A Novel GNSS-based Positioning System to Support Railway Operations

Sophie Damy

Centre for Transport Studies

Department of Civil and Environmental Engineering

Imperial College London

A thesis submitted for the degree of

Doctor of Philosophy of Imperial College London

2016

Declaration of contribution

During the research presented in this thesis, I have received advice from my supervisors, Prof. Washington Y. Ochieng and Dr Arnab Majumdar. Their contributions were in line with the role defined by Imperial College London for supervisors. Furthermore, I had discussions with various colleagues, researchers and academics relevant to the research presented in this thesis. However, they have minor roles and none constituted a significant input to this research. Given the above, I hereby declare that apart from the statutory role played by the supervisors, the work presented in this thesis has been carried out by myself.

Copyright declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

While railway relies on the knowledge of train position to manage operations, systems currently used to locate trains have a limited accuracy, which limits the network capacity. A positioning system combining Global Navigation Satellite Systems (GNSS) with additional sensors has the capability to provide high accuracy positioning information to support railway operations. However, to reach such accuracy, the different sources of error affecting GNSS measurements need to be mitigated. Among these error sources, multipath signals are a prime concern as they are particularly challenging to mitigate due to their high variability along the train path.

The aim of this thesis is to mitigate the effect of multipath on GNSS code based single point positioning in the railway environment. To achieve this aim, the code multipath error in the railway environment is characterised. Measurement weighting techniques traditionally used to mitigate the effect of multipath on the positioning solution are investigated and a novel weighting method based on the train heading and the satellite position is developed and tested.

The results show variable levels of performance, with no single method outperforming the rest along the entire track dataset analysed. In summary, the elevation and the new heading based methods are shown to complement each other, with the latter performing better (34% of the time) where there are trackside obstacles.

Table of Contents

1 Introduction 21

1.1 Background to the research 21

1.2 Research aim and objectives 23

1.3 Structure of the thesis 24

1.4 Datasets 26

1.4.1 Static dataset 26

1.4.2 High-speed line 26

1.4.3 Suburban mainline 28

2 Railway Operations and Location Based Services 32

2.1 Railway operations and challenges 32

2.1.1 Role of railway transport 32

2.1.2 Railway operations 34

2.1.3 Limitations and challenges 37

2.2 Future railway systems 38

2.2.1 Addressing the current limitations 38

2.2.2 The example of ERTMS 39

2.2.3 Positioning within ERTMS 42

2.3 The role of Positioning, Navigation and Timing in the railway domain 42

2.3.1 Location-based railway applications and services 42

2.3.2 Positioning requirements 46

2.3.3 Existing railway positioning requirements 48

2.4 Conclusions 51

3 Current and Future Positioning Systems to Support Railway Operations 52

3.1 Positioning systems and their performance 52

3.1.1 Railway positioning systems 52

3.1.2 Satellite-based positioning systems 57

3.1.3 Terrestrial sensors 62

3.2 GNSS-based systems in the railway industry 65

3.2.1 Applications targeted by European Union projects 66

3.2.2 Sensors used 68

3.2.3 Performances 70

3.3 Selection of sensors 73

3.3.1 Positioning drivers 74
3.3.2 Performance 75
3.3.3 Multi-sensors integration techniques 77
3.3.4 Railway positioning system physical architecture 79
4 GNSS Errors and Multipath Mitigation 81
4.1 Error sources and mitigation techniques 81
4.1.1 GNSS measurements 81
4.1.2 Error sources and models 83
4.1.3 Error mitigation 86
4.1.4 Position estimation and positioning error 90
4.1.5 Conclusion 97
4.2 Multipath error and mitigation 97
4.2.1 Multipath model 97
4.2.2 Multipath mitigation 101
4.3 Weighting methods 104
4.3.1 Weighted least squares estimate 105
4.3.2 Computation of the weights 107
4.3.3 Limitations 114
4.4 Analysis 114
4.4.1 Static case 115
4.4.2 High-speed train 117
4.4.3 Mainline railway 120
4.5 Conclusion 124
5 A Novel Heading based Weighting Method 126
5.1 Novel weighting method 126
5.1.1 Orthogonal distance 127
5.1.2 Definition of weights 130
5.1.3 An adaptive model 134
5.1.4 Heading determination 137
5.2 POINT Software 140
5.2.1 Cycle slip detection 142
5.2.2 Multipath estimation 144
5.2.3 Hatch filter 156
5.3 Summary 160

6 Characterization of the Novel Heading-based Weighting Method Positioning Performance 161
6.1 Preliminary analysis 161
6.1.1 Reference trajectory 161
6.1.2 Positioning solution 163
6.2 Carrier-smoothed code 167
6.3 Existing weighting techniques 171
6.4 Novel heading based weighting 174
6.4.1 Orthogonal distance-based weighting method 174
6.4.2 Heading based weighting method 177
6.4.3 Adaptive weighting method 180
6.5 Conclusion 182
7 Conclusions and Recommendations 183
7.1 Conclusions 183
7.2 Recommendations for future works 185
References 187
Appendix I – Linearization of the pseudorange equations 195
Appendix II – Least squares estimation 198
Appendix III – Presentations and Papers 200

List of Figures

Figure 1.1 – Structure of the thesis 25

Figure 1.2 - Overlapped train trajectories 27

Figure 1.3 – Train trajectory between Four Oaks and Birmingham New Street 29

Figure 1.4 – Train speed as measured by the INS, after filtering and smoothing 29

Figure 2.1 - Visualisation of the GHG emissions trends of various transport modes between 1990 and 2012 (*** Excluding indirect emissions from electricity consumption), from (European Commission 2011b) 33

Figure 2.2 - EU-28 performance for freight transport (left) and for passenger transport (right), between 1995 and 2013, adapted from (European Commission 2011b) 34

Figure 2.3 - Railway operational architecture 36

Figure 2.4 - Signalling block principle, adapted from (Kyriakidis 2013) 38

Figure 2.5 - Moving block principle 38

Figure 2.6- ETCS level 1 architecture, adapted from (Bloomfield 2006) 41

Figure 2.7 - ETCS level 2 architecture, adapted from(Bloomfield 2006) 41

Figure 2.8 - ETCS level 3 architecture, adapted from (Bloomfield 2006) 41

Figure 2.9 - Maturity of different location-based railway services, from (Marais 2014) 45

Figure 2.10 – Architecture of virtual balises generation to support ERTMS, adapted from (SaPPART 2015) 47

Figure 3.1 – Track circuit principle, the orange arrows represent the current intensity, adapted from (Schon et al. 2013) 53

Figure 3.2 – Axle counter principle, adapted from (Schon et al. 2013) 54

Figure 3.3 – Error in position of the balise centre, adapted from (UNISIG 2012) 55

Figure 3.4 - GBAS principle, adapted from (Julien 2011) 59

Figure 3.5- Accuracy of GNSS positioning techniques, adapted from (Julien 2011) 61

Figure 3.6 - E-GNSS enabled railway signalling – from vision to action, (European GNSS Agency 2016) 73

Figure 3.7 - GPS/INS loose integration, adapted from (Gebre-Egziabher 2007). 78

Figure 3.8 - GPS/INS Tight integration, adapted from (Gebre-Egziabher 2007). 78

Figure 3.9 – Physical architecture of a novel GNSS-based positioning system to support railway location-based services. 79

Figure 4.1 – Carrier phase measurement, adapted from (GIS Resources 2013) 82

Figure 4.2 - Description of the multipath effect 85

Figure 4.3 - Kalman filter data flow, adapted from (Groves 2013) 93

Figure 4.4 – A two-layer neural network 94

Figure 4.5- Relative geometry and dilution of precision: geometry with a low DOP (left), geometry with the high DOP (right), adapted from (Kouwenhoven 2011) 96

Figure 4.6 – Geometry of a single specular multipath, adapted from (Stolagiewicz 2009)..... 98

Figure 4.7 - Correlation functions of a signal subject to constructive interference (orange) and destructive interference (red), the direct signal being represented in blue and the reflected signal is green, from (P. Groves 2013) 99

Figure 4.8 - Code multipath error envelope for different spacing of the correlator d, from (Misra & Enge, 2011, p423) 102

Figure 4.9 - The effect of satellite elevation on the atmospheric error 108

Figure 4.10 – Code measurement error standard deviation as a function of the satellite elevation, for $\sigma_{\rho 0}$= 0.20 m 110

Figure 4.11 - Code noise standard deviation as a function of elevation angle, from (Al-Fanek et al. 2007) 110

Figure 4.12 – Phase error standard deviation as a function of the carrier to noise power density ratio (C/N0) 112

Figure 4.13 - Error variance as a function of the carrier to noise power density ratio (C/N0) and the satellite elevation 113

Figure 4.14- Carrier to noise power density ratio as a function of the satellite elevation..... 115

Figure 4.15 – C/A code noise and multipath on L1 as a function of the satellite elevation angle 116

Figure 4.16 – C/A code noise and multipath on L1 as a function of the C/N0 117

Figure 4.17- Carrier to noise power density ratio as function of the satellite elevation 118

Figure 4.18 - C/A code multipath estimate on L1 as a function of the satellite elevation angle . 119

Figure 4.19 - C/A code multipath estimate on L1 function of the carrier to noise power density ratio 119

Figure 4.20- Skyplot representing the satellite elevation and azimuth angles. 120

Figure 4.21- Carrier to noise power density ratio as a function of the satellite elevation..... 121

Figure 4.22- C/A code multipath estimate as a function of the satellite elevation..... 121

Figure 4.23 – Train position at epoch 476986, (Google Earth 2012) 122

Figure 4.24 – C/A code noise and multipath on L1 and satellite elevation as a function of time for (1) a high elevation satellite PRN 32, (2) a medium elevation satellite PRN 20 and (3) a low elevation satellite PRN 28. 123

Figure 4.25- C/A code multipath estimate as a function of the carrier to noise power density ratio 124

Figure 5.1 - Definition of the orthogonal distance between the satellite and the train heading, as viewed on a skyplot. 127

Figure 5.2 - Orthogonal distance as a function of the satellite azimuth and elevation angles for a heading of 45°. 128

Figure 5.3 – Carrier to noise power density ratio as a function of the orthogonal distance..... 128

Figure 5.4 – Position of the train at Wylde Green Station (orange pin) with the azimuth of satellite 3, (Google Earth 2012) 129

Figure 5.5 – C/A code multipath estimate as a function of the satellite orthogonal distance 130

Figure 5.6 - Measurement noise variance as a function of the satellite elevation angle for different weighting methods 131

Figure 5.7 - Measurement noise variance based on the squared orthogonal distance exponential decay term as a function of the satellite elevation and azimuth, for a heading of 45°, as viewed on a skyplot (right) and as view in 3D (left)..... 131

Figure 5.8 - Measurement noise variance based on the Eisner and Black method as a function of the satellite elevation and azimuth, as view in 3D..... 133

Figure 5.9 - Measurement noise variance of the novel weighting method as a function of the satellite elevation and azimuth, as view in 3D for: (top-left) C =1, (top-right) C = 2, (bottom left) C = 3 and (bottom-right) C = 10..... 133

Figure 5.10 - Measurement noise variance for the novel heading based weighting method and the Eisner & Black method, for different values of C/N0 135

Figure 5.11- Value of parameter D as a function of the pseudorange multipath estimate..... 135

Figure 5.12- Measurement noise variance assigned the heading weighting method as a function of the satellite elevation and azimuth, as view in 3D for (C/N0, multipath estimate) = top-left (45 dBHz, 0.01m), bottom-left (35 dBHz, 0.01m), top-middle (45dBHz, 0.1m), bottom-middle (35dBHz, 0.1m), top-right (45dBHz,1m), bottom right (35dBhz, 1m) 136

Figure 5.13 – Comparison of the train heading measured by the INS (after smoothing) with the GNSS-based heading estimate. 138

Figure 5.14 – Mainline suburban train trajectory at epoch 300 (indicated by the white arrow), as the train exits the Sutton Coldfield tunnel (indicated in red), (Google Earth 2012)..... 139

Figure 5.15 – Architecture of the POINT software, adapted from (Jokinen 2014) 140

Figure 5.16 – Position estimation flowchart 141

Figure 5.17 – Cycle slip effect on (1) the carrier phase measurement, (2) the geometry-free carrier, (3) 1st order and (4) 2nd order geometry-free carrier differences, for a high elevation satellite. 143

Figure 5.18 - Cycle slip effect on (1) the carrier phase measurement, (2) the geometry-free carrier, (3) 1st order and (4) 2nd order geometry-free carrier differences, for a low elevation satellite. 143

Figure 5.19 – L1 C/A code and L2 P(Y) code multipath estimates before averaging for PRN32144

Figure 5.20 - Comparison of the code multipath mean value with the values of the moving averages for different averaging windows and for the cumulative moving average 146

Figure 5.21 – Error between the actual averaged multipath value and the estimated averaged multipath using moving averages 146

Figure 5.22 - Comparison of the code multipath mean value with the values of the moving averages for different averaging windows and for the cumulative moving average, in the presence of a cycle slip. 147

Figure 5.23 - Error between the actual averaged multipath value and the estimated averaged multipath using moving averages, in the presence of a cycle slip 147

Figure 5.24 – Duration of uninterrupted data 148

Figure 5.25 - L1 C/A code and L2 P(Y) code multipath estimates after averaging for PRN32 .. 149

Figure 5.26 – Train speed 149

Figure 5.27 – 3D view and picture of Four Oaks station, from (Google Earth 2012) and (Optimist on the run 2008) 150

Figure 5.28 – 3D view of Gravelly Hill station from (Google Earth 2012) and (Optimist on the run 2007) 150

Figure 5.29 – C/A code multipath estimate on L1 probability density function 153

Figure 5.30 – P(Y) code multipath estimate on L2 probability density function 153

Figure 5.31 – Characterization of the normal distribution fit to the code multipath estimate on L1 for all satellites 154

Figure 5.32 - Measurement error standard deviation as a function of the code multipath estimate 156

Figure 5.33 – Carrier-smoothed code minus carrier observable, for different value of the Hatch filter time constant, when the receiver is static 157

Figure 5.34 – Carrier-smoothed code minus carrier observable, for different value of the Hatch filter time constant, when the receiver is dynamic 157

Figure 5.35 – C/A code multipath estimate on L1 distribution for different values of the Hatch filter time constant 158

Figure 6.1 – Tightly coupled positioning solution standard deviation as estimated by Inertial Explorer 162

Figure 6.2 – Number of visible satellites as a function of the epoch 162

Figure 6.3 – Traditional least square solution North, East and Up positioning errors 164

Figure 6.4 – Position (PDOP), Horizontal (HDOP) and Vertical (VDOP) dilution of precision values 164

Figure 6.5 – Satellite visibility 164

Figure 6.6 – Presence of a cutting between epoch 214 and 242, from (Google Earth 2012) 165

Figure 6.7 – Pedestrian bridge above the track at epoch 412, from (Google Earth 2012) 165

Figure 6.8 – Large highway bridge at epoch 841, from (Google Earth 2012) 165

Figure 6.9 – Positioning error probability density function, cumulative distribution function and 95th percentile 166

Figure 6.10 - Positioning error probability density function (blue), cumulative distribution function (yellow) and 95th percentile (red) over sections of track with a good satellite visibility 167

Figure 6.11 – Positioning error for different values of the Hatch filter time constant 167

Figure 6.12 - Positioning error for different values of the Hatch filter time constant, between epoch 290 and 670 168

Figure 6.13 – Train trajectory between epoch 353 and 374, from (Google Earth 2012) 169

Figure 6.14 – Positioning error distribution with and without carrier-smoothed code, for different values of the filter time constant 169

Figure 6.15 – Hatch filter time constant producing the minimum positioning error as a function of the train position. 170

Figure 6.16 – Train trajectory between Chester Road and Erdington station, from (Google Earth 2012) 171

Figure 6.17 – Weighting method producing the minimum positioning error as a function of the train position..... 173

Figure 6.18 – Weighting method providing the smallest positioning error at each epoch, between the elevation method (blue) and the orthogonal distance (green)..... 176

Figure 6.19 - Weighting method providing the smallest positioning error at each epoch, between the sigma-ε method (blue) and the orthogonal distance (green)..... 176

Figure 6.20 - Weighting method providing the smallest positioning error at each epoch, between the sigma-ε method (blue) and the heading based method (green)..... 179

Figure 6.21 - Weighting method providing the smallest positioning error at each epoch, between the elevation method (blue) and the heading based method (green) 179

Figure 6.22 - Weighting method providing the smallest positioning error at each epoch, between the elevation method (blue) and the adaptive method (green)..... 181

Figure 6.23 - Weighting method providing the smallest positioning error at each epoch, between the sigma-ε method (blue) and the adaptive method (green)..... 182

List of Tables

Table 1.1 - Train journeys analysed 27

Table 1.2 – Dataset used in each chapter 30

Table 2.1 - ETCS levels of operations 40

Table 2.2 – Location-based railway applications, adapted from (RSSB 2012b) 44

Table 2.3 – Railway application positioning requirements, adapted from (RSSB 2012b) 49

Table 2.4 – Selected railway positioning requirements 50

Table 2.5 – Additional positioning requirements defined in (RSSB 2012b) 51

Table 3.1 - Performance of railway positioning systems 57

Table 3.2- Positioning performance of GNSS, adapted from (Scot 2014) 61

Table 3.3- Inertial navigation systems grade, performance and cost, from (Groves 2013) 62

Table 3.4 - Signal of Opportunity positioning accuracy, adapted from (Groves 2013) 64

Table 3.5 - GNSS-based European railway projects 67

Table 3.6 - Positioning sensors used in GNSS-based European railway projects 69

Table 3.7 – Satellite visibility 71

Table 3.8 – Duration of GNSS outages 72

Table 3.9 - Accuracy of GNSS-based positioning systems in the railway environment 72

Table 3.10 - Positioning performances and railway requirements 76

Table 4.1 - A summary of the errors in GPS measurements (estimated based on a 10 km baseline and a signal latency of 10 sec), adapted from (Misra & Enge 2010) 89

Table 4.2 - Typical User Equivalent Range Error budget for single frequency GNSS, dual frequency GNSS and differential GNSS at one sigma, adapted from (Hegarty 2013; Misra & Enge 2010) 96

Table 4.3 – Values of the coefficient in the stochastic elevation model developed by Elsobeiey & El-Rabbany (2010) 109

Table 5.1 – Code multipath estimate availability per satellite 151

Table 5.2 – Mean value μ and standard deviation σ of the normal distribution used to represent the code multipath estimate distribution per satellite and for all satellites 155

Table 5.3 – Effect of the Hatch filter time constant on the code multipath estimate standard deviation 159

Table 5.4 – Carrier-smoothed code availability per satellite 159

Table 6.1 – Error sources mitigation 163

Table 6.2 – Characterization of the Hatch filter time constant on the positioning error over the entire track 169

Table 6.3 - Characterization of the Hatch filter time constant on the positioning error for over part of the track with a good satellite visibility (DOP<5) 170

Table 6.4 - Characterization of the impact of the different weighting methods on the positioning error over the entire track 172

Table 6.5 - Characterization of the impact of the different weighting methods on the positioning error over selected sections of track with a good satellite visibility (DOP<5) 172

Table 6.6 - Percentage of epochs each weighting method produces the smallest positioning error 173

Table 6.7 - Characterization of the different weighting methods impact on the positioning error over the entire track 175

Table 6.8 - Characterization of the different weighting methods impact on the positioning error over selected sections of track with a good satellite visibility (DOP<5) 175

Table 6.9 – Comparison of the orthogonal distance-based weighting method with the current weighting methods 175

Table 6.10 - Characterization of the impact of the different weighting methods on the positioning error over the entire track 178

Table 6.11 - Characterization of the impact of the different weighting methods on the positioning error over selected sections of track with a good satellite visibility (DOP<5) 178

