

---

UNIVERSITY OF TAMPERE

Master's Thesis

---

Laura Koski

Positioning with Bayesian coverage  
area estimates and location  
fingerprints

---

Department of Mathematics and Statistics

Mathematics

March 2010

---

University of Tampere

Department of Mathematics and Statistics

KOSKI, LAURA: Positioning with Bayesian coverage area estimates and location fingerprints

Master's Thesis, 59 pages, 6 appendix pages

Mathematics

March 2010

---

## Abstract

A variety of commercial location based services have appeared during recent years. Location awareness is becoming more important also in environments where satellite-based positioning are not available, such as urban areas and indoors. In this work, a method to estimate the coverage area of a wireless communication node is presented. Also a method to use a database of such coverage area estimates for personal positioning is presented. Coverage area estimates are computed using location fingerprinting.

The coverage area is solved by forming the posterior distribution of the parameters using Bayes' rule. The coverage area of a communication node is modeled as an ellipse and is assumed to follow a multivariate normal linear model, which is presented as a general case. The coverage area estimate is derived using both noninformative and informative priors. Also a model which assumes a possibility of outliers and a Bayesian method for detecting outliers are presented.

A positioning method which uses the coverage area estimates is presented. The distribution of a position estimate is derived using Bayes' rule. The position estimate is weighted average of the centers of ellipses and the weights are determined by the shape parameters of ellipses.

Finally, the accuracy and consistency of a position estimate are studied using different coverage estimates.

**Keywords:** multivariate linear models, Bayesian analysis, location fingerprinting, coverage area estimation



## Preface

I wish to thank my supervisor Professor Robert Piché for giving me opportunity to work in the Personal Positioning Algorithms Research group and for giving me this interesting research topic and also for his ideas and comments. I would also like to thank my supervisor Ph.D Jarkko Isotalo for his comments.

I owe a big thank to my co-workers in the research group and especially Ph.D. Simo Ali-Löytty for his patience in answering my many questions. I would like to thank Nokia Inc. for founding this work. I wish to acknowledge M.Sc. Lauri Wirola from Nokia Inc. for helpful discussions and for giving me different perspective to the topic.

Finally, my family deserves a special acknowledgment for constantly supporting me during my studies. Lastly, I wish to thank Lauri for taking good care of me.



# Contents

<b>Abbreviations</b>	<b>vii</b>
<b>Symbols</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Fingerprinting . . . . .	2
1.2 Existing solutions . . . . .	5
1.3 Scope and structure of thesis . . . . .	7
<b>2 Preliminaries</b>	<b>8</b>
2.1 Distributions . . . . .	8
2.2 Bayesian inference . . . . .	14
2.2.1 Bayes' rule . . . . .	14
2.2.2 Bayes' estimators . . . . .	16
2.3 Bayesian filtering . . . . .	18
<b>3 Coverage area models</b>	<b>20</b>
3.1 Multivariate normal linear model . . . . .	20
3.1.1 The likelihood function . . . . .	20
3.1.2 Noninformative prior . . . . .	21
3.1.3 Informative prior . . . . .	23
3.2 Mixture model . . . . .	25
3.2.1 A special linear model . . . . .	25
3.2.2 Outlier detection . . . . .	27
3.3 Updating posterior . . . . .	29
3.3.1 Time invariant updating . . . . .	30
3.3.2 Time variance updating . . . . .	31
3.4 Coverage area estimation . . . . .	32
<b>4 Bayesian positioning</b>	<b>34</b>
<b>5 Evaluation</b>	<b>38</b>
5.1 Coverage area of a communication node . . . . .	38
5.2 Updating coverage area estimate . . . . .	39
5.3 Positioning . . . . .	42
5.3.1 Simulations . . . . .	42
5.3.2 Location accuracy using different coverage area models	44
5.3.3 Consistency of location estimates . . . . .	48
<b>6 Conclusion and future work</b>	<b>52</b>
<b>References</b>	<b>54</b>

<b>A</b>	<b>Energy consumption with GPS and WLAN</b>	<b>60</b>
<b>B</b>	<b>Matrix properties</b>	<b>62</b>
B.1	Matrix derivatives . . . . .	63
B.2	Jacobians of matrix transformations . . . . .	64

# Abbreviations

2D	Two Dimensional
3D	Three Dimensional
3GPP	Third Generation Partership Project
A-GPS	Assisted GPS
AOA	Angle Of Arrival
DCM	Database Correlation Method
CDMA	Code Division Multiple Access
CEP	Circular Error Probable
CID	Cell Identity
CN	Communication Node
DVB-SH	Digital Video Broadcasting - Satellite to Handheld
EDGE	Enhanced Data rates for Global Evolution
E-UTRAN	Evolved UTRAN
FDD	Frequency Division Duplex
GERAN	GSM/EDGE RAN
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile communications
IC	Integral Circuit
ID	Identity Code
iid	independent and identically distributed
LTE	Long Term Evolution
MAP	Maximum a Posterior
MCMC	Markov Chain Monte Carlo
NEES	Normalized Estimation Error Squared
OMA	Open Mobile Alliance
PDA	Personal Digital Assistance
pdf	probability distribution function
RAN	Radio Access Network
RF	Radio Frequency
RM	Radio Map
RSS	Received Signal Strength
RTT	Round Trip Time
spd	symmetric positive definite
TA	Timing Advance



TDOA	Time Difference of Arrival
TOA	Time Of Arrival
UE	User Equipment
UMTS	Universal Mobile Telecommunication System
UTRAN	UMTS Terrestrial RAN
WCDMA	Wideband CDMA
WLAN	Wireless Local Area Network

# Symbols

$\int$	Integral
$\otimes$	Kronecker product
$-1$	Inverse of matrix
$T$	Transpose of matrix
$\frac{\partial y}{\partial x}$	Partial differential of $y$ with respect to $x$
$\mathbf{1}_n$	$n$ variate column vector of ones
$\mathbf{c}$	List of communication nodes
$c_i$	$i$ th communication node
$\text{cof}(x_{ij})$	Cofactor of $x_{ij}$
$dx$	Differential of $x$
$dX$	Matrix of differentials
$E(\mathbf{x})$	Expectation value of random variable $\mathbf{x}$
$\exp(x)$	Exponent function
$e_i$	$i$ th eigenvalue of covariance matrix
$\Gamma(x)$	Gamma function
$\Gamma_p(x)$	Generalized gamma function
$I(x)$	Fishers information of $x$
$I_p$	$p$ -variable identity matrix
$I_v(x)$	Modified Bessel function
$J_{y \rightarrow x}$	Jacobian of transformation $x = f(y)$
$l(\mathbf{x})$	Likelihood function of random variable $\mathbf{x}$
$\log$	Natural logarithm
$L(x)$	Loss function
$\text{mode}(\mathbf{x})$	Mode of random variable $\mathbf{x}$
$\mu$	Matrix valued random variable
$\boldsymbol{\mu}$	Vector valued random variable or expected value
$(\Omega, \mathcal{F}, P)$	Probability space
$p(\mathbf{x}), p_X(x)$	Probability density function of random variable $\mathbf{x}$
$p(\mathbf{x}, \mathbf{y})$	Joint probability density function of random variables $\mathbf{x}$ and $\mathbf{y}$
$p(\mathbf{x} \mathbf{y})$	Conditional probability density function of random variable $\mathbf{x}$ given $\mathbf{y}$
$\pi(\mathbf{x})$	Posterior distribution of random variable $\mathbf{x}$
$\mathbb{R}$	set of real numbers
$R$	Radio map

rank	Rank of matrix
$\sigma_{ij}$	Element from $i$ th row and $j$ th column of matrix $\Sigma$
$\sigma^{ij}$	Element from $i$ th row and $j$ th column of matrix $\Sigma^{-1}$
$\Sigma$	Covariance matrix or matrix valued random variable
$\sum_i$	Sum of components with index $i$
$\theta$	Parameter vector
tr	Trace of matrix
$\mathbf{v}$	State model error
$\text{Var}(\mathbf{x})$	Variance of random variable $\mathbf{x}$
$x$	Scalar valued random variable
$\mathbf{x}$	Vector valued random variable or vector
$\mathbf{x}^{(i)}$	Vector containing $i$ th row of matrix $\mathbf{X}$
$\mathbf{x}_i$	Vector containing $i$ th column of matrix $\mathbf{X}$
$\mathbf{x}_t, \mathbf{x}(t)$	Stochastic process of state
$X_{ij}$	Minor of $x_{ij}$
$X_{(s)}$	Matrix containing $s$ rows of matrix $\mathbf{X}$
$X_{ij}$	$i \times j$ submatrix of $X$
$X^{ij}$	$i \times j$ submatrix of $X^{-1}$
$X_{22.1}$	Schur complement of $X_{11}$ in $\mathbf{X}$
$  \mathbf{X}  $	Absolute value of $ \mathbf{X} $
$ \mathbf{X} $	Determinant of matrix $\mathbf{X}$
$\mathbf{x} \sim N_p(\mu, \Sigma)$	Random variable $\mathbf{x}$ follows $p$ -variate normal distribution with parameters $\mu$ and $\Sigma$
$\mathbf{X} \sim N_{qp}(\mu, \Sigma)$	Random variable $\mathbf{X}$ follows $q \times p$ -variate matrix normal distribution with parameters $\mu$ and $\Sigma$
$x \sim \Gamma(\alpha, \beta)$	Random variable $x$ follows gamma distribution with parameters $\alpha$ and $\beta$
$x \sim \chi_{m,p}^2$	Random variable $x$ follows chi-square distribution with parameters $m$ and $p$
$x \sim \chi_p^2$	Random variable follows central chi-square distribution with parameter $p$
$\mathbf{X} \sim W_p(\Sigma, n)$	Random variable $\mathbf{X}$ follows $p$ -variate Wishart distribution with parameters $\Sigma$ and $n$
$\mathbf{X} \sim W_p^{-1}(\Sigma, n)$	Random variable $\mathbf{X}$ follows $p$ -variate inverse Wishart distribution with parameters $\Sigma$ and $n$

$\mathbf{x} \sim t_p(\mu, A, d)$	Random variable $\mathbf{x}$ follows $p$ -variate $t$ distribution with parameters $\mu$ , $A$ and $d$
$X \sim t_{pq}(\mu, A, B, d)$	Random variable $X$ follows $p \times q$ -variate matrix $t$ distribution with parameters $\mu$ , $A$ , $B$ and $d$

**Boldface symbols** ( $\mathbf{x}, \mathbf{y}$ ) refer to a vector valued random variable, a column vector or a stochastic process depending on context.

**Capital letters** ( $X, Y, A$ ) refer to a matrix valued random variable or a matrix depending on context.



# 1 Introduction

Easy access to and availability of location information have both led to unforeseen increase in location based services. Location services are no longer used just for locating emergency calls or in turn-by-turn navigation. Nowadays location awareness is more often in touch with commercial services, where coarser positioning accuracy is acceptable. Examples of such of location based services include a local weather forecast, finding the nearest restaurant and checking for events in certain area.

The aim of this thesis is to introduce a Bayesian approach to model a coverage area of a CN (Communication Node) using location fingerprinting and introduce how coverage area information can be used to locate a UE (User Equipment) in a wireless communication network. Coverage area of a CN refers to a geographical area where at least the ID (identity) of the CN can be decoded. Location of a UE refers to two or three dimensional coordinates, either global or local, of the UE. CN may be for example a base station in a cellular network, a WLAN (Wireless Local Area Network) access point, a radio station or a TV station or some other local wireless network like in [1]. A UE to be located may be for example laptop, mobile phone or PDA (Personal Digital Assistant) or some other device equipped with a radio. Motivation for this work is to find a positioning method which is available where current positioning methods lack availability, is a low cost and a low power consumption solution and can be used in UEs with lack of satellite-based positioning methods.

Satellite-based (GNSS - Global Navigation Satellite System) positioning methods such as GPS (Global Positioning System) are usually very accurate outdoors, but in dense urban and indoor environment the line of sight paths to satellites may be blocked leading to degraded availability and multipath of signals leading to degraded accuracy. However, these are environments where consumers spend most of their time and use location services. If the UE is connected to some wireless communication network, the network-based positioning is a good solution in urban areas, where the density of different CNs, like cellular base stations and WLAN access points is high.

UEs which are equipped with positioning capability usually need special hardware, for example GNSS-based positioning requires a special IC (Integral Circuit) which decodes navigation data from satellites and tracks their ranging signals. Fingerprinting-based positioning uses the information which is freely available in a communication network and uses the resources anyhow in the UE. Fingerprinting-based method is thus a software solution only and no additional hardware is needed, which reduces costs of producing UEs. This is why this positioning technique can be utilized in all the UEs.

In GNSS-based positioning the UE decodes and tracks navigation data from satellites, which requires energy. In contrast, in fingerprinting-based positioning the UE is always aware of the wireless environment and

measurements come for free in terms of energy. Thus it requires no extra energy to do the required measurements from the network for fingerprint-based positioning. In WLAN-based positioning the WLAN scan costs energy but is less power hungry than GNSS, see Appendix A.

## 1.1 Fingerprinting

Location fingerprinting is a positioning method that determines the location of a UE using a database of radio characteristics. A fingerprint can be defined as a set of radio characteristics recorded from a variety of radio networks together with a position of the record. The characteristics can be used either individually or in combination and may include a frequency of a signal, an identifier of the transmitter, that is the ID of CN, signal strength, a time stamp, a time difference between CNs and time delays. A RF (Radio Frequency) fingerprint includes in addition to location, only the parameters of the received signal, such as RSS (Received Signal Strength), a frequency of signal and the ID of CN and can be considered as a subset of the characteristics defined before. In this work, only RF fingerprints are considered.

Fingerprinting method contains two phases, a data collecting phase and a position determining phase. In the data collecting phase fingerprints are collected and stored in a database. Fingerprints may be collected doing calibration measurements in various locations in an area of interest, fingerprints may be generated from some model or as a combination of previous. The fingerprint database is processed and used to generate an RM (Radio Map), which covers the area of interest and contains information about radio signal properties as a function of position. In the position determining phase the UE samples RF information from CNs and searches for similar patterns in the RM. Position estimate is made by comparing the RM with the measured fingerprint using different kinds of algorithms. Figure 1 illustrates the structure of fingerprinting method.

If the data collecting is done using calibration measurements, fingerprints may be collected with a UE with a positioning capacity (for example GNSS). Indoors GNSS is not usually available or accurate and data collecting could be done for example as a documented walk through the building as described in [2]. If fingerprints are generated using models, parameters estimated by a model are stored in the database together with location report. If the transmit power, for example, is known, fingerprints may be generated using different signal propagation models [3], [4]. Fingerprints may also be generated as a combination of previous, when some actual calibration measurements are done and new fingerprints are interpolated using existing measurements [5]. Usually, an RM consist of estimated coverage areas of CNs but it may also be just a set of grid points, where every grid point contains location and radio characteristics at that point. In the latter case for example

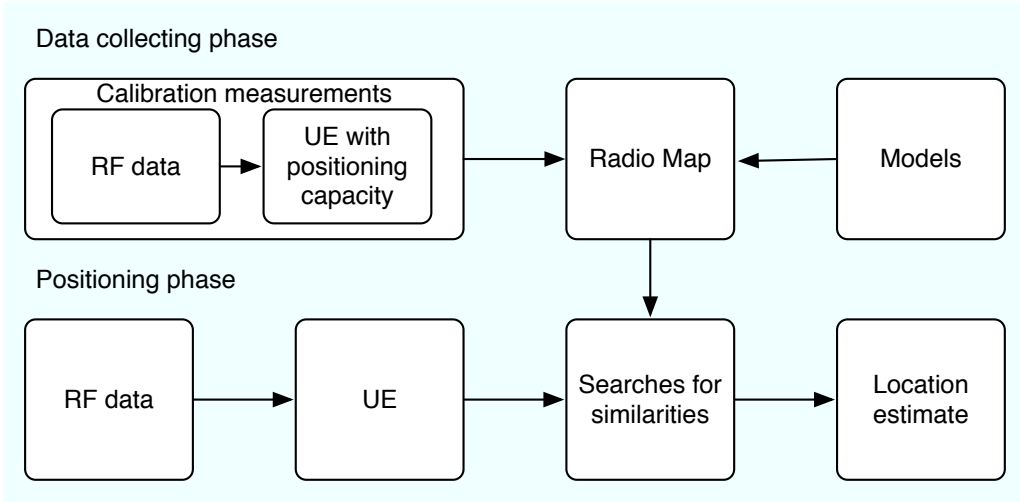


Figure 1: Structure of fingerprint method. Data collecting and positioning are individual processes and operate in parallel in real world deployments.

several signal strength measurements are collected from the same location but only mean of signal strengths is stored in the RM [6]. If the RM is built as a set of grid points the positioning accuracy depends highly on the density of grid points [6], [7]. RMs can be designed to take into account the dynamic variations in the RF signal characteristics. The transmit power of the CN, for example, may vary according to the number of users, which reflects also on coverage area of CN. For this reason the time at which a fingerprint was captured is stored so RMs may accommodate such dynamic changes. One may assume, for example, that during night time the user density is lower reflecting to decreased transmit powers which lead to smaller coverage areas.

Several other location systems have been proposed in wireless networks, but standardized technologies are available only in the cellular network. In following, when it is discussed about cellular GSM (Global System for Mobile communications), WCDMA (Wideband CDMA, CDMA-Code Division Multiple Access) and LTE (Long Term Evolution) networks, it is referred to 3GPP-standardized (Third Generation Partnership Project) GERAN (GSM/EDGE RAN, EDGE-Enhanced Data rates for Global Evolution, RAN-Radio Access Network), UTRAN-FDD (UMTS Terrestrial RAN Frequency-Division Duplex, UMTS-Universal Mobile Telecommunications System) and E-UTRAN (Evolved UTRAN) networks, respectively [8],[9],[10].

All standardized technologies involve using characteristics of radio signals transmitted by, or received by a base station at known locations to determine the location of a UE. Measurements may be for example signal strength, angle of arrival, time of arrival, time-difference of arrival and time delay of signals from the UE at various CNs. Easiest location technique is to use an ID of CN to determine the position of the UE. The UE estimates its location to



be same as that of the CN to which it is associated. This method is very reliable, however the accuracy of the position is poor due to the possibly large cell size. This is used in all cellular networks and is referred to a CID-method (Cell Identity). Publication [11] cites 1350-meter 95% accuracy for CID-method in a sub-urban environment and 600-meter 95% accuracy in an urban environment in the GSM network. Results show that in the urban areas, where density of base stations is higher and thus sizes for cells is smaller, the accuracy of the CID-method improves significantly.

TA (Timing Advance), in WCDMA networks RTT (round-trip time), is used in cellular networks also for positioning. TA measurements are available only with respect to the serving base station and creates a ring around the serving base station and spread of the ring depends on the used RAN. Generally positioning accuracy is poor using TA measurements, but it can be combined to other measurements, for example coverage area information to increase accuracy. It is found in [11] that in the GSM network TA measurements combined with coverage area information results 800-meter 95% accuracy in the sub-urban environment and 550-meter 95% accuracy in the urban environment.

Signal attenuation methods, which are based on signal strength measurements, have been found to be promising in outdoor environments but these methods cannot be used reliably to compute the position of a UE in indoor environments with obstructions and reflecting objects [12]. Most common model to estimate path loss of signal is to use the Okamura-Hata model, which is based on experimental measurements [13]. This model requires lot of information about the environment, such as height of a CN, height of a UE, size of a city and transmission power of the signal. Signal strength is used in cellular network as assistance to an ECID (Enhanced CID) method. In publication [11] authors achieve 960-meter 95% accuracy in sub-urban and 420-meter 95% accuracy in urban areas in the GSM network for positioning method, which combines signal strength and TA measurements with coverage area information.

Location schemes based on TOA (Time Of Arrival) requires fine time synchronization between the transmitters, which is done in cellular network only in CDMA/CDMA2000 networks in the US [14]. Time difference-based methods like TDOA (Time Difference Of Arrival) are used in the cellular GSM (E-OTD - Enhanced Observed Time Difference), WCDMA (IPDL-OTDOA - Idle Period DownLink Observed TDOA) and LTE networks and could be possible also in the mobile TV network based on European DVB-SH (Digital Video Broadcasting - Satellite to Handheld) standard as described in [15]. In the TDOA UE-CN time offset is removed from measurement equations but the method needs real time difference measurements between base stations. In cellular networks there are location measurement units, which measure real time differences between base stations. The TDOA method requires at least three heard base stations to provide 2D location

and four heard base station to provide 3D location. AOA (Angle Of Arrival) method needs a special hardware at base stations and thus is not used in current positioning techniques, but it is becoming a positioning solution in the cellular LTE network [10]. Positioning accuracy varies in all previous methods and depends highly on the environment, because all these methods require the line of sight conditions to ensure an acceptable accuracy and this condition is not always met in the indoor and urban environments.

During the data collection phase fingerprinting takes into account the effects that obstructions will have on the RF signal, such as reflection and attenuation. This makes the fingerprinting method more detailed, precise and reliable even in very complex environments. Drawbacks in the fingerprinting method compared with others are how to perform the data collecting; also the size and location of the database may have impact on the cost and accuracy [16]. The database has to be well designed and it has to be up-to-date so that the position estimate could be calculated in real time and be accurate. Especially in urban environment, landscape changes constantly when new buildings, for example, are built. They block and may reflect radio signals so that RM would not be accurate any longer. So although accuracy is good when RM is up-to-date, integrity is poor because landscape changes constantly [17, p. 100]. Location algorithms should be designed in a way that position estimate would be easy to calculate so that also a UE with low computing capacity could make the calculations.

Fingerprinting techniques can be categorized in two different categories based on their positioning algorithms. Deterministic techniques model the measured data of an access point at location by a scalar value and use a non-probabilistic approach to infer the location of a UE. Location estimate can be found by finding a fingerprint from RM with the closest match or by taking several fingerprints from the database and use algorithms to find the best location estimate. Usually deterministic approaches use the nearest neighbor found method, that is, a set of database points with the closest match are found and the position estimate is a weighted mean of those points using different kinds of weights [18]. Deterministic techniques include also neural networks based fingerprinting methods [19]. Probabilistic techniques store measured data as distributions from the CNs in the RM and use probabilistic techniques to estimate the location of a UE. For example [20] and [21] use a location estimation method based on a statistical signal strength model.

