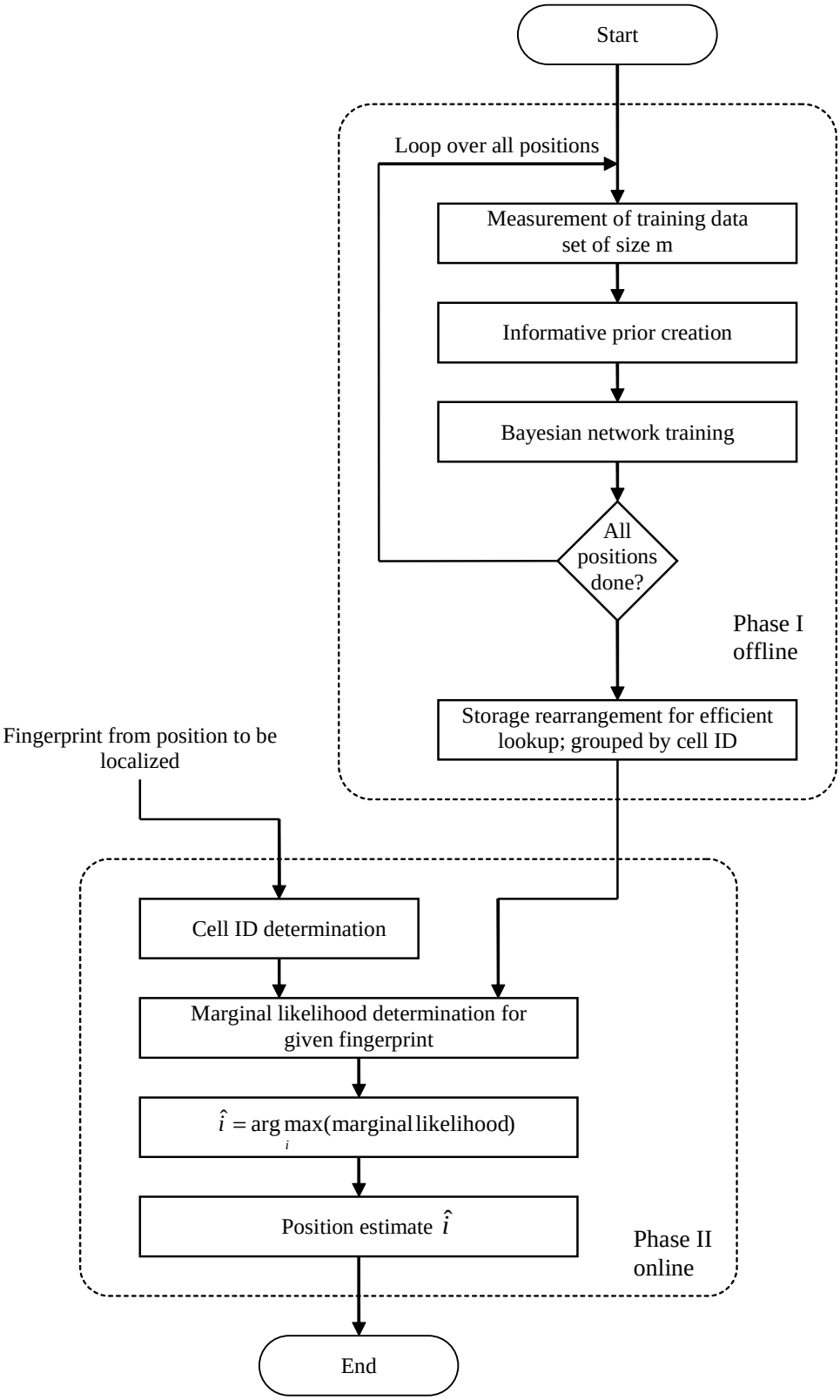


Figure 6.1: Bayesian network based positioning algorithm. Phase I can completely be computed off line. Only Phase II is required to be computed for every position estimate separately.

spaced in a distance of 5 meters. The time period between measuring the training and test patterns was at least 2 days.

Theoretic training set size

Analysis of the database $\boldsymbol{D}$ indicates the cardinality of the nodes's alphabet $|\mathcal{S}| < 28$ in 95% of the measured cases. If we assume no multiple occurrence of a certain symbol within a fingerprint pattern, this yields $(V!)*(|\mathcal{S}|-V+1) = 110880$ different possible fingerprints. In the worst case, the probability density of all samples is uniform, which would require a training set size of $m = 110880$ to represent $0.95^V \approx 70\%$ of the measured cases.

Measured training set size

In Section 6.3 we will analyze the long-term behavior of the measured fingerprints and define a long-term time constant representing the fingerprints' validity over time. By assuming that at least 50% of the fingerprints that we want to localize are still valid, a mean of 371 different fingerprints is measured at all tested sites.

However, simulations with measurements show that an even lower number of different samples is sufficient for feasible accuracy results. Fig. 6.2 depicts the dependence of the accuracy error of the number of training samples in urban environment. With $m = 10$ training samples we achieve a mean positioning error of less than 94 meters in 95% of all cases. A further increase of the training samples' number shows only little improvement.

6.2.3 Discriminating power of the model

To evaluate the discriminating power of the Bayesian networks for different positions we use a measure of how probable a set of fingerprints, taken from position i, occur at all the other positions. We apply a logarithmic score, the marginal logarithmic probability (or negative logarithmic penalty [14]), according to

$$\begin{aligned} L = \log \mathcal{L} &= \log p(\tilde{\boldsymbol{X}}^i | \mathcal{G}_{NMR}) \\ &= \log \prod_{v=1}^{V} \prod_{p=1}^{p_v} \left[\frac{\Gamma(\sum_{s=1}^{S} \alpha_{vps})}{\Gamma(\sum_{s=1}^{S} (\alpha_{vps} + \eta_{vps}))} \prod_{s=1}^{S} \frac{\Gamma(\alpha_{vps} + \eta_{vps})}{\Gamma(\alpha_{vps})} \right], \end{aligned} \tag{6.16}$$

which we have derived in (6.6) and where $\tilde{\boldsymbol{X}}^i \in \tilde{\boldsymbol{X}}$ denotes the subset of fingerprints at the position i. The hyperparameters $\boldsymbol{\alpha}$ denote the fully updated Bayesian network.

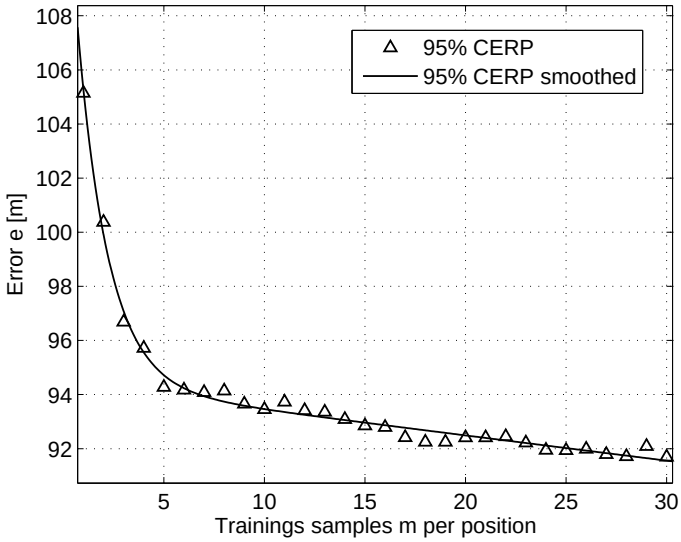


Figure 6.2: Mean accuracy vs. number of training samples (urban area, 1800 test cases). Using a training set of size $m = 10$ achieves already a mean error $e < 94m$.

Figure 6.3 shows the relative marginal logarithmic probability $L^{ref} - L$ for Vienna. This measure denotes the difference of the model scores for a Bayesian network tested with fingerprints taken from the true position (training fingerprints and test fingerprints at the same position) and fingerprints taken at distant positions respectively. The solid curve represents the NMR based method, with the optimized model structure $\mathcal{G}_{NMR}$. The probability for the reference model to emit the test fingerprint is, on average, 600 times higher than for the other models in a distance of 10 meters. The same test is shown for models where the dependencies between the nodes were omitted (dashed curve). The discriminating power is less, but larger than the discriminating power for the standard pattern matching reference model (dotted curve) as e.g. proposed by [41]. This model uses the absolute received power level of all neighboring cells. The serving cell's power level is omitted because of power control.

We use the common 2-dimensional circular error probability (CERP) as a measure for the accuracy. It is defined as the probability of multiple location measurements being inside a virtual circle with radius r. For example, 95% CERP within $r = 56$ meters means that 95% of the location measurements are within 56 meters from the true location. The NMR based localization method achieves a 67% CERP of 13.75m and a 95% CERP of 93.45 meters in the city of Vienna (solid curve in Fig. 6.4). This is fully compliant with the FCC E911 requirements [3], which indicate a 67% CERP of 100m and a 95% CERP of 300 meters for network

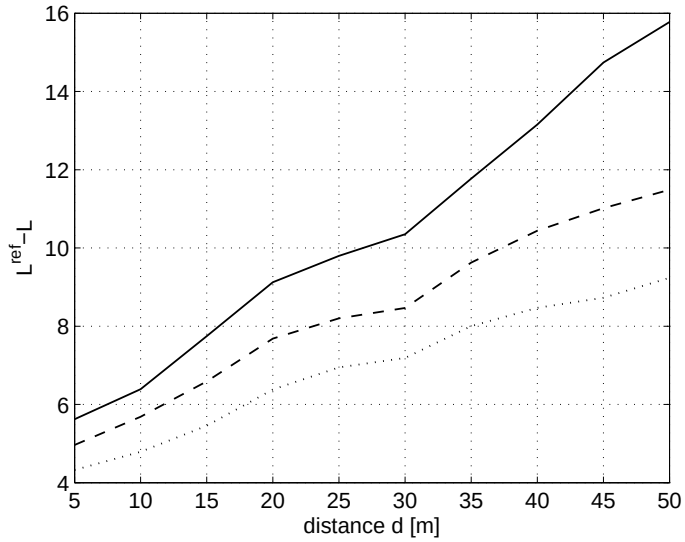


Figure 6.3: Relative marginal probability of the Bayesian model with structure $\mathcal{G}_{NMR}$ for fingerprints at the true position and fingerprints at positions in a distance d.

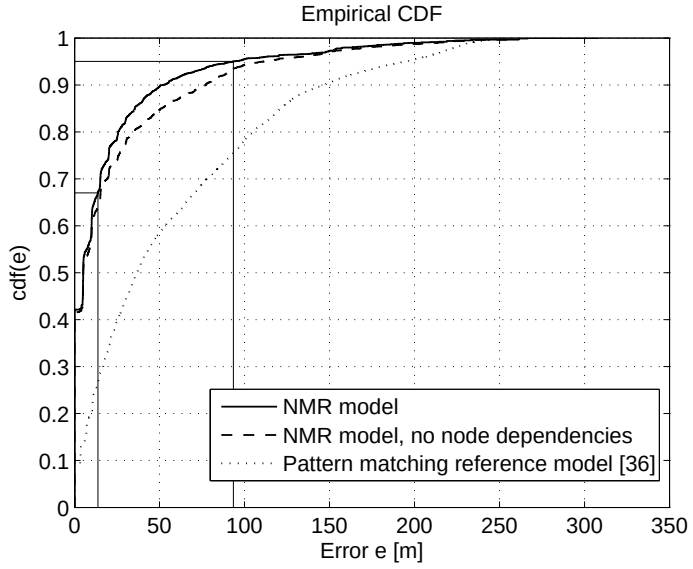


Figure 6.4: Accuracy of different pattern matching methods in urban areas.