Table 6.12 - Comparison of the heading based weighting method with the current weighting methods 178

Table 6.13 - Characterization of the impact of the different weighting methods on the positioning error over the entire track 180

Table 6.14 – Characterization of the impact of the different weighting methods on the positioning error over selected sections of track with a good satellite visibility (DOP<5) 180

Table 6.15 - Comparison of the adaptive based weighting method with the current weighting methods 181

Acronyms

3INSAT	Train Integrated Safety Satellite System
4Cs	Four challenges
AoA	Angle of Arrival
AoD	Age of Data
AM	Amplitude Modulation
APOLO	Advanced Positioning Locator
ATO	Automatic Train Operations
ATP	Automatic Train Protection
AWS	Automatic Warning System
CO2	Carbon Dioxide
C/A	Coarse/Acquisition
C/N0	Carrier to noise power density ratio
CCTV	Closed-circuit television
DAS	Driver Advisory System
DGNSS	Differential GNSS
DOP	Dilution of Precision
EATS	ETCS Advanced Testing and Smart Train Positioning System
ECORAIL	EGNOS Controlled Railway Equipment
EGNOS	European Geostationary Navigation Overlay Service
EGNOS-R	EGNOS for Railway
EOR	European Operational Rules
ERA	European Railway Agency
ERTMS	European Railway Traffic Management System
ERSAT	ERTMS on Satellite – Enabling Application Validation
ETCS	European Train Control System
ETML	European Traffic Management Layer
FAA	Federal Aviation Authority
GADEROS	Galileo Demonstrator Railway Operational System
GALILEO	European Global Navigation Satellite System
GALOROI	Galileo Localization for Railway Operation Innovation
GBAS	Ground-based Augmentation System
GDOP	Geometric Dilution of Precision
GEORAIL	Railway geodesy

GHG	Greenhouse Gases
GIRASOLE	Galileo Integrated Receiver for Advanced Safety of Life Equipment
GLONASS	Russian Global Navigation Satellite System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GRAIL	GNSS introduction in the rail sector
GRAIL2	GNSS-based enhanced odometry for railway
GSA	European GNSS Agency
GSM-R	Global System for Mobile Communications - Railway
HDOP	Horizontal Dilution of Precision
ICAO	International Civil Aviation Organisation
IGS	International GNSS Service
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
INTEGRAIL	Intelligent Integration of Railway Systems
IRISS	Intelligent Railway via Integrated Satellite Services
KF	Kalman Filter
LOCOPROL	Low-cost satellite-based train location system for signalling and train protection for low-density traffic railway lines
LAMBDA	Least-squares AMBiguity Decorrelation Adjustment
LBS	Location Based Services
LHCP	Light-Hand Circular Polarization
LOCOLOC	Locomotive Location
LTE	Long Term Evolution
MoU	Memorandum of Understanding
MEDLL	Multipath Estimation Delay Lock Loop
MEMS	Micro Electro Mechanical Sensor
MTBF	Mean Time Between Failures
NGTC	Next Generation Train Control
NLOS	Non-line-of-sight
NSL	Nottingham Scientific Limited
OEM	Original Equipment Manufacturers
PDOP	Position Dilution of Precision
PNT	Position, Navigation, Timing
PPP	Precise Point Positioning
PRN	Pseudo-Random Noise

PVT	Position, Velocity, Time
R	Pseudorange
RAMS	Reliability, Availability, Maintainability, and Safety
RINEX	Receiver Independent Exchange Format
RHCP	Right-Hand Circular Polarisation
RMSE	Root mean square error
RNP	Required Navigation Performance
RSS	Received Signal Strength
RSSB	Rail Safety and Standards Board
RTK	Real Time Kinematic
RUNE	Railway User Navigation
SaPPART	Satellite Positioning Performance Assessment for Road Transport
SATLOC	Satellite-based operation and management of local low traffic lines
SBAS	Satellite-based Augmentation System
SIL	Safety Integrity Level
SOOP	Signals of Opportunity
SV	Satellite Vehicle
T510	Obtaining data to assess the dependability of GNSS information and accuracy of odometry
T740	Global navigation satellite systems data coverage analysis for railway operations
T892	Data analysis for cost-effective GPS-locator with simple augmentations
TDoA	Time Difference of Arrival
TDOP	Time Dilution of Precision
TEC	Total Electron Content
TECU	Total Electron Content Unit
TEN-T	Trans-European Transport Networks
ToA	Time of Arrival
TPWS	Train Protection and Warning System
TSAG	Technical Strategy Advisory Group
TSLG	Technical Strategy Leadership Group
TTA	Time to Alert
TTFF	Time to First Fixed
UERE	User Equivalent Range Error
UK	United Kingdom
UNB	University of New Brunswick
UNIFE	Association of the European Rail Industry

UNISIG	Union of Signalling Industry
VDOP	Vertical Dilution of Precision
WAAS	Wide Area Augmentation System

Acknowledgements

First and foremost I would like to thanks my supervisors Professor Washington Yotto Ochieng and Dr Arnab Majumdar for giving the opportunity to conduct my PhD at Imperial College and for providing me support, advice and encouragements during my studies and the writing of this thesis. My deepest appreciation and gratitude goes to them. I also would like to thanks Dr Khalid Nur, Dr Ramin Moradi and Dr Shaojun Feng for their friendly help and advice.

I am exceedingly grateful to the Lloyd's Register Foundation for funding my PhD and to Nottingham Scientific Ltd, in particular to Michael Hutchinson and Mark Dumville, for providing me with the data used in this thesis.

I would like to thanks everyone in the Department of Civil Engineering who made this experience enjoyable, through both the bad and good times. Special thanks to Fionnuala for her constant support and to all my great friends from the 6th floor: Severine, Ramandeep, Sarah, Thalis, Ruben, Pablo, Elena, Nicolo, Francesco, Sally, Nils, Bianca, Shane, George Koudis, Ara, Georges Goldberg, Sebastian, Wenlong, Seo.

To my lovely Milena, many thanks for your help, your support and for being so great!

Special thanks go to Simone for supporting me patiently (especially during the writing up period!) and for being by my side throughout this journey that started in Canada; let's see where it takes us next.

To my beloved family, Aude, Isabelle and Eric, for their unconditional love, support and encouragements in all the projects I ever wanted to carry out. This is for you.

Chapter 1

Introduction

Railway operations increasingly rely on the real-time knowledge of train position to manage railway traffic on their network. Global Navigation Satellite Systems (GNSS), in combination with additional sensors (e.g. inertial sensors) and a spatial database, have the capability to provide the level of position accuracy required to support these operations. However, to reach such accuracy, the different error sources impacting GNSS measurements need to be mitigated. Prime amongst these errors are the multipath signals created by signal reflection on various obstacles that surround the receiver. Such errors are particularly challenging to mitigate as they vary along the path of a train's travel. The work presented in this thesis addresses this challenge by mitigating the effect of code multipath on the estimated position in the railway environment.

The first chapter presents a brief background of the research followed by the aim and objectives. It highlights the main contributions of this thesis and describes its structure. Finally, a description of the data used in the following chapter is provided.

1.1 Background to the research

The demand for railway passenger and freight transport is expected to grow in the future. In Europe over the next 15 years, an annual increase of 1.5% in passenger traffic and 2.3% in the freight traffic is foreseen (European Commission 2010). In order to accommodate this growth, the capacity of the European railway network needs to be increased. This could be done building new infrastructures. However, a more efficient use of the current network already has the potential to increase capacity in both the short and medium terms, while generating the information required to justify the need for new infrastructure. To enable the optimisation of the railway traffic flow in real time, the provision of each train's position on the network is required. This knowledge, to be used in future intelligent train control systems, allows the traffic controller to run trains based on their braking distances, safely reducing the current separation distances imposed by the signalling system design. Intelligent train control systems, such the European Railway Traffic Management System (ERTMS), have the potential to increase the line capacity by up to 40% (UNIFE 2012a). Part of this increase can be imputed to the knowledge of the train position, however, the attribution of this budget to the different elements of the system is complex and still remains to be done.

In addition to supporting traffic control, the knowledge of train position on the network can benefit numerous railway applications. These range from customer services, such as informing the passengers of the estimated arrival time of the train, to automated infrastructure maintenance (see details in section 2.3.1). Each of these applications has its own set of requirements defined in terms of accuracy, integrity, continuity and availability of the position.

However, current railway positioning systems, such as track circuits, odometers and axle counters, are unable to provide the performance required by most of the applications. Moreover, they have a high failure rate and are costly to maintain. An alternative positioning system, able to improve accuracy and continuity and also to detect and exclude failures (i.e. position errors larger than a defined protection level), is then required to support the wide range of railway applications envisaged.

Among potential alternative systems, GNSS technology has been identified as the core element for a novel railway positioning system, due to its interoperability with existing systems as well as its low cost and wide coverage (European Commission et al. 2012). Since 2000, numerous projects have been carried out to assess the performance of GNSS-based positioning in the railway domain. Several studies have attempted to define the required navigation performance for railway operations (section 2.3.3), however, no standard positioning requirements are currently available. Several positioning systems using combination of sensors were also developed and tested, with varying performance (section 3.2.3). One of the main challenge to the use of GNSS still resides in the environment the trains travel through. Numerous obstacles such as cuttings, trackside buildings and tunnels can significantly attenuate or block the GNSS signals. These obstacles can also be responsible for signal reflections, named multipath signals, causing the signal coming from a satellite to reach the GNSS antenna via two or more paths, which in turn create a bias in the positioning solution. In order to improve the performance and maintain the continuity of the positioning service in the railway environment, the GNSS receiver has to be integrated with additional systems, sensors and spatial databases. Such systems and sensors include map matching, inertial sensors, wheel sensors and the positioning using signals of opportunity, such as Wi-Fi.

While acknowledging the need to integrate the GNSS receiver with additional sensors, this thesis focuses on improving the performance of the GNSS receiver itself, in the railway environment, with a particular focus on code (pseudorange) based positioning. To do so, the different sources of errors that affect the code measurements are mitigated. While GNSS modernization and augmentation have enabled the mitigation of numerous sources of errors coming from the satellite and the atmosphere, multipath remains a significant contributor to the positioning error. This is especially true in the railway environment where regular obstacles along (e.g. buildings, cuttings) and above (e.g. bridge) the track have the potential to generate a large amount of reflections.

Several techniques have been developed to mitigate the effect of code multipath on the position, including antenna design, receiver design and measurement processing techniques. While the two first methods require modification of the hardware and software of the receiver, the later techniques can be implemented with any type of receivers (providing GNSS raw measurements) and are the focus of this thesis.

One measurement processing method to mitigate the effect of multipath signals on the estimated position relies on the weighting of the GNSS measurements. After estimating which measurements are more likely to be affected by errors, their contribution to the final positioning solution is constrained using weights. The measurement errors are estimated as a function of different parameters, such as the satellite elevation angle or its signal strength (i.e. carrier to noise power density ratio). For example, low elevation satellites are assumed to be more likely affected by measurement errors due to their larger atmospheric residual errors and their increased probability to be affected by multipath. While these models are adapted to measurements generated by a static receiver, it is shown in section 4.4 that their validity is reduced in the case of a dynamic receiver, mainly due to the high variability of the environment, which makes relations between the parameters and the measurement noise less predictable. The incorrect determination of the measurement noise can, in turn, decrease the performance of the position estimation and reduce GNSS positioning accuracy in the railway domain.

To date, the only weighting method for code multipath in the railway domain was developed by (Marais et al. 2015) and relies on the use of a camera and image processing methods to determine the satellite state of reception (i.e. line and non-line sight). No weighting method relying on GNSS alone has yet been developed. This is the focus of this thesis.

1.2 Research aim and objectives

Given the effect of code multipath on the train position estimation and the limitations of existing weighting methods, the aim of this thesis is to develop an improved code multipath mitigation technique based on the traditional methods but taking into account the unique features of the railway, in particular, the track geometry. To achieve this aim, the following objectives have been defined:

1. Undertake a critical review of the railway operations and identify the railway applications that require the knowledge of the train position, together with their required navigation performance.

2. Identify the existing railway positioning systems and techniques, assess their positioning performance with respect to the requirements identified in the first objective.

3. Review the alternative positioning systems and techniques along with the corresponding required navigation performance, and proposed an architecture for a possible railway positioning system.

4. Review GNSS error sources and associated mitigation techniques with a focus on code multipath mitigation techniques and their validity in the railway environment.

5. Develop a novel weighting technique to mitigate the effect of code multipath signals on the estimated position, in the railway environment and implement the new algorithm in positioning software to enable testing.

6. Test and validate the developed algorithms using GNSS data collected onboard a train, in order to develop a strategy for the implementation of the novel positioning system

1.3 Structure of the thesis

This thesis is organised into seven chapters, starting with the introduction. Chapter 2 reviews the opportunities and challenges facing the railway industry and explains the role of train positioning to support future railway traffic management systems and other location-based railway applications. Existing railway positioning requirements are presented at the end of the chapter and a set of requirements is selected as the basis for a future railway positioning function, addressing the first objective presented in section 1.2 and contributing to the understanding of the railway industry needs in terms of positioning.

Chapter 3 reviews the systems and sensors currently used in the railway industry to determine the train position, and their performance in terms of accuracy, integrity, continuity and availability. It also reviews alternative positioning systems that could be used to determine train position. Based on the requirements defined in chapter 2, a positioning system architecture, based around a GNSS receiver, is specified, addressing the second objective defined in section 1.2. The choice is made to focus on the GNSS receiver to address the key limitation that is presented by code multipath error in the railway environment.

Chapter 4 reviews the different sources of error that can affect the GNSS receiver and identifies multipath as the major contributor to the positioning error in the railway environment. Several multipath mitigation techniques are then presented, including weighting-based techniques. This chapter addresses the fourth objective and identifies the limitations of existing weighting methods in the railway environment.

Based on these limitations, chapter 5 develops a novel heading based weighting method, addressing the fifth objective. It also introduces the software used to assess the performance of this novel method and the different functions implemented in the software to perform the analysis. The results on multipath mitigation in the measurements domain are presented as well as the characterization of the multipath effect on the code measurements.

Chapter 6 presents the performance of the different multipath mitigation techniques in the positioning domain, addressing the sixth objective and contributing to the validation of the novel method introduced in Chapter 5. In this chapter, the effect of code smoothing using carrier phase is presented as well as a comparison between the traditional and the novel weighting methods.

Finally, chapter 7 presents the conclusions of this thesis and discusses developments to improve the estimation of code multipath in the railway environment and the weighting of the measurements.

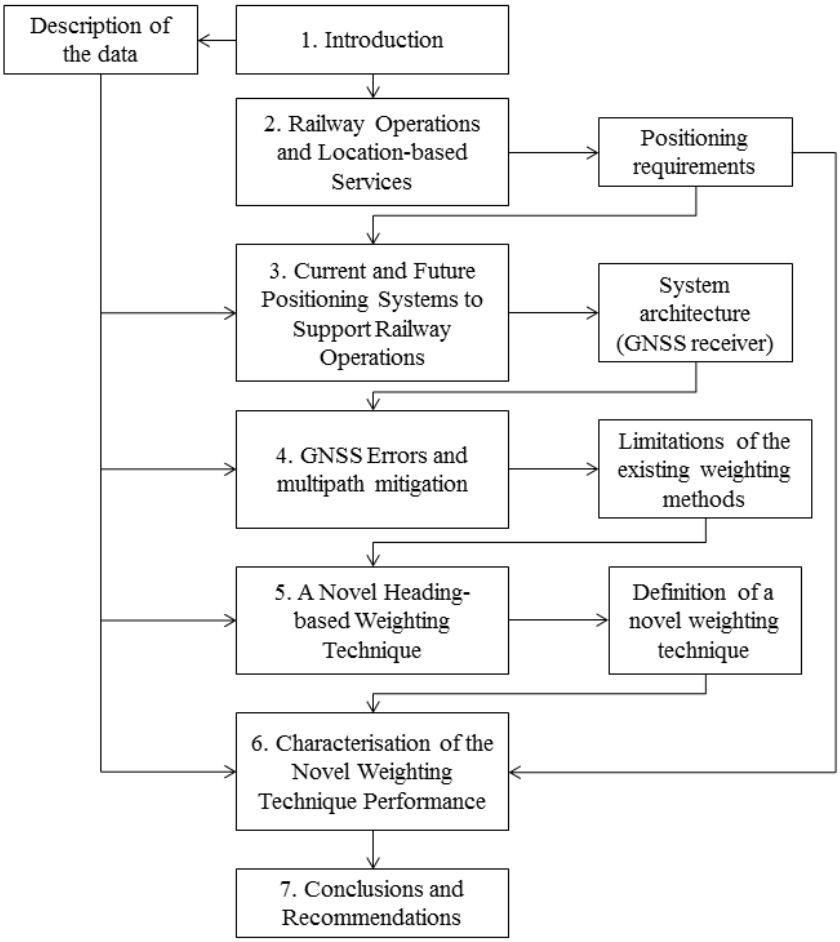


Figure 1.1 – Structure of the thesis

Figure 1.1 represents the overall structure of the thesis and explains how the main contribution of each chapter (on the right-hand side) is used as the basis for the following chapter. The left-hand-side of the

figure shows the chapter in which the data presented in section 1.5 are used. More details on the use of the data are provided in Table 1.2.

In addition to the work presented in this thesis, a list of presentations and publications relative to this research can be found in Appendix III.

1.4 Datasets

The following chapters 2 and 3 present a review of the literature regarding positioning in the railway industry. This review is supported by preliminary analysis of datasets further used in chapters 4, 5 and 6. In order to provide the reader with an understanding of the datasets throughout the thesis, the following subsection provides a description of their characteristics. Three datasets are used and are referred to as 'the static dataset', 'the high-speed line' and 'the suburban mainline'. More technical details on the positioning systems described in this section can be found in Chapter 3.

1.4.1 Static dataset

This first dataset was collected by (Moradi 2014) in Silwood Park on the 9th of May 2012 using a Leica GS15 receiver. The receiver was static for approximately eight hours and was installed in a relatively open area, but the presence of trees and a sloping terrain below antenna level could cause signal reflections.

These data are used in this thesis to compare GNSS measurements characteristics of a static (i.e. at a fixed location) and a dynamic (i.e. in movement) receiver.

1.4.2 High-speed line

This second dataset was collected in the scope of the GNSS introduction in the rail sector (GRAIL) project (GRAIL 2006b), outlined further in Chapter 3. The GNSS receiver (described in 1.4.2.1) was installed onboard a Spanish high-speed train, class 100 AVE (Alta Velocidad Española), operated by Renfe (National Network of Spanish Railway) and capable of reaching a speed of 300 km per hour.

1.4.2.1 Receiver

The receiver used for this data collection is a Novatel Original Equipment Manufacturers (OEM) dual frequency GPS receiver. According to its specification (Novatel 2016a), this receiver can reach a horizontal accuracy (RMS) of 1.2 m in single point positioning using dual-frequency (L1 and L2), which can be further reduced to 0.4 m if differential GPS is used. The receiver expected time for signal reacquisition on L1 is 0.5 second and 1.0 second on L2.

The raw data collected by the receiver is stored in Receiver Independent Exchange Format (RINEX) format and contains the pseudorange, carrier phase, Doppler frequency and carrier to noise ratio (C/N0) values for each satellite on L1 and L2.

1.4.2.2 Dataset description

The dataset covers eight different journeys, mainly on the Madrid-Levante high-speed rail line between Madrid and Albacete but also between Madrid and Valencia, and between Madrid and Barcelona. These different journeys can be visualised on Figure 1.2 and are summarised in Table 1.1.