## 1.2 Existing solutions

Although fingerprint method is not currently in any standards, several RF-fingerprinting-based solutions are already available, at least which uses WLAN access points or cellular network. For indoor positioning there are for example RADAR [18], Ekahau [22], and Horus [21] which use WLAN access points. For outdoor positioning and tracking there are solutions such

as Polaris Wireless [23], Place Lab [24], Navizon [25], DCM [26] and Skyhook [27]. Some of these solutions are commercial services, such as [23] [25], [27] and [22]. A few of these are presented in following sections. Fingerprinting method is proposed to be standardized and work will start most likely in early 2010 in OMA (Open Mobile Alliance) [28], [29], [30], [31].

Fingerprinting method called DCM (Database Correlation Method) was first introduced in [26]. The DCM measures signal strength in various location to create a database and it was designed so that it would be possible to utilize in GSM networks. Difference between place of the user equipment to be located and reference fingerprint is calculated as

$$d(k) = \sum_i (f_i - g_i(k))^2 + p(k),$$

where  $f_i$  is the signal strength UE measures from the  $i$ th cell and  $g_i(k)$  is the signal strength of  $k$ th fingerprint from same cell. Summation is made over all cells that were hearable both to UE and to reference fingerprint. Penalty term  $p(k)$  is added to all the cells that are found only once in fingerprint database. In [32] DCM is presented for WCDMA networks, where in addition to signal strength also RTT measurements are made to limit the area where UE is located. Improvements for DCM, like modification of penalty term and filtering position estimate, are presented in [33]. Also methods where database is formed using predicted fingerprints are presented. For example, the database may be formed using different signal propagation models [3],[4]. In these approaches it is less expensive and faster to form a database, but it doesn't take as accurately into account multipath and non-line-of-sight conditions of radio signals in inbuilt environment.

The first WLAN-based system using the fingerprinting approach was RADAR, described in [18] and this can be considered as a fundamental research for location fingerprinting indoors as a first RF-based system for locating and tracking users inside buildings. RADAR measures signal strength in various calibration points and calculates means, medians and variances of signal strength for every calibration point. A location estimate is determined by comparing measurement with the nearest neighbors in RM using Euclidean distance.

Ekahau is a commercial location system using signal strength measurements from WLAN access points in fingerprinting. Ekahau is capable of locating mobile phones, laptops, PDAs and any other WLAN enabled devices [22]. Ekahau is used for example in hospitals to locate devices or patients and in supermarkets to locate shopping carts [34], [35], [36]. Ekahau's solution is for accurate positioning and is the RM is tailored for every target and is available only in those areas of interest.

Skyhook has their own global database of WLAN access points. To develop this database, Skyhook has employees who find WLAN access points and cellular base stations in cities and towns worldwide plotting their precise

geographical locations. Also the service users can submit their own access point to the WLAN database. Skyhook uses this database to provide location estimate and the system claims 10-meter accuracy [27]. Skyhook's solution has good accuracy but is available only in places where WLAN access points are scanned.

Polaris Wireless provides location services especially to track emergency calls. The system collaborates with operators and gets information, such as the locations of base stations and transmit powers, from the core network. An RM is generated mostly using signal attenuation models but also some calibration measurements are made. Polaris Wireless achieves 44-meter 67% accuracy and 135-meter 95% accuracy in urban environment using GSM network [37]. Polaris Wireless's solution has good integrity, because it is aware of all the changes in network, such as changes in transmit powers, new CN IDs and disappearing CN IDs.

### 1.3 Scope and structure of thesis

Scope of this thesis is to present a low cost, software-based positioning technique, which uses measurements from a radio network. Positioning technique uses fingerprinting method in order to create an RM, which includes the coverage area estimates of wireless CNs. During data collecting, only IDs of communication nodes are stored in a database together with the coordinates of a fingerprint and fingerprints are used to estimate coverage areas of communication nodes in a Bayesian framework using a multivariate normal linear model. The Bayesian framework is used in order to be able to use the prior information about coverage area. Positioning is done using the coverage area information of CNs.

Section 2 deals with preliminaries needed in the thesis. It provides a summary of distributions, together with some properties used in this work, and Bayesian inference, like Bayes' rule, some useful prior distributions and Bayes' estimators. Finally, a brief presentation about Bayesian filtering is made.

In Section 3 multivariate normal linear model is presented in a Bayesian framework. The multivariate normal linear model is used in estimating coverage area of communication node and it is presented using two different priors. Also novel methods for updating the coverage area estimate are presented. A Bayesian method for positioning a UE using coverage area of communication node is presented in Section 4.

In Section 5 methods presented in previous sections are used to estimate coverage area of communication node, to study how the coverage area model of a CN is updated when new observations come in and to explore, how position accuracy and consistency vary using different coverage area estimates.

## 2 Preliminaries

Preliminaries needed in this thesis are presented briefly in this section. Section 2.1 deals with distributions used in this work and some of their properties are derived. Section 2.2 considers Bayesian inference, including presentation of Bayes' rule and discussion of the choice of prior distribution and Bayes' estimator. This section ends with a short presentation of Bayesian filters.

### 2.1 Distributions

**Definition 2.1.** Let  $A \in \mathbb{R}^{p \times p}$  be a symmetric square matrix and  $\mathbf{x} \in \mathbb{R}^p$ .  $A$  is said to be a symmetric positive definite (spd) matrix, if

$$\mathbf{x}^T A \mathbf{x} > 0 \quad (2.1)$$

for all  $\mathbf{x} \neq 0$ .

**Definition 2.2.** Let  $\mathbf{x}$  be a  $p$ -variate vector valued random variable with an expected value  $E(\mathbf{x}) = \boldsymbol{\mu}$  and a covariance matrix  $\text{Var}(\mathbf{x}) = \Sigma$ , where  $\Sigma$  is a  $p \times p$  spd matrix. Then a random variable  $\mathbf{x}$  is said to have  $p$ -variate normal distribution with parameters  $\boldsymbol{\mu}$  and  $\Sigma$ , denote  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$ , if its pdf (probability density function) is

$$p(\mathbf{x}) = |2\pi\Sigma|^{-1/2} \exp \left[ -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right]. \quad (2.2)$$

**Definition 2.3.** Let  $X$  be a  $n \times p$  matrix whose rows  $\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_n^T$  are mutually independent and distributed as  $N_p(\boldsymbol{\mu}, \Sigma)$ . Then  $X$  has a matrix normal distribution with pdf

$$\begin{aligned} p(X) &= |2\pi\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu}) \right] \\ &= |2\pi\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1}(X - \mathbf{1}_n \boldsymbol{\mu}^T)^T (X - \mathbf{1}_n \boldsymbol{\mu}^T) \right]. \end{aligned} \quad (2.3)$$

Denote  $X \sim N_{np}(\mathbf{1}_n \boldsymbol{\mu}^T, \Sigma)$ .

**Theorem 2.1.** If  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Sigma)$  and  $\mathbf{y} = A\mathbf{x} + \mathbf{b}$ , where  $A$  is any  $q \times p$  matrix with  $\text{rank } A = q$  and  $\mathbf{b}$  is a  $q$ -variate vector,  $\mathbf{b} \in \mathbb{R}^q$ , then  $\mathbf{y} \sim N_q(A\boldsymbol{\mu} + \mathbf{b}, A\Sigma A^T)$ .

*Proof.* See [38, p. 67]. □

If  $\text{rank } A < q$ , then  $\text{rank } A\Sigma A^T < q$  and  $\mathbf{y}$  follows singular normal distribution. Singular normal distribution is not considered in this work, but definition and some properties can be found in [39, p. 41–42].

**Definition 2.4.** A random variable  $x > 0$  is said to be gamma-distributed with a shape parameter  $\alpha > 0$  and a scale parameter  $\beta > 0$ , denote  $x \sim \text{Gamma}(\alpha, \beta)$ , if its pdf is

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp[-\beta x], \quad (2.4)$$

where  $\Gamma$  is a gamma function [40, p. 165].

If  $x \sim \text{Gamma}(\alpha, \beta)$  its expected value is  $E(x) = \frac{\alpha}{\beta}$  and variance is  $\text{Var}(x) = \frac{\alpha}{\beta^2}$  [41, p. 451].

**Definition 2.5.** Let  $\mathbf{y}$  be a  $p$ -variate random vector and  $\mathbf{y} \sim N_p(\boldsymbol{\mu}, I_p)$ . Then  $x = \mathbf{y}^T \mathbf{y}$  is said to have a chi-square distribution, with a noncentrality parameter  $m = \boldsymbol{\mu}^T \boldsymbol{\mu}$  and a freedom parameter  $p$ , denote  $y \sim \chi_{p,m}^2$ , with pdf

$$p(x) = \frac{1}{2} (x/m)^{(p-2)/4} I_{(p-2)/2}(\sqrt{mx}) \exp[-(m+x)/2], \quad (2.5)$$

where  $I_v$  is modified Bessel function [42].

When  $\boldsymbol{\mu} = \mathbf{0}$ , the probability distribution function becomes

$$p(x) = \frac{1}{2^{p/2} \Gamma(\frac{1}{2}p)} x^{(p/2-1)} \exp[-x/2], \quad (2.6)$$

which is a central chi-square distribution, denote  $x \sim \chi_p^2$ . Note that central chi-square distribution is a special case of gamma distribution, namely  $\text{Gamma}(p/2, 1/2)$ .

**Definition 2.6.** Let  $X$  be a  $n \times p$  data matrix with iid (independent and identically distributed) rows from  $N_p(\mathbf{0}, \Sigma)$ . Then  $U = X^T X$  is said to have a Wishart distribution with a scale matrix  $\Sigma$  and a freedom parameter  $n$ , denote  $U \sim W_p(\Sigma, n)$ , where  $\Sigma > 0$  and  $n \geq p$ , and has a pdf

$$p(U) = \frac{|U|^{(n-p-1)/2} \exp\left[-\frac{1}{2} \text{tr} \Sigma^{-1} U\right]}{2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^p \Gamma(\frac{1}{2}(n+1-i))}. \quad (2.7)$$

Note that the expected value of  $U$  is  $E(U) = E(X^T X) = \sum_{i=1}^n E(\mathbf{x}_i \mathbf{x}_i^T) = n\Sigma$ . When  $p = 1$ , Wishart distribution reduces to gamma distribution  $\text{Gamma}(n/2, 1/2\sigma)$ . Moments of the components of a Wishart matrix  $U$  are given in [38, p. 115].

**Theorem 2.2.** If  $U \sim W_p(\Sigma, n)$  and  $B$  is a constant  $p \times q$  matrix, then  $B^T U B \sim W_q(B^T \Sigma B, n)$ .

*Proof.* Since  $B^T U B = B^T X^T X B = Y^T Y$ , where  $Y = X B$  and rows of  $X$  are iid  $N_p(\mathbf{0}, \Sigma)$  so according to Theorem 2.1 rows of  $Y$  are iid  $N_q(\mathbf{0}, B^T \Sigma B)$ . Using Definition 2.6,  $Y^T Y$  has the stated distribution.  $\square$

**Theorem 2.3.** If  $U \sim W_p(\Sigma, n)$ , then  $V = U^{-1}$  has pdf

$$p_V(V) = \frac{|V|^{-(n+p+1)/2} \exp\left[-\frac{1}{2} \text{tr} \Sigma^{-1} V^{-1}\right]}{2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n+1-i)\right)}. \quad (2.8)$$

*Proof.*

$$\begin{aligned} p_V(V) &= p_U(V^{-1}) J_{U \rightarrow V} \\ &\stackrel{(B.12)}{=} \frac{|V|^{-(n-p-1)/2} \exp\left[-\frac{1}{2} \text{tr} \Sigma^{-1} V^{-1}\right]}{2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n+1-i)\right)} |V|^{-p-1} \\ &= \frac{|V|^{-(n+p+1)/2} \exp\left[-\frac{1}{2} \text{tr} \Sigma^{-1} V^{-1}\right]}{2^{np/2} \pi^{p(p-1)/4} |\Sigma|^{n/2} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n+1-i)\right)}. \end{aligned}$$

□

$V = U^{-1}$  is said to follow an inverse Wishart distribution with a scale matrix  $\Sigma^{-1}$  and a freedom parameter  $n$ , denote  $V \sim W_p^{-1}(\Sigma^{-1}, n)$ .

In following, some properties of an inverse Wishart distribution will be presented. Consider the partition of a Wishart matrix  $V$

$$V = \left( \begin{array}{c|c} V_{11} & V_{12} \\ \hline V_{21} & V_{22} \end{array} \right), \quad (2.9)$$

where  $V_{11}$  is a  $q \times q$  submatrix of  $V$  and  $1 \leq q \leq p$ . Denote

$$\Sigma^{-1} = \left( \begin{array}{c|c} \Sigma^{11} & \Sigma^{12} \\ \hline \Sigma^{21} & \Sigma^{22} \end{array} \right). \quad (2.10)$$

**Theorem 2.4.**  $V_{11} \sim W^{-1}(\Sigma^{11}, n - (p - q))$ .

*Proof.* Because  $V$  is nonsingular, it can be written in form

$$V = \begin{pmatrix} I & 0 \\ V_{21} V_{11}^{-1} & I \end{pmatrix} \begin{pmatrix} V_{11} & 0 \\ 0 & V_{22} - V_{21} V_{11}^{-1} V_{12} \end{pmatrix} \begin{pmatrix} I & V_{11}^{-1} V_{12} \\ 0 & I \end{pmatrix}. \quad (2.11)$$

Denoting  $V_{22.1} = V_{22} - V_{21} V_{11}^{-1} V_{12}$ , which is the Schur complement of  $V_{11}$  in  $V$ , Equation (2.11) shows that

$$|V| = |V_{11}| |V_{22.1}| \quad (2.12)$$

and

$$V^{-1} = \begin{pmatrix} V_{11}^{-1} + V_{11}^{-1} V_{12} V_{22.1}^{-1} V_{21} V_{11}^{-1} & -V_{11}^{-1} V_{12} V_{22.1}^{-1} \\ -V_{22.1}^{-1} V_{21} V_{11}^{-1} & V_{22.1}^{-1} \end{pmatrix}. \quad (2.13)$$

Writing  $\dot{V} = V_{11}^{-1}V_{12}$ , Equation (2.13) can be written in form

$$V^{-1} = \begin{pmatrix} V_{11}^{-1} & 0 \\ 0 & 0 \end{pmatrix} + M, \quad (2.14)$$

where

$$M = \begin{pmatrix} \dot{V}V_{22.1}^{-1}\dot{V}^T & -\dot{V}V_{22.1}^{-1} \\ -V_{22.1}^{-1}\dot{V} & V_{22.1}^{-1} \end{pmatrix} \quad (2.15)$$

and the distribution in Equation (2.8) can be written in form

$$p(V) \propto [|V_{11}||V_{22.1}|]^{-(n+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{11} V_{11}^{-1} - \frac{1}{2} \text{tr} \Sigma^{-1} M \right]. \quad (2.16)$$

The Jacobian of the transformation from  $(V_{12}, V_{22})$  to  $(V_{22.1}, \dot{V})$  for fixed  $V_{11}$  is

$$J_{(V_{12}, V_{22}) \rightarrow (V_{22.1}, \dot{V})} = \left\| \frac{\partial(V_{12}, V_{22})}{\partial(V_{22.1}, \dot{V})} \right\| = \left\| \begin{array}{cc} \frac{\partial V_{12}}{\partial V_{22.1}} & \frac{\partial V_{22}}{\partial V_{22.1}} \\ \frac{\partial V_{12}}{\partial \dot{V}} & \frac{\partial V_{22}}{\partial \dot{V}} \end{array} \right\| = \left\| \begin{array}{cc} J_{11} & \mathbf{I} \\ \frac{\partial V_{12}}{\partial \dot{V}} & 0 \end{array} \right\|. \quad (2.17)$$

$V_{22}$  does not depend on  $\dot{V}$  and thus  $J_{22}$  is equal to zero. This is why  $J_{11}$  in Jacobian (2.17) is not calculated, since on  $J_{11}$  it would be multiplied with zero. Thus  $J_{(V_{12}, V_{22}) \rightarrow (V_{22.1}, \dot{V})}$  is equal to  $J_{V_{12} \rightarrow \dot{V}}$ . Using (B.13), the Jacobian becomes  $J_{(V_{12}, V_{22}) \rightarrow (V_{22.1}, \dot{V})} = J_{V_{12} \rightarrow \dot{V}} = |V_{11}|^{p-q}$ .

Noting that  $M$  doesn't depend on  $V_{11}$ , the joint distribution of  $(V_{11}, V_{22.1})$  can be written in form

$$\begin{aligned} p(V_{11}, V_{22.1}) &\propto |V_{11}|^{-(n-(p-q)+q-1)} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{11} V_{11}^{-1} \right] \\ &\quad \times |V_{22.1}|^{-(n+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} M \right] \\ &= p(V_{11})p(V_{22.1}), \end{aligned} \quad (2.18)$$

so  $V_{11} \sim W^{-1}(\Sigma^{11}, n - (p - q))$ .  $\square$

Equation (2.18) shows that  $V_{11}$  and  $V_{22.1}$  are mutually independent and  $V_{22.1} \sim W^{-1}(M, n)$ .

**Theorem 2.5.** *If  $V \sim W_p^{-1}(\Sigma^{-1}, n)$  and  $A$  is a  $q \times p$  matrix with  $\text{rank } A = q$ , then  $AVA^T \sim W_q^{-1}(A\Sigma^{-1}A^T, n - p + q)$ .*

*Proof.* Because  $V \sim W_p^{-1}(\Sigma^{-1}, n)$ ,  $U = V^{-1} \sim W_p(\Sigma, n)$ . Any  $q \times p$  matrix of rank  $q$  can be written as  $A = B [I_q, 0] \Lambda \Sigma^{1/2}$ , where  $B$  is  $q \times q$  non-singular matrix, and  $\Lambda = \begin{bmatrix} \Lambda_1 \\ \Lambda_2 \end{bmatrix}$  is an orthogonal  $p \times p$  matrix, that is

$$\Lambda^T \Lambda = \Lambda \Lambda^T = I_p \Leftrightarrow \Lambda^{-1} = \Lambda^T. \quad (2.19)$$



$\Lambda_1$  consists of  $q$  first rows of  $\Lambda$  and is an orthonormal basis for the rows of  $A\Sigma^{-1/2}$ .  $\Lambda_2$  is a  $(p-q) \times p$  matrix and its rows are orthogonal to the rows of  $A\Sigma^{-1/2}$ , so  $A\Sigma^{-1/2}\Lambda_2^T = 0$ . Let  $C = \Lambda\Sigma^{-1/2}U\Sigma^{-1/2}\Lambda^T$ . According to Theorem 2.2,  $C \sim W_p(\Lambda\Sigma^{-1/2}\Sigma\Sigma^{-1/2}\Lambda^T, n) = W_p(I_p, n)$ . Now

$$\begin{aligned} AU^{-1}A^T &= B [I_q, 0] \Lambda\Sigma^{1/2}U^{-1}\Sigma^{1/2}\Lambda^T [I_q, 0]^T B^T \\ &= B [I_q, 0] C^{-1} [I_q, 0]^T B^T \\ &= BC^{qq}B^T, \end{aligned} \quad (2.20)$$

where  $C^{qq}$  is the upper left hand matrix of  $C^{-1}$ . According to Theorem 2.4  $C^{qq} \sim W_q^{-1}(I_q, n - (p - q)) = W_q^{-1}(I_q, n - p + q)$ , so  $(C^{qq})^{-1} \sim W_q(I_q, n - p + q)$ . Now  $(AU^{-1}A^T)^{-1} = (B^T)^{-1}(C^{qq})^{-1}B^{-1}$ . Using again Theorem 2.2 results in  $(AU^{-1}A^T)^{-1} \sim W_q((BB^T)^{-1}, n - p + q)$  and therefore  $AU^{-1}A^T \sim W_q^{-1}(BB^T, n - p + q)$ . But

$$\begin{aligned} B &= A\Sigma^{-1/2}\Lambda^T [I_q, 0]^T \\ &= A\Sigma^{-1/2}\Lambda_1^T + A\Sigma^{-1/2}\Lambda_2^T \\ &= A\Sigma^{-1/2}\Lambda_1^T \end{aligned} \quad (2.21)$$

and since

$$\begin{aligned} BB^T &= A\Sigma^{-1/2}\Lambda_1^T\Lambda_1\Sigma^{-1/2}A^T \\ &= A\Sigma^{-1}A^T, \end{aligned} \quad (2.22)$$

and one can notice that  $AU^{-1}A^T = AVA^T \sim W_q^{-1}(A\Sigma^{-1}A^T, n - p + q)$ . □

**Theorem 2.6.** *If  $V \sim W_p^{-1}(\Sigma^{-1}, n)$ ,  $n \geq p + 2$ , then  $E(V) = \frac{\Sigma^{-1}}{n-p-1}$ .*

*Proof.* It follows from Theorem 2.5 that for any  $q \times p$  matrix  $A$  with rank  $A = q$ ,  $AVA^T \sim W_q^{-1}(A\Sigma^{-1}A^T, n - p + q)$ . Choosing  $A = \mathbf{a}^T$ , where  $\mathbf{a}$  is a  $p$ -variate vector, it follows that  $\mathbf{a}^T V \mathbf{a} \sim W_1^{-1}(\mathbf{a}^T \Sigma^{-1} \mathbf{a}, n - p + 1)$ . Thus

$$\begin{aligned} (\mathbf{a}^T V \mathbf{a})^{-1} &\sim W_1((\mathbf{a}^T \Sigma^{-1} \mathbf{a})^{-1}, n - p + 1) \\ &= \text{Gamma}((n - p + 1)/2, (\mathbf{a}^T \Sigma^{-1} \mathbf{a})/2). \end{aligned} \quad (2.23)$$

Next it is shown that, if  $x \sim \text{Gamma}(\alpha, \beta)$ , then  $E(x^{-1}) = \frac{\beta}{\alpha-1}$ . The pdf of  $y = x^{-1}$  is

$$\begin{aligned} p_Y(y) &= p_X(y^{-1}) \left| \frac{\partial x}{\partial y} \right| \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha-1)} \exp[-\beta/y] \left| -\frac{1}{y^2} \right| \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} \exp[-\beta/y]. \end{aligned} \quad (2.24)$$

So expected value of  $y$  is

$$\begin{aligned}
E(y) &= \int y p_Y(y) \, dy \\
&= \int y \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)} \exp[-\beta/y] \, dy \\
&= \frac{\beta^\alpha \Gamma(\alpha - 1)}{\beta^{\alpha-1} \Gamma(\alpha)} \\
&= \frac{\beta}{\alpha - 1}.
\end{aligned} \tag{2.25}$$

Thus using Equations (2.23) and (2.25) result in

$$\begin{aligned}
E(\mathbf{a}^T \mathbf{V} \mathbf{a}) &= \frac{\mathbf{a}^T \Sigma^{-1} \mathbf{a}}{n - p + 1 - 2} \\
\Leftrightarrow \mathbf{a}^T E(\mathbf{V}) \mathbf{a} &= \mathbf{a}^T \frac{\Sigma^{-1}}{n - p - 1} \mathbf{a}.
\end{aligned} \tag{2.26}$$

Because Equation (2.26) holds for all constant vectors  $\mathbf{a}$  and both matrices are symmetric it follows that  $E(\mathbf{V}) = \frac{\Sigma^{-1}}{n-p-1}$ .  $\square$

**Theorem 2.7.** *If  $V \sim W_p^{-1}(\Sigma^{-1}, n)$ , then  $\text{mode}(V) = \frac{\Sigma^{-1}}{n+p+1}$ .*

*Proof.* To find the mode of  $V$ , the maximum of  $p(V)$  must be found. This can be done by differentiating  $\log p(V)$  with respect to  $V^{-1}$  because  $\log p(V)$  is maximized at same  $\Sigma$  as  $p(V)$ .  $p(V)$  is regarded as a function of  $V^{-1}$  instead of  $V$  and is thus differentiated with respect to  $V^{-1}$  instead of  $V$

$$\frac{\partial \log p(V)}{\partial V^{-1}} \stackrel{(B.9), (B.10)}{=} (n + p + 1)V - (\Sigma^{-1})^T. \tag{2.27}$$

Because  $\Sigma^{-1}$  is symmetric, equating Equation (2.27) to zero gives

$$V = \frac{\Sigma^{-1}}{n + p + 1}. \tag{2.28}$$

Function  $p(V)$  should really be maximized over the set of all positive definite symmetric matrices, but since unconstrained solution in Equation (2.28) is only turning point of function in Equation (2.8) and it is positive definite symmetric matrix, the constraint does not need to be applied. Thus solution in Equation (2.28) is a unique mode.  $\square$

In the Bayesian theory of multivariate normal sample with both mean and covariance unknown, a distribution called *normal-inverse-Wishart* is usually discussed. Normal-inverse Wishart is a joint distribution of two parameters and is defined next.