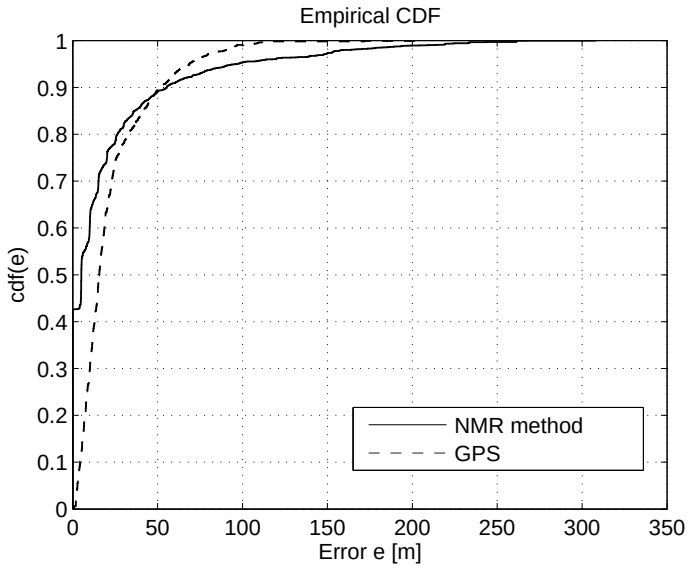


Figure 6.5: Accuracy of the NMR based localization method vs. accuracy of GPS. GPS had a reliability of only 53% , while the NMR based method shows a reliability of 100%.

based methods. Omitting the conditional dependencies between the nodes lead to a 95% CERP of 111.4m. The standard pattern matching model achieves a 95% CERP of 196.2m in this environment.

In Fig. 6.5 the NMR based localization accuracy (solid) is compared with the GPS accuracy of the same area (dashed). Note that only successful GPS fixes were taken into account. The GPS reliability[2] was 53%, while the NMR based localization method shows a reliability of 100%. We also mention that the NMR based localization method's accuracy is upper bounded by the radius of the cell size in cellular networks, which is typically several 100 meter in urban areas in GSM.

6.3 Long-Term Stability

For an efficient usage of pattern matching based location methods, a long validity of the database is required. It is desirable that the data do not depend on time. To describe the time variation of the received power levels we collect fingerprints over a period of several days at several sites in urban environment. See Appendix D for a complete list of the sites as well as a description of the measurements.

For the NMR based pattern matching method it is in particular important

[2]We use the definition of reliability given in [66] which is simply a ratio of successful positioning attempts out of all attempts made.

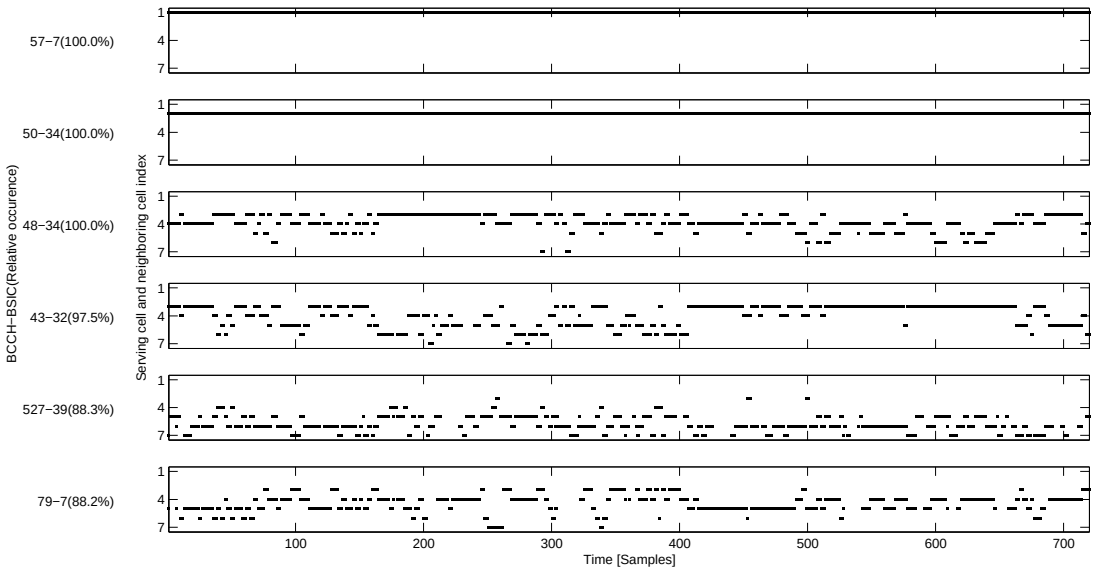


Figure 6.6: Broadcast control channel of serving and neighboring cells in the time domain. Sample frequency 0.1 Hz.

that there is little or no relative change among the received power levels of all received base stations. The received fingerprint shall remain constant over time.

To measure the long-term stability we observe a sequence of training-patterns $\boldsymbol{D'} = \{\boldsymbol{X'}(1), \boldsymbol{X'}(2), \ldots, \boldsymbol{X'}(k)\}$ at time $t < t_0$ and compare it with observed patterns $\boldsymbol{D} = \{\boldsymbol{X}(1), \boldsymbol{X}(2), \ldots, \boldsymbol{X}(l)\}$ at time $t_0 < t < t_0 + \Delta t$.

We define

$$q(\Delta t) = \frac{n_{\Delta t}}{l}, \tag{6.17}$$

where $n_{\Delta t}$ is the number of elements in $\{\boldsymbol{x}(j) \in \boldsymbol{D} | \boldsymbol{x}(j) \in \boldsymbol{D'}; j = 1 \ldots l\}$. $q(\Delta t) \in [0, 1]$ then accounts to a measure of recognizable patterns within the received sequence of patterns. To get a notion about how long the patterns can be considered constant we define, similar to the long-term investigations of mobile channels, a long-term time constant according to:

Definition: *The long-term time constant τ_{LT} is the time difference Δt for which the measure $q(\Delta t)$ is less than 0.5 in 90% of all cases.*

This time constant corresponds to the time Δt where 90% of the cdfs measured from different samples in the database crosses[3] the point $(0.5; 90\%)$. This definition seem reasonable, albeit it is initially arbitrary. If we use training samples

[3]The definition does not require an exact crossing, it only requires a value less than (0.5) in 90% of all tested cases.

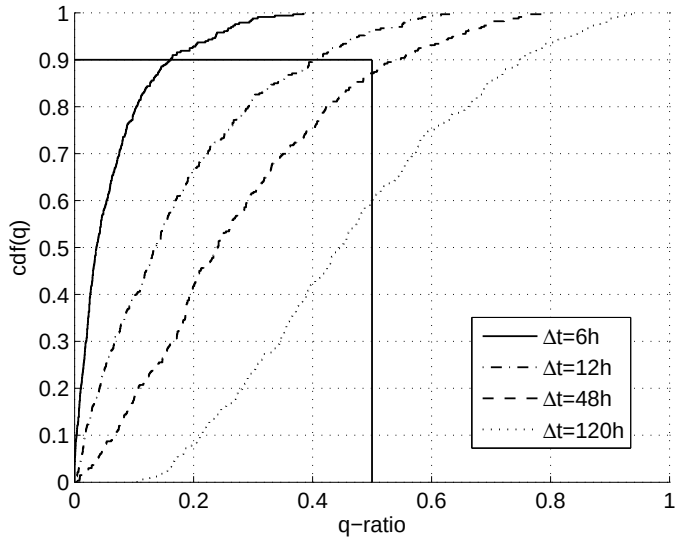


Figure 6.7: Empirical cdf of the q-ratio in urban environment (k = 3000).

for the models taken at time $t < t_0$ still 50% of all samples would be accurately recognizable before $t+\Delta t$ (in 90% of all cases). Other definitions of the long-term time constant change the results quantitatively, although not qualitatively.

After having defined the measures, we will now apply the described evaluation mechanism to all measured patterns. An off-the-shelf Siemens S55 mobile phone was used for the measurements. The required data (Broadcast Control Channel ID, Received Power Level and Base Station Identity Code) were read from the standardized Network Measurement Report (NMR) via the serial interface. The record time was several days for several urban sites. One example is shown in Fig. 6.6. The plot shows six subplots, one for each received cell ID at this site ordered according to its relative occurrence. The number depicted in parenthesis denote the Broadcast Control Channel ID and the Base Station Identity Code as well as the probability of occurrence during the measurements. The y-axis of each sub-plot corresponds to the rank of the cell ID within the current fingerprint, where 1 denotes the serving cell, 2 the strongest neighboring cell, etc. The x-axis shows the sample index, measured with a sampling frequency of 0.1 Hz. We find that the serving cell remains the same in 98% of all cases in this particular example. Note, however, that only the first 2 hours of the measurement period are shown. For the full graph, a description and a discussion of all sites the reader is referred to Appendix D.

An aggregated figure for urban environment is given in Fig. 6.7 where the cdfs of q are plotted.

We highlighted the $(0.1; 50\%)$ point in the graphs, which determines the long-term constant τ_{LT}. For urban environment τ_{LT} is 42 hours. After 120 hours the successful recognition of half of the tested patterns decreases to 60%.

In urban areas the propagation is in particular characterized by the surrounding buildings, i.e. static scatterers and NLOS. The variation of the pattern results from additional time variant scatterers, i.e. cars which cause appearing and disappearing paths. The decreasing value of q for larger Δt indicates that a short observation cannot properly describe a certain position. The longer a reference database is used without update, the larger the number of unknown patterns grows. A continuous update of the database will thus be required for all pattern matching methods.

Chapter 7

Conclusions

The research reported in this thesis addresses fingerprint positioning and introduces the network measurement report based positioning method. The research is closely tied to currently available cellular network technologies, most importantly GSM, which makes the results usable and applicable in practice. All results found, have been published at conferences and as journal contributions.

This chapter now summarizes the main results. It concludes with a discussion and a proposal for applicable positioning technologies and methods for future personal positioning.

7.1 Main Results

The research presented in this thesis introduces a novel localization method which is based on pattern matching. We have used Bayesian networks to represent a position by describing dependencies between the different measured cell IDs at a single position. The serving cell and ordered neighboring cells, according to the received power levels, define hereby a fingerprint pattern.

The resulting ambiguity of the pattern, caused by propagation effects like small- and large-scale fading and the time variance of the environment, are modelled by the capability of Bayesian networks to encode uncertainty.

The structure of the Bayesian network is optimized for urban areas in accordance with the requirements for pattern matching positioning and is compared with other models that are currently available.

We further present a positioning algorithm based on a comparison of the mobile's current fingerprint with all models in the expected target area (the area of the serving cell). The best match defines the estimated position. We show that the complexity of this algorithm can be reduced to a single maximization of

an off-line computed database data vector, if the positioning is based on a single fingerprint.

From this point of view, the NMR based method can be seen as an enhanced cell ID method.

It is further shown that a low number of reference samples is sufficient for the training of the Bayesian network as well as for the creation of a prior.

The prior is thereby constructed from "expert knowledge". Instead of counting past samples we use a certain circled area around the current position and count the number of equal realizations within this area. Informally spoken, we deduct the prior hyperparameters of the current position from knowledge about neighboring positions. We empirically show that the impact of the prior distribution, compared with the non-informative Jeffreys prior, results in a better fit of the model in over 90% of the tested cases.

Trial results in the urban city of Vienna are presented, meeting all requirements of the FCC E911 regulations. The 95% CERP is found to be 93.45 meters with a reliability of 100%. The 67% CERP is 13.75 meters. The accuracy results are compared with currently available reference pattern matching models and GPS. While the reference pattern matching models perform worse, GPS shows a similar accuracy to the NMR based method, but a weaker reliability of 63%. The basic limitation of GPS in urban areas, its necessity of line-of-sight to the signal transmitters is shown to be the strength of the NMR based localization method.

Long-term measurement results are presented to prove the long-term stationarity of the fingerprints, showing a time constant of 42 hours. During that period, it is still possible to determine 50% of all received fingerprints at the measurement site in 90% of all cases.

7.2 Future Possibilities for Personal Positioning

Personal positioning has suffered from technological limitations during the past years. Latest research, however, has shown that most limitations have been overcome. In particular the FCC's regulatory pressure in the United States has lead to immense effort to improve all types of positioning methods. While terrestrial time based localization methods currently loose importance in outdoor environments, satellite based positioning (GPS and Galileo) gain importance, because of better accuracy and availability. In indoor environments, signal level based positioning is favored because of its cheap implementation, especially when combined with all types of wireless local area network (LAN) standards. Ultra wideband time based positioning allows higher accuracy, but requires additional hardware.

Currently, however, none of the technologies available can provide satisfactory

service for all environments and applications. The best solution is therefore still dependent on the situation and might also happen to be a combination of different methods. It thus remains to be seen which method will find its way to a large number of handheld devices and thus becomes the "standard" for personal positioning.

Appendix A

Dirichlet Distribution

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ be a point in the $(k-1)$ simplex with $0 < \theta_i < 1$ and $\sum_{i=1}^{k} \theta_i = 1$.