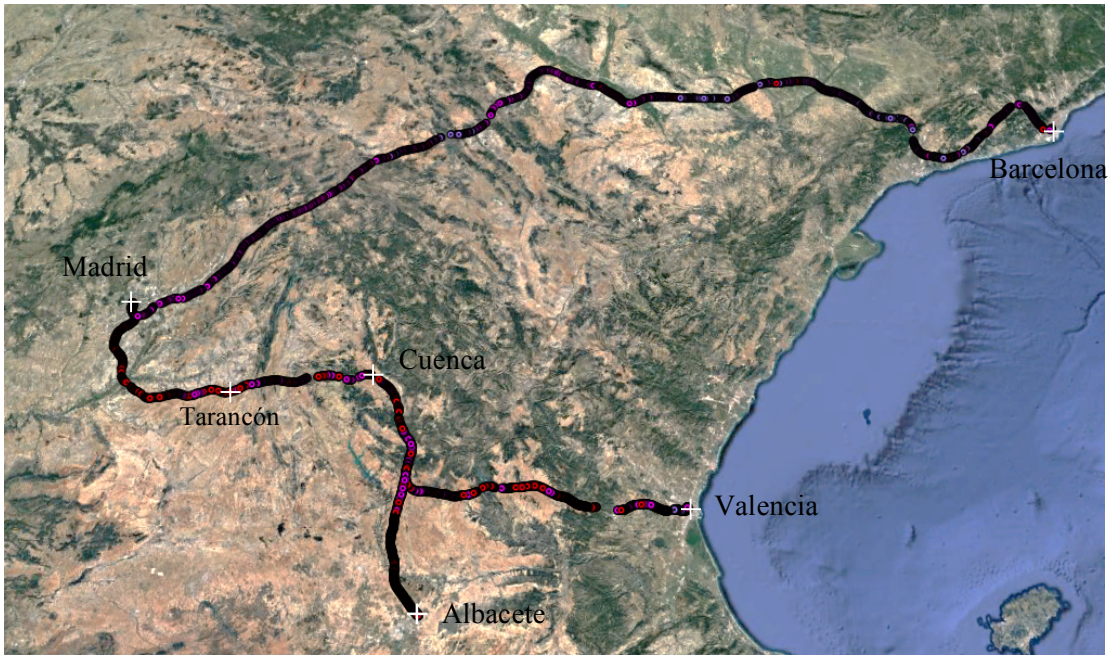


Figure 1.2 - Overlapped train trajectories

Table 1.1 - Train journeys analysed

Date	Origin / Destination	Duration (seconds)
17/11/2011	Madrid ↔ Albacete	23340
20/11/2011	Madrid ↔ Cuenca	5311
21/11/2011	Madrid → Valencia → Albacete → Madrid	11541
29/11/2011	Madrid ↔ Albacete	26578
30/11/2011	Madrid ↔ Albacete	7498
04/03/2012	Madrid ↔ Tarancón	4368
08/03/2012	Madrid ↔ Barcelona	12268
22/03/2012	Madrid ↔ Ocana	12035

A pre-selection process was performed on the data to remove long static periods both at the beginning and at the end of each journey, as these correspond to train preparation for data collection and do not reflect regular train operations. After this selection, a total of 102,939 epochs were analysed, corresponding to over 28 hours of data.

1.4.2.3 Limitations

Due to the lack of a reference trajectory to perform the comparison in the positioning domain, but given the extensive amount of data available, this dataset is used to assess the characteristics of GNSS measurements in the railway environment (e.g. satellite visibility, availability of the measurements).

1.4.3 Suburban mainline

This dataset was collected by Nottingham Scientific Limited (NSL) within the scope of the T892 project entitled "Data and analysis for a cost-effective GPS-based locator with simple augmentations" (RSSB 2012b). The receiver (described in section 1.4.3.1) was installed onboard a London Midland Class 323 (323-202) train (maximum operational speed of 90 mph or 145 km per hour) operating in the West Midlands, in particular, the suburban areas around Birmingham.

1.4.3.1 Receiver

A GPS and Inertial Measurement Unit (IMU) Novatel SPAN-CPT receiver was installed on board the train. The SPAN-CPT receiver provides a hybrid GPS and inertial solution that is post-processed to generate the reference trajectory. The GPS receiver is dual-frequency and can provide a GPS only position. The IMU has six degrees of freedom, which means it measures accelerations and angular rates along all three axes. According to its specification (Novatel 2016b), the Novatel SPAN-CPT can reach 1.2 m horizontal accuracy (RMS) in single point positioning using L1 and L2. The gyroscope bias is 20 degrees per hour and the accelerometer bias 50 mg, equivalent to approximately 0.5 m/s^2. During GNSS outages of 10 seconds, the receiver can maintain a horizontal accuracy (RMS) of 1.20 m and a vertical accuracy (RMS) of 0.75 m. For an outage of 60 seconds, the horizontal accuracy (RMS) is limited to 7 m and vertical accuracy (RMS) to 2.60 m.

A second receiver was installed on board the train, the OEMv1G receiver which acquired data from both GPS and Global Navigation Satellite System (GLONASS) on L1. According to its specification (Novatel 2011), it can achieve a horizontal accuracy (RMS) of 1.5 m in single point positioning. If differential GPS is used, this accuracy can be reduced to 0.4 m. Its expected time of signal reacquisition on L1 is 0.5 second.

The 1553 epochs of raw data collected by both receivers are stored in Receiver Independent Exchange Format (RINEX) format. The SPAN-CPT file contains the coarse/acquisition (C/A) code pseudorange

on L1, the P(Y) code pseudorange on L2, the carrier phase on L1 and L2, and Doppler frequency and carrier to noise ratio (C/N0) values for each satellite. The OEMv1G file contains similar information, but only on L1.

1.4.3.2 Dataset description

The journey analysed in this thesis lasted for approximately 25 minutes (1553 epochs) and took place between Four Oaks and Birmingham New Street on the 14th of December 2012, as shown in Figure 1.3.

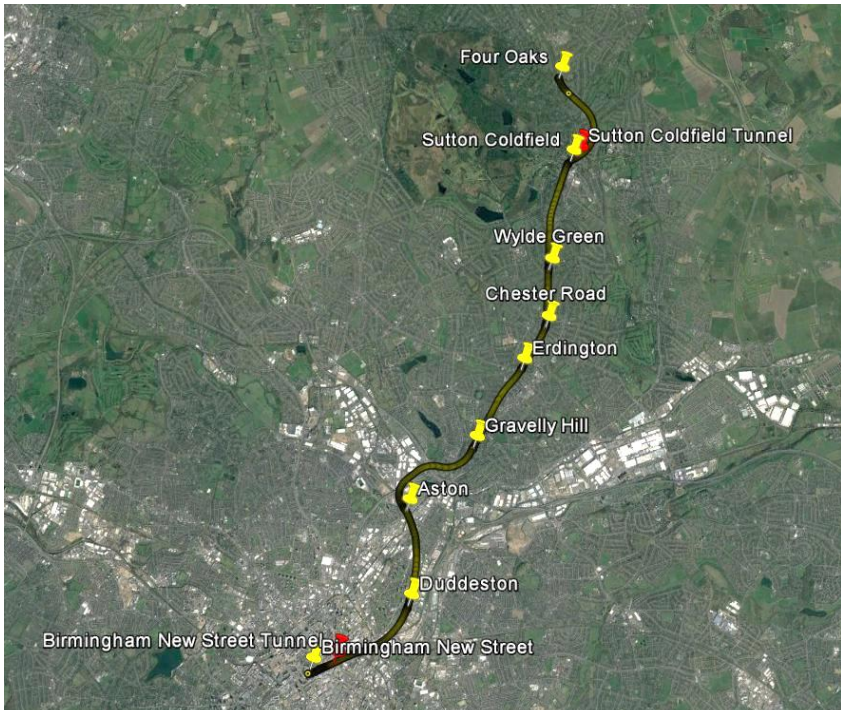


Figure 1.3 – Train trajectory between Four Oaks and Birmingham New Street

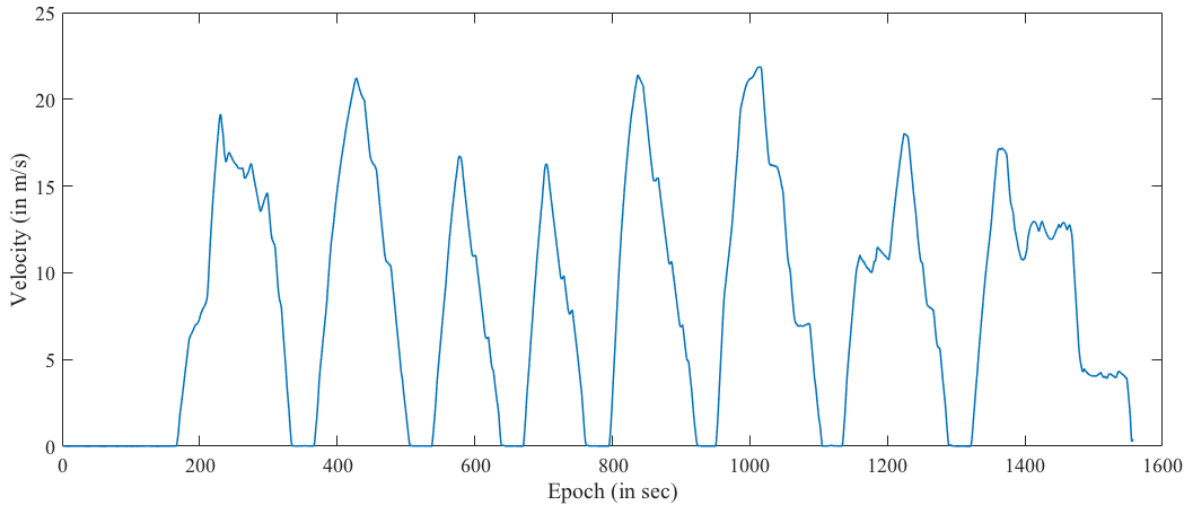


Figure 1.4 – Train speed as measured by the INS, after filtering and smoothing

During its journey, the train stopped at nine stations (indicated by yellow pins on Figure 1.3). It also travelled through two tunnels (indicated by red pins), one in Sutton Coldfield with a length of approximately 160 meters, and one when entering Birmingham New Street, where the train stopped. The train average speed between stops is 12.3 m per seconds (approximately 44 km per hour), with a maximum speed of 25 m per second (90 km per hour), as shown in Figure 1.4.

1.4.3.3 Limitations

The suburban mainline dataset is used to assess the impact of the code multipath mitigation techniques investigated in this thesis on the position. The IMU data is integrated with the dual frequency GPS receiver data in order to produce a reference trajectory (see section 6.1.1 for detail). However, the same GPS data is used for the analysis as dual frequency measurements are needed to compute the code multipath estimate (see section 4.2.1.3). This implies that errors in the GNSS measurements impact the analysed data and also have the potential to impact the reference trajectory in a similar manner. The assessment of the accuracy of the reference trajectory is discussed in detail in section 6.1.1.

Additionally, the suburban mainline dataset is relatively short (25 minutes) compared to the typical duration of a train journey. This could potentially limit the statistical significance of the results presented in this thesis. However, due to the diversity of the environment the train travels through during these 25 minutes (e.g. tunnels, along track-side buildings and trees), the results should provide a good indication of GNSS positioning performance in the railway domain.

Table 1.2 – Dataset used in each chapter

Analysis	Static dataset	High-speed rural railway	Suburban mainline railway (SPAN-CPT)	Suburban mainline railway (OEMv1G)
Satellite availability in the railway environment (Table 3.7)		✓	✓	✓
Duration of GNSS outages (Table 3.8)		✓	✓	✓
Carrier to noise ratio function of the satellite elevation angle (Section 4.4)	✓	✓	✓	
Code multipath function of the satellite elevation and carrier to noise ratio (Section 4.4)	✓		✓	
A novel heading based weighting method (Chapter 5)			✓	
Characterization of the novel heading-based weighting method positioning performance (Chapter 6)			✓	

The three datasets are used in different analyses and Table 1.2 summarises which dataset is used for each analysis and the sections of this thesis outlining the results.

Chapter 2

Railway Operations and Location Based Services

This chapter demonstrates the significant role played by positioning in supporting railway operations. It starts by explaining the role of rail transport in Europe and the challenges it faces. The principles of railway operations are then reviewed and their limitations identified. Subsequently, future railway systems addressing these limitations are presented, with the example of the European Rail Traffic Management System (ERTMS). The importance of positioning within these systems is then demonstrated, leading to the review of the location-based railway services along with their positioning requirements. The main contribution of this chapter is to outline a set of positioning requirements used to specify the positioning system architecture in chapter 3.

2.1 Railway operations and challenges

2.1.1 Role of railway transport

A railway is a land-based, guided form of transport which aims to move people and goods while minimising the occupation of the tracks, the cost of operations and the impact it has on the environment (Fontgalland 1984). Currently, in Europe, railways account for 6% of the passenger market and approximately 16% of the inland freight market (European Commission 2011b). In contrast, road transport accounts for more than 70% of all passenger and freight traffic. However, railways present numerous benefits compared to road transport. Firstly, it is one of the safest forms of transport with a risk of fatality seven times lower than that of road transport (Profillidis 2013). Additionally, its environmental impact is relatively low. Not only do trains produce less greenhouse gas (GHG) emissions (in 2008, only 0.8% of the transport total carbon dioxide (CO2) emission in Europe was due to the railway while road transport was responsible for 70.7%, see Figure 2.1), they also consume less energy (in 2008 in Europe, railways accounted for 2.5% of the total energy consumption while road transport was responsible for 81.3% (European Commission 2011b)).

Despite its current lower share of the transport demand and the difficulties presented by the planning and construction of new railway infrastructure, the benefits of rail transport present an opportunity to address some the challenges posed by road transport.

The objective of the European Commission to significantly reduce GHG emissions to 80% below their 1990 levels by 2050 (European Commission 2012) and the need to address congestion should lead the

rail transport capturing a larger share of the transport demand. In addition to the benefits related to emissions and energy consumption, a railway is also space efficient. It has been estimated that a journey from home to work by metro consumes ninety times less urban space than the journey by car (European Commssion 2004). As cities grow bigger and more crowded, urban space becomes scarce, making space efficient transport valuable. These factors have the potential to make railway transport the preferred mode of travel.

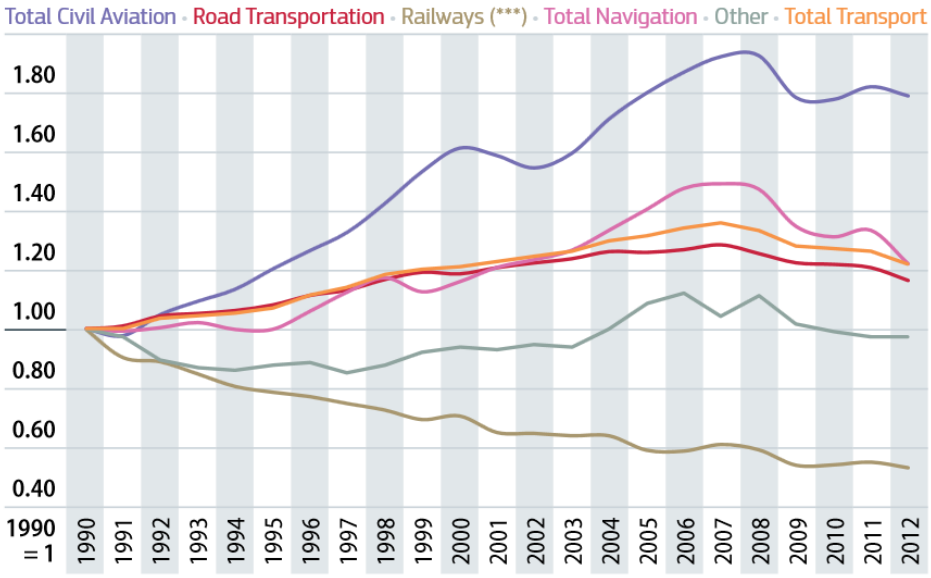


Figure 2.1 - Visualisation of the GHG emissions trends of various transport modes between 1990 and 2012 (*** Excluding indirect emissions from electricity consumption), from (European Commission 2011b)

Based on these many benefits, the European Commission predicts an eight times increase in freight moved by rail and a twelve times increase in passenger travel by 2050 (European Commission 2011a). In short term, travel in passenger kilometres is expected to increase from 19.3 billion in 2013 to 26 billion in the United Kingdom (UK) in 2030 (UK Parliament 2012). Over the same period, movement by freight is also expected to increase from 31 to 50.4 billion of tonnes (Freight Transport Association 2008).

However, railway transport faces a significant competition, mainly from the road sector, as shown in Figure 2.2. To realise the predicted capture of a larger share of the transport demand, railways need to increase the capacity of the current rail networks must be increased and traffic management optimised to improve efficiency. The European Commission (2011) states that structural changes are required for rail to compete effectively and take a greater share of medium and long distance transport, for both passengers and freight. This includes changes to the current infrastructure and the concept of operations.

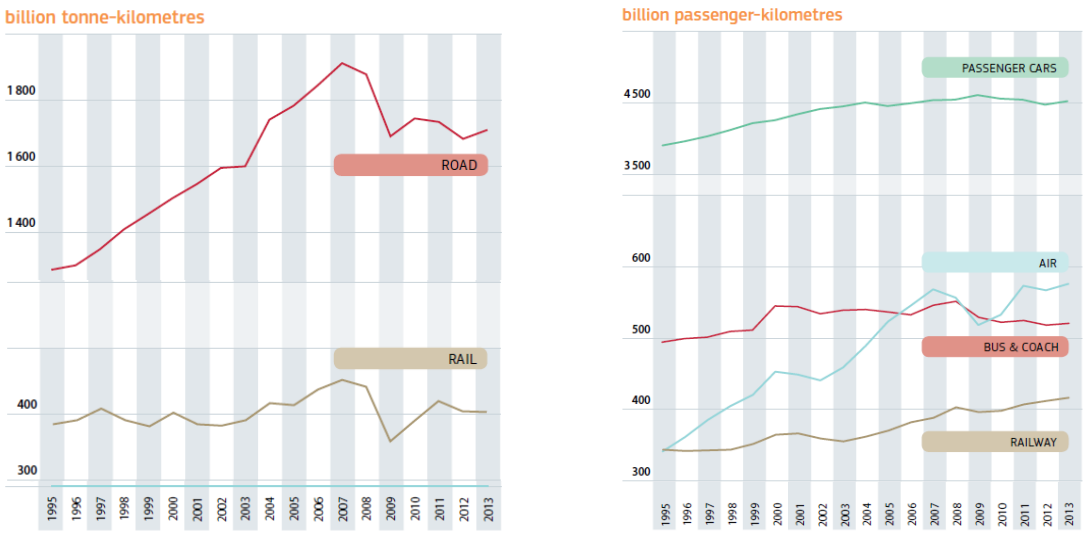


Figure 2.2 - EU-28 performance for freight transport (left) and for passenger transport (right), between 1995 and 2013, adapted from (European Commission 2011b)

Similar trends in rail passenger and freight demands are observed outside of Europe. The railway supply market is forecast to increase by 2.7% between 2013 and 2019, with the highest growth expected in Latin America, Asia and the Pacific (UNIFE 2015), making the optimisation of the network capacity usage a global challenge.

The following section presents the concept of railway operations and identifies the current limitations to be addressed to tackle the expected growth in traffic.

2.1.2 Railway operations

Railways differ from other modes of transport as the train wheels are guided on rails, implying that the train cannot change its course if there is any danger. Additionally, due to the adhesion of steel wheel on steel rail, braking cannot be immediate and relatively long distances are required for the train to stop, e.g. a train travelling at 160 km per hour needs approximately 1400 m to stop (Schon et al. 2013). A warning is, therefore, required to allow the train driver to stop before meeting any obstacle, and this is the main role of the control and signalling systems.