**Definition 2.7.** If a random vector  $\boldsymbol{\mu}|\Sigma \sim N(\hat{\boldsymbol{\mu}}, \Sigma)$  with a fixed  $\hat{\boldsymbol{\mu}}$  and random matrix  $\Sigma \sim W^{-1}(S, m)$ , then the joint distribution of

$$p(\boldsymbol{\mu}, \Sigma) = p(\boldsymbol{\mu}|\Sigma)p(\Sigma) \quad (2.29)$$

is said to have normal-inverse-Wishart distribution.

**Definition 2.8.** A  $p$ -variate random vector  $\mathbf{x}$  is said to have a multivariate  $t$  distribution with  $d > 0$  degrees of freedom if its pdf is

$$p(\mathbf{x}) = \frac{\Gamma(\frac{1}{2}(d+p))}{(\pi d)^{p/2} \Gamma(\frac{d}{2}) |\mathbf{T}|^{1/2}} \left[ 1 + (\mathbf{x} - \boldsymbol{\mu})^T (d\mathbf{T})^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]^{-(d+p)/2}, \quad (2.30)$$

where  $\mathbf{T}$  is  $p \times p$  symmetric positive definite matrix, denote  $\mathbf{x} \sim t_p(\boldsymbol{\mu}, \mathbf{T}, d)$ .

The mean and variance of vector  $\mathbf{x}$  are  $E(\mathbf{x}) = \boldsymbol{\mu}$  and  $\text{Var}(\mathbf{x}) = \frac{d}{d-2} \mathbf{T}$ , respectively [43, p. 445].

Next the matrix  $t$  distribution is defined, which is a generalization of the multivariate  $t$  distribution.

**Definition 2.9.** A random  $p \times q$  matrix  $\mathbf{X}$  is said to have a matrix  $t$  distribution with  $d > 0$  degrees of freedom if its pdf is

$$f(\mathbf{X}) = \frac{\Gamma_p(\frac{1}{2}(d+p+q-1))}{(\pi)^{pq/2} \Gamma_p(\frac{1}{2}(d+p-1))} |\mathbf{B}|^{p/2} |\mathbf{A}|^{-q/2} \times |\mathbf{I}_p + \mathbf{A}^{-1}(\mathbf{X} - \boldsymbol{\mu})' \mathbf{B}(\mathbf{X} - \boldsymbol{\mu})|^{-(d+p+q-1)/2}, \quad (2.31)$$

where  $\mathbf{A}$  is a  $q \times q$  spd matrix and  $\mathbf{B}$  is a  $p \times p$  spd matrix and  $\Gamma_p(b)$  is generalized gamma function, see definition [44, p. 427]. Denote  $\mathbf{X} \sim t_{pq}(\boldsymbol{\mu}, \mathbf{B}, \mathbf{A}, d)$ .

The mean of matrix  $\mathbf{X}$  is  $E(\mathbf{X}) = \boldsymbol{\mu}$  and the variance is  $\text{Var}(\mathbf{X}) = \frac{1}{d-2} \mathbf{B}^{-1} \otimes \mathbf{A}$  [44, p. 447]. Operation  $\otimes$  denotes Kronecker product. Definition and properties can be found from [45, p. 104].

## 2.2 Bayesian inference

In Bayesian statistics, all parameters are assumed to be random variables. In Bayesian statistics unknown parameters are explored by forming statistical models from an observed sample and by using prior information about the unknown parameters.

### 2.2.1 Bayes' rule

Bayesian statistics is based on Bayes' rule and its applications.

**Definition 2.10** (Bayes' rule). Let  $\mathbf{y}$  be an observation whose probability distribution  $p(\mathbf{y}|\boldsymbol{\theta})$  depends on the values of parameter vector  $\boldsymbol{\theta}$ . Given the observed data  $\mathbf{y}$ , the conditional distribution of parameter  $\boldsymbol{\theta}$  is

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})}. \quad (2.32)$$

In Bayes' rule,  $p(\boldsymbol{\theta})$  is a prior distribution of an unknown parameter  $\boldsymbol{\theta}$ , that is knowledge about  $\boldsymbol{\theta}$  without knowledge of data. Given the data  $\mathbf{y}$ ,  $p(\mathbf{y}|\boldsymbol{\theta})$  may be considered as a function not of  $\mathbf{y}$  but of  $\boldsymbol{\theta}$ . When the observation  $\mathbf{y}$  is known and fixed, the function  $p(\mathbf{y}|\boldsymbol{\theta})$  is called a likelihood function of  $\boldsymbol{\theta}$  for given  $\mathbf{y}$  and can be written  $l(\boldsymbol{\theta}|\mathbf{y})$ .

Using

$$p(\mathbf{y}) = \int p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}, \quad (2.33)$$

it can be noticed that integral in Equation (2.33) depends only on  $\mathbf{y}$ , which is assumed to be fixed and known, and behaves merely as a normalizing constant that ensures that the probability distribution  $p(\boldsymbol{\theta}|\mathbf{y})$  integrates to 1. Thus the Bayes' rule can be written as

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{y}). \quad (2.34)$$

The likelihood function updates the prior knowledge about unknown parameters. It can be regarded as representing the information about  $\boldsymbol{\theta}$  coming from the data.

Probability density function  $p(\boldsymbol{\theta}|\mathbf{y})$  is called a posterior distribution of  $\boldsymbol{\theta}$  given  $\mathbf{y}$ . Posterior distribution is an updated information about  $\boldsymbol{\theta}$  and tells us what is known about  $\boldsymbol{\theta}$  given  $\mathbf{y}$ . Posterior contains all available information about  $\boldsymbol{\theta}$  and should be used for making decisions and estimates. In future it is referred to the prior distribution and posterior distribution simply as the 'prior' and the 'posterior', respectively and denote the posterior distribution as  $\pi(\boldsymbol{\theta}|\mathbf{y})$ .

Thus the prior tells what is known about unknown parameters without knowledge of data. If any information about unknown parameters is available, then this information should be use to formulate informative prior on the parameters. One convenient way to do this is to choose the prior to be in the same distribution family with the likelihood. This kind of prior is called a natural conjugate prior. The author in [38, p. 80] states two advantages of using the natural conjugate prior: the posterior is in the same form with prior and the posterior is generally tractable mathematically.

When no prior information about parameters is available one can choose a prior that affects the posterior as little as possible. These kind of priors are called noninformative or vague priors [38, p. 47]. If the state of knowledge about parameters is poor, it can be assumed that prior is uniformly distributed and is thus proportional to some constant  $c$ . One way

to define the noninformative prior is to use the prior introduced by Jeffreys [38, p. 77]:

$$p(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2},$$

where  $\mathbf{I}(\boldsymbol{\theta})$  is the Fisher information of  $\boldsymbol{\theta}$ ,

$$\mathbf{I}(\boldsymbol{\theta}) = \mathbb{E} \left[ \left( \frac{\partial \log l(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta}} \right)^2 \middle| \boldsymbol{\theta} \right] = - \mathbb{E} \left[ \frac{\partial^2 \log l(\boldsymbol{\theta}|\mathbf{y})}{\partial \boldsymbol{\theta}^2} \middle| \boldsymbol{\theta} \right]. \quad (2.35)$$

Key feature of the Jeffreys prior is that it is invariant under changes in parameterization, see [46, p. 12]. Prior distribution is called proper if it integrates to one, otherwise it is improper. Problem using the Jeffreys prior is that it is not always proper.

Posterior distribution can be updated recursively one observation at time. If the posterior distribution of  $\boldsymbol{\theta}$  given the observation  $\mathbf{y}$  is  $p(\boldsymbol{\theta}|\mathbf{y})$  and a new observation  $\mathbf{z}$  arrives, the posterior of  $\boldsymbol{\theta}$  given  $\mathbf{y}, \mathbf{z}$  can be obtained using Bayes' rule:

$$\pi(\boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) \propto \pi(\boldsymbol{\theta}|\mathbf{y})l(\mathbf{z}|\boldsymbol{\theta}). \quad (2.36)$$

### 2.2.2 Bayes' estimators

The result of the Bayesian analysis is the posterior distribution which combines information about the prior and observations. Although the posterior distribution provides the means of making all relevant inference about parameters and this can be summarized for example by plotting density function as shown in Figure 2, sometimes information on the posterior distribution is convenient to summarize as point estimates like mean, mode or median.

Bayes' estimator depends on the used loss function. If  $\hat{\boldsymbol{\theta}}$  is some estimator for parameter  $\boldsymbol{\theta}$  and loss which is caused by using this estimator is  $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ . The loss function measures the loss from choosing  $\hat{\boldsymbol{\theta}}$  when  $\boldsymbol{\theta}$  holds and thus

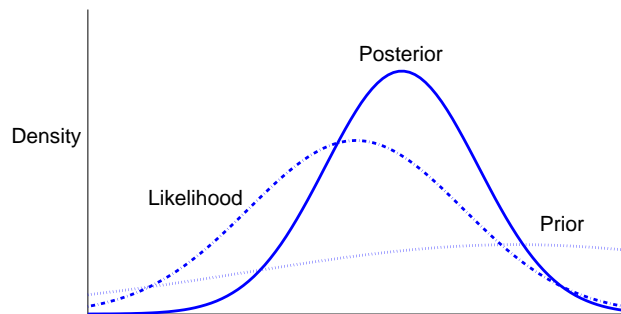


Figure 2: Posterior distribution of parameter  $\boldsymbol{\theta}$

measures how good the estimator  $\hat{\boldsymbol{\theta}}$  is. The Bayes' estimator is defined to be the estimator which minimizes expected value of the loss function [47, p. 116]

$$\mathbb{E}(L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})|\mathbf{y}) = \int L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}. \quad (2.37)$$

Three Bayes estimators are presented which are obtained by using different loss functions.

*Quadratic loss function*,  $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T C (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ , where  $C$  is a spd matrix. Assuming that the expected value of the posterior distribution is  $\boldsymbol{\mu}$ , the expected value of loss function becomes

$$\begin{aligned} \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T C (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})|\mathbf{y}] &= \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}) - (\boldsymbol{\theta} - \boldsymbol{\mu})]^T C [(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}) - (\boldsymbol{\theta} - \boldsymbol{\mu})|\mathbf{y}] \\ &= \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})^T C (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})|\mathbf{y}] \\ &\quad - \mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})^T C (\boldsymbol{\theta} - \boldsymbol{\mu})|\mathbf{y}] \\ &\quad - \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu})^T C (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})|\mathbf{y}] \\ &\quad + \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu})^T C (\boldsymbol{\theta} - \boldsymbol{\mu})|\mathbf{y}] \\ &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})^T C (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}) - 2(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})^T C \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu})|\mathbf{y}] \\ &\quad + \mathbb{E}[\text{tr}(C(\boldsymbol{\theta} - \boldsymbol{\mu})(\boldsymbol{\theta} - \boldsymbol{\mu})^T)] \\ &= (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})^T C (\hat{\boldsymbol{\theta}} - \boldsymbol{\mu}) - 2(\hat{\boldsymbol{\theta}} - \boldsymbol{\mu})^T C \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu})|\mathbf{y}] \\ &\quad + \text{tr}(C \mathbb{E}[(\boldsymbol{\theta} - \boldsymbol{\mu})(\boldsymbol{\theta} - \boldsymbol{\mu})^T|\mathbf{y}]). \end{aligned} \quad (2.38)$$

Second term in Equation (2.38) is equal to zero and third term does not depend on  $\hat{\boldsymbol{\theta}}$ . So expected loss is minimized when  $\hat{\boldsymbol{\theta}} = \boldsymbol{\mu}$  and the Bayes' estimator using quadratic loss function is posterior mean of  $\boldsymbol{\theta}$ .

*Zero-one loss function*,

$$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \begin{cases} 0 & \text{if } \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in K \\ 1 & \text{if } \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \notin K \end{cases}, \quad (2.39)$$

where  $K$  is some compact set containing the origin. The expected value of the loss function becomes

$$\mathbb{E}(L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})|\mathbf{y}) = \mathbb{p}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \notin K|\mathbf{y}) = 1 - \mathbb{p}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in K|\mathbf{y}), \quad (2.40)$$

which is minimized when  $\mathbb{p}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \in K|\mathbf{y})$  is maximized. The maximum is achieved when  $K$  is shrink to an arbitrarily small neighborhood of origin and the Bayes' estimate  $\hat{\boldsymbol{\theta}}$  converges to mode of posterior. Mode of posterior distribution is usually called *maximum a posteriori* -estimate, that is MAP-estimate.

*Absolute difference loss function*, for scalar random variable,  $L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$ . So  $\hat{\theta}$  is chosen to minimize

$$\int |\hat{\theta} - \theta| \pi(\theta|\mathbf{y}) d\theta = \int_{-\infty}^{\hat{\theta}} (\hat{\theta} - \theta) \pi(\theta|\mathbf{y}) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta}) \pi(\theta|\mathbf{y}) d\theta. \quad (2.41)$$

Minimum can be obtained by differentiating Equation (2.41) with respect to  $\hat{\theta}$  and the result equated to zero. A result is that  $\hat{\theta}$  must satisfy

$$\int_{-\infty}^{\hat{\theta}} \pi(\theta|\mathbf{y}) d\theta = \int_{\hat{\theta}}^{\infty} \pi(\theta|\mathbf{y}) d\theta. \quad (2.42)$$

Thus the Bayes' estimator of  $\theta$  using absolute difference loss function is median of  $\pi(\theta|\mathbf{y})$ .

For vector valued random variable there is no natural definition of the median. Authors in [48, p. 51] proposed an alternative loss function:

$$L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_i |\hat{\theta}_i - \theta_i|. \quad (2.43)$$

Then  $E(L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})|\mathbf{y})$  is also a sum, whose components are minimized independently by letting each  $\hat{\theta}_i$  be the median of the marginal distribution of  $\theta_i$ .

## 2.3 Bayesian filtering

Time series are used to measure how a process might evolve under time. Mathematical model of time series is a stochastic process.

**Definition 2.11.** Let  $(\Omega, \mathcal{F}, P)$  be a probability space and let  $T$  be a parameter set. Stochastic process is a mapping  $\mathbf{x} : \Omega \times T \rightarrow \mathbb{R}$ , such that for every fixed  $t \in T$ ,  $\mathbf{x}(t)$  is a random variable and is denoted as  $\mathbf{x}_t$ .

Stochastic process is a dynamic process and it amounts to a sequence of random variables. One type of stochastic processes is the Markov process which is defined next.

**Definition 2.12.** Markov process is a sequence of variables  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots$  with the Markov property

$$p(\mathbf{x}_k | \mathbf{x}_1, \dots, \mathbf{x}_{k-1}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}). \quad (2.44)$$

In filtering approach the states  $(\mathbf{x}_t, t \in \mathbb{N})$  and measurements  $(\mathbf{y}_t, t \in \mathbb{N})$  of a dynamic process are modeled to be a stochastic process. It is assumed to have an initial state with known distribution, and a state model and a measurement model is defined to estimate the posterior distribution for state parameters given current observations.

Initial state of the process is denoted as  $p(\mathbf{x}_0)$ . Dynamic of the process is modeled with a state model:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}). \quad (2.45)$$

The state model is an transition equation which connects the old states to new ones and is used as a prior distribution to parameters. The measurement model models the connection between state and measurements:

$$p(\mathbf{y}_t|\mathbf{x}_t). \quad (2.46)$$

Observations  $(\mathbf{y}_t, t \in \mathbb{N})$  are assumed to be conditionally independent given the process  $(\mathbf{x}_t, t \in \mathbb{N})$ .

The posterior distribution  $f(\mathbf{x}_t|\mathbf{y}_{1:t})$  of state  $\mathbf{x}_t$  can be computed recursively in time using Bayesian filtering. The Bayesian filter includes two steps, a prediction step and an update step. Assuming that conditional pdf  $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$  is known, the prediction step gives the prior distribution for time step  $t$  given observations  $\mathbf{y}_{1:t-1}$  [49, p. 5]

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}. \quad (2.47)$$

Updating step gives the posterior distribution of state  $\mathbf{x}_t$  given all observations  $\mathbf{y}_{1:t}$  and is done using Bayes' rule:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{x}_t|\mathbf{y}_{1:t-1})p(\mathbf{y}_t|\mathbf{x}_t). \quad (2.48)$$



### 3 Coverage area models

Multivariate linear models are used to measure how different variables depend on each other. In this work they are used to estimate parameters of the coverage area of a CN. Coverage area of a CN is modeled as an ellipse and the parameters of interest are  $\mu$ , a place parameter of coverage area and  $\Sigma$ , a shape parameter of coverage area. The multivariate linear model is widely studied and can be found in literature using both Bayesian and frequentist approach [44, Section 8],[43, Section 8],[39, Section 6] and [38, Sections 8.4–8.6]. In this work multivariate linear model is presented using Bayesian approach and two different priors. Box and Tiao [44] presented a multivariate normal linear model assuming the parameters to be mutually independent and used noninformative Jeffreys prior to form a posterior distribution. Broemeling [43] used a likelihood function of parameters to form an informative conjugate prior to the unknown parameters. In both cases the derivation of the posterior is presented in a general case although in this application only a vector-valued case is used. These are presented in Sections 3.1.2 and 3.1.3, respectively.

#### 3.1 Multivariate normal linear model

In this section a multivariate normal model is analyzed in a Bayesian viewpoint, using both noninformative and informative priors on the parameters. Suppose there are  $p$  correlated dependent variables and  $q$  independent variables and the relationship between the two sets is studied. The linear model can be written in the form

$$Y = X\mu + \epsilon \quad (3.1)$$

where  $Y = (\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(n)})^T$  is a  $n \times p$  matrix representing  $n$  mutually independent observations on the  $p$  dependent variables and  $\mathbf{y}_{(i)}$  is the column vector representing the transpose of  $i$ th row of matrix  $Y$ ,  $X$  is a known  $n \times q$  matrix of  $n$  observations on  $q$  independent variables,  $\mu$  is a  $q \times p$  matrix of unknown parameters and  $\epsilon$  is an  $n \times p$  matrix of random errors. The rows of  $\epsilon = (\boldsymbol{\epsilon}_{(1)}, \boldsymbol{\epsilon}_{(2)}, \dots, \boldsymbol{\epsilon}_{(n)})^T$  are assumed to be mutually independent normal random vectors with zero mean and unknown variance  $\Sigma$ . A joint posterior distribution and marginal posterior distributions for unknown parameters  $\mu$  and  $\Sigma$  are calculated.

