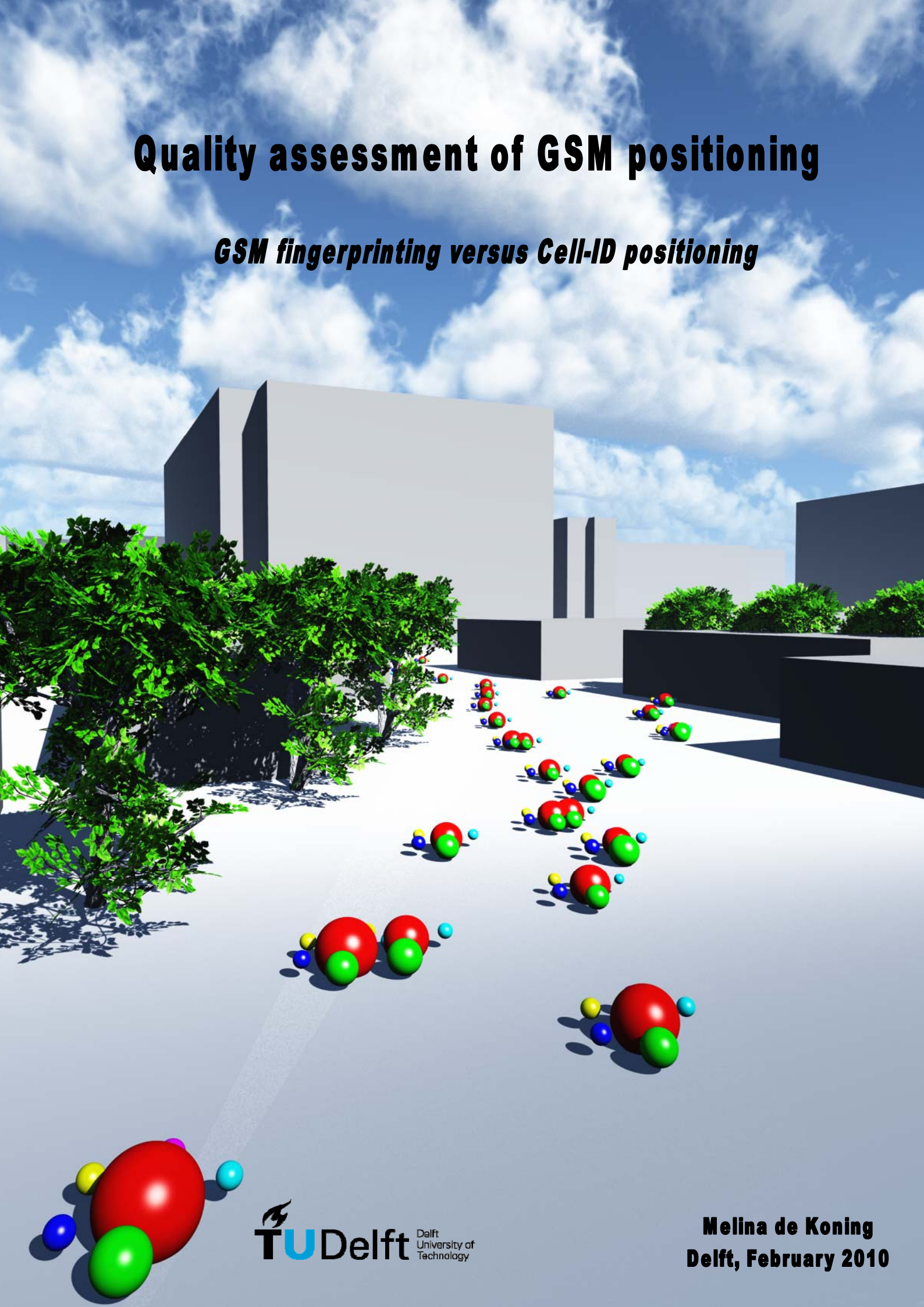
Quality assessment of GSM positioning

GSM fingerprinting versus Cell-ID positioning



TU Delft Delft University of Technology

Melina de Koning

Delft, February 2010

Quality assessment of GSM positioning
GSM fingerprinting versus cellid positioning

by

Melina de Koning

A thesis
Submitted in partial fulfillment of the requirements for the degree

Master of Science
in
Geomatics

Professor: prof. dr. ir. P.J.M. van Oosterom
Supervisor: ir. E. Verbree

Delft University of Technology
Delft, The Netherlands

January 2010

Abstract

Quickly finding a 911 caller can save lives. For the emergency caller it is not always possible to tell the operator where he/she is. If the caller uses a landline, than it is easy to determine the position of the caller, because the landline belongs to an address. But for a mobile phone this is not the case and because of this it is more difficult to locate a mobile caller. This thesis discusses different positioning techniques that do not need any network changes. For this thesis only information that is already available will be used.

There exist different techniques to use the GSM network information for positioning. An easy technique is cellid positioning. This technique makes use of the knowledge of the location of a GMS cell. Measurements are done to determine a point in the cell, which will be used to position the mobile station (MS) to. The position where the same current cellid occurs is averaged to come to the reference point. Our test data is captured in such way that the mobile unit is plotted approximately in the centre of the cell. And because of this the maximum error made is approximately the radius of the cell.

The position of a cell will not change very quickly, because of the complexity and cost to create a GSM network. For this reason one only need to determine ones a position in the cell that will be used to plot the MS to. This makes cellid positioning a robust and easy to implement positioning technique. A drawback of this positioning technique is that it is not possible to determine motion within a cell, because the MS will be plotted to the same point while in the cell.

Another GSM positioning techniques is the database correlation method (DCM) a form of GSM fingerprinting. The set of cellids and corresponding signal strengths received at one position is characteristic for that position and for this reason called a fingerprint. First a reference database is created, this database contains fingerprints and the position where this fingerprint occurs. After the creation of the database the mobile unit can be positioned by comparing the fingerprint of the MS with the ones in the database. The position of the reference fingerprint that is most similar will be the position where the mobile unit is positioned to. Different algorithms are tested to find the most similar fingerprint. More points within a cell are used to position the MS to and for this reason it is possible to detect motion within a cell.

Because of the reason that cellid positioning and DCM give similar results, a new technique is introduced, another form of GSM fingerprinting namely the signal space distribution method (SSDM). The main difference between DCM and SSDM is that DCM does not necessarily uses all information from the fingerprint, only the information that occurs in the fingerprint of the MS and occurs in the reference fingerprint will be used. SSDM always uses all available information in the

fingerprint. With SSDM received signal strength (RSS) distribution maps are created from the reference points in the database for ever cellid. The RSS distribution maps of the cellids that occur in the fingerprint of the MS will be combined to come to the position of the MS. This RSS distribution maps are used to determine the position of the mobile unit.

SSDM makes us of RSS distribution maps and because of this the MS can be positioned everywhere in the cell. This makes it also with this method possible to determine motion within a cell.

With cellid positioning the error made is 211 meters or less for 67% of the measurements. For DCM different algorithms were tested. The algorithm that gave the best result makes an error of 172 meters or less for 67% of the measurements. Looking at SSDM this error made is only 123 meters or less for 67% of the measurements.

Finally we can conclude that SSDM gives the best results. Comparing cellid positioning with SSDM an accuracy improvement of 88 meters can be achieved. Furthermore with SSDM it is possible to detect motion of the MS within a cell. This makes this technique for more application suitable.

Samenvatting

Het snel vinden van iemand in nood die 112 belt kan levens redden. Helaas is het voor de beller niet altijd mogelijk om een duidelijke omschrijving van zijn locatie te geven. Indien iemand van een vaste telefoon belt kan deze snel gelocaliseerd worden, omdat bij een vaste telefoon een adres hoort. Bij gebruik van een mobiele telefoon gaat dit niet en wordt het lastiger om de beller te localiseren. Deze afstudeerscriptie gaat over het plaatsbepalen van de mobiele telefoon gebruikmakend van informatie van het GSM netwerk op een dusdanige wijze dat er geen extra toevoegingen aan het net gedaan hoeven worden. Er wordt alleen gebruik gemaakt van die informatie die reeds in het netwerk aanwezig is.

Verschillende technieken zijn mogelijk om de GSM informatie te gebruiken voor plaatsbepaling. Een simpele techniek is cellid plaatsbepaling. Informatie over waar een gsmcel zich bevindt wordt gebruikt voor de plaatsbepaling. Aan de hand van eerdere metingen wordt een positie in een cel bepaald waar de mobiele telefoon naar toe gepositioneerd zal worden. Onze testdata is dusdanig ingewonnen dat drie cellen volledig zijn gedekt met metingen. Hierdoor zal de mobiele telefoon ongeveer naar het midden van de cel wordt gepositioneerd worden.De maximale fout die dan gemaakt wordt is dan gelijk aan de straal van de cel.

De locatie van een cel zal niet snel veranderen, aangezien het voor een netwerkbeheerder duur en gecompliceerd is om een netwerk op te bouwen. Er hoeft dan ook maar eenmalig een positie in de cel bepaald te worden om cellid plaatsbepaling te kunnen gebruiken. Dit maak cellid plaatsbepaling een robuste en eenvoudige plaatsbepalingstechniek. Een nadeel van deze techniek is dat beweging binnen een cel niet is waar te nemen, aangezien de mobiele telefoon binnen een cel altijd naar hetzelfde punt gepositioneerd wordt.

Een andere techniek die gebruikt wordt voor GSM plaatsbepaling is de database correlation method (DCM) een vorm van GSM fingerprinting. De set van cellids en signaal sterktes die op een bepaald punt worden ontvangen zijn karakteristiek, zoals ook een vingerafdruk karakteristiek is, daarom worden dit fingerprints genoemd (Engels voor vingerafdruk). Eerst wordt een database met referentiepunten aangelegd. Deze database bevat fingerprints en de positie waar deze fingerprint voorkomt. Vervolgens wordt er met een mobiele telefoon weer een fingerprint in het veld waargenomen. Deze fingerprint wordt vergeleken met de fingerprints die reeds in de database staan en de fingerprint die het meest op de net gemeten fingerprint lijkt wordt gekozen als de positie waar de mobiele telefoon op dat moment is. Verschillende algorithmes worden getest om de juiste fingerprint uit de database te halen. Omdat binnen een cel de mobiele

telefoon op meerdere plekken gepositioneerd kan worden is het met deze techniek wel mogelijk om beweging van de mobiele telefoon binnen een cel waar te nemen.

De resultaten van beide technieken vallen tegen en DCM is niet echt significant beter dan cellid plaatsbepaling. Daarom wordt er nog naar een derde nieuwe techniek gekeken, ook een vorm van GSM fingerprinting, de signal strength distribution method (SSDM). Het grote verschil met DCM is dat bij DCM alleen de informatie van de cellids wordt meegenomen die in de gemeten vingerafdruk voorkomt van de mobiele telefoon en ook in de vingerafdruk van de referentie database. Hierdoor wordt in veel gevallen niet alle beschikbare informatie gebruikt. SSDM maakt wel gebruik van alle beschikbare informatie. Voor SSDM worden distributiekaarten van de ontvangen signaal sterkte gemaakt van elke cellid die voorkomt in de referentiedatabase, de RSS distributiekaarten. Voor de plaatsbepaling worden de distributiekaarten gecombineerd van de cellids welke in de vingerafdruk van de mobiele telefoon voorkomen. Door deze combinatie wordt uiteindelijk een plek bepaald waar de mobiele telefoon gepostioneerd wordt.

Omdat SSDM gebruik maakt van distributiekaarten kan de mobiele telefoon op elke plek in de cel gepositioneerd worden. Ook als hier geen waarneming is gedaan. Hierdoor is het ook met deze techniek mogelijk om beweging binnen een cel waar te nemen.

Voor cellid plaatsbepaling geldt dat 67% van de metingen maakt een fout van 211 meter of minder. Voor DCM zijn verschillende algoritmes getest om tot een positie te komen. Het algoritme dat de beste resultaten oplevert, maakt een fout van 172 meter of minder voor 67% van de waarnemingen. Terwijl als we kijken naar SSDM de fout gemaakt voor 67% van de waarnemingen is 123 meter of minder.

Uiteindelijk kan geconcludeerd worden dat SSDM de beste resultaten geeft. Vergeleken met cellid plaatsbepaling wordt er een nauwkeurigheidsverbetering gehaald van 88 meter. Tevens is het met SSDM mogelijk om beweging binnen een cel waar te nemen, waardoor deze techniek voor meer toepassingen geschikt is.

Acknowledgements

This thesis is the final step in getting my degree in Geomatics. The Geomatics master is done at Delft University of Technology and this masters research at the GIS technology department. The research is carried out in cooperation with the company WebIntegration in Rotterdam. Lots of people have contributed to it and I would like to thank them for their help.

Prof. dr. ir. P.J.M. van Oosterom for his interest and enthusiasm.

My supervisor ir. E. Verbree for his endless patience, his trust and his motivation; specifically for the outstanding technical support with respect to the signal strength distribution method.

Dr. ir. F. Kleijer for his programming with Matlab and his help starting this thesis project.

Dr. ir. G.J.M. Janssen (Delft University of Technology) for taking the time to share his knowledge about GSM networks with me.

E.L. Poppe for his creativity and cartographic ideas.

A. Damhuis for reading and giving comments on my thesis.

All colleagues of WebIntegration for their assistance with the hard- and software. Especially when I spilled my tea over my laptop and I was convinced that I had lost everything. But luckily you could save my laptop and recover my files.

My family and friends for their love and support. The trajectory to this thesis was a long and hard road, but you all kept believing in me. Especially my husband, A. de Boer, for his help, his never-ending believe and love. The figure on front of this thesis is made by him, thank you!

Mom and dad, you were my prop and stay during my complete time at University. Thanks for all you have done for me.

Table of Contents

List of abbreviations

3GPP	third Generation Partnership Project
A-GPS	Assisted GPS
AVC	Authorization and Validation Centre
BD	Billing Database
BSC	Base Station Controller
BSS	Base Station Subsystem
BTS	Base Transceiver Station
CDF	Cumulative Distribution Function
DCM	Database Correlation Method
EIR	Equipment Identity Register
ETSI	European Telecommunications Standards Institute
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSM	Global System for Mobile communications
HDOP	Horizontal Dilution Of Precision
HLR	Home Location Register
IMEI	International Mobile Equipment Identity
IMSI	International Mobile Subscriber Identity
KNN	K Nearest Neighbour
LAC	Local Area Code
MCC	Mobile Country Code
MNC	Mobile Network Code
MS	Mobile Station
MSC	Mobile Switching Centre
NN	Nearest Neighbour
NSS	Network Switching System
POLS	Privacy Observant Location Software
PSTN	Public Switched Telephone Network
RD	RijksDriehoekstelsel
RF	Radio Frequency
RFID	Radio Frequency IDentification
RSS	Received Signal Strength

SIM	Subscriber Identity Module
SSDM	Signal Strength Distribution Method
UMTS	Universal Mobile Telecommunications System
VLR	Visitor Location Register
WDI	Weighted Distance Inverse
WGS84	World Geodetic System 1984
WKNN	Weighted K Nearest Neighbour

List of terms

Accuracy is the combination of reliability and precision (Polman en Salzmann 1996).

Broadcast channels are used to transfer system information such as timing references and synchronization information. The broadcast provides system information, system configuration information, and lists of neighbouring radio channels to all mobile devices operating within its radio coverage area (Harte, 2005).

A *cell* refers to the geographic coverage area of a base station where one or more radio frequencies are present.

The *cellid* is the unique identification number of a cell.

Cellid positioning is a method in which only the cellid of the current cell is used to determine the position of the mobile phone.

The *current cell* is the cell that belongs to the base station to which the mobile phone is connected to during data transmission, a.k.a. *serving cell* or *active cell*.

A *fingerprint* is a list of position dependent network characteristics, namely cellids and signal strengths of up to seven distinct cells.

Frequency allocation is the amount of frequency bands that is assigned by a regulatory agency for use for specific types of radio services.

GSM fingerprinting is a positioning technique that matches a unique combination of GSM network characteristics, the fingerprint, in the field with the reference fingerprints in the database to determine the position of the mobile phone in the field. In this thesis the fingerprint consists of cell ids and signal strengths of up to seven distinct cells.

Handover is the process of changing from one base station to another base station, which implies that the current cell is changed.

A *location* is a description of a point in place, for example the point is in the shopping mall or is in Kingstreet.

A *position* is the x and y coordinates of a point in place.

Precision is the closeness to the true position.

Reliability the sensitivity of the measurements for mistakes in the model (Polman en Salzmann 1996).

A *signal* is defined as a function of one or more variables, which conveys information on the nature of a physical phenomenon (Haykin, 1999).

The *spatial median* is defined as the bivariate position measure to minimize the sum of absolute distances to observations (Gower, 1974).

A *system* (or in this case network) is defined as an entity that manipulates one or more signals to accomplish a function, thereby yielding new signals (Haykin, 1999).

A *wide fingerprint* is a fingerprint consisting of GSM network characteristics of more than seven distinct cells. These fingerprints cannot be measured with a consumer mobile phone and is for this reason not relevant for this research.

1. Introduction

"Fulton county officials say a woman died while waiting for help that was delayed because a 911 operator sent emergency responders to the wrong address. … Fulton County's 911 director, Alfred Moore, says it began Saturday afternoon when Darlene Dukes called 911 for help because she was 'in respiratory distress.' … Moore says the 911 operator misheard the address Dukes gave and sent crews to Wells Street in Atlanta when Dukes was actually at home on Wales Street in Johns Creek, north of Atlanta. … (and) the operator should have noticed that the call was coming from a cell tower in north Fulton County, not Atlanta. The mistake caused a 25-minute delay in response." (CBS news, August 6, 2008)

Knowing where an emergency is, is vital information for emergency responders. Unfortunately as the example above shows this information can be easily misheard by the 911 operator due to for example a bad telephone connection, different linguistic accents or subdued voices. In the example is mentioned that the operator could have seen from which base transceiver station the call was coming from, but even when the operator had looked at this it could give the wrong impression. A base transceiver station can handle a maximum number of calls when this maximum is reached the call is redirected to another base transceiver station. GSM positioning techniques can overcome these problems.

Most positioning techniques using the GSM network need expensive network changes or have a very limited accuracy. In this thesis different positioning techniques are evaluated that do not require any network changes. The most simple technique is cellid positioning. With cellid positioning the mobile unit is plotted to a point in the cell where the mobile unit is connected to. Another technique examined is the database correlation method (DCM), a form of GSM fingerprinting. The basis of this technique is a database with reference points that is used as a kind of look-up table. The position of the reference points are stored together with some GSM network information. Later in the field these network information is measured again and matched with the already stored data in the look-up table.

Earlier research is done at cellid positioning and DCM, which will be explained in chapter 3. New in this research is that for DCM the number of matching neighbours is taken into account. And both techniques are not earlier research on a network in the Netherlands. Because both techniques give not satisfying results a new technique is introduced, the signal strength distribution method (SSDM). This method is also a form of GSM fingerprinting, because it makes

use of the characteristics of the combination of cellids and signal strengths, the fingerprint. From the fingerprints in the reference database received signal strength (RSS) distribution rasters are created. These rasters are used to find the position of the mobile unit.

Knowing where a mobile caller is located can be used for more than only public safety. An example is for social networking, like a friendfinder application. This thesis is done in cooperation with the company WebIntegration. This company has a friendfinder application for which they are looking for a better GSM positioning technique. The application is LiveContacts (www.livecontacts.com sd), it works like msn messenger. People using this application can chat with each other. Additional to this chatting one can also see where their friends are located. Some new phones like the Iphone from Apple have a friendfinder application standard installed on it.

Other applications can be tracking, billing or advertising. Tracking can be the tracking of personal assets. For example a laptop with a GSM chip integrated in it. But also tracking of persons, i.e. a parent that gets an alert when their child leaves the playground. Billing one can think of road tolling or parking. The last mentioned possible application, advertising, for example someone gets a message when walking nearby a store with the promotions that are in that store at the moment. Besides these examples, many more are possible.

Some new mobile phones are standard equipped with GPS. The positioning techniques discussed in this thesis can be used complementary to GPS. Because GPS does not work well indoors and does not work well urban environments, because of high buildings that block the GPS signal, a.k.a. the urban canyon problem.

1.1 Research question

The challenge in this research is to find a GSM positioning technique that does not require any knowledge of the network infrastructure. The main advantage of such positioning technique is that it works independently from the network operator. A simple technique for doing this is to define one point in every cell and to position the mobile unit to that point.

A more complex method is using more information that can be received by the mobile unit. The method chosen for this is GSM fingerprinting. This technique is tested before in other countries, mainly Scandinavian countries, with good results. Most of these researches were done in indoor environments. We selected a heavily urbanized area in Delft (district Voorhof) as our study area, because we expect that GSM positioning is particularly an alternative for GPS positioning in urban areas (i.e. GPS fix is disturbed by high buildings, a.k.a. the urban canyon problem).

All methods are tested in a real world scenario. Knowing all this the research question for this thesis becomes:

How much accuracy improvement can be achieved with GSM fingerprinting compared to GSM cell-id positioning in a real world scenario?

For answering this research question the following sub questions are used:

1. How does the GSM network work?
2. What are the characteristics of signal strength?
3. How does cell-id positioning work? Both in theory and in practice.
4. How does GSM fingerprinting work? Both in theory and in practice.
5. What is the accuracy improvement (if any) of GSM fingerprinting compared to cell-id positioning?

1.2 Context

This research is conducted in cooperation with WebIntegration bv. This company has developed a friendfinder application for the consumer market. With this application, LiveContacts, it is possible to chat with your friends using your mobile phone. Besides chatting it is also possible to see where your friends are. Therefore this application needs a positioning technique, WebIntegration already used a positioning technique, namely cellid positioning. The software sends data from the users of Livecontacts who use a mobile phone with GPS to a database of WebIntegration. This data contains coordinates and the base transceiver station where the mobile phone is connected to. With this information a point in the cell is calculated and later used for the users that do not have a mobile phone with GPS. This research focuses on how this positioning technique can be improved.

While working for an independent company it is not possible to make use of knowledge about the position of the base transceiver stations owned by the network operator. For developing a positioning technique this has to be kept in mind. General knowledge about the GSM network is available. All networks have the same structure, this structure can be used in the positioning technique.

For this research measurements are taken in the field. The field test is conducted in Delft, because of its urban character. It is expected that GSM fingerprinting performs well in a urban environment, because of the small cell sizes. It was planned to do a similar test in a rural environment, but unfortunately the hard- and software let us down. The measurements are done using the Privacy-Observant Location System (POLS) application (http://pols.sourceforge.net/ sd). This application measures signal strength and cellids of the current cell and of up to six neighbouring cells. All measurements are done on the same network, in this case the network of Orange, nowadays this network does not exist anymore and became part of the T-Mobile network. Orange was chosen because the mobile phone used had a simlock, which means that the phone only operates on the Orange network. The choice of network is not relevant for this thesis, because all GSM networks in the Netherlands have the same structure.