For a train to move safely, several conditions have to be fulfilled: the route ahead should be free, the points (i.e. moveable rails that enable trains to cross from one track to another) at junctions should be set correctly and the route should be protected against other train movements (Schon et al. 2013). These conditions cannot be checked by the train driver who has a very limited awareness of the traffic around him. Instead, a control centre, responsible for a portion of the railway network, ensures that these conditions are met for each train that travels through its regulated part of the network. The control centre aims to minimise the number of necessary stops while maximising the train business speed, i.e. “the

ratio between the distance that separates the two ends of a business trip and the average time needed to achieve it, taking into account all stops" (Schon et al. 2013).

In the control centre, the first task of the signalman is to determine the position of each train on the network and to preserve sufficient separation between trains to avoid a collision. For this purpose, the railway network is divided into sections of track, called blocks, with only one train allowed in a block at any one time. Signals are used to prevent entry by other trains into an occupied block. From the control centre, the signalman sets the train route by selecting the required blocks on the signalling layout display. In modern control rooms, a computer is used which displays the data and the network layout on a screen. By selecting the different blocks, signals and points along the route are set to allow for train movement, with points in their required positions and no conflicting routes set or in use (RSSB 2003).

Signals are used to visually give an instruction to the driver, in particular, the permission to proceed, also called movement authority. Traditional signals can display three aspects. Green means that the line is clear and that the train can proceed normally at the speed limit. Yellow indicates to the driver to start braking and prepare to stop at the next signal. Finally red, also called danger, indicates that the train should stop. The distance between the yellow and the red signal is computed on the basis of the braking distance and an additional safety margin. On railway lines with higher speed limits, two yellow signals can be used: a double yellow gives an early warning, followed by a single yellow, warning that the next signal is at danger.

While signals give the driver the permission to proceed, a system is needed to confirm to the signalman whether the track ahead is free or occupied. This is the role of the detection systems, installed on the track between a pair of consecutive signals. These systems detect whether a train enters or exits a block and send the information to the control centre, which displays it on the signalling layout display. Such systems are designed to be fail-safe and if a fault occurs, send an indication that the block is occupied and thus, prevent any subsequent train from entering. Certain detection systems are directly linked to the signals and change their aspect without requiring any action from the signalman. These automatic signals are commonly used on portions of lines with no junction. Others signals are controlled by the signalman while setting the route. Further details on detection systems are given in section 3.1.1.

The signalling system plays a key role in providing for the "safe routeing, spacing and control of trains" (RSSB, 2003, p10). To ensure the safety of operations and to prevent signals and points from being set in error, an interlocking system is used. The interlocking system controls the setting and releasing of signals and points to prevent conflicting movements and unsafe conditions. For instance, it makes it impossible for the signalman to turn a signal to proceed if the route ahead has not been proven safe. Interlocking aims to make railway operations clearer and more logical and ensures that movement authority is only given when it is safe to do so and points are set and locked correctly. It also guarantees movement authority can be withdrawn and route locking is only released when it is safe to do so (RSSB

2003). First implemented in 1843 using hard-wired circuitry (ALSTOM Signaling Inc. 2004), the interlocking function is now performed using digital logic implemented by software.

While interlocking ensures that no unsafe condition arises from the signalling and control systems, it neither prevents the driver from over-speeding nor from failing to follow the instructions displayed by the signals. To prevent these events, train warning and protection systems have been developed. Automatic warning systems (AWS) use visual and audio warnings to inform the driver of the aspect of the signal being approached. The device uses magnetic inductors installed on the track ahead of the signal. As the train travels over the device, a warning is generated in-cab. The driver is given two seconds to cancel the warning otherwise, the emergency brakes are applied. However, the driver can cancel the warning and still fail to stop. This is the reason why the train protection and warning system (TPWS) was developed. The TPWS automatically applies the brakes if the train passes the track mounted sensors at a speed higher than required.

Figure 2.3 shows the interactions between the different elements of the railway systems.

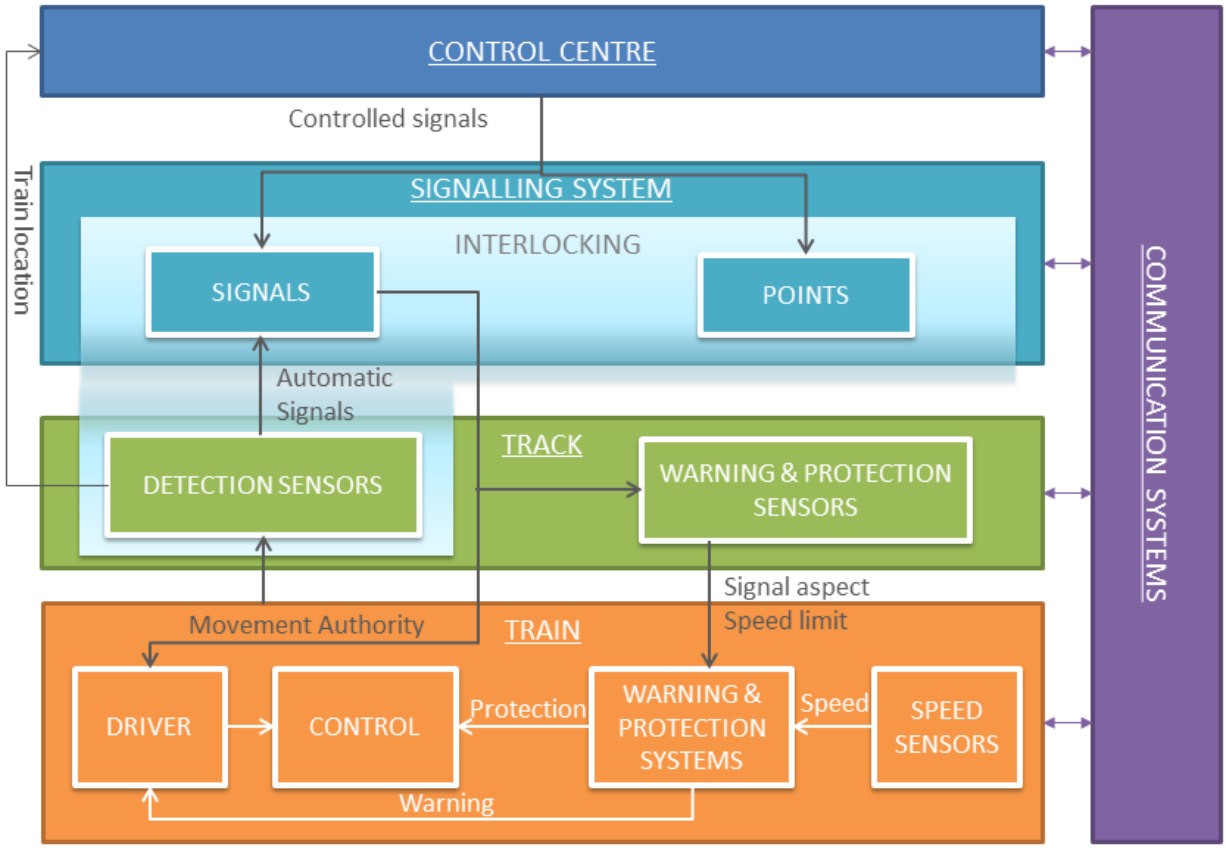


Figure 2.3 - Railway operational architecture

In summary, in the control centre, signallers and controllers monitor train position, set routes and resolve route conflicts by means of movement authorisations. Signalling and control systems use track equipment to provide the control centre with train location and display movement authorisations to train drivers through trackside and/or in-cab signals. Finally, the communication systems enable information exchange between the control centre, the signalling system, the track and the train.

Signalling and train control are vital functions for the safety of railways. They prevent collisions when trains are travelling on the same track and permit the safe movement of trains as they cross from one track to another. However, these functions currently have some serious limitations, as discussed below.

2.1.3 Limitations and challenges

The UK railway Technical Strategy Advisory Group (TSAG) has identified four main challenges, commonly referred to as the 4Cs, and that are seen as the main drivers of change. These challenges are to increase the network capacity and customer satisfaction, and to reduce the levels of carbon emissions and the cost of operation and maintenance while maintaining the safety of operations (TSLG 2012). To address these challenges, the current limitations of railway operations should be addressed.

Firstly, the capacity of the railway network, i.e. the number of trains that can be incorporated in a timetable without conflict and while complying with the regulatory requirements (TSLG 2012), is limited. This is due to the use of blocks that can be several kilometres long. Following the current regulation, and to ensure the safety of operations, only one train is allowed in a block, even if it might be safe for two trains to travel through this block at the same time (i.e. separation distance longer than the safe braking distance).

Secondly, the cost of operating and maintaining the current railway infrastructure is very high. The net operating cost of the railway network in the UK in 2015 was approximately £1.5 billion and the maintenance cost £1.3 billion, representing 40% of Network Rail (owner and infrastructure manager) expenditure (National Audit Office 2015). These figures exclude renewal and enhancement costs. These high costs are partly due to the extensive amount of trackside equipment used, such as signals and track sensors, and are reflected in the fare paid by customers. In addition to being costly, these systems also experience high failure rates resulting in disruptions and in turn, poor customer satisfaction.

In summary, the limited network capacity, high operating and maintenance costs, and high failure rates of the system are limitations that must be overcome for the railway to become competitive in terms of price, safety and quality of service. The following section shows how future railway systems address these limitations.

2.2 Future railway systems

2.2.1 Addressing the current limitations

Given the challenges outlined in the previous section, the first limitation to be addressed is the network capacity. The capacity can be increased by building new tracks or by lengthening trains and platforms. However, a more efficient use of the existing infrastructures can already offer extra train paths, at a lower cost and without much disruption (Glover 2013). An optimal use of the network will also generate the information required to justify the need for new infrastructure.

To optimise the existing capacity, train traffic management needs to be improved. This can be done by the implementation of intelligent traffic management systems (TSLG 2012) using real-time positioning of trains to enforce moving blocks. Unlike the static blocks discussed in the previous section, moving blocks are dynamic and follow the train motion. They include the train from rear to front, its braking distance and a safety margin. Figure 2.4 and Figure 2.5 illustrate the difference between the current signalling block principle and the novel moving block principle.

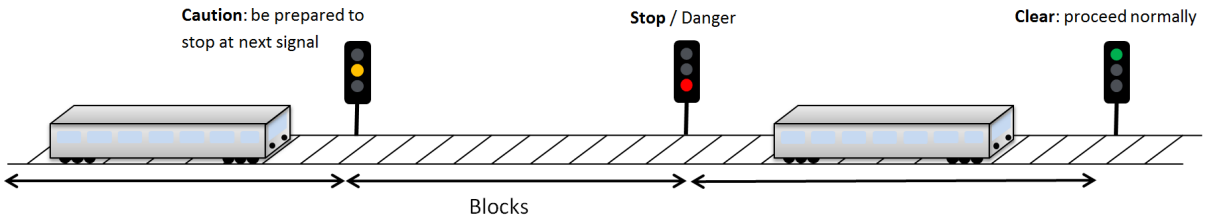


Figure 2.4 - Signalling block principle, adapted from (Kyriakidis 2013)

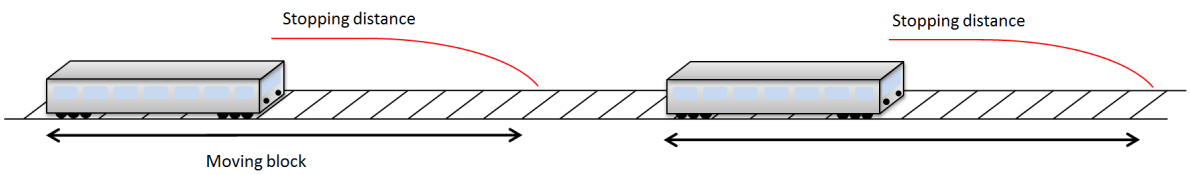


Figure 2.5 - Moving block principle

The train continuously monitors its position and reports it to the control centre. The signalman can then follow in real time the movement of each train on the network and deliver movement authorities to each of them directly via an in-cab display. Therefore, trackside equipment such as signals and detection sensors are not required.

Moving blocks will enable trains to operate closer to each other and reduce the number of times they have to stop, making a more efficient use of the railway network and thus increasing the capacity of each line. In addition, by moving the detection and signalling functions from the track to the train, the number of trackside equipment can be considerably reduced. The removal of such equipment will enable the cost of their operation and maintenance to be reduced. In summary, intelligent real-time traffic management systems have the potential to deliver higher capacity (TSLG 2012), at a reduced cost.

One intelligent real-time traffic management system is the European rail traffic management system (ERTMS). Initially developed for use in Europe, 38 countries around the world are currently running or contracted to be equipped with ERTMS (UNIFE 2012b), including China, Australia, India, Brazil and Russia. ERTMS architecture, operational principles and benefits are described in detail in the following section.

2.2.2 The example of ERTMS

The European Rail Traffic Management System, ERTMS, is a European initiative to create a single standard for train control, command, signalling and communication systems. It is designed to address the increasing demand for cross-border traffic within Europe by establishing a standardised traffic management system. In the UK, Network Rail plans to progressively equip its network with ERTMS over the next thirty years (Glover 2013).

ERTMS is composed of a suite of systems including control centres, onboard systems, trackside systems and communications. These systems are traditionally divided into four categories (RSSB 2010):

- The train control component, European Train Control System (ETCS). This provides the train with a movement authority (i.e. the maximum distance it can travel), the speed profile of the track and other track information.
- The communication system, Global System for Mobile Communications - Railway (GSM-R). This is a radio system which provides voice and data communication between the track and the train. It is based on the standard GSM using frequencies specifically reserved for rail applications.
- The European Traffic Management Layer (ETML), to optimise the use of the railway layout by reducing scheduling conflicts.
- A set of rules, the European Operational Rules (EOR), to standardise railway operations over Europe.

The ETCS supervises the train so that it does not go past the movement authority or exceed the maximum allowable speed. It is composed of two sub-systems: the onboard equipment and the trackside

equipment. These systems can be deployed in different configurations to provide several levels of operation as presented in Table 2.1.

Table 2.1 - ETCS levels of operations

ETCS level	Operations
Level 0	This level covers the operations of trains equipped with ETCS onboard equipment operated on tracks not fitted with the ETCS trackside equipment. It provides supervision of the trains' speed against the maximum permitted speed.
Level 1	Level 1 is based on track-to-train communications and is used in combination with the existing signalling system. The movement authority is transmitted to the train through a eurobalise (a transponder mounted on the track which can communicate with a train passing over it), while the train position is provided by on-track equipment, as shown in Figure 2.6.
Level 2	In level 2, a radio block centre is introduced at the traffic management centre (Figure 2.7). Electronic messages are sent to and received from the onboard train equipment via GSM-R. The continuous communications between the train and the radio block centre via GSM-R enable the transmission of the movement authority directly to the driver. The signal aspect is displayed in-cab and hence, the system can work with or without lineside signals.
Level 3	Level 3 provides a continuous update of the train's location and movement authority. It uses positioning data from the train rather than from the track equipment and relies on the moving block principle to enforce train separation. Therefore, signals and track detection systems are not required. The train location is continuously estimated onboard using sensors such as the odometer and information sent from the eurobalises. The train position is then sent to the control centre via GSM-R, as illustrated in Figure 2.8.

These different levels enable a smooth transition to the ERTMS, with the possibility of the gradual installation of systems rather than a single upgrade of the entire network.

ERTMS operates as follows (RSSB 2010). Similar to current control systems, the signaller sets the route and the interlocking function ensures that there is no conflict. A movement authority is then delivered to the onboard equipment. It contains the maximum distance the train is allowed to travel, the speed limit and the gradient of the line. Using this information, the onboard system computes a speed profile that the driver should not exceed and presents it on the in-cab display. If the driver attempts to

exceed the distance or the speed permitted, the ETCS intervenes to either slow down or stop the train. Information is provided to the onboard systems through balises and a GSM-R data radio link. The train also uses GSM-R to send information back to the trackside, such as its position and status.

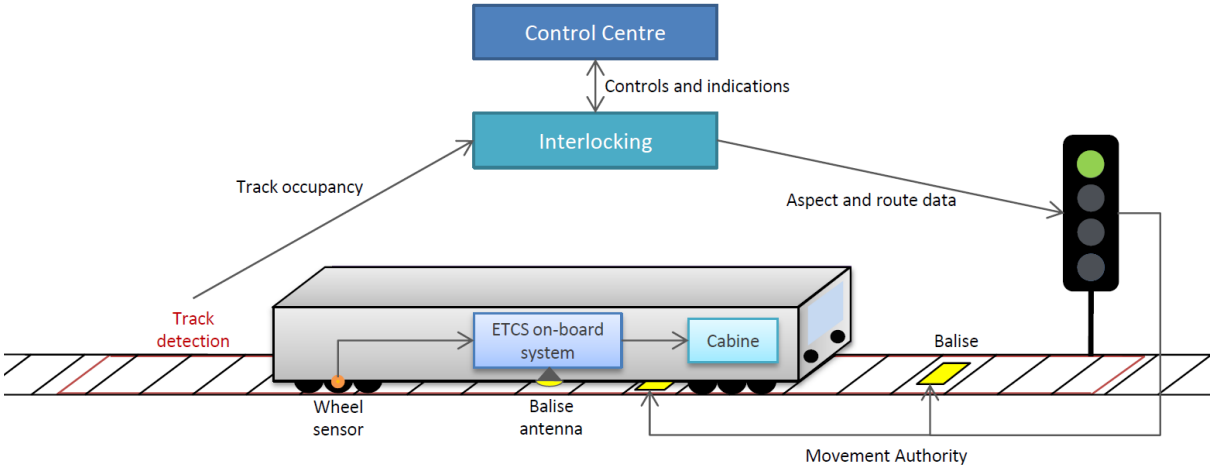


Figure 2.6- ETCS level 1 architecture, adapted from (Bloomfield 2006)

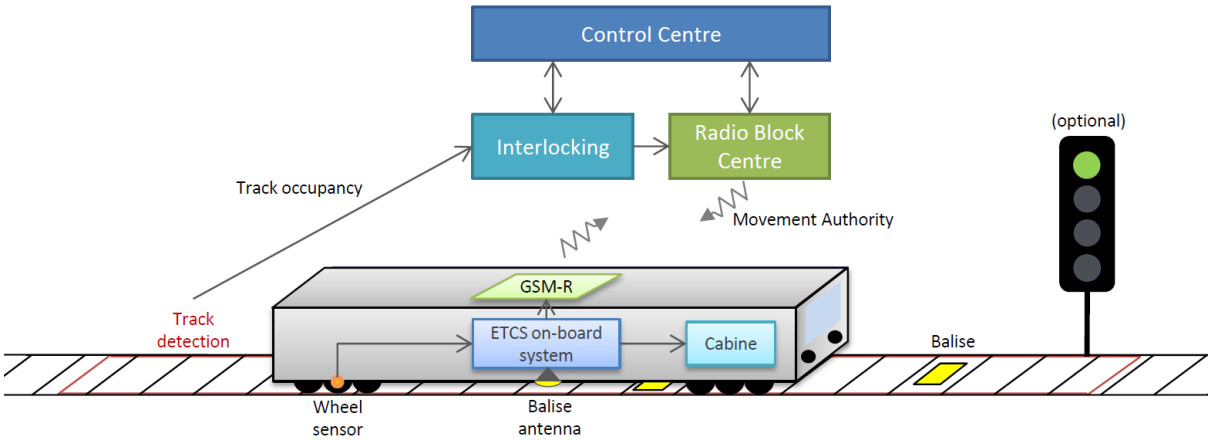


Figure 2.7 - ETCS level 2 architecture, adapted from(Bloomfield 2006)

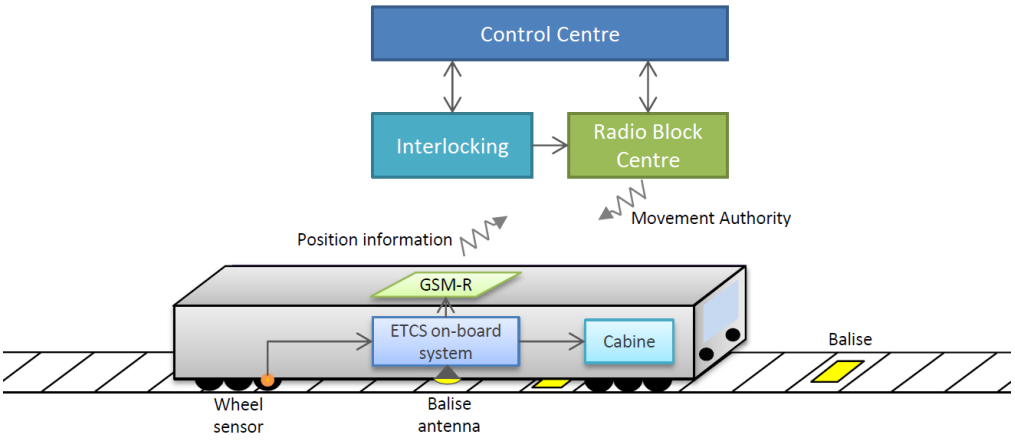


Figure 2.8 - ETCS level 3 architecture, adapted from (Bloomfield 2006)

The ERTMS operating principles present numerous benefits when compared to the current operations (European Commission 2016). In addition to increasing capacity and safety and reducing operation costs and malfunctions, ERTMS will enhance interoperability at the European level. By allowing traffic to flow seamlessly between borders, the movement of larger volumes of passenger and freight between countries will be enabled. Bringing the different national railway operators together will strengthen the market position of European railway globally and make the sector more competitive. Cross-border interoperability will, for instance, be the main enhancer to the development of the core European freight network Trans-European Transport Networks (TEN-T) envisioned by the European Commission (European Comission 2011), enabling railway transport to capture a larger portion of the in-land freight traffic.