Let further $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ be a set of parameters with $\alpha_i > 0$. The Dirichlet density is then defined as

$$\mathcal{D}(\boldsymbol{\theta}|\alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1} = c \prod_{i=1}^{k} \theta_i^{\alpha_i - 1} \tag{A.1}$$

with $\alpha^+ := \sum_{i=1}^{k} \alpha_i$. The non-negative quantities α_i determining the specific member of the Dirichlet family are hyperparameters.

Any aggregation of a subset of the Dirichlet variables $\boldsymbol{\theta}$ results again in a Dirichlet distribution with the corresponding aggregation of the parameters according to

$$(\theta_1, \ldots, \theta_q) \sim \mathcal{D}(\alpha_1, \ldots, \alpha_q, \sum_{i=q+1}^{k} \alpha_i). \tag{A.2}$$

Proof:

With the assignment of $\theta_{k-2}^+ = \sum_{i=1}^{k-2} \theta_i$ we write

$$\begin{aligned}
\int \mathcal{D}(\theta_1, \ldots, \theta_{k-1}) d_{\theta_{k-1}} &= c \int_0^{1-\theta_{k-2}^+} \theta_1^{\alpha_1 - 1} \ldots \theta_{k-1}^{\alpha_{k-1} - 1} \left(1 - \sum_{i=1}^{k-1} \theta_i\right)^{\alpha_k - 1} d\theta_{k-1} \\
&= c\theta_1^{\alpha_1 - 1} \ldots \theta_{k-2}^{\alpha_{k-2} - 1} \int_0^{1-\theta_{k-2}^+} \theta_{k-1}^{\alpha_{k-1} - 1} \left(\frac{1 - \theta_{k-2}^+ - \theta_{k-1}}{1 - \theta_{k-2}^+}(1 - \theta_{k-2}^+)\right)^{\alpha_k - 1} d\theta_{k-1} \\
&= c\theta_1^{\alpha_1 - 1} \ldots \theta_{k-2}^{\alpha_{k-2} - 1} \int_0^{1-\theta_{k-2}^+} \theta_{k-1}^{\alpha_{k-1} - 1} \left(1 - \frac{\theta_{k-1}}{1 - \theta_{k-2}^+}\right)^{\alpha_k - 1} (1 - \theta_{k-2}^+)^{\alpha_k - 1} d\theta_{k-1}.
\end{aligned}$$

Assigning $u = \frac{\theta_{k-1}}{1-\theta^+_{k-2}}$ and $\frac{d\theta_{k-1}}{du} = 1 - \theta^+_{k-2}$ yield

$$\begin{aligned} &= c\theta_1^{\alpha_1-1} \dots \theta_{k-2}^{\alpha_{k-2}-1}(1-\theta^+_{k-2})^{\alpha_k} \int_0^1 u^{\alpha_{k-1}-1}(1-\theta^+_{k-2})^{\alpha_{k-1}-1}(1-u)^{\alpha_k-1}du \\ &= c\theta_1^{\alpha_1-1} \dots \theta_{k-2}^{\alpha_{k-2}-1}(1-\theta^+_{k-2})^{\alpha_{k-1}+\alpha_k-1} \int_0^1 u^{\alpha_{k-1}-1}(1-u)^{\alpha_k-1}du. \end{aligned} \tag{A.3}$$

The integral in (A.3) is Dirichlet's integral and has the following solution [70]

$$\int, \dots, \int \prod_{i=1}^{k} \theta_i^{n_i} d\theta_1 \dots d\theta_k = \frac{\prod_{i=1}^k n_i!}{((\sum_{i=1}^k \alpha_i) + k - 1)!}. \tag{A.4}$$

By applying (A.4) we receive from (A.3) for the marginal probability

$$\begin{aligned} p(\theta_1, \dots, \theta_{k-2}) &= \int p(\theta_1, \dots, \theta_{k-1}) d\theta_{k-1} \\ &= \tilde{c}\theta_1^{\alpha_1-1} \dots \theta_{k-2}^{\alpha_{k-2}-1}\big(1-\theta^+_{k-2}\big)^{\alpha_{k-1}+\alpha_k-1} \\ &= \mathcal{D}(\alpha_1, \dots, \alpha_{k-2}, \alpha_{k-2}+\alpha_k) \sim (\theta_1, \dots, \theta_{k-2}). \end{aligned} \tag{A.5}$$

It is easily shown that repeated application of these proof leads to (A.2).

The marginal probability of a certain parameter θ_i is then a Dirichlet distribution $Dir(\alpha_i, \sum_{j=1;j\neq i}^k \alpha_j)$ with the density

$$(\theta_i) \sim c\theta_i^{\alpha_i-1}(1-\theta_i)^{\sum_{j=1;j\neq i}^k \alpha_j} = Be(\alpha_i, \sum_{j=1;j\neq i}^{k}) \tag{A.6}$$

which is equal the density of the Beta distribution.

The expectation $\mathbb{E}[\theta_i|\boldsymbol{\alpha}]$ is then

$$\mathbb{E}[\theta_i|\boldsymbol{\alpha}] = \frac{\alpha_i}{\alpha_i + \sum_{j=1;j\neq i}^k \alpha_j} = \frac{\alpha_i}{\alpha^+}. \tag{A.7}$$

The variance of θ_i is given by

$$\begin{aligned} var(\theta_i) &= \frac{\alpha_i(\sum_{j=1;j\neq i}^k \alpha_j)}{(\alpha_i + \sum_{j=1;j\neq i}^k \alpha_j)^2(\alpha_i + \sum_{j=1;j\neq i}^k \alpha_j^+ 1)} \\ &= \frac{\alpha_i(\alpha^+ - \alpha_i)}{(\alpha^+)^2(\alpha^+ - 1)}. \end{aligned} \tag{A.8}$$

Appendix B

Multinomial Process

The following section is a brief description of the multinomial process as we use it throughout this thesis. For a more thorough discussion refer e.g. to [14].

Assume a discrete random variable X that can take on one of a set of k mutually exclusive and exhaustive values $x_1, x_2, \ldots, x_k$ with probabilities $\theta \equiv (\theta_1, \ldots, \theta_k)$. Furthermore assume $\theta_i > 0$ for all i and the sum of all θ_i equals 1. That is,

$$P(X = x_i|\theta) = \theta_i \quad i = 1, \ldots, k. \tag{B.1}$$

Suppose a sequence d of n independent observations of X. We denote n_i the number of observations of type x_i for $i = 1, \ldots, k$. The likelihood is then given by

$$p(d|\theta) = \prod_{i=1}^{k} \theta_i^{n_i}. \tag{B.2}$$

If one does not keep the full information of d, but instead only keeps the numbers n_i and the observation scheme is non-informative, e.g. d contains all information available when deciding to stop the observation, one gets for the likelihood

$$p(n_1, \ldots, n_k|\theta) = \frac{n!}{\prod_{i=1}^{k} n_i!} \prod_{i=1}^{k} \theta_i^{n_i} \tag{B.3}$$

for n being fixed in advance.

Now let the prior uncertainty about θ be the Dirichlet distribution given by

$$\mathcal{D}(\theta|\alpha_1, \ldots, \alpha_k) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1} \tag{B.4}$$

with hyperparameters $\alpha = (\alpha_1, \ldots, \alpha_k)$, $\alpha_+ := \sum_{i=1}^k \alpha_i$ and each $\alpha_i > 0$.

Note that the Dirichlet distribution is conjugate for the multinomial sampling process. The posterior distribution for θ given the sufficient statistic $(n_1, \ldots, n_k)$ of an observed sequence d is according to $p(\theta|d) \propto p(d|\theta)p(\theta)$

$$p(\theta|d) \propto \prod_{i=1}^{k} \theta_i^{n_i+\alpha_i-1}. \tag{B.5}$$

The posterior distribution is again Dirichlet $\mathcal{D}(\alpha_1 + n_1, \ldots, \alpha_k + n_k)$ and can thus be obtained by simply updating each of the hyperparameters, by adding the number of cases observed to yield the associated outcome.

For prediction or Bayesian model comparison, the marginal probability of the data can be used, given by

$$\begin{aligned} p(d) &= \int_\Theta p(d|\theta)p(\theta)d\theta \\ &= \frac{\Gamma(\alpha_+)}{\prod_{i=1}^k \Gamma(\alpha_i)} \int_\Theta \prod_{i=1}^{k} \theta_i^{n_i} \prod_{i=1}^{k} \theta_i^{\alpha_i-1} \\ &= \int_\Theta \prod_{i=1}^{k} \theta_i^{n_i+\alpha_i-1}. \end{aligned} \tag{B.6}$$

By applying the solution of Dirichlet's integral (A.4) we receive from (B.6) for the marginal probability

$$\begin{aligned} p(d) &= \frac{\Gamma(\alpha_+)}{\prod_{i=1}^k \Gamma(\alpha_i)} \frac{\prod_{i=1}^k (n_i + \alpha_i - 1)!}{(\alpha_+ + n - 1)!} \\ &= \frac{\Gamma(\alpha+)}{\prod_{i=1}^k \Gamma(\alpha_i)} \frac{\prod_{i=1}^k \Gamma(n_i + \alpha_i)}{\Gamma(\alpha_+ + n)}. \end{aligned} \tag{B.7}$$

Appendix C

Parent-Child monitor

This section describes the parent-child monitor as introduced by [61] and [15]. For a brief introduction we refer the reader also to [14].

Let $\boldsymbol{X} = (X(m) : m = 1, \ldots, M)$ denote a sequence of discrete random variables. Let $p(X_v|pa(X_v))$ denote the conditional probability specification of each node v of a Bayesian model, where we associate a set of parent-child monitors. These monitors detect discrepancies between prior beliefs in X_v and the observed value x_v given a certain parent configuration $pa(X_v) = \boldsymbol{\rho}_{vp}$. Now let $p_m(.)$ denote the conditional probability distribution for X_v given its parents after $m-1$ cases have been observed. With observation x_m we associate a logarithmic score $L_{m,v}$ given by

$$L_{m,v} = -\log p_m(x_m|pa(X_m) = \boldsymbol{\rho}_{vp}) \tag{C.1}$$

which increments the parent-child monitor for the parent configuration $\boldsymbol{\rho}_{vp}$ and node v.

By cumulating over the sequence of M cases, the total logarithmic score is given by $\sum_{m=1}^{M} L_{m,v}$. If we coherently update $p_m(.)$ on the basis of the joint probability distribution of $\boldsymbol{X}$, then $p_m(x_m|pa(X_m) = \boldsymbol{\rho}_{vp}) = p(x_m|x_1, \ldots, x_{m-1}, pa(X_m) = \boldsymbol{\rho}_{vp})$ and thus

$$\begin{aligned} L_v &= -\log \prod_{m=1}^{M} p_m(x_m|pa(X_m) = \boldsymbol{\rho}_{vp}) \\ &= -\log \prod_{m=1}^{M} p(x_m|x_1, \ldots, x_{m-1}, pa(X_m) = \boldsymbol{\rho}_{vp}) \\ &= -\log p(x_1, \ldots, x_m|pa(X_m) = \boldsymbol{\rho}_{vp}). \end{aligned} \tag{C.2}$$

This is just the negative logarithm of the joint probability of all the data

of one node v observed given its parents. That is, the logarithmic score, L_v is invariant to the order in which the data was observed.

Appendix D

Long-Term Stability

For an efficient usage of a pattern matching based location method, a long validity of the database is required. It is desirable that the data does not depend on time, and if (e.g. because of construction work), a change can be discovered automatically to allow an easy update of the database. To describe the time variance of the Rx-levels, measurements over a time period of about 10 days have been performed at several sites (D) in Vienna (urban) and around Vienna (suburban). The following sections describe the measurement sites in more detail.

D.1 Measurement Set-Up

An off-the-shelf Siemens S55 mobile phone equipped with an external mini magnet-base antenna is used for the measurements. The required data (Tab. D.1) is read from the standardized NMR via the serial interface. To protect the measurement equipment from environmental impacts it is stored in a fibre-box for the duration of measurement as shown in Fig. D.1. The antenna is placed at a height of 1.5 meters above ground. A 12V/144Ah battery serves as power supply at the outdoor sites, avoiding a recharge process during the time of measurement. At

Site	Section	Environment
Aspern	D.2.1	urban
Hoher Markt	D.2.2	urban
Hörlgasse Türkenstrasse	D.2.3	urban
Hörlgasse Wasagasse	D.2.4	urban
Neuer Markt	D.2.5	urban
Soldanellenweg	D.2.6	urban
Templergasse	D.2.7	suburban

Table D.1: Measurement sites for long-term measurements

(a) Grit container, where the measurement equipment was stored during measurements.