The field test consists of cycling through a test area to collect test data. Later these data are used for testing cellid positioning and the different GSM fingerprinting techniques. The mathematical computations are done using Mathworks Matlab. The diagrams in this thesis are a result of this. The cartographic representations and the raster calculations are done in ESRI's ArcGIS 9.3.

1.3 Structure of the thesis

Chapter 2 GSM network: in this chapter information is given about the GSM network that is relevant for this thesis. Special attention will be given to the signal strength. Subquestions 1 and 2 will be answered.

Chapter 3 Related work: other research carried out in the field of mobile positioning is treated, especially research that is relevant for this thesis.

Chapter 4 Collection of sample data: for answering subquestion 3 and 4 test data has to be collected, in this chapter is explained how this is done.

Chapter 5 GSM Cellid positioning: the theory and practice of cellid positioning will be explained. This answers subquestion 3.

Chapter 6 Database correlation method: a form of GSM fingerprinting will be introduced, the database correlation method (DCM). The theory and practice of DCM will be explained. This answers subquestion 4.

Chapter 7 Signal strength distribution method: using the data collected in chapter 4 a new GSM fingerprinting technique is introduced, the signal strength distribution method (SSDM). This technique makes use of RSS distribution rasters to determine the position of the mobile unit.

Chapter 8 Conclusions: the answer on the research question is given in this chapter. This answers subquestion 5 and the research question of this thesis.

2. GSM network

The GSM network is a very complex system. Only a few parts of the system are relevant for this thesis. In this chapter an overview of these relevant parts of the GSM network are given. In section 2.1 a general introduction is given. Understanding of the structure of the network is needed to understand the different positioning techniques; in section 2.2 this is explained. Some GSM positioning techniques make use of the received signal strength (RSS), in section 2.3 the characteristics of the RSS is given.

2.1 Introduction

The first mobile phones, introduced in the late 1970s, were the so called car phones. This phone was a two-way radio with a connection to the landline telephone system. Because of their popularity and very few frequencies available callers encountered long waiting times before able to make a call. This problem was solved by introducing a cellular system. From this time on almost every country used to have its own cellular system for mobile communication.

Mobile phones were only able to work on one of the many different systems, so roaming to a neighbouring country was impossible. To overcome this problem a European standard was created. The development of this standard started in 1982 in Stockholm by the Groupe Spécial Mobile (GSM). It is initially created to provide a single industry standard for European cellular systems. At that time the European Telecommunications Standards Institute (ETSI) became the managing body. The first commercial system came available in 1991. With the introduction of the commercial network the acronym GSM changes to Global System for Mobile communications. The first non-European country that implements GSM is Australia and did this in 1993. Now, in 2010, 219 countries (www.GSMWorld.com sd) worldwide are using the GSM standard.

In 1998 the third generation partnership project (3GPP) group was formed to create the next evolution of mobile specification, Universal Mobile Telecommunications System (UMTS). The main difference between GSM and UMTS is that UMTS has faster data transfer rates, because of different use of the available channels. UMTS is also a cellular network and probably the proposed positioning techniques in this thesis can also be applied. In this research we limit ourselves to the GSM network. The 3GPP has now taken over the management of GSM specifications. GSM specifications (and evolved versions of the specification) can be obtained at www.3GPP.org.

GSM operates in the 900 MHz and 1800 MHz frequency bands, these are licensed bands. Advantage of this is that less interference occurs. The 900 MHz band is divided in two 25 MHz

bands, namely 890 – 915 MHz for uplink communication and 935 – 960 MHz for downlink communication. For the 1800 MHz band this is 1710 – 1785 for uplink communication and 1785 – 1880 for downlink communication. Uplink means that the mobile phone transmits and for downlink the transmission is done by the base transceiver station, which means that the mobile unit receives. The uplink and downlink bands are further more divided in channels with a frequency bandwidth of 2500 kHz. This results in 500 frequencies in total, 125 for the 900 MHz band and 375 for the 1800 MHz band. The different channels are assigned to the different operators in a country. In this thesis the Orange network in the Netherlands is used. They operate on 880,2-880,8, 882,4-886,4 and 1740,2-1755,0 for uplink communication and 925,2-925,8, 927,4-931,4 and 1835,2-1850,0 for downlink communication. The different frequencies are reused within the network. This frequency reuse is done in such way that the frequencies cause no interference with each other. The purpose of frequency reuse is to increase the number of customers that can be served.

2.2 Network structure

The GSM network consists of three subsystems, the mobile station (MS), the base station subsystem (BSS) and the network switching system (NSS), a schematic overview of the network is given in Figure 1. The BSS controls the communication between the MS and the NSS. The NSS controls the communication between different mobile users, but also the communication between a mobile user and a fixed network user. The MS consists of two parts, the mobile equipment, which can uniquely be identified with the international mobile equipment identity (IMEI) number and the subscriber identity module (SIM) card, which can be uniquely identified with the international mobile subscriber identity (IMSI) number.

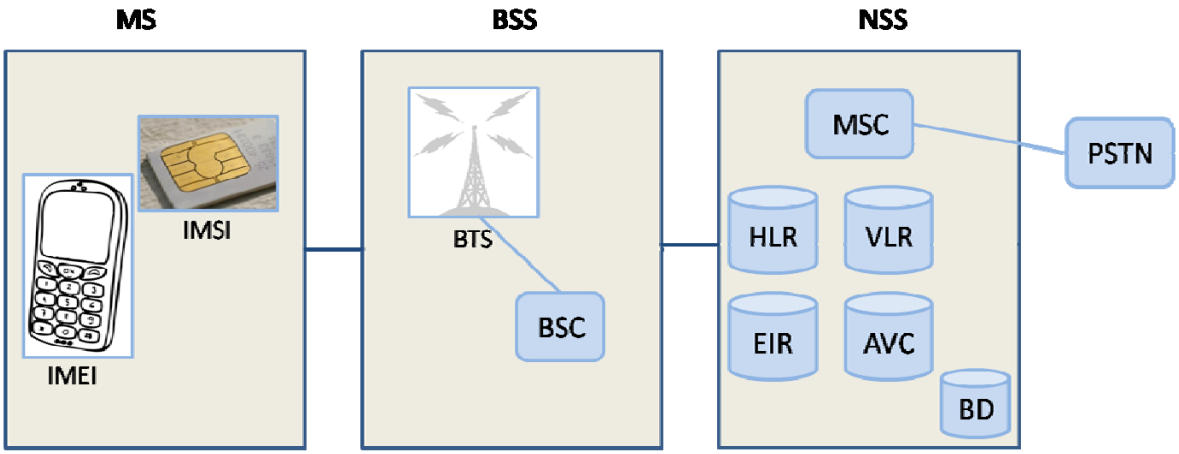


Figure 1 overview of the structure of a GSM network with the three different subsystems, namely the mobile station (MS), the base station subsystem (BSS) and the network switching system (NSS).

One of the components of the BSS is the base transceiver station (BTS). BTS are stand alone transmission systems that define a cell site and is composed of an antenna system, building, and base station radio equipment. The BTS is connected to a base station controller (BSC) that coordinates the radio channel assignments and thus also handles with handover procedures.

The main parts of the NSS are the mobile switching centres (MSC) and the different databases. The MSC processes requests for service from mobile devices and land line callers, and routes calls between the BTS and the public switched telephone network (PSTN). The key network databases include a master subscriber database (home location register, HLR), temporary active user subscriber database (visitor location register, VLR), unauthorized or suspect user database

(equipment identity register, EIR), billing database (BD), and authorization and validation centre (AVC).

The GSM network is a cellular network. This means that the total area covered is subdivided in different cells. Different terms are used for cells of different sizes. The term macrocell is used for cells with a range of 10 – 35 km, these cells occur mostly in rural areas. For urban areas the microcell is used and has a range of 200 m up to 3 – 5 km.

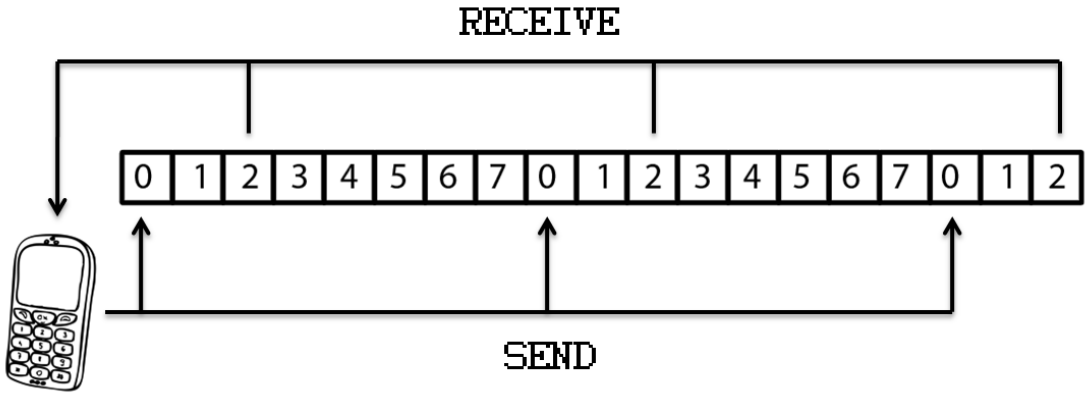


Figure 2 the mobile phone sends in timeslot 0 and receives in timeslot 2. The other timeslots are used to listen to other BTS.

The smallest division of the channel is the timeslot. The duration of a timeslot is 577 microseconds and they are numbered from 0 to 7. During GSM communication the MS transmits in one timeslot, receives on one timeslot and has 6 idle timeslots available, see Figure 2. During these idle time periods the MS tunes to other frequencies and measures their signal strength. In this way a list is generated with receiving neighbouring cellids and their received signal strength. This list is used when a handover is needed. During communication the BTS tries to keep the MS as long as possible connected. When a threshold value is reached than the MS will switch to a different BTS. This procedure is called mobile assisted handover (MAHO). By this handovers are minimized and also the risk for losing the connection.

Actual coverage of a cell depends on three factors, height of the antenna at the BTS, type of antenna and the power level emitted. The number of costumers that can be served by one cell is limited. In urban areas cell sizes will be smaller than in rural areas, because in urban areas there are more potential callers.

2.3 Signal strength

Signal strength is measured in dBm, this is decibel with reference scale milliwatt. In this way it is an absolute unit not to confuse with dB which is a dimensionless unit. For GSM it ranges from -48 dBm to -110 dBm, with -48 dBm as strongest value. The minus sign indicates that the signal strength is decreasing. Another way to express signal strength is the received signal level (RxLev). Relationship between the RxLev and dBm values reads:

$$RxLev = dBm + 110 .$$

Therefore RxLev values are never negative and range between 0 and 62.

Two types of channels exist, traffic channels and control channels. Traffic channels are used for communication. On these traffic channels radio frequency (RF) power control is applied. RF power control objective is to control the transmission power level of the mobile phone during calls, to fulfil the required uplink power level and quality, while keeping transmission power as small as possible. Because of this power control the signal strength is not constant. The strength depends on the number of callers in the current cell and neighbouring cells and the distance of the mobile

terminal to the transmitter. Power control is not used on the control channels. For this reason the signal strength on the control channels are relatively constant in time. The control channel used in this research is the broadcast control channel (BCCH). The BCCH transmits its information only on the first timeslot, in the other timeslots the BTS sends a dummy burst. The MS can listen at any time to different BCCH carriers and decide, based on signal strength, whether a handover is needed.

Signal strength variation within a cell is caused by distant dependent fading. There are three types of fading that occur within a cellular network namely absorption, free-space loss and multipath fading (a.k.a. Rayleigh fading). Free-space loss is the attenuation of the signal of a given distance. It is also said that it is the path length of the signal. Objects, like buildings and trees, absorb part of the signal, this process is called absorption. This can be compensated for by using higher-gain antennas and higher power levels, by doing this the same geographical area will be covered than if there was no absorption. Compensation can also be done by creating more and smaller cells.

Signals arriving at the MS are direct and indirect signals. Indirect signals are signals that are reflected from objects before it arrives at the MS, this is called multipath or Rayleigh fading. Direct signals arrive at the MS without reflecting from an object, this happens when the MS is in direct line of sight with the BTS. Because of this signals arrive at the MS out of phase, this result in either that signal cancel each other out or they supplement each other.

Signal strengths of the control channel are relatively constant in place and time, as can be seen in Figure 3. This figure is created using a mobile phone with POLS running on it. This software takes every second a measurement of the signal strength on the control channel. The phone is left at one place for 24 hours. Figure 3 is only a part of these measurements, namely one hour. POLS rounds the measurements off to whole values.

Signal strength only changes when major changes in the environment occur, for example demolishing of a big building. Another characteristic of the signal strength is that it varies in space, because of the previous mentioned fading. This means that a base transceiver station heard at a position may be heard stronger or not at all a few meters away.

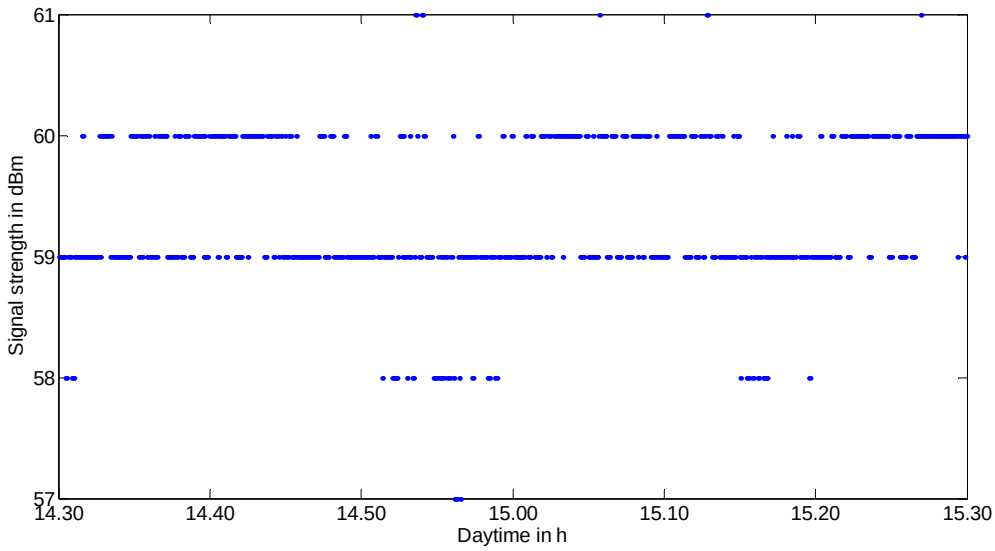


Figure 3 signal strength of the current cell at one position in Delft.

3. Related work

There are different techniques for finding a mobile user. This chapter gives an overview of relevant previous research done in this field. Starting with techniques using radio frequencies (RF) in chapter 3.1. In 3.2 systems using WiFi signals are discussed and finally in 3.3 techniques using the GSM signal.

There are techniques that make use of knowledge of the infrastructure of the network, the so called network based technologies. The goal of our research is to find a positioning technique that does not make use of such knowledge. For this reason the network based technologies are not discussed for more information on these technologies the reader is referred to (Karimi 2004).

3.1 Radio Frequency systems

The Active Badge system (Want 1992) is one of the first systems able to locate mobile users; in this case the mobile users are persons wearing the Active Badge, a so called tag. The system is designed for the receptionist of an office building and as a replacement of the pager. The receptionist is able to see where a person is and which telephone is located nearest to this person. Using this information he/she is able to direct phone calls to the nearest phone by the person for whom the call is. The principle of the Active Badge system is that the badge sends a unique signal to a beacon every 15 seconds, the network processes this signal to a location where that person is. This location is stored and displayed at the monitor of the receptionist.

A second system using radio frequencies is RADAR. The main difference between RADAR and Active Badge is that the beacons within the RADAR system sends signals while the beacons of the Active Badge system only receive signals and the tag of the Active Badge only sends a signal while the tag of the RADAR system can send and receive signals. The advantage of this is that RADAR can make use of the received signal strength. RADAR records and processes signal strength information of multiple base stations, and in this way it provides a positioning and tracking system for indoor environments. Different processing techniques are tested, the best results were found with the empirical method. The median resolution of the RADAR system is 2 to 3 meters (Bahl en Padmanabahn 2000).

Systems, like Active Badge and RADAR, that use a tag that can send and/or receive RFs for identification and tracking of goods, persons or animals is called radio frequency identification (RFID). A good example of RFID that still is in use is the Dutch police who use this system for tracking and tracing evidence collected at crime scenes. The main disadvantage of RFID systems is

that a complete infrastructure is needed within the building. This can be very expensive to build and maintain. Another drawback of the system is that it can only be used within a predefined area, namely the area where the different beacons are located.

3.2 WiFi systems

As is said in the previous part positioning using RF is expensive because of the need of radio beacons through the positioning area. With the introduction of the Wireless Local Area Networks (WLAN) 'beacons' sending a signal become very common around the world. The signal that is used in todays WLAN is called WiFi.

WiFi can also be used as a RFID system, in that way that the tags used comply with the IEEE 802.11 standard (http://www.ieee802.org/ sd). A big advantage of this is that WiFi access points are widely available nowadays, because of the inexpensive development costs of WiFi networks. Recent research done by (Moen 2007) shows that WiFi RFID can also be used in outdoor environments. An accuracy of 50 meters is found when using multiple access points, using only one access point the accuracy found was approximately 100 meters. The actual location determination is performed by the hardware using a combination received signal strengths.

WiFi fingerprinting is another possibility than WiFi RFID to use the WiFi signal for positioning. This WiFi fingerprinting is done by (Li 2005). For fingerprinting two steps are needed; the training phase in which a database is created containing measurements at some reference points and the positioning phase in which the location of the mobile unit can be identified by comparing its signal strength measurements with the reference data. For this last phase three different algorithms are used, nearest neighbour (NN), k-nearest neighbour (KNN) and weighted k-nearest neighbour (WKNN). With NN the reference point with the closest distance is signal space is chosen from the database to be the position where the mobile unit is. KNN makes use of *k* different neighbours to determine the position of the mobile unit. WKNN also uses k nearest neighbours but uses the distance in signal space as a weight to calculate the position of the mobile unit. These three algorithms are also used in these thesis to find the mobile unit in the GSM network and therefore the algorithms are explained in more detail in chapter 5. Li concluded that the best results are found using the KNN and WKNN algorithms with *k* equal to 3 or 4. These results indicates that only using two nearest neighbours is not enough, but too many nearest neighbours could decrease the accuracy of the estimator since some of the nearest neighbours are too far from the estimated points.

If a denser database can be generated efficiently by interpolation based on a small number of reference points, labour effort and time can be saved during the training phase. Two methods, weighted distance inverse (WDI) and kriging, are chosen here to easily fill the database without extra measurements. It is shown by (Li 2005) that WDI cannot successfully estimate the received signal strength from the known information and thus is not a good interpolation technique for this purpose. Kriging on the other hand gives very good results.

There are some major drawbacks of using WiFi for locating mobile users, namely

1. The received signal strength varies significantly. The received signal strength depends on the orientation of the antenna of the receiver and of movement of people nearby the receiver.
2. WiFi works on the 2.4 GHz frequency bands, other devices also work on this band, for example the microwave oven or Bluetooth devices. Because of this the signal suffers from interferences of these devices.
3. WiFi networks are easy and relative cheap to build, because of this it is more likely that the network will change often or completely vanish.
4. WiFi networks are small networks.

3.3 GSM systems

GSM networks are large networks and not likely to change very often, because it is expensive and complicated to build. Furthermore GSM works on a licensed frequency band which reduces interferences of other devices and as will be shown in chapter 3, the received signal strength is relatively stable. The structure of WiFi networks and GSM networks are very similar, this makes it reasonable to believe that the techniques developed for WiFi networks will also work for GSM networks.

(Laitinen, Lähteenmäki en Nordström 2001)(Khalaf-Allah en Kyamakya 2006)(Takenga, Peng en Kyamaka 2007)(Otsason, et al. 2005) and (Varshavsky, Chen, et al. 2006) use the fingerprinting method in combination with the GSM signal. Laitinen used $(k) = \sum_i (f_i - g_i(k))^2 + p(k)$, where f_i is the signal strength of the request fingerprint, $g_i(k)$ is the corresponding signal strength of the kth database fingerprint, to calculate the position of the mobile user and $p(k)$ is a penalty term for cellids that do occur in one of the fingerprints but not in the other.. The accuracy found is 90 meters or less in 90% of the cases and was 44 meters or less for 67% in urban areas, for the suburban area this is 190 meters and 74 meters for 90% and 67% respectively.

(Anderson en Muller 2006) used the GSM signal to distinguish different states of movement, like walking, running and driving. It is found that it is possible to distinguish between these different states when looking at the change of cell information. When the list of neighbouring cells varies rapidly the most likely movement is driving, when it does not change the mobile user stands still.

(Varshavsky, Chen, et al. 2006) uses two types of positioning, namely centroid and fingerprinting. Both need a training phase. The first method, centroid, predicts the position of the base station in the training phase. During the retrieval phase the position is calculated by averaging the positions of the base stations that appear in the measurement. Fingerprinting matches every measurement in the testing set to one or more measurements observed during the training phase and then averages the true positions of the best matched measurements. Results are that with fingerprinting that it is possible to achieve 70 – 200 meters median error outdoors, and by observing the Euclidean distance in signal strength for two consecutive measurements it is possible to determine whether a mobile is moving or not (this for place detection purposes).

SkyLoc (Varshavsky, LaMarca, et al. 2007) determines the floor on which a user is located using GSM signal strength fingerprinting. The system identifies the correct floor in 73% of the cases and is within 2 floors correct in 97% of the cases. To find the correct floor the following two steps are performed, first calculate the Euclidean distance between the current fingerprint and all fingerprints in the database, second the fingerprint with the smallest Euclidean distance predicts the correct floor. They claim that the system is robust, because the training and testing sets are collected with different hardware and one month apart. With different hardware it is meant that different phones are used, but the type of the phone stayed the same (Varshavsky, LaMarca, et al. 2007).

The main goal of the research of (DeBlauwe 2008) is to examine the possibilities of GSM-based positioning for location based services. One of the techniques used is cellid positioning. His definition of cell id positioning is, how the maximum of information can be taken out of the knowledge that the mobile phone is located in cell *i* so that the position of the mobile phone can be estimated with the highest possible accuracy. DeBlauwe states that cell-id positioning is not an all-purpose positioning method, which means that for some applications cell-id positioning can be used, but for most applications the accuracy of this positioning method is too low. For this reason he also examines other techniques which use received signal strength (RSS). These techniques make also use of knowledge about the GSM network, for example the angle of arrival, because of this these other techniques are left out here. The interested reader is referred to his thesis for further information (DeBlauwe, 2008).

4. Collection of sample data

The next chapters deal with the actual positioning techniques. Before these techniques can be examined test data has to be collected. This chapter deals with the collection of this test data. A short description of the test area is given in chapter 4.1. Details about the used hardware en software are given in 4.2 and 4.3 respectively. In chapter 4.4 it is explained how the test data is collected.