2.2.3 Positioning within ERTMS

To address the railway traffic growth and to capture a larger share of transport demand, railway systems need to be upgraded. Flexible, intelligent traffic management systems combined with in-cab signalling have the potential to increase the capacity of the network while maintaining the safety of operations.

However, the implementation of such intelligent traffic management system will only be possible if the position of each train on the network is accurately determined. In fact, ERTMS level 3 requires that the train continuously estimates its position and velocity, and provides them to the control centre in order to optimise operations. Furthermore, the computation of the moving block length also requires the knowledge of the train position. It is, therefore, apparent that the determination of the train's position with sufficient accuracy, integrity, continuity and availability levels is critical for enhancing future railway traffic management systems. Furthermore, numerous other railway applications could benefit from the knowledge of the train position, as discussed in the following section.

2.3 The role of Positioning, Navigation and Timing in the railway domain

Despite trains being guided by rails and thus having very restricted dynamics, the knowledge of a train's position is critical to the safe and efficient operations of a railway. As explained in section 2.2.3, train control and management rely extensively on this knowledge and furthermore such positioning information can enhance applications in addition to train control and management.

2.3.1 Location-based railway applications and services

A study conducted by Rail Safety and Standards Board (RSSB) (RSSB 2012a), based on previous literature (GRAIL 2006a; APOLO 1999; GNSS Rail User Forum 2000), identified 54 different railway applications or services that rely, to a different extent, on the knowledge of the train's position.

The complete list of these applications is presented in Table 2.2, organised in five categories, reflecting the principal beneficiaries:

- signalling and control applications, which enable rail traffic to flow safely and efficiently
- operation applications, which manage the use of the route, the rolling stock and crews in the most efficient way
- customer applications, which provide information to passengers or freight users
- engineering applications, which benefit the engineering teams within the train companies
- infrastructure applications, which deliver services of interest to monitor and maintain the railway infrastructure

These applications were identified through consultations with different UK railway stakeholders such as Network Rail (the network manager), infrastructure managers, train operating companies and rolling stock companies. For a complete description of each application, see (Dumville & Hutchinson 2012).

Location-based railway applications range from the safety critical for signalling and control to the liability critical for cargo monitoring, and non-safety critical for passenger information systems. The safety critical applications are described as having the potential to directly or indirectly cause harm to humans, destroy the system, damage property external to the system or damage the environment (NASA 2004). Such a safety critical application is automatic train protection (ATP) that requires the monitoring of both the train's speed and distance to the end of the movement authority to prevent the driver from over speeding or passing signals at danger. For this purpose, the ATP function requires the train position and location of the end of movement authority.

In the case of moving blocks, the location of the end of movement authority depends on the location and speed of the preceding train. Thus, an estimate of each train's position on the network with sufficient accuracy, integrity, continuity and availability is required to ensure the safety of operations.

Liability critical applications are those that can generate legal or economic consequences. For example, the monitoring of cargo enables freight users to determine the time of arrival of goods at a depot. This time of arrival can then be used to plan subsequent services such as loading cargo on other vehicles. The knowledge of the cargo location can help optimise these tasks and thus limit the economic impact delays can create. In certain cases, when the goods carried by rail freight can cause a threat to the society or damage the environment (e.g. toxic or corrosive chemicals), cargo monitoring is considered safety-critical.

Table 2.2 – Location-based railway applications, adapted from (RSSB 2012b)				
Signalling and control applications	Operations applications	Customer applications	Engineering applications	Infrastructure applications
Automatic Train Protection (ATP) On-train ERTMS interface Train awakening Cold movement detector Train integrity and train length monitoring Location of GSM-R reports User worked crossings Secondary line signalling Trackside personnel protection Possessions management Track circuit diversity during leaf fall Alternative temporary block working Diverse positioning systems Detonator replacement Tilting trains	Centralised clock Traffic management & regulation Eco-driving Driver Advisory System (DAS) Fast and slow lines discrimination Automatic train operation Driverless train Temporary speed restrictions & emergency speed restrictions Fleet management Cargo monitoring Terminal management Door operations Passenger count Pantograph control Driver route knowledge assistant On-train ticketing, retail and authentication On-train reservations On-train catering & services Train crew information services Infrastructure charges Delay attribution Incident management response Logistics planning and monitoring	Customer information systems Passenger information systems Personal journey assistant Location based services & points of interest Passenger broadband Train approaching warning	On-train monitoring recorder On-train closed-circuit television (CCTV) On-train monitoring On-train automation On-train braking measurement probes Odometer calibration	Digital route map creation Structural monitoring Gauging surveys Automated infrastructure maintenance

Finally, non-safety critical applications are those that do not present any economic, legal or safety implications, such as passenger information systems. These systems include indicator boards at the station but also on-train display and pre-trip information. The accurate, safe, continuous and available knowledge of the train position can provide more up-to-date and reliable information than at present, thus providing improved support to customers travel plans.

Not only will the enhancement of the current railway positioning function support existing location-based applications, it will also encourage the development of new applications. Figure 2.9 shows existing location-based railway services as well as concepts that can be supported by the knowledge of the train position.

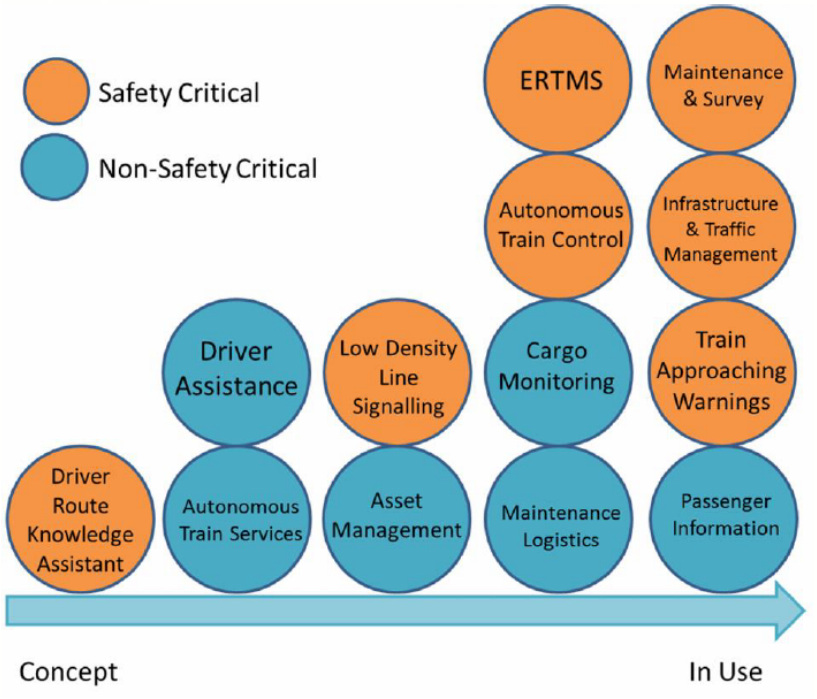


Figure 2.9 - Maturity of different location-based railway services, from (Marais 2014)

For instance, it will support the implementation of automatic train operation (ATO), a well-established function in urban transport. Used by London Underground on the Central, Northern, Jubilee and Victoria lines, this system is still to be used on any European main railway lines. ATO systems in mass transit include several elements such as an ATP system, a built-on ATO package for automatic driving, a traffic management system and a robust telecommunication network. Most ATO systems implement Communication-Based Train Control (CBTC) which continuously estimate the positions of the different trains and communicate their status to the trackside equipment (Pascoe & Eichorn 2009). Sensors such as odometers, accelerometers or GPS receivers are used to estimate the train position. The implementation of ATO on mainlines is more complex than for mass transits due to the mixed railway traffic (trains with different speeds), the complexity of the networks (numerous tracks and junctions)

and to the different types of trains. However, a Memorandum of Understanding (MoU) was signed in 2012 for ATO to be included in the ERTMS documentation (European Commission et al. 2012). As defined by (RSSB 2012b), while using the ATO, the driver will still be present in the cab to monitor the system and mitigate risks in case of an emergency. However, if the system is proved to be reliable enough, it could support the implementation of fully driverless trains.

Many rail applications can benefit from the knowledge of the train's position. However, in order to adequately design the positioning function and to ensure the safety of operations, the positioning requirements of each application need to be understood and defined.

2.3.2 Positioning requirements

Requirements are necessary conditions the positioning system and application should fulfil. At the application level, they describe operational needs such as determining the track the train is on, its mileage and its direction of travel. These operational requirements are then used to compute the positioning requirements, describing the level of performance the positioning system should achieve in order to support the application. No standard method is currently available to translate the operational requirements into positioning requirements. However, the ongoing Satellite Positioning Performance Assessment for Road Transport (SaPPART) project (SaPPART 2015) attempts to define such a method for intelligent traffic systems (ITS) and mobility services.

To illustrate further the difference between operational and positioning requirements, the example of virtual balise generation is taken. Virtual balises aim to replace their track-mounted counterparts which are both expensive and required in a large number to support ERTMS (on average two balises per kilometre (Gurnik & Trégl 2005)).Virtual balises use the train position together with a database containing the location of the virtual balises. When the train position corresponds to a location stored in the database, a message, similar to the one sent by a track-mounted balise, is sent to the onboard systems.

The virtual balise operational requirements can be defined in terms of missed detection, e.g. the train travels 'above' a virtual balise but no message is sent to the onboard systems, or false detection, e.g. the train is detected on the wrong track. These probabilities of missed and false detections can then be used to determine the required level of accuracy, integrity, continuity and availability the positioning system should provide. A positioning system that can achieve the positioning requirements is then specified. During operations, this system provides the virtual balise application with the train position, velocity, time and quality indicators, as illustrated in Figure 2.10. The application then uses these values to determine operational quantities, such as the position of the train on the track. Thus, requirements are defined from the application level to the positioning level, in contrast to operations.

In this thesis, the requirements are defined for the positioning function. Originally defined by the International Civil Aviation Organization (ICAO) to support civil aviation operations, the navigation requirements in terms of accuracy, integrity, continuity and availability are commonly used for any positioning application (ICAO 2006).

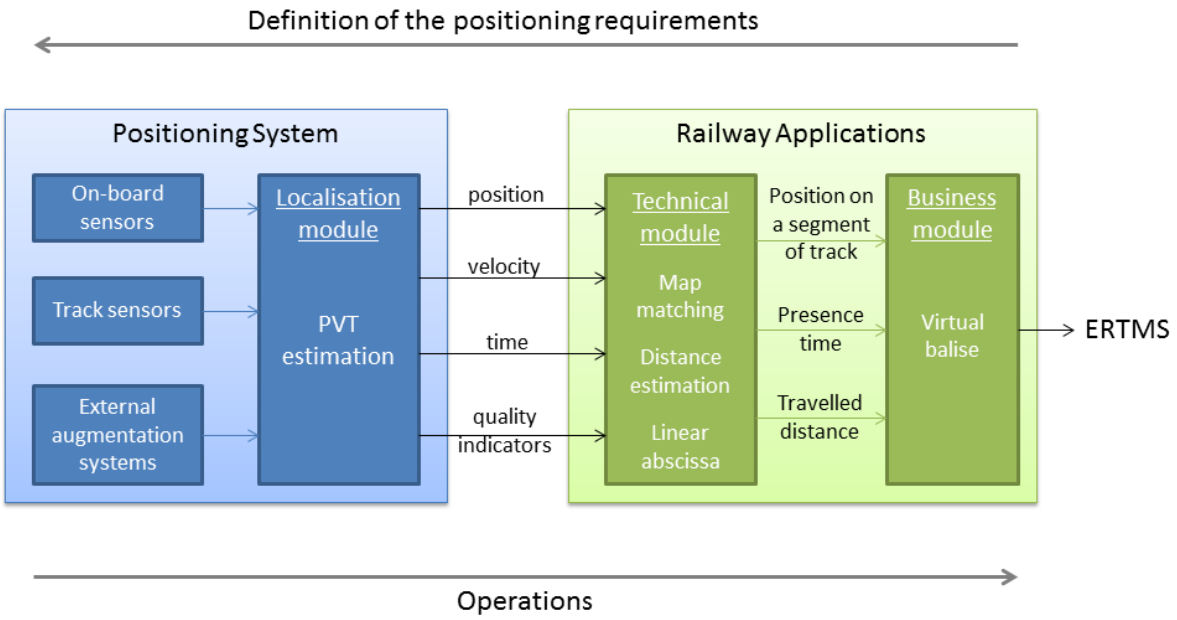


Figure 2.10 – Architecture of virtual balises generation to support ERTMS, adapted from (SaPPART 2015)

Accuracy measures the quality of the position. The position error is the difference between the position estimated by the system and the actual position. For an estimated position at a specific location, the probability that this error is within the accuracy requirement should be at least 95 percent (ICAO 2006). If required by the railway operational needs, a different percentage could be selected to define the accuracy of the train position.

Integrity is a measure of the trust that can be placed in the correctness of the information supplied by the total system, in this case, the navigation system. It includes the ability of the system to provide a timely and valid warning to the user when the system must not be used for the intended operation. Integrity quantifies the concept of safety (ICAO 2006) and integrity requirements are defined using three parameters:

- The alert limit which represents the largest position error that may occur without being detected.
- The integrity risk which represents the probability that an error exceeds the alert limit without the user being informed within the time to alert.

• The time to alert which is the maximum allowable elapsed time from the onset of a positioning failure until the equipment announces the alert.

The continuity of service of a system is the ability of the total system to perform its function without unscheduled interruption during the intended operation (ICAO 2006). More specifically, it is the probability that performance will be maintained (i.e. navigation solution within the alert limits) for the duration of an operation assuming the system was available at the beginning of that operation and was predicted to operate throughout. Continuity risk is the probability that a system that is available at the start of an operation, is interrupted during the intended period of operation.

The availability of a navigation system is an indication of the ability of the system to provide usable service within the specified coverage area. It is defined by the percentage of time the system is to be used for navigation, during which reliable navigation information is presented to the user (ICAO 2006). This means that the accuracy requirements are met, integrity alert limits are protected and continuity provided.

These definitions are commonly used to define the navigation performance required to support any positioning application.

2.3.3 Existing railway positioning requirements

Several studies have attempted to define the required navigation performance for railway operations (GRAIL 2006a; APOLO 1999; GNSS Rail User Forum 2000; GAARDIAN 2009; RSSB 2012b). According to these studies, unlike aviation, railway operations cannot be divided into phases. Instead, requirements are defined for each of the applications identified in the previous section.

It is important to note that none of these requirements is currently recognised as standard and that they only provide recommended values. The most complete set of requirements is presented by the RSSB (RSSB 2012b) and on the basis of discussions with the different railway stakeholders (infrastructure operator and managers, train operating companies and rolling stock companies). The definition of these requirements is based on the characterization of each location-based railway application in terms of its purpose, operational principles and reliance on positioning information. A subset of these requirements in presented in Table 2.3.

RSSB (2012) defines the traditional requirements of accuracy, integrity, availability and continuity (except for the missing integrity risk), but in addition, introduces the following:

- the coverage, which is the geographical area where the application operates;
- the mission time, which is the expected duration of the operation;
- the time accuracy, which is computed on the basis of the position accuracy and the distance travelled;

- the speed accuracy, which is similar to the position accuracy and is expressed at 95%;
- the direction accuracy which corresponds to the heading accuracy and is also expressed at 95%;
- the tolerable start-up time which defines the time required to initiate the application (similar to the time to first fix TTFF);
- the number of tolerable failures which is similar to the mean time between failures (MTBF).

Table 2.3 – Railway application positioning requirements, adapted from (RSSB 2012b)

Application	Coverage	Time accuracy (s)	Position accuracy (m)	Speed accuracy (kph)	Direction accuracy (deg)	Integrity Alert Limit (m)	Integrity TTA (s)	Mission time (s)	Continuity (%)	Availability (%)	Tolerable start-up time (s)	Number of tolerable outages per day
Signalling and control applications												
Automatic Train Protection	known routes	0.01	2	2	5.0	5	2	3600	94	99	1	1
On-train ERTMS interface	known routes	0.01	2	2	5.0	5	2	3600	94	99	1	1
Train integrity	known routes	0.01	2	5	5.0	5	10	3600	94	99	5	1
Operation applications												
Traffic Management & Regulation	sub-network	0.1	17	5	22.5	42	25	3600	68	99	30	6
Driverless Train	known routes	0.01	2	2	5.0	4	10	3600	88	99	5	2
Door operations	known stations	0.1	1	2	5.0	3	25	120	99	99	5	6
Customer applications												
Customer information system	GB network	0.1	33	5	22.5	83	25	3600	68	99	30	6
Passenger broadband	GB network	0.1	6	5	5.0	14	25	3600	28	99	5	20
Engineering applications												
On-train monitoring recorder	GB network	0.1	14	5	5.0	35	25	3600	28	99	5	20
On-train automation	GB network	0.1	6	5	5.0	14	25	3600	68	99	5	6
Infrastructure applications												
Digital route map creation	GB network	1	0.3	na	1.0	0.8					1	
Structural monitoring	known routes	1	0.02	na	na	0.1					1	

A closer examination of the requirements in Table 2.3 shows that infrastructure related applications (digital route map creation, structural monitoring and automated infrastructure maintenance) present the most stringent requirements in terms of the accuracy and integrity alert limit. However, no time to alert and continuity or availability risk are stated. Among the other applications, automatic train protection is one of the most stringent in terms of accuracy, integrity, availability and continuity. A positioning accuracy of 2 m is required with a time to alert of 2 s. Compared with the 16 m horizontal accuracy and 6 sec time to alert required in aviation for a category I approach (ICAO, 2006), these values will be challenging to achieve. However, the continuity risk remains lower than that defined for a category I approach ($1\text{-}1x10^{-6}$ in any 15 sec) due to the fact that trains can be stopped in the case of a failure of the system.