##### 3.1.1 The likelihood function

It is assumed that the error vectors

$$\boldsymbol{\epsilon}_{(i)} = \mathbf{y}_{(i)} - \mu^T \mathbf{x}_{(i)} \quad i = 1, \dots, n, \quad (3.2)$$

are, for given  $\mu$  and  $\Sigma$ , distributed as  $N_p(\mathbf{0}, \Sigma)$  and mutually independent. Now a likelihood function for  $\epsilon$  can be formed, which is matrix normal distribution

$$l(\mu, \Sigma|Y) \propto |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (Y - X\mu)^T (Y - X\mu) \right]. \quad (3.3)$$

When  $\hat{\mu} = (X^T X)^{-1} X^T Y$  is the least square estimator of  $\mu$ , the quadratic form  $(Y - X\mu)^T (Y - X\mu)$  in the exponent of Equation (3.3) can be written in form

$$\begin{aligned} (Y - X\mu)^T (Y - X\mu) &= [(Y - X\hat{\mu}) + X(\hat{\mu} - \mu)]^T [(Y - X\hat{\mu}) + X(\hat{\mu} - \mu)] \\ &= (Y - X\hat{\mu})^T (Y - X\hat{\mu}) + (\mu - \hat{\mu})^T X^T X (\mu - \hat{\mu}). \end{aligned} \quad (3.4)$$

When  $S = (Y - X\hat{\mu})^T (Y - X\hat{\mu}) = Y^T Y - \hat{\mu}^T X^T X \hat{\mu}$  is the residual sum of squares, the likelihood (3.3) becomes

$$l(\mu, \Sigma|Y) \propto |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} ((\mu - \hat{\mu})^T X^T X (\mu - \hat{\mu}) + S) \right]. \quad (3.5)$$

### 3.1.2 Noninformative prior

For the prior distribution of the parameters  $(\mu, \Sigma)$ , it is assumed that  $\mu$  and  $\Sigma$  are mutually independent so that

$$p(\mu, \Sigma) = p(\mu)p(\Sigma) \quad (3.6)$$

and the Jeffreys invariant prior is developed by applying Equation (2.35) to one parameter at time. Logarithm of the likelihood density is

$$\log l(\mu, \Sigma|Y) = c - \frac{n}{2} \log |\Sigma| - \frac{1}{2} \text{tr} \Sigma^{-1} (Y - X\mu)^T (Y - X\mu). \quad (3.7)$$

First assume that  $\Sigma$  is constant. Then

$$\begin{aligned} \frac{\partial^2 \log l(\mu, \Sigma|Y)}{\partial \mu \partial \mu^T} &\stackrel{(B.9)}{=} \frac{\partial}{\partial \mu^T} \left( -\frac{1}{2} (2X^T Y \Sigma^{-1} + 2\Sigma^{-1} \mu^T X^T X) \right) \\ &\stackrel{(B.11)}{=} -X^T X \otimes \Sigma^{-1} \end{aligned} \quad (3.8)$$

and

$$\begin{aligned} -E \left( \frac{\partial^2 \log l(\mu, \Sigma|Y)}{\partial \mu \partial \mu^T} \right) &= -E(-X^T X \otimes \Sigma^{-1}) \\ &= X^T X \otimes \Sigma^{-1}. \end{aligned} \quad (3.9)$$

Hence, the information matrix is constant and the Jeffreys invariant prior for  $\mu$  is

$$p(\mu) \propto 1. \quad (3.10)$$

Next assume that  $\mu$  is constant. Denote  $\Sigma = (\sigma_{ij})$  and  $\Sigma^{-1} = (\sigma^{ij})$ ,  $i, j = 1, \dots, p$ . Differentiating Equation (3.7) with respect to  $\sigma^{ij}$ , where  $i \geq j$ , results in

$$\begin{aligned} \frac{\partial \log l(\mu, \Sigma | Y)}{\partial \sigma^{ij}} &= \frac{n}{2|\Sigma^{-1}|} \frac{\partial |\Sigma^{-1}|}{\partial \sigma^{ij}} - \frac{1}{2} (Y - X\mu)^T (Y - X\mu) \\ &= \frac{\text{cof}(\sigma^{ij})}{2|\Sigma^{-1}|} - \frac{1}{2} (Y - X\mu)^T (Y - X\mu), \end{aligned} \quad (3.11)$$

where  $\text{cof}(\sigma^{ij})$  is the cofactor of  $\sigma^{ij}$ . So according to Definition B.2, the first term in (3.11) is simply  $\frac{n}{2}\sigma_{ij}$ . Second derivatives are

$$\frac{\partial \log l(\mu, \Sigma | Y)}{\partial \sigma^{ij} \partial \sigma^{kl}} \propto \frac{\partial \sigma_{ij}}{\partial \sigma^{kl}}, \quad i, j, k, l = 1, \dots, p, \quad i \leq j, k \leq l, \quad (3.12)$$

and the determinant of the information matrix of  $\Sigma^{-1}$  is

$$|\mathbf{I}(\Sigma^{-1})| = \left| -\mathbf{E} \left[ \frac{\partial \log l(\mu, \Sigma | Y)}{\partial \sigma^{ij} \partial \sigma^{kl}} \right] \right| \propto \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|. \quad (3.13)$$

Using Jacobian of transformation from  $\Sigma^{-1}$  to  $\Sigma$ , the Jeffreys invariant prior for  $\Sigma$  becomes

$$p(\Sigma) \propto |\mathbf{I}(\Sigma)|^{1/2} = |\mathbf{I}(\Sigma^{-1})|^{1/2} J_{\Sigma^{-1} \rightarrow \Sigma} = \left| \frac{\partial \Sigma}{\partial \Sigma^{-1}} \right|^{1/2} = |\Sigma|^{-\frac{1}{2}(p+1)}. \quad (3.14)$$

Combining Equations (3.5), (3.6) and Bayes' theorem, the joint posterior of  $(\mu, \Sigma)$  is

$$\begin{aligned} \pi(\mu, \Sigma | Y) &\propto p(\mu, \Sigma) l(\mu, \Sigma | Y) \\ &= p(\mu) p(\Sigma) l(\mu, \Sigma | Y) \\ &= |\Sigma|^{-(n+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} ((\mu - \hat{\mu})^T X^T X (\mu - \hat{\mu}) + S) \right]. \end{aligned}$$

After rewriting elements in the exponent of Equation (3.15) in form

$$\Sigma^{-1} ((\mu - \hat{\mu})^T X^T X (\mu - \hat{\mu}) + S) = (\Sigma \otimes (X^T X)^{-1})^{-1} (\mu - \hat{\mu})^T (\mu - \hat{\mu}) + \Sigma^{-1} S, \quad (3.15)$$

it can be noticed that the conditional distribution of  $\mu$  given  $\Sigma$  is a normal distribution  $N_{qp}(\hat{\mu}, \Sigma \otimes (X^T X)^{-1})$ . The marginal posterior of  $\mu$  and  $\Sigma$  can be derived by integrating Equation (3.15) with respect to  $\Sigma$  and  $\mu$ , respectively. To obtain marginal posterior of  $\mu$  one must notice, that posterior distribution with respect to  $\Sigma$  is a Wishart distribution, so marginal posterior of  $\mu$  becomes

$$\begin{aligned} \pi(\mu | Y) &= \int \pi(\mu, \Sigma | Y) d\Sigma \\ &\propto |S + (\mu - \hat{\mu})^T X^T X (\mu - \hat{\mu})|^{-n/2} \\ &\propto \left[ \mathbf{I}_p + S^{-1} (\mu - \hat{\mu})^T X^T X (\mu - \hat{\mu}) \right]^{-n/2}. \end{aligned}$$

That is,  $\pi(\mu|Y)$  follows a matrix t-distribution  $t_{pq}(\hat{\mu}, X^T X^{-1}, S, n - p - q + 1)$ .

The posterior distribution in Equation (3.15) with respect to  $\mu$  is a normal distribution when  $\Sigma$  is known, so by integrating Equation (3.15) with respect to  $\mu$ , the marginal posterior of  $\Sigma$  becomes

$$\begin{aligned} \pi(\Sigma | Y) &= \int \pi(\mu, \Sigma | Y) d\mu \\ &\propto \int |\Sigma|^{-(n+p+1)/2} \\ &\quad \times \exp \left[ -\frac{1}{2} \text{tr}((\mu - \hat{\mu})^T \Sigma^{-1} \otimes X^T X (\mu - \hat{\mu}) + \Sigma^{-1} S) \right] d\mu \\ &\propto |\Sigma|^{-(n+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right] |\Sigma|^{q/2} \\ &\propto |\Sigma|^{-(n-q+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right]. \end{aligned}$$

That is,  $\pi(\Sigma|Y)$  follows an inverse Wishart distribution  $W_p^{-1}(S, n - q)$ .

### 3.1.3 Informative prior

Suppose that there is a prior information assessable about the parameters. Then it is suitable to use informative prior, which takes account the prior knowledge about the parameters.

The natural conjugate prior is found by interchanging the roles of observable and unobservable random variables and replace the observable parameters with some arbitrary parameters. The likelihood function in Equation (3.5) can be expressed in the form

$$\begin{aligned} l(\mu, \Sigma | Y) &\propto |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} S \right] \\ &\quad \times \exp \left[ \frac{1}{2} \text{tr} (\Sigma \otimes (X^T X)^{-1})^{-1} (\mu - \hat{\mu})^T (\mu - \hat{\mu}) \right], \end{aligned} \quad (3.16)$$

which is, according to Definition 2.7, a normal-inverse-Wishart distribution. The natural conjugate prior for parameters is also a normal-inverse-Wishart distribution, namely

$$p(\mu, \Sigma) = p(\mu|\Sigma)p(\Sigma), \quad \text{where} \quad (3.17)$$

$$\begin{aligned} p(\mu|\Sigma) &= \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (\mu - \mathbf{m})^T A (\mu - \mathbf{m}) \right] \quad \text{and} \\ p(\Sigma) &= |\Sigma|^{-(v+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} B \Sigma^{-1} \right], \end{aligned} \quad (3.18)$$

where  $\mathbf{m}$  is a  $q \times p$  known matrix,  $A$  is known symmetric positive definite  $q \times q$  matrix,  $v \geq p$  known constant and  $B$  is  $p \times p$  positive symmetric matrix. Thus  $\mu|\Sigma \sim N_{qp}(\mathbf{m}, \Sigma \otimes A^{-1})$  and  $\Sigma \sim W^{-1}(B, v)$ .

The joint posterior density for  $\mu$  and  $\Sigma$  is found by using Bayes' theorem and multiplying Equations (3.16) and (3.17)

$$\begin{aligned}
\pi(\mu, \Sigma | \mathbf{Y}) &\propto l(\mu, \Sigma | \mathbf{Y}) p(\mu, \Sigma) \\
&\propto |\Sigma|^{-(n+v+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} ((\mathbf{Y} - \mathbf{X}\mu)^T (\mathbf{Y} - \mathbf{X}\mu) \right. \\
&\quad \left. + (\mu - \mathbf{m})^T \mathbf{A} (\mu - \mathbf{m}) + \mathbf{B}) \right] \\
&= |\Sigma|^{-(n^*+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} \mathbf{Q} \right],
\end{aligned} \tag{3.19}$$

where  $n^* = n + v$  and

$$\begin{aligned}
\mathbf{Q} &= (\mathbf{Y} - \mathbf{X}\mu)^T (\mathbf{Y} - \mathbf{X}\mu) + (\mu - \mathbf{m})^T \mathbf{A} (\mu - \mathbf{m}) + \mathbf{B} \\
&= \mu^T (\mathbf{X}^T \mathbf{X} + \mathbf{A}) \mu - \mu^T (\mathbf{X}^T \mathbf{Y} + \mathbf{A} \mathbf{m}) \\
&\quad - (\mathbf{X}^T \mathbf{Y} + \mathbf{A} \mathbf{m})^T \mu + \mathbf{Y}^T \mathbf{Y} + \mathbf{m}^T \mathbf{A} \mathbf{m} + \mathbf{B} \\
&= (\mu - \hat{\mathbf{m}})^T \hat{\mathbf{A}} (\mu - \hat{\mathbf{m}}) + \hat{\mathbf{S}},
\end{aligned}$$

in which

$$\hat{\mathbf{m}} = (\mathbf{X}^T \mathbf{X} + \mathbf{A})^{-1} (\mathbf{X}^T \mathbf{Y} + \mathbf{A} \mathbf{m}) \tag{3.20}$$

$$\hat{\mathbf{A}} = \mathbf{X}^T \mathbf{X} + \mathbf{A} \tag{3.21}$$

$$\hat{\mathbf{S}} = \mathbf{Y}^T \mathbf{Y} + \mathbf{B} + \mathbf{m}^T \mathbf{A} \mathbf{m} - \hat{\mathbf{m}}^T \hat{\mathbf{A}} \hat{\mathbf{m}}. \tag{3.22}$$

Equation (3.19) shows that also the joint posterior density follows normal-inverse-Wishart distribution.

The marginal posterior of  $\mu$  can be derived by integrating joint posterior (3.19) with respect to  $\Sigma$ .

$$\begin{aligned}
\pi(\mu | \mathbf{Y}) &= \int \pi(\mu, \Sigma | \mathbf{Y}) d\Sigma \\
&\propto |\hat{\mathbf{S}} + (\mu - \hat{\mathbf{m}})^T \hat{\mathbf{A}} (\mu - \hat{\mathbf{m}})|^{-(n^*)/2} \\
&\propto \left[ \mathbf{I}_p + \hat{\mathbf{S}}^{-1} (\mu - \hat{\mathbf{m}})^T \hat{\mathbf{A}} (\mu - \hat{\mathbf{m}}) \right]^{-(n^*)/2},
\end{aligned} \tag{3.23}$$

which has a matrix t distribution  $t_{pq}(\hat{\mathbf{m}}, \hat{\mathbf{S}}, \hat{\mathbf{A}}^{-1}, n^* - p - q + 1)$ . Marginal posterior of  $\Sigma$  becomes

$$\begin{aligned}
\pi(\Sigma | \mathbf{Y}) &= \int \pi(\mu, \Sigma | \mathbf{Y}) d\mu \\
&\propto |\Sigma|^{-(n^*+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} \hat{\mathbf{S}} \right] |\Sigma|^{q/2} \\
&\propto |\Sigma|^{-(n^*-q+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} \hat{\mathbf{S}} \right],
\end{aligned} \tag{3.24}$$

which is again an inverse-Wishart distribution,  $W^{-1}(\hat{\mathbf{S}}, n^* - q)$ .

## 3.2 Mixture model

Most statistical procedures are arrived with the assumption that each one of the given set of observations is generated by a specific stochastic model. This assumption does not correspond to reality, because usually there are observations that do not follow the pattern of the majority of the data and show some extremeness relative to basic model. Box and Tiao [50] called these observations outliers. One way to deal with outliers is to reject the outlier being observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model assumed. A more convenient approach is to assume that there was a small prior probability  $\alpha$  that any given observation was not generated by the central stochastic model as well as a complementary prior probability  $(1 - \alpha)$  that it was so generated.

A special linear model is considered, where most of the observations are normally distributed about its mean with variance  $\Sigma$  and an outlier is normal with the same mean but a larger variance  $k\Sigma$ . Also a Bayesian method for detecting outliers is given. This method uses the random errors of the model and uses the posterior distribution of the errors to detect the outliers

### 3.2.1 A special linear model

When assuming that some of the observations may be outliers, a different kind of linear model is needed. Box and Tiao [50] introduced a special linear model where they explore the consequences of such a supposition for the normal theory linear model. They introduced this special linear model as univariate case and it is generalized here to a multivariate case. Natural conjugate prior is used to derive the posterior of parameters, so this model can be considered as a special case of the model introduced in Section 3.1.3.

Suppose that there are observations  $Y = (\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(n)})^T$ . Supposing that the random error  $\epsilon$  associated with each observation  $\mathbf{y}_{(i)}$  could have been drawn from two sources, from a central model  $N_p(\mathbf{0}, \Sigma)$  or from an alternative model  $N_p(\mathbf{0}, k\Sigma)$ , with probabilities  $(1 - \alpha)$  and  $\alpha$  respectively. Supposing that the observation matrix  $Y$  is partitioned into two matrixes  $Y_{(s)}$  and  $Y_{(r)}$ , where  $s + r = n$ , two linear models can be formed

$$Y_{(s)} = X_s \mu + \epsilon_{(s)}, \text{ where } \epsilon_{(s)} \sim N_p(\mathbf{0}, \Sigma) \quad (3.25)$$

$$Y_{(r)} = X_r \mu + \epsilon_{(r)}, \text{ where } \epsilon_{(r)} \sim N_p(\mathbf{0}, k\Sigma), \quad (3.26)$$

where  $Y_{(s)}$  is a  $s \times p$  and  $Y_{(r)}$  is a  $r \times p$  observation matrix.

Supposing that  $k$  is fixed and observations are mutually independent, the likelihood of  $(\mu, \Sigma)$  with a particular partition  $(s)$  of the observation matrix can be expressed as a product of two linear models

$$l(\mu, \Sigma | Y_{(s)}, Y_{(r)}, (s)) = l(\mu, \Sigma | Y_{(s)}) l(\mu, k\Sigma | Y_{(r)}). \quad (3.27)$$

Considering all the possible partitions, the entire likelihood can be expressed as a weighted average of all the possible likelihoods

$$l(\mu, \Sigma | Y) = \frac{\sum_{(s)} w_{(s)} l(\mu, \Sigma | Y_{(s)}, Y_{(r)}, (r))}{\sum_{(s)} w_{(s)}}. \quad (3.28)$$

where  $w_{(s)}$  is weight corresponding to combination  $(s)$ . Thus the entire likelihood function consists of  $2^n$  expressions of the type given in Equation (3.27) corresponding to the  $2^n$  possible combinations of the errors  $\epsilon$ .

Consider a particular combination  $(s)$ . Under the assumption of two normal distributions given in Equations (3.25) and (3.26) the likelihood function of a certain partition becomes

$$\begin{aligned} l(\mu, \Sigma | Y_{(s)}, Y_{(r)}) &\propto |\Sigma|^{-s/2} \exp \left[ -\frac{1}{2} \Sigma^{-1} (Y_{(s)} - X_{(s)} \mu)^T (Y_{(s)} - X_{(s)} \mu) \right] \\ &\quad \times |\Sigma|^{-r/2} \exp \left[ -\frac{1}{2k} \Sigma^{-1} (Y_{(r)} - X_{(r)} \mu)^T (Y_{(r)} - X_{(r)} \mu) \right] \\ &= |\Sigma|^{-n/2} \exp \left[ -\frac{1}{2} \Sigma^{-1} (Y_{(s)} - X_{(s)} \mu)^T (Y_{(s)} - X_{(s)} \mu) \right. \\ &\quad \left. + \frac{1}{k} (Y_{(r)} - X_{(r)} \mu)^T (Y_{(r)} - X_{(r)} \mu) \right]. \end{aligned} \quad (3.29)$$

Exponent of Equation (3.29) can be written in form

$$\begin{aligned} &(Y_{(s)} - X_{(s)} \mu)^T (Y_{(s)} - X_{(s)} \mu) + \frac{1}{k} (Y_{(r)} - X_{(r)} \mu)^T (Y_{(r)} - X_{(r)} \mu) \\ &= \mu^T (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)}) \mu - \mu^T (X_{(s)}^T Y_{(s)} + \frac{1}{k} X_{(r)}^T Y_{(r)}) \\ &\quad - (X_{(s)}^T Y_{(s)} + \frac{1}{k} X_{(r)}^T Y_{(r)})^T \mu + Y_{(s)}^T Y_{(s)} + \frac{1}{k} Y_{(r)}^T Y_{(r)} \\ &= (\mu - \hat{\mu})^T (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)}) (\mu - \hat{\mu}) + S, \end{aligned} \quad (3.30)$$

where

$$\hat{\mu} = (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)})^{-1} (X_{(s)}^T Y_{(s)} + \frac{1}{k} X_{(r)}^T Y_{(r)}) \quad \text{and} \quad (3.31)$$

$$S = Y_{(s)}^T Y_{(s)} + \frac{1}{k} Y_{(r)}^T Y_{(r)} - \hat{\mu}^T (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)}) \hat{\mu}. \quad (3.32)$$

So likelihood in Equation (3.29) becomes

$$\begin{aligned} l(\mu, \Sigma | Y_{(s)}, Y_{(r)}) &\propto |\Sigma|^{-n/2} \\ &\quad \times \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (\mu - \hat{\mu})^T (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)}) (\mu - \hat{\mu}) + S \right]. \end{aligned} \quad (3.33)$$

It can be seen that the likelihood in Equation (3.33) of parameters  $(\mu, \Sigma)$  follows normal-inverse-Wishart distribution, which is a convenient choice for natural conjugate prior for  $(\mu, \Sigma)$ . So prior becomes

$$p(\mu, \Sigma) \propto |\Sigma|^{-(v+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (\mu - \mathbf{m})^T \mathbf{A} (\mu - \mathbf{m}) + \mathbf{B} \right], \quad (3.34)$$

where  $\mathbf{m}$  is a known  $q \times p$  matrix,  $\mathbf{A}$  is a known  $q \times q$  symmetric positive definite matrix,  $\mathbf{B}$  is a known  $p \times p$  symmetric positive definite matrix and  $v \geq p$  is a known constant. By multiplying (3.29) and (3.34), the posterior for  $(\mu, \Sigma)$  becomes

$$\begin{aligned} \pi(\mu, \Sigma | Y_{(s)}, Y_{(r)}) &\propto l(\mu, \Sigma | Y_1, Y_2) p(\mu, \Sigma) \\ &\propto |\Sigma|^{-(n+v+p+1)/2} \\ &\times \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (Y_{(s)} - X_{(s)} \mu)^T (Y_{(s)} - X_{(s)} \mu) \right. \\ &\quad + \frac{1}{k} (Y_{(r)} - X_{(r)} \mu)^T (Y_{(r)} - X_{(r)} \mu) \\ &\quad \left. + (\mu - \mathbf{m})^T \mathbf{A} (\mu - \mathbf{m}) + \mathbf{B} \right] \\ &\propto |\Sigma|^{-(n+v+p+1)/2} \exp \left[ -\frac{1}{2} \text{tr} \Sigma^{-1} (\mu - \hat{\mathbf{m}})^T \hat{\mathbf{A}} (\mu - \hat{\mathbf{m}}) + \hat{\mathbf{S}} \right], \end{aligned} \quad (3.35)$$

in which

$$\hat{\mathbf{m}} = (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)} + \mathbf{A})^{-1} (X_{(s)}^T Y_{(s)} + \frac{1}{k} X_{(r)}^T Y_{(r)} + \mathbf{A} \mathbf{m}) \quad (3.36)$$

$$\hat{\mathbf{A}} = (X_{(s)}^T X_{(s)} + \frac{1}{k} X_{(r)}^T X_{(r)} + \mathbf{A}) \quad (3.37)$$

$$\hat{\mathbf{S}} = Y_{(s)}^T Y_{(s)} + \frac{1}{k} Y_{(r)}^T Y_{(r)} + \mathbf{B} + \mathbf{m}^T \mathbf{A} \mathbf{m} - \hat{\mathbf{m}}^T \hat{\mathbf{A}} \hat{\mathbf{m}}. \quad (3.38)$$

Also the joint posterior of parameters  $(\mu, \Sigma)$  follows normal-inverse-Wishart distribution. As in Section 3.1.3, a marginal distribution of parameters can be solved, and marginal posteriors becomes  $\mu | \Sigma, Y \sim N_{qp}(\hat{\mathbf{m}}, \Sigma \otimes \mathbf{A}^{-1})$ ,  $\mu | Y \sim t_{pq}(\hat{\mathbf{m}}, \hat{\mathbf{S}}, \hat{\mathbf{A}}^{-1}, n + v - p - q + 1)$  and  $\Sigma | Y \sim W_p^{-1}(\hat{\mathbf{S}}, n + v - q)$ .

If it is known which observations are outliers, it can be chosen the particular combination  $(s)$  and put  $w_{(s)} = 1$  and choose all the other weights to be zero.

### 3.2.2 Outlier detection

The previous section showed that after observations are sorted, it is straightforward to form the posterior distributions of unknown parameters.



The problem is how to detect the outliers from the set of observations. It is assumed that observations may come from two different distributions, the underlying model  $N_p(\mu, \Sigma)$  and the alternative model  $N_p(\mu, k\Sigma)$ . An observation will be accepted as an outlier if the posterior probability that the realized error of an observation is greater than  $k$  is higher than the prior probability of the realized error of an observation is greater than  $k$ . So if  $p(\boldsymbol{\epsilon}_{(i)} > k | Y) > p(\boldsymbol{\epsilon}_{(i)} > k)$ , then the  $i$ th observation will be accepted as an outlier. This Bayesian approach to outlier detection was first proposed by Chaloner and Brant [51]. They presented a method in order to detect an outlier in univariate linear model. Later Varbanov [52] generalized this approach to a multivariate linear model.