4.1 Study area

A field test is conducted to test the methods that will be proposed in chapter 5, 6 and 7. Delft, in the western part of the Netherlands, is chosen as study area, because of its urban character. In urban areas cell sizes are small, because of the high number of potential callers. It is expected that cellid positioning and GSM fingerprinting both are positively influenced by this. Testing is done in the area named Voorhof, located in the south west from the city centre of Delft. Voorhof is chosen because it is one of the most densely populated areas of the Netherlands. Figure 4 gives an impression of the testing area.

Figure 4 Visual impression of Voorhof in Delft.

4.2 Hardware

The hardware used for the field test is a SPV C500 mobile phone, an Orange simcard and an AD-300 Adapt Bluetooth GPS device, see Figure 5. The software used to acquire the GSM data only works on a limited number of mobile phones. For this reason the SPV C500 was chosen. The mobile was simlocked by Orange, which means that the phone can only be used with a Orange simcard. The choice of the network is not relevant for this research. All network operators have a

GSM network with the same structure. The only difference in network structure that can be relevant for this thesis is the use of a different antenna. All network operators in the Netherlands make use of directed antennas. Thus the choice of network is not very important.

The GPS device is needed for two reasons. The first reason is that the coordinates collected by the GPS is taken as true position. The results of the different GSM positioning techniques are compared to the GPS position to objectively verify the results. The second reason is that information of where the reference data is collected is needed for the used GSM positioning techniques.

Figure 5 hardware used for the field test; a mobile phone SPV C500, Orange simcard and Adapt Bluetooth GPS device.

4.3 Software

On the telephone privacy observant location software (POLS) is installed made by Intel Research in Seattle together with the University of Washington (http://pols.sourceforge.net/ sd). This software makes it possible to log GPS and GSM data on a mobile phone at the same time.

Table 1 gives an overview of how the data output of POLS looks like. The first row gives the data collected from the Bluetooth GPS device as is indicated by the first part of the row (TYPE=GPS). After this the time of the measurement is given in epoch, this means the seconds since January 1, 1970 and the time that the Bluetooth GPS device had its first fix. NUMSAT gives the number of satellites that are used for positioning and HDOP gives an idea about the precision of the positioning. Than the latitude (LAT) and longitude (LON) are given, these are the coordinates of the GPS device in WGS84 reference system. Finally the date of the measurement is given.

Rows 2 to 8 of *Table 1* gives the data collected from the mobile phone. Row 2 gives the data about the current cell, because ATTACHED is stated to be TRUE, and rows 3 to 7 gives the data about the neighbouring cells, in this example 6 neighbouring cells are measured. The GSM data consists of a code in which country the mobile phone is located, the mobile country code (MCC), for the Netherlands this is 204. A code for the network which the mobile phone is currently using, the mobile network code (MNC), the Dutch code for Orange is 20. A code for the area the mobile phone is working in, the local area code (LAC). The LAC is a collection of multiple cells. Voorhof is located in code 10601. Per cell the cellid, signal strength and the channel the cell is operating on is given.

Table 1 output example of POLS. The first row gives the GPS data, the second row data about the current cell and rows 3 to 8 gives the data about the neighbouring cells.

TYPE=GPS|TIME=1201629651|TIMEOFFIX=180100.271|NUMSAT=06|HDOP=01.4|LAT=515954.7|LON =42160.0666666667|DATE=290108|
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=11802| SIGNAL=52|CHANNEL=715|ATTACHED=TRUE
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=11801| SIGNAL=31|CHANNEL=734|ATTACHED=FALSE
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=41782 |SIGNAL=15|CHANNEL=735|ATTACHED=FALSE
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=41783 |SIGNAL=19|CHANNEL=674|ATTACHED=FALSE
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=37791| SIGNAL=25|CHANNEL=712|ATTACHED=FALSE
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=2544 2|SIGNAL=11|CHANNEL=722|ATTACHED=FALSE
TYPE=GSM|TIME=1201629651|GSMTYPE=CELL|MCC=204|MNC=20|AREAID=10601|CELLID=16691| SIGNAL=18|CHANNEL=702|ATTACHED=FALSE

Table 2 result of the POLS output of Table 1.

Description	*Value*
Time	1201629651
Latitude	515954.7
Longitude	42160.0666666667
Mobile country code	204
Mobile network code	20
Local area code	10601
Current cellid	11802
Signal strength current cellid	52
Cellid neighbour 1	11801
Signal strength neighbour 1	31
Cellid neighbour 2	41782
Signal strength neighbour 2	15
Cellid neighbour 3	41783
Signal strength neighbour 3	19
Cellid neighbour 4	37791
Signal strength neighbour 4	25
Cellid neighbour 5	25442
Signal strength neighbour 5	11
Cellid neighbour 6	16691
Signal strength neighbour 6	18

Rearranging the needed data from Table 1 gives Table 2. In this table only the data needed for this thesis are collected. In this example data of 6 neighbouring cells are collected, this will not always be the case. The number of neighbours received depends on how many cells the mobile unit can receive. High buildings can block signals, because of this it can happen that the mobile unit is not able to receive six neighbouring cells. Another reason that it is not possible to receive neighbouring cells is that cell sizes are large and because of this the signal does not reach the mobile unit.

4.4 Execution

When collecting the sample data the mobile phone is in idle mode. This means that all received signal strengths are signal strength of the control channel on which no power control is used and therefore is constant in time. One measurement consists of the current cellid and its received signal strength and of the cellid of up to six neighbouring cells and their received signal strength. This results in a maximum of 7 cellids and 7 received signal strengths (RSS) per measurement.

The database is created by cycling all roads of the testing area. It is tried to cycle every road once. Also a big loop around the area is cycled to make sure that three cells are completely covered. The data is collected on different times of the day. This results in a database with 3389 points. In Figure 6 the points of the reference database are plotted. The different colours represent the different current cells, as can be seen the testing area is covered by four cells that are used for communication.

Figure 6 current cellid plotted for the reference database. Each colour represents a current cellid.

A week after the database is created the testing area is visited again, but this time to cycle a test route. The measurements collected are used to test the different positioning techniques. The route is cycled three times on different times of the day. The first and the last time the route is cycled anti-clockwise and the second time the route is cycled clockwise. During the second time the route is cycled the GPS makes large errors as can be seen in Figure 7A, namely the blue dots, for this reason this test route is left out.

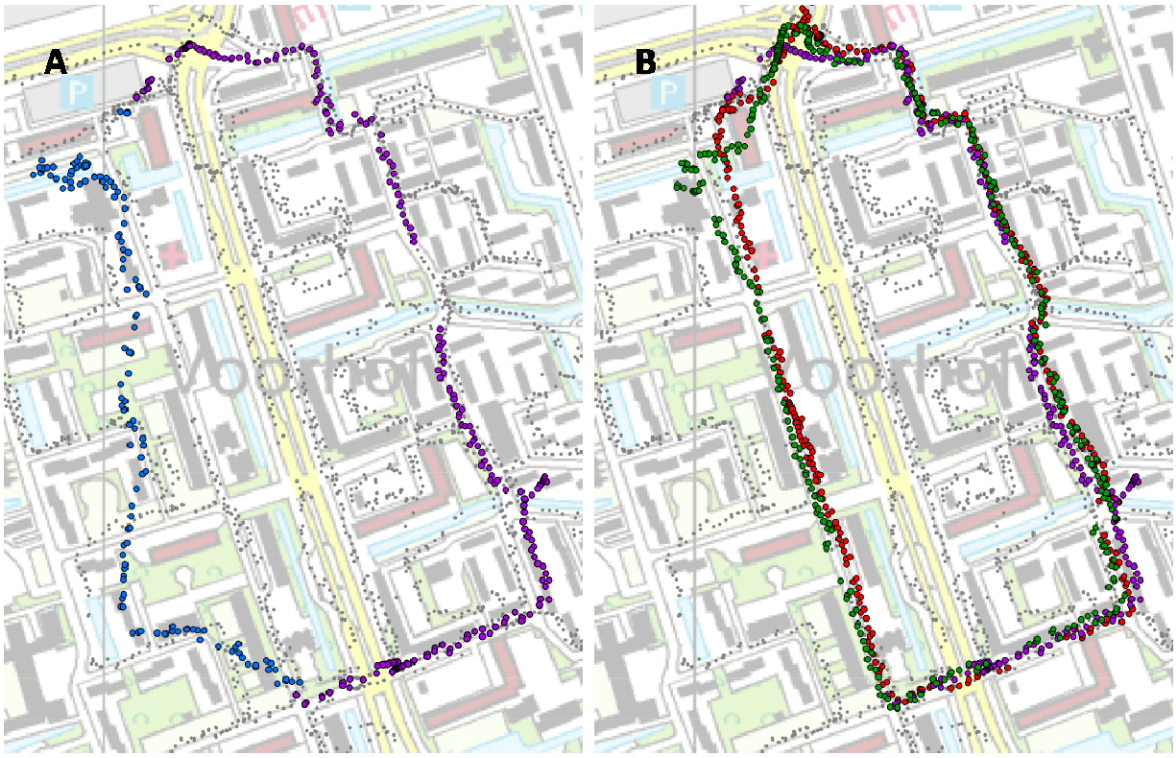


Figure 7 overview of the testroute. 6A) in blue the points that are left out of the analysis, because of the large error made by the GPS. 6B) All points of the three testroutes. In green route 1, in purple route 2 and in red route 3.

Figure 8 shows the points collected for the test route. The colours represent again the different current cellids. The same colour scheme is used as in Figure 6. It can be seen that the testing route is covered by four current cells and comparing Figure 6 with Figure 8 it is concluded that the current cells of the reference database are the same as for the testing route. It can be concluded that the current cellid is constant in time and place.

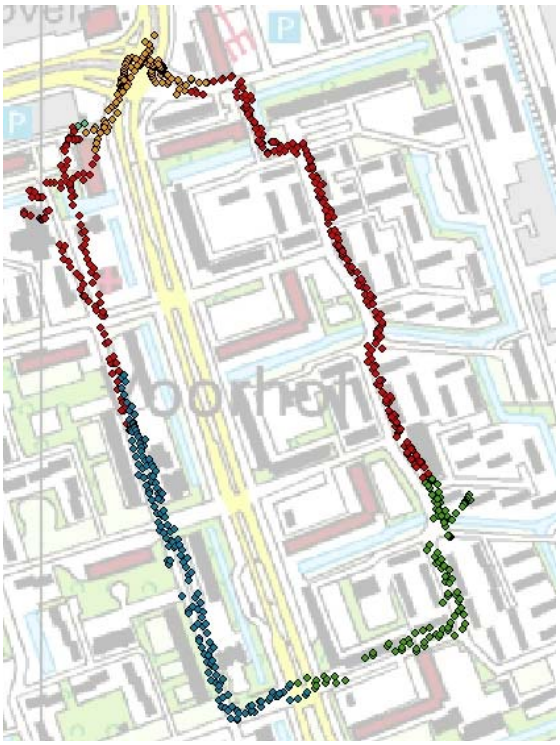


Figure 8 current cellid plotted for the test route. Each colour represents a current cellid.

5. GSM cellid positioning

GSM network consists of different cells. Each of these cells has its unique identification number, the cellid. When it is known where this cellid is received this information can be used for positioning. This chapter gives an overview of how this information can be used to determine one's position, this is done in paragraph 5.1. In paragraph 5.2 the results are given when cellid positioning is applied to our test data.

5.1 Cellid positioning

In the Netherlands the mobile operators make use of directed antennas. This means that the antenna, which is attached to the base transceiver station (BTS) emits in an angle of 120 degrees. Each BTS has three antennas connected to it, so that the area around the base transceiver station is fully covered by all antennas. For most operators the antennas are directed north, southeast and southwest and the corresponding cellids end with a 1, 2 or 3 respectively. Figure 9 shows that this is also the case for our selected study area Voorhof: the BTS is located approximately at the yellow dot and the different cells are drawn, i.e. the north-directed cell (11801), the southeast-directed cell (11802) and the southwest-directed cell (11803).

The simplest GSM positioning technique is the antenna mast position. This means that the mobile user is positioned at the position of the BTS. It can be clearly seen in Figure 9 that this gives a problem when directed antennas are used, because for cell 11801, 11802 and for 11803 the user will be positioned on the same spot, namely the yellow dot. When omnidirectional antennas are used, these antennas send in 360 degrees, this problem does not occur, because than the BTS and the antenna is situated in the centre of the cell.

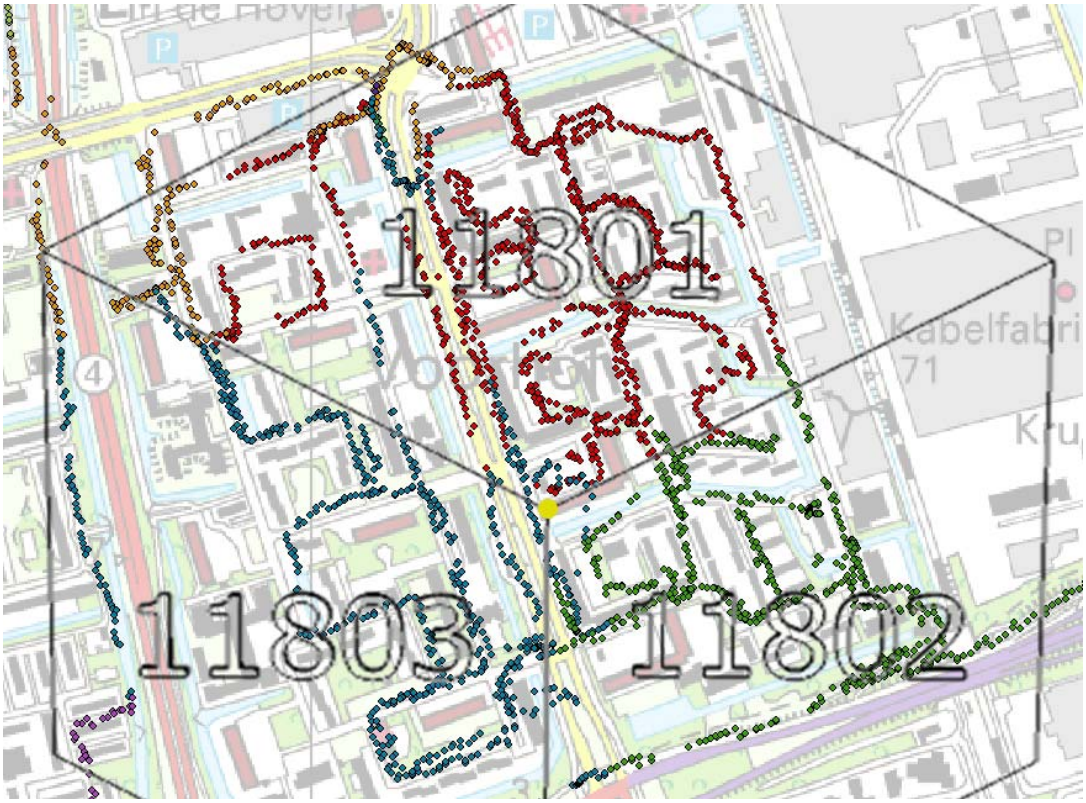


Figure 9 cell coverage of the study area. The yellow dot is the approximate location of the base station. The different dotted colours represent the different measured current cellid. Here north of the base station cell 11801 is located, southeast cell 11802 and southwest of the base station 11803.

Another drawback of positioning the mobile station (MS) at the position of the BTS is that the position of the BTS needs to be known. Telecom operators are not willing to share the information of the network infrastructure, because of the severe completion. For commercial applications and services, telecom operators try to keep the knowledge of the exact position of the base transceiver station and the corresponding cells as much as possible to themselves.

With only limited knowledge of the infrastructure of the GSM network, it still is possible to use the current cellid to position the MS. In order to do this, measurements should be collected. These measurements should include at least coordinates of where the measurement is taken and the current cellid. By averaging all the collected coordinates related to one current cellid a point can be calculated, which will be the position the MS will be positioned to when cellid positioning is used. If the measurements are evenly spread over the cell and completely cover the cell this calculated position will approximately be the centre of the cell. This positioning technique is currently used at the company WebIntegration.

GSM networks are stable, because of the complexity and cost to create the network. This means that the location of a cell does not change very often. It will only change when major changes occur in the area of a cell. Because of this stability the point where the mobile unit will be plotted to needs only to be determined once per cell. This makes cellid positioning a technique that is stable and very easy to implement.

As mentioned in paragraph 2.2 cell sizes varies from 200 meters in urban areas to 35 kilometres in rural areas. Assuming that the mobile unit will be plotted in the centre of the cell this means that the maximum error to be made will be 100 meters for urban areas and 17,5 kilometres for rural areas. Because our test area is in an urban area it is expected that the maximum error made is not more than a few hundred meters.

5.2 Real world scenario

In chapter 4 the collection of sample data for building our reference database is discussed. It was shown that the testing area is covered by four current cells. For each of these four cells a point can be calculated using cellid positioning and the reference dataset. In Figure 10 the result of this can be seen, the four big blue dots give the position to where a mobile unit will be positioned when the mobile unit is located in that current cell.

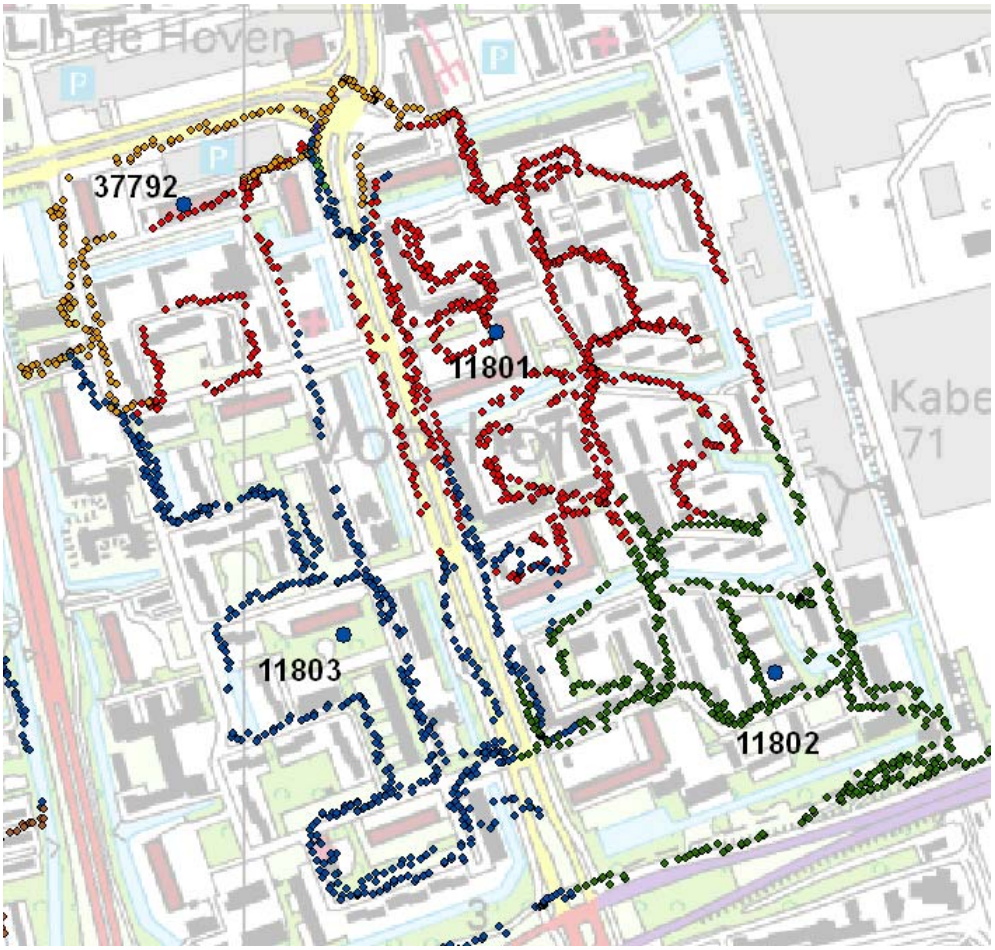


Figure 10 The four big blue dots are the positions where the mobile station will be plotted to, when using the cellid positioning method.

Accuracy can be split into two parts, namely precision and reliability. Precision is the closeness to the true position. In this case the true position is the position where the mobile unit actually is. Figure 11 gives an impression of the precision of cellid positioning. On the x-axis the distance between the reference point and the true position is given, on the y-axis the percentage of occurrence. Distance to the true position means the distance between the point where to mobile unit is positioned and the point where the mobile unit actually is. The most right bar belongs to two measurements in the third time the route is cycled. These two measurements have a different current cellid than all others, namely 16683. This cell is not completely covered by the test database, for this reason the reference point lies relatively far from the true position. Ignoring these values it is noticed that the positioning error never exceeds approximately 300 meters. This can be explained by the fact that cell sizes are small in the test area, because of the high number of potential callers. To be more precise, cell size in Voorhof is 600 meters which corresponds with the maximum error found. Using cellid positioning the measurement is postitioned in the centre of the cell when the reference points are evenly spread over the cell. The maximum error that is made than is when the measurement is taken at the border of the cell.

Another remarkable thing is that the positioning error is never lower than approximately 40 meters. The reason for this is the way the testing route is chosen. The test route never goes exactly through the centre of the cell as can be seen in Figure 7. This makes it for cellid positioning impossible to have a positioning error of 0 meters. If a test route would go through the centre of the cell than a positioning error of 0 meters will be found.

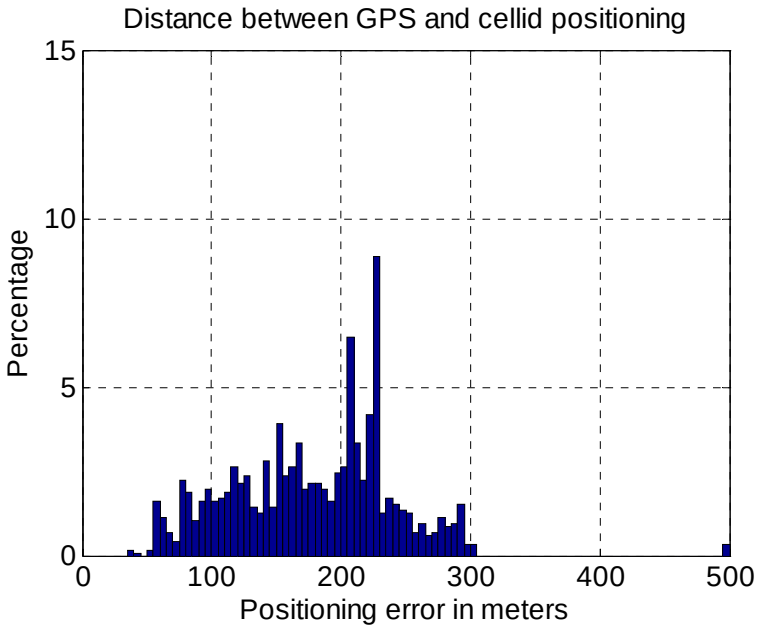


Figure 11 positioning error in meters for cellid positioning for all points collected in the three different test routes. The x-axis gives the error in meters, the y-axis the percentage of occurrence.

