ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA DOPRAVNÍ

Prokop Houda

PERFORMANCE OF MAP MATCHING AND ROUTE TRACKING DEPENDING ON THE QUALITY OF THE GPS DATA

Diplomová práce

2016

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE
Fakulta dopravní
děkan
Konviktská 20, 110 00 Praha 1

K616...Ústav dopravních prostředků

ZADÁNÍ DIPLOMOVÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení studenta (včetně titulů):

Bc. Prokop Houda

Kód studijního programu a studijní obor studenta:

N 3710 – IS – Inteligentní dopravní systémy

Název tématu (česky): **Map matching a route tracking v závislosti na kvalitě GPS dat**

Název tématu (anglicky): Map Matching and Route Tracking Depending on the Quality of the GPS Data

Zásady pro vypracování

Při zpracování diplomové práce se řiďte osnovou uvedenou v následujících bodech:

- Modelujte několik map matching a route tracking algoritmů.
- Zhodnoťte výkonnost map matching a route tracking algoritmů a jejich kombinací v závislosti na GPS datech s různými vstupními chybami a vzorkovacími frekvencemi.
- Analyzujte potenciál využití map matching a route tracking algoritmů na základě zhodnocení jejich odezvy na různě kvalitní GPS data.
- Specifikute jaká je výkonnost map matching algoritmů za použití specifické route tracking evaluační metriky.
- Zaměřte se na online map matching metody.

Rozsah grafických prací: podle pokynů vedoucího diplomové práce

Rozsah průvodní zprávy: minimálně 55 stran textu (včetně obrázků, grafů a tabulek, které jsou součástí průvodní zprávy)

Seznam odborné literatury: D. Jianjun and C. Xiaohong, "An Improved Map-Matching Model for GPS Data with Low Polling Rates to Track Vehicles," in TMEE, Changchun, 2011.

J. Rife and B. Pervan, "Overbounding Revisited: Discrete Error-Distribution Modeling for Safety-Critical GPS Navigation," IEEE Transaction on Aerospace and Electronic Systems, vol. 48, no. 2, April 2012.

Vedoucí diplomové práce: **Ing. Milan Sliacky**

David Gundlegård, MSc.

Datum zadání diplomové práce: **29. června 2015**
(datum prvního zadání této práce, které musí být nejpozději 10 měsíců před datem prvního předpokládaného odevzdání této práce vyplývajícího ze standardní doby studia)

Datum odevzdání diplomové práce: **30. listopadu 2016**
a) datum prvního předpokládaného odevzdání práce vyplývající ze standardní doby studia a z doporučeného časového plánu studia
b) v případě odkladu odevzdání práce následující datum odevzdání práce vyplývající z doporučeného časového plánu studia

L.S.

doc. Ing. Petr Bouchner, Ph.D.
vedoucí
Ústavu dopravních prostředků

prof. Dr. Ing. Miroslav Svítek, dr. h. c.
děkan fakulty

Potvrzuji převzetí zadání diplomové práce.

Bc. Prokop Houda
jméno a podpis studenta

V Praze dne .... 2. června 2016

CZECH TECHNICAL UNIVERSITY IN PRAGUE
Faculty of Transportation Sciences
Dean's office
Konviktská 20, 110 00 Prague 1, Czech Republic

K616..**Department of Vehicle Technology**

MASTER'S THESIS ASSIGNMENT

(PROJECT, WORK OF ART)

Student's name and surname (including degrees):

Bc. Prokop Houda

Code of study programme code and study field of the student:

N 3710 – IS – Intelligent Transport Systems

Theme title (in Czech): **Map matching a route tracking v závislosti na kvalitě GPS dat**

Theme title (in English): Map Matching and Route Tracking Depending on the Quality of the GPS Data

Guides for elaboration

During the elaboration of the master's thesis follow the outline below:

- Model the several map matching and route tracking algorithms.
- Evaluate the performance of map matching and route tracking algorithms and their combinations depending on the GPS data sets with different error rates and sampling frequencies.
- Analyze the potential of using of map matching and route tracking algorithms based on the evaluation of their response to the GPS data sets of different qualities.
- Specify how well the map matching algorithms perform using the route tracking evaluation metric.
- Focus on online map matching methods

Graphical work range: pictures, tables

Accompanying report length: minimum 55 pages

Bibliography:

D. Jianjun and C. Xiaohong, "An Improved Map-Matching Model for GPS Data with Low Polling Rates to Track Vehicles," in TMEE, Changchun, 2011.

J. Rife and B. Pervan, "Overbounding Revisited: Discrete Error-Distribution Modeling for Safety-Critical GPS Navigation," IEEE Transaction on Aerospace and Electronic Systems, vol. 48, no. 2, April 2012.

Master's thesis supervisor: **Ing. Milan Sliacky**
David Gundlegård, MSc.

Date of master's thesis assignment: **June 29, 2015**
(date of the first assignment of this work, that has be minimum of 10 months before the deadline of the theses submission based on the standard duration of the study)

Date of master's thesis submission: **November 30, 2016**
a) date of first anticipated submission of the thesis based on the standard study duration and the recommended study time schedule
b) in case of postponing the submission of the thesis, next submission date results from the recommended time schedule

L. S.

doc. Ing. Petr Bouchner, Ph.D.
head of the Department
of Vehicle Technology

prof. Dr. Ing. Miroslav Svítek, dr. h. c.
dean of the faculty

I confirm assumption of master's thesis assignment.

Bc. Prokop Houda
Student's name and signature

Prague ....June 2, 2016

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o etické přípravě vysokoškolských závěrečných prací.

V Praze, 24.11.2016

Abstrakt

Satelitní navigační měření nejsou nikdy dokonale bezchybná. Kvůli několika typům chyb, které ovlivňují přenos signálu otevřeným prostorem i urbanizovanou krajinou každé jednotlivé měření obsahuje určitou míru nejistoty. Přijímače satelitního signálu zároveň nepřijímají signál spojitě, ale diskrétně. Vzorkovací frekvence a původní poziční chyba vnášejí nejistotu do všech pozičních algoritmů používaných při v lokalizačních, logistických a ITS systémech. Tato práce zkoumá efekt původní poziční chyby a vzorkovací frekvence na geometrické a topologické map matching algoritmy a na přesnost route trackingu v rámci těchto algoritmů. Zároveň jsou hodnoceny i efekty rozličné hustoty dopravní sítě a je vytvořena platforma pro simulaci a evaluaci map matching algoritmů.

Map matching je proces přiřazování původních navigačních měření na síť. Značné množství autorů v posledních dekádách představilo své algoritmy, což dokládá komplexnost celého tématu, především z důvodu měnících se prostředí a dopravních sítí v nichž tyto algoritmy operují. Geometrické a topologické map matching algoritmy jsou vybrány, modelovány a simulovány a jejich reakce na rozličné kombinace vstupů je ohodnocena. Zároveň jsou vzneseny doporučení pro možná ITS využití především v podobě shrnutí požadavků na potenciální přijímač.

Výsledky potvrzují obecný předpoklad, že map matching vylepšuje původní měření a že jej můžeme do jisté míry brát jako formu „error mitigation“. Zároveň je zjištěna univerzální závislost mezi zvýšením původní chyby a chyby po map matchingu pro všechny porovnávané algoritmy. Nicméně porovnání algoritmů také ukázalo velké rozdíly mezi topologickými a geometrickými algoritmy v jejich schopnosti vyrovnat se s zkreslenými vstupními měřeními. Zatímco topologické algoritmy vykazovaly jasně lepší výsledky ve scénářích s menšími původními chybami a menší vzorkovací frekvencí, geometrické algoritmy se ukázaly být efektivnějšími při velkých chybách měření a vyšších frekvencích. To je způsobeno především schopností jednoduššího opuštění chybně namapované pozice ve srovnání s topologickými algoritmy, což se ukázalo jako komparativní výhoda v těchto datových scénářích.

Navazující práce by se měly soustředit na zahrnutí většího množství algoritmů do porovnání, což by přineslo cennější výsledky. Zároveň simulace chyb za použití simulace velikosti chyby se známou distribucí by mohlo být vylepšeno – komplexnější modelování vstupní chyby by zvýšilo obecnou platnost výsledků.

KLÍČOVÁ SLOVA

Map matching, Route tracking, GPS, chyba, vzorkování, hustota

Abstract

Satellite positioning measurements are never perfectly unbiased. Due to multiple types of errors affecting the signal transmission through an open space and urban areas each positioning measurement contains certain degree of uncertainty. Satellite signal receivers also do not receive the signal continuously, but the localization information is received discretely. Sampling rate and positioning error provide uncertainty towards the various positioning algorithms used in localization, logistics and in intelligent transport systems applications. This thesis examines the effect of positioning error and sampling rate on geometric and topological map matching algorithms and on the precision of route tracking within these algorithms. Also the effects of the different network density on the performance of the algorithms are evaluated. It also creates the platform for simulation and evaluation of map matching algorithms.

Map matching is the process of attaching the initial positioning measurement to the network. A number of authors presented their algorithms during past decades, which shows how complex topic the map matching is, mostly due to the changing environmental and network conditions. Geometric and topological map matching algorithms are chosen, modelled and simulated and their response to the different input combinations is evaluated. Also the recommendations for possible ITS applications are carried out in terms of proposed requirements of the receiver.

The results confirm general expectation that the map matching overall improves the initial position error and that map matching serves as a form of error mitigation. Also the correlation between the increase of the original positioning error and the increase of the map matching error is universal for all the algorithms in the thesis. But the comparison of the algorithm also showed large differences between the topological and geometric algorithms and their ability to cope with distorted input data. Whereas topological algorithms were clearly performing better in scenarios with smaller initial error and smaller sampling rate, geometric matching proves to be more effective in heavily distorted or very sparsely sampled data set. That is caused mostly by the ability to easily leave the wrongly mapped position which is in these situations comparative advantage of simple geometric algorithms.

Following work should concentrate on involving even more algorithms into the comparison, which would produce more valuable results. Also the simulation of the errors using the error magnitude simulation with known distribution could be improved - an improved error modelling could increase the generalization of the results.

KEYWORDS

Map matching, Route tracking, GPS, Error, Sampling, Density

Used abbreviations

ABS	Anti-lock brake system
BDS	BeiDou navigation satellite system
C2C	Curve-to-curve
CEP	Circular error probability
CSV	Comma separated values
DRMS	Distance root mean squared
ESA	European space agency
ETRS	European terrestrial reference system
GEO	Geostationary orbit
GIS	Geographic information system
GLONASS	Globalnaja navigacionnaja sputnikovaja sistěma (Russian)
GNSS	Global navigation satellite system
GPS	Global positioning system
GSO	Geosynchronous orbit
HMM	Hidden Markov model
ICAO	International civil aviation organization
IERS	International Earth rotation and reference systems
IR	Intersection rule
IRNSS	Indian regional navigation satellite system
IRP	Intersection rule parameter
ISRO	Indian space research organization
ITS	Intelligent transport system
LAT	Latitude
LiU	Linköping University
LON	Longitude
MEO	Medium Earth orbit
MP	Measurement point
NAD	North American datum
NOAA	National oceanic and atmospheric administration
OSM	Open street map
P2C	Point-to-curve
P2P	Point-to-point
SCR	Stopped car rule
SCRP	Stopped car rule parameter
WGS	World geodetic system

Table of Contents

List of Tables

List of Figures

Acknowledgments

In this chapter I would like to express my deepest gratitude towards my supervisors David Gundlegård from Linkoping University and Milan Sliacky from Czech technical university in Prague for their help during one and a half year of its creation. Without them and their guidance, I cannot imagine how much harder and longer the process would be. In addition, I would like to mention the utmost valuable guidance received from Dr. Carl Henrik Häll, whose expert advices regarding positioning systems and school administration process proved to be extremely helpful. I would also like to thank both universities for allowing me to participate in such a wonderful and inspiring double degree program.

Největší poděkování ovšem patří mým blízkým, kteří mi pomáhali v průběhu celých šesti let mého studia. Jedná se o mé dobré přátele převážně z řad spolužáků, bývalých spolužáků a fotbalových spoluhráčů, kteří mi připomínali, že nejen studiem živ je člověk. Největší inspiraci i oporu jsem ovšem nacházel v rodině. Především chci poděkovat svým rodičům, díky jejichž neutuchající duševní i materiální podpoře v průběhu celého studia jsem mohl sepsat tuto práci a stát se vysokoškolsky vzdělaným člověkem. Nic z toho by navíc nebylo možné bez idylického domova plného respektu a lásky, který mi společně poskytli od mého dětství až doteď a který mě formoval v člověka jakým jsem dnes.

1. INTRODUCTION

1.1. BACKGROUND

Global navigation satellite system (GNSS) is a global positioning system providing precise real-time navigation using continuous satellite signal emission that is received by user on or near the Earth's surface. Even though it is used in plenty of areas, namely in telecommunications, agriculture, mining industry or environmental studies, this thesis focuses on the road traffic applications.

Basically all the road traffic positioning applications need to estimate, using the knowledge about the road network, the real position of the user on the predefined road link from somehow biased and distorted measurements. The process of attaching the measurements to a map is called map matching. Several techniques of map matching are used in the positioning applications, yet there is no generally accepted "correct" solution to this problem. The reason for that is mostly the variety of received signal parameters, namely positioning error and sampling rate. But the received signal is also largely different in urban and rural areas (mostly due to multipathing) and lastly – the density of the road network plays the huge role in the precision of map matching computations.

1.2. PURPOSE AND OBJECTIVES

The thesis aims to compare the characteristics of several map matching techniques in different situations with different quality of received data. The biggest focus is put on online algorithms (performing the mapping in real time), as only one offline version is evaluated. It also aims to analyse, based on that behaviour, the strengths and limitations of each examined approach. It does not have the ambition to find "the best" universal map matching algorithm for all the possible applications.

The aim of route tracking and optimizing problem examination is to specify how well the map matching algorithms perform using the specific metric for results. With regards to these results, the aim of the thesis is also to provide recommendations for some of the applications in terms of the GPS device receiver and map matching technique.

1.3. METHOD

As the name of the work suggests, GPS, as the most widely used GNSS is evaluated. More precisely the degree to which the quality of the GPS data affects the behaviour of positioning devices and positioning and optimizing algorithms is studied. Quality of the GPS data set was represented by two main parameters – positioning error and sampling rate, but also the density of the road network and its effects are examined.

Comparison of the algorithms based on their quality and effectiveness is achieved firstly by modelling several variations of noisy data set, link network and compared algorithms (mostly in MATLAB software) and then by the evaluation of the results. The algorithms of three geometric and four topological map matching methods are evaluated in 7 different network density scenarios in rural, suburban and urban areas. Each of these scenarios is distorted with initial positioning errors of different magnitudes ranging from error mean of 1 to 30 meters. Sampling rate is simulated in range between 1 second and 150 seconds. Several metrics were used for the evaluation of the results. The standard positioning metrics described purely geometrical/geographical performance, expressing the distance of mapped position from the true position.

The route tracking problem is described by different performance metric, through the calculation of percentage of measurements mapped on the correct link. These results play more important role for most of the route tracking and optimizing applications compared with mere distance from the true position. The possible real world route tracking and optimizing applications were described and recommendations for the quality of their GPS signal receiver and used map matching algorithm based on the upper mention techniques were carried out.

1.4. OUTLINE

The thesis is structured as follows: Chapter 2 consists of the literature review that explains the global positioning systems, maps and road networks and specifies the terminology. It also describes the different known approaches towards the map matching. Chapter 3 covers the system description. The data processing is described as well as the implementation of simulated map matching algorithms. The performance metrics is specified. Chapter 4 evaluates and analyses the results of the simulations. Chapter 5 covers the conclusions and finally, chapter 6 presents the discussion regarding the results and their meaning and recommendations for possible future work.

2. THEORETICAL BACKGROUND

2.1. GLOBAL POSITIONING SYSTEMS

2.1.1. OVERVIEW

Global Navigation Satellite Systems (GNSS) are positioning systems using three segments (space, control and user) to determine the user position with a precision up to several centimetres. It is not the purpose of this work to thoroughly explain the GNSS history, specifics and qualities, nevertheless, at least a brief overview of the systems is necessary for understanding the following work.

Space segment of GNSS systems consists of a set of satellites orbiting the Earth. The number of satellites is different for different systems – GPS uses, by the time of this thesis making (March 2016), 31, GLONASS uses 24, Galileo plans to have 30 satellites and BeiDou 35. In general, all the systems need their satellites orbiting in high altitudes in order to cover most of the globe. The GPS, GLONASS and Galileo are using MEO, BeiDou MEO in combination with GSO and IRNSS GEO and GSO in their attempt to cover most of the world or the important parts of the world.

These satellites are constantly and continuously emitting electromagnetic signal that is received by the user (receiver). Since the signal has a time tag and is travelling at the speed of light, the receiver is able to tell his distance from the satellite. From the geometrical point of view, that knowledge allows him to determine the sphere circumscribed around the satellite on the surface of which the user lies. In combination with the signal from the second satellite is received, receiver can determine that his position is on the intersection of the two spheres – circle. The third satellite determines two possible points of location (as three spheres intersect in two points) and the fourth satellite determines the only possible location.

Despite of the fact that Global Positioning System (GPS) is in the media and in general public sometimes viewed as the synonym for satellite navigation, other space agencies and/or military organizations also run or will run their own navigation systems. On a global scale, mostly Russian GLONASS is used beside of GPS. Or, to be more precise, GLONASS is very often used in cooperation with GPS in civil applications and that cooperation has remarkable success, as is shown i.e. by Defraigne and Hermegnies [1] or by Kovar et al.[2]. But with GPS upgrading its satellites to L5 signals [3] and with European Galileo also signalling in L5 [4], GLONASS future for civil global use is questionable (GLONASS has not expressed a desire to upgrade to L5 signal).

Galileo is a civil project of European Space Agency (ESA) planned for civil applications and civil use[5] which makes it different from the two upper mentioned systems that were constructed and deployed as military systems and serve primarily military purpose[6][7]. Now (March 2016) 10 satellites are in orbit [5] and since ESA is able to deploy two satellites with one rocket (and 4 in the future with Ariane 5 rocket [8]), Galileo is currently on the fast pace towards being functional.

Another global positional system is being built in China and is called BeiDou Navigation Satellite System (BDS). In fact, the first version of BeiDou called BeiDou-1 is already functional (since 2011), but it cannot be considered global, since it is functional only in south-eastern Asia [9]. The global system is supposed to be functional in the early 2020s [10].

The different approach from the GNSS systems mentioned above was chosen by Indian Space Research Organization (ISRO). Their system called Indian Regional Navigation Satellite System (IRNSS) uses satellites orbiting in higher altitudes (after completion of deployment phase 3 satellites in GEO and 4 in GSO will all be approximately in 36 000 km above the surface). The geometry of the orbits will allow IRNSS to precisely locate users in the area of Indian mainland and approximately 1500 kilometres around India. That makes this satellite system regional; therefore it cannot be classified as GNSS. But it is mentioned because it is using L5 signal band and is compatible with Galileo and GPS (as described in article by Majithiya et. al [11].

2.1.2. Performance of GNSS systems

The task of expressing the performance of various GNSS is not trivial, since the performance varies it time due to the satellite configuration, errors occurring due to multiple reasons (see chapter 2.1.4) that cannot be, in most cases, predicted and, of course, due to the characteristics of the receiver. That is why it is impossible to tell the exact precision of positioning – even the precision considering the ideal conditions and ideal receiver needs to be expressed as a range or on a certain confidence level.

On their official website, Russian federal space agency shows the precision of navigation at chosen locations at any given day using GLONASS and also the comparison to GPS at the confidence level of 95% [12]. The horizontal precision is to the day of the creation of this work ranging from 4.6m to 8.5m and vertical precision is between 11.8m and 23.9m. By comparing these values with values achieved with GPS data we can see that the GPS provides better results at these given locations (all in Russia), but not larger by the orders of magnitude – both horizontal and vertical precision are better by approximately 1-2 meters[12]. That is caused mostly by the number of reachable satellites that tend to be by the day larger by 2-3 compared to GLONASS.