Infrastructure related applications may require a dedicated positioning system delivering a high level of accuracy but need to be time-independent, as no time to alert or continuity risk are requested. Regarding the other requirements, a positioning system able to fulfil the automatic train protection requirements can also be used by a wide range of other applications.

Most studies offer comparable values for these requirements. A study conducted by (Filip & Rispoli 2014a) also defines the value of the integrity risk at 10^{-9} per hour. This value comes from the tolerable hazard rate of 10^{-9} dangerous undetected failures per hour with which new train control systems have to comply. Even if this risk is derived from a solid theoretical basis, such a stringent requirement has limited practical use as it will be difficult to achieve in real-life scenarios. In comparison, the current integrity risk defined in aviation for all phases of flight (excluding the approach) is 10^{-7} per hour, while the environment in which aircraft navigate in makes positioning easier (no obstacles). However, it might be possible to redistribute the integrity risk across the components of the train control system, reducing the part allocated to the positioning function. The distributions of the railway risk between the different functions, including the positioning function, is currently being investigated by the satellite positioning working group with the Next Generation Train Control (NGTC) project (NGTC 2016). The project findings should be available by the end of 2016.

The aim of this thesis is to develop a positioning system that can support the widest range of applications. For that purpose, the most stringent requirements found in the literature are selected and summarised in Table 2.4 and Table 2.5. Infrastructure application requirements are however excluded as their needs are mainly in terms of accuracy (see Table 2.3), and relate more to surveying applications than to navigation.

Table 2.4 – Selected railway positioning requirements

Coverage	Position accuracy	Integrity risk	Integrity alert limit	Integrity TTA	Mission Time	Continuity	Availability
known routes	1.0 m	10^{-9}/h	3.0 m	2 s	3600 s	10^{-6}	99.99%

Table 2.5 – Additional positioning requirements defined in (RSSB 2012b)

Time accuracy (s)	Speed accuracy (m/s)	Direction accuracy (deg)	Tolerable start-up time	Number of tolerable outages per day
0.01	2	1.0	1	1

Railway requirements are particularly stringent and require the use of precise positioning techniques and integrated positioning solutions to ensure the required levels of accuracy, integrity, continuity and availability are achieved.

2.4 Conclusions

The growing demand for transportation together with the need to capture a larger share of this demand in Europe requires railways to evolve and implement intelligent traffic management systems to increase their capacity and efficiency. Such systems require the knowledge of each train's position on the network with the required levels of accuracy, integrity, continuity and availability, and cannot be delivered by current railway systems, as seen in 2.1.3.

While supporting traffic management, a dedicated train positioning platform has the potential to support the large range of additional location-based services presented in Table 2.2. To prevent the development of ad hoc solutions for each location-based service, a holistic approach must be followed in designing the novel positioning platform. While at present, no positioning standards have been produced, several studies identified the numerous railway applications that can benefit from a dedicated positioning function as well as their associated requirements. Based on these studies, a set of requirements reflecting the most stringent values is selected with the purpose of developing a system that can address the widest range of railway location-based applications. The following chapter reviews the existing positioning technologies and solutions used both in the railway and in other domains and assesses their ability to address the selected requirements.

Chapter 3

Current and Future Positioning Systems to Support Railway Operations

This chapter reviews the existing railway systems used to determine the position of the train and compares their performance to the requirements defined in chapter 2. Alternative positioning sensors, including terrestrial and satellite-based sensors, are then presented along with their performance. The existing positioning systems tested for railway applications are subsequently reviewed. Finally, sensors are selected based on the railway positioning requirements and previously reviewed sensors and existing solutions. The main contribution of this chapter is the definition of a possible architecture for a railway dedicated positioning system.

3.1 Positioning systems and their performance

This section reviews the positioning technologies currently relevant to the railway industry including terrestrial and satellite-based systems. The review covers the principles of operation, performance and limitations.

3.1.1 Railway positioning systems

This first subsection reviews the positioning systems and sensors used in the railway industry. Currently, train position is provided by track-mounted equipment such as track circuits and axle counters which detect the presence of a train within a block and set the signal accordingly (e.g. signal set at danger if a train is detected). The velocity of the train is estimated using either an odometer or a Doppler radar onboard.

3.1.1.1 Track circuits

The track circuit, invented in 1872, detects the presence of a train electrically by the passage of a weak current through a section of two running rails, insulated from adjoining sections (Glover 2013). As seen in Figure 3.1, it consists of a power supply on one side of the block, connected to both rails, and of a track relay connected to both rails on the other side of the block.

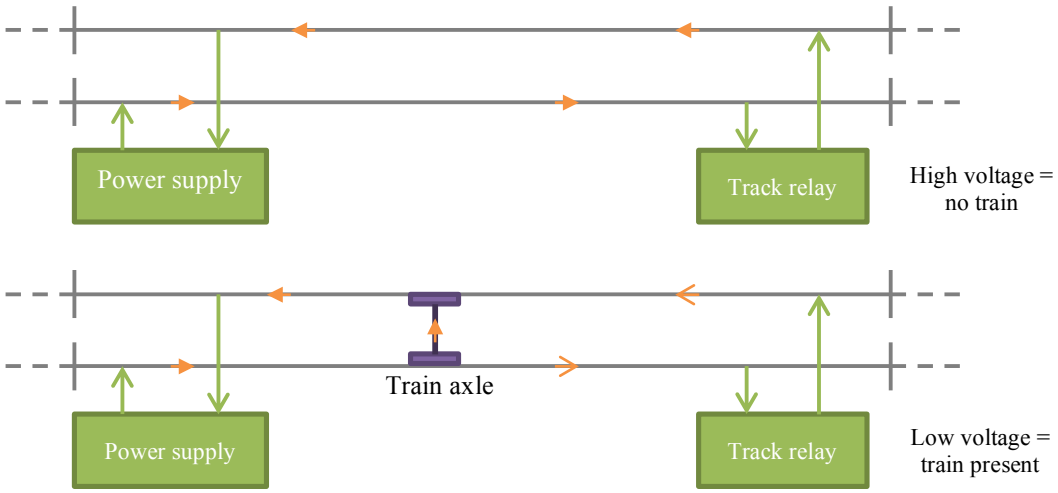


Figure 3.1 – Track circuit principle, the orange arrows represent the current intensity, adapted from (Schon et al. 2013)

The power supply sends a current through one of the rails, which circulates to the track relay and back to the power supply through the second rail. If a train enters the section, its metallic axles create a short circuit, meaning that only a very small portion of the current reaches the track relay. Thus, by using a voltage detection threshold at the track relay, the presence of a train in the section is detected, providing an indication of the train location as explained in section 2.1.2 (Schon et al. 2013).

Track circuits present several advantages. They are, by design, fail-safe. Any failure or break in the circuit leads to a low voltage at the track relay, indicating that the block is occupied and preventing the entry of any train to the block. They can also detect the absence of a train as well as its presence, providing permanent detection. Broken rails can also be detected, as a cracked rail does not conduct current as efficiently. Finally, the current sent by the power supply can be modulated and used for track-to-train transmission (Schon et al. 2013).

However, track circuits are sensitive to the rail-wheel contact. The presence of rust or leaves on the rail reduces its conductivity. This partly explains why these systems are prone to failures, especially in wet areas (Hall 2005), leading to the development of axle counters to compensate for the weakness.

3.1.1.2 Axle counter

Axle counters simply count the number of axles that enters a section of track and deduct the number that leaves. If the result of this operation is zero, the block is free. Junction boxes are used to store the counts at each end of the section. The counts are then sent to a central unit which performs the calculation and sends the result to the interlocking system, as shown in Figure 3.2.

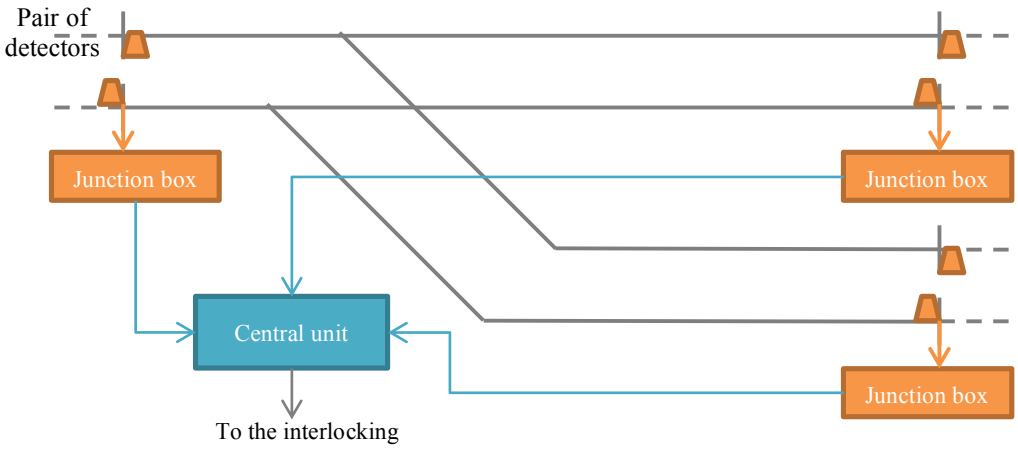


Figure 3.2 – Axle counter principle, adapted from (Schon et al. 2013)

Axle counters are not limited in length and can be installed on several tens of kilometres of track, as opposed to track circuits which can only reach 1500 to 2000 meters (Schon et al. 2013). Furthermore, they are insensitive to the rail-wheel contact and can be used in wet areas or with metallic sleepers (i.e. elements that support the rails). They are also less expensive than track circuits.

However, they also present several drawbacks. Firstly, problems can arise at low speeds and if the train axle stops directly above the detectors. Also, they are reset to zero after a fault such as a power failure (Schon et al. 2013). This implies that after such a fault, a visual inspection is required to ensure that the block is free. Finally, unlike track circuits, they do not detect broken rails and cannot be used for track-to-train transmission.

Given their strengths and weaknesses, track circuits and axle counters are usually used in combination. Track circuits are still used as the main means of detection but are replaced by axle counters in wet areas such as tunnels.

Both track circuits and axle counters offer a coarse estimation of the train location by indicating its presence in a block. However, they do not offer the possibility to continuously monitor the train movement, as planned in the ETCS level 3. For this purpose, balises have been introduced.

3.1.1.3 Balise

As explained in section 2.2.2, the future ERTMS plans to phase out track circuits and axle counters and to instead (or in complement with) use balises. A balise is a transponder mounted on the track, which is energised by the train passing above it. Once energised, it transmits an electronic message (telegram) to the train containing data such as the balise location. A group of balises can be used to help the train

determine its direction of travel. The train must be equipped with a balise reader which can energise the balise and receive the telegram (RSSB 2010).

As described in (UNISIG 2012), the balise is required to deliver its location with an accuracy that is a function of the train speed. Below 40 km per hour, the location accuracy should be within ±0.20 m, with a confidence interval of 0.998. For speeds of between 40 to 500 km per hour, the accuracy should be ± (1.1 x10^{-3}.velocity + 0.15) m, as illustrated in Figure 3.3.

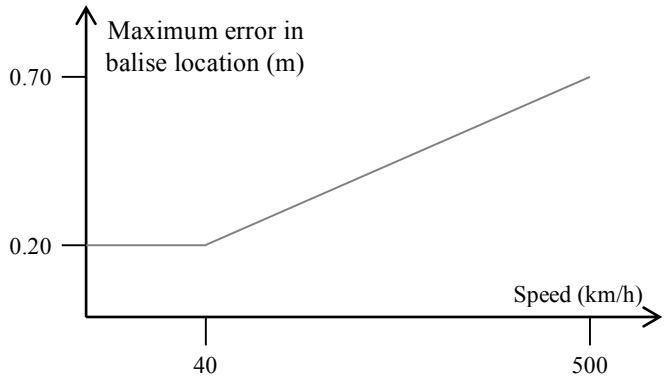


Figure 3.3 – Error in position of the balise centre, adapted from (UNISIG 2012)

Balises provide positioning information when a train travels above it. This means their operational requirements have to be stringent since should the information provided be either incorrect or missing, the system has to wait until the next balise to get a new fixed location. Balises are required to operate with a tolerable hazard rate of 10^{-9} per hour. However, with a continuous positioning system, it might be possible to relax this requirement due to the increased positioning availability.

Balises are one of the main elements of the ETCS architecture. They are relatively expensive to install and the fact that low-density lines with little revenue may be unable to afford them has initiated the search for alternative solutions (Rousseau et al. 2004; Barbu & Marais 2014). However, it is worth noting that balises are not exclusively used for positioning as they can also send additional information to the train such as the track description (e.g. track gradient, maximum allowable speed). Should an alternative system be used to determine the train position, this additional information will still need to be provided, for example by using a communication link.

3.1.1.4 Odometer

An odometer is a sensor installed on a train wheel that measures the number of full rotations performed by this wheel as well as their frequency (i.e. the number of rotations per unit of time). Using both this information and the known radius of the wheel, the distance travelled and the speed of the train can be estimated.

However, odometers have a limited accuracy as sliding and skidding of the wheels can cause an over-reading or under-reading of the distance travelled. Moreover, measurements deteriorate with the wear of the wheels. Odometers used in the railway industry are required to provide an accuracy (95%) of ± 5 m ± 5% of the distance travelled from the last reference point (GRAIL 2008), for instance since the last balise. Odometers are also required to meet a speed accuracy requirement of ±2 km per hour for speed lower than 30 km per hour, that then increases linearly up to +/- 12 km per hour at 500 km per hour (ERTMS Users Group 1998). These requirements are defined so that the speed error is bound, ensuring the correct estimation of the train velocity. Finally, several odometers can be used to provide redundancy and ensure the integrity of the measurements.

3.1.1.5 Doppler radar

Doppler radar is an alternative sensor to determine the train speed. It is mounted under the train to provide independent speed and distance measurements. It uses pulses of electromagnetic waves transmitted by a directional antenna. When these waves meet an object, some of the pulses are reflected back to the sensor which can then determine a range. The velocity of an object can be estimated by applying the Doppler principle (APOLO 1999).

Nevertheless, such devices are expensive and the signal is disturbed by bad weather conditions, for example by snow. Doppler radars are required to have a performance equivalent to that of the odometer, i.e. an accuracy (95%) of ± 5 m ± 5% of the distance travelled since the last reference point.

3.1.1.6 Eddy current sensors

A speed estimation technique based on Eddy current sensors is currently under development, based on research by (Geistler 2002). This technique uses two sensors mounted on the train and separated by a known distance. As these sensors travel above the same elements of the track (e.g. sleeper or point), they generate an electromagnetic wave and record the reflected wave. As they travel above similar elements, the two sensors record similar signals, with a time offset due to the separation distance. By estimating this time offset and knowing the distance between the two sensors, the train speed can be estimated. Additionally, these sensors can be used to detect track elements such as junctions, which create a recognisable signature on the recorded signal. This information can then be used in combination with a topographic database of railway junctions to perform map matching.

However, these sensors are still under development and are yet to be commercially viable for railway applications. Compared to odometers, eddy current sensors offer the advantage of using non-contact measurement methods and are not affected by the wheel behaviour (i.e. in the case of sliding of skidding). They also come at a lower cost than Doppler radars. However, more testing is required to assess their positioning performance.

In summary, several sensors and systems, both track-mounted and onboard the train, are currently used to estimate train position. However, most systems were not originally designed for positioning and thus their capabilities are limited (see

Table 3.1). Track circuits and axle counters only provide a coarse estimation of the train position by indicating its presence in a block that can be several kilometres long. Balises provide accurate (sub-metre level) but discontinuous positioning information. And while odometers and Doppler radars are able to continuously estimate the train velocity and position along the track, their current performance, however, cannot fulfil the accuracy requirement of 1 m identified in section 2.3.3.

Table 3.1 - Performance of railway positioning systems

Technology	Performance
Track circuits and axle counters	Depend on the length of the block, up to several tens of metres
Balise	Only provide a position when the train is above the balise (discrete positioning system)
Odometer, Doppler radar	Provide an accuracy of +/-5m +/- 5% of the distance travelled from the last reference point (i.e. balise)

To achieve the positioning requirements in Table 2.4, research has been directed towards alternative positioning methods, including the satellite-based and terrestrial systems discussed below.

3.1.2 Satellite-based positioning systems

3.1.2.1 Global navigation satellite systems

Satellite-based navigation systems offer an attractive alternative solution to the current railway positioning systems reviewed in section 3.1.1, as they provide autonomous and continuous (in some environments) positioning service with a global coverage. Furthermore, they offer a solution for the railway system that is independent of track-side equipment and interoperable across borders.

The satellite-based positioning market has witnessed rapid growth in the last decade mainly due to the reduction in both size and cost of the receivers. Satellite positioning makes use of the global navigation satellite systems (GNSS). Currently, only the United Stated GPS and the Russian GLONASS are fully operational. The European Union is developing its own system, Galileo, that is planned to be completed by 2020 (Blanchard 2012) and the Chinese system, Beidou, should also be fully operational by 2020 (Ding & Yang 2012).

A GNSS receiver is able to estimate its position autonomously, in a stand-alone mode. However, its estimation is affected by numerous error sources such as the receiver and satellite clock biases, the

ionospheric and tropospheric delays (due to the fact that the signal is travelling through the atmosphere), multipath due to the receiver surrounding environment and receiver noise (Kaplan & Hegarty 2006). The expected accuracy of GPS standalone positioning is 9 m horizontally and 15 m vertically (95%) (Scot 2014) using pseudorange measurements. Pseudorange measurements are one of the two types of observables provided by GNSS. The pseudoranges of each satellite are obtained by multiplying the time taken for each signal to reach the receiver by the speed of light. While this measurement is robust, different error sources limit its accuracy to about one metre.

However, signals are transmitted by satellites on several frequencies. For example, GPS currently uses three frequencies named L1 (1575.42 MHz), L2 (1227.60 MHz) and L5 (1176.45 MHz). By using measurements on several frequencies, it is possible to mitigate some errors, such as the ionospheric delay and to increase the accuracy to 6 m horizontally and 10 m vertically (95%) (Scot 2014). This improvement may, however, be insufficient for many applications. Additional techniques, known as GNSS augmentations, have been developed to enhance the accuracy, integrity, continuity and availability of GNSS positioning.

3.1.2.2 GNSS augmentation systems

GNSS augmentation systems use differential GNSS (DGNSS) and supplement the receiver with additional ranging signals, differential corrections and integrity alerts (Groves, 2013). DGNSS differences measurements to remove spatially and temporally correlated errors such as atmospheric errors, satellite orbit error and satellite clock offset. The use of additional receivers at known fixed locations enables corrections to be computed and sent to the user via a data link (Kaplan & Hegarty 2006). Augmentation systems also monitor the integrity of the signals sent by the satellites and deliver an integrity message to the user on whether or not a satellite should be used. Depending on their coverage, DGNSS can be classified into local area and wide area DGNSS.

Two types of GNSS augmentation systems exist, the Satellite Based Augmentation Systems (SBAS) and the Ground Based Augmentation Systems (GBAS). Both systems are certified (or in the process of being certified) for use in guiding aircraft during certain flight phases, including approach for SBAS and approach, landing, departure and surface operations for GBAS (USA Department of Defense 2008b).

SBAS apply the wide area GNSS concept and offer services over a country or small continent. Existing SBAS includes the American WAAS, covering North and a part of Central America and EGNOS covering Europe. These systems use a network of reference stations to compute corrections that are then broadcast to the user via geostationary satellites (Kaplan & Hegarty 2006). The purpose of SBAS is to increase the accuracy, integrity and availability of GNSS. Whilst they were primarily designed to

support aviation, railway operations can also benefit from EGNOS services (European GNSS Agency 2012).