Another representation of the results is given in Figure 12 where the points are plotted around the true position. This plot is made by subtracting the true position from the calculated position. On the x-axis is the error in the x-direction and on the y-axis the error in y-direction, both in meters. Here again it can be seen that the positioning error is never 0, because there are no points exactly at the true position.

The x-axes in Figure 12 represents a North-south line through the calculated position, with North at the bottom, because we subtracted the true position from the calculated position. The y-axes represents a West-East line through the calculated position, with West at the right side. Knowing this one can recognize parts of the test route in the figure, namely the parts that are in the same current cell. From this figure it is also clear that in most cases the centre of the cell is situated West from the test route, this can also be verified in Figure 10.

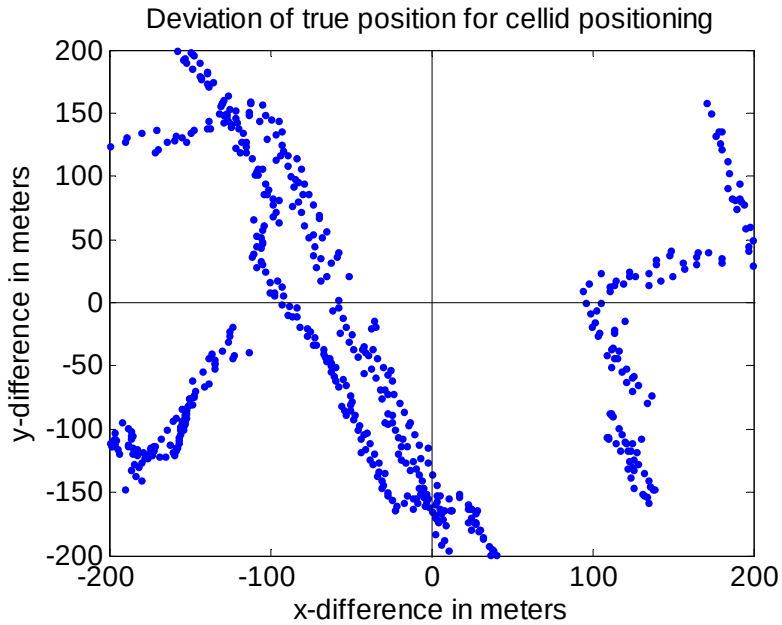


Figure 12 deviation from true position for cellid positioning.

The other part of the accuracy description is reliability. From the results of the cellid positioning technique no reliability measure is given. The point used to position the mobile unit to can be said to be an arbitrary point in the cell. When the measurements to come to this point are taken differently, for example not evenly spread over the cell like what happened with cellid 16683, the position of this point will be completely different. With other words, with cellid positioning one can position the mobile unit somewhere in the cell, but nothing is known about how good this position is.

In this chapter a GSM positioning technique is introduced that only uses the current cellid. With POLS more information about the network can be collected at a position. In the next chapter all information is used to calculate the position of the mobile unit. This means current cellid and its received signal strength (RSS), but also the cellid of up to six neighbouring cells and their corresponding RSS. Adding this extra information to the positioning better results are expected.

6. Database correlation method

In the previous chapter cellid positioning is discussed as a form of GSM positioning technique. In this chapter another technique for positioning using the GSM network is proposed, namely the database correlation method (DCM). This technique makes use of a reference database. In 6.1 the reference database creation is explained. In 6.2 it is explained how this database is used for positioning. In 6.3 the relationship between the signal space distance and the Euclidean distance is examined. The set of neighbours received is the topic of 6.4. Finally in 6.5 all theory from this chapter is applied on our test data.

6.1 Training phase

The database correlation method (DCM) is a form of GSM fingerprinting. The set of received cellids of the current cell and of up to six neighbouring cells and their corresponding received signal strengths (RSS) is called a fingerprint, because of its unique characteristics at one position. For this reason techniques that make use of these fingerprints is called GSM fingerprinting.

The basis for DCM is a database that works as a kind of lookup table. The creation of this database is the first step in DCM and is called the training phase. The database consists of some reference points in the area of interest. At these reference points a fingerprint is measured. This fingerprint consists of the current cellid and the cellid of up to six neighbouring cells and their corresponding received signal strength (RSS).

A way to visualize a fingerprint can be seen in Figure 13. In this figure the fingerprints are represented by up to seven dots. The dot in the middle represents the current cell, the dots around this middle dot represent the neighbouring cells. The colour of the dot corresponds to the cellid and the size of the dot depends on the received signal strength, a larger dot means a higher RSS. The neighbouring cells are ordered by the RSS value in counter clockwise direction. This means that the position of the dot says nothing about the location of the neighbouring cell.

The collection of the data for the reference database is explained in chapter 4. The result is a database with 3389 reference points. Figure 13 gives a small example of how the data looks like. In this example two different current cellids occur, because two different colours are found in the centre of the fingerprints, blue and red. The fingerprints in the upper half of the figure only have five neighbouring cells. The GSM signals for these positions are probably blocked by the two buildings at the top of the figure, the two dark shapes.

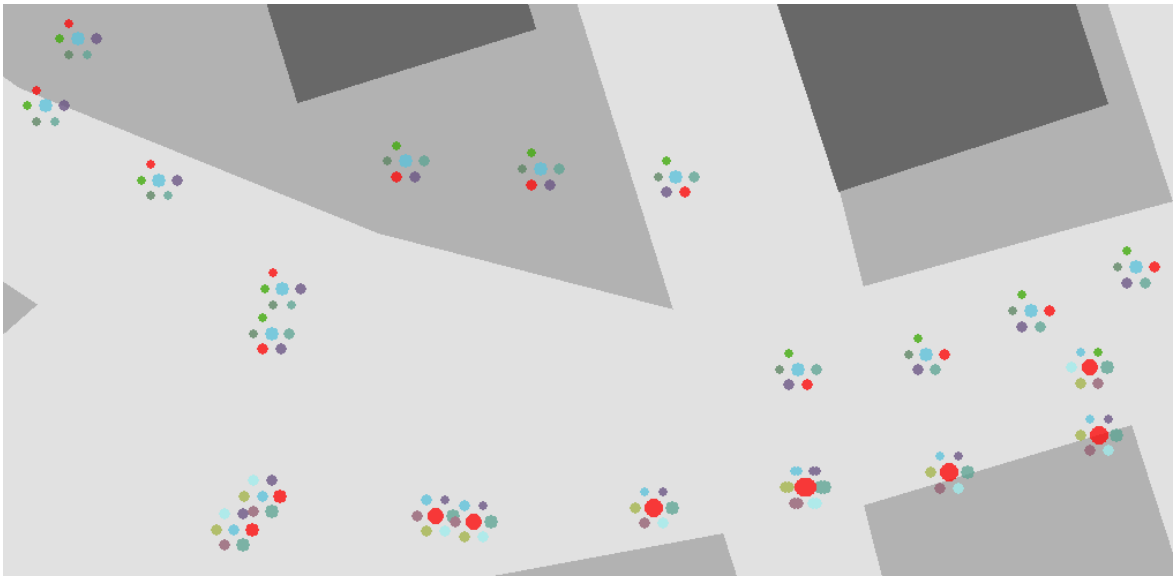


Figure 13 an example of GSM fingerprints collected for the reference database. The dot in the middle represents the current cell and the dots around it represent the neighbouring cells. The colour of the dot represents the cellid and the size of the dot corresponds to the received signal strength. The neighbouring cells are ordered by the RSS value in counter clockwise direction.

6.2 Retrieval phase

Now the database is created a method is needed to use this database to come to the position of the mobile station (MS). This is done in the retrieval phase, a.k.a. the online phase. The coordinates of the MS is determined by matching the fingerprint of the MS with the once of the reference points in the database. Different algorithms can be used to perform this matching and will be explained in this part.

For matching the distance in signal space is used. This distance is not the same as the distance in Euclidean space, i.e. the metric distance. The distance in signal space can be calculated as

$$d(\underline{x}_j, \underline{x}) = \sqrt{\sum_i (\underline{x}_j - \underline{x})^2}$$

where $\underline{x}$ is the vector with signal strengths of the MS, with a maximum of 7 signal strengths, and $\underline{x}_j$ is the vector of signal strengths of the reference point in the database. In words this means that the distance is calculated by subtracting signal strengths with the same cellid and adding these differences. This means that not necessarily all cellids in a fingerprint are used. Only the cellids that correspond to a cellid of the fingerprint in the reference database are used.

The simplest method is to take the reference point from the database with the closest distance in signal space. The position of this point is used as the position of the MS. This method is called the nearest neighbour method. The search can be limited to the reference points with the same current cellid as the measured fingerprint, because we have seen in chapter 4.2 that the current cellid is constant in time and place.

It will not always be the case that the measured point is on the exact position of one of the reference points in the database. A possibility to take this into account is to select not only the point with the closest distance in signal space, but choose the *k* closest points in the signal space from the database. The coordinates of these nearest neighbours are averaged to come to the position of the MS. This is done in the k nearest neighbours algorithm (KNN). The coordinates of the MS can be calculated by

$$\underline{y} = \frac{\sum_{j=1}^{k} \underline{y}_{ij}}{k}$$

with $\underline{y}_{ij}$ the vector of coordinates of the reference point with the closest distance in signal space and k the number of nearest neighbours.

By using k reference points, it can happen that a point relatively far away in signal space is chosen. To take this into account a weighted average could be calculated, with the weight depending on the distance in signal space. This is called the weighted k nearest neighbour algorithm (WKNN), and can be calculated as follows

$$w_j = \frac{1}{d(\underline{x}_{ij}, \underline{x}) + d_0}$$

with $d(\underline{x}_{ij}, \underline{x})$ the distance in signal space and d_0 a small constant, for instance 0.01, to prevent dividing by zero. The formula of the WKNN than becomes

$$\underline{y} = \frac{\sum_{j=1}^{k} \underline{w}_j \underline{y}_{ij}}{\sum_{j=1}^{k} \underline{w}_j}$$

with again $\underline{y}_{ij}$ the vector of coordinates of the reference point with the closest distance in signal space, k the number of nearest neighbours and $\underline{w}_j$ a vector of weights.

Using the KNN and WKNN one have to keep in mind that the k nearest neighbours are the nearest neighbours in signal space, this does not necessarily mean that all these point are nearest neighbours in physical space. In (Brunato, M., R. Battiti, 2005) is shown that changing *k* in the range from 5 to 15 only changes the estimation error by 1%. For this reason k can be kept small so as not to corrupt the averaging. Keeping *k* small has also the advantage that searching time and thus computing time is saved.

6.3 Distance in signal space versus Euclidean distance

The different searching algorithms in this chapter make use of the distance in signal space. It is assumed that a close distance in signal space means that the reference point is also close in physical space. A reliability test is performed to check if this statement is true. For this test a reference point is taken from the database. The position of this point is recalculated using the nearest neighbour method, discussed in paragraph 6.2, on the remaining reference points in the database.

In Figure 14 the signal space distance is plotted against the Euclidean distance. The Euclidean distance is measured between the true position of the point and the position of the reference point with the closest distance in signal space. The signal space distance is calculated as is shown in section 6.2. Looking at Figure 14 it seems that the points are very distributed, but actually most of the points lay on the y-axis.

In Figure 3 it is shown that the received signal strength is relatively stable in time. It can be seen that the RSS fluctuates between 3 dBm. The points in Figure 14 not on the y-axis can be explained by this fact. If a measured RSS is slightly different than the signal space distance will not be equal to zero. For this reason the points not on the y-axis can be neglected.

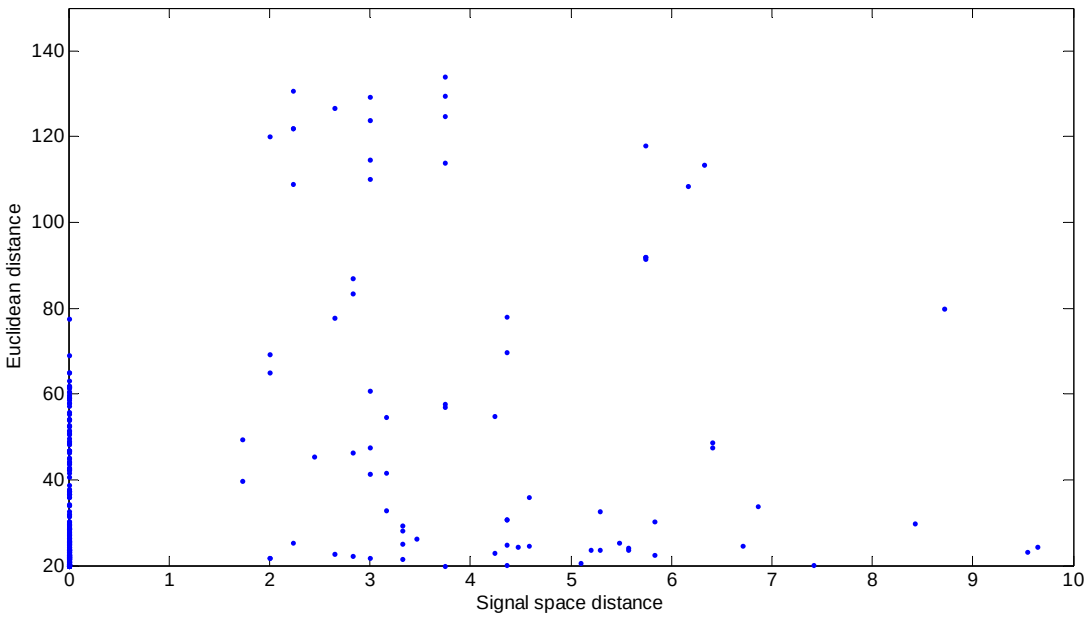


Figure 14 signal space distance plotted against the Euclidean distance, Euclidean distance is measured in meters.

Out of the 3389 points in the database and thus used for the reliability test, 3155 lie on the y-axis, this means that these points have a signal space distance of 0. Making a bar plot of these zero-points, as is done in Figure 15 where the Euclidean distance is plotted on the x-axes and on the y-axes the percentage of occurrence, it is clear that most points have a mistake of 10 meters or less. To be more precise 50 % of these points have a Euclidean distance of 8,5 meter or less and for 90% of the points the Euclidean distance is 18,3 meters or less. From this it can be concluded that there exists a correlation between the signal space distance and the Euclidean distance and therefore it is fair to assume that DCM is a good technique for positioning.

Furthermore one can conclude that a fingerprint is unique for an area with radius 4,25 meters in 50% of the cases and a radius distance of 9,15 for 90% of the cases. This supports the idea of the unique characteristics of the fingerprints.

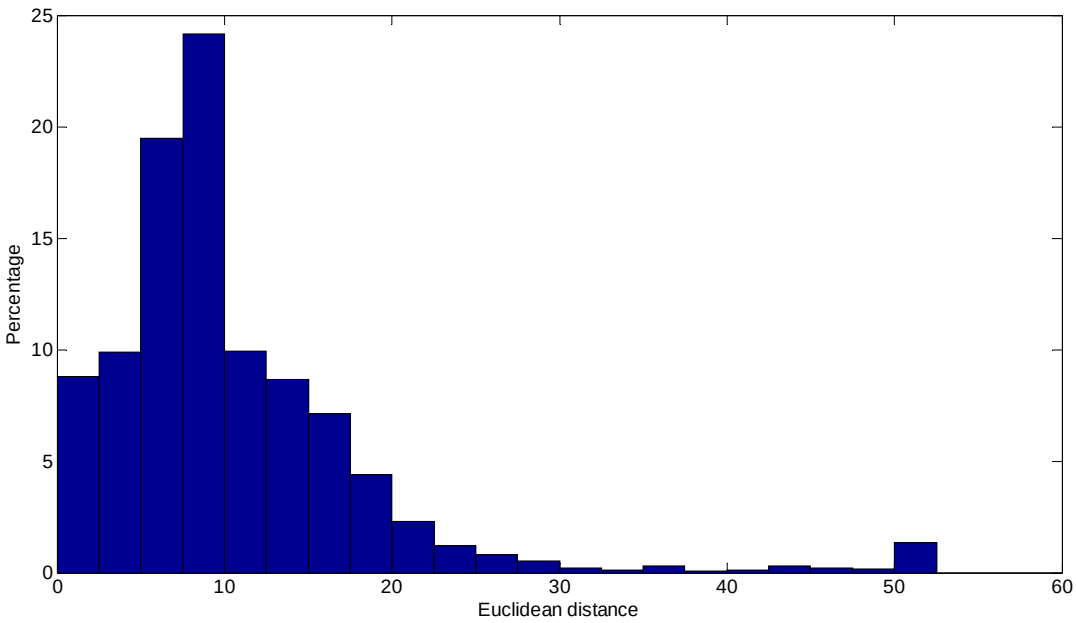


Figure 15 bar plot of the points with signal space distance of 0. On the x-axis the Euclidean distance and on the y-axis the percentage of occurrence.

6.4 Set of neighbours

In paragraph 4.2 it was shown that the current cellid is constant in time and place. The question arises if this is also the case for the set of neighbours received at one position. In Figure 16 fingerprints are plotted of the database and of the test route. The fingerprints of the test route are connected with a striped line. The different colours represent the different cellids and the sizes of the dots represent the corresponding RSS. In this figure it is not possible to find an exact match of a point of the test route with a reference point from the database. To see if an exact match can be found the database points are compared with the points in the test route.

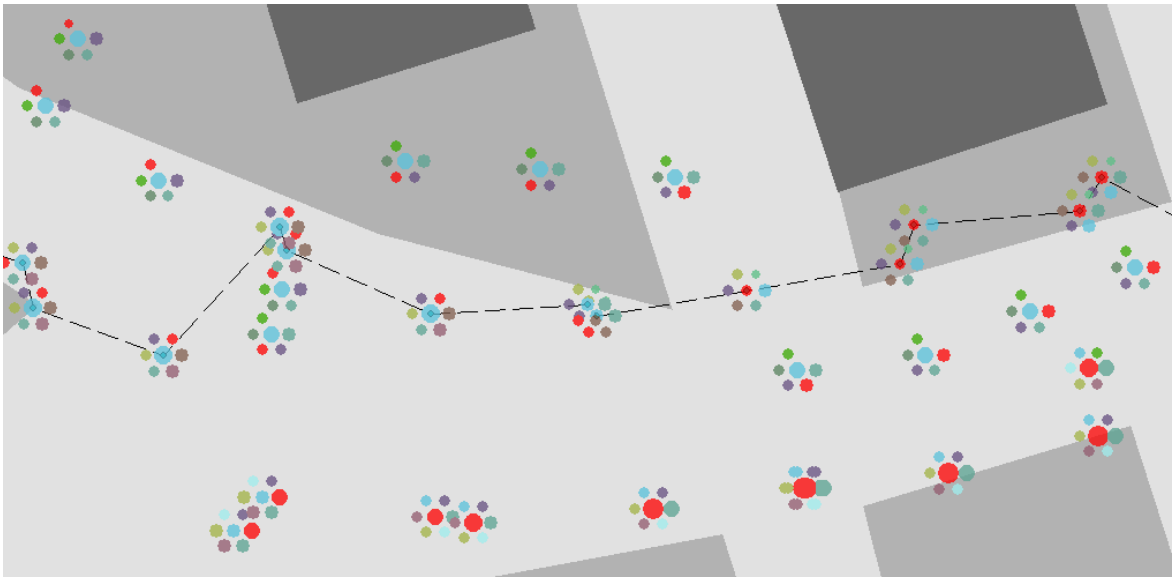


Figure 16 fingerprints of the test route and of the reference points in the database, one fingerprint consist of a center point with at most six points around it. The middle dot represents the current cell and the dots around it represent the neighbouring cells. The striped line represents the test route and the points on this line are the fingerprint of the test route.

In Figure 17 the reference points of the database are plotted. Each colour represents a different set of neighbours; the RSS is not taken into account. This means that the same colour does not necessarily mean that the fingerprints are exactly the same, because a fingerprint is the combination of cellids and RSSs. The same is done for the test route, using the same colour scheme of Figure 17, so the same colours in both figures represent the same set of neighbours. The result of this can be seen in Figure 18. Comparing Figure 18 with Figure 6 it is clear that not all points of the test route are drawn. The points that are left out in this figure are the points for which no exact match could be made with the database.

That it is not possible to make an exact match with a reference point of the database does not mean that DCM does not work. It means that one or more of the cellids in the fingerprint of the test route or in the fingerprint of the reference point in the database is not used to calculate the signal space distance. For this reason it seems better to take the number of matching cellids into account.

The distance in signal space is calculated by adding the difference of the RSS for cellids that occur in the measured fingerprint at the MS and in the reference fingerprint. This means that when more cells have the same cellid more is added to the signal space distance. Doing this gives points with fewer cells in common an advantage for cells that have more cells in common. For this reason it is better to calculate an average signal space distance, by dividing the signal space distance by the number of matching neighbours.

Figure 17 reference points of the database. Each colour represents a different set of neighbours.

Figure 18 points of the test route that can be exact matched to a set of neighbours of the reference points of the database. Each colour represents a different set of neighbours.

6.5 Real world scenario

Now we know that DCM is a technique that could work, because there exists a correlation between the signal space distance and the Euclidean distance, as is shown in 6.3. But that probably the number of matching neighbours has to be taken into account, because the set of neighbours is not constant, as is shown in 6.4. Now we have to apply this on our test data and compare the results of the different DCM algorithms with the GPS position.