GPS precise positioning standard [13] specifies the global average of precision at confidence level of 95% at less than 5.9m with dual-frequency code and less than 6.3m with single frequency code. That corresponds well with the upper mentioned GLONASS comparison.

The precision of other mentioned systems cannot be measured since they are not yet fully operational. According to the official documents [5], Galileo aims to be more precise than public GPS (with precision of 1m), which, if achieved, could certainly provoke the big changes regarding the positioning applications market. BeiDou's goal presented on the official website is 10m accuracy [14].

2.1.3. Performance of GNSS applications

The performance of the GNSS application often needs to be quantified, although it is not the simple thing to achieve. This thesis divides the performance characteristics according to the official ICAO document "Global Navigation Satellite System (GNSS) Manual" [15]. That means the performance characteristics are divided to four categories: accuracy, integrity, continuity and availability. Different users and different applications have different needs regarding these characteristics, so it is not possible to distinguish which one of them is more important in general.

Accuracy is perhaps the most intuitive of the four, since it expresses the actual difference of measured and real position in units of distance [15]. And since the modelling done in this work have been only on the algorithmic level, without any actual hardware or service testing, it was the main navigational performance metric for determining the quality of tested positioning algorithms.

Integrity expresses the level of trust that user can place into the correctness of the provided information. That characteristic may be then used for informing the user whether the provided position is to be trusted or not (in cases of exceeding the set level of integrity) [15]. Continuity (as the name suggests) expresses the system's ability to continuously provide its function [15]. That is being described in the terms of probability, whereas availability, the time system is working with required accuracy, integrity and continuity, are expressed in the time variables.

2.1.4. Errors

Sources

GNSS errors have several sources. Generally accepted division used i.e. in He's et al. article [16] is the division into the following categories:

- Satellite orbital prediction errors
- Satellite clock bias
- Ionospheric refraction
- Tropospheric refraction
- Signal multipath
- Receiver clock offset
- Receiver noise error

First four of the upper mentioned sources may be mitigated by differencing or heavily reduced by modelling and almost the same applies for clock offset. What is really challenging and must be taken into consideration in every GNSS system is the multipath propagation that is dependent on the receiver surrounding environment and therefore cannot be mitigated by simple solution applicable in the same manner for all the measurements. Signal multipaths when it is reflected by big solid objects in a way between the receiver and satellite. In that case signal takes less direct path taking more time to reach receiver and causes error.

Distribution
The thesis aims to examine and explore the effects of sampling rate, error rate and network density to the effectiveness of the several map matching and route tracking algorithms by error and sampling rate simulation applied to the reference data. Therefore it is necessary and very important to choose correct distribution of the error of the GNSS measurement.

Most of the other authors who have done somewhat similar work used normal distribution for their simulations of the spatial measurement error [17][18][19][20], but some of them (such as Goh et al.[18]) in the same time agree that normal distribution is the simplification of the problem and that error distribution is in reality non Gaussian.

The real distribution of error is usually described as Rayleigh distribution, or generally Weibull distribution (since Rayleigh is a specific case of Weibull distribution). That is quite nicely shown and described in research made by University of Sao Paulo [21], where authors back that statement up by their numerous measurements. But Rayleigh (or Weibull) distribution is not universally accepted as the only correct distribution and some studies suggest that there might be different distributions that are more fitted for position error of GNSS. One example of a study of that kind is the Ted Driver's paper [22] that shows Gamma distribution as the best fit (ahead of Rayleigh).

Due to these discrepancies indicating three distributions (four if we count Weibull separately) as suitable for positioning error estimation, own independent research and measurements were decided to be conducted by a mobile GPS receiver as a part of the thesis. Results of the research were then used in the simulation as well as theoretically suitable distributions (chapter 4.2). Detailed description of the error distribution measurement can be found in chapter 3.2.3 together whereas general explanation of distribution used for simulations.

2.1.5. Sampling Rate

The term sampling rate (or polling rate i.e. in [23]) can be defined as a time difference between the positioning measurements. The term sampling rate was chosen for the purposes of the work, since that allows the usage of time units that are more illustrational compared with units of frequency (Hz). Sampling rate was chosen despite of the fact that the vast majority of related papers (such as [18]) describe the frequency in positioning in terms of seconds – meaning that the frequency in the true sense of the word is not used.

Since one of the main goals of the thesis is to examine the effects of rate on the map matching precision, the sampling rate was simulated in fairly large interval, ranging from a second to hundreds of seconds. The process of sampling rate simulation is described thoroughly in chapter 3.2.4.

2.2. Maps and road network models

2.2.1. Overview

The crucial part of almost any positioning related work is projecting the results of raw measurements and existing physical transport network (or any other relevant network) to useful and somehow workable information. That is being done by projecting three dimensional information about the location to a two dimensional plane – map. Map is then used not only for mere displaying of measurements, but in the vast majority of transport application also for positioning enhancements known as map matching (i.e. [24],[23][25])– the positions are then considered as 2D points for the rest of the positioning process.

That obviously makes map precision very important, since every error made in the measurement phase would be magnified by map matching calculations if these were to be done on the wrongly chosen map. The necessary thing to realize is that there is no completely precise way of map construction, since due to dimension reduction some part of the information is always lost. That is why multiple projections that are useful for different parts of world and for different types of areas exist.

It is not the purpose of this work to examine in much detail such a complex topic, but since the map matching is the core of the thesis, at least a brief introduction to the map projections and geodetic systems is necessary in order to understand the modelling work done in the next chapter.

2.2.2. Geodetic systems

The Earth is often simplified to the round shape. That is what all of us learn at elementary school – the round Earth, together with 7 other round planets, orbits the round Sun. But in the same way the simplified version of elementary school physics is not sufficient for solving the problems in the quantum micro-world, the round Earth is not sufficient for advanced geodesy. Firstly – the Earth is, mostly due to its rotation, oblate spheroid and the difference between the radius at the equator and at the pole is approximately 21 kilometres in the most widely used geodetic model WGS 84 [26]

But that would not be enough for the precise navigation and positioning, so other actions and assumptions need to be taken in order to produce the geodetic system. The oblate spheroid that has specified nominal level of ocean surface is called the geoid and it is the enhanced version of the upper mentioned oblate spheroid. The geoid is shaped in such a manner that the oceans would take if only the rotational forces and gravitation were the influential factors in shaping their surface, as described i.e. by NOAA at their website [27]. Therefore all the other and rather unstable or not so easily predictable factors such as tidal forces or weather are not considered.

Obviously, only the shape of the geoid would not be sufficient for any projection and positioning activities, since there is a need for a coordinate system stating the origin of the datum and the axis. Usually, the Earth's centre of mass is the origin of the coordinates [26]. The geoid is then defined by meridians and parallels. Meridian is a line of the same longitude connecting the Earth's poles, whereas parallel is a line of a constant latitude that is parallel to Earth's equator [26]. The prime meridian also needs to be set as a reference meridian with the longitude of 0 degrees. The normalized prime meridian has for hundreds of years been the Greenwich meridian that passes through the Royal Observatory Greenwich in London [28], even though it had been shifted a bit by International Earth Rotation and Reference Systems

Service (by 5.3 arcseconds, which is approximately 100 meters east [29]) and that IERS Reference Meridian is used in the most of the geodetic systems.

Most widely used geodetic system is called WGS 84. It has been developed in the United States and is used in GPS [13]. It is used for the global applications, but the smaller area is explored, the more specific datum is usually used and that is why many different geodetic systems exist around the world (such as NAD 83 [30] for the North America or ETRS89 [31] for Europe). Despite of the fact that the model used in this thesis is evaluating the data from Sweden, the WGS 84 is used, since the map source (OpenStreetMap [32]) uses this datum.

2.2.3. Coordinates

The coordinates that are used in the vast majority of geodetic and positioning systems and in cartography are called longitude and latitude. Longitude is being defined as an angle between the plane containing the prime meridian and a plane containing the poles and the measured location [33], whereas latitude is the angle between the plane containing the centre of the Earth that is orthogonal to the rotation axis of the Earth (equatorial plane) and the line containing the measured point and the centre of the Earth [34]. The longitude and latitude are shown in the figure below.

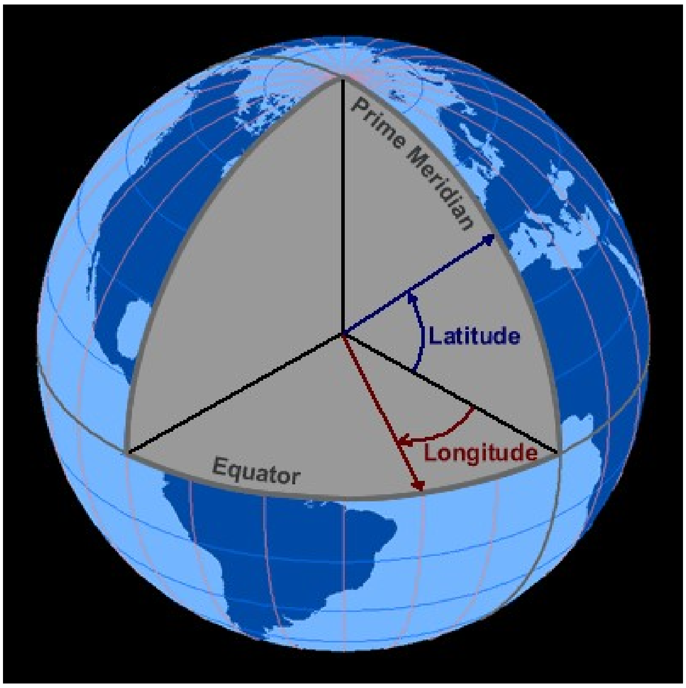


Figure 1: Longitude and Latitude, image courtesy of Dennis Ward/UCAR

The longitude and latitude coordinates are very useful in the sense of covering all of the Earth's surface with just two coordinates, but that is being achieved by using angles and therefore the units are not linear in the distance sense (1 degree of longitude is not the same distance at different latitudes) which makes it difficult for distance calculations. And that is one of the reasons for the usage of projections to 2D plane.

2.2.4. Projections

As was mentioned before in this paper, the projection of Earth's surface to the two dimensional plane is called map and it is the crucial part of geodesy and most of the positioning systems. To simplify the explanation of projection it is possible to imagine the plane that will serve as a map attached to the geodetic system's datum sphere and the points of the sphere projected perpendicularly to it. The way how and where the plane is attached influences the magnitude and manner of the distortion after the plane is unfolded.

The basic projection division (that is used for example in a GIS and Map collection tutorial of the Ball State University [35]) is the division to projections by presentation of a metric property and projections created from different surfaces. The most used projections of a metric property are gnomonic and various compromise projections. Gnomonic projection is projected from the Earth's centre to a tangent plane. It is used mostly for naval navigation since it correctly displays the distance (the shortest path between two points is the corresponding with the reality). All the other distortions increase as the distance from the centre point increases.

But the most relevant projections for the purposes of this thesis are the ones created from different surfaces. The notoriously known are cylindrical, azimuthal and conic projections. Conic projections use cone as a surface of the plane. Cone can be placed either as a tangent to the one chosen parallel (that parallel is then not distorted) or as intersecting the Earth's surface at two chosen parallels that are non-distorted. With increasing distance from these tangential/intersection parallels the distortion increases as well. The meridians are straight lines, crossing the parallels (arcs of a circle) at right angle. The conical projections are used mostly for regions far from the equator with bigger east-western than south-northern size, such as Alaska or parts of the Russia and Europe [35].

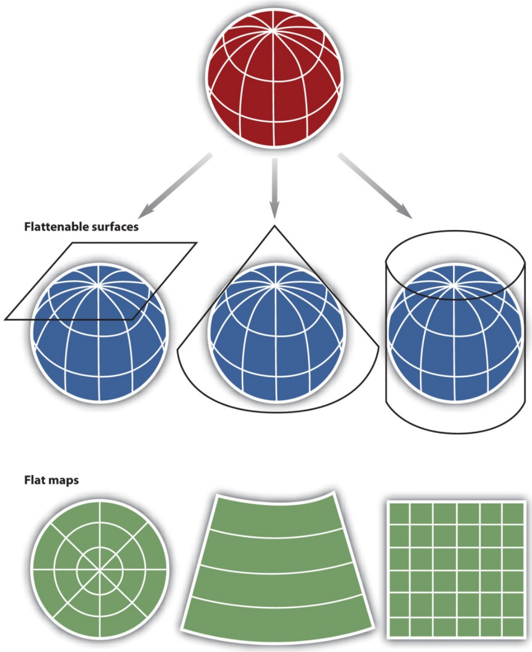


Figure 2: Azimuthal, Conic and Cylindrical projections (from left to right) [36]

Azimuthal projections are used mostly for displaying important point of interest in the centre of the area, since they are constructed by putting the plane of the map on one particular point that is then undistorted and distortions then gradually increases with are receding from that point. It is often used for displaying of a city in the context of its surroundings or for displaying of the Earth's poles [35].

Cylindrical projections use surface of a cylinder as a projection plane and are most widely used for displaying the whole world, but not only for that. The general idea of the undistorted line is the same as for the conic projection – the undistorted line is the tangent line or the intersecting line, but the difference is that distortion is then spread equally in both directions. The undistorted line may be the equator (Mercator projection), meridian (Transverse Mercator) or any diagonal lines with respect to north (Oblique Mercator).

The projection used in the thesis is the Mercator projection with undistorted equator. Even though Sweden is more precisely mapped by Transverse Mercator projection (due to its long south-northern shape), the thesis works with data from smaller areas and, most importantly, map data from OpenStreetMaps[32], use that projection.

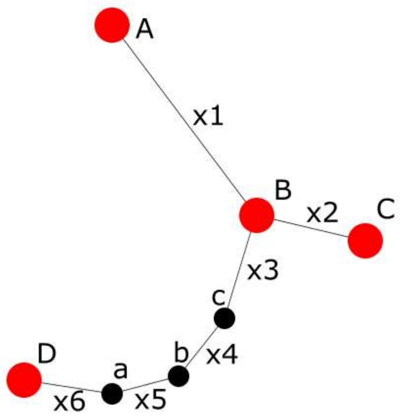
2.2.5. Route, Road network, topology and density

As mentioned in chapters above, maps simplify the complex reality into the two dimensional world through map projections. But the outcome of the projection is still too complex for any meaningful road network analysis and modelling, including map matching. In order to simplify the topological relations vector representations of spatial objects are constructed.

This subchapter will provide the basic terminology used in this thesis with regards to topology and road network. It is necessary to specify the terminology, since the unexplained usage of the words like point, node, arc etc. might lead to confusions. The terminology used in this work is generally based on Alex Kupper's terminology explained in his book [37].

Map point is the simplest element of the network and represents the place where at least one segment ends.
Segment is a straight line connecting two points that represents straight part of the road.
Node is a point where at least two links intersect and where the user makes a route choice or a point where at least one link ends.
Link is a set of segments connecting two nodes. Can be viewed as a polyline.
Route is a set of links that the user travels on.

Example of topological structures is shown in Figure 3 below



Nodes: {A, B, C, D}

Map points: {A, B, C, D, a, b, c}

Segments: {x1, x2, x3, x4, x5, x6}

Links: {x1, x2, [x3, x4, x5, x6]}

Figure 3: Topology

The density of the road network is (together with positioning error and sampling rate) the examined parameter influencing the map matching performance. It is easy to understand, as the dependency between the density of the network and the quality of the matching is almost instinctive, since we can, even without any quantification, say that the sparser the network is, the fewer possibilities for mapping there are and therefore the mapping is easier and in general more precise.

Nevertheless, for any kind of analysis, the vague term "density of the road network" needs to be specified and quantified, since the term itself has been used for a number of different purposes. In road traffic analysis, the term has been usually used as the length of the roads in the examined area. That can be seen for example in European Road Statistics created by European Union Road Federation [38]. Authors working with networks and their models in general usually use more "node-oriented" definition. Rosenblatt uses the definition based on the percentage of the realized connection compared with potential connections [39] (and therefore describes the network density as a degree to which the network is connecting its nodes). Eroglu et al [40] use much simpler method of describing the network in terms of number of nodes per area.

The method of network's density description chosen for the purposes of this work combines Eroglu's et al approach with the one described by Yang et al. [41] present in their paper the network density definition using dynamic density. The dynamic density is obtained by measuring the number of nodes and edges within the measured distance from each position of the route, which is providing much more valuable information compared with the mere measurement of the nodes and edges in the whole map. Yang et al. [41] are using the comparison between the number of edges and nodes as a metric for density description (which is not different from the Rosenblatt's description [39]), but that is not suitable for the purpose of this work, since it needs to express the number of links in proximity to the measured point and the utilization of the theoretical network of all nodes connected to each other is not important.

For these reasons the used density description metric consists of dynamic measurement of the number of intersection nodes in the proximity of each measured position of the route. Also the average density is calculated for each simulation scenario. That can be described by two simple equations.

$$d_{ij} = \frac{\sum_1^k 1}{S} \quad (1)$$

$$D_j = \frac{\sum_{i=1}^m d_{ij}}{m} \quad (2)$$

Where:

d_{ij} is the dynamic density in the proximity of the i-th measured position in j-th scenario
k is the number of nodes in the proximity of the i-th measured position in j-th scenario
S is the area of proximity [m^2]
D_j is the average density in the j-th scenario
m is the number of measured positions in the j-th scenario

The density computation is described in chapter 3.2.5. Results of simulations and the dependence of map matching on the road network density are shown in chapter 4.

2.3. MAP MATCHING AND ROUTE TRACKING

2.3.1. OVERVIEW

The fundamental problem evaluated in this thesis is the problem of map matching. The map matching is (as the name suggests) a method for user's position assignment to map based on received location data. Map matching algorithms are supposed to be able to determine which link and which segment of the link is the user travelling on and on which position on the link it is. Clearly, the map matching serves as a basic form of error mitigation (especially within the road traffic where we can easily rule out the cases of vehicle driving outside of the road) and is widely used in the state-of-the-art navigation systems.

The fact that the map matching algorithms are used very frequently (and are almost a necessity in the road traffic navigation systems) does not make the problem any easier. In fact, there is no universally recognized "best map matching algorithm" that would be the best fit for all the possible inputs and applications and that is why the purpose and the objective of this work is to model several map matching algorithms of different approaches (ranging from simple point-to-point matching to more sophisticated selective look ahead methods [25] and to compare their performance based on different inputs, namely to see how they cope with different measurement errors, sampling rates and network densities. That is described in more detail in the chapter 3.3.

2.3.2. Classification of Algorithms

There is no normalized way of map matching algorithm classification, therefore it is necessary to explain the division used in this section of the thesis. In reality, different authors divide algorithms differently. Some authors use the general/special division that distinguishes between the algorithms that are applied for all users and all links (general) and algorithms used just for specific links and/or specific users (i.e. in [43]), another way of map matching algorithms division is the division to centralized and locally performed algorithms. It is also possible to divide the algorithms based on the time of their execution (real-time/post-processing), as can be seen for example in Goh's et al. work [18].

But this work does not use the upper mentioned divisions, since they are not focusing on the method of matching itself (which is the most important in this thesis) but instead mostly on the method of working with the results. Therefore this thesis works with following map matching division.

P2P

Point – to – point algorithm is the basic and purely geometrical map matching method. It is simply projecting the measurement to the closest point in the network. It is not considering not only the way or the direction of travel, but not even the links. In reality that means that often several measured points may be mapped to the same map point.

That algorithm is the simplest of all the map matching methods and (understandably) also the least precise of them. But in the same time, its simplicity can be viewed as an advantage, since it can work very fast and with very simple maps.