(b) Outdoor measurement setup

Figure D.1: Measurement equipment for long-term measurements.

indoor sites additional AC/DC adaptors were used for supplying the equipment from the 230V/50Hz mains.

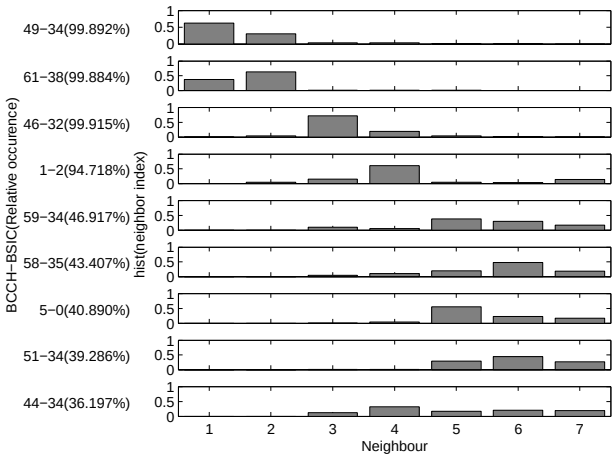
D.2 Measurement Results

Sections D.2.1-D.2.7 show the results of the measurement campaign at several urban and suburban sites. Each section is divided in five figures and a table giving additional information about the site. Figure 1,2 and 5 of each section show nine subplots (rows) corresponding to the received base stations identified by their BCCH and their BSIC. The received base stations are ordered according to their probability of occurrence (shown in parenthesis) at the measurement site. For instance, "(57-7(99.988%)" denotes the base station with BCCH=57 and BSIC=7, which occurs in 99.998% of all measured samples. Figure 1 of each section shows the pmf of the neighbor index. For the site Hoher Markt, for example, a quasi-stationary pattern for the serving cell (BCCH=57) and the strongest neighbor (BCCH=50) is shown (Fig. D.2.7). Their probability of occurrence within the measured samples is 99.8% and 93.2% respectively. Figure 2 of each section shows the pmf of the Rx-levels. Figure 3 and 4 show a map and a photo of the measurement site. The actual site is marked with a black circle. Figure 5 of each section depicts the cell ID's current neighbor index (y-axis) over all collected samples (x-axis). At site Hoher Markt (Fig. D.2.11) the quasi stationary pattern of the serving cell and the strongest neighbor is nicely observable. Note, the fast change for the neighbor index of the cell with ID 50 between sample $20k$ and $30k$. Since the time corresponds with the weekend, this

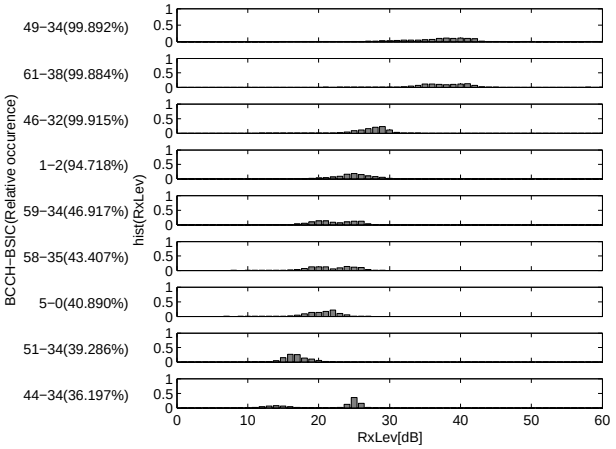
effect might be caused by a blocking obstacle, parked in one of the free parking lots at the weekend. A similar effect, however, has not been measured at any other site used for the long-term measurements.

In Section D.2.7 the measurement results for the suburban site Templergasse in Mödling are shown. The receiver was placed in the garden of a house surrounded only by streets with medium traffic volume. Noticeably is Fig. D.2.32 with the high probability of occurrence for the first six received base stations. The cell with the BCCH IDs 43,56,54,51,41 and 131 are practically present in every measured sample at this position. Hence, the time variance at this position results mainly from the change of their neighbor indices.

D.2.1 Aspern

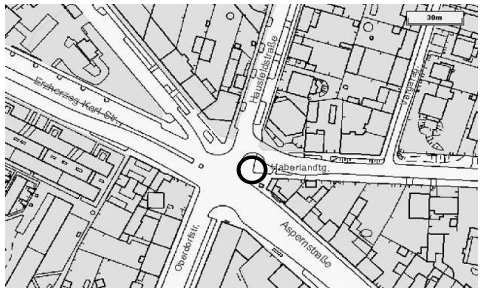


D.2.2: Pmf of neighboring index.



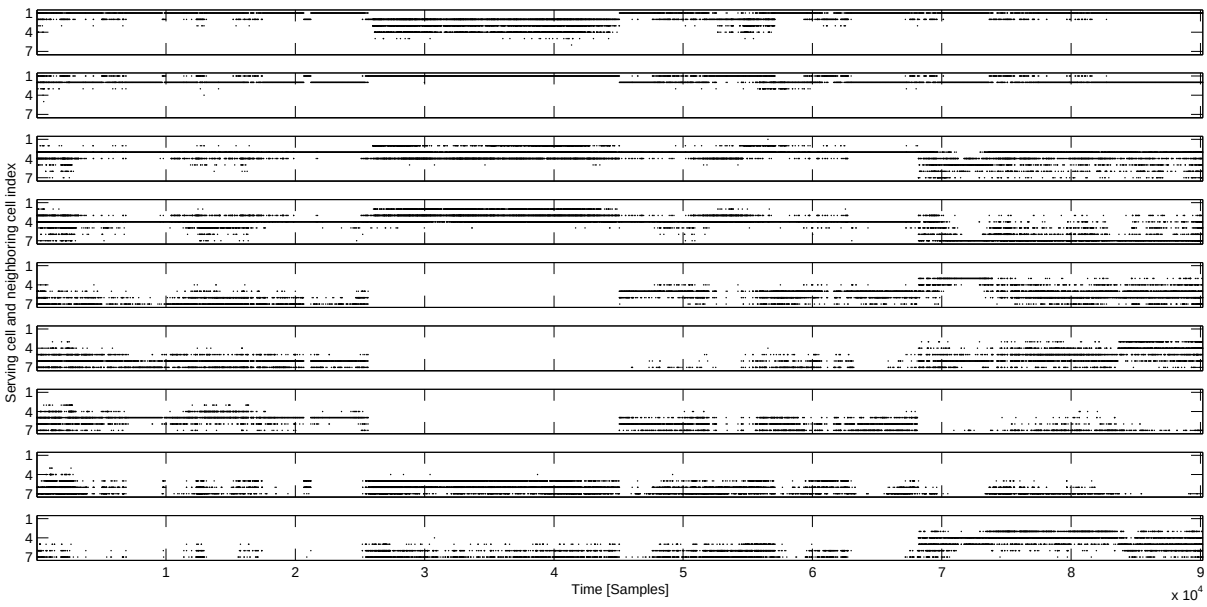
D.2.3: Pmf of Rx-levels.

File name:	Aspern.txt
Sample number:	90178
Sample frequency:	0.10
Start date:	July 16, 2004
Duration:	10 days, 10 hours
Avg Temperature:	17 °C
Site Type:	Urban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



D.2.4: Map.

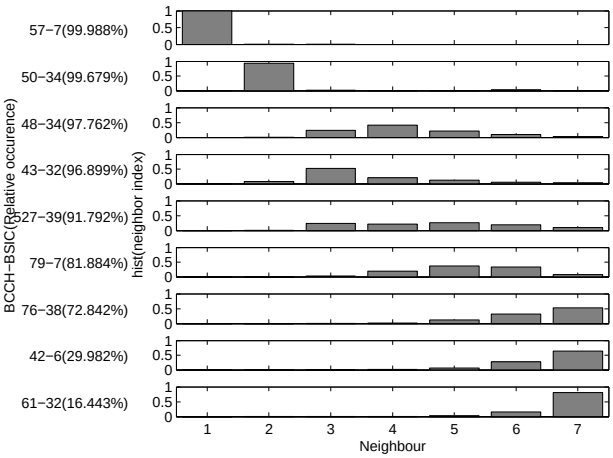
D.2.5: Photo of site.



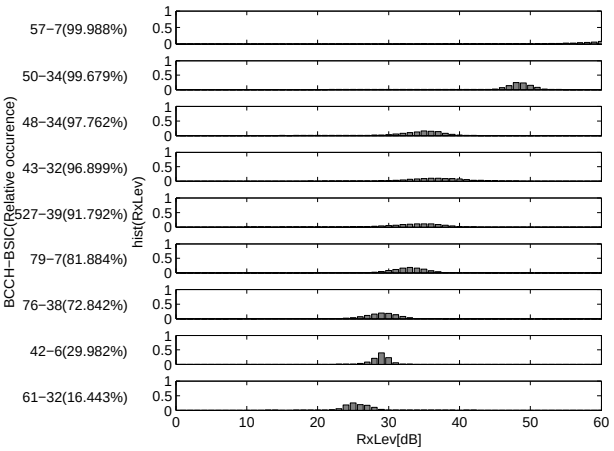
D.2.6: Neighbors over time.

D.2.2 Hoher Markt

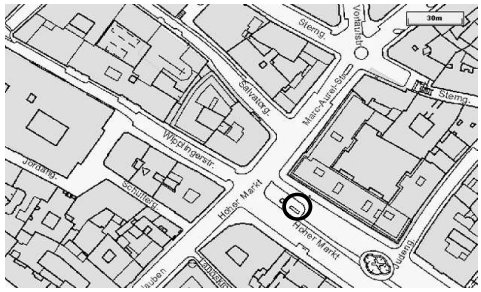
File name:	HoherMarkt.txt
Sample number:	110239
Sample frequency:	0.10
Start date:	July 25, 2004
Duration:	12 days, 18 hours
Avg Temperature:	17.5 °C
Site Type:	Urban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



D.2.7: Pmf of neighboring index.

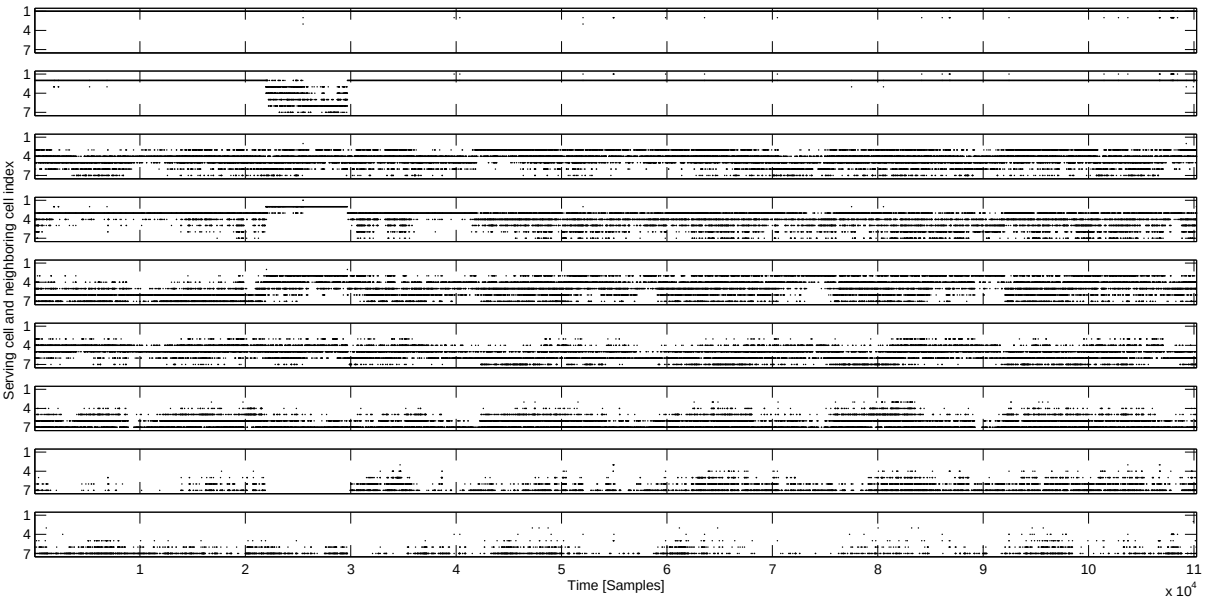


D.2.8: Pmf of Rx-levels.