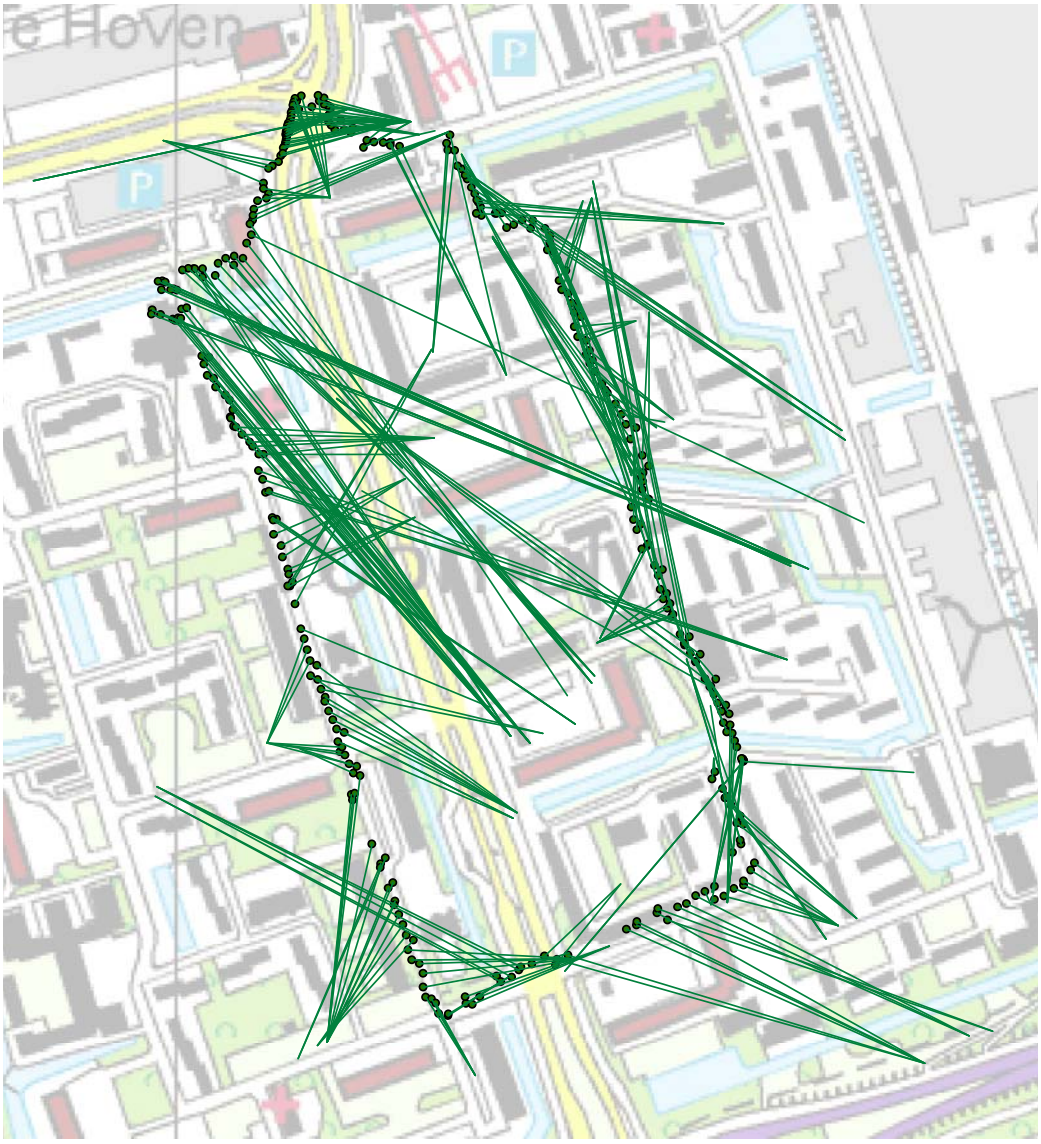


Figure 19 example of one of the test routes. The green dots represent the true position measured with GPS. The green line represents the positioning error made with the nearest neighbour algorithm, the end of the green line is the position calculated with this algorithm.

In Figure 19 the results of one of the test routes is given. The green dots represent the GPS position, the connecting green line represents the error made using the nearest neighbour algorithm, end of the line string is the position calculated with this algorithm. Comparing this figure with how it would look like for cellid positioning, where the MS only is plotted to four different points, it is clear that the MS is positioned at more different positions. This means that with DCM motion of the MS within a cell can be detected, while with cellid positioning this is not possible.

Looking at Figure 19 it can be seen that some of the points of the test route are positioned at the same position with DCM. This means that the fingerprint are very similar, note that the fingerprint does not need to be exactly the same, because only the cellids that occur in the reference fingerprint and in the fingerprint of the MS are used for the positioning.

Accuracy is again split up in two parts, namely precision and reliability, as is done for cellid positioning. In Figure 20 the histogram is given for the nearest neighbour method, the histograms of the KNN and WKNN methods give similar results as can be seen in Appendix B. On the x-axis

the positioning error in meters is given and on the y-axis the percentage of occurrence. The bar most to the right represents the occurrence of errors of 495 meters or more.

Comparing Figure 11 with Figure 20, both figures can also be seen in appendix B, it is directly clear that the positioning error of the nearest neighbour method is more spread than the error made with cellid positioning. Which can be explained by the fact that more points are used to position the MS to. With cellid positioning the MS is positioned at 5 different positions namely a point in the five different current cellids that occur in the test routes. While with DCM all database points that are in those 5 current cells can be used for positioning. On the other hand with the nearest neighbour method it is possible that the mobile unit is plotted at the true position.

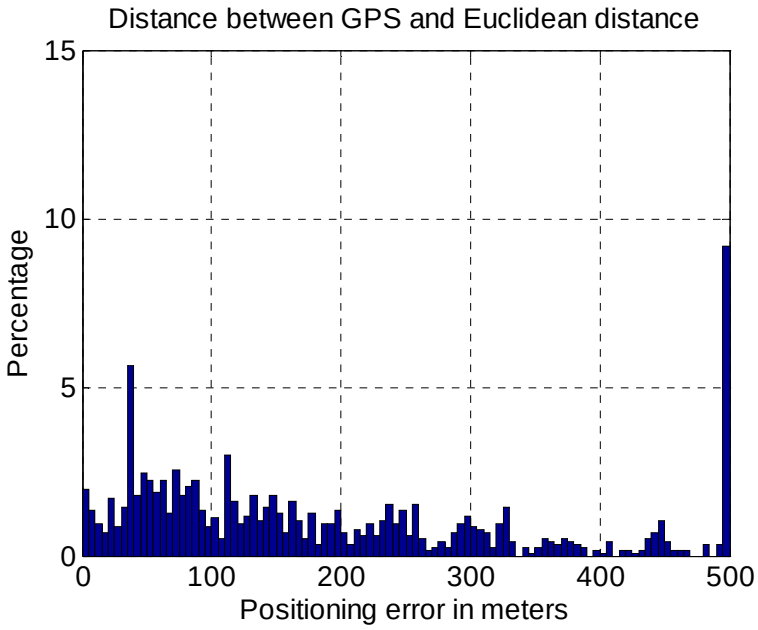


Figure 20 positioning error in meters for the nearest neighbour method without taking the number of neighbours into account. On the x-axis the distance from the true location in meters and on the y-axis the percentage of occurrence. The last bar represents the percentage that have a positioning error of 495 meters or more.

A visualization of the precision is given in Figure 21 where the spread around the true position is plotted for the nearest neighbour method, the plots for the KNN and WKNN methods give similar as can be seen in Appendix A. On the x-axis stands the error in the x-direction and on the y-axis the error in the y-direction, both in meters. The errors are more located directly around the true position compared to Figure 12. The deviation of the true position is more random than with cellid positioning. This can again be explained by the fact that more points are used for positioning.

With the DCM method errors larger than 300 meters also occur, as can be seen in Figure 20. While with cellid positioning the error almost never exceeds 300 meters. The reason that larger errors occur is that in the worst case scenario the mobile unit can be plotted on the edge of the cell, while it is located across the cell on the other edge. This would make an maximum error of around 600 meters for our test area.

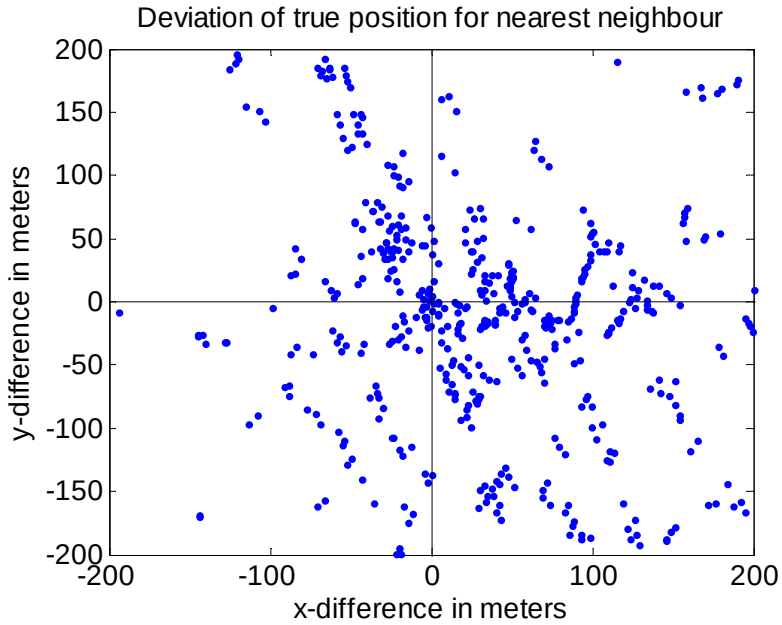


Figure 21 deviation from true position for the nearest neighbour method, without taking the number of neighbours into account.

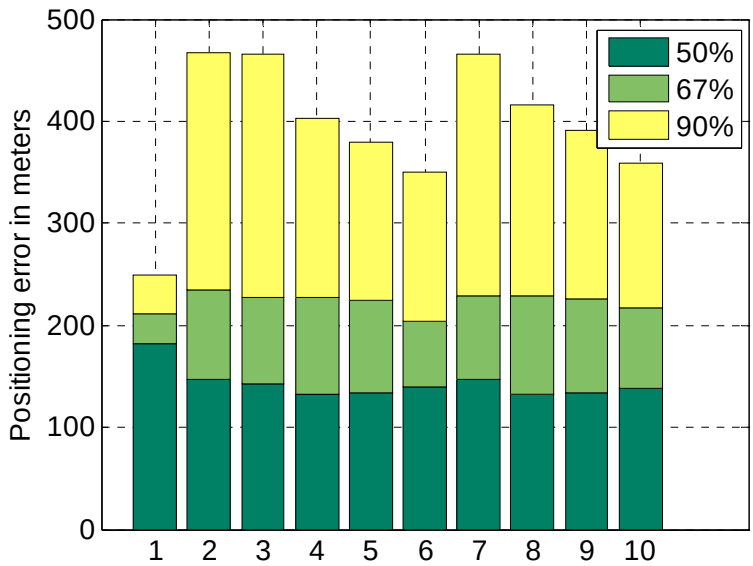


Figure 22 summary of the different positioning techniques and their accuracy. 1: cellid positioning. 2: nearest neighbour. 3: 2 nearest neighbour. 4: 3 nearest neighbour. 5: 4 nearest neighbour. 6: 5 nearest neighbour. 7: 2 weighted nearest neighbour. 8: 3 weighted nearest neighbour. 9: 4 weighted nearest neighbour. 10: 5 weighted nearest neighbour.

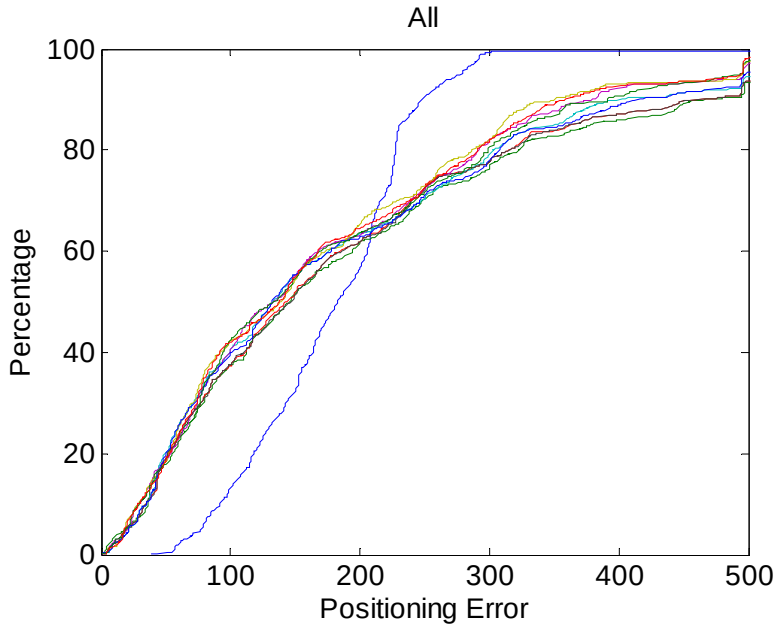


Figure 23 cumulative distribution function of cellid positioning, the blue line, and the different GSM fingerprinting methods. For approximately two third of the measurements the GSM fingerprinting techniques give better results.

Looking at Figure 22 and at 67% of the measurements the best results are obtained with the 5 nearest neighbour method. This means that using this algorithm in 67% of the calculations the error made is 2047 meters or less. For cellid positioning the error made is 211 meters or less for 67% of the measurements. This results in an accuracy improvement of only 7 meter.

In Table 3 the numerical results are given for all methods, the minim and maximum value, the average and the standard deviation. The large maximum error made with cellid positioning can be explained by a point with current cellid 16683. The maximum value for the different DCM algorithms do not really differ that much. The 5 nearest neighbour method gives the best result taken all values and Figure 22 into account.

The histogram given in Figure 22 does not give an idea about the distribution of the positioning errors. For this the cumulative distribution function (CDF) is given in Figure 23. The steeper a line is in a CDF the better the results are, because than most positioning errors are small. The blue line represents DCM. The other lines represent the different algorithms for DCM, it is clear that the differences for these different algorithms are almost none. At approximately two third of the measurements the error made with al techniques is the same, but the lines for DCM are steeper. This means that for two third of the measurements DCM give better results.

In (Li, 2005) it was stated that including 3 or 4 neighbours in the GSM fingerprinting method give the best results. From Figure 22 it is observed that the accuracy still improves taking more than 4 neighbours into account. According to (Brunato, M., R. Battiti, 2005), the accuracy would only improve with 1% when including more than 5 neighbours and because of that we have limited ourselves to only include not more than 5 neighbours in our field test.

We have looked at the precision of GSM fingerprinting, now we have to take a look at the reliability of GSM fingerprinting. The reliability of GSM fingerprinting can be defined by the signal space distance, because a low signal space distance implies that the mobile unit is close to the reference point. In other words a low signal space distance implies a good reliability

Table 3 numerical results of all DCM algorithms.

Method	*Minimum*	*Maximum*	*Average*
Cellid positioning	38,9	721,3	178,9
Nearest neighbour	1,3	649,7	193 6
2 nearest neighbour	1,7	648,1	189,2
3 nearest neighbour	2,2	650,7	182,0
4 nearest neighbour	1,1	650,5	175,0
5 nearest neighbour	3,7	649,4	170,8
Weighted 2 nearest neighbour	1,7	648,1	189,7
Weighted 3 nearest neighbour	2,2	650,7	183,4
Weighted 4 nearest neighbour	1,1	650,5	176,9
Weighted 5 nearest neighbour	3,9	649,4	172,8

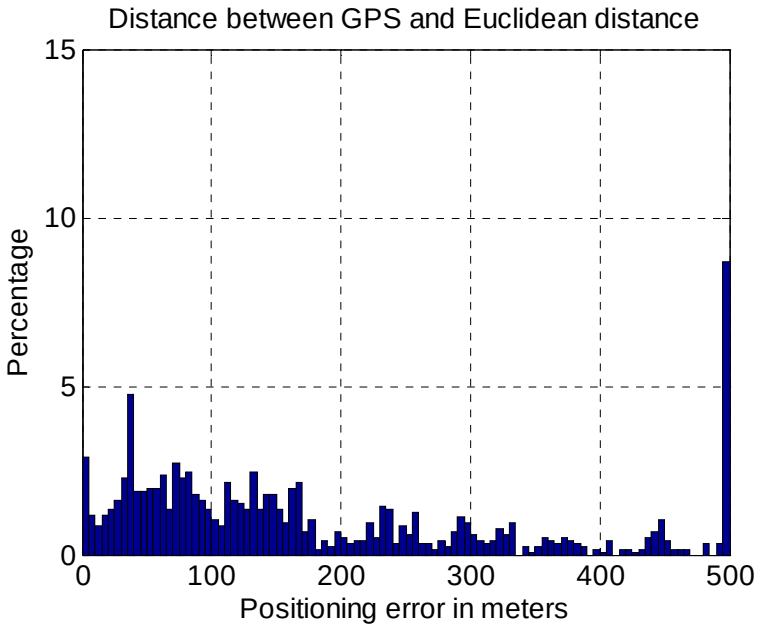


Figure 24 positioning error in meters for the nearest neighbour method taking the number of neighbours into account. On the x-axis the distance from the true location in meters and on the y-axis the percentage of occurrence. The last bar represents the percentage that have a positioning error of 495 meters or more.

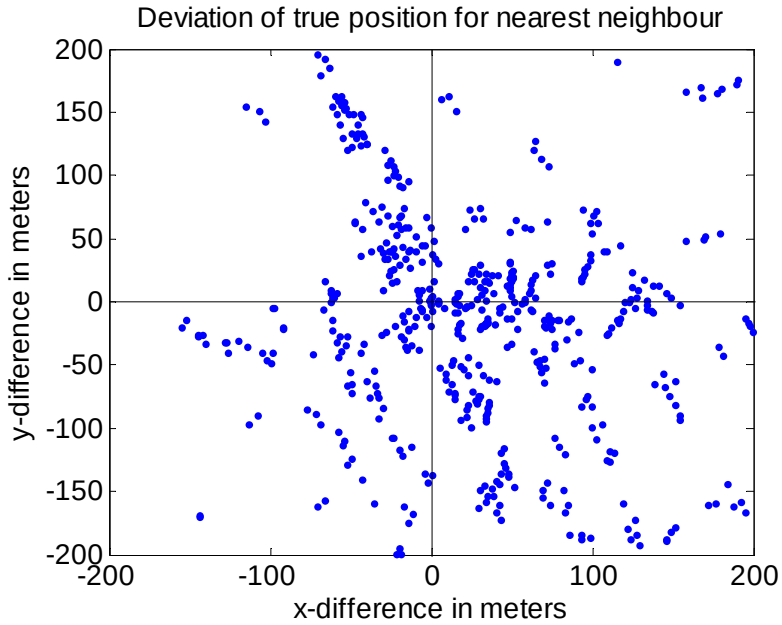


Figure 25 deviation from true position for the nearest neighbour method, taking the number of neighbours into account.

In 6.4 we concluded that DCM could probably be improved when the number of matching neighbours is taken into account by calculating an averaged signal space distance. In Figure 24 the result is given when this is done for the nearest neighbour method, the results of the other algorithms are given in appendix C. There seems almost no difference when comparing this figure with Figure 20.

From a comparison of Figure 26 with Figure 22, it becomes clear that a precision improvement is obtained when taking the number of matching neighbours into account. Especially for the KNN methods a precision improvement is obtained. For the nearest neighbour method the improvement is not very significant. Further, there is almost no precision improvement for the weighted nearest neighbour methods which can be explained by the fact that the signal space distance is divided by the number of matching neighbours. Doing this makes the signal space distance very small. This means that averaging is not very useful any more.

In Figure 26 the results of all positioning techniques are given, Table 4 gives the numerical results. Looking at 67% cellid positioning performs better than the nearest neighbour algorithm, namely an error of 211 meters or less for cellid positioning and 214 meters or less for the nearest neighbour algorithm. The other algorithms perform better than cellid positioning namely an error between 172 to 189 meters for 67% of the measurements. In Appendix E a table is given with all results.

Looking at Figure 26 and at 67% of the measurements the best results are obtained with the weighted 4 nearest neighbour technique. The error made with this technique is 172 meters or less for 67% of the measurements. For cellid positioning the error made is 211 meters or less for 67% of the measurements. This results in an accuracy improvement of 39 meter.

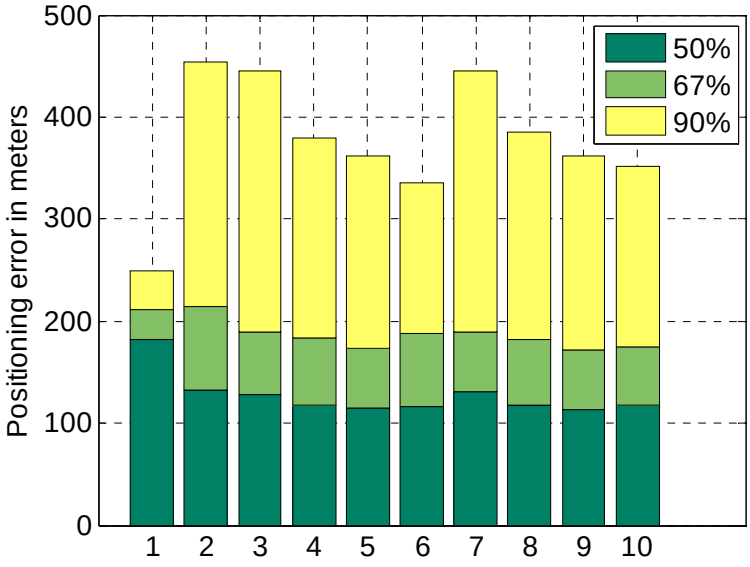


Figure 26 summary of the different positioning techniques and their accuracy taking into account the number of matching cellids. 1: cellid positioning. 2: nearest neighbour. 3: 2 nearest neighbour. 4: 3 nearest neighbour. 5: 4 nearest neighbour. 6: 5 nearest neighbour. 7: weighted 2 nearest neighbour. 8: weighted 3 nearest neighbour. 9: weighted 4 nearest neighbour. 10: weighted 5 nearest neighbour.

The signal space distance can again be taken as a measure for the reliability. Care should be taken to not compare the averaged signal space distance with the signal space distance used for the 'normal' DCM methods.

Table 4 numerical results for all DCM algorithms.

Method	*Minimum*	*Maximum*	*Average*
Cellid positioning	38,9	721,3	178,9
Nearest neighbour	1,1	649,7	184,3
2 nearest neighbour	2,4	648,1	177,0
3 nearest neighbour	0,6	650,7	167,9
4 nearest neighbour	3,9	650,5	161,2
5 nearest neighbour	2,1	649,4	158,3
Weighted 2 nearest neighbour	2,4	648,1	178,1
Weighted 3 nearest neighbour	2,3	650,7	169,1
Weighted 4 nearest neighbour	3,9	650,5	162,8
Weighted 5 nearest neighbour	3,5	649,4	159,8

7. Signal strength distribution method

In chapter 5 and 6 we have seen two different techniques to use the GSM data for positioning. With the first technique, cellid positioning, information about the current cellid is used to determine the position of the MS. The second technique, DCM, makes use of up to 7 cellids and of their corresponding signal strength, but we have also seen that in most cases only a part of the 7 cellids and RSSs is used. In this chapter a method is introduced that in all cases uses all received cellids and their corresponding signal strengths. In 7.1 the new method, the signal strength distribution method (SSDM), is explained. To see if this SSDM gives better results, the method is tested on our test data in chapter 7.2.

7.1 Introduction

In this new method, the signal strength distribution method (SSDM), the collected fingerprints for the reference database are used to create signal strength distribution rasters. In the positioning phase these maps are combined to create a map which contains the possible positions of the mobile unit. The minimal footprint on this map is the most likely position of the MS. The SSDM is a form of GSM fingerprinting, because we still make use of the unique characteristics of the combination of cellids and signal strengths, the fingerprint.

The RSS distribution raster is created by doing a nearest neighbour interpolation over all points in the database with the same cellid, this means that one fingerprint in the database can be used for the creation of up to seven RSS distribution maps. It makes no difference if the cellid belongs to the current cell or belongs to one of the neighbour cells.