Computations used in the P2P algorithm are simple – the measured point is compared with a set of candidate map points by measuring the distance to them. The measurement point is then mapped to the map point to whom it has the shortest distance, as can be seen in the following equation.

$$M_i = \left\{ N_k | d_k = \min_{1 \le j < n} \left\{ \sqrt{\left(P_{ix} - N_{jx}\right)^2 + \left(P_{iy} - N_{jy}\right)^2} \right\} \right\} \tag{3}$$

Where:

M_i is the i-th mapped point
N_j is the j-th map point
P_{ix}, P_{iy} are the x and y coordinates of the measured point
n is the number of candidate nodes
d_k is the minimal distance to the measured point from any node

The important part of the algorithm is obviously the selection of the candidate map points. That may be achieved by multiple ways such as construction of increasingly large circles or by map segmentation, which is the method used in this work. Each map point is in the beginning of the process equipped with the number of rectangular map segment it fits into.

During the mapping part of the algorithm the measured point is compared only with map points in the same segment and in the adjacent ones.

P2C

Point – to – curve matching is another case of geometrical map matching method using the measurement of distance as the only metric for decision making. Measured point is mapped to the closest link segment in the closest point.

The distance between point and an infinite straight line is on a 2D plane measured by construction of perpendicular line that goes through the measured point. But in the case of a road network all lines (link segments) are finite and therefore there may not be the intersection between the perpendicular line going through the measured point and the link segment. In that case straight lines need to be constructed to the ending points of link segment and their lengths compared in order to find the shortest distance (as is shown in the equations below).

$$\overrightarrow{u_j} \perp \overrightarrow{v_j} \quad (4)$$

$$c_j\left(P_i, \overrightarrow{u_j}\right) \quad (5)$$

$$d_i = \begin{Bmatrix} (P_i, A_j) & if\ (A_j, B_j) \cap c_j = \emptyset\ and\ |a_j| \leq |b_j| \\ (P_i, B_j) & if\ (A_j, B_j) \cap c_j = \emptyset\ and\ |a_j| > |b_j| \\ (P_i, I_j) & else \end{Bmatrix} \quad (6)$$

Where:

A_j, B_j are starting and ending points of the examined link segment
$\overrightarrow{v_j}$is a directional vector of (A_j, B_j)
$P_i$is a measured point
d_i is the shortest straight line from measured point to any link
$a_j, b_j$are straight lines connecting $P_i$with $A_j$and $B_j$respectivelly

That is obviously applicable only for networks that consist of straight link segments. But these networks are used by the vast majority of algorithms (i.e. [24][18][23]), including all the algorithms used in the next chapter of this thesis. In case of networks using curve instead of straight lines essentially same methods are used for the map matching. The point is matched to the link segment that is closest to it and to the intersection of shortest possible straight line going from the point to the curve with the given curve. Only the methods of finding this shortest line are different.

In the similar way as in P2P matching, the set of candidate links needs to be chosen prior to the matching itself. Usually it is done by the simple construction of circle around the measurement point and all the links containing inside of that circle or crossing the circle are considered to be the candidate set. This method may be also called “error buffering”, such as in Jianjun’s and Xiaohong’s paper [23]. The static map segmentation, as explained in P2P matching may be used as well, but due to the higher complexity of computations within the algorithm (compared to “simple” distance measurement) it may not be the most effective way.

The P2C method is undoubtedly more precise than simple P2P matching, but the main concerns and sources of the error still remain. The reason for that is that simple P2C matching is still only geometrical matching that does not reflect the network configuration and possibilities or the route of vehicle (either passed or – in case of offline matching – route that is the vehicle going to take). The very common problems can be shown in following figures. Figure 4 and Figure 5 used by White et al. in [44] show quite nicely major difficulties. The Figure 4 shows how "un-smartness" of algorithm causes the inability to map points properly in areas close to intersection, since the point should clearly be mapped to the link A, but because it is equally close to the link B it may be mapped to the wrong location. Figure 5 is emphasising the problem of instability of the algorithm. In case of the measurements being in almost the same distance from the two links the mapped position will likely oscillate between them, which is obviously wrong.

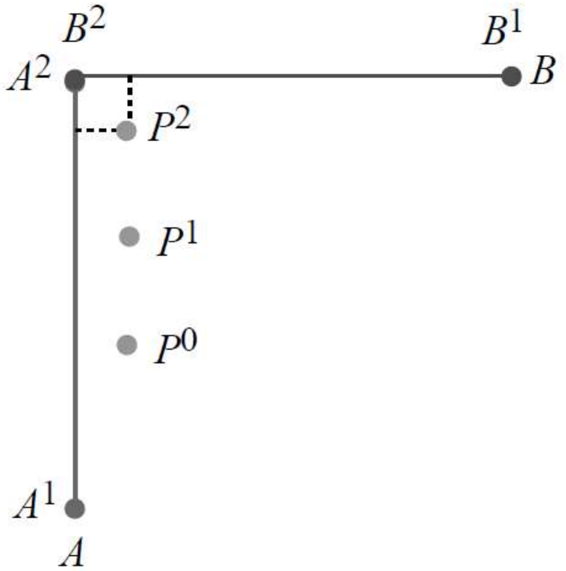


Figure 4: Example of incorrect mapping [44]

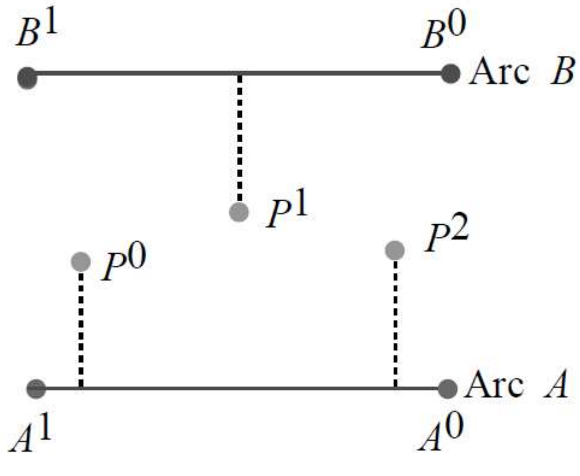


Figure 5: Example of mapping oscillation [44]

The described P2C matching is strictly geometrical form of matching, not using any information about the topology (except for the obvious information between which two points lies the link segment) or the routing. But since there is no normalized way of map matching division, it is possible to find authors who talk also about the enhanced P2C mappings using for example connectivity information to select candidate nodes [44] as a P2C matching. Also, it is worth noting that most of the matching algorithms use this P2C method for actual projection of the point to the link – to the closest point on that link.

C2C

Curve –to –curve matching is a bit more sophisticated map matching method compared to the previous two, since it adds one more feature to the decision making process – it measures the distance not only of the measured point from the link, but the distance of the line between the measured point and the point that has been measured before that and the link. The mapping of the point on the link itself is being made in the exactly same manner as in the P2C matching – to the intersection of the perpendicular line to the link segment going through the measured point or to one of the finishing points of the link segment (in case the intersection lies outside of the link) – the difference is solely in the process of the selection of the link.

As White et al. [44] state, there are two basic ways how to determine the distance of the links of different lengths. It is possible to either simply add up the distances of the finishing points or to add up the distances between the finishing points of the shorter link with the projected point at the longer link. The curve-to-curve distance can be computed using one of the following equations:

$$d_{ij} = \left|P_i - A'_j\right| + \left|P_{i-1} - B'_j\right| \tag{7}$$

Or:

$$d_{ij} = \left|P_i - A_j\right| + \left|P_{i-1} - B_j\right| \tag{8}$$

Where:

d_{ij} is the curve to curve distance
P_i, $P_{i-1}$are the two following measured points
A'_j, B'_j are two projected points of P_i and P_{i-1}
A_j, B_j are two finishing points of the link

Incremental

Incremental method described in this chapter is generally not (in contrast with P2P, P2C and C2C method) well defined by the other authors. Perhaps the most well known example of the term itself comes from Piotr Szwed and Kamil Pekala [19]who used it for their algorithm of the Hidden Markov model for map matching, which is not the meaning of the term that is used for the purposes of this work. This work uses the term incremental map matching for the algorithms heavily influenced by Quddus et al [43] comparing weighted scores for various route characteristics. Generally, the distance and angular score is being used in incremental algorithms (i.e. by Yin and Wolfson [45]), but it is possible to also implement the consideration of the other characteristics. Quddus et al. [43], for example, uses the score for the relative position to the link. The problem with adding more characteristics into the consideration is that it makes the weighting parameter estimation more difficult. Proper and precise parameter estimation is vitally important for the functionality of the algorithm and it is necessarily empirical.

The accepted definition for the purposes of this work is that the incremental method is the map matching solution using the map topology together with simple geometry and basically can be viewed as an enhanced version of curve-to-curve mapping. After the process of the candidate links selection, the links are evaluated based on their distance from the point and based on the angular difference between the route and the link. The distance is computed in the same manner as in the previously described algorithm. Candidate links selection process and partial score computation are, as specific parameters of individual algorithms, described in detail in chapter 3.3.5. What is common for all of the incremental method's algorithms modelled for this thesis is the computation of their overall score:

$$S(P_i, c_j) = \sum_{k=1}^{n} s_k(P_i, c_j) \tag{9}$$

Where:

$S(P_i, c_j)$ is the overall score for mapping the measured point P_i to the link c_j
$s_k(P_i, c_j)$ is the score of one eval. characteristics (i.e. distance, angle) for mapping P_i to c_j
n is the number of evaluated characteristics

Other methods

The P2P, P2C, C2C and various incremental algorithms are modelled in the following chapters. But there have been a number of algorithms using different methods and approaches described in literature. Perhaps the most widely used group of methods can be called **statistical approach** or **probabilistic algorithms** (which is the name used by Quddus et al. in map-matching algorithms summary created in 2006 [46]).

The general idea of probabilistic algorithms is that confidence area around a position measurement is constructed based on the error variances. The confidence area is then projected on the network and the link segment within the area is identified as the segment the user is travelling on. In case more segments are found within the confidence area, criteria similar to the ones used in topological incremental map matching, such as proximity and connectivity criteria are used to identify the correct segment.

The general overview of probabilistic method is described by Quddus et al. [46]. Probabilistic algorithms themselves were developed during the last three decades by multiple authors. One of the first to introduce probabilistic matching was Honey [47], who patented his algorithm for matching dead reckoning's sensor signals to a network. Ochieng et al. [48] later developed algorithm for GPS position matching that uses knowledge about network in order to increase computational effectiveness – the confidence area is constructed only near the intersections where is a possibility of link change.

Many other sophisticated and advanced map matching methods do not fit any of the upper mentioned categories. Algorithms that base their decision making on Hidden Markov model (HMM), such as the ones introduced by Szwed and Pekala [19], Torre et al. [49] or Goh et al. [18], differ from the previously categorized methods by the logic of the decision making. In general, the algorithm using HMM works with road segmentation – road segment and projection of positioning measurement on that segment represent HMM state. Algorithm starts with mapping the first measurement "conventionally" .Then the emission probability (likelihood of measurement point observation on the candidate segment being true) and transition probability to another is HMM states are computed. Transition probability computation can use different variants of weights and constraints. Szwed and Pekala use a simple maximal speed constraint[19], whereas Torre et al. and Goh et al. use more sophisticated topological constraints [49] [18].

Interesting concepts using fuzzy logic model is presented in Fu et al's paper [50], since it produces one simple output (probability of matching to the candidate link) from mere geometrical information about road network and the position measurements – difference between the user's route direction and link direction and proximity of the measurement from the link is analyzed.

Other authors use the combination of GNSS based knowledge with other source of information about the user's movement. Najjar and Bonnifait [51] created an algorithm combining differential GPS with ABS sensors in order to have continuous information about the location in the road network. Based on the combination of these information sources, two criterions are created – proximity criterion and heading criterion that serve as an input to the map matching system. Then Belief theory is applied for each positioning measurement – based on the upper mentioned criterion, each link is assigned particular degree of belief and the link with the largest degree is selected as the correct one.

Gustafsson et al. [52] presented a method with ambition to limit or to entirely eliminate GPS from the road network positioning. Their algorithm works with information from particle filters and with wheel speed and combines the information with the digital map. Authors state that even with very faulty initial position estimation, final achieved accuracy was one meter. The advantage of this approach compared to GPS based system is the performance in urban area where GPS's accuracy declines due to high buildings shading. The disadvantage is the necessity of at least somewhat correct initial position estimation.

Nevertheless, algorithms described in this subchapter ("Other methods") serve only an illustrative role with the purpose to show the complexity of the topic. Despite of the fact that all the upper mentioned methods are interesting and well thought it would have been difficult to model and evaluate the performance of all of them, all the more so when not all of them present their structure in detailed manner that could be used for simple enough transformation to MATLAB model. None of the algorithms described in this subchapter is modeled and evaluated in latter chapters.

3. SYSTEM DESCRIPTION

3.1. OVERVIEW

The system for map matching and route tracking algorithms evaluation is described in this section. The inputs and outputs of the system are described thoroughly in the chapters 3.2 and 3.4, whereas the algorithms are described in the chapter 3.3.

The figure below depicts the system with its inputs and outputs. It is important to realize that the figure (and the system described in this chapter) does not consider inputs into algorithms whose effects are not evaluated (such as map or speed of the user/vehicle) as inputs into the "evaluation system". Therefore only road network density, positioning error magnitude and distribution and sampling rate, as only evaluated parameters, are considered to be inputs into the system. Output consists of positioning error expressed generally in unit of distance of error. The other output, correctly identified links, describe results in route tracking and route optimizing category and its purpose is to show the suitability of examined map matching algorithms with given inputs for different route tracking and optimizing applications.

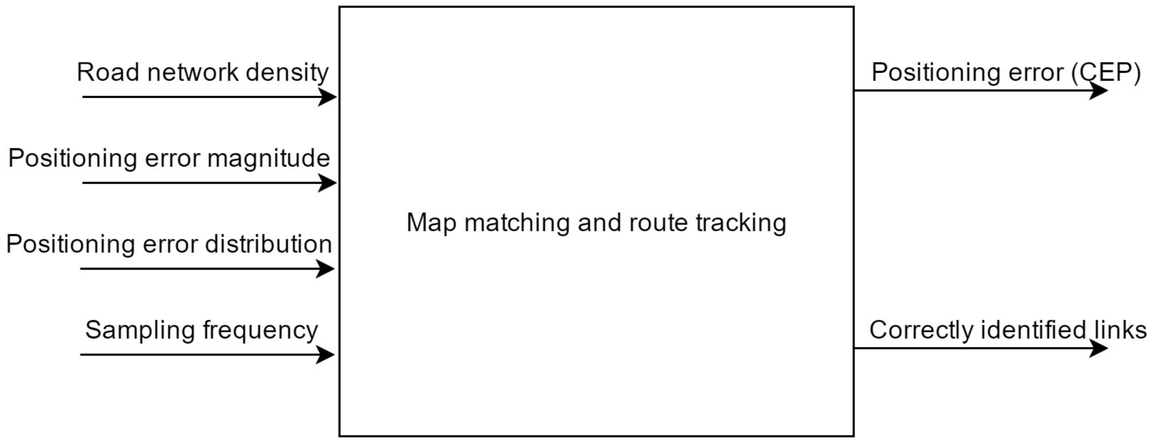


Figure 6: System description

3.2. DATA PROCESSING

3.2.1. DATA GENERATION

Necessary data set for simulations within this thesis consisted of the two basic parts – map and position measurements. The area of the city of Norrköping in the south-eastern Sweden was chosen as suitable location for the modelling, mostly due to the availability of the real positioning measurements. Map was extracted from OpenStreetMap [32] in a form of two points connected by straight lines, thus creating link segments. Points are specified in space by longitude and latitude. Each couple of points also has an ID of the link segment (all the links specified by couples of points with the same ID form the link segment) and topological information such as speed limit or link segment length. Extracted map used in all the simulation consists of 11601 links and covers the area between Linköping and Norrköping (including both cities).

Positioning measurements were provided from Linköping University's database and consisted of 45349 measurements collected in a time span of almost 5 and a half months between the August of 2011 and the April 2012 in 159 separate measurements mostly in the city of Norrköping, but also in the areas exceeding the mapped area. After the measurements taken outside of the upper mentioned mapped area were filtered out, only 39913 of them remained. The measurements consisted of the position in longitude/latitude and also the date and the time of the event, which is the information that proved to be very useful especially in the sampling rate simulation (see chapter 3.2.4).

3.2.2. Generation of route scenarios

This chapter explains the actions performed on the original data set in order to achieve normalized and comparable results and to allow the modelling of desired scenarios. Figure 7 below shows the chosen approach and the data processing leading from the original set of real world measurements to the prepared simulation scenario.

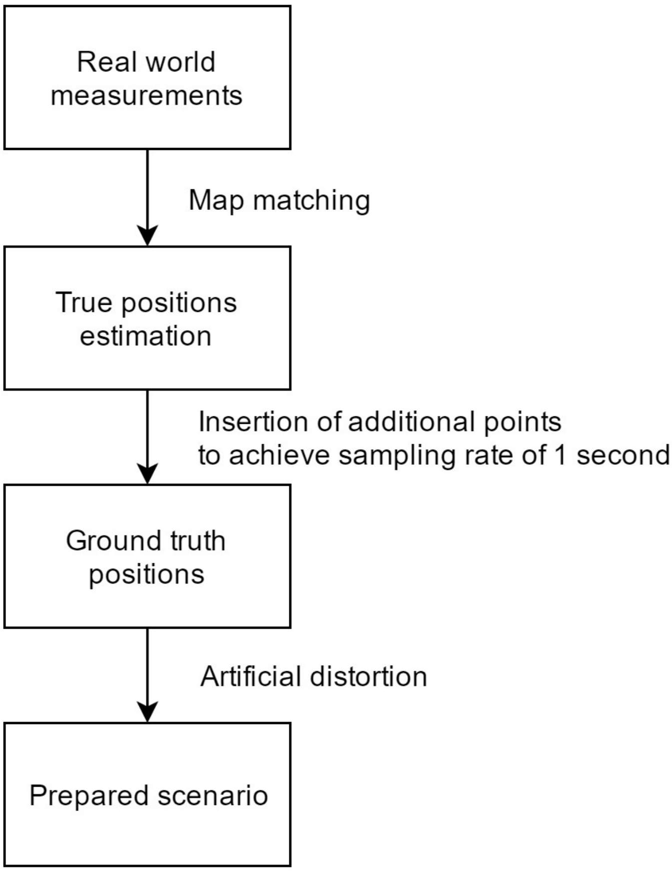


Figure 7: Data processing

The beginning of the data processing lays in having the real world measurements prepared according to the requirements described in the chapter 3.2.1. After that, the most sophisticated algorithm method – the offline incremental method – is applied to the data set. It matches the measurements to the links and these are then considered to be the true positions of the measured points. Since true positions are always just estimates, also some minor manual changes are possible during that step.

With true positions estimated, maximal examined sampling rate needs to be achieved by the addition of points between each two positions (in case they are not already the lowest examined time period apart). That process is explained in detail in the chapter 3.2.4. After the induction of points, ground truth data set – the set that is in the end compared with simulation results - is prepared.

But that is not the end of data processing, since it would not make sense to examine the results of map matching on perfectly precise data. That is the reason for an application of artificial distortion. The distortion consists of lowering the sampling rate (which is quite a simple task) which is described in detail in the chapter 3.2.4 and of error simulation. The error simulation is described in the chapter 3.2.3. After these two tasks are completed, the simulation scenario is prepared.

3.2.3. Error simulation

Independent error research and measurements

Due to the inconsistency of the optimal positioning error distribution description in the relevant literature (explained in the chapter 2.1.4), independent measurements were performed within the thesis to explore the topic and to clarify which distribution ought to be most precise in simulating the horizontal positioning error. The idea was that either one of the distributions proposed by literature (normal i.e. [17], Rayleigh [21] or Gamma [22]) would be clearly confirmed and used in the thesis, or none of the upper mentioned distributions would be confirmed and thus results of the measured error distribution would be used as well as theoretically correct distributions and results would be compared.