GBAS applies the local area DGNSS concept and offers service over a small area such as an airport, as shown in Figure 3.4. It broadcasts correction information to users via ground stations and provides a higher accuracy than SBAS. GBAS has been approved by the Federal Aviation Administration (FAA) for CAT I precision approach operations, which means it can reach an accuracy of 16 m horizontally while ensuring an integrity risk of 1-2x10^{-7} per approach and a time to alert of 6 seconds. It is however expected to reach an accuracy of 3.6 m horizontally with an integrity risk of 1-1x10^{-9} per 15 sec and a time to alert of 1 to 2 seconds (Eurocontrol 2013).

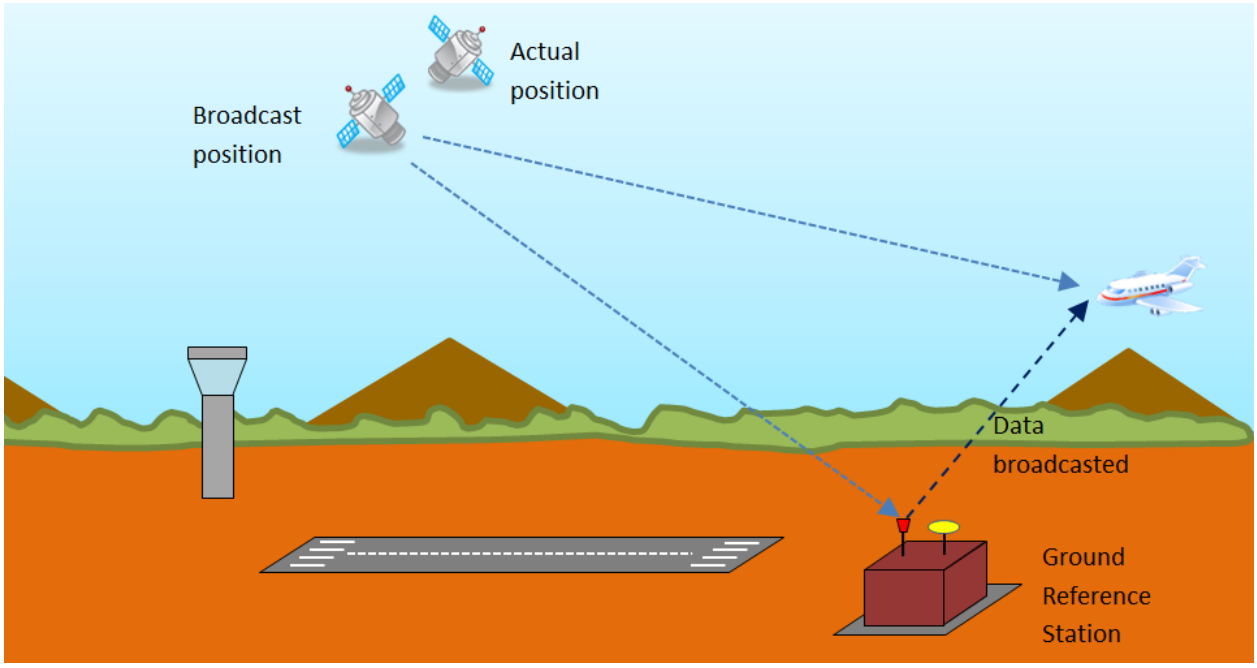


Figure 3.4 - GBAS principle, adapted from (Julien 2011)

The idea of an augmentation network dedicated to railway operations, known as EGNOS-R, has been investigated by Filip & Rispoli (2014b). Such a system would use the existing EGNOS service as well as additional reference stations installed along the railway tracks in order to provide corrections to the onboard GNSS unit and attempt to reach the tolerable hazard risk of 10^{-9} per hour. The need for additional trackside installations will, however, need to be precisely quantified as one objective of the future railway systems discussed in section 2.2.1 is to reduce the amount of trackside equipment.

Despite ensuring a certain level of integrity and mitigating some error sources, the use of GNSS combined with SBAS and GBAS cannot achieve the one-metre accuracy required to support railway operations (see Table 2.4) and higher accuracy positioning techniques have to be used.

3.1.2.3 High accuracy GNSS-based positioning

High accuracy positioning with GNSS is achieved with carrier phase measurements, the second GNSS observable. The carrier phase corresponds to the accumulated phase offset between the receiver and the satellite signal. Whilst it can provide centimetre level accuracy, the initial number of integer cycles in the carrier phase measurement, called integer ambiguities, has to be determined. Carrier phase measurement is more precise than the pseudorange but is ambiguous and less robust. Indeed in high dynamics applications, cycle slips occur which should be detected and recovered to maintain the correct ambiguity solution.

There are two main positioning methods using carrier phase measurements: Real Time Kinematic (RTK) and Precise Point Positioning (PPP).

RTK requires at least two receivers to estimate the differenced carrier phase ambiguity. One receiver, the base station, is located at an accurately surveyed location. Another receiver in the vicinity of the base station (10 to 20 km), the rover, can then receive information broadcast by the base (code and carrier measurements) via a real-time communication channel. The proximity of the receivers enables a number of spatially and temporally correlated errors to be either cancelled or mitigated by differencing the base and rover measurements. A convergence time of a few seconds is then required to fix the integer ambiguities. Once the ambiguities have been correctly fixed, RTK can achieve accuracy performance in the range of a few centimetres (Sanz Subirana et al. 2013).

Precise Point Positioning is performed using a single receiver with a single or dual-frequency. Satellite clock, orbit and bias corrections are provided via a data link. Single frequency PPP uses ionospheric maps to compute the ionospheric delay correction. Decimetre level accuracy can be achieved, though this requires a long convergence time. Dual-frequency PPP uses a combination of measurements on both frequencies to cancel the first order ionospheric delay. Centimetre level accuracy can then be achieved in real time. However, a long initialization time (about 20 minutes) is required each time the system starts in order to estimate the integer ambiguities. A significant time may also be required for re-convergence (Jokinen 2014).

The performance of the different GNSS systems and techniques are presented in Table 3.2 and Figure 3.5

As discussed above, the use of augmented GNSS effectively removes a number of error sources and can result in improved accuracies at the metre level. Using carrier phase measurements can further improve accuracies to the decimetre or centimetre levels.

However, GNSS-based positioning remains sensitive to the loss of the signal between the receiver and the satellite due to obstruction and radio frequency interference. Therefore, it alone cannot offer a

continuous positioning service that fulfils the stringent requirements of the railway. Hence, the use of additional sensors is required.

Table 3.2- Positioning performance of GNSS, adapted from (Scot 2014)

		Open Services				Safety of Life	High Precision	
		Galileo	GPS single-frequency	GPS dual-frequency	EGNOS OS (observed)	EGNOS SoL	RTK	PPP
Accuracy (95%)	Horizontal	4 m	9 m	6 m	3 m (1 m)	3 m	~ 1 cm	~10 cm
	Vertical	8 m	15 m	10 m	4 m (1.5 m)	4 m	~ 1 cm	~10 cm
	Time	30 ns	40 ns	40 ns	13 ns (13 ns)	13 ns	-	-
Availability		99.5%	99%	99%	99% (99%)	99%	local	95%
Integrity		Yes*	Yes**	Yes**	No	Yes	No	No

(*) available with Galileo Public Regulated Service (PRS), (**) as defined in USA Department of Defense (2008, p11)

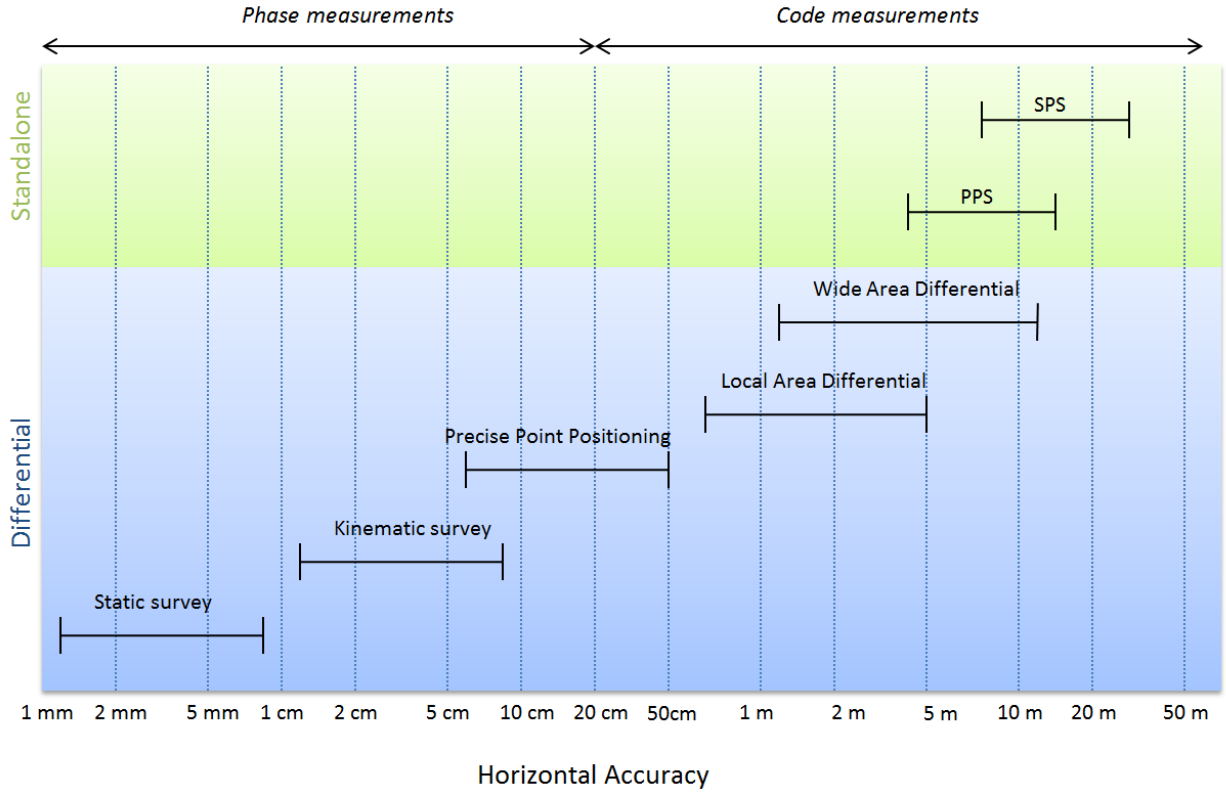


Figure 3.5- Accuracy of GNSS positioning techniques, adapted from (Julien 2011)

3.1.3 Terrestrial sensors

Integrated systems (i.e. composed of several sensors) are commonly used in aviation and road transport. They usually involve inertial sensors and map matching algorithms. New systems employing signals of opportunity and image matching are also being developed. The following section reviews these navigation techniques.

3.1.3.1 Inertial sensors

Inertial sensors comprise of both accelerometers and gyroscopes. Accelerometers sense a vehicle's rectilinear motion while gyroscopes sense the rotational motion (i.e. the vehicle angular rotation rate) without using an external reference. Inertial measurement units (IMU) are commonly used to obtain a three-dimensional navigation solution. They are composed of three accelerometers and three gyroscopes that enable the movements of a vehicle along all three axes to be measured. Since train movement is restricted by the rails, a simpler inertial system than IMU might be used to determine the train position (RSSB 2012a). Inertial sensors present the advantages of being autonomous, non-radiating and insensitive to external perturbations. However, they are expensive and subject to an accumulation of errors leading to a position drift (Groves 2013). The performance of the IMU depends on its grade and cost, as shown in

Table 3.3.

Table 3.3- Inertial navigation systems grade, performance and cost, from (Groves 2013)

Inertial Navigation System	Position drift	Cost
High accuracy system	0.1 NM/hr = 185m/hr	>$1 million
Medium accuracy system	1 NM/hr = 1.85 km/hr	$5 000 to $100 000
Low accuracy system	10 NM/hr = 18.5 km/hr	<$10

Inertial sensors have been extensively used in combination with GNSS receivers as they provide complimentary performance. The GNSS position can be used to align and calibrate the IMU, while the IMU can provide a solution when GNSS measurements are unavailable.

3.1.3.2 Digital maps

Digital maps are used to perform map-matching, that is to determine the location of the train on the rail network. A map-matching algorithm compares the position solution from the navigation system with a rail tracks database and supplies a correction if the navigation solution drifts off the track (Groves 2013). Different types of correction exist. The solution can simply be perpendicularly projected on the track

or the estimation can take into account the heading and speed information of the vehicle (Quddus et al. 2005). Map matching requires a digital, accurate and up-to-date map of the network. Some of the weaknesses of map matching include incorrect identification of the right track, especially when the network is dense, for example, at a station with many parallel tracks.

Map-matching is required to put the positioning information in context. For instance, the signalman cannot directly use the latitude and longitude of the train. The track mileage (i.e. the position of the train along the track) is required and can be estimated by matching the train position to a digital map of the network. Map-matching can also be used as a positioning aid, for instance, to determine the heading of the train based on its position and the track map.

Three-dimensional maps can also be used to improve the accuracy of conventional positioning techniques in dense urban areas. These maps containing information about the position, size and shape of buildings can be used to predict whether signals are blocked or reflected. Non-line-of-sight (NLOS) signals, i.e. only received via reflections, can then be excluded from the positioning estimation process (Groves et al. 2012), leading to a reduction of 69.2% on average of the cross-street positioning error (Wang et al. 2013). This technique is of particular interest for railway applications as it can help determine on which track the train is travelling on in urban areas.

3.1.3.3 Signals of Opportunity

Although designed for other purposes, signals of opportunity (SOOP) can be used for positioning. Examples include radio, television broadcasts and Wi-Fi (Groves 2013). Measurements can be taken based on: Time of Arrival (ToA), Time Difference of Arrival (TDoA), Angle of Arrival (AoA) and Received Signal Strength (RSS) (Moghtadaiee et al. 2011). Transmitters situated at known locations broadcast information regarding their identity and the transmission time, which can be used to compute the user position. The RSS can also be used to perform location fingerprinting in which a database of parameters dependent on location (generated in a training phase) is used to estimate the user position. Examples of accuracies achieved by different SOOP system are shown in Table 3.4.

Signals of opportunity are useful in urban and indoor environments where a GNSS-based position is more difficult to compute. However, as the signals are not originally designed for positioning, they may not contain the transmitter location. In that case, an up-to-date database with sufficient resolution is required.

A positioning system that combines GNSS and GSM-R, the railway dedicated radio signals (see section 2.2.2) has been investigated by (Faragher 2011). It relies on the GSM-R signals used by the train driver to communicate with the control centre via radio. These signals are available on the entire rail network, including in tunnels. The times of arrival of the signals are used to determine the speed of the train and a dead-reckoning algorithm is used to estimate the position. The main challenge is to determine the

offset between the GSM-R signals as the masts are not synchronised. However, these masts are usually installed so that the areas they cover are overlapping, enabling the offset to be estimated. In the future, GSM-R will be replaced by LTE (Long Term Evolution), which is expected to further improve communication-based positioning for railways.

Table 3.4 - Signal of Opportunity positioning accuracy, adapted from (Groves 2013)

System	Accuracy (95%)
Analog broadcast (e.g. Amplitude Modulation (AM))	about 2% of the wavelength (60 m for a frequency of 100 kHz)
Digital broadcast (e.g. Rosum Television Positioning System)	5 to 20 m
Wireless Local Area Network (also known as Wi-Fi or IEEE 802.11)	1 to 5 m
Ultra Wide Band Positioning	up to 50 cm

3.1.3.4 Vision-based positioning

Image matching techniques require a camera acquiring images that can then be used in different ways. They can be compared with a database in order to determine the camera's viewpoint (i.e. its location), or they can be used to identify individual features. However, both these techniques require a large database. The images can also be used to perform a 'visual odometry' by comparing successive images (Groves 2013). However, it is difficult in this case to differentiate linear and rotational movements and additional sensors may be needed to calibrate the system.Cameras can also be used to improve GNSS-based positioning performance by characterising the environment through which the vehicle is travelling. In (Marais et al. 2012), cameras are installed on the vehicle roof, covering a large field of view (180 degrees). The images obtained by these cameras are then processed to characterise regions of “sky” and “non-sky”. This information is then used along with the satellites elevation and azimuth angles to exclude non-line-of-sight (NLOS) satellites, which reach the receiver only after reflections. This method enhances conventional positioning accuracy. Sánchez et al. (2016) showed that a combination of GNSS, IMU and fisheye camera in an urban environment improves the position accuracy from 10.60 m (95%) for a GNSS/IMU solution only, to 6.93 m (95%) with the camera.

However, vision based positioning techniques are affected by weather conditions and it can be difficult to make the difference between “sky” and “non-sky” regions if the luminosity is too high or too low, or if it is raining. Also, frequent maintenance is required to clean the lens.

3.1.3.5 Conclusion

Numerous positioning sensors and systems, with different positioning performance, are currently available. They can be used in combination to achieve even higher levels of performance. However, the choice of sensors should be optimised as a function of the positioning application requirements in terms of accuracy, integrity, continuity and availability.

It can, however, be noticed that satellite-based navigation systems offer an attractive solution to support railway operations, as they provide an autonomous and continuous positioning service with a global coverage. Furthermore, they offer a solution that is both independent of track-side equipment and inter-operable across borders. Finally, their cost is relatively low and they are easy to install (i.e. they do not require to retrofit the train cab). In light of these advantages, a Memorandum of Understanding (MoU) was signed by the European Commission, the European Railway Agency (ERA) and the European Rail Sector Associations (UNIFE) in April 2012, recognising that Global "Navigation Satellite System (GNSS) can play a major role in the rail sector, both for fleet management and rail safety (signalling and train control)" (European Commission et al. 2012). This MoU places GNSS as a core element of the future railway positioning system, and the rest of this thesis focuses on the development of a GNSS-based railway position solution.

The following section presents research projects which have investigated such solutions. It reviews the applications targeted by these projects as well as the sensors selected and the positioning performance reached.

3.2 GNSS-based systems in the railway industry

The railway industry has realised the benefits of GNSS-based navigation systems and several studies have been carried out in Europe to investigate the performance of a GNSS-based positioning system in the railway environment.

As a result of a comprehensive literature review, this section presents the different GNSS-based railway projects in Europe. It identifies the different applications targeted by these projects as well as their aim and the positioning solutions they investigated.

While this literature review focuses on projects carried out in Europe, other countries such as China and the USA are adopting intelligent train traffic management systems similar to ERTMS and using GNSS (Marais et al. 2017). ERTMS itself is being adopted worldwide, in countries such as Saudi Arabia, South Korea, Australia and India. However, all these projects face the same challenge relative to the standardisation of the use of GNSS in the railway domain, as illustrated in the following section, on the case of Europe.

3.2.1 Applications targeted by European Union projects

Research in the field of GNSS positioning applied to railway operations started in 1999 with the project Advanced Positioning Locator (APOLO). This project reviewed the different railway applications that could benefit from an enhanced positioning function and was the first to acknowledge the fact that each application needed its own set of positioning requirements (APOLO 1999). Subsequent projects built on the foundations laid down by the APOLO project, most of them focusing on a dedicated application, as can be seen in Table 3.5.

Numerous GNSS-related railway projects have investigated the potential of satellite navigation to support railway operations. A large number of projects (LOCOPROL 2005; Urech 2002; GaLoROI 2012; Barbu & Marais 2014) aimed to develop a GNSS-based alternative positioning system to support traffic management and control on low-density lines. Such a system will have a lower cost than standard ERTMS equipment (e.g. balise) while delivering an equivalent level of performance.