The following formula is used to calculate the raster with the possible positions of the MS:

$$result\ raster = \sqrt{\sum_{n} (raster_i - RSS_i)^2}$$

where $raster_i$ is the RSS distribution raster of cell *i*, RSS_i is the received signal strength of cell *i* at the current position, *n* is the number of cellids that occur in the fingerprint. The most likely position of the mobile unit is at the minimum raster value of this result raster.

The mobile unit receives a maximum of seven different cellids, the RSS distribution maps of these cellids are use to determine the position of the mobile unit. This means that all information retrieved is used for the positioning, while with DCM only the information of the cellids that are the same in the reference fingerprint and the measured fingerprint is used.

As stated the most likely position of the MS is at the minimum of the raster. With the above formula this will result in an area where the minimum value occurs. To come to a position of where the MS is, the raster is converted to points. The coordinates of the points that have the minimal raster value are averaged to come to the most likely position of the mobile unit.

7.2 Real world scenario

The histogram in Figure 27 gives the results when using SSDM for our test routes. As can be seen the maximum error made is around 415 meters, which is much less than found with the different DCM algorithms, where the maximum error exceeds the 500 meters in 5 to 9 percent of the cases. The distribution of the errors also differs comparing DCM with SSDM. The distribution of DCM is evenly spread with a small peak around 40 meters. SSDM on the other hand has high peaks on the left of the image, which means lower errors. From this it can be concluded that the error made is in most of the cases for SSDM much smaller than for DCM. Taking also cellid positioning in account, Figure 11, than also SSDM give better results. The peak for cellid positioning is around 200 meters.

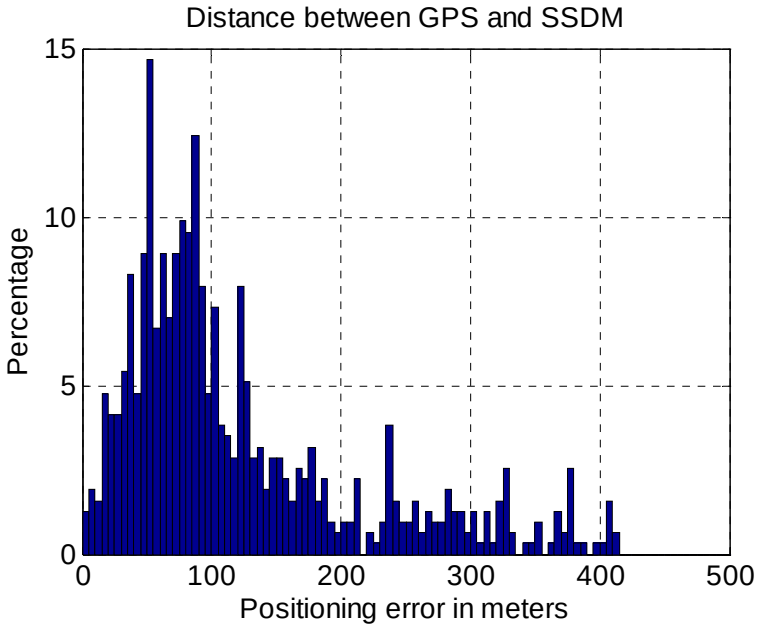


Figure 27 positioning error in meters for the signal strength distribution method.

In Figure 28 the precision of SSDM is given. There seems not much difference between this figure and similar figures for the DCM algorithms. The difference in the figures is mostly the number of points falling in the rang of -200 to 200 meters. With SSDM much more points have an error of 200 meters or less. This results in a higher precision for SSDM.

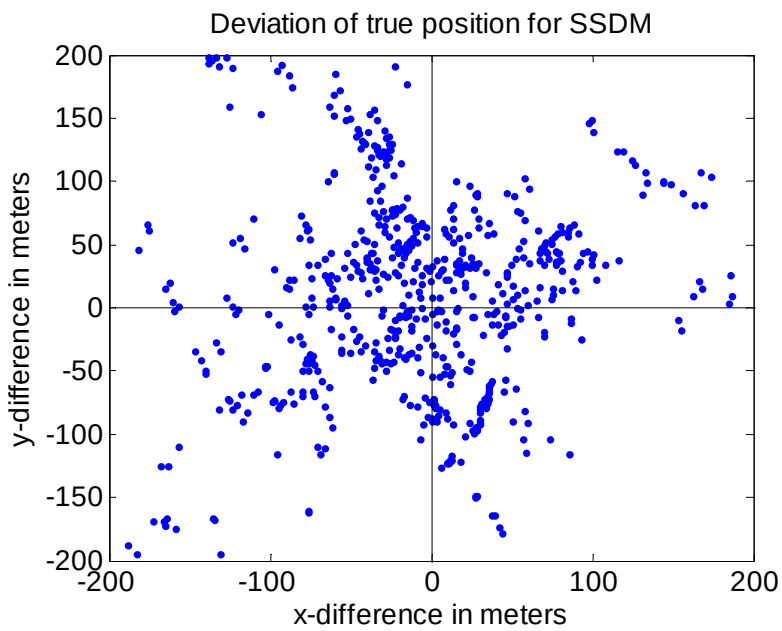


Figure 28 deviation from true position for the signal strength distribution method.

Figure 29 and Figure 30 give a summary of all the results, in Figure 29 the results of the normal DCM are given and in Figure 30 the results of DCM taking the number of matching neighbours into account. From these figures it is directly clear that SSDM give the best results. Looking at 67% of the measurements the error made is 123 meters or less for SSDM. Comparing this with cellid positioning, 211 meters or less, and the different DCM algorithms, ranging from 172 meter to 229 meters, it can be concluded that SSDM is a much better positioning method. An accuracy improvement of 88 meters is made using SSDM instead of cellid positioning.

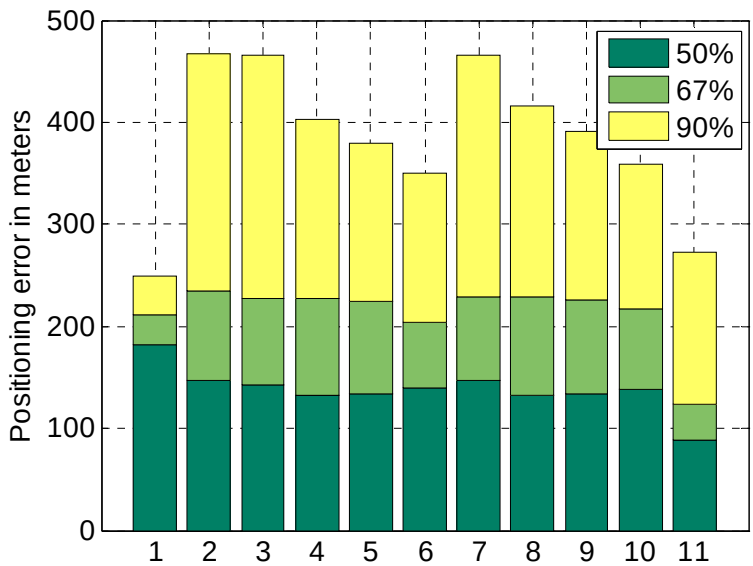


Figure 29 summary of the different positioning techniques and their accuracy. 1: cellid positioning. 2: nearest neighbour. 3: 2 nearest neighbour. 4: 3 nearest neighbour. 5: 4 nearest neighbour. 6: 5 nearest neighbour. 7: 2 weighted nearest neighbour. 8: 3 weighted nearest neighbour. 9: 4 weighted nearest neighbour. 10: 5 weighted nearest neighbour. 11: signal strength distribution method.

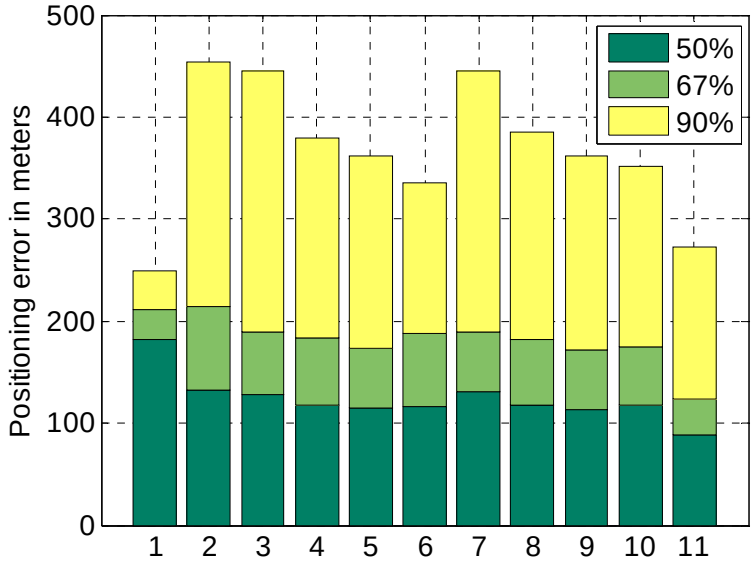


Figure 30 summary of the different positioning techniques and their accuracy taking into account the number of matching cellids. 1: cellid positioning. 2: nearest neighbour. 3: 2 nearest neighbour. 4: 3 nearest neighbour. 5: 4 nearest neighbour. 6: 5 nearest neighbour. 7: weighted 2 nearest neighbour. 8: weighted 3 nearest neighbour. 9: weighted 4 nearest neighbour. 10: weighted 5 nearest neighbour. 11: signal strength distribution method.

	Value
Minimum	4,1
Maximum	414,4
Average	120,7

Figure 31 gives the results of one of the test routes. Again in this figure it can be seen that the errors made are smaller than with DCM. Having in mind that with cellid positioning the MS will only be positioned at five different positions and for this reason it is not possible to detect motion within a cell, SSDM plots the MS at much more different positions and for this it is possible to detect motion within a cell. This makes SSDM for more application suitable than cellid positioning.

Some points of the test route are positioned at the same position using SSDM, as can be seen in Figure 31. For these points the fingerprint are the same, because all information is used for the positioning. Looking better at the figure one can see that the points that are positioned at the same position are all points that are close to each other. And the same can be concluded as was done with the reliability test in chapter 6.3, that a GSM fingerprint has unique characteristics in a small area. Note that the points at the top of the route seem to be positioned all at the same position. This is not the case, but because of the scale it seems like this. But it does mean that the fingerprints at that points do not differ that much, to be more precise the cellids are the same only the RSS differs.

Comparing Figure 19 with Figure 31 it is clear that the errors made are smaller with SSDM. Better in most cases one is positioned near the true position and it is easier to recognize the test route in the calculated positions.

Figure 31 example of one of the test routes. The green dots represent the true position measured with GPS. The green line represents the positioning error made with SSDM, the end of the green line is the position calculated with this method.

8. Conclusion

8.1 Discussion

We evaluated different types of GMS positioning, see Figure 32 for an overview of the different methods. Our objective is to assess the positioning quality of cellid positioning versus GSM fingerprinting. Therefore, we collected some sample data using GPS/GSM together with POLS software, this sample data is used to test the different GSM positioning methods. Our work deviates from that of previous research, in that we applied our research on a real-world outdoor scenario instead of a lab or campus environment. We selected a heavily urbanized area in Delft (district Voorhof) as our study area, because we expect that GSM positioning is particularly an alternative for GPS positioning in urban areas (i.e. GPS fix is disturbed by high buildings, a.k.a. the urban canyon problem).

Using cellid positioning, the positioning for 67% of the measurements is 211 meters or less. This enables us to get a caller's position on a neighbourhood-level. For particular location-based services, this positioning accuracy would be already sufficient (e.g. commercial advertising services). However, and particularly with respect to the E911 European regulations, we need to determine a callers' position on a street-level accuracy.

Two different forms for GSM fingerprinting techniques are examined. First the traditional technique, the database correlation method (DCM). And because DCM did not give satisfying results a new form of GSM fingerprinting is introduced namely the signal strength distribution method (SSDM).

Two general remarks can be made for DCM. First there exists a relationship between the signal space distance and the Euclidean distance. This means that it is likely that a point with a low signal space distance is the closest point in physical space. Therefore it is fair to assume that the signal space distance is a good measure for testing if a reference point makes a good match.

Secondly the set of neighbours is investigated. For most of the points of the test route the set of neighbours could not be exactly matched with a point in the database. The signal space distance is only calculated for points with the same cellid. In the 'normal' fingerprinting methods the number of matching neighbours is not taken into account. It is likely that a reference point with more matching neighbours makes a better match than a reference point with less matching neighbours. To take this into account an averaged signal space distance is introduced.

DCM gives in some cases a little better result than cellid positioning; though not significantly better. Looking at 67% of the measurements the positioning error ranges between 172 to 235

meters. This means that in most cases we are able to determine a caller's location on a subneighbourhood-level only.

As alternative for the database correlation method, we therefore introduced a new type of GSM fingerprinting: the signal strength distribution method (SSDM). For this method, we computed RSS distribution rasters per cellid. Next, we extracted a spatial intersection using all the distribution rasters for which we received a cellid at an observation point to find a unique spatial footprint (i.e. fingerprint) for that specific location. This means that all information retrieved is used for the positioning, while with DCM only the information of the cellids that occur in the reference fingerprint and in the fingerprint of the observation point is used.

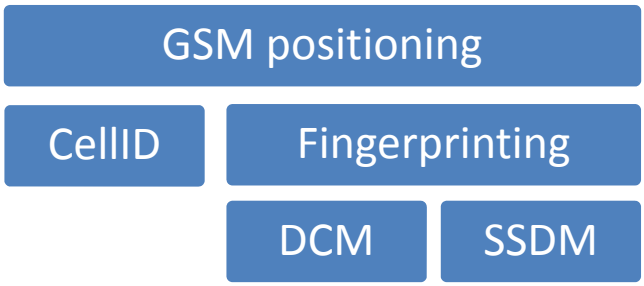


Figure 32 overview of GSM positioning methods.

We experienced that the SSDM gave significantly better results compared to the other GSM positioning methods. Looking at 67% of the measurements the positioning error made was 123 meters or less. This makes this method for most cases a suitable positioning method for street-level services.

An advantage of the different GSM fingerprinting techniques compared to cellid positioning is that motion can be detected. As long as the MS stays in one cell it is not possible with cellid positioning to detect that the MS is in motion, because the position will always be plotted to the same spot. With GSM fingerprinting the MS will be positioned at different spots in the cell and for this it can be detected that the MS is in motion.

8.2 Conclusions

The objective of this research was to find a GSM-based positioning technique that improves the accuracy compared to cellid positioning in which it was proposed to evaluate the possibilities of applying GSM fingerprinting. The research question is

How much accuracy improvement can be achieved with GSM fingerprinting compared to cellid positioning in a real world scenario?

From our results, we conclude the following:

- Cellid is a very easy to implement, stable and consistent positioning method. Using this method, a caller's location can be determined with an accuracy of 211 meters for 67% of the measurements.
- DCM gives relatively better results than cellid positioning, however not significantly better. Using this method the caller's location can be determined with an accuracy of 172 meters for 67% of the measurements. This can be achieved by using the weighted 4

nearest neighbour method and taking the number of matching neighbours into account. This results in an accuracy improvement of 39 meters compared to cellid positioning.

- SSDM gives the highest positioning accuracy compared to the other GSM positioning methods. Using this method, the caller's location can be determined with an accuracy of 123 meters for 67% of the measurements. This results in an improvement of 88 meters compared to cellid positioning.
- With the different GSM fingerprinting techniques motion within a cell of the MS can be detected, while for cellid positioning this is not possible.

8.3 Recommendations

In this thesis it was shown that the results obtained with SSDM are much better than cellid positioning and DCM for an urban environment. It would be interesting to test if this is also the case for a rural area. In rural areas cell sizes are larger. This means that the maximum error made by cellid positioning is also larger, again assuming that the reference points are collected evenly spread over the cell, thus that the mobile unit is plotted in the centre of the cell. Rural areas consists of less high buildings and for this the fluctuations of the RSS will be less this will negatively influence GSM fingerprinting.

In this thesis positioning is done taken each point individually, but actually a route was cycled. This means that the position of the previous point can tell something about the current position. A method that uses the position of the previous point needs to be initialized. This means that the position of the first point has to be determined as precise as possible. Because if the positioning error made at the first point is very large, than all succeeding points also have a large positioning error. The result map of the SSDM can probably be used to initialize, because in this map an overview of where the mobile unit can be positioned and also an indication of how likely it is that the mobile unit is positioned there is given. A low value in the result map indicates that it is more likely that the mobile unit is at that position than a high value.

Focus on implementation details of GSM fingerprinting. As we approached the GSM positioning from a geosciences perspective, we did not include any implementation details. All results were derived from post-processing and local computations on our sample dataset for our study area. We experienced that both fingerprinting methods a very computational intensive, which require good configuration and implementation focusing on post-processing of reference databases and client-server connections (i.e. computer science).

We limited ourselves in this thesis to the GSM network. The UMTS network is also a cellular network. For this reason it can be assumed that the same positioning techniques can be applied. Further research should be conducted to verify if this is true.

References

http://craftwww.epfl.ch/research/catchbob/ (accessed March 2009).

http://www.gsmworld.com/technology/gsm/ (accessed December 24, 2009).

www.livecontacts.com (accessed December 24, 2009).

http://www.ieee802.org/ (accessed December 2009, 24).

Ahas, R., J. Laineste, A. Aasa, and Ü. Mark. "The spatial accuracy of mobile positioning: Some experiences with geographical studies in Estonia." *Location based services and telecartography.* 2007. 445 - 460.

Anderson, I., and H. Muller. "Context awareness via GSM signal strength fluctuation." *Conference on pervasive computing.* Dublin, 2006. 27 - 31.

Bahl, Paramvir, and Venkata N. Padmanabahn. "RADAR: An in-building RF-based user location and tracking system." *Infocom*, 2000.

Borkowski, J., and J. Lempiäinen. "Pilot correlation positioning method for urban UMTS networks." *European wireless conference.* Nicosia, Cyprus, 2005.

Brunato, M., R. Battiti. "Statistical learning theory for location fingerprinting in wireless LANs." *Computer Networks* 47, no. 6 (2005): 825 - 845.

Chen, M. Y., et al. "Practical metropolitan-scale positioning for GSM phones." *Ubiquitous computing.* 2006.

DeBlauwe, N. *GSM-based positioning: Techniques and application.* Brussel: VUBPress, 2008.

Gower, J. C. "The mediancentre." *Applied Statistics*, no. 23 (1974): 466 - 470.

Harte, L. *Introduction to Global System for Mobile communication (GSM): Physical channels, logical channels, network, and operation.* Althos Publishing Inc., 2005.

Haykin, S., B. van Veen. *Signals and Systems.* John Wiley & Sons, 1999.

http://pols.sourceforge.net/. Intel Research Seattle. http://pols.sourceforge.net/ (accessed December 26, 2009).

http://www.hydro.nl/pgS/NL/pctrans_nl.htm. http://www.hydro.nl/pgS/NL/pctrans_nl.htm (accessed December 2009, 27).

Juurakko, S., and W. Backman. "Database correlation method with error correction for emergency location." *Wireless personal communications*, no. 30 (2004): 183 - 194.

Karimi, H.A., A. Hammad. *Telegeoinformatics; Location-Based Computing and Services.* Boca Raton, Florida: CRC Press LLC, 2004.

Khalaf-Allah, M., and K. Kyamakya. "Mobile location in GSM networks using database correlation with Bayesian estimation." *Symposium on computers and communications.* Sardinia, 2006.

Laitinen, H., J. Lähteenmäki, and T. Nordström. "Database correlation method for GSM location." *Vehicular Technology Conference.* Rhodes, Greece, 2001.

Li, B., Y. Wang, H. K. Lee, A. Dempster, C. Rizos. "A new method for yielding a database of location fingerprints in WLAN." *IEE proceedings - communications*, October 2005.

Moen, H.L. "A Study of Wi-Fi RFID Tags in Citywide Wireless Networks." Department of Telematics, Norwegian University of Science and Technology, Trondheim, 2007.

Otsason, V., A. Varschavsky, A. LaMarca, and E. de Lara. "Accurate GSM indoor localization." *Conference on Ubiquitous computing.* Tokyo, Japan, 2005.

Polman, J., and M.A. Salzmann. *Handleiding voor de technische werkzaamheden van het Kadaster.* Apeldoorn: Kadaster, 1996.

Takenga, C., T. Peng, and K. Kyamaka. *Conference on Advanced Communication Technology.* Phoenix park, Korea, 2007.

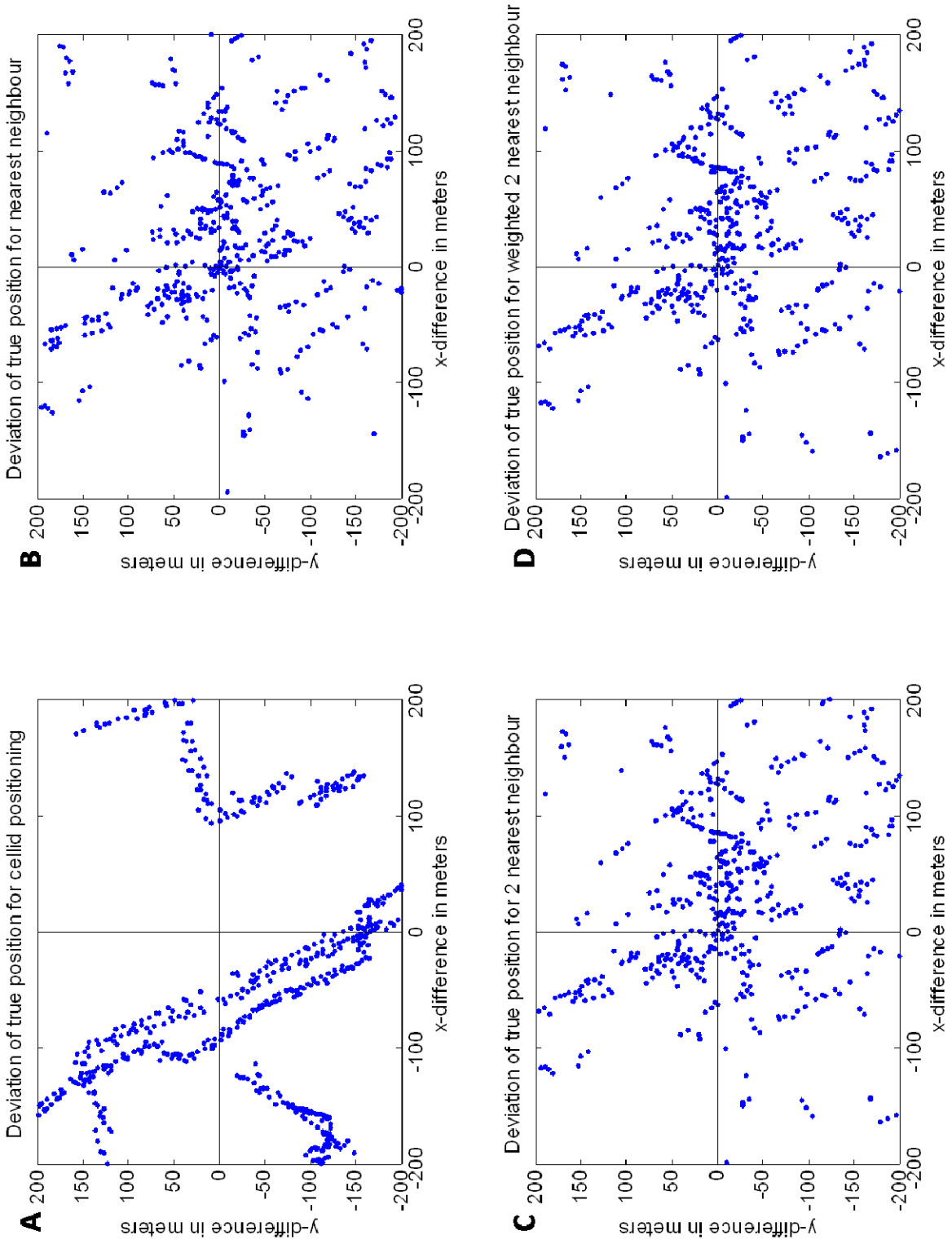
Varshavsky, A., A. LaMarca, J. Hightower, and E. de Lara. "The SkyLoc floor localization system." *Conference on Pervasive Computing and Communications.* New York City, US, 2007.