As a part of the research several measurements were done both in open space and in urban areas by two mobile GPS receivers (Sony Xperia M4 Aqua smart phones). Longest measurement (time wise) took place for 11 hours, whereas shortest measurements were approximately one hour long. Both the devices were recording their positions in the interval of 0.5 seconds. Ultimately, one (the longest) measurement was chosen for evaluating errors in the urban area and one measurement was chosen for evaluating errors in the open space.

The most delicate and problematic part of the measurement was establishment of the correct “ground truth” position. That was done by extracting the position from orthographic Google maps [53] based on known position in relation to other visible objects. Nevertheless, despite of the best efforts, it might have caused a bit of discrepancy, since the position was extracted only visually. After the reference “ground truth” was established, the mobile GPS receivers measured location in stationary positions for 11 hours. After that, measured data were extracted in CSV format and the analysis of the results was performed. The results were analysed in MATLAB software.

Since the measured data consisted of the longitude and latitude coordinates (and timestamp) and the “ground truth” was extracted also in the longitude/latitude form, it was necessary to transfer the information into the metric expression – to calculate the distance between the “ground truth” and each measured location, which serves as a location error. The haversine formula (thoroughly explained i.e. in Sinnott’s article [54] and on several websites using its application – i.e. [55]) calculating the distance over the Earth's surface was used to do that. The haversine formula applied to the examined case is shown below in equations 1 to 3.

$$a = \sin^2\left(\frac{\Delta LAT}{2}\right) + \cos(LAT_1) \cdot \cos(LAT_2) \cdot \sin^2\left(\frac{\Delta LON}{2}\right) \quad (10)$$

$$c = 2 \cdot \arctan\left(\frac{\sqrt{\mathrm{a}}}{\sqrt{(1-\mathrm{a})}}\right) \quad (11)$$

$$d = R \cdot c \quad (12)$$

Where:

ΔLAT is the difference between the measured point's and the ground truth's latitudes [deg]
LAT_1 is the latitude of the measured point [deg]
LAT_2 is the latitude of the ground truth [deg]
ΔLON is the difference between the measured point's and the ground truth's latitudes [deg]
R is the Earth's radius [meters]
d is the distance between the measured point and the ground truth [meters]

Following results showed immense discrepancy between the expected values of the error and the actual results. There are several possible explanations for that. The first might be biased ground truth extraction from Google maps (since it had to be done manually) and not sufficiently long measurement (even though the literature [22] suggested that the several hours long measurement is sufficient). Also, smart phones that were used as mobile receivers are not primarily positioning equipment and that may have caused the erroneously measured or erroneously logged positions.

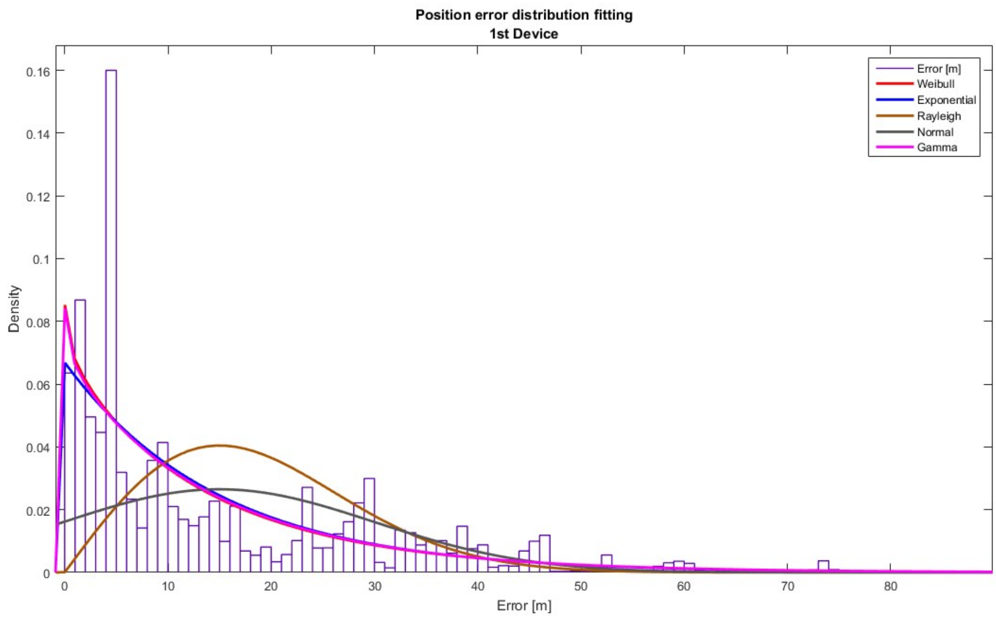


Figure 8: Position error distribution fitting, 1st Device. Created in MATLAB software.

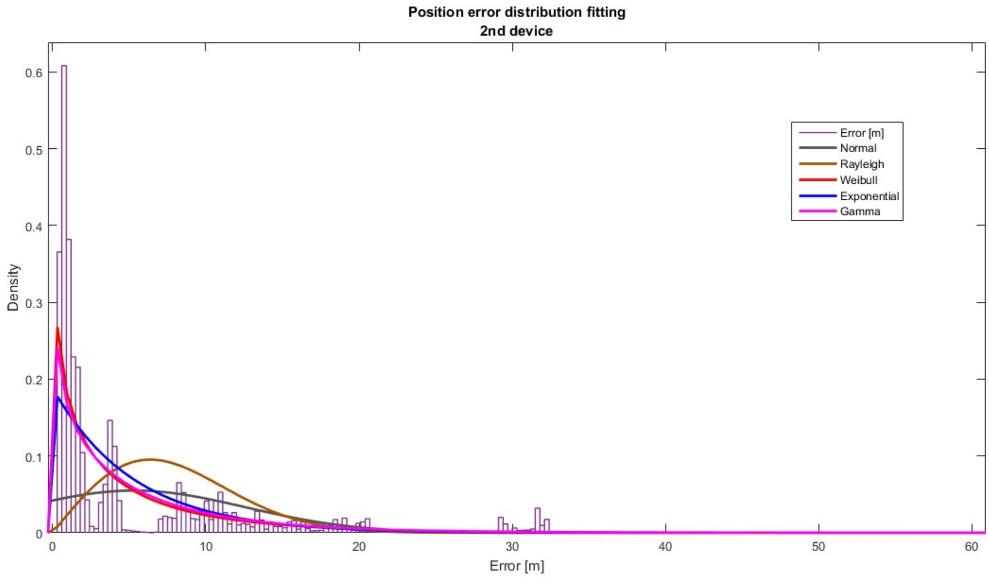


Figure 9: Position error distribution fitting, 2nd Device.

Figure 8 and Figure 9 show Rayleigh distribution, generally considered as the most suitable for GNSS error distribution [21], to be completely wrong (at least confronted with our measurements). The same can be said about normal distribution. Weibull distribution and Gamma distributions have the best results with very similar parameters. That is well documented in Table 1. The reason for that is the closeness of shape parameter (2^{nd} parameter) to 1, where both of these distributions take form of exponential distribution, as explained for example in Engineering and Statistical Handbook of U.S. National Institute of Standard and Technology [56].

Table 1: Error distribution parameters

Distribution	1^{st} Parameter		2^{nd} Parameter	
	1^{st} Device	2^{nd} Device	1^{st} Device	2^{nd} Device
Normal	14.944	5.3102	15.0457	7.2598
Rayleigh	14.9949	6.3601	N/A	N/A
Weibull	14.6497	4.5587	0.9575	0.7923
Exponential	14.944	5.3102	N/A	N/A
Gamma	15.7128	7.1318	0.9511	0.7446

Table 1 shows the parameters of examined distributions shown in the Figure 8 and Figure 9. Since Weibull and Gamma distributions had relatively better results compared to the other considered distributions and Gamma distribution was identified as the correct fit by Driver [22], it was chosen as the distribution for error simulations. 7 different combinations of parameters were used in order to simulate the different accuracy of the receiver. The combinations can be seen in Table 2

Error implementation

The error is added to ground truth positions by the easiest way possible. As was described in the chapter 2.1.4, the positioning error consists of a lot of parts and its correct description and modelling is a very complex process. That is why the thesis uses the simulation of the total error magnitude (which is the method described by Newson and Krum [20]). Even though that is the simpler way of error simulation (compared with modelling of parts of errors individually), it produces a few problems. The difficulties with error distribution choice and the reasons for Gamma distribution choice itself are described thoroughly in chapter 2.1.4

Gamma distribution is represented by two parameters describing their scale and shape. These parameters were in the beginning set in accordance with literature review and own measurements and then changed in order to represent the different error magnitudes. This was done by random numbers simulation in Matlab software. The error's mean averaged from one meter to approximately 30 meters in 10 different scenarios. The variants of used parameters are shown in the Table 2.

The big problem and one of the biggest sources of the result's discrepancies (as addressed in discussion in chapter 5) is that the obtained errors represent only the magnitude of the error – therefore they take the form of scalar. But that is not enough for the purposes of this work, since we need to express them in a 2D space and the vectorization, which is done using the uniform distribution random number generator and moving average filter takes place.

In the end, 10 different parameter combinations were used and are shown in the following table.

Table 2: Used parameters of Gamma distribution for error simulations

k **– shape parameter**	***Θ*** **- scale parameter**	**Theoretical average error (k*Θ) [m]**
3.147	0.462	1.454
4.725	0.462	2.183
4.725	0.924	4.366
9.450	0.924	8.732
9.450	1.848	17.463
5.235	0.475	2.487
10.470	0.475	4.973
10.470	0.951	9.957
15.710	0.951	14.940
15.710	1.902	29.880

3.2.4. Sampling rate implementation

Sampling rate induction is not a difficult task, once the ground truth data are created, since they are created with the highest examined rate (one measurement each second) and all it takes to create data set with different frequency is to selectively choose only certain points. The examined scenarios use frequency of 1, 2, 3,4,5,10,15,20,35,50 and some scenarios also 100 and 150 seconds.

In contrast with that, the induction of points to true points estimation data set is not trivial. Since, as stated above, the ground truth data set needs to have one point every second, it is clear that points need to be added based on the original measurements rate. In order to keep the points realistic, the speed in two bordering points is estimated, averaged and then points are added in between them as if they kept the speed. Since the real world measured data used in this work have sampling frequencies in low numbers of seconds (usually 3-4), this method is sufficient. If sampling rate in original data set was lower, the averaging of the speed would produce highly distorted results.

3.2.5. Network density computation

As described in the chapter 2.2.5, the road network density is for the purposes of this work defined as an average number of network nodes in the proximity of the route. The dynamic density is computed in the proximity of each measurement point and all the dynamic density values are then averaged in order to get the average density of the given scenario.

Proximity area around the measurement point takes form of a hexagon. The circle would provide more precise results, but hexagon was chosen due to computational reasons. All the points within the proximity hexagon are evaluated; nodes are identified and then added together. After that, the value of the dynamic density (number of nodes per area of proximity hexagon) is converted to normalized unit – number of nodes per square kilometre.

Scenarios are not classified according to the density value but according to the type of area that is represented by the given density. The network density scale is depicted in the Table 3

Table 3: Network density scale

Density [nodes/km^2]	Area
0-100	Rural
100-300	Suburban
300 and more	Urban

3.2.6. Summary

The results of the previously described process are simulated route scenarios that take place in various parts of the network and possess various error and sampling rate characteristics. They also differ from one another by the length of the route and by the travel time. Scenarios are summarized in the table below.

Table 4: Simulated scenarios summary

Route no.	Network	Density [nodes/km^2]	Travel time [s]	Route length [m]	No. of scenarios		
					Errors	Sampling	Total (Err*Sampl)
1	Urban	342	751	6044	10	12	120
2	Rural	32	233	7052	10	10	100
3	Suburban	228	165	2394	10	10	100
4	Suburban	238	212	2011	10	10	100
5	Urban	338	701	6721	10	12	120
6	Urban	444	405	4425	10	10	100
7	Urban	355	369	4724	10	10	100

Following figures show three of the used scenarios. Each of these scenarios belongs to different network group according to the scale specified in Table 3.

Figure 11: Route no.1, Urban network

Figure 10: Route no.4, Suburban network

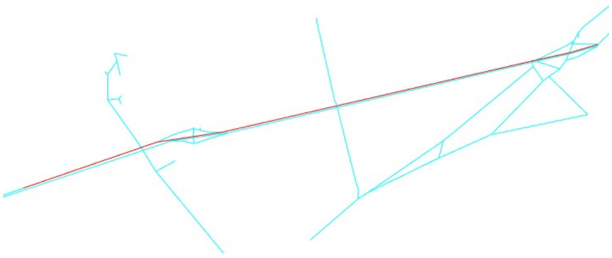


Figure 12: Route no.2, Rural network

3.3. DESCRIPTION OF USED MAP MATCHING ALGORITHMS

3.3.1. OVERVIEW

Theoretical background involving map matching regarding general attributes of several well known methods and their geometric expressions are thoroughly explained and described in the chapter 2.3. This part of the thesis is focusing on the description of the actual algorithms used for simulations in MATLAB software as well as on explaining certain attributes, strengths and limitations of these algorithms.

All of the map matching algorithms used in the simulation part of the thesis (except for one version of P2P matching, see chapter 3.3.2) follow one basic structure depicted in Figure 13, where only certain inputs and outputs differ as well as the map matching cycle itself. But following the same structure of the algorithms allowed an easy transition between the map matching methods as well as comparing the results (see chapter 4.2).

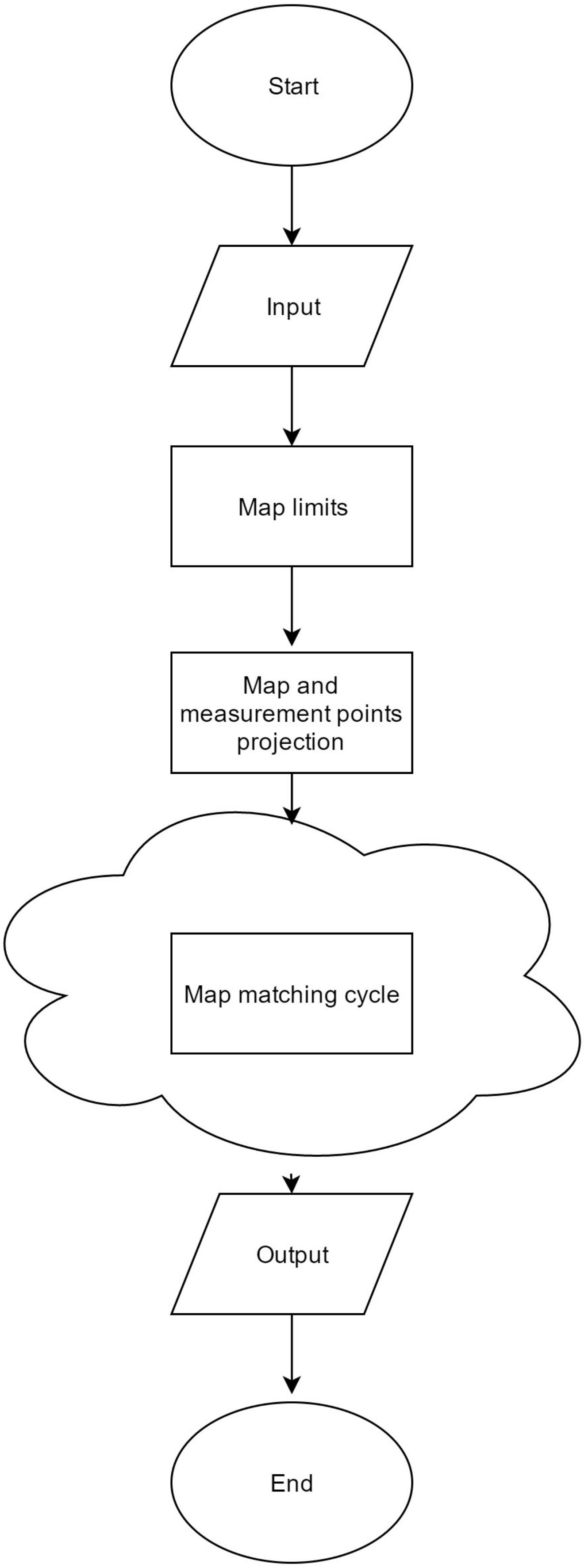


Figure 13: Overview of map matching algorithm.

3.3.2. P2P

System concept

The simplest of examined algorithms uses different structure of system concept compared with the other algorithms. While P2C, C2C and all types of incremental algorithms follow the structure depicted in Figure 13 with only map matching cycle being different, P2P uses entirely different structure, shown in Figure 14 below. There are two main differences between P2P and other types of algorithms. First one is that P2P does not use map projection, since it does not need to work with straight lines representing road network and so it keeps both measurement points and points of map in their longitude and latitude coordinates.

The second of the main differences is the absence of proximity hexagon that is used in all the other algorithms. In order to achieve maximal algorithmic simplicity, static map segmentation is used instead (as is explained in more detail in "Algorithm description" subchapter). The basic idea of P2P method is described in the chapter 2.3.2. The important factor for the matching performance is obviously also the density of the candidate points. The approach described in Quddus et al. [46] was chosen for the purposes of this work in that regard – all the map points are considered as candidate points (for the specification of map points, see chapter 2.2.5).

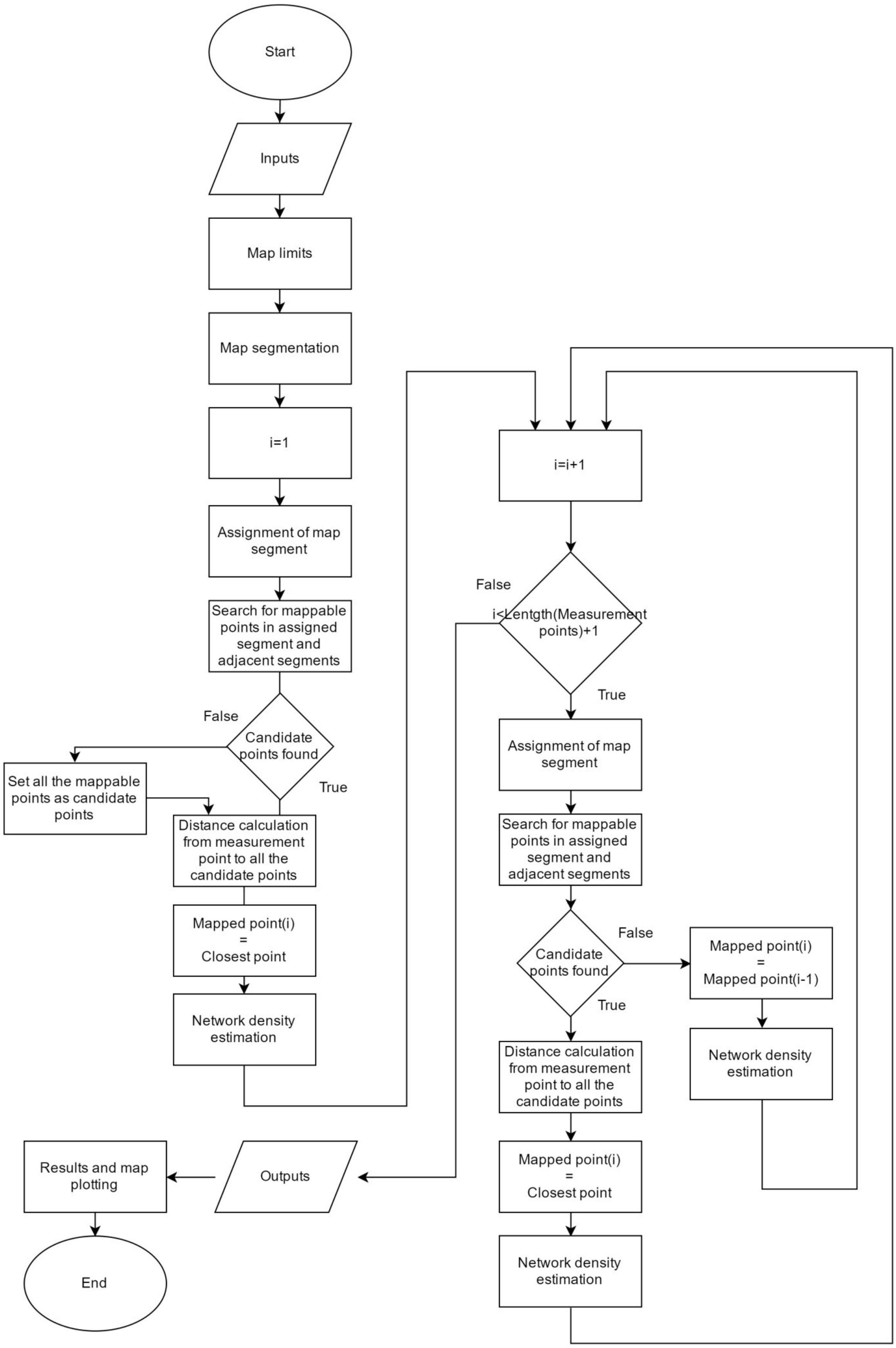


Figure 14: System elements of P2P algorithm

Inputs
Set of input variables and parameters reflects the simplicity of the algorithm, since it is smaller compared to the other geometric methods and (of course) to the more sophisticated topological incremental methods. Only variables necessary to perform the algorithmic computation are latitude and longitude of measurement points and coordinates of starting and ending points of link.