D.2.9: Map.

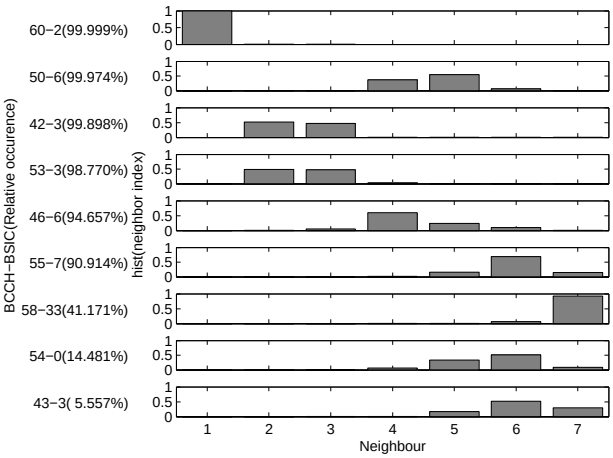
D.2.10: Photo of site.



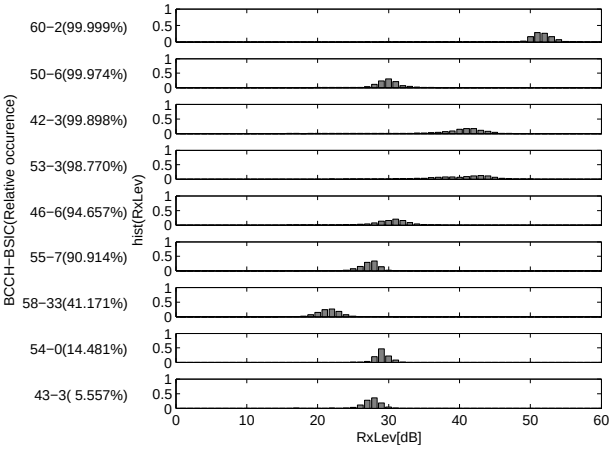
D.2.11: Neighbors over time.

D.2.3 Hörlgasse-Türkenstrasse

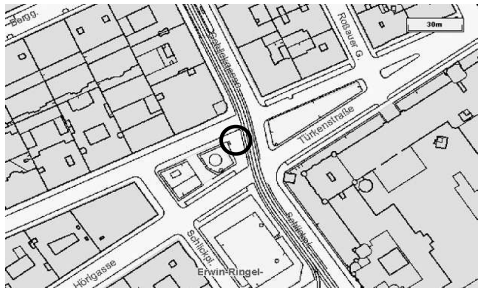
File name:	HoerlgTuerkenstr.txt
Sample number:	107072
Sample frequency:	0.10
Start date:	August 12, 2004
Duration:	12 days, 9 hours
Avg Temperature:	15.8 °C
Site Type:	Urban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



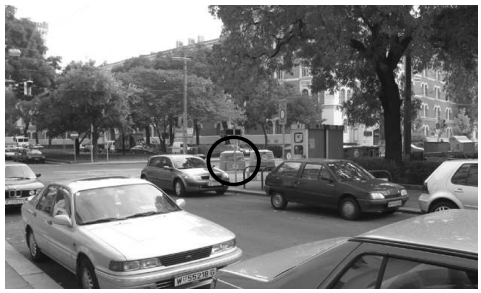
D.2.12: Pmf of neighboring index.



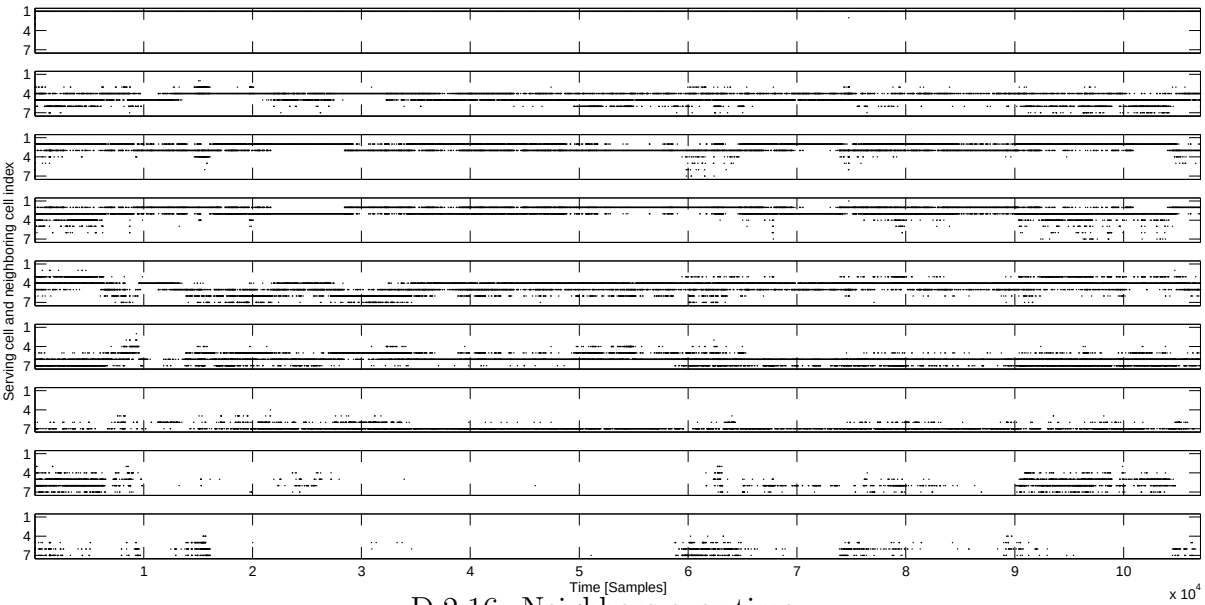
D.2.13: Pmf of Rx-levels.



D.2.14: Map.



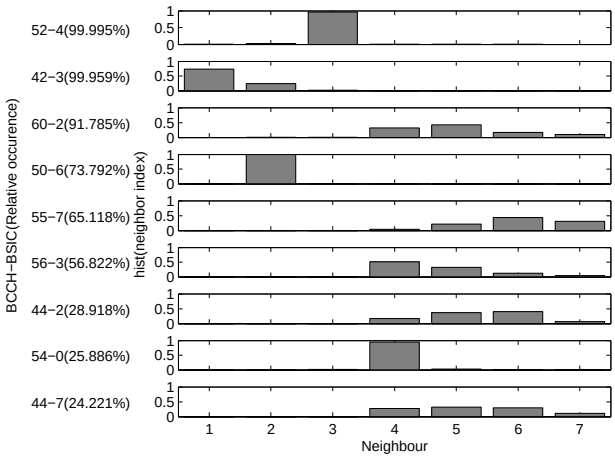
D.2.15: Photo of site.



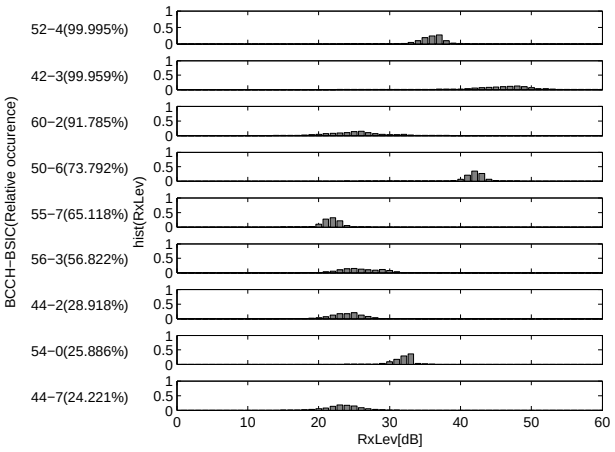
D.2.16: Neighbors over time.

D.2.4 Hörlgasse-Wasagasse

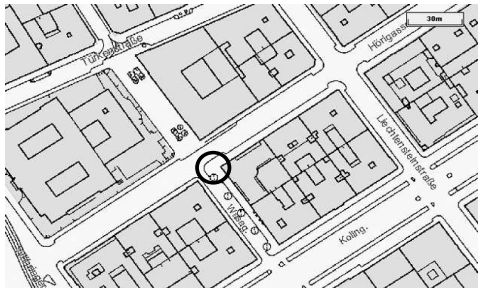
File name:	HoerlgWasag.txt
Sample number:	81959
Sample frequency:	0.10
Start date:	August 25, 2004
Duration:	9 days, 12 hours
Avg Temperature:	21.2 °C
Site Type:	Urban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



D.2.17: Pmf of neighboring index

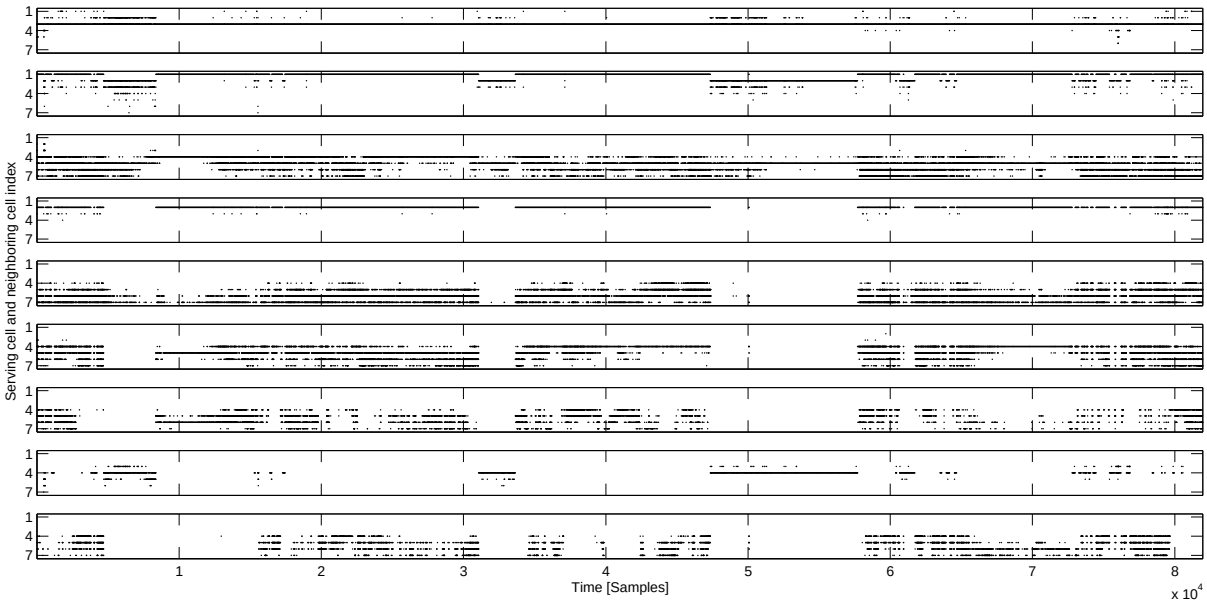


D.2.18: Pmf of Rx-levels.



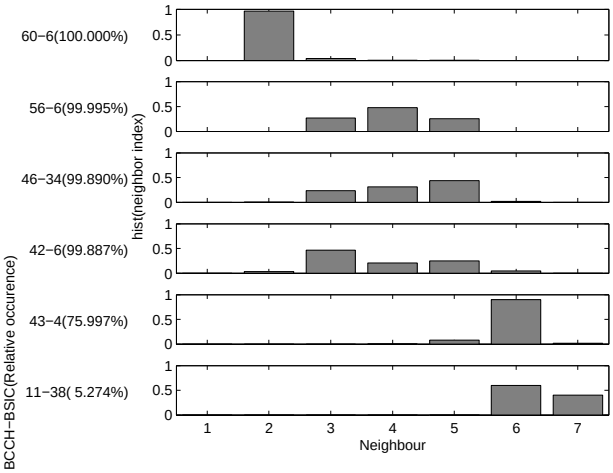
D.2.19: Map.

D.2.20: Photo of site.

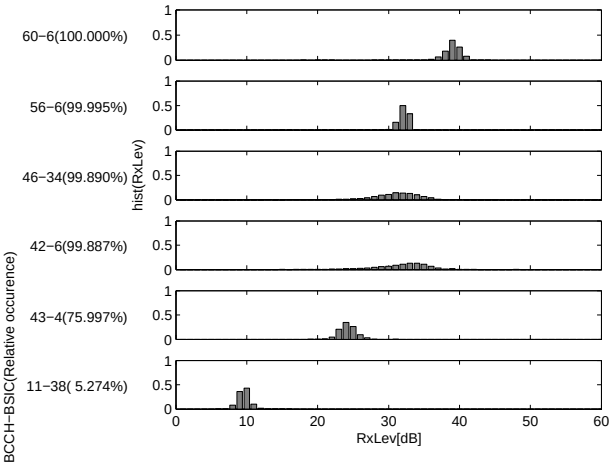


D.2.21: Neighbors over time.