Suppose that there is an observation set  $Y = (\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(n)})^T$ , where  $\mathbf{y}_{(i)}$ s are  $p$ -variate mutually independent observations. Linear model can be written in form

$$Y = X\mu + E, \quad (3.39)$$

where  $X$  is  $n \times q$  known matrix and the  $n \times p$  matrix  $E$  contains the random errors for the observation set  $Y$ ,

$$E = \begin{pmatrix} \boldsymbol{\epsilon}_{(1)}^T \Sigma \\ \boldsymbol{\epsilon}_{(2)}^T \Sigma \\ \vdots \\ \boldsymbol{\epsilon}_{(n)}^T \Sigma \end{pmatrix}. \quad (3.40)$$

It is assumed, that random errors are mutually independent and are normally distributed,  $\boldsymbol{\epsilon}_{(i)} \sim N_p(\mathbf{0}, I_p)$ . The model (3.39) can be written also as

$$\mathbf{y}_{(i)} = \mu^T \mathbf{x}_{(i)} + \Sigma^{1/2} \boldsymbol{\epsilon}_{(i)}, \quad (3.41)$$

and the random error  $\boldsymbol{\epsilon}_{(i)}$  becomes

$$\boldsymbol{\epsilon}_{(i)} = \Sigma^{-1/2} (\mathbf{y}_{(i)} - \mu^T \mathbf{x}_{(i)}). \quad (3.42)$$

Define

$$\delta_i = \boldsymbol{\epsilon}_{(i)}^T \boldsymbol{\epsilon}_{(i)} = (\mathbf{y}_{(i)} - \mu^T \mathbf{x}_{(i)})^T \Sigma^{-1} (\mathbf{y}_{(i)} - \mu^T \mathbf{x}_{(i)}), \quad (3.43)$$

which is used to test whether the  $i$ th observation is an outlier or not. The  $i$ th observation  $\mathbf{y}_i$  is declared to be an outlier, if  $\delta_i > k$  for an appropriate choice of  $k$ . Notice that because random errors  $\boldsymbol{\epsilon}_{(i)}$  are normally distributed with zero mean and variance  $I_p$ , according to Definition 2.5  $\delta_i$  will have a central chi-square distribution of  $p$  degrees of freedom.

The value of  $k$  can be chosen so that prior probability of no outliers is large. Suppose that probability of finding an outlier is  $\alpha$ . So probability that there are no outliers is  $1 - \alpha$ , i.e

$$\begin{aligned} 1 - \alpha &= p(\delta_i < k, \text{ for all } i) \\ &= p(\delta_1 < k) p(\delta_2 < k) \dots p(\delta_n < k) = F_p(k)^n, \end{aligned} \quad (3.44)$$

where  $F_p(k)$  denotes cumulative distribution function for central chi-square distribution with  $p$  degrees of freedom. The solution for Equation (3.44) is

$$k = F_p^{-1}((1 - \alpha)^{1/n}). \quad (3.45)$$

If  $p(\delta_i > k|Y)$ , the posterior probability that the squared  $\delta_i$  of the realised but unobserved error exceeds the critical value  $k$ , is larger than  $p(\delta_i > k)$ , the probability that the  $\delta_i$  exceeds  $k$  under the hypothesis of the model, then the  $i$ th observation may be an outlier.

In order to obtain the posterior distribution of  $\delta_i$ , posterior distribution of random error  $\epsilon_{(i)}$  must be derived. Since  $\epsilon_{(i)} = \Sigma^{-1/2}(\mathbf{y}_{(i)} - \mu^T \mathbf{x}_{(i)})$  is a linear function of parameters  $\mu$  and  $\Sigma$ , the joint posterior of  $\mu$  and  $\Sigma$  must be obtained. Section 3.1.3 shows that using informative priors the posterior of  $\mu$  given  $\Sigma$  and the marginal posterior of  $\Sigma$  are obtained respectively as follows

$$\begin{aligned} \mu|\Sigma, Y &\sim N_p(\hat{\mathbf{m}}, \Sigma \otimes \hat{\mathbf{A}}^{-1}) \\ \Sigma|Y &\sim W^{-1}(\hat{\mathbf{S}}, n + v - q). \end{aligned}$$

The posterior of  $\epsilon_{(i)}|\Sigma, Y$  can be obtained by using following information. According to Theorem 2.1 the posterior distribution of  $\epsilon_{(i)}$  is normally distributed  $N_p(\hat{\epsilon}_{(i)}, \mathbf{x}_{(i)}^T \hat{\mathbf{A}} \mathbf{x}_{(i)})$ , where  $\hat{\epsilon}_{(i)} = \Sigma^{-1/2}(\mathbf{y}_{(i)} - \hat{\mathbf{m}}^T \mathbf{x}_{(i)})$ . Define

$$\sigma_i = \mathbf{x}_{(i)}^T \hat{\mathbf{A}} \mathbf{x}_{(i)} \quad (3.46)$$

$$\lambda_i = \sigma_i^{-1}(\mathbf{y}_{(i)} - \hat{\mathbf{m}}^T \mathbf{x}_{(i)})\Sigma^{-1}(\mathbf{y}_{(i)} - \hat{\mathbf{m}}^T \mathbf{x}_{(i)}) \quad (3.47)$$

$$T_i = \frac{\delta_i}{\sigma_i}. \quad (3.48)$$

Distribution of  $T_i$  is non-central chi-square with  $p$  degrees of freedom and non-centrality parameter  $\lambda_i$  and  $p(\delta_i > k|Y) = p(T_i > \sigma_i k|Y)$ . Choosing  $\hat{\Sigma} = E(\Sigma)$  the posterior probability  $p(\delta_i > k|Y)$  can be calculated.

In this work, the method for detecting outliers presented in this section is used to find the observations which do not follow the same model as the majority of observations. This is used as a prior information on which partition ( $s$ ) is best and the weight for that partition is set equal to one. Thus the likelihood of the parameters  $\mu$  and  $\Sigma$  presented in (3.28) reduces to the likelihood presented in likelihood (3.27). More formal analysis to detect posterior estimates for parameters  $\mu$  and  $\Sigma$  could be done using MCMC (Markov Chain Monte Carlo) methods, like Gibbs sampler [41]. However, one aim of this thesis is to keep a computational burden as low as possible and for that reason the approach used in this work is chosen.

### 3.3 Updating posterior

In the Bayesian statistics the posterior distribution of unknown parameters may be updated recursively when new observations come in, as shown in

Equation (2.36). The updating formula for posterior distribution is derived in the common case in Section 3.3.1. However, Equation (2.36) assumes that all observations are generated from the same distribution and does not take into account that the parameters are changing in time. For that reason, an additional time variant formula is derived in Section 3.3.2, where the posterior distribution of the parameters is computed as time series. This takes also into account all the previous measurements in addition to new ones but it takes account of possibility that the parameters are dynamic.

### 3.3.1 Time invariant updating

It is assumed that a new observation  $\check{Y} = (\check{\mathbf{y}}_{(1)}, \check{\mathbf{y}}_{(2)}, \dots, \check{\mathbf{y}}_{(l)})^T$  arrives and the posterior distribution  $\pi(\mu, \Sigma | Y)$  derived with a knowledge of previous observations  $Y = (\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(n)})^T$  is to be updated. As in previous sections, the measurement model of new observation can be written in form

$$\check{Y} = Z\mu + \epsilon, \text{ where } \epsilon_{(i)} \sim N_p(\mathbf{0}, \Sigma) \quad (3.49)$$

and a likelihood function of  $(\mu, \Sigma)$  given observation  $\check{Y}$  can be written as

$$l(\mu, \Sigma | \check{Y}) \propto |\Sigma|^{-l/2} \exp \left[ -\frac{1}{2} \Sigma^{-1} (\check{Y} - Z\mu)^T (\check{Y} - Z\mu) \right]. \quad (3.50)$$

The posterior distribution of  $(\mu, \Sigma)$  given observations  $Y, \check{Y}$  can be derived using Equation (2.36):

$$\begin{aligned} \pi(\mu, \Sigma | Y, \check{Y}) &\propto \pi(\mu, \Sigma | Y) l(\mu, \Sigma | \check{Y}) \\ &\propto |\Sigma|^{-(n^*+l+p+1)/2} \exp \left[ -\frac{1}{2} \Sigma^{-1} ((\mu - \hat{\mathbf{m}})^T \hat{\mathbf{A}} (\mu - \hat{\mathbf{m}}) + \hat{\mathbf{S}} \right. \\ &\quad \left. + (\check{Y} - Z\mu)^T (\check{Y} - Z\mu)) \right] \\ &= |\Sigma|^{-(n^*+l+p+1)/2} \exp \left[ -\frac{1}{2} \Sigma^{-1} (\mu - \check{\mathbf{m}})^T \check{\mathbf{A}} (\mu - \check{\mathbf{m}}) + \check{\mathbf{S}} \right], \end{aligned} \quad (3.51)$$

where

$$\check{\mathbf{m}} = (Z^T Z + \hat{\mathbf{A}})^{-1} (Z^T \check{Y} + \hat{\mathbf{A}} \hat{\mathbf{m}}) \quad (3.52)$$

$$\check{\mathbf{A}} = Z^T Z + \hat{\mathbf{A}} \quad (3.53)$$

$$\check{\mathbf{S}} = \check{Y}^T \check{Y} + \hat{\mathbf{S}} + \hat{\mathbf{m}}^T \hat{\mathbf{A}} \hat{\mathbf{m}} - \check{\mathbf{m}}^T \check{\mathbf{A}} \check{\mathbf{m}}. \quad (3.54)$$

The updated posterior distribution of parameters  $(\mu, \Sigma)$  follows also normal-inverse-Wishart distribution.

Marginal posteriors of  $\mu$  and  $\Sigma$  can be derived as in Sections 3.1.2 and 3.1.3. The marginal posterior of  $\mu$  follows a matrix t distribution  $t_{pq}(\check{\mathbf{m}}, \check{\mathbf{S}}, \check{\mathbf{A}}^{-1}, n^* + l - p - q + 1)$  and marginal posterior of  $\Sigma$  follows an inverse-Wishart distribution  $W^{-1}(\check{\mathbf{S}}, n^* + l - q)$ .

### 3.3.2 Time variance updating

In filtering approach the development of the coverage area of a communication node is assumed to be a Markov chain. In this section it is assumed that  $\boldsymbol{\mu}$  is a  $p$ -variate vector.

As an initial state it is assumed that the parameters  $(\boldsymbol{\mu}, \Sigma)$  follow an inverse Wishart distribution, that is

$$p(\boldsymbol{\mu}_0, \Sigma_0) = p(\boldsymbol{\mu}_0 | \Sigma_0) p(\Sigma_0), \quad (3.55)$$

where

$$\boldsymbol{\mu}_0 | \Sigma_0 \sim N(\hat{\boldsymbol{\mu}}_0, \frac{1}{a_0} \Sigma_0) \text{ and } \Sigma_0 \sim W^{-1}(S_0, n_0). \quad (3.56)$$

The state model for the parameters  $\boldsymbol{\mu}_t, \Sigma_t$  at time step  $t$  is defined as

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \mathbf{v}_t, \quad \text{where } \mathbf{v}_t \sim N(0, c \Sigma_{t-1}), \quad c \geq 0, \quad (3.57)$$

$$\Sigma_t = k \Sigma_{t-1}, \quad \text{where } k \geq 1. \quad (3.58)$$

Used state model gives less weight to the old place parameter and it assumes that the prior for step  $t$  is obtained by increasing the uncertainty of previous time step.

It is assumed that for every time step  $t$  also a set of observations may come in instead of only a single observation at a time. However, the observations are assumed to be mutually independent and the measurement model is defined as previous

$$Y_t = X_t \boldsymbol{\mu}_t + \boldsymbol{\epsilon}_t, \quad \text{where } \boldsymbol{\epsilon}_{(i)} \sim N(0, \Sigma_t) \quad (3.59)$$

and  $Y_t$  is a  $n_t \times p$  observation matrix. Because it is restricted to the case where  $\boldsymbol{\mu}$  is a  $p$ -variate vector, it follows that a design matrix  $X_t$  is an  $n_t$ -variate vector.

A prediction step is done first to  $\boldsymbol{\mu}$  using Equation (2.47) and then to  $\Sigma$  using transformation described in Equation (3.58).

$$\begin{aligned} p(\boldsymbol{\mu}_t, \Sigma_{t-1} | Y_{1:t-1}) &= \int p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \Sigma_{t-1}) p(\boldsymbol{\mu}_{t-1}, \Sigma_{t-1} | Y_{1:t-1}) d\boldsymbol{\mu}_{t-1} \\ &= \int p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \Sigma_{t-1}) p(\boldsymbol{\mu}_{t-1} | \Sigma_{t-1}, Y_{1:t-1}) p(\Sigma_{t-1} | Y_{1:t-1}) d\boldsymbol{\mu}_{t-1} \\ &= p(\Sigma_{t-1} | Y_{1:t-1}) \int p(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \Sigma_{t-1}) p(\boldsymbol{\mu}_{t-1} | \Sigma_{t-1}, Y_{1:t-1}) d\boldsymbol{\mu}_{t-1} \\ &= p(\boldsymbol{\mu}_t | \Sigma_{t-1}, Y_{1:t-1}) p(\Sigma_{t-1} | Y_{1:t-1}), \end{aligned} \quad (3.60)$$

where  $\boldsymbol{\mu}_t | \Sigma_{t-1}, Y_{1:t-1} \sim N_p(\hat{\boldsymbol{\mu}}_{t-1}, (c + \frac{1}{a_t}) \Sigma_{t-1})$ . Prediction to  $\Sigma$  becomes

$$\begin{aligned} p(\boldsymbol{\mu}_t, \Sigma_t | Y_{1:t-1}) &= p(\boldsymbol{\mu}_t, \frac{1}{k} \Sigma_t | Y_{1:t-1}) \left| \frac{\partial \Sigma_t}{\partial \Sigma_{t-1}} \right| \\ &\propto p(\boldsymbol{\mu}_t | \Sigma_t, Y_{1:t-1}) p(\Sigma_t | Y_{1:t-1}), \end{aligned} \quad (3.61)$$

where  $\boldsymbol{\mu}_t | \Sigma_t, Y_{1:t-1} \sim N_p(\hat{\boldsymbol{\mu}}_t, \frac{1}{k} (c + \frac{1}{a_t}) \Sigma_t)$  and  $\Sigma_t | Y_t \sim W_p^{-1}(kS_t, m_t)$ .

In an update step the prior is updated with likelihood function as in Equation (2.48)

$$\begin{aligned} \pi(\boldsymbol{\mu}_t, \Sigma_t | Y_{1:t}) &\propto p(\boldsymbol{\mu}_t, \Sigma_t | Y_{1:t-1}) l(\boldsymbol{\mu}_t, \Sigma_t | Y_t) \\ &\propto |\Sigma_t|^{-(n+p+1)/2} \exp \left[ \Sigma_t^{-1} (a_t (\boldsymbol{\mu}_t - \hat{\boldsymbol{\mu}}_t)^T (\boldsymbol{\mu}_t - \hat{\boldsymbol{\mu}}_t)) + S_t \right], \end{aligned} \quad (3.62)$$

in which

$$a_t = k \frac{a_{t-1}}{1 + a_{t-1}c} + X_t^T X_t, \quad (3.63)$$

$$\hat{\boldsymbol{\mu}}_t = a_t^{-1} \left( k \frac{a_{t-1}}{1 + a_{t-1}c} \hat{\boldsymbol{\mu}}_{t-1} + X_t^T Y_t \right), \quad (3.64)$$

$$S_t = S_{t-1} + Y_t^T Y_t + k \frac{a_{t-1}}{1 + a_{t-1}c} \hat{\boldsymbol{\mu}}_{t-1}^T \hat{\boldsymbol{\mu}}_{t-1} - a_t \hat{\boldsymbol{\mu}}_t^T \hat{\boldsymbol{\mu}}_t, \quad (3.65)$$

$$n = \sum_{i=0}^t n_i. \quad (3.66)$$

So also the posterior distribution follows a normal-inverse-Wishart distribution and the marginal distribution of  $\boldsymbol{\mu}$  and  $\Sigma$  can be derived.

### 3.4 Coverage area estimation

In this work the multivariate normal linear model is used to model a coverage area of a wireless communication node using location reports. For every CN  $c_i$ , there is a list of measurements  $Y = (\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(n)})^T$ , which refers to positions where a CN  $c_i$  is heard, that is every measurement  $\mathbf{y}_{(i)}$  is a  $p$ -variate vector of coordinates. Measurement equation can be written as a multivariate linear model

$$Y = \mathbf{1}_n \boldsymbol{\mu}^T + \boldsymbol{\epsilon}, \quad (3.67)$$

where  $\boldsymbol{\epsilon}_{(i)}$  is an error vector distributed as  $N_p(\mathbf{0}, \Sigma)$ . The unknown parameters of the CN that are of interest are  $\boldsymbol{\mu}$ , the location parameter of the CN, and  $\Sigma$ , the shape parameter of the CN. The location parameter describes where measurements are located and the shape parameter measures how latitude and longitude coordinates of measurements are correlated and describes the size and shape of a coverage area and how it is oriented.

Result of the Bayesian analysis is a posterior distribution. In this kind of application it is appropriate to summarize the information the posterior distribution gives to some Bayes' estimate presented in Section 2.2.2. For this model the posterior is completely specified by the Bayes' estimates. Probability for the location  $\mathbf{p}$  of a UE in the coverage area  $c$  follows normal distribution, namely

$$\mathbf{p} | c \sim N_p(\hat{\boldsymbol{\mu}}, \hat{\Sigma}), \quad (3.68)$$

where  $\hat{\boldsymbol{\mu}}$  and  $\hat{\Sigma}$  are Bayes' estimates for  $\boldsymbol{\mu}$  and  $\Sigma$ , respectively. As an example, if an informative prior is used to generate the posterior and mean estimates of parameters are used, the parameters of a coverage area would be

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \text{E}(\boldsymbol{\mu}) = \hat{\mathbf{m}} \text{ and} \\ \hat{\Sigma} &= \text{E}(\Sigma) = \frac{1}{n^* - p - 2} \hat{S}.\end{aligned}\tag{3.69}$$

## 4 Bayesian positioning

Next a Bayesian method for determining a position estimate for a UE using existing coverage area information about the CNs is presented. Assume that there exists an RM, which contains estimated coverage areas of CNs. The goal is to generate a position estimate of a UE using this RM and information that the UE measures i.e. the identification codes of heard communication nodes. Assume that coverage areas are modeled as multivariate normal distributions with known place and shape parameters  $(\boldsymbol{\mu}, \Sigma)$ .

Let  $\mathbf{c} = (c_1, c_2, \dots, c_n)$  be a list of CNs heard by a UE and  $c_i = (\boldsymbol{\mu}_i, \Sigma_i)$  where  $\boldsymbol{\mu}_i$  and  $\Sigma_i$  are the place and the shape parameter of  $i$ th heard CN, respectively. Let  $\mathbf{p}$  be a  $p$ -variate vector containing the position of the UE. Assumed that  $\mathbf{p}|c_i \sim N_p(\boldsymbol{\mu}_i, \Sigma_i)$ . The probability that the UE located at  $\mathbf{p}$  hears the  $i$ th CN can be obtained using Bayes' rule

$$\begin{aligned} p(c_i \in \mathbf{c}|\mathbf{p}) &\propto p(\mathbf{p}|c_i \in \mathbf{c})p(c_i \in \mathbf{c}) \\ &\propto p(\mathbf{p}|c_i \in \mathbf{c}). \end{aligned} \quad (4.1)$$

The last approximation is obtained using an assumption that prior  $p(c_i \in \mathbf{c})$  is uniformly distributed, so prior probability that UE hears CN  $c_i$  is equal for all  $i$ s. Assuming that observations are mutually independent the likelihood function for  $\mathbf{p}$  can be formed

$$\begin{aligned} l(\mathbf{p}|\mathbf{c}) &= \prod_{i=1}^n p(c_i \in \mathbf{c}|\mathbf{p}) \\ &\propto \prod_{i=1}^n \exp \left[ -\frac{1}{2}(\mathbf{p} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{p} - \boldsymbol{\mu}_i) \right] \\ &= \exp \left[ -\frac{1}{2} \sum_{i=1}^n (\mathbf{p} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{p} - \boldsymbol{\mu}_i) \right]. \end{aligned} \quad (4.2)$$

The exponent in Equation (4.2) becomes

$$\begin{aligned} &\sum_{i=1}^n (\mathbf{p} - \boldsymbol{\mu}_i)^T \Sigma_i^{-1}(\mathbf{p} - \boldsymbol{\mu}_i) \\ &= \sum_{i=1}^n \left[ \mathbf{p}^T \Sigma_i^{-1} \mathbf{p} - \mathbf{p}^T \Sigma_i^{-1} \boldsymbol{\mu}_i - \boldsymbol{\mu}_i^T \Sigma_i^{-1} \mathbf{p} + \boldsymbol{\mu}_i^T \Sigma_i^{-1} \boldsymbol{\mu}_i \right] \\ &= \mathbf{p}^T \left( \sum_{i=1}^n \Sigma_i \right) \mathbf{p} - \mathbf{p}^T \left( \sum_{i=1}^n \Sigma_i^{-1} \boldsymbol{\mu}_i \right) - \left( \sum_{i=1}^n \boldsymbol{\mu}_i^T \Sigma_i^{-1} \right) \mathbf{p} + \sum_{i=1}^n \boldsymbol{\mu}_i^T \Sigma_i^{-1} \boldsymbol{\mu}_i \\ &= (\mathbf{p} - \hat{\mathbf{p}})^T S^{-1}(\mathbf{p} - \hat{\mathbf{p}}) + \text{constant}, \end{aligned} \quad (4.3)$$

in which

$$\begin{aligned}\hat{\mathbf{p}} &= \left(\sum_{i=1}^n \Sigma_i^{-1}\right)^{-1} \left(\sum_{i=1}^n \Sigma_i^{-1} \boldsymbol{\mu}_i\right) \\ \mathbf{S} &= \left(\sum_{i=1}^n \Sigma_i^{-1}\right)^{-1}.\end{aligned}\tag{4.4}$$

Using normal prior  $\mathbf{p} \sim N_p(\boldsymbol{\mu}_0, \Sigma_0)$ , the posterior density becomes

$$\begin{aligned}\pi(\mathbf{p}|\mathbf{c}) &\propto p(\mathbf{p})l(\mathbf{p}|\mathbf{c}) \\ &\propto \exp\left[-\frac{1}{2}(\mathbf{p} - \hat{\mathbf{p}})^T \mathbf{S}^{-1}(\mathbf{p} - \hat{\mathbf{p}})\right],\end{aligned}\tag{4.5}$$

which is a normal density  $N_p(\hat{\mathbf{p}}, \hat{\mathbf{S}})$ , where  $\hat{\mathbf{p}} = (\sum_{i=0}^n \Sigma_i^{-1})^{-1}(\sum_{i=0}^n \Sigma_i^{-1} \boldsymbol{\mu}_i)$  and  $\hat{\mathbf{S}} = (\sum_{i=0}^n \Sigma_i^{-1})^{-1}$ . So the position estimate is a weighted mean and the weights are inverses of the covariance matrixes.