Algorithm description
In the beginning, based on the extreme values of longitude and latitude of the measurements, the initial boundaries of the area of interest are set. Then all the mappable points within this area are extracted and then divided into 64 similar rectangular segments (with respect to longitude and latitude – the straight line is connecting points with the same longitude or latitude).

After that the cycle runs through all the measurement points. First it assigns the ID number of the segment to each measurement point according to its position. Then it calculates the distance to all the mappable points in the same segment and in all the neighbouring segments. In case no mappable point is found in the assigned segment or in neighbouring segments, the measurement is mapped to the previously mapped position. In case there are no mappable points found for the first point of the route, all the points within the initial boundaries are considered to be candidates.

The distance is calculated from the longitude and latitude coordinates, so no projection is used (since it does not need to be due to considering only points and not lines.). Haversine formula, that is thoroughly explained in the chapter 2.1.4 and in equations 10-13 is used for that. Then all the distances are compared and the examined measurement point is mapped to the closest mappable candidate point.

Computation of the network density is slightly different compared to the other algorithms since it does not use hexagon around each measurement point, but it counts the number of nodes in each segment and then averages the number based on the rate of use of given segment. Later the average number of nodes is, of course, transferred to the normalized unit (number of nodes per square kilometre). The detailed explanation of the network density computations are to be found in chapter 2.2.5.

3.3.3. P2C

System concept
P2C algorithm differs from P2P by not considering the map as a set of points, but also taking connectivity of these points into consideration. That means it is not possible (or at least not precise) to work with points in longitude/latitude coordinates, but projection needs to be used in order to produce correct 2D map. As mentioned in chapter 2.2.4, Mercator projection was used for P2C and also for all the other algorithms with an exception of P2P.

The general idea of the method (described thoroughly in the chapter 2.3.2) is that the measurement point is mapped to the closest point of the closest link. The algorithm searches in proximity of the given measurement point for the set of candidate links and then chooses the closest one for the actual mapping. The P2C method for mapping the point to a link is used in all the other algorithms (except for P2P) as the others differ only with a way of choosing the proper link. System concept of the mapping cycle is shown in Figure 15 below, the P2C algorithm respects the generally used algorithmic scheme depicted in Figure 13.

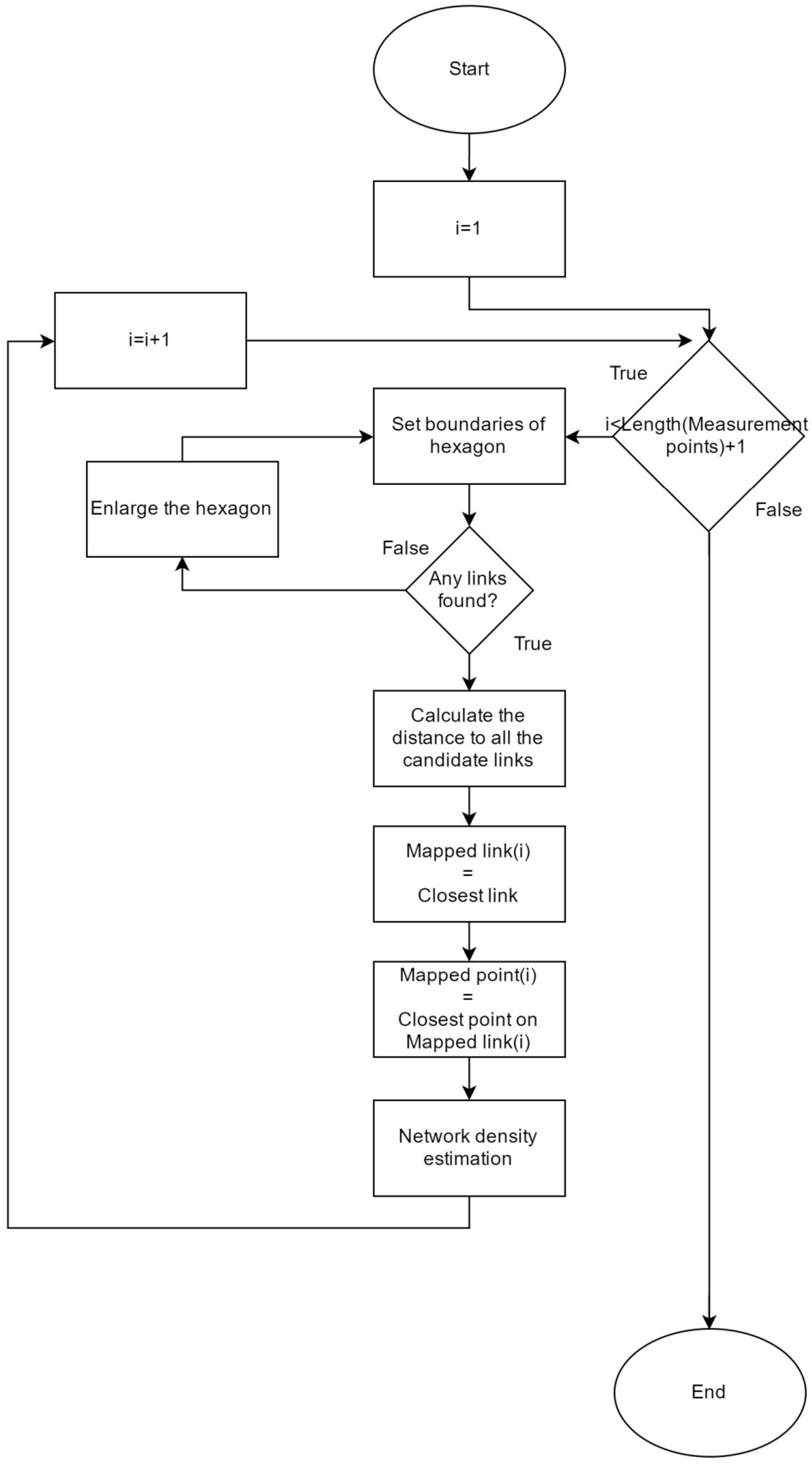


Figure 15: System elements of the P2C algorithm's mapping cycle

Inputs
Inputs of the geometric map matching method are reflecting (compared to P2P matching) the need to involve the information about connectivity of the link nodes, since P2C does not map measurements solely to nodes or map points, but also to links. Also proximity hexagon is used; therefore it is necessary to specify the radius of circumscribed circle of that hexagon.

Altogether, inputs are longitude and latitude coordinates of measurement points, longitude and latitude coordinates of map points and the IDs of the link segment that connects the given map point with the other map point. Another input is the radius of circumscribed circle of the proximity hexagon.

Algorithm description
The algorithm starts with transferring all the measurement points and map points to metric 2D system using Mercator projection. Then the initial boundaries of the map are set based on the extreme values of the measurement points (route) and map points within these boundaries are extracted.

After that, the matching cycle that goes through all the measurement points of the route begins. It starts with the creation of the proximity hexagon around the measurement point and the extraction of all the links that are either within the hexagon or that cross it. In case none is found, the radius of the hexagon is enlarged twice and the same is done once again if still no link is found. If there is no link within the proximity hexagon even after it has been enlarged two times, all the links within the map boundaries are considered to be the candidate links.

Then the distance to all the candidate links is calculated using the simple method described thoroughly in subchapter "P2C" in chapter 2.3.2. After that, the distances are compared and the link closest to the given measurement point is selected for the mapping and the measurement point is mapped to the closest position on that link. Also the dynamic network density is measured for the given measurement point (the method is thoroughly described in chapter 2.2.5.

3.3.4. C2C

System concept
In general, the system concept of C2C algorithm differs from the P2C algorithm only in one key aspect – instead of calculating the distance of the measurement point from the map links, it calculates the distance of the route (in a form of the straight line) between the two last measurements from the map links. Other than that, the mapping of the points is similar to the P2C algorithm. In the Figure 16 below, the system elements of the mapping cycle are depicted.

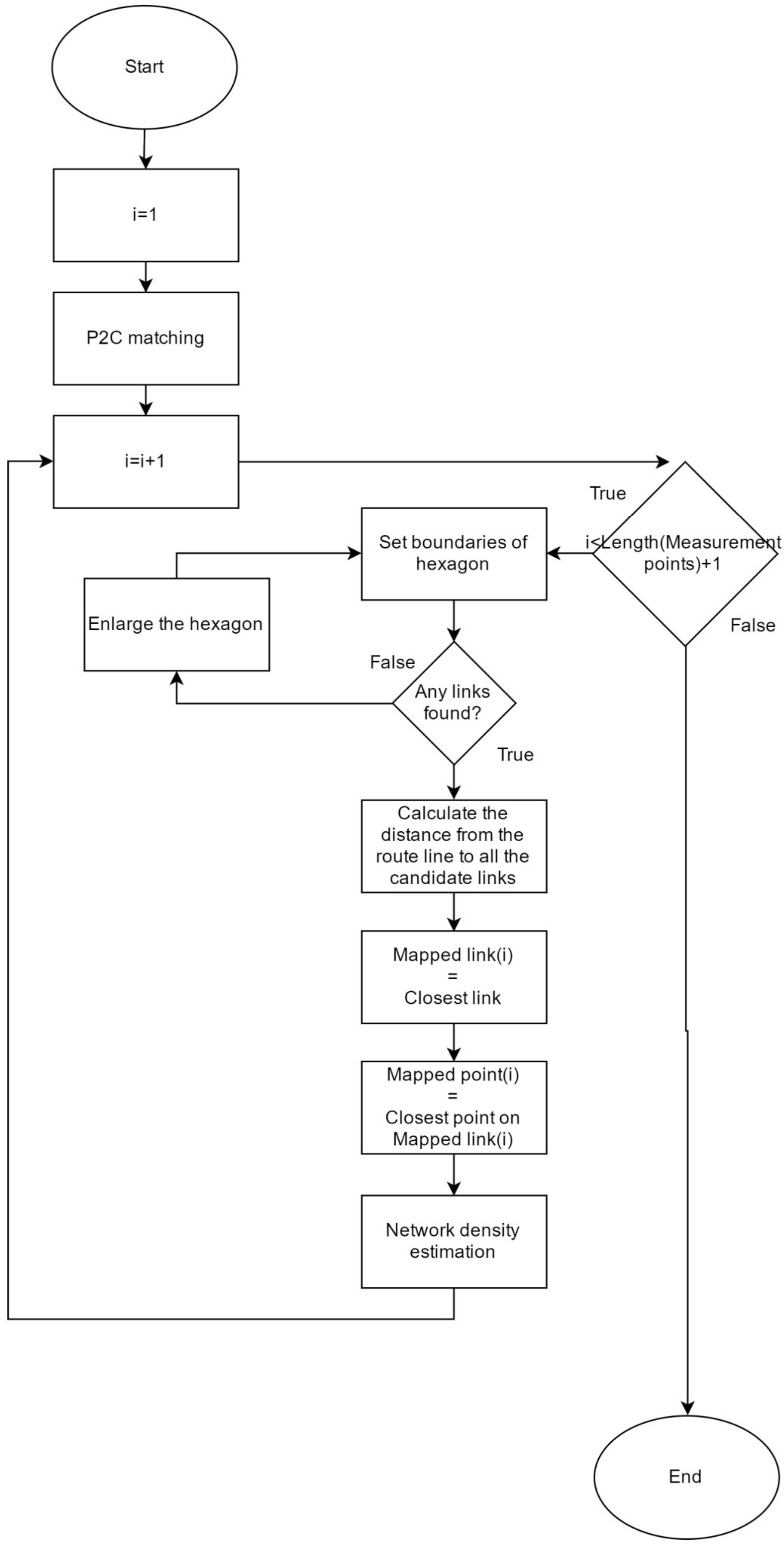


Figure 16: System elements of C2C algorithm

Inputs

Inputs of the C2C algorithm are exactly the same as the inputs of P2C algorithm described above in chapter 3.3.3. These are the coordinates in the form of longitude and latitude of both measurement points and map points together with the ID of the link for describing connectivity of points and radius of circumference circle of proximity hexagon are necessary.

Algorithm description

As the C2C algorithm respects the general map matching algorithm structure depicted in Figure 13 and all the system elements except for map matching cycle are similar to the ones already described in chapter 3.3.3, only the map matching cycle is the focus of this subchapter.

The map matching cycle starts with mapping the first point with P2C method, since there is no route yet that can be used for C2C matching. After that it moves to a C2C regime that is working for the rest of the route. It is generally not vastly different from the P2C matching cycle, since it draws the proximity hexagon around the measurement point in a same manner, looking for links within or crossing this hexagon. In case no links are found, hexagon is twice enlarged. In case still no links are found, all the map points within the initial map boundaries are considered to be candidate links.

Then the distance between each candidate link and the straight line connecting last two measurement points is measured, using metric described in chapter 2.3.2, subchapter "C2C". After that, these distances are compared and the link that is closest is chosen to be the mapped. When link is chosen, the measurement point is mapped to this link using the simple geometric technique described in 2.3.2, subchapter "P2C". Also the dynamic network density is saved, based on the number of the number of the nodes within the proximity hexagon.

C2C holds advantages over P2C in a sense that it respects the direction of passed travel. On the other hand that also produces problems, since, especially with greater differences between the link segment length and passed route length, the computation of distance between lines become quite distorted..

3.3.5. INCREMENTAL

System concept

This chapter does not explain the meaning and general understanding of the term "incremental map matching" since it has already been explained in chapter 2.3.2. The goal of this chapter is to thoroughly describe used variants of the incremental map matching algorithms, their strengths and limitations. Incremental method is very broad term and 4 different types of incremental map matching algorithms were evaluated. The basic division is to online (without look-ahead) and offline (with look-ahead) algorithms, since looking ahead hugely affects the functionality. Nevertheless, both types of incremental algorithms (offline and online) share plenty of the system elements. Three versions of online and one version of offline algorithms were created and evaluated within the scope of this work. The flowchart in the Figure 17 presents the explanation of the algorithm's system elements.

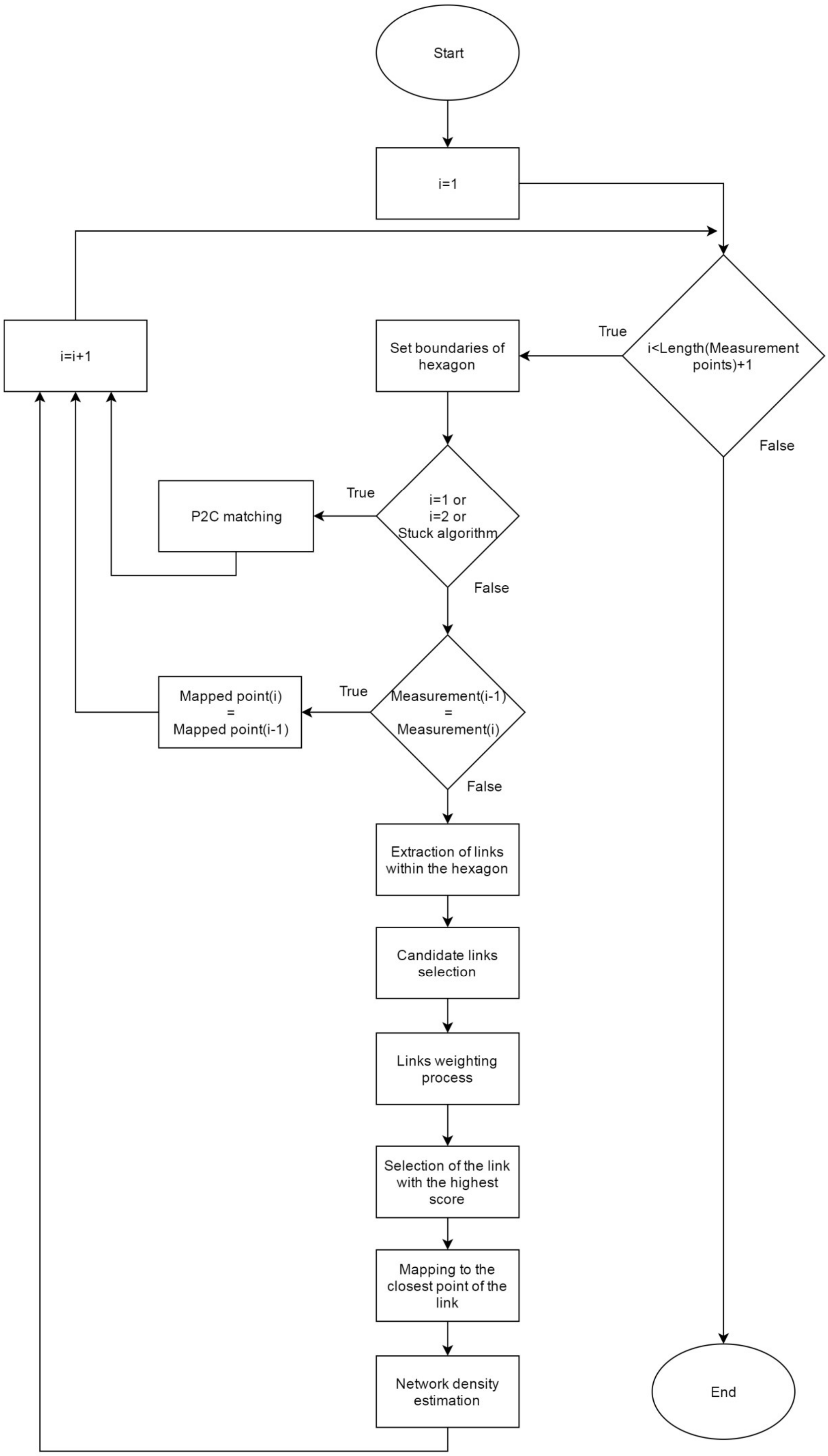


Figure 17: System elements of the incremental algorithm.

Inputs

All of the incremental algorithms have specific inputs compared to simple geometric map matching methods, because, as described in chapter 2.3.2, possible links are evaluated based on their score. The score of the link segment is always computed through the addition of the angular score and the distance score and both of these scores need to be weighed together through weighting parameters. These parameters (explained as a part of the score equation in chapter 2.3.2) are the specific incremental inputs. These inputs are used in all the incremental algorithms together with inputs used in P2C and C2C algorithms – measurement and map points coordinates, ID determining the link segment connectivity and the radius of the proximity hexagon's circumference circle.