D.2.5 Neuer Markt

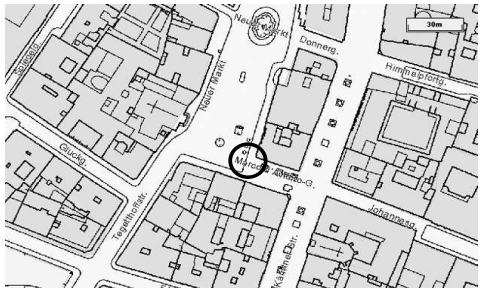


D.2.22: Pmf of neighboring index.



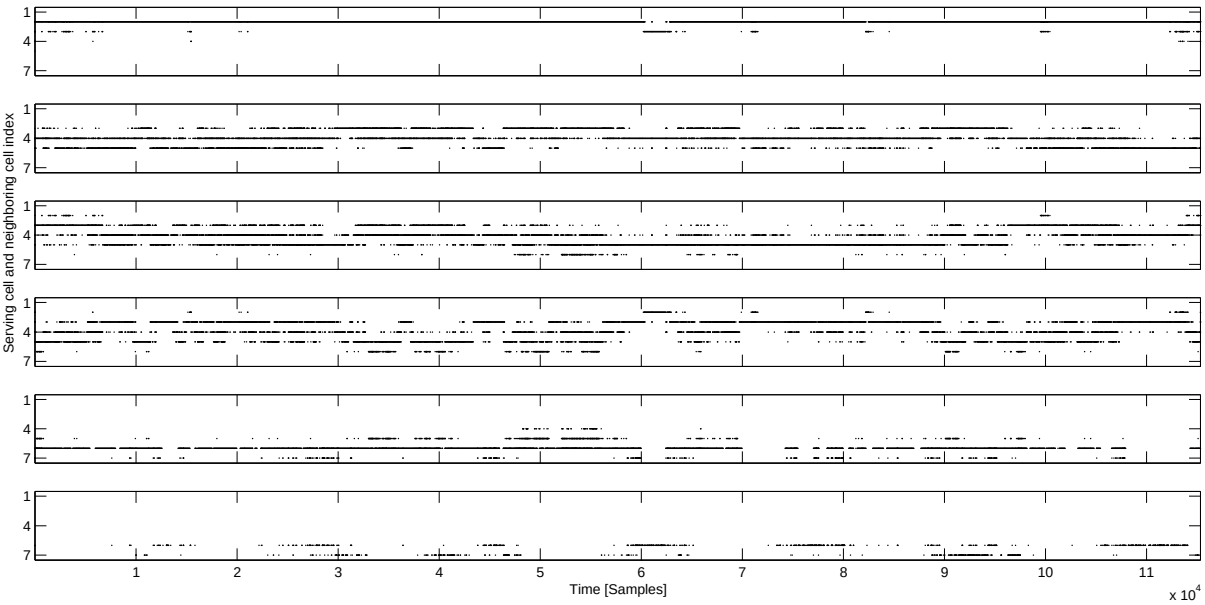
D.2.23: Pmf of Rx-levels.

File name:	NeuerMarkt.txt
Sample number:	115353
Sample frequency:	0.10
Start date:	September 6, 2004
Duration:	13 days, 8 hours
Avg Temperature:	16.2 °C
Site Type:	Urban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



D.2.24: Map.

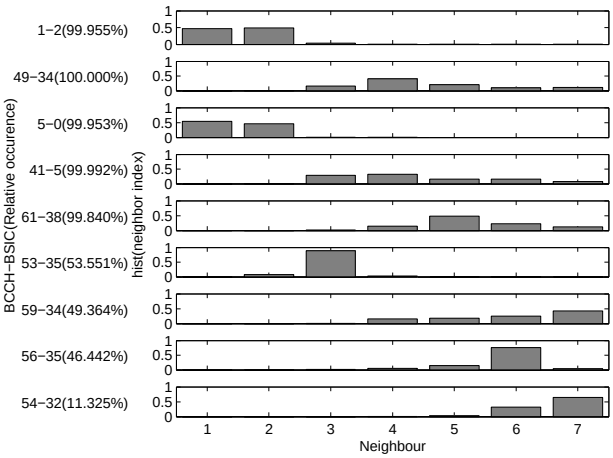
D.2.25: Photo of site.



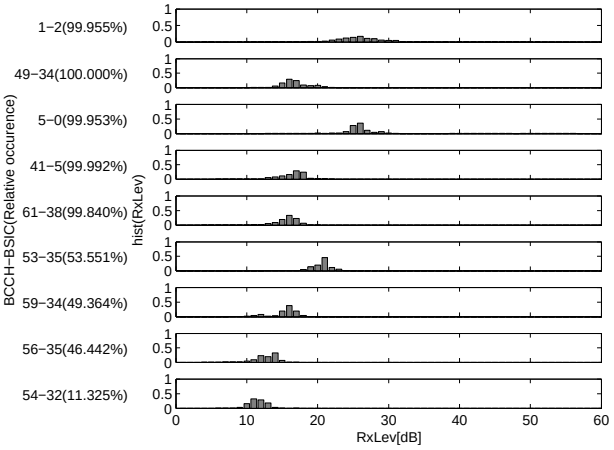
D.2.26: Neighbors over time.

D.2.6 Soldanellenweg

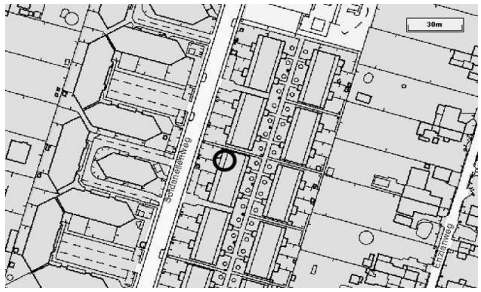
File name:	Soldanellenweg.txt
Sample number:	105947
Sample frequency:	0.10
Start date:	September 20, 2004
Duration:	12 days, 6 hours
Avg Temperature:	15.1 °C
Site Type:	Urban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



D.2.27: Pmf of neighboring index.

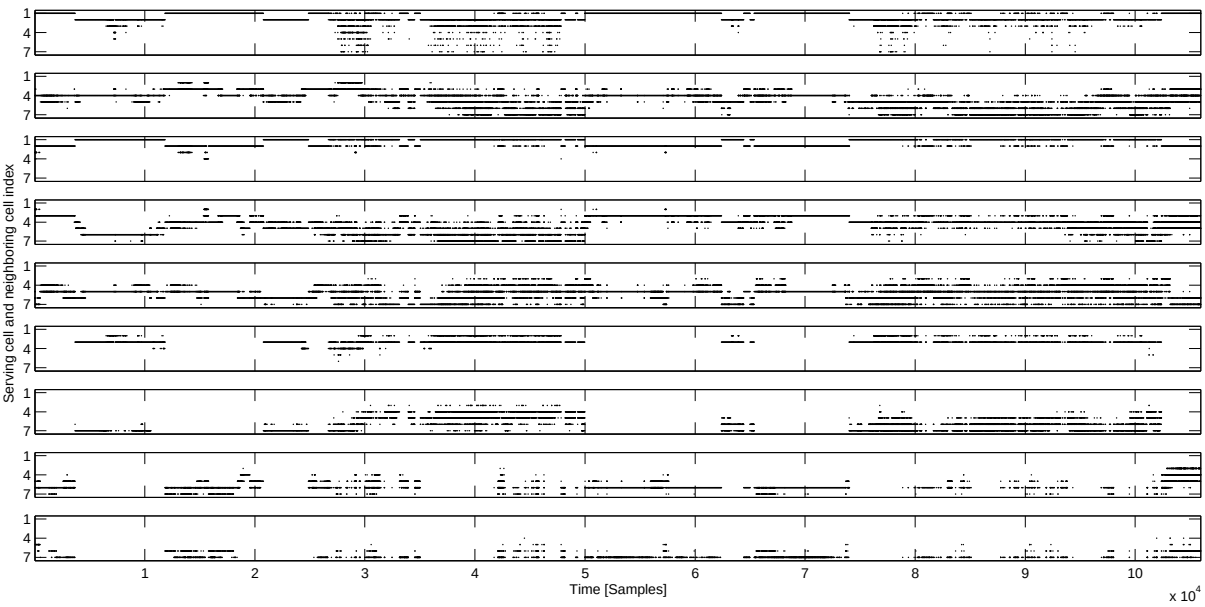


D.2.28: Pmf of Rx-levels.



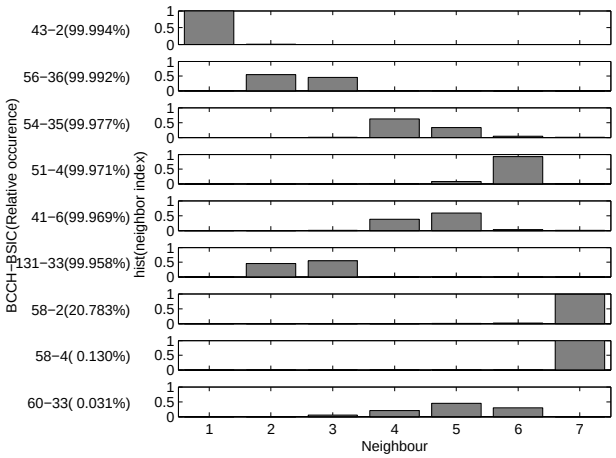
D.2.29: Map.

D.2.30: Photo of site.

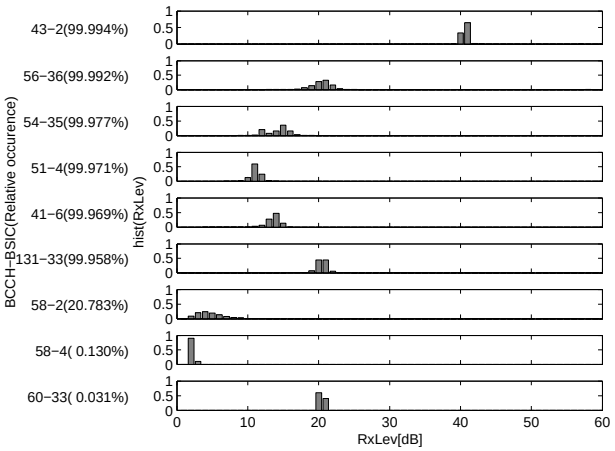


D.2.31: Neighbors over time.

D.2.7 Templergasse

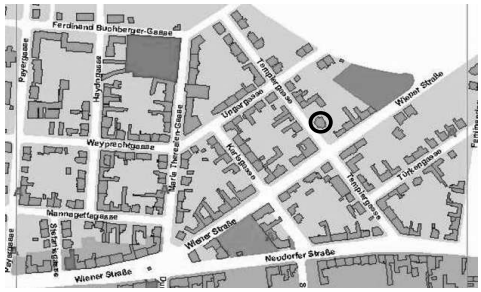


D.2.32: Pmf of neighboring index.