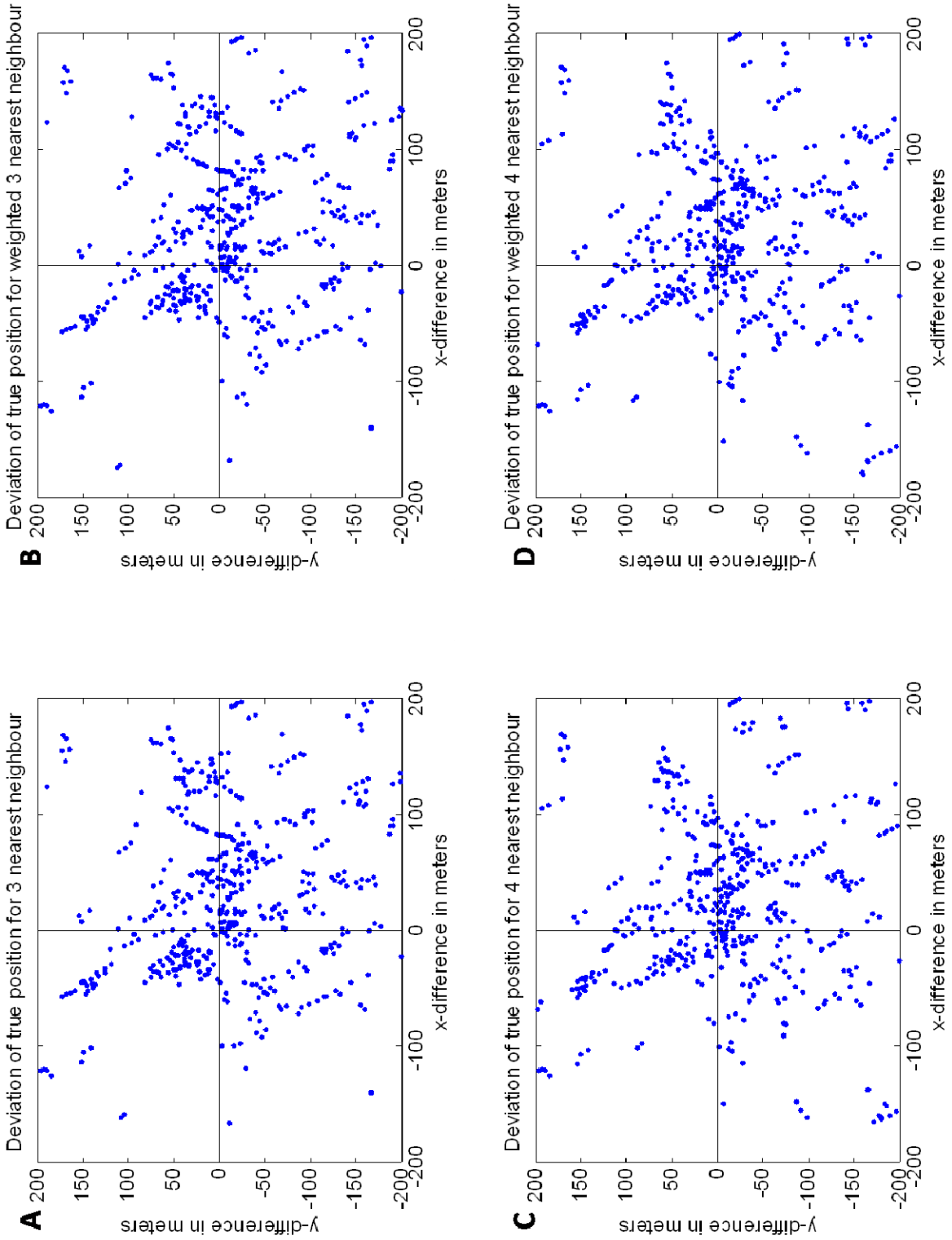
Varshavsky, A., et al. "Are GSM phones THE solution for localization?" *Workshop on Mobile Computing Systems and Applications.* Washington, USA, 2006.

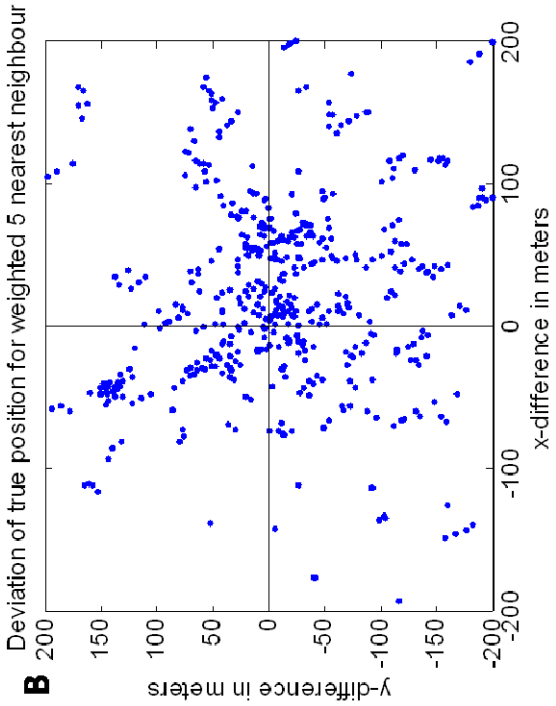
Want, R, A. Hopper, V. Falcao and J. Gibbons. "The Active Badge location system." 1992.

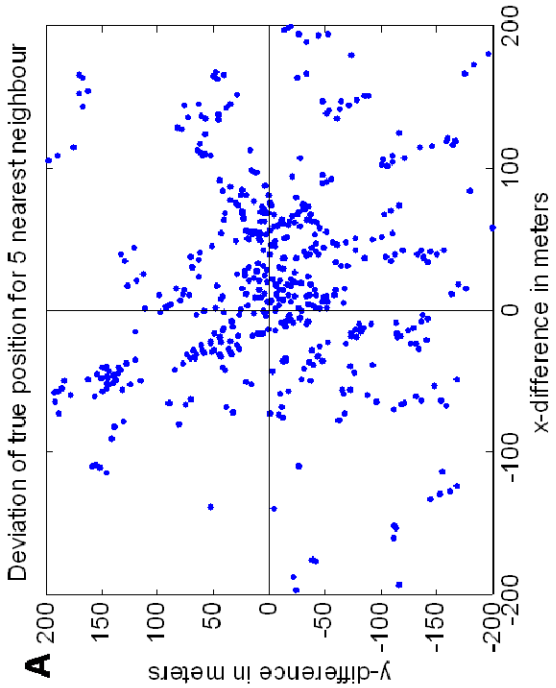
Appendix A: Deviation of true position

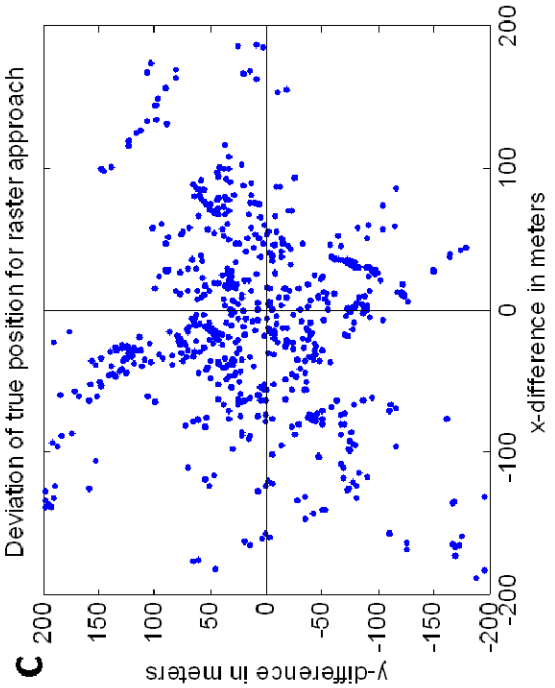
For these figures the number of matching neighbours is not taken into account.




A
Deviation of true position for 3 nearest neighbour
B
Deviation of true position for weighted 3 nearest neighbour
C
Deviation of true position for 4 nearest neighbour
D
Deviation of true position for weighted 4 nearest neighbour
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200

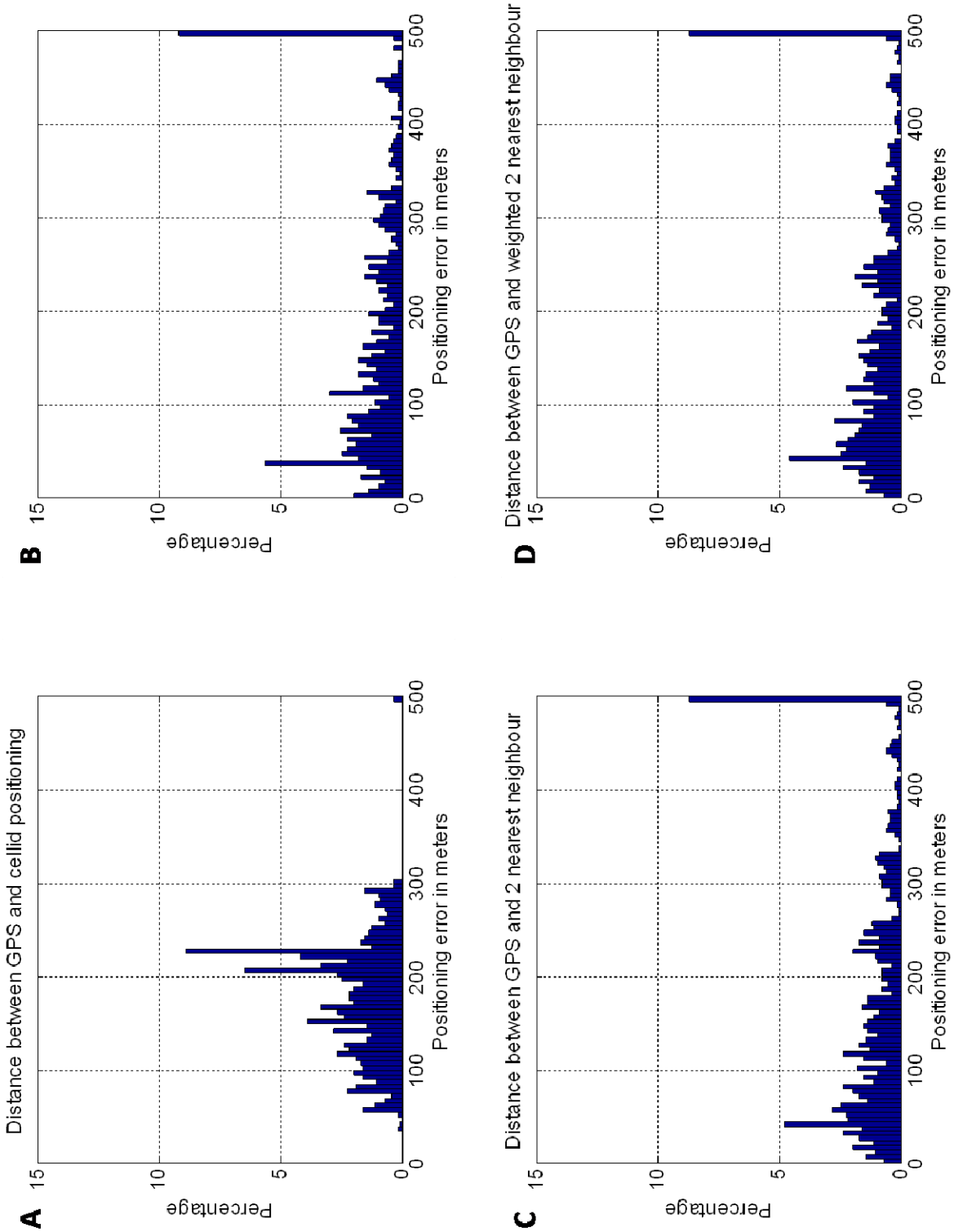

B
Deviation of true position for weighted 5 nearest neighbour
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200
-200
-100
0
100
200

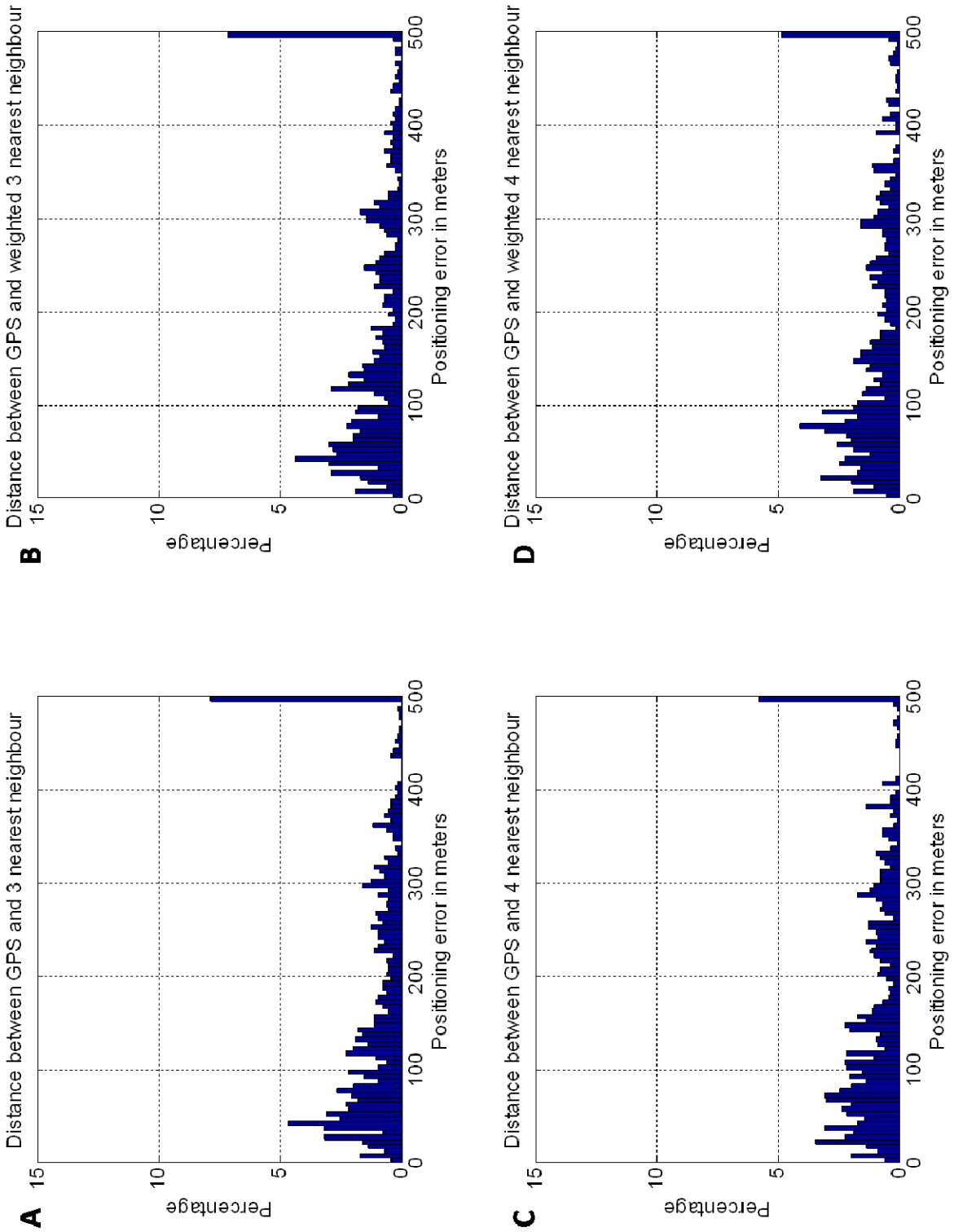

A
Deviation of true position for 5 nearest neighbour
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200
-200
-100
0
100
200

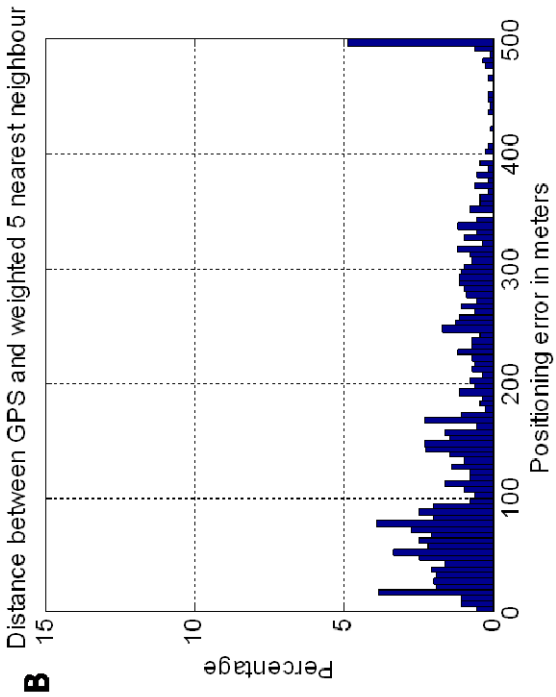

C
Deviation of true position for raster approach
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200
-200
-100
0
100
200

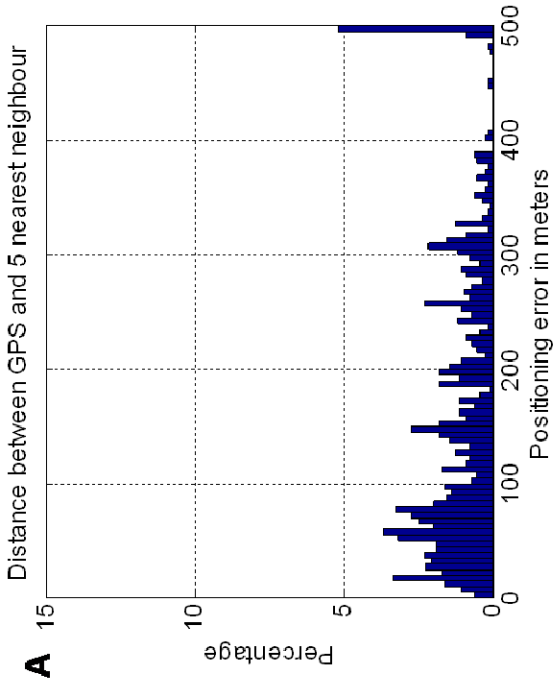
Appendix B: Histogram with positioning error

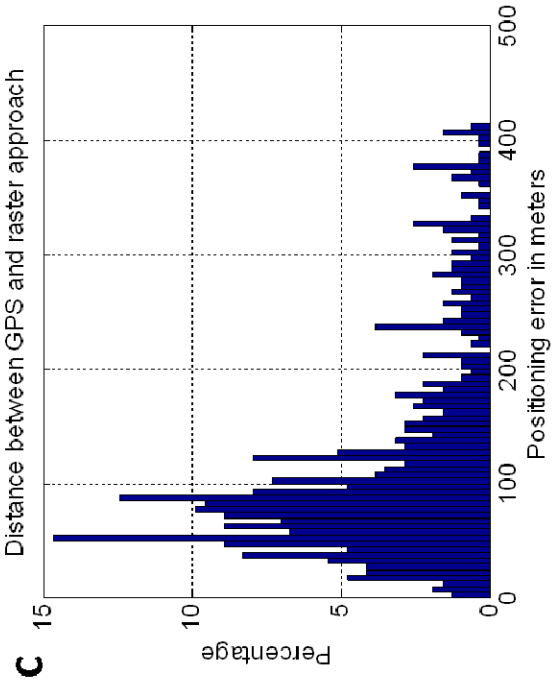
For these figures the number of matching neighbours is not taken into account.