Other important inputs for the advanced versions of the incremental method (both online and offline versions) are an information about passed time between the measurements in the form of timestamp, speed limits on each link and the distance buffer parameter, specifying how far can the user reach for the time between the last two measurements. The most advanced online and offline methods also use parameters called "stopped car rule parameter" and "intersection rule parameter" that are both explained in latter subchapters.

Algorithm description

As mentioned above, four versions of the incremental algorithm are used in this work. They are called **Incremental1 (I1), Incremental2 (I2), Incremental3 (I3)** and **IncrementalLA1 (LA1)**. These versions follow the same structure (depicted in Figure 13), but differ in few key aspects. LA1 is using look-ahead, whereas I1, I2 and I3 are not. Offline algorithms differ from one another by the candidate link segment selection process (described in latter subchapters).

All of the map matching cycles of examined incremental algorithms start with mapping the first two measurement points with P2C method. The same simple geometric method is used also whenever the algorithm is "stuck" and keeps mapping the points to the same point. In other cases the incremental method regime begins. It starts with construction of proximity hexagon and extraction of all the links within it. Then the candidate link segments selection process specific for every version of the incremental algorithm is applied.

After the candidate links are selected, all of them are assigned the score that reflects their potential for precise matching. The algorithms used within the scope of the work were evaluated in two categories – distance (proximity) and angle. That is similar to the work of Yin and Wolfson [45]. For example Quddus et al. [43] use also other means to evaluate the score, but since all the partial scores need to be weighed together and weight is added empirically, more scores equal more uncertainty.

Distance score applied in all used incremental algorithms is calculated by the equation 13 below.

$$DSC_{ij} = \frac{p_1}{(p_2)^{d_{ij}}} \tag{13}$$

Where:

DSC_{ij} is distance score of the i-th measurement point and the j-th link segment
p_1 is the first parameter of distance score
p_2 is the second parameter of distance score
d_{ij} is the shortest distance from the i-th measurement point and the j-th link in meters

Angular score applied in all used incremental algorithms is calculated by the equation 14 below.

$$ASC_{ij} = p_3 \cdot \cos\theta_{ij} \tag{14}$$

Where:

ASC_{ij} is angular score of the i-th measurement point and the j-th link segment
p_3 is the angular score parameter
θ_{ij} is the angle between the j-th link segment and the straight line connecting i-th and (i-1)-th measurement point

The general approach of the offline version is not very different from the online version, since the used offline algorithm uses the same logic of score computation and candidate links choice methods as the 3rd version of online incremental algorithm. The only difference is that the offline version adds the maximal score of the next measurement point (or measurement points – based on the look-ahead "depth") if mapped to the evaluated link or to any link reachable from it. The following equation explains it.

$$S(P_i, c_j) = s(P_i, c_j) + \sum_{k=1}^{n} s(P_{i+k}, c_m) \tag{15}$$

Where:

$S(P_i, c_j)$ is the overall score for mapping the measured point P_i to the link c_j
$s(P_i, c_j)$ is the score for mapping P_i to c_j
n is the "depth" of the look-ahead that sets how far (time wise) algorithm looks
$s(P_{i+k}, c_m)$ is the maximal score for mapping following point to the link c_j or to link that is reachable from link c_j

After that the link segment with the highest score is selected as the correct one and the measurement point is mapped to it by the P2C logic.

Candidate link segments selection

Candidate link segments selection is a crucial process of all the examined and simulated incremental algorithms. It is also the selection of the candidate links that distinguishes four versions of the algorithm from one another.

Incremental 1

The first variant, called for the purposes of the work “the discrete method”, uses purely topological method of considering the discrete number of link segments adjacent to the lastly mapped link in a direction of the travel. Illustration in Figure 18 shows how the candidate links selection works in case two adjacent link segments are considered as candidate links. Also the previously mapped link is considered to be among the set of candidate links.

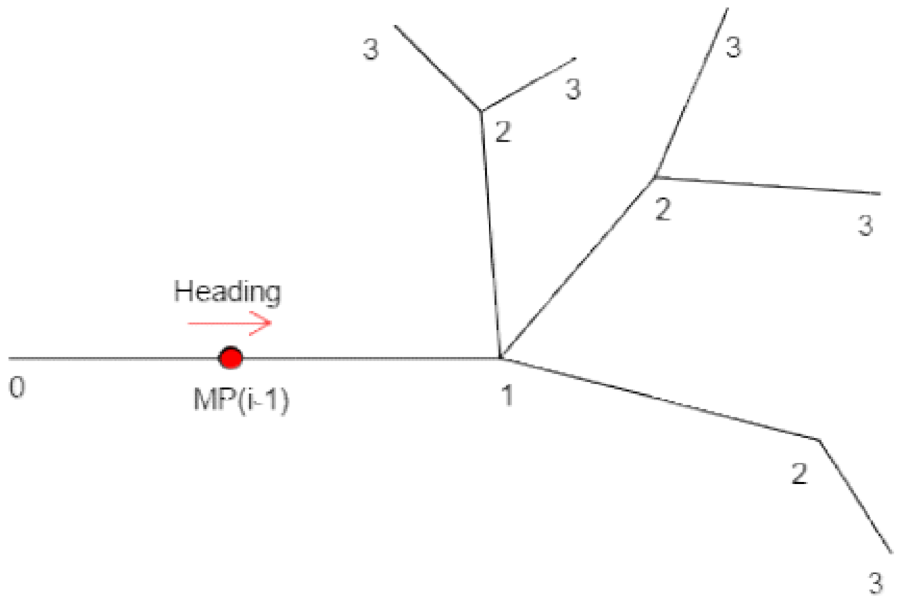


Figure 18: Topographical method of candidate links choosing (used in variant Incremental 1). Figure depicts considering 2 adjacent links to previously mapped link segment and previously mapped link segment itself as candidate links. MP (i-1) is previously mapped point and after its heading is determined, two more ending points of adjacent links are added to the ending point of mapped link.

Incremental 2

Other variants use more advanced method that considers adjacent links to be candidate links only in case it is possible to reach them given the time difference between the previously mapped point and examined measurement point and speed limits of links plus certain buffer (in case the speed limit was exceeded or point was incorrectly mapped). The purpose of these activities is not to evaluate which link segment is the most probably the correct link segment, but only to mitigate mapping to clearly impossible links. Possible distance from previously mapped point is calculated by the equation 16 below.

$$l_{lim} = (v_{lim} + v_{buff}) \cdot \Delta t \qquad (16)$$

Where:

l_{lim} is the possible distance from previously mapped point (distance limit) in meters
v_{lim} is the speed limit of the examined link in meters per second
v_{buff} is empirically determined speed buffer in meters per second
Δt is the time difference between the previously mapped point and examined measurement point

The distance limit calculation is also important in terms of setting the limit on the actually considered link segment, not only for the link segment choice. The candidate link segments choice using speed-distance method is shown in the Figure 19 below.

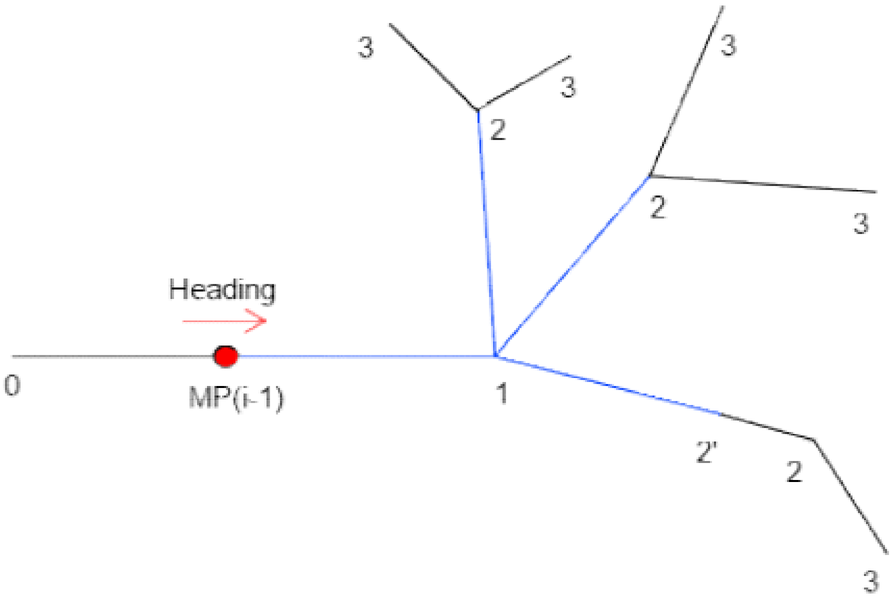


Figure 19: Speed-distance limit method of candidate links choice (used in variants Incremental 2 and 3 and Incremental LA1).
Blue lines depict candidate links. Notice that the lowest link was reduced (since it would be impossible to reach the original ending point) and the new ending point 2' was constructed.

Incremental 3 and Incremental LA1
Even those enhancements do not necessary improve the mapping in all cases. Map matching is in general most vulnerable in the close proximity to an intersection (since the score differences between the link segments closely behind intersections tend to be small, but mapping to the correct one is crucial) and in cases the vehicle is moving slowly or not at all (since the short route heavily influenced by even small error drastically alters the angle between the measurements and therefore angular score is not correct).

That is why the most advanced methods called I3 and LA1 use not only the speed-distance limit method described in the subchapter above, but also two measures to mitigate these faulty mappings. Two measures were for the purposes of this work called the "Intersection rule" and the "Stopped car rule".

Intersection rule is applied in case the previously mapped position is within the small distance from the intersection upstream. In that case the possibility that the wrong mapping occurred is considered and all the link segments starting in the point of intersection are also considered to be candidate links. For that reason the IRP (intersection rule parameter) is stating how far behind the last mapped point the algorithm looks for possible intersection.

$$Dec_{IR} = \begin{cases} 1 & if \sqrt{\left(x_{MP(i-1)} - x_I\right)^2 + \left(y_{MP(i-1)} - y_I\right)^2} < IRP \\ 0 & else \end{cases} \qquad (17)$$

Where:

Dec_{IR} is the decision made by the algorithm regarding the consideration of links from previous intersection. 1 means that links are considered, 0 that they are not
$x_{MP(i-1)}, y_{MP(i-1)}$ are the coordinates of previously mapped point
x_I, y_I are the coordinates of the first intersection upstream from the previously mapped point
IRP is the intersection rule parameter, expressed in meters

Stopped car rule prevents the car to change the link if only small difference in position is detected (car is expected to be still or to move very slowly). For that reason the SCRP stopped car rule parameter) is introduced and it decides what is considered as stopped car/very slowly moving car. The algorithm measures the time distance between the last two measurements and multiplies it with the SCRP. Resulting number is the limiting distance considered as "no movement". It is clear that the parameter can be also viewed as a limit speed, considered as "zero speed". The decision-making equation 15 is expressed below.

$$Dec_{SCR} = \begin{cases} 1 & if \sqrt{\left(x_{MSP(i)} - x_{MSP(i-1)}\right)^2 + \left(y_{MSP(i)} - y_{MSP(i-1)}\right)^2} < SCRP \\ 0 & else \end{cases} \qquad (18)$$

Where:

Dec_{SCR} is the decision made by the algorithm regarding the constraint of measurement point to the same link as previously mapped point. 1 means the point is forced to stay on the previously mapped link segment, 0 that it is not
$x_{MSP(i)}, y_{MSP(i)}$ are the coordinates of currently examined measurement point
$x_{MSP(i-1)}, y_{MSP(i-1)}$ are the coordinates of previous measurement point
$SCRP$ is the stopped car rule parameter in meters per second

Specifics of the method
Online version of the algorithm works with the device that does not know the future positions. That makes the algorithm generally vulnerable to positioning errors in specific cases, where it's usually useful "gravitation" preventing changing links in improbable moments causes the error. That case can be called "inability to react to turning on intersection" and it is shown in the Figure 20 below.

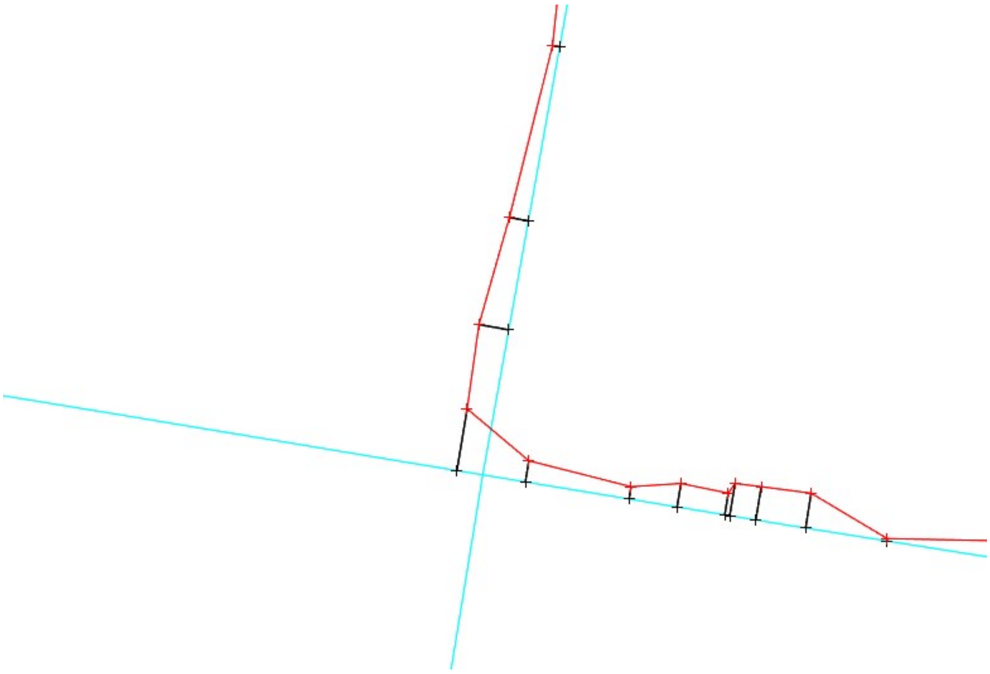


Figure 20: The Intersection with measurements mapped to links by incremental map matching algorithm. Links are depicted by blue color, red crosses are measurements and they are mapped by black line.

Figure 20 shows very accurately the upper mentioned problem, since the “intelligence” of the algorithm did not serve the purpose there and the first measurement after the turn on the intersection didn’t react to the slight change of angle and mapped the measurement in the straight direction. In this case, simple geometric P2C and C2C algorithms map the measurement correctly, since they map it to the closest (correct) link. That can be seen in Figure 21 below. Also, these problems are easily solved in an offline version of incremental matching, which is described in the following subchapter.

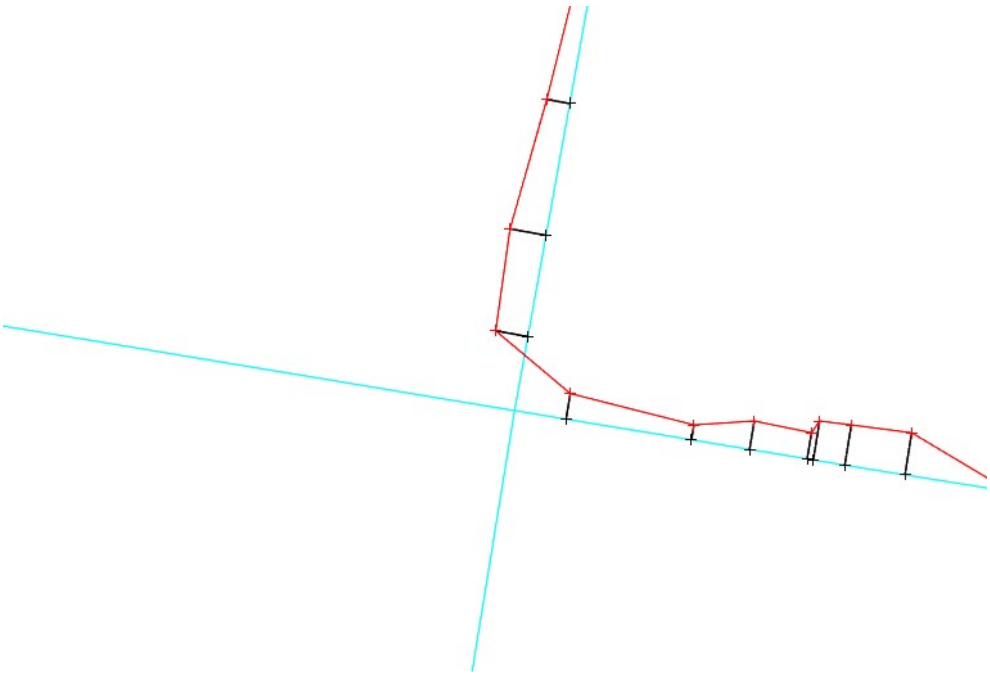


Figure 21: The Intersection with measurements mapped to links by geometric P2C map matching algorithm. Links are depicted by blue color, red crosses are measurements and they are mapped by black line.

Specific characteristics of the incremental map matching are also the outcome of backwards route tracking that needs to be implemented for the proper functionality of the algorithm. The knowledge on which links and link segments the vehicle has travelled on the way from the last mapped point to the evaluated measurement point is crucial for the candidate links selection process. But in the same time the algorithm's main function is the map matching, therefore the route tracking is simplified compared to the route tracking applications and algorithms such as Kim's and Moon's tracking of moving magnetic sensor objects[57] or Chen's and Li's tracking based on fuzzy control[58].

In case both of the evaluated points – previously mapped location and currently evaluated measurement point are within the proximity hexagon, standard brute force shortest path algorithm is used for calculating the possible routes between the mapped point and potentially mapped positions of the currently evaluated measurement point.

But in case the previously mapped point is not inside or on the border of the proximity hexagon of the evaluated measurement point, simplification actions need to be performed, since the previously mapped link may not be within the set of extracted links. In that case previously mapped position is projected to the border of the proximity hexagon to an intersection between the hexagon's boundaries and the straight line between the previously mapped position and currently evaluated measurement. The newly projected point is called the substitute point. Then the substitute point is mapped to the one of the links within the hexagon using the same weighting process as for the standard mapping. The timestamp of the point changes proportionally to the distance between the previously mapped position and substitute point. The process is shown in the Figure 22.

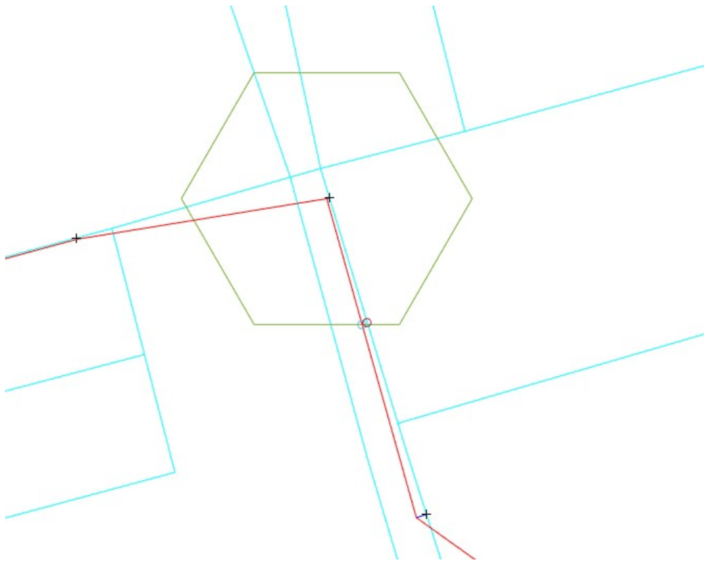


Figure 22: Proximity hexagon is constructed around the measurement point (black cross in the centre of the hexagon). The previously mapped point is not the part of the hexagon and therefore the substitute point is constructed (small blue circle) and then mapped to the position inside or on the border of hexagon (small red circle).