Figure 3 illustrates how position estimate is distributed given heard CNs. Figure 3 shows distributions of three CNs and Figure 3c shows contours for corresponding coverage areas. Figure 3b shows the distribution of position estimate, given these three CNs, is drawn and Figure 3b shows the contours for position estimate.

If no prior information about a position estimate is known, it is justified to use a prior with a large covariance. In that case, the prior distribution approaches uniform distribution. When the prior covariance is large, inverse of the covariance approaches to a zero matrix, and effect of prior mean to posterior mean vanishes. As an example, if the location of the UE is known with accuracy of the country, a coarse prior could be a circle or an ellipse, which covers that country. If prior knowledge about the location of the UE improves, prior location can also be with accuracy of a city or of a building.

A weakness of the model is the assumption that observations, that is, distributions of CNs are mutually independent and reception probability follows a normal distribution. In some case this assumption may be misleading. Suppose, for example, that there are  $n$  observations and coverage area of observations are identically distributed with same parameters  $\boldsymbol{\mu}$  and  $\Sigma$ . Then the position estimate is the place parameter  $\boldsymbol{\mu}$  of communication node, as one may expect, but the uncertainty of position estimate is no longer same as the uncertainty of CN, but has decreased to  $\frac{1}{n}\Sigma$ . Also, if it known that two coverage areas  $c_i$  and  $c_j$  overlap, it is misleading to assume that they are independent on each other. In this case, if it is known that UE is on a coverage area of  $c_i$ , there is also a probability of being on coverage area of  $c_j$ .

It could also be taken account all CNs in RM, instead of just taking account of heard CNs, because it may also give convenient information when particular CN is *not* heard at location  $\mathbf{p}$ . It could be expected that if some



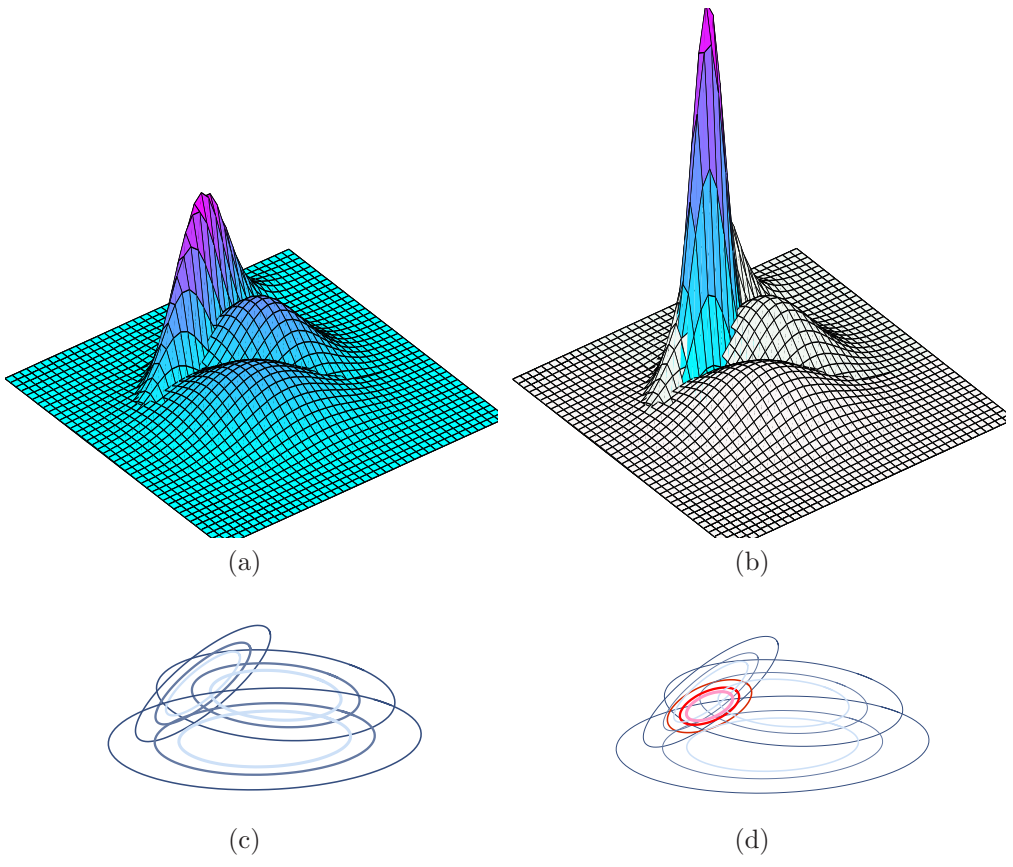


Figure 3: Distribution of coverage areas and position estimate. 50%, 68% and 90 % contours for distributions are drawn.

CN is not heard in location  $\mathbf{p}$  then the probability of being on its coverage area is small. In this work this information is not taken into account. One reason for this is that CNs UE reports to be heard is not always the maximum number of CN. For example in GSM network UE is mandated to report the six strongest, although actual number of heard cells may be larger. Other reason is to reduce computational requirements, because especially when RM is large, the computational burden would be huge compared to the case when only heard CNs are taken account.

## 5 Evaluation

In this section all the methods discussed in Sections 3 and 4 are implemented and tested using simulated data. The multivariate linear model is used to estimate the coverage area of a communication node using location reports. Coverage areas are estimated in three different conditions. First assuming that there is no prior information available about the coverage area as in Section 3.1.2, secondly assuming that some prior information is available as in Section 3.1.3 and thirdly, as a special case of the second one, assuming that there is prior information available but there is a probability that some measurements may be outliers as in Section 3.2.1. It is simulated how coverage area of communication node is updated using methods presented in Sections 3.3.1 and 3.3.2, when new reports come in. Positioning method described in Section 4 is tested and it is studied how positioning accuracy changes using different models to estimate the coverage area. Also the consistency of position estimate is studied.

### 5.1 Coverage area of a communication node

The resulting coverage area estimate is not the actual geographical coverage area of CN and even more importantly, the estimated place parameter is not the location of the communication node. Actually, the coverage area estimate gives information about how data collectors are distributed in the coverage area of CN. In data collecting phase data collectors are located in places where GNSS positions are available, usually outdoor environment, such as highways and city centers. Thus, the coverage area estimate gives the most likely location of the users in the given coverage area of CN, although actual coverage area overlaps a bigger area. This model is, however, convenient to use because the scope is to present a positioning technique which provides a good positioning for the majority of users.

Because the coverage area is assumed to be normally distributed, the location parameter  $\mu$  of the coverage area converges to the sample mean when the sample size increases, especially when flat prior is used. There may be a case where a fingerprint collector stands at the same location for a long time and reports a large cluster of fingerprints from a single location. This may lead to a false coverage area model because it drags the sample mean toward that single point and reduces deviation of sample, and thus leads to a smaller coverage area model. This problem could be handled for example by somehow pre-filtering the data. One way for pre-filtering could be to combine fingerprints close enough each other so that the cluster of fingerprints reduces to just one fingerprint. This, however, requires that all the fingerprints should be stored in the database. The author in [53, p. 58] proposed two other solutions for this problem. One solution was to use a velocity information from a GNSS receiver to recognize when the data collector is not moving

and omit the fingerprint collection for those time instants. Other solution was to 'freeze' fingerprint collection after some threshold value of incoming fingerprints. Both solutions have, however, assumes that all the data comes from the same user. There may also be several users in the same location and the second solution does handle the situation when coverage area changes.

In case there are only few location reports from a single CN it is advantageous to use the prior information for estimating the coverage area of a CN. It is also convenient to use the information that is available about the coverage area. Information about a network type, such as WLAN, cellular, is always available. This information is essential, because different network types uses different technologies. A radius of a cellular GSM cell, for example, can be up to 35 km [54] whereas the coverage of a WLAN access point is only few hundreds of meters.

Using priors is essential because positioning accuracy suffers when the coverage of a CN is modeled incorrectly. If the coverage of the CN is modeled to be too big, the uncertainty of the position estimate increases. Too small coverage areas decreases the uncertainty of the position estimate, but may lead to an incorrect position estimate. Use of priors is vital especially when there are only a few observations from the CN because in such a case there is not enough information about the whole coverage area. When sample size increases the effect of prior decreases.

When estimating a prior distribution of a location parameter, information about a physical place of some of heard CN could be used. Unfortunately, this information is not usually available and it is preferred to use the flat prior distribution for the location parameter, which prior has minimal effect on the posterior distribution.

Use of the prior when estimating the coverage area is illustrated in Figure 4. Figure 4a shows a real coverage area of a CN and observations reported from the communication node are drawn in red. Figure 4b shows an estimated coverage area when the Jeffreys prior is used. Figure 4c shows an estimated coverage area when an informative prior is used. The prior coverage area is drawn with dashed line. Figure 4 shows that use of the informative prior leads to better coverage area estimates.

## 5.2 Updating coverage area estimate

The coverage areas of CNs changes all the time and it is not realistic to assume that at some point the coverage area model is ready. There can be changes in transmission power and the coverage area may increase or decrease. There may also be changes in antennas, for example a direction of an antenna may change, which leads to changes in coverage area. In some cases, the CN may change its location altogether, for example mobile access points or mobile cells may be transported to an area which requires an extra network capacity, for example for music festivals. There may also be a case

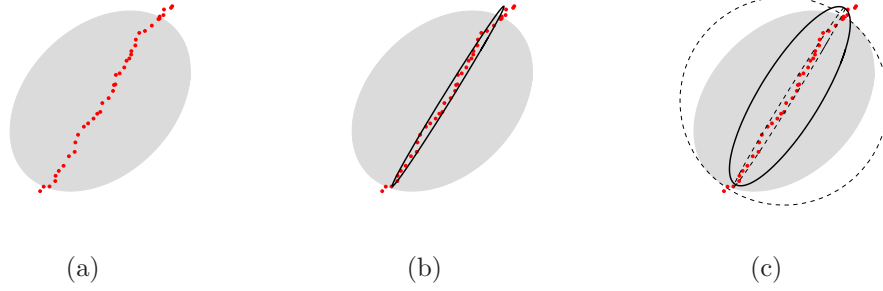


Figure 4: Effect of prior to coverage area estimate

where one coverage area is split in two in such a way that the original ID of CN remains with one of the two coverage areas and the other coverage area gets new ID.

All the location reports do not come at the same time and new data comes in after the first cell estimate is made. It is important to be able to update the coverage area model when new data comes. The coverage area model can be updated recursively using formulas derived in Section 3.3. Section 3.3.1 assumes the coverage area to be static. In that case it is assumed that there are no changes in the coverage area and all the measurements come from the same distribution. Section 3.3.2 assumes that the coverage area may evolve under time and gives more weight to new observations. There are additional parameters  $c$  and  $k$  where  $c$  is a 'decay factor' and it gives less weight to old measurements and  $k$  is a factor which increases the uncertainty of the shape parameter of older measurements. In both cases, the measurement model is the same as in Equation (3.67). In the special case of  $c = 0$  and  $k = 1$ , the state model reduces to the static coverage area model presented in Section 3.3.1. This section illustrated how the coverage area is updated using different values of  $c$  and  $k$ .

Figure 5 shows an example how the updating formulae works. Figure 5 illustrates a situation where the direction of antenna is changed. There are 200 location reports from an original coverage area — only a part of observations is drawn in Figure 5. After that the direction of the antenna is turned 70 degrees clockwise and 20 location reports arrive from a new coverage area. It is assumed that one fingerprint arrives for every time step.

Figure 6 illustrates the movement of the coverage area estimate. In each figure, movement is presented with one value of  $k$  and three values of  $c$ . Dashed ellipses drawn in figures are coverage area estimates after 200 location reports and solid ellipses are coverage area estimates after 220 location reports. Coverage area estimates drawn in figures are 68% contours of the distributions of coverage area estimates. Figure 6a shows that when  $c = 0$ , which is the static coverage area model, the coverage area estimate does not move significantly when only 20 location reports have arrived from

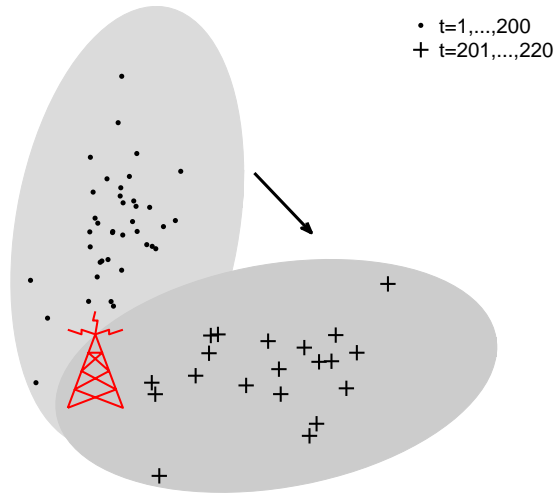


Figure 5: Test situation

a new coverage area. However already with the small values of the decay factor  $c$  the coverage area estimate moves towards the new coverage area and forgets the old location reports. When the value of  $c$  increases, the coverage area estimate seems to follow new location reports closely, which also affects on the size of the coverage area estimate leading to smaller coverage area estimates. Figures 6b and 6c shows that already with small increase of  $k$ , the coverage area estimate grows rather fast.

In real life static updating may not be as straightforward as presented in this section. If, for example, outlier detection is in use, observations from changed coverage area may be detected as outliers.

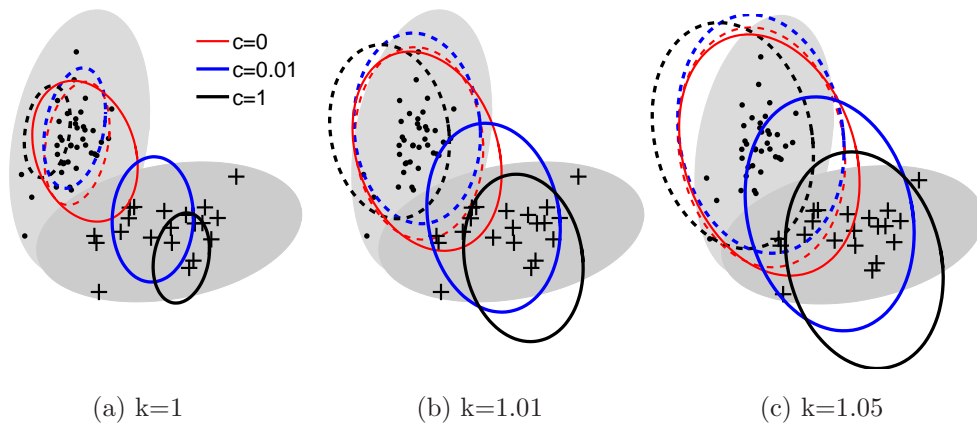


Figure 6: Example of update. Dashed lines ellipses are coverage area estimates after 200 locations. Solid ellipses are coverage area estimates after 220 location reports.

## 5.3 Positioning

In this section the different models derived in Section 3 are used to form a coverage area estimate on the basis of location reports to compare how the models affect a position estimate. The effect of different Bayes estimates on the positioning accuracy is studied. Two Bayes estimates for posterior distribution are compared, Bayes estimate using quadratic loss function, that is a mean of the marginal posteriors of parameters and Bayes estimate using zero-one loss function, that is MAP-estimate. Because t-distribution, marginal posterior of  $\mu$ , is symmetric, mean and MAP estimates are same for the place parameter, but for the shape parameter the used point estimates are different. Using these different coverage area models and estimates for parameters the positioning is made using the formula presented in Section 4. The position estimate is both a mean and MAP estimate, because posterior of position estimates are normally distributed, which is a symmetric distribution. The position estimates are compared with the exact positions. The measure used in comparing different coverage area models is the Euclidean norm of a 2-dimensional error vector.

### 5.3.1 Simulations

Fingerprints include a list of different CN IDs from various networks. In cellular network, for example, fingerprints include IDs of both serving and neighboring cells. Serving cell is the one that is associated with UE and neighboring cells are those around serving cell whose identity can be decoded, but are not currently used. During the data collection phase, fingerprints are collected with a UE with positioning capacity. One fingerprint consists of a list of base station ID's heard and coordinates of the UE. Fingerprints are stored in the fingerprint database. Fingerprints are processed so that fingerprints are split into location reports by a CN. So for every communication node ID, there is a a list of coordinates, where the communication node is heard. These location reports are used to estimate the coverage area of each communication node. The ellipse-shaped coverage areas are modeled to be multivariate normal distributions. An RM is generated using coverage area information about each CN, see Figure 7. In the positioning phase the UE hears IDs of both serving and neighboring communication nodes and compares this information with the RM and generates the best position estimate.

Position estimation is done using generated networks. First a cellular network is built which includes a GSM and a WCDMA network. The GSM network is generated so that it covers the whole area of interest. The WCDMA network is built parallel with the GSM network and is used whenever it is available. The WCDMA network is located in urban areas, where the density of users is higher. Also some WLAN access points are

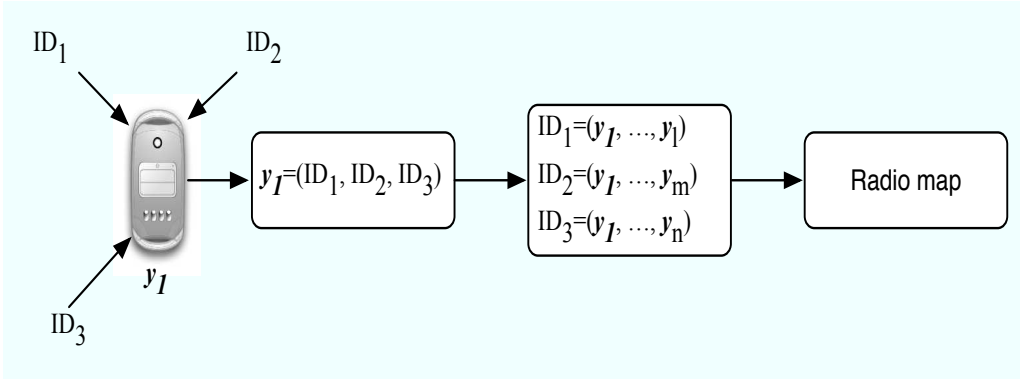


Figure 7: Radio map generation

generated and can be used to improve position accuracy. WLAN access points are mostly located in urban environment, where population density is high. Coverage areas are assumed to be ellipses and different priors are used for different RANs.

In real life, fingerprints are not uniformly distributed, not to mention normally distributed, in the coverage areas. As mentioned before, fingerprints are distributed in places where during data collecting phase GNSS is available, for example along roads. For this reason, a road map is generated on a test map and fingerprints, are taken uniformly from roads. Some fingerprints are also generated using a random walk. Generated fingerprints are sorted by CN and a RM is generated using coverage area models presented in Sections 3.1.2, 3.1.3 and 3.2.1. Another RM is generated, where fingerprints are distributed in coverage areas as assumed by the model, i.e. fingerprints are normally distributed in coverage areas of CNs. In both cases, circular priors, where the radius depends on RAN, are used.

Test cases to be located are generated uniformly from the test map, and come also from places where fingerprints were not generated, for example inside buildings.

In following, notation  $R_{\text{non}}$  is used for the RM, where coverage areas are calculated using normal model with Jeffreys prior,  $R_{\text{inf}}$  is used for the RM when prior information about coverage areas is used and  $R_{\text{out}}$  is used for the RM, when possibility of outliers is considered when coverage areas are modeled. When  $R_{\text{non}}$  is generated, in the cases where there is only one incoming observation or all observations are from the same point, the shape parameter  $\Sigma$  of coverage area estimate would be zero matrix. This may cause problems in positioning phase because zero matrix is not invertible. For that reason in those cases a small covariance is used for a shape parameter, so that computations would be possible. CNs with zero covariance matrix could also be left out from the RM. Bayes' estimate used for shape parameter  $\Sigma$  (mean or MAP) is marked as superscript.



### 5.3.2 Location accuracy using different coverage area models

Results show that the accuracy of location estimate depends on the coverage area model. In urban areas positioning accuracy is slightly better using  $R_{\text{inf}}^{\text{mean}}$  compared to  $R_{\text{out}}^{\text{mean}}$ . Mean error using  $R_{\text{inf}}^{\text{mean}}$  is 135 meters, whereas mean error using  $R_{\text{out}}^{\text{mean}}$  is 151 meters. However, there is a significant decrease in performance when  $R_{\text{non}}^{\text{mean}}$  is used. Mean error has increased to 1.62 kilometers. Figure 8 presents a few test cases to illustrate why positioning errors are bigger using  $R_{\text{non}}$ . There are cells with only a few incoming observations and coverage area estimate is so small that positioning algorithm drags positioning estimate onto it, as shown in Figure 8a. There is a WLAN access point with only one observation and position estimate turns out to be that observation. Also, if all observations come from almost collinear points, uncertainty of the coverage area estimate is small in one direction but large in other direction, which may lead to large positioning error, as shown in Figure 8b. Cases illustrated in Figure 8b cause the biggest errors in  $R_{\text{non}}$ . For urban areas  $R_{\text{non}}$  results in 21.4-kilometer 95% accuracy whereas  $R_{\text{inf}}$  and  $R_{\text{out}}$  achieve 450-meter 95% accuracy. Therefore, Figures 8a and 8b show that the prior information about coverage area improves position estimate.

In this simulation outliers are location reports which come from locations that are far away from the majority of location reports, as illustrated in Figure 9.  $R_{\text{inf}}^{\text{mean}}$  results in a slightly better positioning accuracy than  $R_{\text{out}}^{\text{mean}}$ .  $R_{\text{out}}^{\text{mean}}$  may have treated some 'good' observations as outliers. Difference in positioning accuracy is so small that the conclusion is that although the outlier detection has made some falls, it does not jeopardize the system performance. As an example outlier detection could work wrong by treating all incoming observation during radio map generation as outliers and rejecting them.