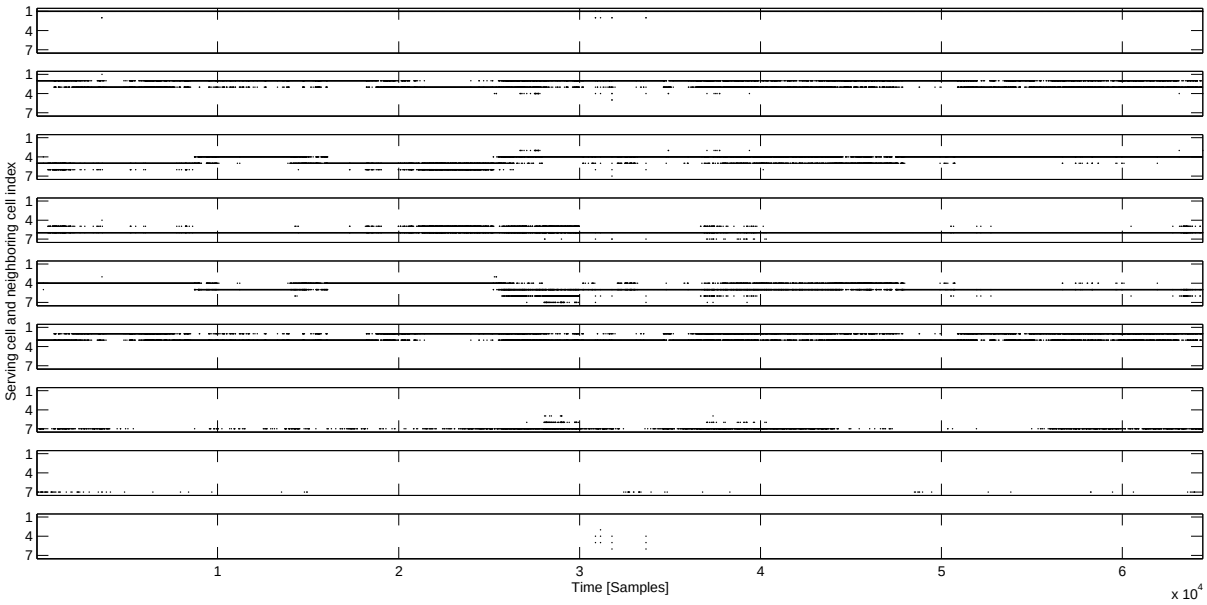
D.2.33: Pmf of Rx-levels.

File name:	Templergasse.txt
Sample number:	64448
Sample frequency:	0.10
Start date:	October 5, 2004
Duration:	7 days, 12 hours
Avg Temperature:	13.1 °C
Site Type:	Suburban
Receiver:	Siemens S55
Antenna:	External, 0/3db (900/1800MHz)
Power Supply:	12V
Antenna height	1.5m above ground



D.2.34: Map.

D.2.35: Photo of site.



D.2.36: Neighbors over time.

Used in method	NMR field	Description
✓	Date	Current date.
✓	Time	Current time of the day.
	MCC	Mobile Country Code.
	MNC	Mobile Network Code.
	LAC	Location area.
✓	CID	Cell Identification; 16 bit number identifying a cell within the GSM Location Area Identiy. Together with the MCC, the MNC and the LAC a unique identification of the cell is possible.
✓	RxLev_Full	Reception Level (Received Signal Strength Indication) of serving cell.
	RxQual_Full	Reception Quality RXQUAL (without Discontinuous Transmission).
	RxQual_Sub	Reception Quality RXQUAL (with Discontinuous Transmission).
✓	BCCH_N1	Broadcast Control Channel of neighboring cell 1.
✓	BSIC_N1	Base Station Identity Code of neighboring cell 1, used to ensure the distinction between different BCCHs in areas where multiple BCCHs transmit at the same frequency.
✓	RxLev_N1	Reception Level of neighboring cell 1.
✓	BCCH_N2	Broadcast Control Channel of neighboring cell 2.
✓	BSIC_N2	Base Station Identity Code of neighboring cell 2.
✓	RxLev_N2	Reception Level of neighboring cell 2.
✓	BCCH_N3	Broadcast Control Channel of neighboring cell 3.
✓	BSIC_N3	Base Station Identity Code of neighboring cell 3.
✓	RxLev_N3	Reception Level of neighboring cell 3.
✓	BCCH_N4	Broadcast Control Channel of neighboring cell 4.
✓	BSIC_N4	Base Station Identity Code of neighboring cell 4.
✓	RxLev_N4	Reception Level of neighboring cell 4.
✓	BCCH_N5	Broadcast Control Channel of neighboring cell 5.
✓	BSIC_N5	Base Station Identity Code of neighboring cell 5.
✓	RxLev_N5	Reception Level of neighboring cell 5.
✓	BCCH_N6	Broadcast Control Channel of neighboring cell 6
✓	BSIC_N6	Base Station Identity Code of neighboring cell 6.
✓	RxLev_N6	Reception Level of neighboring cell 6.

Table D.2: Network Measurement Report fields

Appendix E

Abbreviations and Acronyms

2D :	Two Dimensional
3D :	Three Dimensional
3G :	Third Generation
3GPP :	The 3rd Generation Partnership Project Agreement
AFLT :	Advanced Forward Link Trilateration
AGPS :	Assisted Global Positioning System
AOA :	Angle of Arrival
AWGN :	Additive White Gaussian Noise
BCCH :	Broadcast Control Channel
BER :	Bit Error Rate
BS :	Base Station
BSIC :	Base Station Identity Code
C/A :	Coarse Acquisition
cdf :	Cumulative Distribution Function
CDMA :	Code Division Multiple Access
CERP :	Circular Error Probability
cmf :	Cumulative Mass Function
COST :	European CO-operation in the field of Scientific and Technical Research
DCM :	Database Correlation Method
DGPS :	Differential GPS
DOA :	Direction of Arrival
DOD :	Department of Defense

DR :	Dead Reckoning
DSP :	Digital Signal Processor
EC :	European Commission
EOTD :	Enhanced Observed Time Difference
ESA :	European Space Agency
ETSI :	European Telecommunications Standards Institute
FCC :	United States Federal Communication Commission
GDOP :	Geometric Dilution of Precision
GLONASS :	Global Navigation Satellite System (Russian)
GNSS :	Global Navigation Satellite System
GPS :	Global Positioning System
GSM :	Global System for Mobile Communications
GTD :	Geometric Time Difference
Hz :	Hertz
ID :	Identity
IPDL :	Idle Period Downlink
LBS :	Location Based Service
KL :	Kullback Leibler
LAN:	Local Area Network
LCS :	Location Service
LF :	Location Fingerprinting
LIF :	Location Interoperability Forum
LMU :	Location Measurement Unit
LORAN :	Long Range Radio Navigation
LOS :	Line of Sight
LS :	Least Square
MCT :	Mobile Communications Terminal
MS :	Mobile Station
NA :	Not Applicable
NLOS :	None Line of Sight
NMR :	Network Measurement Report
OMA :	Open Mobile Alliance
OTD :	Observed Time Difference
OTDOA :	Observed Time Difference of Arrival

P-Code :	Precision Code
pdf :	Probability Density Function
PRN :	Pseudorandom Noise
PSAP :	Public Safety Answering Point
QASPR :	Quallcomm Automatic Satellite Position Reporting
RF :	Radio Frequency
RF :	Radio Frequency
RMSE :	Root Mean of Squared Errors
RTD :	Round Trip Delay
RTD :	Real Time Difference (see EOTD)
RTT :	Round Trip Time
Rx-level :	Received Power Level
SA :	Selective Availability
SNR :	Signal to Noise Ratio
SPS :	Standard Positioning Service
SS :	Signal Strength
SV :	Space Vehicles
TA :	Timing Advance
TDMA :	Time Division Multiple Access
TDOA :	Time Difference of Arrival
TOA :	Time of Arrival
UMTS :	Universal Mobile Telecommunications System
UWB :	Ultra Wide Band
WLAN :	Wireless Local Area Network
WLS :	Wireless Location Signatures

Appendix F

List of Variables

We list symbols that are generally used throughout the thesis. Locally used symbols are omitted. Symbols are ordered according to case and alphabet.

$\boldsymbol{\alpha}$:	Hyperparameters of Dirichlet distribution
$\tilde{\alpha}$:	Posterior hyperparameters of Dirichlet distribution
Δt:	Time difference
ϵ:	Epsilon
η:	Multiplicity of samples
$\boldsymbol{\theta}$:	Parameter vector
λ:	Discrete Bayesian network
$\boldsymbol{\rho}$:	Parent configuration
σ:	Location variability
τ_{LT}:	Long-term time constant
ϕ_C:	Potential
$\mathcal{C}$:	Family of cliques
$\mathcal{G}$:	Bayesian network structure
$\mathcal{G}_{NMR}$:	Optimized Bayesian network structure
$\mathcal{I}$:	Set of positions
$\mathcal{P}$:	Probability set of a Bayesian network
$\mathcal{S}$:	Alphabet of variables X
$\mathcal{V}$:	Node set of Bayesian network
i:	Position
m:	Sample number
n:	Uniform fraction of prior distribution

p:	Enumerated parent configuration
$pa(X_v)$:	Parent set of X_v
r:	Radius
r_m:	Distance between positions
t:	Time
v:	Node
C:	Clique
$\boldsymbol{D}$:	Database
$\boldsymbol{D}_0$:	Database for prior construction
E:	Set of edges of a Bayesian network
$\boldsymbol{w}$:	Weighting factors
$\hat{\boldsymbol{w}}$:	Optimal weighting factors
$\boldsymbol{x}$:	Realizations of $\boldsymbol{X}$
x^s:	Element of $\mathcal{S}$
K:	Kullback Leibler Information
L:	Marginal likelihood
M:	Length of Database
P:	Received power levels
S:	Cardinality of alphabet $\mathcal{S}$
U:	Believe Universe
V:	Maximal number of nodes
$\boldsymbol{X}$:	Random variables representing nodes

Bibliography

[1] Revision of the commissions rules to ensure compatibility with enhanced 911 emergency calling systems. Notice of Proposed Rulemaking, 9 FCC Rcd. 6170 (E911 NPRM) CC Docket 94-102, RM-8143, United States Federal Communication Commission (FCC), 1994.

[2] RTCM recommended standards for differential navstar GPS service. RTCM special committe no. 104, Radio Technical Commission For Maritime Services, 1800 Diagonal Road, Suite 600, Alexandria, Virginia 22314-2840 USA, January 1994.

[3] The commissions rules to ensure compatibility with enhanced 911 emergency calling systems. Report and Order and Further Notice of Proposed Rulemaking, 11 FCC Rcd. 18676 (1996)("First E911 Report and Order and FNPRM") CC Docket No. 94-102, United States Federal Communication Commission (FCC), 1996.

[4] On universal service and users' rights relating to electronic communications networks and services (universal service directive). Directive 2002/22/EC of the european parliament and of the council, European Commission, March 2002.

[5] Anonymos. 1990 federal navigation plan. Copies available from NTIS, 5285 Port Royal Road, Springfield, MA 22161 as Document DOT-VNTSC-RSPA-90-3/DOD-4650.4, 1990.

[6] J. M. Bernardo and A. F. M. Smith. *Bayesian theory.* John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA, May 1994. ISBN: 0-471-92416-4.

[7] E. Bonek and A. F. Molisch. Mobilkommunikation. Lecture notes of Prof. E. Bonek, Vienna University of Technology, 1998.

[8] G. E. P. Box. Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799, December 1976.

[9] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference - A Practical Information Theoretic Approach*. Springer-Verlag New York, Inc., 175 Figth Avenue, New York, NY 10010, USA, 2 edition, 2002. ISBN: 0-387-95364-7.

[10] Y. T. Chan and K. C. Ho. An efficient closed form localization solution from time difference of arrival measurements. In *Proceedings on IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP-94*, volume 2, pages 393–396, Adelaide, Australia, April 1994.

[11] C. Cook and M. Bernfeld. *Radar Signals: An Introduction to Theory and Applications*. Artech House Radar Library Series. Artech House, Inc., 685 Canton Street, Norwood, MA 02062, USA, 1993. ISBN: 0-89-006733-3.

[12] G. Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42(23):393–405, 1990.

[13] G. F. Cooper and E. Herskovits. A bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4):309–347, 1992.

[14] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems*. Statistics for Engineering and Information Science. Springer-Verlag New York, Inc., 175 Figth Avenue, New York, NY 10010, USA, 1999. ISBN: 0-387-98767-3.

[15] R. G. Cowell, A. P. Dawid, and D. J. Spiegelhalter. Sequential model criticism in probabilitic expert systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(3):209–219, March 1993.

[16] E. Damosso and L. M. Correia, editors. *COST Action 231: Digital mobile radio towards future generation systems*. European Commission, Directorate-General Telecommunications, Information Society, Information Market, and Exploitation of Research, 1999. ISBN: 92-828-5416-7.

[17] R. J. Danchik. An overview of transit development. *Johns Hopkins APL Technical Digest*, 19(1):18–26, 1998.

[18] F. Evennou, F. Marx, and S. Nacivet. Sensor fusion for UWB and Wifi indoor positioning systems. In *Proceedings of 8th International Symposium on Advanced Radio Technologies (ISART)*, Boulder, Colorado, USA, March 2005.

[19] M. Feuerstein. Wireless signatures technology for position location. Presentation at the 7th International Symposium on Advanced Radio Technologies (ISART), Boulder, Colorado, USA, March 2004.