3.3.6. SUMMARY

The used methods and their key characteristics are summarized and compared in the Table 5

Table 5: Summary of the used algorithms

Algorithm	Abbreviation	Type	Candidate link choice
Point-to-point	P2P	Geometric	-
Point-to-curve	P2C	Geometric	PH
Curve-to-curve	C2C	Geometric	PH
Incremental1	I1	Topological	PH+topological method
Incremental2	I2	Topological	PH+SD
Incremental3	I3	Topological	PH+SD+IR+SCR
Incremental look-ahead 1	LA1	Topological	PH+SD+IR+SCR

- PH = proximity hexagon, SD = speed-distance method, IR – intersection rule, SCR – stopped care rule

3.4. EVALUATED CHARACTERISTICS

3.4.1. POSITIONING ERROR

Position error of mapping describes the behaviour of algorithms in the simplest and the most illustrative manner since it measures the error of the mapped position in distance unit. That is done for each of the positions and then aggregated by CEP (circular error probability).

CEP is defined as “the radius of circle centred at the true position, containing the position estimate with probability of 50%” [59]. That explains the position error more complexly compared to other methods such as DRMS [59], since one very faulty measurement does not drastically influence the overall result.

Nevertheless, the used CEP method is CEP 67, which is essentially the same as the upper described CEP; the only difference is the probability of containing the position estimates. CEP 67 works with 67%. The CEP 67 was not calculated, but estimated from the empirical cumulative density functions of positioning errors at p (67).

3.4.2. CORRECTLY IDENTIFIED LINKS

Correctly identified links serve as the indication of route tracking. The output is simply the percentage of correctly mapped link. That is computed by very simple equation.

$$CIL_{ij} = \frac{CL_{ij}}{n_{ij}} \quad (19)$$

Where:

CIL_{ij} is the percentage of correctly identified links in the i-th algorithm and j-th scenario
CL_{ij} is the number of correctly identified links in the i-th algorithm and j-th scenario
n_{ij} is the number of measurements in the i-th algorithm and j-th scenario

Compared with the correctly identified distance and time correctly identified links do not consider the network and its connectivity at all for the computations. The simple 1 (is on the link) or 0 (is not on the link) are calculated for each measurement point and after the simulation the value of CIL_{ij} is divided by a number of measurement points, the average of correctly mapped links is set.

4. RESULTS

4.1. OVERVIEW

Overall, 7 routes in urban, suburban and rural environment are examined within this work. These routes are transformed into the 740 simulated scenarios with combinations of varying input positioning errors and sampling rates. The variety of the network density in these scenarios is extremely important for gaining meaningful data sets, since the description of the road network density's effect on the map matching performance is one of the goals of the work and it was in the same time the only effect that was not artificially simulated – therefore the road environment needed to be chosen carefully.

It is necessary to group the resulting data in order to work with them properly. They are grouped according to their inputs – that means the results of every simulation scenario with given environment, sampling rate and error, are compiled and grouped. Every scenario is then represented by the set of values, describing their positioning error (represented by CEP 67), wrongly and correctly mapped link, distance and time (represented by value in meters/seconds and by percentage).

All the simulations are performed in Matlab software.

4.2. EVALUATION OF THE SIMULATION RESULTS

4.2.1. EFFECTS OF THE ORIGINAL POSITIONING ERROR

When original error and final map matching error were compared, they generally showed quite expected results, as it turned out very reliable linear dependency was formed for the geometric algorithms. The linear dependency is nicely visible in Figure 24. Whereas topological incremental algorithms produced error in a more unpredictable manner, due to the fact that one wrong mapping can influence long part of the route and produce multiple faulty mappings. That is shown in the Figure 24. Note that mapping errors larger than 60m are not shown.

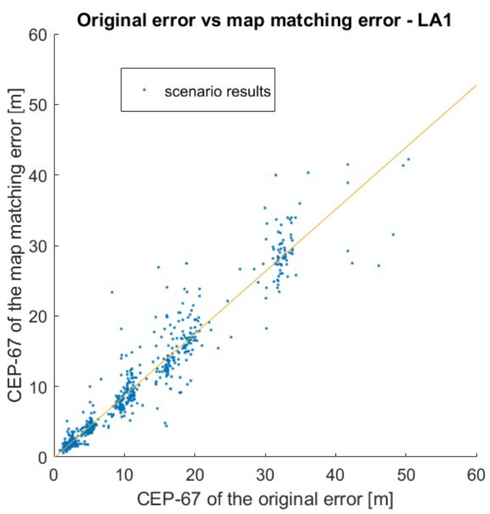


Figure 23: Original error vs. Map matching error - LA1

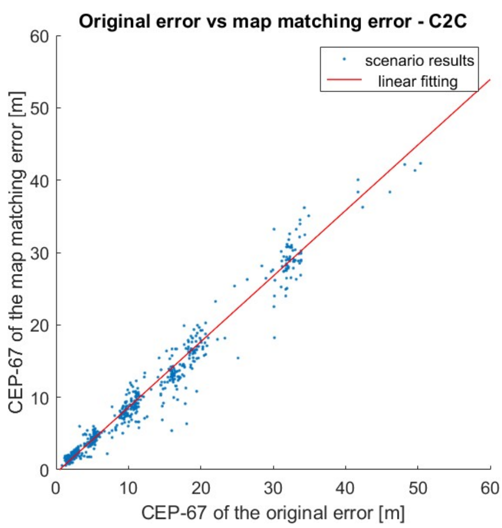


Figure 24: Original error vs. Map matching error - C2C

Table 6: The average improvement of map matching error compared with the original error

Average improvement of MM error against original error	All scenarios	Orig Error CEP-67 < 10m	Sampling rate < 16s	Orig Error CEP-67 < 10m and Sampling rate < 16s
P2P	-893%	-2100%	-910%	-2099%
P2C	14%	26%	15%	23%
C2C	14%	19%	15%	18%
I1	-25%	-42%	22%	31%
I2	-139%	-399%	18%	16%
I3	24%	31%	24%	30%
LA1	32%	36%	34%	40%

Table 6 confirms the results shown in the Figure 24 and the Figure 24. Due to the simplicity of geometric algorithms, the mapped location is never very far from the correct location. That is the reason P2C and C2C algorithms perform much better if the average over all scenarios is considered compared with basic topological algorithms (I1 and I2). The basic topological incremental algorithms meanwhile perform much worse and worsen the original error by magnitude of hundreds of percent in average over all scenarios. The P2P method worsens the localization quite significantly.

If only smaller original errors are considered (CEP-67 smaller than 10 meters), the average improvement of P2C, C2C, I3 and LA1 increases, while all the other algorithms perform worse. On the other hand, consideration of only lower sampling rate increases the performance of topological algorithms very significantly, whereas geometric algorithms improve only slightly.

4.2.2. Effects of the sampling rate

The problem of sampling rate's effect evaluation is less complex than positioning error's, since sampling is artificially set (the same as positioning error) as discrete values (which is different from positioning error). Samplings of 1,2,3,4,5,10,15,20,35, 50 and for two scenarios also 100 and 150 seconds are simulated, and when aggregated and grouped, effects on individual algorithms are evaluated.

One of the basic metrics of map matching performance is the percentage of correctly chosen links. In the Figure 25, the sampling effect is visible, as the average percentage of correctly mapped links steadily declines between 5 second to 35 second rate for all the algorithms except for P2P matching (topological algorithms decline more steeply than P2C and C2C algorithms). Surprisingly enough, incremental algorithms improve their performance between 35 and 50 second rate, but that may be caused by insufficient amount of data (shorter scenarios included only few measurement points with rate of 50 seconds). Incremental algorithms map links more correctly than P2C and C2C in an interval between 1 and 10 second rate, at rates higher than 15 seconds the opposite is true.

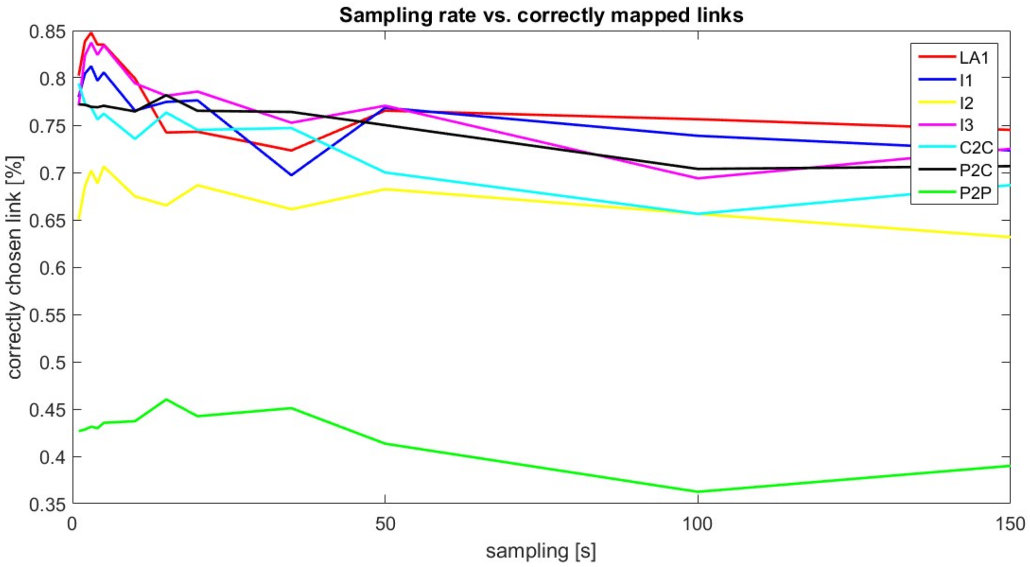


Figure 25: Sampling rate vs. correctly mapped links

Figure 26 below shows in more detail the results of algorithm in terms of correctly mapped links for sampling rates between one and five seconds. The advantage of smart topological algorithms is clearly visible, with LA1 and I3 being the most precise. Interesting fact is that whereas P2C and C2C algorithms performance is highest at the lowest rate and then slowly declines, the topological methods improve the percentage of correctly chosen links up to sampling rate of 3 seconds where they all culminate. That is caused by the combination very small sampling rate – very large original error. The large error often causes the measurement points to not follow the direction of travel, which is difficult to deal with for the topological algorithms, but not a problem at all for the geometric algorithms.

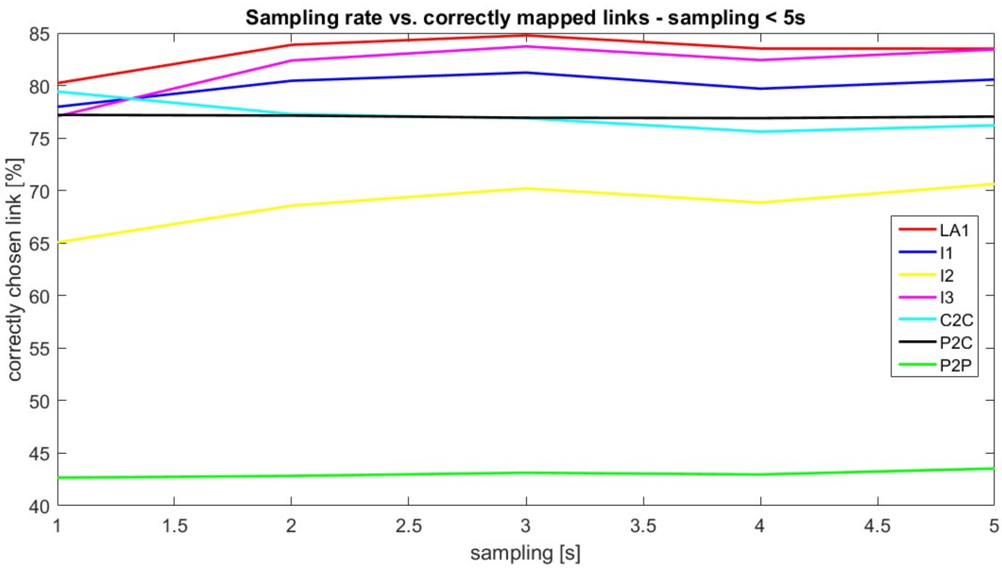


Figure 26: Sampling rate vs. correctly mapped links for sampling rate < 5s

Even though the difference is still not big between the simple geometric and more complex topological methods and the percentage of correctly chosen links may still seem quite low for highly sophisticated algorithms the thing to realize is that the average of correctly chosen links is done over all the original positioning errors which (due to the reasons described in the upper paragraph) favors geometric methods.

4.2.3. Effect of network density

The network density is evaluated by the initial classification of the simulated routes into one of the three network categories based on their density: urban, suburban and rural. The details of the density computation are described in the chapter 3.2.5 and the classification in the chapter 3.2.6. The simulation results are then grouped according to these classifications and according to the type of the used algorithm. The initial expectation of the better performance in sparser networks has been confirmed (as can be seen in the Figure 27 and Figure 28), but certain somewhat surprising results are registered nevertheless.

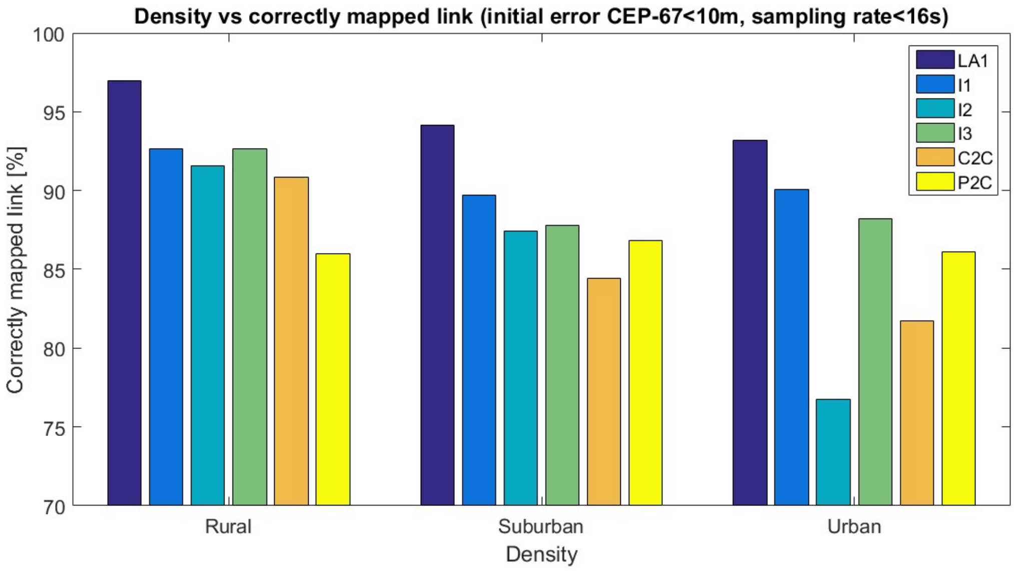


Figure 27: Density vs. correctly mapped link (initial error CEP-67 < 10m, sampling rate < 16s

The Figure 27 depicts the percentage of correctly mapped link of all the algorithms except for P2P for which the route tracking performance metric do not make sense. This figure contains only scenarios where the initial error CEP-67 was smaller than 10 meters and the sampling rate was less than 16 seconds. The effect of the network density is clearly visible, as the algorithms decrease their respective correctly mapped distance percentages with an increasing density, even though there are two exceptions. P2C algorithm performs really badly in rural area and I1 performs slightly better in suburban environment than in urban environment. The LA1 algorithm proves to be the most effective in terms of correctly mapped link in all the network categories, followed by the I3. Geometric algorithms perform worse.

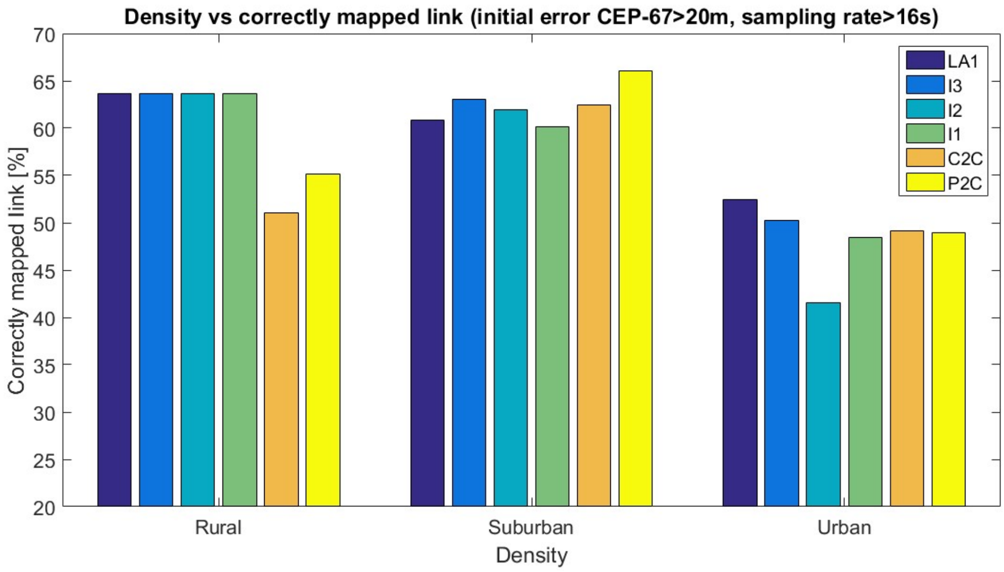


Figure 28: Density vs. correctly mapped link (initial error CEP-67 > 10m, sampling rate > 16s

The Figure 28 depicts the similar variables as the Figure 27; the difference is in the constraints applied on the dataset. Whereas the previous figure shows results of scenarios with small errors and low sampling rates, the Figure 28 depicts the results of scenarios with large errors (CEP-67>20m) and high sampling rates (>16s). Interesting result is the severe drop of the performance of the I2 algorithm in the urban network. Generally, the algorithms once again decrease their correctly mapped link percentage as the density increases, apart from the P2C and C2C algorithms, which perform better in suburban than in rural areas. Apart from the rural network, the topological algorithms do not perform significantly better compared to the geometric ones.

4.2.4. Route tracking performance, usage in ITS applications

The route tracking in terms of backwards tracking of the movement of the vehicle between the two mapped positions is not only necessary, at least to some extent, in order for topological algorithms to be functional, but also provide the interesting results that may be used in ITS applications for their requirements specification.

The metrics for measurement of correctly mapped link is thoroughly explained in chapter **Chyba! Nenalezen zdroj odkazů.**. The important thing to realize is that the evaluated algorithms are map matching algorithms and the route tracking is not their primary goal. Nevertheless, the results provide valuable insight into the performance of the evaluated algorithms in a bigger complexity than the mere positioning error measurements. Figure 27 and Figure 28 show the correctly mapped link percentage in dependency on the network density. Following tables, Table 7 and Table 8 provide the information relatable to the application requirements since they show the percentage of simulated scenarios achieving certain degree of the evaluated characteristics. This was done for all the scenarios, but only the important ones are shown in this chapter, the rest of them may be found in the Appendix of the thesis.

Table 7 shows the number of achievements of the LA1 given the constraint of 95% of correctly mapped link. It is clearly visible that the constraint is very harsh, as only scenarios with lowest frequencies (between 1 and 3) and the lowest original positioning errors (CEP-67< 7m) fulfill it in more than 50% of cases. That indicates the fact that the applications requiring this type of performance would have to either use very precise positioning receiver and to send positioning measurements very often, or the different map matching algorithm, not used in this thesis, would have to be used.