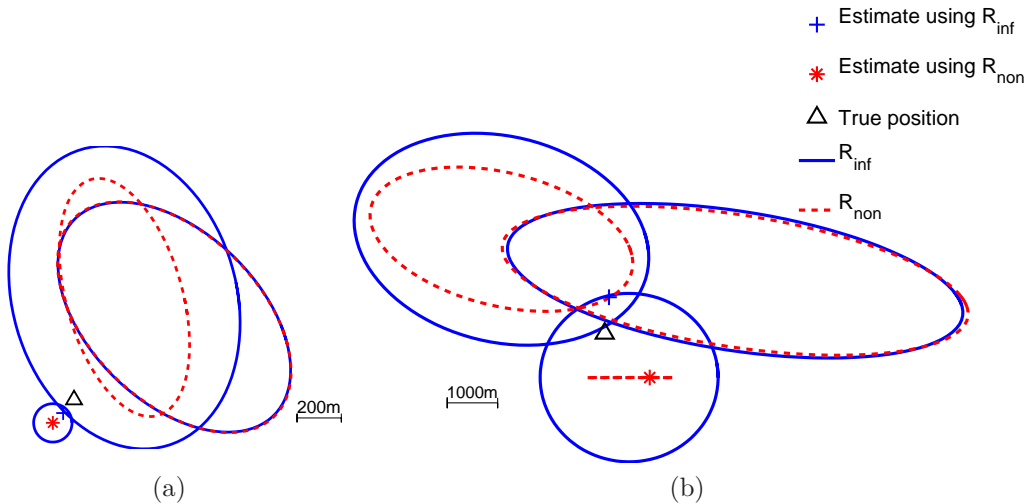


Figure 8: Examples of test cases.

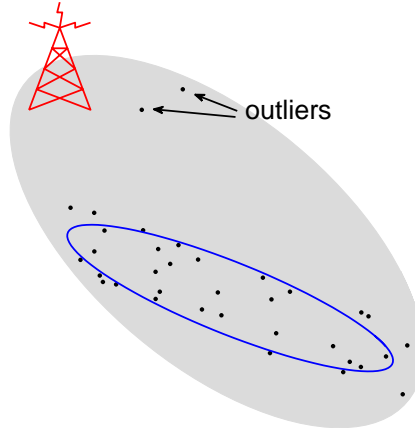


Figure 9: Example of outliers.

In rural areas positioning accuracy is not as good as in urban areas. This result is expected, because the GSM network is used and no WLAN access points are available. In rural areas, there is no difference in positioning accuracy between different coverage area models. In rural areas, mean error is 1.96 kilometers with  $R_{\text{inf}}^{\text{mean}}$  and  $R_{\text{out}}^{\text{mean}}$  and mean error using  $R_{\text{non}}^{\text{mean}}$  is 1.94 kilometers. Results show that the clear advantage of the use of a prior seen in urban areas is not achieved in rural areas.

Altogether results show that use of the prior leads to better positioning accuracy. There is not large difference in positioning accuracy, whether outlier detection is used or not. The computational burden, however, is significantly larger using outlier detection. It required almost 14 times more time to form  $R_{\text{out}}^{\text{mean}}$  than  $R_{\text{inf}}^{\text{mean}}$  or  $R_{\text{non}}^{\text{mean}}$ . MAP-estimate gives smaller coverage area estimates than mean-estimate and in urban areas MAP estimate gives slightly better positioning accuracy. Improvement is shown especially in test cases where positioning error is already small,  $R_{\text{inf}}^{\text{mean}}$  results in 122-meter 68% accuracy whereas  $R_{\text{inf}}^{\text{mean}}$  achieves 94-meter 68% accuracy. In rural areas there are no clear differences in positioning accuracy between different estimates for  $\Sigma$ . Results for positioning accuracy are summarized in Tables 1 and 2.

In an ideal case fingerprints are normally distributed in the coverage area as assumed by the model. Hence the generated coverage areas estimate the real coverage areas better. It is tested if these coverage area estimates improve positioning performance. Positioning accuracy is tested using the same network and the same test cases as in previous.

Results show that positioning accuracy is improved in urban areas, when fingerprints are generated as assumed by coverage area model. Mean error using  $R_{\text{inf}}^{\text{mean}}$  has decreased to 93 meters. In rural areas, however, results are opposite. Mean error using  $R_{\text{inf}}^{\text{mean}}$  has increased to 2.12 kilometers. Results show the positioning accuracy benefits from modeling the distribution of

Table 1: Positioning errors using mean-estimate in RM generators. Fingerprints are generated from roads.

Urban	68% [m]	95% [m]	mean [m]	rmse [m]
$R_{\text{inf}}^{\text{mean}}$	122	451	135	261
$R_{\text{out}}^{\text{mean}}$	177	451	151	270
$R_{\text{non}}^{\text{mean}}$	111	21452	1621	5677
Rural	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{mean}}$	2.17	4.22	1.96	2.31
$R_{\text{out}}^{\text{mean}}$	2.17	4.22	1.96	2.31
$R_{\text{non}}^{\text{mean}}$	2.17	4.21	1.94	2.30
All	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{mean}}$	1.12	3.50	0.85	1.46
$R_{\text{out}}^{\text{mean}}$	1.12	3.50	0.86	1.46
$R_{\text{non}}^{\text{mean}}$	1.57	5.17	1.75	4.65

Table 2: Positioning errors using MAP-estimate in RM generators. Fingerprints are generated from roads.

Urban	68% [m]	95% [m]	mean [m]	rmse [m]
$R_{\text{inf}}^{\text{MAP}}$	94	452	129	258
$R_{\text{out}}^{\text{MAP}}$	150	452	144	265
$R_{\text{non}}^{\text{MAP}}$	110	21452	1615	5677
Rural	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{MAP}}$	2.17	4.22	1.96	2.31
$R_{\text{out}}^{\text{MAP}}$	2.17	4.22	1.96	2.31
$R_{\text{non}}^{\text{MAP}}$	2.17	4.21	1.94	2.30
All	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{MAP}}$	1.12	3.50	0.85	1.46
$R_{\text{out}}^{\text{MAP}}$	1.12	3.50	0.86	1.46
$R_{\text{non}}^{\text{MAP}}$	1.57	5.17	1.75	4.65

network users rather than the actual coverage area. Positioning accuracy for the ideal case is summarized in Tables 3 and 4. Results show that the positioning performance does not suffer even though in real life incoming data is not similar as it is assumed by the model and in some cases benefits from it.

Table 3: Positioning errors using mean-estimate in RM generators. Fingerprints are generated as assumed by model.

Urban	68% [m]	95% [m]	mean [m]	rmse [m]
$R_{\text{inf}}^{\text{mean}}$	97	221	93	153
$R_{\text{out}}^{\text{mean}}$	97	226	94	153
$R_{\text{non}}^{\text{mean}}$	106	237	99	155
Rural	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{mean}}$	2.23	4.34	2.15	2.45
$R_{\text{out}}^{\text{mean}}$	2.23	4.34	2.15	2.45
$R_{\text{non}}^{\text{mean}}$	1.89	4.64	2.08	2.39
All	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{mean}}$	1.25	3.21	0.82	1.46
$R_{\text{out}}^{\text{mean}}$	1.25	3.21	0.82	1.46
$R_{\text{non}}^{\text{mean}}$	1.01	3.05	0.80	1.42

Table 4: Positioning errors using MAP-estimate in RM generators. Fingerprints are generated as assumed by model.

Urban	68% [m]	95% [m]	mean [m]	rmse [m]
$R_{\text{inf}}^{\text{MAP}}$	95	216	92	152
$R_{\text{out}}^{\text{MAP}}$	95	221	93	152
$R_{\text{non}}^{\text{MAP}}$	103	230	96	153
Rural	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{MAP}}$	2.12	4.36	2.12	2.41
$R_{\text{out}}^{\text{MAP}}$	2.12	4.36	2.12	2.41
$R_{\text{non}}^{\text{MAP}}$	1.77	4.64	2.02	2.36
All	68% [km]	95% [km]	mean [km]	rmse [km]
$R_{\text{inf}}^{\text{MAP}}$	1.25	3.19	0.81	1.44
$R_{\text{out}}^{\text{MAP}}$	1.25	3.19	0.81	1.44
$R_{\text{non}}^{\text{MAP}}$	1.00	3.03	0.78	1.40

Coverage area estimates and positioning algorithms are not restricted to only cellular network and WLAN but can be applied also in other wireless networks. Accuracy of the positioning method presented in this thesis is also studied indoors in [55]. The methods are tested using a prototype wireless sensor network. One commonly used positioning method is just to calculate a mean of the coordinates of CNs [56]. Results show that the use of coverage area information in position calculation improves positioning accuracy. In [55] also the signal strength information is used to improve the positioning accuracy, but the improvement in positioning accuracy is slight.

### 5.3.3 Consistency of location estimates

In previous, only accuracy of a position estimate was used to evaluate the positioning algorithm and the coverage area estimate. In addition to accuracy, also the consistency of the position estimate can be used to evaluate the quality of a model. The position estimate is said to be consistent if its predicted error is at least as large as actual errors [57].

Covariance of the position estimate can be used to evaluate consistency. If the covariance of the position estimate is too small, it no longer represents a reliable measure for the uncertainty of the position estimate [58]. Too large covariance as compared to real uncertainty may decrease the usability of position estimate. In this work the consistency of the position estimate is studied using CEP (Circular Error Probable) values and NEES (Normalized Estimated Error Squared). CEP value is defined to be a radius of a circle that contains a certain percentage of the error distributions when centered at the correct location [59]. As an example, for a two-dimensional Normal distribution,

$$\text{CEP68} = 0.75(\sqrt{e_1} + \sqrt{e_2}) \text{ and} \quad (5.1)$$

$$\text{CEP95} = 1.22(\sqrt{e_1} + \sqrt{e_2}), \quad (5.2)$$

where  $e_1$  and  $e_2$  are eigenvalues of the covariance matrix. NEES is defined as follows

$$\text{NEES} = (\mathbf{p} - \hat{\mathbf{p}})^T \hat{\mathbf{S}} (\mathbf{p} - \hat{\mathbf{p}}) \quad (5.3)$$

and consistency is checked using a chi-squared test:

$$\text{NEES} \leq \chi_p^2(1 - \alpha), \quad (5.4)$$

where  $\alpha$  is the desired confidence level [60, p. 345].

In urban areas the covariance of position estimate tends to be small. Position estimate believes that, when coverage area estimates are small, the position estimate has to be accurate. Both 68% and 95% quantile of positioning error using  $R_{\text{inf}}^{\text{mean}}$  (see Table 1) lies between corresponding CEP68 and CEP95 values. CEP68 shows 80-meter 68% quantile and 377-meter

95% quantile, whereas CEP95 shows 130-meter 68% quantile and 612-meter 95% quantile.  $R_{out}^{mean}$  shows similar results. In urban areas, when  $R_{non}^{mean}$  is used, both CEP68 and CEP95 values are below corresponding values in positioning accuracy.  $R_{non}^{mean}$  shows 39-meter and 62-meter 68% CEP68 and CEP95 quantiles, respectively. In test cases as described in Figure 8 the uncertainty of position estimate is very small although positioning error may be big.

In rural areas the uncertainty estimate gives more reliable estimates for errors. For all RMs 68% quantile of position error is below 68% quantile for both CEP68 and CEP95. Also, 95% quantile of position error is below corresponding CEP95 value. Altogether Figure 10 shows that CEP95 satisfies the definition of consistency. It is above position error in all cases. CEP values are summarized in Table 5.

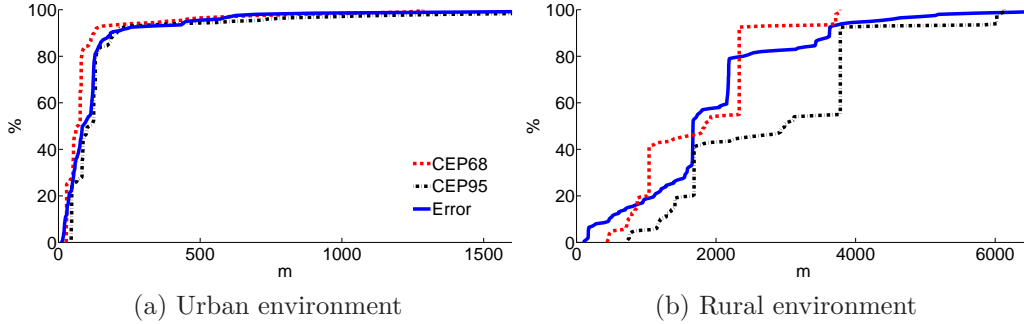


Figure 10: Quantile plots of error and CEP68 and CEP95 values in urban areas and in rural areas when  $R_{out}^{mean}$  is used.

NEES test shows similar results with CEP. In urban areas position estimate has difficulties with consistency, especially when  $R_{non}$  is used. In that case position estimate is consistent only for 8% of test cases with confidence level  $\alpha = 0.05$  and 11% of test cases with confidence level  $\alpha = 0.01$ . For  $R_{inf}^{mean}$  corresponding values are 60% and 66%, which are also quite low. NEES test shows acceptable consistency only in rural areas, when  $R_{inf}^{mean}$  or  $R_{out}^{mean}$  is used. In both cases position estimate is consistent for 97% of test cases with confidence level  $\alpha = 0.01$ .

When fingerprints are ideally distributed on coverage areas, the uncertainty of position estimate is in the same order with actual position error. In urban areas, when  $R_{inf}^{mean}$  is used, 99% of test cases were consistent with confidence level  $\alpha = 0.01$  and in rural areas all test cases were consistent with confidence level  $\alpha = 0.01$ . This shows that consistency of position estimate suffers, when in real life incoming data is not as assumed by model. Results in consistency using NEES test are summarized in Table 6.

Table 5: CEP68 and CEP95 values using different RM generators.

Urban		68% [m]	95% [m]
CEP68	$R_{\text{inf}}^{\text{mean}}$	80	377
	$R_{\text{out}}^{\text{mean}}$	80	377
	$R_{\text{non}}^{\text{mean}}$	39	249
CEP95	$R_{\text{inf}}^{\text{mean}}$	130	612
	$R_{\text{out}}^{\text{mean}}$	130	612
	$R_{\text{non}}^{\text{mean}}$	62	404
Rural		68% [km]	95% [km]
CEP68	$R_{\text{inf}}^{\text{mean}}$	2.33	3.71
	$R_{\text{out}}^{\text{mean}}$	2.33	3.71
	$R_{\text{non}}^{\text{mean}}$	2.32	3.70
CEP95	$R_{\text{inf}}^{\text{mean}}$	3.78	6.02
	$R_{\text{out}}^{\text{mean}}$	3.78	6.02
	$R_{\text{non}}^{\text{mean}}$	3.76	6.00
All		68% [km]	95% [km]
CEP68	$R_{\text{inf}}^{\text{mean}}$	1.04	2.33
	$R_{\text{out}}^{\text{mean}}$	1.04	2.33
	$R_{\text{non}}^{\text{mean}}$	1.03	2.32
CEP95	$R_{\text{inf}}^{\text{mean}}$	1.68	3.78
	$R_{\text{out}}^{\text{mean}}$	1.68	3.78
	$R_{\text{non}}^{\text{mean}}$	1.67	3.76

Table 6: Percentages of test cases where position estimate is consistent

		Fingerprints are generated from roads		Fingerprints are generated from coverage areas	
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
Urban	$R_{\text{inf}}^{\text{mean}}$	60	66	90	99
	$R_{\text{out}}^{\text{mean}}$	59	65	90	99
	$R_{\text{non}}^{\text{mean}}$	8	11	85	91
Rural	$R_{\text{inf}}^{\text{mean}}$	73	97	99	100
	$R_{\text{out}}^{\text{mean}}$	72	97	99	100
	$R_{\text{non}}^{\text{mean}}$	67	86	76	78
All	$R_{\text{inf}}^{\text{mean}}$	65	78	93	99
	$R_{\text{out}}^{\text{mean}}$	64	77	93	99
	$R_{\text{non}}^{\text{mean}}$	31	40	82	86

Consistency tests show that the positioning algorithm has more problems in consistency in urban areas than in rural areas. One reason for this is that in urban areas the coverage area estimates are smaller compared to the actual coverage area than in rural areas. It is possible that in urban area location reports arrive only from one certain area of coverage area. Location reports may, for example, come only from one road, which cuts the coverage area. In rural areas, the coverage areas are larger, it is more likely, that location reports arrive from larger area, for example from two roads crossing each other. Naturally, because in urban areas coverage area estimates are of lower quality, also consistency suffers.



## 6 Conclusion and future work

In this work a Bayesian framework for estimating the coverage area of a wireless CN is introduced. Coverage areas are modeled using location reports measured by UEs. Also a method for positioning a UE using a database of coverage area models is presented.

This thesis presents location fingerprinting approach to model the coverage area of wireless communication node. Location fingerprinting is a positioning technique that determines user's location using a database of signal characteristics. Short literature review about location fingerprinting is made together with a survey of existing solutions.

Coverage areas are assumed to be ellipses and the unknown parameters of ellipses are solved. Unknown parameters are the place parameter and the shape parameter of an ellipse. In this work coverage areas are assumed to follow a normal distribution and multivariate normal linear model is used to estimate the coverage areas of CNs. Bayes' rule is used to calculate the posterior distributions of unknown parameters. In this work the information the posterior gives is specified with point estimates of the posterior distribution. Coverage areas of CNs depend highly on used RAN and Bayes' rule offers the possibility to use prior information about the coverage area. Also, the multivariate linear model allows a closed-form solution for coverage area estimates, which leads to a low computational burden.

Three different approaches for modeling the coverage area are presented. First approach uses noninformative prior to calculate the posterior distribution of parameters. Second approach uses the natural conjugate prior. Third approach uses natural conjugate prior, but it assumes that there is a probability that some of the locations reports may be outliers. Also a method to update the coverage area estimates recursively is derived using Bayesian filtering.

For positioning purposes posterior distribution of the UE position, given a set of heard CNs, is derived. Posterior of the position estimate follows a normal distribution with the mean that is a weighted average of the centers of the coverage area ellipses. Weights are determined from the shape parameters of coverage area ellipses.

This work studies how the different coverage area estimates affect on positioning accuracy. Simulations show that use of a prior improves positioning accuracy. Especially useful prior information is when there are few location reports, all location reports are arrived from the same location or from the same line. All the cases lead to too small coverage area estimates, if no prior information is used.

Also the consistency of a position estimate is studied. Prior information about the coverage area estimate improves the consistency of the position estimate. However, even though prior information about coverage area is

used, the positioning algorithm has problems with consistency when location reports are not normally distributed in coverage areas. Consistency tests show that although the positioning accuracy is appropriate the usability of the position estimates, however, should be considered. Especially in real world applications, user should get a realistic estimate for the uncertainty of the position.

Algorithms used in this work are not restricted to only a two-dimensional case and it would be interesting to study how positioning algorithm works when also height measurements are available. This could be useful at least in indoor positioning, for instance in apartment buildings, where WLAN access points are on multiple floors. However, height measurements may not be that straightforward and it may be justified to use only discrete height (i.e. floor) measurements.

In this work positioning is done as a static one-shot location solution. Next step could be to try filtering the solution and testing how navigation would work using coverage area -based positioning. Results may not be that good using the cellular network, but at least in urban environment, where the density of WLAN access points is high, algorithms may give an appropriate continuous navigation-grade position solution.

In the positioning algorithm the assumption about the independent measurements could be studied further. Moreover, different kinds of priors could be used. As an example, in indoor positioning a floor map of a building could be used as prior information to calculate the posterior distribution of the positioning estimate.

## References

- [1] V. A. Kaseva, M. Kohvakka, M. Kuorilehto, M. Hännikäinen, and T. D. Hämäläinen, “A wireless sensor network for RF-based indoor localization,” *EURASIP Journal on Advances in Signal Processing*, vol. 2008, 2008.
- [2] H. Marion, “Indoor positioning with a WLAN access point list on a mobile device,” *Proceedings of WSW’2006 at SenSys’2006*, October 2006.
- [3] D. Zimmerman, J. Baumann, A. Layh, F. Landstorter, R. Hoppe, and G. Wlfe, “Database correlation for positioning of mobile terminals in cellular networks using wave propagation models,” *IEEE 60th Vehicular Technology Conference*, September 2004.
- [4] R. S. Campos and L. Lovisolo, “A fast database correlation algorithm for localization of wireless network mobile nodes using coverage prediction and round trip delay,” *IEEE 69th Vehicular Technology Conference*, April 2009.
- [5] B. Li, J. Slater, A. G. Dempster, and C. Rizos, “Indoor positioning techniques based on wireless LAN,” School of Surveying and Spatial Information Systems, Tech. Rep. A 384, September 2006.
- [6] V. Honkavirta, “Location fingerprinting methods in wireless local aarea network,” Master’s thesis, Tampere University of Technology, October 2008. [Online]. Available: [math.tut.fi/posgroup/honkavirta\\_mscth.pdf](http://math.tut.fi/posgroup/honkavirta_mscth.pdf)
- [7] N. Swangmuang and P. Krishnanmurthy, “Location fingerprinting analyses toward efficient indoor positioning,” *Proceedings of the 2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications*, 2008.
- [8] 3GPP TS 43.059 Functional Stage 2 Description of Location Services (LCS) in GERAN. [Online]. Available: <http://www.3gpp.org>
- [9] 3GPP TS 25.305 Stage 2 Functional Description of User Equipment (UE) Positioning in UTRAN. [Online]. Available: <http://www.3gpp.org>
- [10] 3GPP TS 36.305 Stage 2 functional specification of User Equipment (UE) positioning in E-UTRAN. [Online]. Available: <http://www.3gpp.org>
- [11] M. A. Spiritio, S. Pöykkö, and O. Knuuttila, “Experimental performance to estimate the location of legacy handsets in GSM,” *Proceedings of IEEE 54th Vehicular tchnology conference VTC*, October 2001.