A
Distance between GPS and 3 nearest neighbour
B
Distance between GPS and weighted 3 nearest neighbour
C
Distance between GPS and 4 nearest neighbour
D
Distance between GPS and weighted 4 nearest neighbour
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500

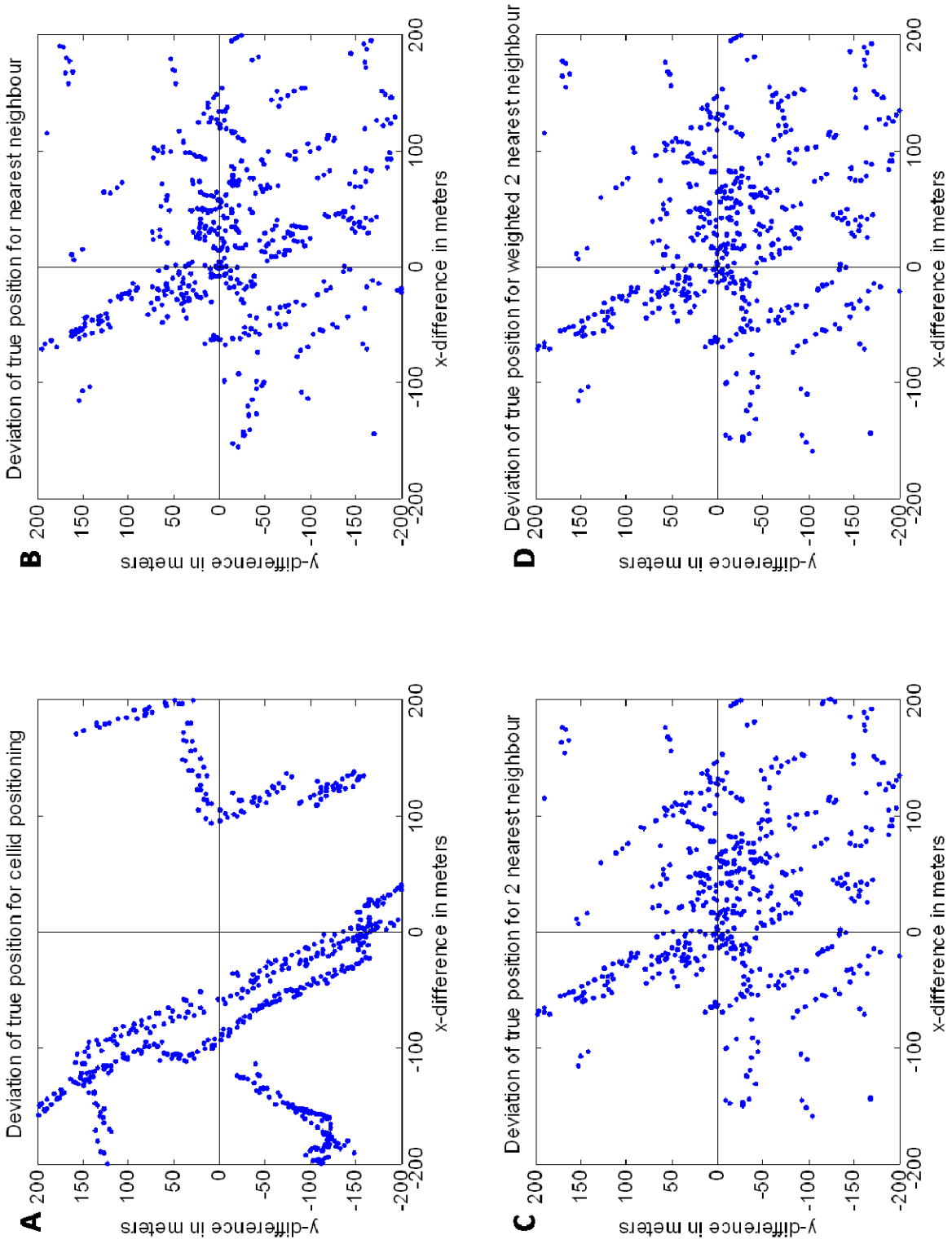

B
Distance between GPS and weighted 5 nearest neighbour
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500

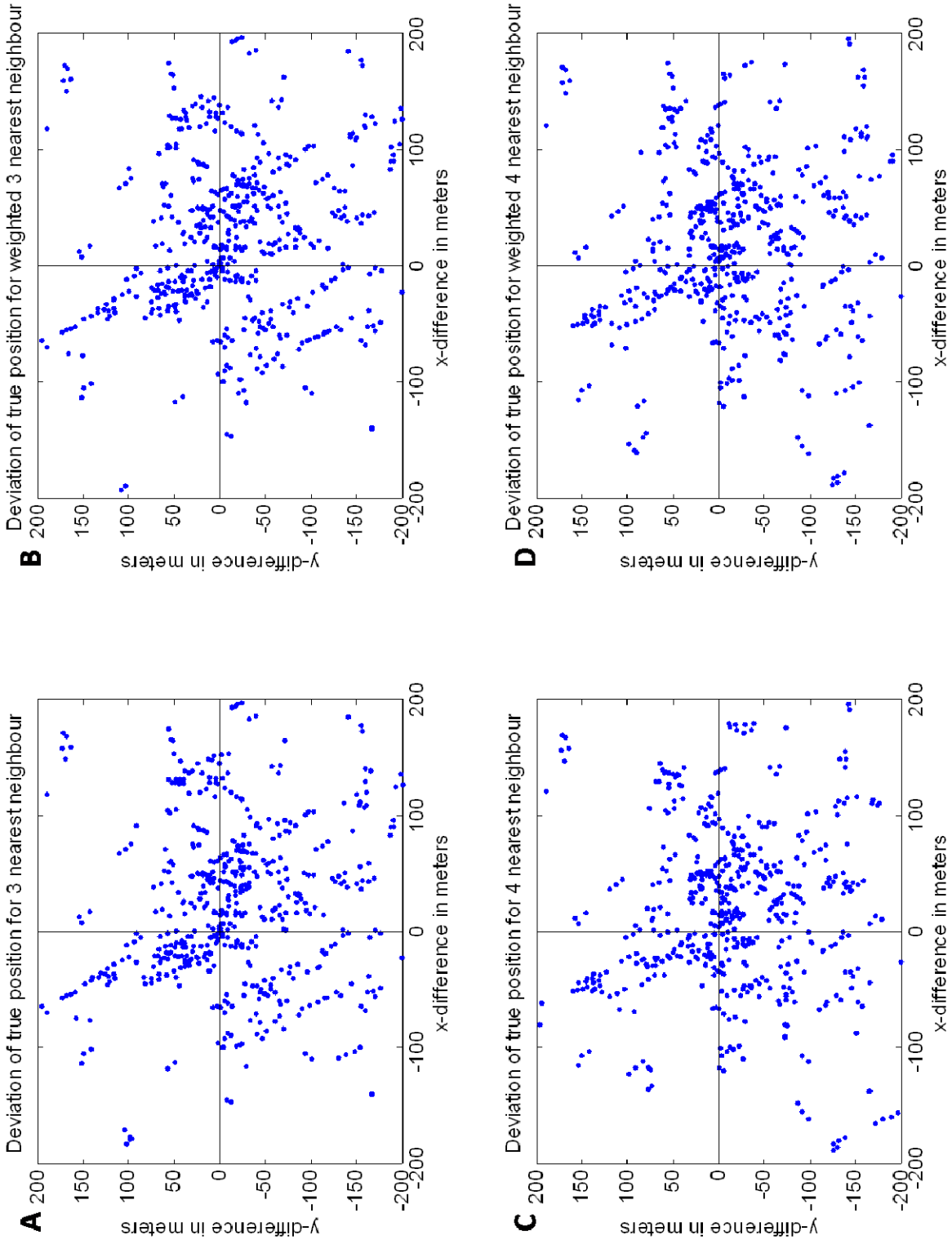

A
Distance between GPS and 5 nearest neighbour
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500

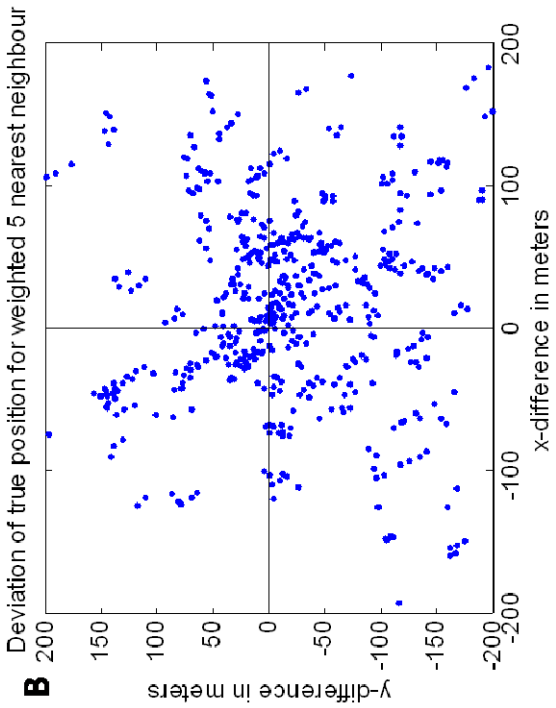

C
Distance between GPS and raster approach
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500

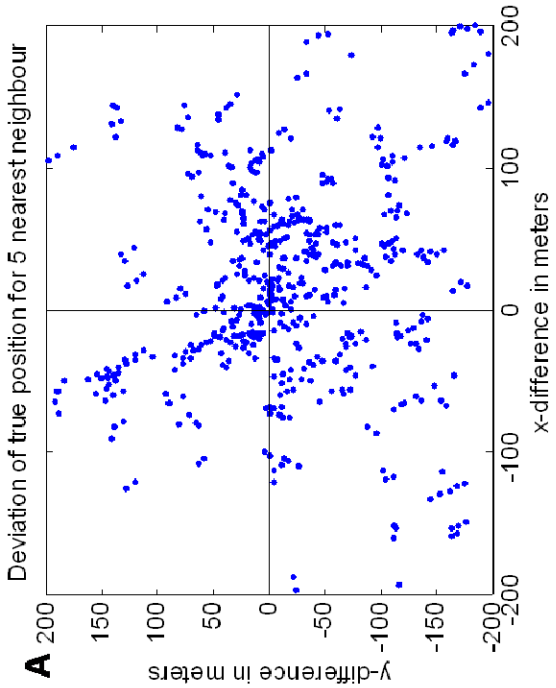
Appendix C: Deviation of true position

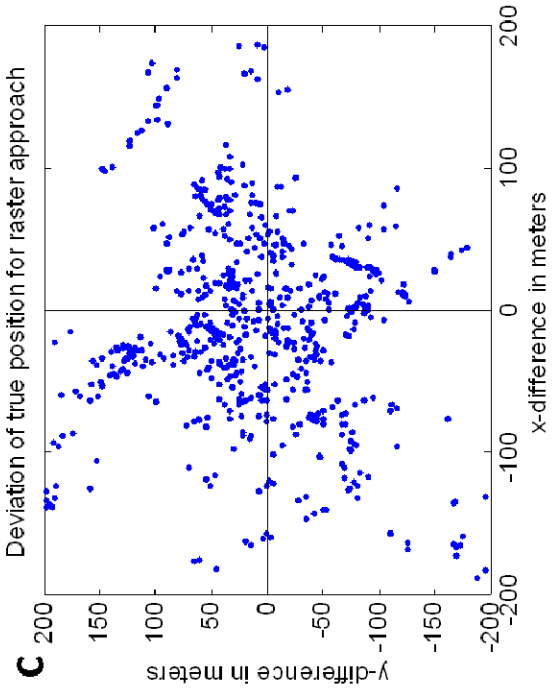
For these figures the number of matching neighbours is taken into account.




A
Deviation of true position for 3 nearest neighbour
B
Deviation of true position for weighted 3 nearest neighbour
C
Deviation of true position for 4 nearest neighbour
D
Deviation of true position for weighted 4 nearest neighbour
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200

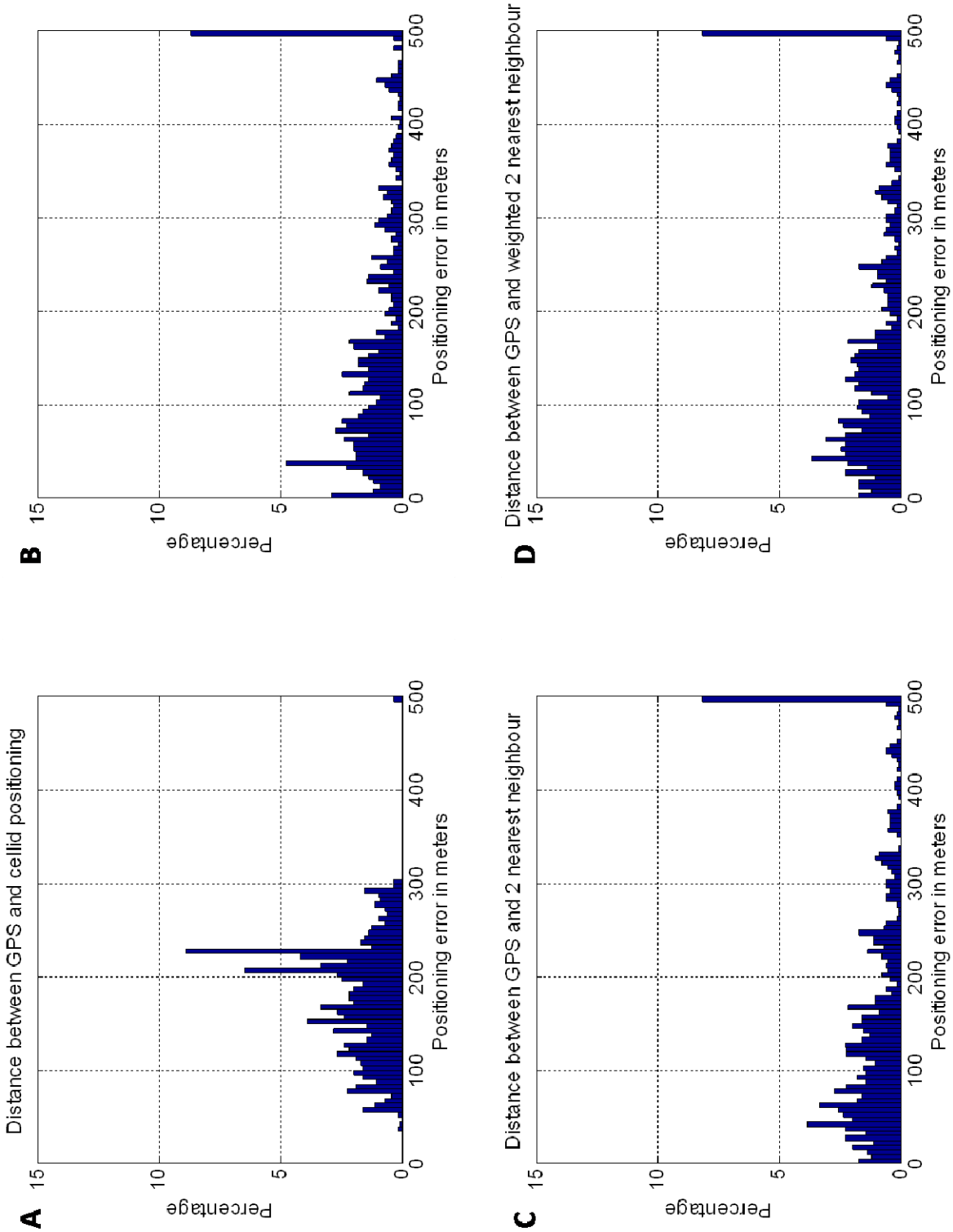

B
Deviation of true position for weighted 5 nearest neighbour
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200
-200
-100
0
100
200

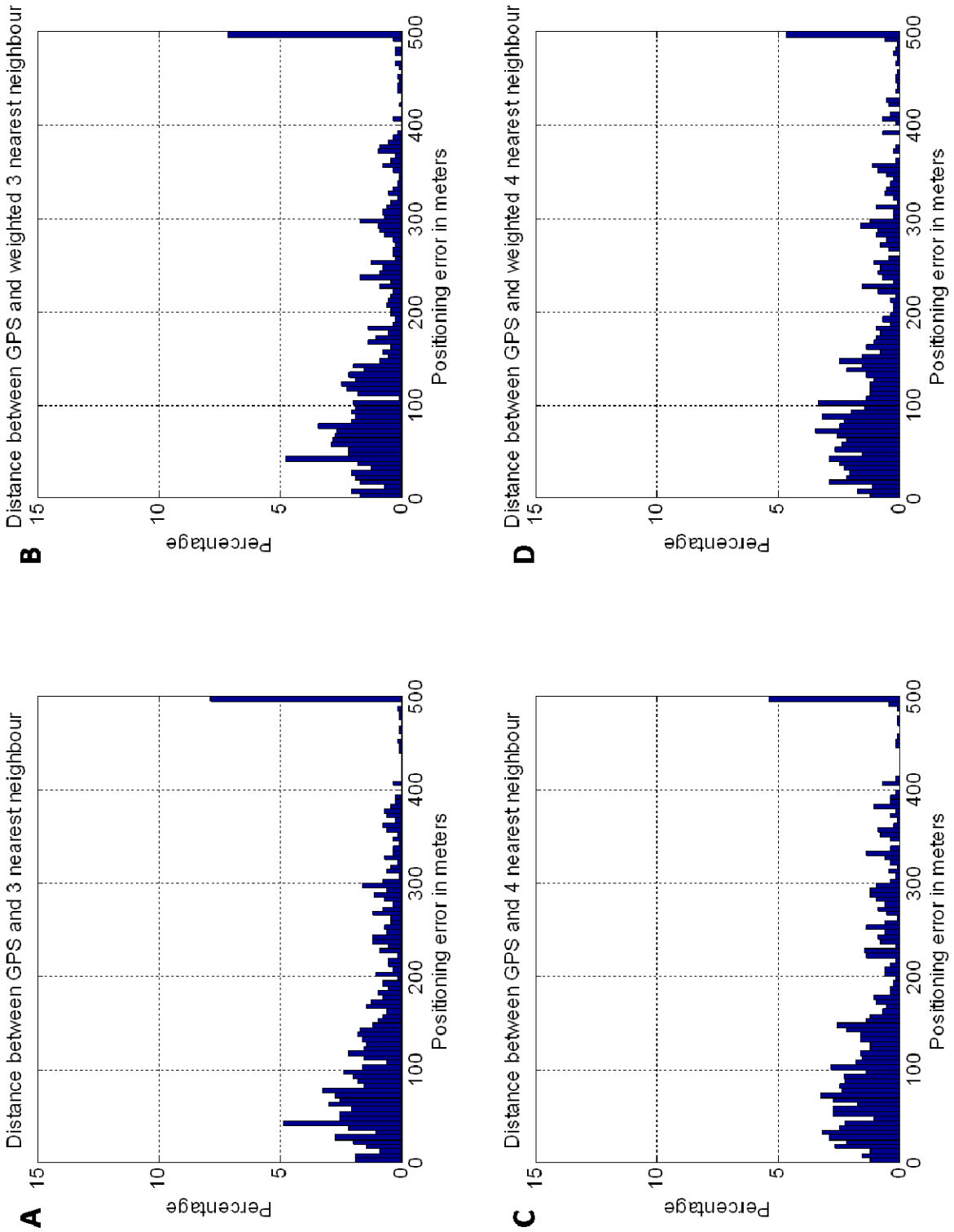

A
Deviation of true position for 5 nearest neighbour
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200
-200
-100
0
100
200

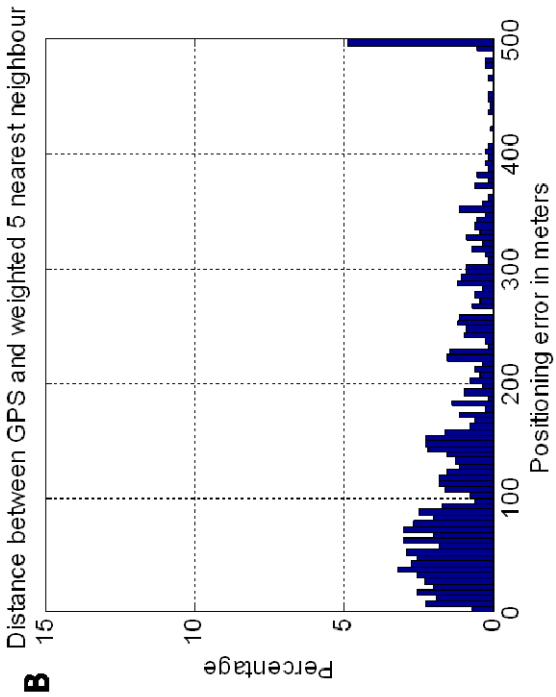

C
Deviation of true position for raster approach
y-difference in meters
x-difference in meters
200
150
100
50
0
-50
-100
-150
-200
-200
-100
0
100
200

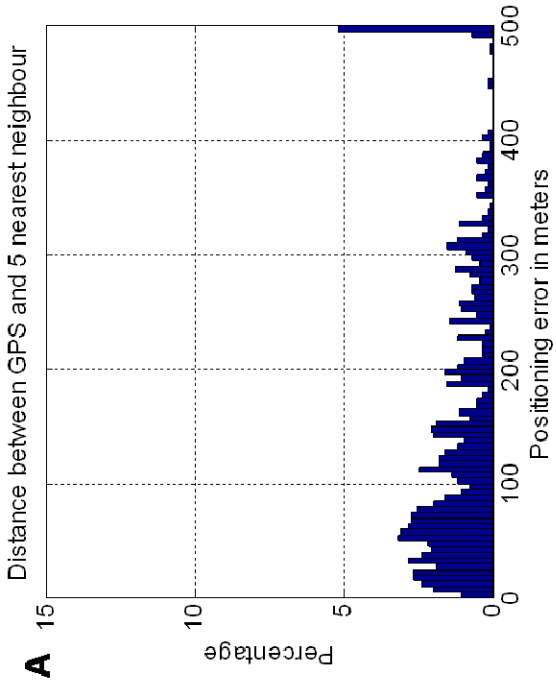
Appendix D: Histogram with positioning error

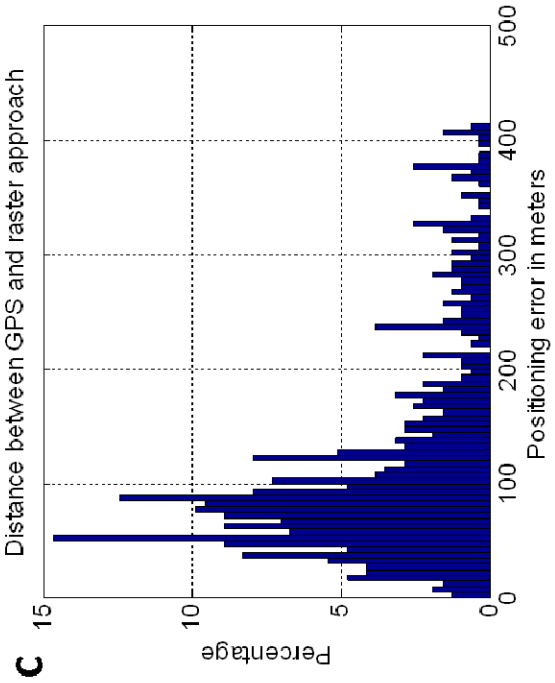
For these figures the number of matching neighbours is taken into account.




A
Distance between GPS and 3 nearest neighbour
B
Distance between GPS and weighted 3 nearest neighbour
C
Distance between GPS and 4 nearest neighbour
D
Distance between GPS and weighted 4 nearest neighbour
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500


B
Distance between GPS and weighted 5 nearest neighbour
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500


A
Distance between GPS and 5 nearest neighbour
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500


C
Distance between GPS and raster approach
Percentage
Positioning error in meters
0
5
10
15
0
100
200
300
400
500

Appendix E: R values of all methods

Method	*50%*	*67%*	*90%*
Cellid positioning	182	211	248
Nearest neighbour	147	235	467
2 nearest neighbour	143	228	465
3 nearest neighbour	131	228	466
4 nearest neighbour	133	224	402
5 nearest neighbour	138	204	380
Weighted 2 nearest neighbour	147	229	351
Weighted 3 nearest neighbour	132	228	466
Weighted 4 nearest neighbour	133	225	391
Weighted 5 nearest neighbour	137	216	359
Signal strength distribution method	88	123	272

Method	*50%*	*67%*	*90%*
Cellid positioning	182	211	248
Nearest neighbour	133	214	454
2 nearest neighbour	127	189	446
3 nearest neighbour	117	183	379
4 nearest neighbour	115	174	362
5 nearest neighbour	116	188	335
Weighted 2 nearest neighbour	130	189	446
Weighted 3 nearest neighbour	118	183	385
Weighted 4 nearest neighbour	113	172	362
Weighted 5 nearest neighbour	118	175	351
Signal strength distribution method	88	123	272