The possible ITS application with these requirements could be the electronic fee collection Whereas most of the European countries use the DSRC based system [60], some of them use satellite based systems (such as Germany). This system needs to work precisely and Quddus et al. [46] specifies the desired precision at 95% correctly chosen link. Since it is not necessary to perform mapping in real time, the offline look-ahead methods can to be chosen.

Table 7: Percentage of LA1 scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	89%	39%	0%	26%
(3,7>	72%	21%	0%	16%
(7,12>	8%	0%	0%	0%
(12,20>	15%	0%	0%	3%
>20	11%	0%	0%	0%

The Table 8 shows the same LA1 algorithm's performance metric as the Table 7, but with different constraint, as only 80% of correctly mapped link is required for scenario to be considered successful. That changes the results quite considerably and allows the possible application to choose reasonable positioning device and to use the sampling rate up to 20s. One of the possible applications is the offline fleet management tracking controlling the proper route choice.

Table 8: Percentage of LA1 scenarios achieving 80% of correctly mapped link

Original error CEP-67 [m]	Sampling period [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	100%	100%	60%	50%
(3,7>	100%	100%	60%	38%
(7,12>	100%	96%	33%	32%
(12,20>	81%	52%	13%	15%
>20	47%	47%	4%	4%

The applications using the route optimization, such as smart winter road maintenance program ASSIST [61] need the maintenance vehicle to identify the link it is travelling on in order to know what part of the road network has been ploughed and salted and when. That means the correctly mapped link is the best metric to decide what the application's device requirements should be. If the 95% precision of correctly mapped link is desired, the Table 9 shows the results – 4 scenarios achieved at least 50% of reliability.

Table 9: Percentage of I3 scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	82%	74%	12%	21%
(3,7>	61%	50%	9%	14%
(7,12>	10%	0%	0%	0%
(12,20>	16%	0%	0%	0%
>20	10%	0%	0%	0%

The applications that need the highest positioning accuracy (the lowest error) need to use the different metric. The example of such application is the public transport priority at intersections, where it is critical to correctly identify the distance of the vehicle from the intersection. According to Quddus et al. the requirements for priority at intersections are the accuracy of 5m (CEP-67) with integrity [46]. Considering the intersection priority happens in urban areas with high network density, the demanded results were achieved for higher original errors than specified 5 meters only by the third incremental algorithm with SCR and IR and only in scenarios where the initial location error sampling rate was less than 16s. The demanded accuracy was also achieved by P2C algorithms, but with less correct link identification precision, which is not desirable. The offline version of the algorithm could not be considered, since the priority at intersections needs to happen in real-time.

5. DISCUSSION

The results of the work showed not only how well the geometric and topological map matching algorithms respond to different quality of the input data, but also how complex topic map matching is and how difficult it is to try to quantify and simulate different real-world positioning scenarios. Nevertheless, the platform for simulation and evaluation of the map matching algorithms has been developed and tested, which may be viewed as the greatest accomplishment of this work.

The comparison of examined algorithms revealed the strengths and limitations of the approaches and that created the knowledge base for stating the recommendations for various ITS applications. Although it is necessary to realize that only the geometric and topological algorithms were evaluated and that the other methods such as various probabilistic methods or Hidden Markov model methods might prove to be more effective than the examined algorithms in some specific situations or for specific applications.

That shows what the main focus of the future work in this field should be. The thesis created the platform for map matching algorithms comparison and for quantification of the results, but compared only fraction of the existing algorithms. Addition of more algorithms into the consideration could provide more valuable results and more valuable information for potential ITS applications.

The artificial error simulation can be viewed as a source of the greatest discrepancy towards the results, since the used method of error magnitude simulation followed by the random error direction simulation with moving average smoothing does not reflect some of the issues the user faces in reality. The change of the error is in reality not as chaotic as in the simulations done in this work, even though when added together, the error magnitudes follow the same distribution. The reason for that are the similar environmental conditions in certain parts of the route. In order to respect that, more complex error simulation would have to take place and also 3D modelling of the road surroundings would have to be used.

6. CONCLUSIONS

The simulations results and their analysis showed in great detail strengths and limitations of examined geometric and topological map matching algorithms, which was one of the main objectives of the thesis. This chapter summarizes the results and reasons behind those results.

The fact that not all the examined map matching algorithms proved to lower the positioning error in general do not confirm the statement that map matching can be viewed as a basic form of error mitigation. The original statement seems logical, since all the simulated users of the road network were at all times either on the link or in the node, whereas distorted positions were simulated throughout the 2D plane (which is mostly outside of the link/node). All the map matching algorithms mapped the position to the link or node and therefore increased the likelihood of correct matching. Nevertheless, the high error distortions combined with either very low or very high sampling rate proved to have such a negative effect on topological algorithms that it changed the overall average improvement of the positioning error into the deterioration.

The comparison of the algorithms performance brought rather expected results. The topological incremental matching performed better than geometrical algorithms at scenarios with smaller original positioning error where their more sophisticated method of finding the correct link proved to be effective. Negative effects of both the location error and sampling rate are smoothed by the network density, which showed to affect geometric and topological algorithms similarly with only few exceptions. In scenarios with larger location error, the lower sampling rate or both the topological incremental algorithms performed worse compared to geometric algorithms due to the inability to react on fast changing situations. Chaotic (high error) or sparse (low sampling rate) data increase the likelihood of wrong matching that persists for a number of measurements, whereas geometric algorithms work with one measurement at a time and therefore generally recovered from the wrong matching faster.

Analysis of correctly mapped link showed how well the algorithms performed in terms of route tracking and therefore allowed the recommendations to be made regarding the various types of ITS applications. The applications with high requirements for correct link choice should use the ultimately most precise algorithm. That proved to be third version of incremental algorithm with SCR and IR. To achieve the desired level of 95% percent of correctly chosen links, the location error should not exceed 7 m and the sampling rate should be between 1 and 5s.

REFERENCES

[1] P. Defraigne and G. P. Harmegnies, “Time and frequency transfer combining GLONASS and GPS data,” in *Frequency Control and the European Frequency and Time Forum (FCS)*, San Francisco, 2011.

[2] P. Kovár, P. Kacmarik and F. Vejrazka, “Interoperable GPS, GLONASS and Galileo Software Receiver,” *IEEE Aerospace and Electronic Systems Magazine,* vol. 26, no. 4, pp. 24-30, 2011.

[3] United States Coast Guard Navigation Center, “CNAV Performance,” 19 March 2015. [Online]. Available: http://www.navcen.uscg.gov/?pageName=CNAV_Performance. [Accessed 27 March 2016].

[4] G. W. Hein, J. Godet, J.-L. Issler, J.-C. Martin, P. Erhard, R. Lucas-Rodriguez and T. Pratt, “Status of Galileo Frequency and Signal Design,” in *Proceedings of the 15th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2002)*, Portland, 2002.

[5] ESA, “What is Galileo?,” 18 December 2015. [Online]. Available: http://www.esa.int/Our_Activities/Navigation/The_future_-_Galileo/What_is_Galileo. [Accessed 27 March 2016].

[6] National Coordination Office for Space-Based Positioning, Navigation and Timing, “GPS Overview,” National Coordination Office for Space-Based Positioning, Navigation and Timing, 11 February 2014. [Online]. Available: http://www.gps.gov/systems/gps/. [Accessed 27 March 2016].

[7] Information and analysis center for positioning, navigation and timing, “GLONASS history,” Information-analytical centre, 2016. [Online]. Available: https://www.glonass-iac.ru/en/guide/. [Accessed 27 March 2016].

[8] ESA, “The story so far,” 18 December 2015. [Online]. Available: http://www.esa.int/Our_Activities/Navigation/The_story_so_far. [Accessed 27 March 2016].

[9] BeiDou Navigation Satellite System, “System Introduction - Development Process,” 2012. [Online]. Available: http://www.beidou.gov.cn/2012/12/14/20121214e14a946961514cce98ab04d876258343.html. [Accessed 27 March 2016].

[10] BeiDou Navigation Satellite System, “System Introduction - Development Plan,” 2012. [Online]. Available: http://www.beidou.gov.cn/2012/12/14/2012121481ba700d7ca84dfc9ab2ab9ff33d2772.

html. [Accessed 27 March 2016].

[11] P. Majithiya, K. Khatri and J. Hota, "Indian Regional Navigation Satellite System - Correction Parameters for Timing Group Delays," *Inside GNSS Magazine,* pp. 40-46, January/February 2011.

[12] Russian institute of space device engineering, "Precision of GLONASS/GPS navigational definitions," Russian institute of space device engineering, 25 March 2016. [Online]. Available: http://www.sdcm.ru/smglo/stparam?version=eng&repdate&site=extern. [Accessed 27 March 2016].

[13] Department of Defense, United States of America, Global Positioning System - Precise Positioning Service Performance Standard, Department of Defense, United States of America, 2007.

[14] BeiDou Navigation Satellite System, "System Introduction - Service Target," BeiDou Navigation Satellite System, 2012. [Online]. Available: http://www.beidou.gov.cn/2012/12/14/20121214617d2d7b70334862a052d58bbf38beea.html. [Accessed 27 March 2016].

[15] International Civil Aviation Organization, Global Navigation Satellite System (GNSS) Manual, 2005.

[16] Z. He, Y. Hu, J. Wu, J. Wang, J. Hou and K. Wang, "A comprehensive method for multipath performance analysis of GNSS navigation signals," in *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, Xi'an, 2011.

[17] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang and Y. Huang, "Map-matching for low-sampling-rate GPS trajectories," in *GIS '09, Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, 2009.

[18] C. Goh, J. Dauwels, N. Mitrovic, M. Asif, A. Oran and P. Jaillet, "Online map-matching based on Hidden Markov model for real-time traffic sensing applications," in *15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Anchorage, 2012.

[19] P. Szwed and K. Pekala, "An incremental map-matching algorithm based on Hidden Markov Model," in *Artificial Intelligence and Soft Computing*, Zakopane, 2014.

[20] P. Newson and J. Krumm, "Hidden Markov map matching through noise and sparseness," in *GIS '09 Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, 2009.

[21] University of Sao Paulo, “GPS Horizontal position accuracy,” 14 February 2007. [Online]. Available: http://www.leb.esalq.usp.br/disciplinas/Molin/leb447/Arquivos/GNSS/ArtigoAcuraciaGPSsemAutor.pdf. [Accessed 17 April 2016].

[22] T. Driver, “Long-Term Prediction of GPS Accuracy: Understanding the Fundamentals,” in *Proceedings of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007)*, Fort Worth, 2007.

[23] D. Jianjun and C. Xiaohong, “An Improved Map-Matching Model for GPS Data with Low Polling Rates to Track Vehicles,” in *International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*, Changchun, 2011.

[24] K. Ghys, B. Kuijpers, B. Moelans, W. Othman, D. Vangoidsenhoven and A. Vaisman, “Map matching and uncertainty: an algorithm and real-world experiments,” in *GIS '09 Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, 2009.

[25] M. Weber, L. Liu, K. Jones, M. J. Covington, L. Nachman and P. Pesti, “On map matching of wireless positioning data: a selective look-ahead approach,” in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, San Jose, 2010.

[26] The Defense Mapping Agency, Datums, Ellipsoids, Grids and Grid Reference Systems, Fairfax: Defense Mapping Agency, 1990.

[27] NOAA, “What is the geoid?,” National Ocean Service, [Online]. Available: http://oceanservice.noaa.gov/facts/geoid.html. [Accessed 27 March 2016].

[28] G. Dolan, “The Greenwich Meridian,” Graham Dolan, 2016. [Online]. Available: http://www.thegreenwichmeridian.org/tgm/articles.php?article=0. [Accessed 27 March 2016].

[29] Federal Agency for Cartography and Geodesy, “International Earth Rotation and Reference Systems Service,” Federal Agency for Cartography and Geodesy, 2013. [Online]. Available: http://www.iers.org/IERS/EN/Home/home_node.html. [Accessed 27 March 2016].

[30] NGS Infocenter, “NAD 83 (NSR2007) National Readjusments,” National Geodetic Survey, 3 May 2013. [Online]. Available: http://www.ngs.noaa.gov/NationalReadjustment/. [Accessed 27 March 2016].

[31] Z. Altamimi, “ETRS89,” 28 April 2011. [Online]. Available: http://etrs89.ensg.ign.fr/. [Accessed 27 March 2016].

[32] OpenStreetMap, “About OpenStreetMap,” 2016. [Online]. Available: https://www.openstreetmap.org/about. [Accessed 17 April 2016].

[33] “Definition of Longitude by Merriam-Webster,” [Online]. Available: http://www.merriam-webster.com/dictionary/longitude. [Accessed 27 March 2016].

[34] “Definition of Latitude by Merriam-Webster,” [Online]. Available: http://www.merriam-webster.com/dictionary/latitude. [Accessed 27 March 2016].

[35] GIS Research and Map Collection, “Maps and Cartography: Map Projections,” Ball State University Libaries.

[36] J. E. Campbell and M. Shin, Geographic Information System Basics, 2011.

[37] A. Küpper, Location based services - Fundamentals and Operations, Wiley, 2005.

[38] European Union Road Federation, European Road Statistics - 10th Edition, Brussels, 2011.

[39] G. Rosenblatt, “What is Network Density - and How Do You Calculate It?,” *The Vital Edge,* 2013.

[40] A. Eroglu, E. Onur and H. Oguztuzun, “Estimating Density of Wireless Networks in Practice,” in *IEEE 26th International Symposium on Personal, Indoor and Mobile Radio Communications - (PIMRC): Mobile and Wireless Networks*, 2015.

[41] L. Yang, J. Hao and J. Yang, “The Cyclic Temporal Network Density and Impact on Information Diffusion for Delay Tolerant Network,” in *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, Chengdu, 2010.

[42] N. Velaga, M. Quddus and A. Bristow, “Developing an enhanced weight-based topological map-matching algorithm for intelligent tranport systems,” *Transportation Research Part C: Emerging Technologies,* vol. 17, no. 6, pp. 672-683, 2009.

[43] M. A. Quddus, W. Y. Ochieng, L. Zhao and R. B. Noland, “A general map matching algorithm for transport telematics applications,” *GPS Solutions,* vol. 7, no. 3, pp. 157-167, 2003.

[44] C. E. White, D. Bernstein and A. L. Kornhauser, “Some map matching alhorithms for personal navigation assistants,” *Transportation research part C - Emerging technologies,* vol. 8, pp. 91-108, 2000.

[45] H. Yin and O. Wolfson, “A Weight-based Map Matching Method in Moving Objects Databases,” in *16th International Conference on Scientific and Statistical Database*

Management, Santorini Island, 2004.

[46] M. A. Quddus, W. Y. Ochieng and R. B. Noland, “Current map-matching algorithms for transport applications: State-of-the art and future research directions,” *Transportation Research,* no. 15, pp. 312-328, 2007.

[47] S. K. Honey, W. B. Zavoli, K. A. Milnes, A. C. Phillips, M. S. White and G. E. Loughmiller, “Vehicle navigational sytem and method”. United States of America Patent 4796191, 3 January 1989.

[48] W. Ochieng, M. Quddus and R. Noland, “Map-matching in complex urban road networks,” *Brazilian Journal of Cartography,* vol. 2, no. 55, pp. 1-14, 2003.

[49] F. Torre, D. Pitchford, P. Brown and L. Terveen, “Matching GPS traces to (possibly) incomplete map data: bridging map building and map matching,” in *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, New York, 2012.

[50] M. Fu, J. Li and M. Wang, “A hybrid map matching algorithm based on fuzzy comprehensive judgment,” in *Proceedings of the 7th International IEEE Conference on Intelligent Transport Systems*, Washington, 2004.

[51] M. E. E. Najjar and P. Bonnifait, “A Road-Matching Method for Precise Vehicle Localization Using Belief Theory and Kalman Filtering,” *Autonomous Robots,* no. 19, pp. 173-191, 2005.

[52] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson and P.-J. Norlund, “Particle filters for positioning, navigation and tracking,” *IEEE Transactions on Signal Processing,* vol. 50, no. 2, pp. 425-437, 2002.

[53] Google, “Google Maps,” Google, 2016. [Online]. Available: www.maps.google.com. [Accessed 18 April 2016].

[54] R. Sinnott, “Virtues of the Haversine,” *Sky and Telescope,* vol. 68, no. 2, p. 159, 1984.

[55] C. Veness, “Calculate distance and bearing between two Latitude/Longitude points using haversine formula in JavaScript,” Chris Veness, 2016. [Online]. Available: http://www.movable-type.co.uk/scripts/latlong.html. [Accessed 18 April 2016].

[56] NIST/SEMATECH, “e-Handbook of Statistical Methods,” U.S. Department of Commerce, 30 October 2013. [Online]. Available: http://www.itl.nist.gov/div898/handbook/. [Accessed 15 July 2016].

[57] H.-K. Kim and J.-J. Moon, “Route Tracking of Moving Magnetic Sensor Objects and Data Processing Module in a Wireless Sensor Network,” in *International Conference*

on Smart Manufacturing Application, Gyenonggi-do, 2008.

[58] G. Chen and Z. Li, "Automatic Route Tracking of Smartcar Based on Fuzzy Control," in *Proceedings of the 7th World Congress on Intelligent Control and Automation*, Chongqing, 2008.

[59] NovAtel, "GPS Position Accuracy Measures," 3 December 2003. [Online]. Available: http://www.novatel.com/assets/Documents/Bulletins/apn029.pdf. [Accessed 16 May 2016].

[60] ETSI, "Electronic Fee Collection," ETSI, 2016. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/intelligent-transport/electronic-fee-collection. [Accessed 15 July 2016].

[61] European Space Agency, "ASSIST(FS) - Advanced Snow plough and salt Spreader based on Innovative Space Technologies," ESA, 6 July 2015. [Online]. Available: https://artes-apps.esa.int/projects/assist-fs. [Accessed 15 July 2016].

[62] M. Srivatsa, R. Ganti, J. Wang and V. Kolar, "Map matching: facts and myths," in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Orlando, 2013.

APPENDICES

PERCENTAGE OF SCENARIOS ACHIEVING AT LEAST 95% OF CORRECTLY MAPPED LINK

Table 10: Percentage of P2P scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling period [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	0%	0%	4%	0%
(3,7>	0%	0%	7%	0%
(7,12>	0%	0%	2%	0%
(12,20>	0%	0%	0%	0%
>20	0%	0%	0%	0%

Table 11: Percentage of P2C scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	46%	45%	54%	72%
(3,7>	0%	10%	29%	46%
(7,12>	0%	0%	0%	22%
(12,20>	0%	0%	0%	10%
>20	0%	0%	0%	4%

Table 12: Percentage of C2C scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	10%	9%	5%	19%
(3,7>	8%	0%	4%	14%
(7,12>	0%	0%	0%	15%
(12,20>	0%	0%	0%	7%
>20	0%	0%	0%	4%

Table 13: Percentage of I1 scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	72%	37%	7%	16%
(3,7>	34%	15%	7%	8%
(7,12>	6%	0%	0%	0%
(12,20>	15%	0%	0%	0%
>20	0%	0%	0%	0%

Table 14: Percentage of I2 scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	61%	25%	12%	16%
(3,7>	40%	17%	9%	16%
(7,12>	9%	0%	0%	0%
(12,20>	19%	0%	0%	0%
>20	10%	0%	0%	0%

Table 15: Percentage of I3 scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	82%	74%	12%	21%
(3,7>	61%	50%	9%	14%
(7,12>	10%	0%	0%	0%
(12,20>	16%	0%	0%	0%
>20	10%	0%	0%	0%

Table 16: Percentage of LA1 scenarios achieving 95% of correctly mapped link

Original error CEP-67 [m]	Sampling rate [s]			
	(1,3>	<4,5>	<10,20>	>20
<3	89%	39%	0%	26%
(3,7>	72%	21%	0%	16%
(7,12>	8%	0%	0%	0%
(12,20>	15%	0%	0%	3%
>20	11%	0%	0%	0%