- [12] T. Tonteri, “A statistical modeling approach to location estimation,” Master’s thesis, University of Helsinki, May 2001. [Online]. Available: [www.cs.helsinki.fi/u/ttonteri/pub/g.pdf](http://www.cs.helsinki.fi/u/ttonteri/pub/g.pdf)
- [13] M. Hata, “Experimental performance to estimate the location of legacy handsets in GSM,” *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, August 1980.
- [14] J. Syrjärinne, “Studies on modern techniques for personal positioning,” Ph.D. dissertation, Tampere University of Technology, December 2001.
- [15] P. Thevenon, O. Julien, C. Macabiau, D. Serant, L. Ries, S. Corazza, and M. Bousquet, “Positioning principles with a mobile TV system using DVB-SH signals and a single frequency network,” in *Digital Signal Processing, 2009 16th International Conference on*, 2009, pp. 1–8. [Online]. Available: <http://dx.doi.org/10.1109/ICDSP.2009.5201204>
- [16] P. Krishnanmurthy, “Position location in mobile environments,” *Proceedings of NSF Workshop on Context-Aware Mobile Database Management (CAMM)*, 2002.
- [17] L. Wirola, “Studies on location technology standards evolution in wireless networks,” Ph.D. dissertation, Tampere University of Technology, 2010.
- [18] P. Bahl and V. N. Padmanabhan, “RADAR: An in-building RF-based user location and tracking system,” *Proceedings of the 19th International conference on Computer Communications (InfoCom). IEEE*, vol. 2, pp. 775–784, March 2000.
- [19] M. Stella, M. Russo, and D. Begusic, “Location determination in indoor environment based on RSS fingerprinting and artificial neural network,” *ConTel 2007. 9th International Conference on Telecommunications.*, 2007.
- [20] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas, and J. Sievänen, “A probabilistic approach to WLAN user location estimation,” *International Journal of Wireless Information Networks*, vol. 9, no. 3, July 2002.
- [21] M. Youssef and A. Agrawala, “The Horus WLAN location determination system,” *MobiSys '05: Proceedings of the 3rd international conference on Mobile systems, applications, and services*, pp. 205–218, 2005.
- [22] “Ekahau.” [Online]. Available: [www.ekahau.org](http://www.ekahau.org)
- [23] “Polaris wireless.” [Online]. Available: [www.polariswireless.com](http://www.polariswireless.com)

- [24] “Place lab.” [Online]. Available: [www.placelab.org](http://www.placelab.org)
- [25] “Navizon.” [Online]. Available: [www.navizon.com](http://www.navizon.com)
- [26] H. Laitinen, J. Lahteenmaki, and T. Nordstrom, “Database correlation method for GSM location,” *IEEE 51st VTC, Rhodes, Greece*, May 2001.
- [27] “Skyhook.” [Online]. Available: [www.skyhookwireless.com](http://www.skyhookwireless.com)
- [28] L. Wirola, I. Halivaara, and J. Syrjärinne, “Requirements for the next generation standardized location technology protocol for location-based services,” *Journal of Global Positioning Systems*, vol. 7, no. 2, pp. 91–103, 2008.
- [29] L. Wirola, “Generic Fingerprinting in SUPL 2.1,” Nokia Inc., Tech. Rep. OMA-LOC-2008-0303, 2008, presented in OMA LOC WG meeting, 19th-21st May, Wollongong, Australia.
- [30] OMA-WID\_0183-SUPL3\_V1\_0-20090114, Work Item Document, Secure User Plane Location Enabler Release (SUPL) V3.0. [Online]. Available: [www.openmobilealliance.org](http://www.openmobilealliance.org)
- [31] L. Wirola, “LPP OMA extensions draft,” Nokia Inc., Qualcomm, Nokia Siemens Networks, ZTE, Telcordia, LG Electronics, Andrew Corporation, Polaris Wireless, Tech. Rep. OMA-LOC-2009-0266R02, 2009, presented in OMA LOC WG meeting, 21st-23rd October, Los Angeles, USA.
- [32] S. Ahonen and H. Laitinen, “Database correlation method for UMTS location,” *IEEE 57th VTC, Jeju, Rhodes, Greece*, April 2003.
- [33] P. Kemppe, “Database correlation method for multi-system location,” Master’s thesis, Helsinki University of Technology, August 2005. [Online]. Available: [http://www.tml.tkk.fi/Julkaisut/dtyot.php?pub\\_type=1](http://www.tml.tkk.fi/Julkaisut/dtyot.php?pub_type=1)
- [34] T. Lehto, “Ekahau paikantaa USA:n sairaaloissa. Langattomat yritysjärjestelmät Eurooppaa edellä,” *Tietokone*, 31.10.2006.
- [35] A. Karkimo and T. Lehto, “Suomalainen Ekahau sai ison paikannusdiilin. USA:n yliopistolliset sairaalat asiakkaiksi,” *Tietokone*, 7.1.2009.
- [36] S. Kotilainen, “Suomalaisen Ekahaun WLAN-paikannus käyttöön. Kaupan kärryihin lähellä olevien tuotteiden tarjouksia,” *Tietokone*, 22.5.2003.

- [37] “Discussion document: Pattern matching location methods,” Polaris Wireless, Tech. Rep. 3GPP-TSG-RAN-PR-091217, 2009, presented in TSG-RAN Meeting #46, 1st-4th December, Sanya, PRC.
- [38] S. Press, *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*, 2nd ed. Malabar, Florida: Robert E. Krieger publishing company, 1982.
- [39] K. Mardia, J. Kent, and J. Bibby, *Multivariate Analysis*. London, San Diego, New York, Barkley, Boston, Sydney, Tokyo, Toronto: Academic Press, 1979.
- [40] E. Liski, “Matemaattinen tilastotiede,” 2008, University of Tampere, lecture notes. [Online]. Available: <http://mtl.uta.fi/tilasto/mtt-perusteet08/>
- [41] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, Berlin, Heidelberg: Springer-Verlag, 1999, vol. 1.
- [42] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 1964. [Online]. Available: <http://www.math.ucla.edu/~cbm/aands/>
- [43] L. Broemeling, *Bayesian Analysis of Linear Models*. New York and Basel: Marcel Dekker, inc, 1985.
- [44] G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis*. New York, Chichester, Brisbane, Toronto, Singapore: A Wiley-Interscience Publication, John Wiley and Sons, inc, 1973.
- [45] J. Isotalo, S. Puntanen, and G. P. Styan, “Formulas useful for linear regression analysis and related matrix theory,” Department of Mathematics and Statistics, Tech. Rep. A 384, September 2008.
- [46] A. Luoma, “Tilastollinen päättely,” 2009, University of Tampere, lecture notes. [Online]. Available: <http://mtl.uta.fi/tilasto/paattely2/>
- [47] P. H. Garthwaite, I. T. Jolliffe, and B. Jones, *Statistical Inference*, 2nd ed. Oxford, New York: Oxford University Press, 2002.
- [48] A. O’Hagan and J. Forster, *Kendall’s advanced theory of statistics, volume 2B: Bayesian Inference*, 2nd ed. Arnold, London, 2004.
- [49] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York, Berlin, Heidelberg: Springer-Verlag, 2002.

- [50] G. Box and G. Tiao, “A Bayesian approach to some outlier problems,” *Biometrika*, vol. 55, no. 1, pp. 119–129, March 1968.
- [51] K. Chaloner and R. Brant, “A Bayesian approach to outlier detection and residual analysis,” *Biometrika*, vol. 75, no. 4, pp. 651–659, December 1988.
- [52] A. Varbanov, “A Bayesian approach to outlier detection in multivariate normal samples and linear models,” *Communication in Statistics-Theory and Methods*, vol. 27, no. 3, pp. 547–557, 1998.
- [53] H. Sairo, “Error detection in personal satellite navigation,” Ph.D. dissertation, Tampere University of Technology, December 2006.
- [54] 3GPP TS 45.010, GSM/EDGE RAN Radio subsystem synchronization. [Online]. Available: <http://www.3gpp.org>
- [55] L. Koski, R. Piché, V. Kaseva, S. Ali-Löytty, and M. Hännikäinen, “Positioning with coverage area estimates generated from location fingerprints,” submitted to 7th Workshop on Positioning, Navigation and Communication 2010 (WPNC’10).
- [56] N. Bulusu, J. Heidemann, and D. Estrin, “GPS-less low cost outdoor localization for very small devices,” *IEEE Personal Communications Magazine*, vol. 7, no. 5, pp. 28–34, October 2000. [Online]. Available: <http://lecs.cs.ucla.edu/~bulusu/papers/Bulusu00a.html>
- [57] S. Ali-Löytty, N. Sirola, and R. Piché, “Consistency of three Kalman filter extensions in hybrid navigation,” *European Journal of Navigation*, vol. 4, no. 1, February 2006.
- [58] T. Lefebvre, H. Bruyninckx, and J. De Schutter, “Kalman filters for non-linear systems: a comparison of performance,” *International Journal of Control*, vol. 77, no. 7, 2004.
- [59] E. Kaplan and C. Hegarty, *Understanding GPS Principles and Applications*, 3rd ed. Artech house, Inc. Norwood, MA, USA, 2006.
- [60] S. S. Ge and F. L. Lewis, *Autonomous Mobile Robots: Sensing, Control, Decision Making and Applications*. New York, Berlin, Heidelberg: Taylor & Francis Group, CRC Press, Boca Raton, 2006.
- [61] Y. Ueno and K. Maehara, “An elementary proof of the generalized laplace expansion formula for permanents,” *The academic reports, the Faculty of Engineering, Tokyo Institute of Polytechnic*, vol. 30, no. 1, pp. 15–18, 20070000. [Online]. Available: <http://ci.nii.ac.jp/naid/110006631245/en/>

- [62] W. L. Leamer and I. Olkin, “The Jacobians of certain matrix transformations useful in multivariate analysis,” *Biometrika*, vol. 38, no. 3, pp. 345–367, December 1951.



# A Energy consumption with GPS and WLAN

Difference of power consumption between GPS and WLAN is studied. Measurements are done with same Nokia 5800 Music Express mobile phone and program used in measuring power consumption is Nokia Energy Profiler, which is a free program and downloadable from the Internet. Program records power consumption and draws a curve from it.

Figure 11a shows a curve for WLAN scan and Figure 11b shows a curve for GPS. Measurements were done at same office indoors and there was a window at office. Energy consumptions can be calculated using equation

$$E = Pt,$$

where  $E$  is energy in joules,  $P$  is power in watts and  $t$  is time in seconds. In both cases, mobile phone was first in idle state, when no other applications were used. Then scan was made and phone returns in idle state. In both cases there is some leaps from the idle state due to background light of mobile phone and usage needed to turn on systems. As shown in Figures 11a and 11b WLAN scan takes four seconds and GPS fix takes up to 20 seconds. There was about 10 hearable WLAN access point at time measurements were done. Also assistance was on, when GPS fix was done, which reduces time for GPS fix. 20 seconds is quite optimal time for GPS fix and it is difficult to get faster results. Although time needed for GPS fix may increase significantly in different environments. Results for energy consumption are shown in Table 7.

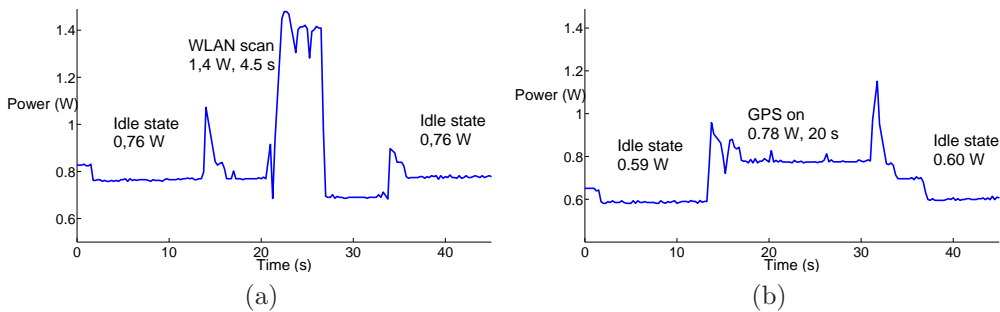


Figure 11: Power consumption with WLAN and GPS

Table 7 shows that GPS requires a bit more energy than WLAN. But as discussed before, time for GPS fix was close to optimal and may increase due to environment, while WLAN scan would not require much more energy in different environments.

Table 7: Times in seconds, powers in watts and energies in Joules required to GPS fix and WLAN scan.

Scan	time [s]	power [W]	Energy [J]
WLAN	4.5	0.64	2.9
GPS	20	0.18	3.6

## B Matrix properties

**Definition B.1.** The determinant of a matrix  $X \in \mathbb{R}^{p \times p}$  is a real valued scalar function of elements of  $X$  and is defined as

$$|X| = \sum (-1)^{|\tau|} x_{1\tau(1)} \times \cdots \times x_{p\tau(p)}, \quad (\text{B.1})$$

where the summation is taken over all permutations  $\tau$  of  $(1, 2, \dots, p)$ , and  $|\tau|$  equals +1 or -1, depending on whether  $\tau$  can be written as the product of an even or odd number of transpositions.

**Definition B.2.** The cofactor of an element  $x_{ij}$  of a matrix  $X \in \mathbb{R}^{p \times p}$  is defined as

$$\text{cof}(x_{ij}) = (-1)^{i+j} X_{ij}, \quad (\text{B.2})$$

where  $X_{ij}$  is a minor of  $x_{ij}$  and is defined to be a determinant of a matrix obtained from  $X$  by deleting the  $i$ th row and  $j$ th column.

**Lemma B.1.** *Determinant of a matrix  $X \in \mathbb{R}^{p \times p}$  can be written using Laplace expansion:*

$$\begin{aligned} |X| &= \sum_{i=1}^p x_{ij} \text{cof}(x_{ij}) \text{ for any fixed } j = 1, \dots, p \\ &= \sum_{j=1}^p x_{ij} \text{cof}(x_{ij}) \text{ for any fixed } i = 1, \dots, p. \end{aligned}$$

*Proof.* See [61]. □

**Definition B.3.** The inverse of a matrix  $X \in \mathbb{R}^{p \times p}$  is the unique matrix  $X^{-1}$  satisfying

$$XX^{-1} = X^{-1}X = I_p. \quad (\text{B.3})$$

The inverse exists if and only if  $X$  is non-singular, that is, if and only if  $|X| \neq 0$

**Lemma B.2.** *If  $X \in \mathbb{R}^{p \times p}$  is invertible, then*

$$X^{-1} = \frac{(\text{cof}(x_{ij}))^T}{|X|}.$$

*Proof.* Suppose that matrix  $X$  is invertible. Then  $|X| \neq 0$  and

$$X(\text{cof } x_{i,j})^T = \begin{pmatrix} \sum_{i=1}^p x_{1i} \text{cof}(x_{1i}) & \sum_{i=1}^p x_{1i} \text{cof}(x_{2i}) & \cdots & \sum_{i=1}^p x_{1i} \text{cof}(x_{pi}) \\ \sum_{i=1}^p x_{2i} \text{cof}(x_{1i}) & \sum_{i=1}^p x_{2i} \text{cof}(x_{2i}) & \cdots & \sum_{i=1}^p x_{2i} \text{cof}(x_{pi}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^p x_{pi} \text{cof}(x_{1i}) & \sum_{i=1}^p x_{pi} \text{cof}(x_{2i}) & \cdots & \sum_{i=1}^p x_{pi} \text{cof}(x_{pi}) \end{pmatrix}. \quad (\text{B.4})$$

According to Lemma B.1, the diagonal values in (B.4) are  $|X|$ . Off diagonal elements are equal to zero [39, p. 457]. So

$$X(\text{cof } x_{ij})^T = |X|I_p, \quad (\text{B.5})$$

and similarly it can be derived that,

$$(\text{cof } x_{ij})^T X = |X|I_p. \quad (\text{B.6})$$

Since  $|X| \neq 0$ , Equations (B.5) and (B.6) can be multiplied on both sides by  $1/|X|$  to obtain

$$X \left( \frac{1}{|X|} (\text{cof } x_{ij})^T \right) = I_p \quad (\text{B.7})$$

or, equivalently

$$\left( \frac{1}{|X|} (\text{cof } x_{ij})^T \right) X = I_p. \quad (\text{B.8})$$

Thus, the matrix  $(1/|X|)(\text{cof}(x_{ij}))^T$  satisfies the definition of an inverse.  $\square$

## B.1 Matrix derivatives

**Definition B.4.** The derivative of a scalar function  $f$  of a matrix  $X = x_{ij}$ ,  $X \in \mathbb{R}^{p \times n}$ , is defined as

$$\frac{\partial f}{\partial X} = \left( \frac{\partial f(X)}{\partial x_{ij}} \right), \quad i = 1, \dots, p, \quad j = 1, \dots, n.$$

**Theorem B.1.** If  $X \in \mathbb{R}^{q \times p}$ ,  $A \in \mathbb{R}^{p \times q}$  and  $Y \in \mathbb{R}^{p \times p}$ ,

$$(i) \quad \frac{\partial \text{tr } AX}{\partial X} = A^T, \quad (\text{B.9})$$

$$(ii) \quad \frac{\partial |Y|}{\partial Y} = |Y|(Y^{-1})^T. \quad (\text{B.10})$$

*Proof.* (i)  $\text{tr } AX$  can be written

$$\text{tr } AX = \sum_{i=1}^q a_{1i}x_{i1} + \dots + \sum_{i=1}^q a_{pi}x_{ip} = \sum_{j=1}^p \sum_{i=1}^q a_{ji}x_{ij}.$$

Hence,

$$\left( \frac{\partial \text{tr } AX}{\partial x_{ij}} \right) = a_{ji}$$

and

$$\frac{\partial \text{tr } AX}{\partial X} = A^T.$$

(ii) Using Laplace expansion of determinant, the determinant of matrix  $Y$  can be written

$$|Y| = \sum_{j=1}^p y_{ij} \operatorname{cof}(y_{ij}) \text{ for any fixed } i = 1, \dots, p.$$

Thus, the derivative becomes

$$\frac{\partial |Y|}{\partial Y} = (\operatorname{cof}(y_{ij})) = |Y|(Y^{-1})^T.$$

□

**Definition B.5.** Derivation of a matrix  $Y \in \mathbb{R}^{p \times p}$  with respect to  $X \in \mathbb{R}^{p \times p}$  is defined as

$$\frac{\partial Y}{\partial X} = \begin{pmatrix} \frac{\partial y_{11}}{\partial x_{11}} & \frac{\partial y_{11}}{\partial x_{21}} & \dots & \frac{\partial y_{11}}{\partial x_{pp}} \\ \frac{\partial y_{21}}{\partial x_{11}} & \frac{\partial y_{21}}{\partial x_{21}} & \dots & \frac{\partial y_{21}}{\partial x_{pp}} \\ \vdots & & & \vdots \\ \frac{\partial y_{pp}}{\partial x_{11}} & \frac{\partial y_{pp}}{\partial x_{21}} & \dots & \frac{\partial y_{pp}}{\partial x_{pp}} \end{pmatrix}.$$

**Theorem B.2.** If  $A \in \mathbb{R}^{q \times p}$ ,  $X \in \mathbb{R}^{p \times r}$ ,  $X \in \mathbb{R}^{r \times s}$ ,

$$\frac{\partial AXB}{\partial X} = B^T \otimes A. \quad (\text{B.11})$$

The result can be derived using direct differentiation presented in Definition B.5.

## B.2 Jacobians of matrix transformations

**Definition B.6.** Suppose that  $X$  and  $Y$  are matrices with  $r$  distinct elements. Then, if  $Y = f(X)$ , the Jacobian of the transformation is defined as

$$J_{X \rightarrow Y} = \|A\|, \quad \text{where } A = \left( \frac{\partial x_i}{\partial y_j} \right), \quad i, j = 1, \dots, r,$$

where  $\|A\|$  means the absolute value of  $|A|$ , and  $(x_1, \dots, x_r)$  and  $(y_1, \dots, y_r)$  denote the distinct values of  $X$  and  $Y$ , respectively.

**Lemma B.3.** [62] If  $X = f(U)$  and  $Y = g(V)$  are transformations from  $U$  and  $V$  to new variables  $X$  and  $Y$ , then

$$J_{(U,V) \rightarrow (X,Y)} = J_{U \rightarrow X} J_{V \rightarrow Y}.$$

**Definition B.7.** If  $f(x_1, \dots, x_n)$  is a differentiable function of  $x_1, \dots, x_n$ , then

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i$$

is called the differential of  $f$ .

**Definition B.8.** If  $A = (a_{ij})$ ,  $A \in \mathbb{R}^{p \times q}$ , where  $a_{ij}$  are functions of  $(x_1, \dots, x_t)$ , then

$$dA = (da_{ij}), \quad i = 1, \dots, p; j = 1, \dots, q,$$

and  $dA$  is called matrix of differentials.

**Lemma B.4.** Let  $A$  and  $B$  be real matrices. Then

$$d(AB) = (dA)B + A(dB).$$

*Proof.* See [62]. □

**Lemma B.5.** Let  $X \in \mathbb{R}^{p \times p}$  be a symmetric matrix and let  $A$  be a nonsingular constant matrix. If  $Y = AXA^T$ ,

$$J_{Y \rightarrow X} = ||A|^{p+1}|$$

*Proof.* See [62] □

**Theorem B.3.** Let  $X, A \in \mathbb{R}^{p \times p}$  and let  $A$  be nonsingular.

(i) Let  $X$  be spd. If  $Y = X^{-1}$ ,

$$J_{X \rightarrow Y} = |Y|^{-p-1}. \quad (\text{B.12})$$

(ii) If  $X = AY$ ,

$$J_{X \rightarrow Y} = ||A|^p|. \quad (\text{B.13})$$

*Proof.* (i) Proof is based on the fact that  $J_{X \rightarrow Y} = J_{dX \rightarrow dY}$ . When  $X$  is nonsingular, it holds that  $XY = I_p$ . Taking differentials on both sides and using Lemma B.4

$$\begin{aligned} d(XY) &= d(X)Y + X dY = 0 \\ \Rightarrow dX &= -X(dY)X. \end{aligned}$$

Now using Lemma B.5 and treating  $Y$  as a constant matrix, the Jacobian becomes  $J_{X \rightarrow Y} = |X|^{p+1} = |Y|^{-p-1}$

(ii) Equation  $X = AY$  can be written in form  $(\mathbf{x}_1 \dots \mathbf{x}_p) = (A\mathbf{y}_1 \dots A\mathbf{y}_p)$ , where  $\mathbf{x}_i$  denotes the  $i$ th column of matrix  $X$ . Thus, the transformation of each column of  $X$  is independent of the others. For the transformation  $\mathbf{x} = A\mathbf{y}$ , each element can be written in form  $x_i = \sum_{k=1}^p a_{ik}y_k$ , and  $\left(\frac{\partial x_i}{\partial y_j}\right) = (a_{ij}) = A$ . Therefore  $J_{\mathbf{x} \rightarrow \mathbf{y}} = ||A||$ . Now using Lemma B.3, the Jacobian becomes  $J_{X \rightarrow Y} = J_{(\mathbf{x}_1, \dots, \mathbf{x}_p) \rightarrow (\mathbf{y}_1, \dots, \mathbf{y}_p)} = J_{\mathbf{x}_1 \rightarrow \mathbf{y}_1} \cdots J_{\mathbf{x}_p \rightarrow \mathbf{y}_p} = ||A|^p|$ . □