[20] M. Feuerstein. Field tests of hybrid wireless location technologies. Presentation at the 8th International Symposium on Advanced Radio Technologies (ISART), Boulder, Colorado, USA, March 2005.

[21] B. Friedlander. A passive localization algorithm and its accuracy analysis. *IEEE Journal of Oceanic Engineering*, 12(1):234–244, 1987.

[22] D. Geiger and D. Heckerman. A characterization of the dirichlet distribution through global and local independence. *Annals of Statistics*, 25:1344–1368, 1997.

[23] S. Gezici, H. Kobayashi, and H. V. Poor. A new approach to mobile position tracking. In *Proceedings on IEEE Sarnoff Symposium on Advances in Wired and Wireless Communications*, pages 204–207, Ewing, NJ, USA, March 2003.

[24] D. Golinelli, D. Madigan, and G. Consonni. Relaxing the local independence assumption for quantitative learning in acyclic directed graphical models through hierarchical partition models. In *Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics (Uncertainty '99)*, Fort Lauderdale, Florida, USA, January 1999.

[25] M. Gudmundson. Correlation model for shadow fading in mobile radio systems. *Electronic Letters*, 27(23):2145–2146, Nov 1991.

[26] G. Gunnarsson, M. Alanen, T. Rantalainen, V. Ruutu, and V. M. Teittinen. Location trial system for mobile phones. In *Proceedings of the IEEE Global*

Telecommunications Conference GLOBECOM98, pages 2211–2216, Sydney, Australia, November 1998.

[27] D. Heckerman. A tutorial on learning with bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, Advanced Technology Division, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, March 1995.

[28] D. Heckerman, D. Geiger, and D. M. Chickering. Learning bayesian networks: The combination of knowledge and statistical data. Technical Report MSR-TR-94-09, Microsoft Research, Advanced Technology Division, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, 1995.

[29] A. K. Jain, R. P. W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.

[30] H. Jeffreys. *Theory of Probability*. Oxford University Press, Great Clarendon Street, Oxford OX2 6DP, 3 edition, 1961. ISBN: 0-19-850368-7.

[31] F. V. Jensen. *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. Springer-Verlag New York, Inc., 175 Figth Avenue, New York, NY 10010, USA, 2001. ISBN: 0-387-95259-4.

[32] O. Kennemann. Locating mobiles in non-flowing traffic. In *Proceedings on 6th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages 274–278, Toronto, Ont., Canada, September 1995.

[33] P. Kontkanen, P. Myllymäki, T. Silander, H. Tirri, and P. Grünwald. On predictive distributions and bayesian networks. *Statistics and Computing*, 10(1):39–54, 2000.

[34] H. Kunczier. Evaluation of existing GSM/UMTS LCS algorithms. Technical report, Forschungszentrum Telekommunikation Wien, ftw., 2001.

[35] H. Kunczier and H. Anegg. LoLa - a prototype of a network independent wireless internet service. In *Proceedings of the 5th International Symposium on Advanced Radio Technologies (ISART)*, Boulder, Colorado, USA, March 2003.

[36] H. Kunczier and H. Anegg. Appliance of optimized bayesian networks for location estimation. In *Proceedings of the 59th IEEE Semiannual Vehicular Technology Conference (VTC)*, Milan, Italy, May 2004.

[37] H. Kunczier and H. Anegg. Enhanced cell id based terminal location for urban area location based applications. In *Proceedings of the 1st IEEE Consumer Communications and Networking Conference*, Las Vegas, Nevada, USA, January 2004.

[38] H. Kunczier and H. Anegg. Enhanced location estimation via pattern matching and motion modelling. In *Proceedings of the 6th International Symposium on Advanced Radio Technologies (ISART)*, Boulder, Colorado, USA, March 2004.

[39] H. Kunczier, A. Friedreich, and R. Zelenka. A bayesian method for mobile positioning in urban areas. In *Proceedings of the International Conference on Communications, Internet and Information Technology (CIIT 2004)*, St. Thomas, USA, November 2004.

[40] H. Laitinen, S. Ahonen, S. Kyriazakos, J. Lähteenmäki, R. Menolascino, and S. Parkkila. Cellular location technology. Cellular Network Optimisation based on Mobile Location (CELLO) WP2 Deliverable, http://www.telecom.ece.ntua.gr/cello/, November 2001.

[41] H. Laitinen, J. Lähteenmäki, and T. Nordström. Database correlation method for GSM location. In *Proceedings on the 53rd IEEE Vehicular Technology Conference, (VTC) 2001 Spring Conference*, volume 4, Rhodes Island, Greece, May 2001.

[42] S. Larder. E-OTD based location service implementation. In *Proceedings of Mobile Location Workshop (MLW 2001)*, Espoo, Finland, June 2001.

[43] B. L. Le, K. Ahmed, and H. Tsuji. Mobile location estimator with NLOS mitigation using kalman filtering. In *Proceedings of IEEE wireless Communications and Networking (WCNC'03)*, volume 3, pages 1969–1973, New Orleans, LA, USA, March 2003.

[44] J. Lee and R. Scholtz. Ranging in a dense multipath environment using an UWB radio link. *IEEE Transactions on Selected Areas in Communications*, 20(9):1677–1683, 2002.

[45] B. Ludden, A. Pickford, J. Medland, H. Johnson, F. Brandon, L. E. Axelsson, K. Viddal-Ervik, B. Dorgelo, E. Boroski, and J. Malenstein. Cgalies work package 1 draft report. *http://www.telematica.de/cgalies/*, May 2001.

[46] S. Mangold and S. Kyriazakos. Applying pattern recognition techniques based on hidden markov models for vehicular position location in cellular networks. In *Proceedings of 50th IEEE Vehicular Technology Conference (VTC-Fall)*, pages 780–784, Amsterdam, Netherlands, September 1999.

[47] W. Mansfeld. *Satellitenortung und Navigation: Grundlagen und Anwendung globaler Satellitennavigationssysteme*. Friedr. Vieweg & Sohn Verlag/GWV Fachverlag GmbH, Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany, 2nd edition, 2004. ISBN: 3-52-816886-2.

[48] M. J. Marsan, G. C. Hess, and S. S. Gilbert. Shadowing variability in an urban land mobile environment at 900MHz. *Electronic Letters*, 26(10):646–648, May 1990.

[49] J. C. Maxwell. *A Dynamical Theory of the Electromagnetic Field*. The Scientific Papers of James Clerk Maxwell, Vol. 2. Dover, New York, 31 East 2nd Street, Mineola, NY 11501-3852, USA, 1865. ISBN: 0-48-649561-2.

[50] G. Mizusawa. Performance of hyperbolic position location techniques for code division multiple access. Master of science, Virginia Polytechnic Insitute and State University, August 1996.

[51] K. P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. Doctor of philosophy, University of California, Berkeley, 2002.

[52] M. A. Pallas and G. Jourdain. Active high resolution time delay estimation for large BT signals. *IEEE Transactions on Signal Processing*, 39(4):781–788, 1991.

[53] B. W. Parkinson and J. J. Spilker Jr., editors. *Global Positioning System: Theory and Applications*, volume 1. American Institute of Aeronautics and

Astronautics, Inc., 370 L'Enfant Promenade, SW, Washington, DC 20024-2518, 1996. ISBN: 1-56347-106-X.

[54] B. W. Parkinson and J. J. Spilker Jr., editors. *Global Positioning System: Theory and Applications*, volume 2. American Institute of Aeronautics and Astronautics, Inc., 370 L'Enfant Promenade, SW, Washington, DC 20024-2518, 1996. ISBN: 1-56347-107-8.

[55] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag New York, Inc., 175 Figth Avenue, New York, NY 10010, USA, 2nd edition, 1994. ISBN: 0-387-94173-8.

[56] J. Rissanen. Stochastic complexity (with discussion). *Journal of the Royal Statistical Society, Series B*, 49(3):223–239 and 253–265, 1987.

[57] D. Rusakov and D. Geiger. On parameter priors for discrete DAG models. In *Proceedings of AI and Statistics - 8th International Workshop on Artificial Intelligence and Statistics*, Key West, Florida, USA, January 2001.

[58] S. R. Saunders. *Antennas and Propagation for Wireless Communication Systems*. John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA, 2000. ISBN: 0-471-98609-7.

[59] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

[60] H. Song. Automatic vehicle location in cellular communications systems. *IEEE Transactions on Vehicular Technology*, 43(4):902–908, Nov. 1994.

[61] D. J. Spiegelhalter, N. L. Harris, K. Bull, and R. C. G. Franklin. Empirical evaluation of prior beliefs about frequencies: methodology and a case study in congenital heart disease. *Journal of the American Statistical Association*, 89(426):435–443, 1994.

[62] S. C. Swales, J. E. Maloney, and J. Stevenson. Locating mobile phones and the us wireless E-911 mandate. In *IEE Colloquium on Novel Methods of Location and Tracking of Cellular Mobiles and Their System Applications*, London, May 1999.

[63] J. Syrjärinne. *Studies of Modern Techniques for Personal Positioning*. Thesis for the degree of doctor of technology, Tampere University of Technology, March 2001.

[64] A. R. Syversveen. Noninformative bayesian priors. interpretation and problems with construction and applications. Technical report, Department of Mathematical Sciences, NTNU, Trondheim, 1998.

[65] F. B. Taub. Book review: Estimating ecological risks. *Ecology*, 74(4):1290–1291, 1993.

[66] Test Plans and Criteria Subgroup CDG Location Technology Forum. CDG test plan document for location determination technologies evaluation (baseline version - accepted). Technical report, Bell Labs, Lucent Technologies, copyright CDMA Development Group., 2000.

[67] G. L. Turin. An introduction to matched filters. *IRE Transactions on Information Theory*, 6(3):785–792, 1960.

[68] M. Umlauft, E. Michlmayr, H. Anegg, H. Kunczier, and G. Pospischil. LoLa: Ein Prototyp für einen UMTS-basierenden mobilen Stadtführer (in german). In H. Meuer and O. Spaniol, editors, *Praxis der Informationsverarbeitung und Kommunikation*. K. G. Saur, München, Germany, September 2002.

[69] S. Watanabe. *Pattern Recognition, Human and Mechanical*. John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA, 1985. ISBN: 0-471-80815-6.

[70] S. S. Wilks. *Mathematical Statistics*. John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0012, USA, 1962. ISBN: 0-47-194644-3.

[71] R. Winkler. The assessment of prior distributions in bayesian analysis. *American Statistical Association Journal*, 62:776–800, 1967.

[72] M. Wylie and J. Holtzman. The non-line of sight problem in mobile location estimation. In *Proceedings on the 5th IEEE International Conference on Universal Personal Communications*, volume 2, pages 827–831, Cambridge, MA, USA, September 1996.

[73] L. Xiong. A selective model to suppress NLOS signals in angle-of-arrival(AOA) location estimation. In *Proceedings on the 9th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages 461–465, Boston, MA, USA, September 1998.

[74] D. Zimmermann, M. Layh, J. Baumann, and F. Landstorfer. Positioning of mobile terminals in cellular networks using wave propagation models. COST 273 TD(04)195, Duisburg, Germany, September 2004.

[75] D. Zimmermann, P. Wertz, J. Bauknecht, and F. M. Landstorfer. Performance of location methods in an urban mobile communication network. COST 273 TD(04)053, Athens, Greece, Jannuary 2004.

Curriculum Vitae

Harald Kunczier

Citizen of Austria

Born on May 22^{nd}, 1976, in Mödling, Austria

Education

2000-2005:	*Technische Universität Wien* PhD studies in electrical engineering
1995-2000:	*Technische Universität Wien* Diploma studies in electrical engineering
2000-2002:	*Donau Universität Krems* Master of Business Administration in General Management
1994-1994:	*Baldwin Wallace College*, Ohio, USA
1990-1995:	*Höhere Technische Lehranstalt*, Mödling
1986-1990:	*Bundesrealgymnasium Franz Keimgasse*, Mödling

Employment/Internships

since Oct. 1999:	*Forschungszentrum Telekommunikation Wien (ftw.)* Research and project management in LBS and positioning technologies
since Mar. 2005:	*Fachhochschule Hagenberg* Lecture and Lab "Location based Services"
Aug. 1998 - Sep. 1999:	*ATOS Origin Information Technology*, Wien SAP to BAAN data migration

National Service

Apr. 2004 - Dec. 2005: Jägerregiment 24, Lienz

Hobbies

mountain biking, travelling, golfing