Other projects (InteGRail 2010; European Space Agency 2005; Albanese & Marradi 2005; GSA 2005; Senesi 2016) focused on safety-critical applications and on the use of EGNOS safety-of-life features to perform train positioning. Some projects also investigated ways GNSS can be used to support ERTMS, for example by enhancing the odometer performance (GRAIL 2006b) or by implementing virtual balises (NGTC 2016). Virtual balises aim to replace physical balises in order to reduce the installation cost of ERTMS. As mentioned in section 2.3.2, such a system compares the train position against a database containing the location of the virtual balises. A message similar to the one sent by the balise is delivered to the onboard system when the train comes across the location of the virtual balise.

Many of these projects investigated the certification aspect required to support the adoption of GNSS in the railway domain. Indeed, any alternative positioning system will have to demonstrate its ability to fulfil railway reliability, availability, maintainability and safety (RAMS) requirements prior to their installation on a train. A link between the railway RAMS requirements and the positioning Required Navigation Performance (RNP) has been derived by (Beugin et al. 2010).

Despite the large number of projects carried out since 1999, the efforts made to introduce GNSS into the railway domain have not been consolidated. Most of the projects focus on one application, (e.g. train control for low-density line, enhanced odometry), leading to the development of ad hoc positioning platforms, fulfilling the needs of only specific services. While this approach develops a better understanding of GNSS performance in specific cases, it restricts its possible overall benefits.

However, the RSSB took an interesting approach by characterising the railway environment for GNSS-based positioning (RSSB 2007) and by testing the performance of different receivers in such an environment (RSSB 2012a). This global approach enables a better understanding of the challenges posed by railway and the development of a positioning solution able to holistically support location based railway services. This approach is followed in this thesis.

Table 3.5 - GNSS-based European railway projects

Project	Period	Aim
APOLO Advanced Positioning Locator	1999-2001	Define and test a GNSS-based prototype for train position
LOCOPROL Low-cost satellite-based train location system for signalling and train protection for low-density traffic railway lines	2001-2004	Develop a GNSS-based train protection, control and command system, with a focus on low-density traffic railway lines
GADEROS Galileo Demonstrator Railway Operational System	2001-2004	Demonstrate the use of GNSS safety-of-life features for defining a satellite-based system to perform train location for safe railway applications on low traffic lines and that can be integrated into ERTMS/ETCS.
INTEGRAIL Intelligent Integration of Railway Systems	2001-2004	Use of EGNOS signal in safety-critical railway traffic management and control, focusing on secondary or rural lines
ECORAIL EGNOS Controlled Railway Equipment	2001-2005	Use of GNSS in safety critical railway application (level crossing demonstration)
RUNE Railway User Navigation	2001-2006	Demonstrate the use of GNSS integrity and safety of life service (EGNOS) to perform train location for safety critical applications
LOCOLOC Locomotive Location	2002-2004	Develop and demonstrate a complete, very low cost, fail-safe train navigation and integrity system based on GNSS, focusing on speed and acceleration, as well as a service centre to complement LOCOPROL
GEORAIL Railway geodesy	2004-2008	Develop a geodetic framework with the use of a Coordinate-based continuous Numerical Track Description
GIRASOLE Galileo Integrated Receiver for Advanced Safety of Life Equipment	2005-2007	Develop a receiver dedicated to rail applications using Galileo and EGNOS Safety of Life services
AIL GNSS introduction in the RAIL sector	2005-2008	Support the introduction of GNSS in the rail market with an emphasis on ERTMS/ETCS and especially on enhanced odometry, absolute positioning, cold movement detection and train awakening, and train integrity
T740 Global navigation satellite systems data coverage analysis for railway operations	2004-2007	Characterise the railway environment and define the contribution required from inertial sensors

T510 Obtaining data to assess the dependability of GNSS information and accuracy of odometry	2006-2009	Understand of the dependability of GNSS-derived position and speed for automatic train protection on low density and rural lines
IRISS Intelligent Railway via Integrated Satellite Services	2009-2012	Use satellite navigation services and communications to improve train operating companies operations
GRAIL2 GNSS-based enhanced odometry for railway	2010-2013	Develop and validate a GNSS-based enhanced odometry for ETCS application in high-speed railway lines
T892 Data analysis for cost-effective GPS-locator with simple augmentations	2011-2013	Test a number of GNSS receivers and augmentations to assess the capacity to fulfil railway positioning requirements
GALOROI Galileo Localization for Railway Operation Innovation	2012-2014	Develop a certifiable safety relevant satellite-based onboard train location unit to be used on low traffic density railway lines
SATLOC Satellite-based operation and management of local low traffic lines	2012-2014	Develop and demonstrate GNSS safety (EGNOS and Galileo) in live rail application for train control, speed supervision, traffic control and management of low traffic lines
EATS ETCS Advanced Testing and Smart Train Positioning System	2012-2016	Provide a model of the complete onboard ERTMS system behaviour and propose a novel positioning system based on GNSS to enable the migration from ETCS level 2 to level 3
NGTC Next Generation Train Control	2013-2016	Standardisation of the virtual balise concept, with the requirement of backwards compatibility with ETCS level 3
3INSAT Train Integrated Safety Satellite System	2014-2016	Develop and validate a new satellite-based platform to be integrated into an ERTMS system by providing a SIL4 compliant positioning solution through the association of different GNSS receivers, onboard sensors, augmentations and integrity monitoring networks and independent onboard integrity monitoring capability.
ERSAT ERTMS on Satellite – Enabling Application Validation	2015-2017	Verify the suitability of EGNSS (Galileo and EGNOS) as the enabler of cost-efficient and economically sustainable ERTMS signalling solutions for safety railway applications

3.2.2 Sensors used

This section reviews the positioning sensors selected in the different projects presented in the previous sub-section. These sensors include multi-frequency, multi-constellation GNSS receivers, satellite augmentations, odometer, inertial sensors, balise, digital maps and Doppler radar. Table 3.6 summarises the findings of the literature review.

Table 3.6 - Positioning sensors used in GNSS-based European railway projects

	GPS L1	GPS L1/L2	GLONASS	GALILEO	EGNOS	Odometer	Doppler radar	INS	Maps
APOLO	✓		✓		✓	✓	✓	✓	✓
LOCOPROL	✓	✓	✓		✓	✓		✓	✓
ECORAIL	✓				✓	✓			✓
RUNE	✓				✓	✓		✓	✓
GRAIL	✓	✓		✓	✓	✓		✓	✓
T510	✓	✓				✓		✓	
IRISS	✓	✓	✓		✓			✓	
GRAIL2	✓	✓			✓	✓	✓	✓	✓
T892	✓	✓	✓			✓		✓	
GALOROI	✓			✓	✓				✓
SATLOC	✓	✓		✓	✓	✓			✓
3INSAT	✓		✓	✓	✓	✓		✓	✓

The majority of projects use dual-frequency GPS receivers, sometimes in combination with GLONASS, as the use of an additional constellation can increase satellite availability in areas with poor visibility. Some projects have also assessed the impact of the future Galileo system (Barbu & Marais 2014) through simulations. The European augmentation system EGNOS is used in most projects to provide corrections and integrity information to the receiver.

Most systems use the odometer due to its availability onboard trains to measure their speed. The odometer data can easily be integrated into the positioning solution. When available, Doppler radar measurements are used also. Inertial sensors, traditionally used in combination with GNSS in road transport and aviation, are also extensively used in these projects.

Maps are recognised as an important element and are used in most projects. The GEORAIL project (GEORAIL 2008), presented in the previous subsection, was entirely dedicated to the development of a geodetic framework for the creation of maps that could subsequently be used to perform map matching.

While most projects present the sensors used to compute the position of the train, the majority fail to explain how the measurements are combined. Additionally, not all the positioning performance information is available.

3.2.3 Performances

This subsection presents the performance of the GNSS-based railway positioning systems developed by the previously presented projects (see Table 3.5). In addition, some preliminary results are presented related to the data that is used in chapter 1.

Despite the large number of projects that are investigating the use of GNSS to support railway applications, the performance of the systems they develop is often unavailable. The results found are presented below. First, results regarding satellite visibility are presented as this has a direct impact on the position availability since a minimum of four satellites is required to estimate the receiver position (if the position is solved in three dimensions). Table 3.7 presents the satellite visibility measured by the LOCOPROL project (LOCOPROL 2005) and the T510 project (RSSB 2007). The two last rows of the table, high-speed train and mainline train, present the results of a preliminary analysis performed using two datasets. The first dataset was collected during for the GRAIL project onboard a high-speed train, in Spain, travelling mostly through rural areas. The second dataset was collected for the T892 project onboard a mainline train travelling in the suburban area of Birmingham. A more detailed description of these datasets is provided in section 1.4. The results presented in this section are used to build a better understanding of GNSS-based positioning system performance in the railway environment.

It can be seen from Table 3.7 that in all scenarios, four and more satellites were visible at least 85% of the time, with a reduction to 75% if integrity is required (six satellites and more). The use of a second constellation (GPS and GLONASS) appears to have limited impact on the satellite visibility, though it slightly improves the availability of the solution if integrity is required (visibility of 6% higher when using GLONASS). The grade of the GPS receiver also appears to have a limited effect on the satellite visibility.

Table 3.8 represents the duration of GNSS outages as assessed by the T510 project (RSSB 2007) and as estimated using the two datasets previously described. It should be noticed that in the case of the high-speed train dataset, over 28 hours of data were analysed, while in the case of the mainline train, only a 25-minute-long dataset was available for analysis and thus the results may not be representative. The dataset used in the T510 project is quite similar to the mainline train as it was also collected on mainline trains travelling in the UK. Both datasets show that a large majority of the GPS outages have a duration shorter than five seconds. An average outage duration of two seconds was derived from the T510 analysis which is most likely related to the presence of obstacles such as overhead bridges or trackside buildings. The 4% of outages longer than three minutes can be caused by either long tunnel or covered stations. Interestingly, outages shorter than five seconds only represent 63% of all outages in the case of the high-speed train, which means that on average outages are longer for the high-speed train than from the mainline train. However, 90% of the outages are shorter than ten seconds. This difference in outage duration can be due to longer obstructions (e.g. tunnels) or to the receiver high

dynamics that makes it harder for signals to be re-acquired after an outage. Finally, from the mainline analysis, the professional grade receiver appears to re-acquire signals faster than the medium/high-grade receiver and the use of GLONASS slightly reduce the duration of the outages.

In terms of positioning accuracy, Table 3.9 summarises the findings of the different research projects APOLO, T510 and T892. The integrated GPS, odometer and gyroscope positioning platform, tested onboard trains in both the Czech Republic and Spain by the APOLO project, achieved an accuracy of 4.8 m (95%), without using EGNOS corrections. Using EGNOS corrections, the accuracy was improved to 1 m (95 %) (Barbu 2001). However, such a level of accuracy was not achieved by the T510 project using a high-grade GPS receiver with a six degrees of freedom INS. It is possible that the performance of the APOLO positioning solution was obtained under good satellite visibility conditions, and without long periods of signal outages, limiting the drift of the gyroscope. The T892 project shows that a single frequency high sensitivity GPS receiver presented an accuracy of 3 to 5 m (95%) in the railway environment.

Table 3.7 – Satellite visibility

Project	Environment	Constellation	Visibility ≥ 4	Visibility ≥ 6	Average
LOCOPROL	Good areas *	GPS and GLONASS	98%	95%	-
	Bad areas *	GPS and GLONASS	96%	80%	-
T510	Urban areas	GPS	85%	75%	-
	Rural areas	GPS	>95%	>95%	-
High-speed train	Rural areas	GPS	85%	82%	7
Mainline train	Suburban areas	Professional grade GPS	91%	84%	7
	Suburban areas	Medium/high grade GPS	90%	84%	7
	Suburban areas	Medium/high grade GPS and GLONASS	91%	90%	12

(*) as defined in (LOCOPROL 2005)

Table 3.8 – Duration of GNSS outages

Project	Environment	Constellation	> 3 min	< 5 sec	Average
T510	Urban areas	GPS	-	85%	-
	Rural areas	GPS	4%	98%	2 sec
High-speed train	Rural areas	GPS	0.01%	63%	6 sec
Mainline train	Suburban areas	Professional grade GPS	0%	83%	6 sec
	Suburban areas	Medium/high grade GPS	0%	74%	7 sec
	Suburban areas	Medium/high grade GPS and GLONASS	0%	78%	6 sec

Table 3.9 - Accuracy of GNSS-based positioning systems in the railway environment

Project	System	Accuracy (95%)
APOLO	GPS (L1), odometer, gyroscope	4.8 m
	GPS(L1), EGNOS, odometer, gyroscope	1 m
T510	High-grade GPS, gyroscopes (3 axes), accelerometer (3 axes)	~ 1 to 2 m
	Low grade (high sensitivity) GPS, single axis gyroscope, odometer, direction of travel	~ 2 to 8 m*
T892	GPS L1 high sensitivity	3 to 5 m

(*) if outages < 5 sec, for longer GPS outages errors up to 30 m were observed

While giving an indication of the performance of different positioning systems, these results also highlight that further research is required to quantify the performance of different sensors in the diverse railway environment.

In conclusion, 16 years of research have paved the way to the development of a certifiable GNSS-based positioning system to support railway operations. However, a piecemeal approach has been followed with each project focusing on a specific application and developing its own positioning system. The European GNSS Agency (GSA) is now coordinating efforts with European Railway Agency (ERA) and has developed a roadmap, shown in Figure 3.6, describing how different projects will lead to the development of a certifiable positioning system by 2020.

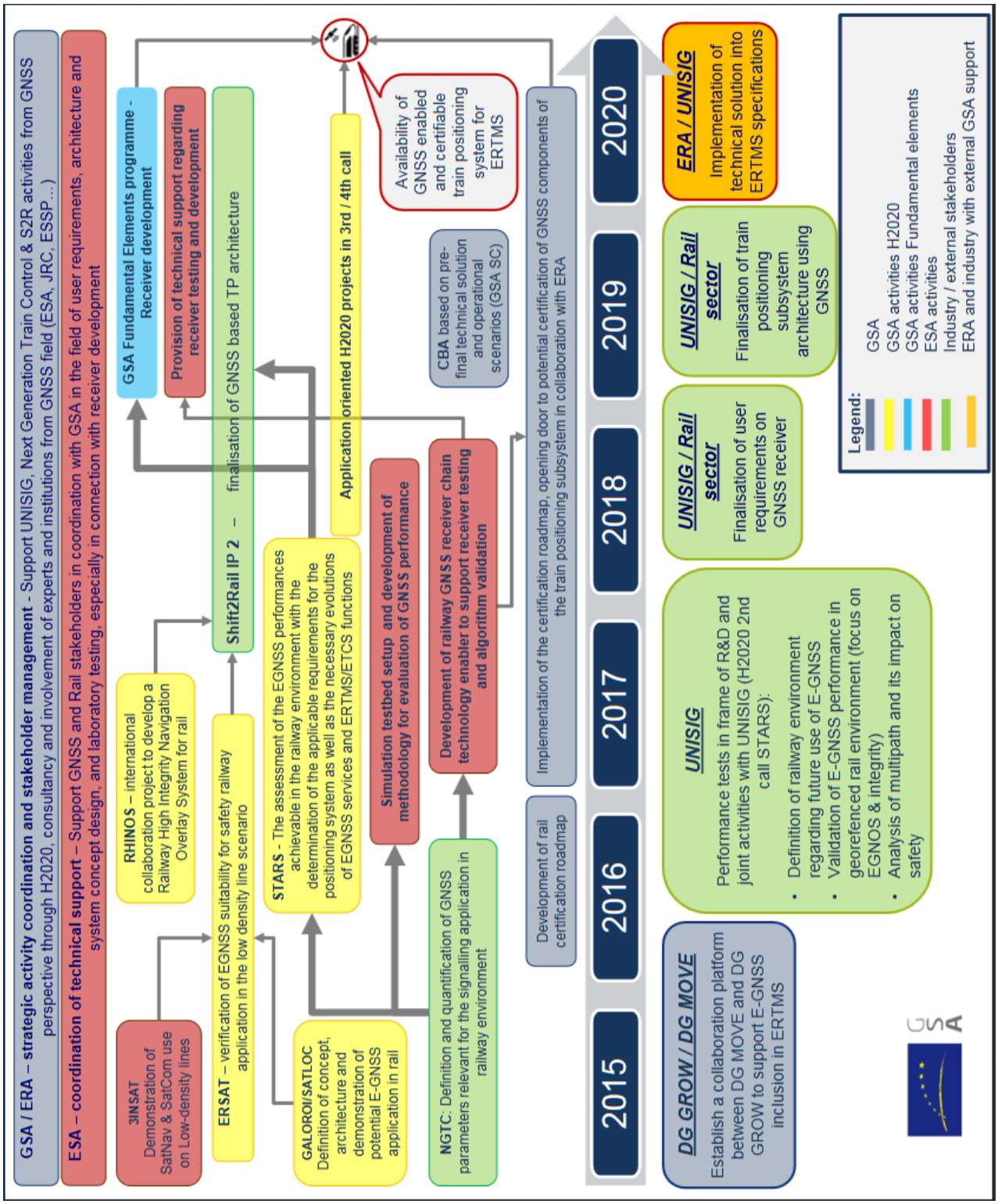


Figure 3.6 - E-GNSS enabled railway signalling – from vision to action, (European GNSS Agency 2016)

3.3 Selection of sensors

Based on the previously reviewed positioning sensors and solutions investigated by research projects, this section discusses a possible architecture for a railway positioning system. Firstly, the positioning

drivers, including the positioning requirements and operational environment constraints, specific to railway applications are reviewed. The second part assesses the capacity of the different positioning sensors to achieve the required performance. Finally, as these sensors will be used in combination, integration techniques are reviewed in terms of their benefits and disadvantages. In concluding this chapter, a potential architecture for railway positioning is presented.

3.3.1 Positioning drivers

In addition to fulfilling the stringent positioning requirements defined in section 2.3.3, the novel positioning system should be adapted to the railway environment and its operations.

The railway environment poses a challenge due to numerous obstacles that may occur along the track. Cuttings, trees, overhead power lines, bridges and gantries are some of the many elements that can limit satellite visibility and reflect the signals sent by the satellites. Station canopies and tunnels can even completely block signals. From 2000 to 2007, the RSSB collected GNSS data along strategic railway lines to assess the GNSS coverage over UK railway lines (RSSB 2007). While most of the UK can see around seven satellites, it was estimated that approximately 10.5% of the network (in stretches of over 150 m) either has poor coverage or GPS outage. It was also estimated that 4% of the network, corresponding to 722 km of track, is covered by tunnels, with most tunnels being less than 150 m in length (RSSB 2007).

Such an obstructed environment makes the use of satellite navigation challenging. However, the guided nature of railway transport can help to reduce the dimensionality of the positioning problem. In fact, if an accurate and up to date digital map of the railway network is available, only one item of information is required, the position of the train along the track, sometimes called mileage or curvilinear abscissa (Neri et al. 2012). The three-dimensional problem of determining the train latitude, longitude and height can then be reduced to one dimension. In turn, fewer measurements are needed, hence two satellites are sufficient in order to estimate the train position and any additional measurements can still be used to check the reliability of the computed position. The digital data map also enables the conversion of the one dimension curvilinear abscissa back to the three dimension position.

Another interesting aspect of railway operations is the low dynamics of the train. Indeed, due to the steel to steel contact, trains accelerate and decelerate over a long period of time. Intercity train commonly have an acceleration of between 0.4 and 0.6 m/s^2 and a deceleration of 0.4 to 0.5 m/s^2 (Profillidis 2013). In comparison, a passenger car typically accelerates at 1.5 m/s^2 and decelerates at 2 m/s^2. Additionally, trains change directions in a very smooth way, as the railway tracks consist of straight sections and curve of large radius connected to each other by a series of variable radius transition curves (Profillidis 2013). On the UK rail network, 97% of the curves have a radius of larger

than 500 m (Profillidis, 2013, p302). This low dynamics and seamless change in the train direction can be used to average the values and smooth the positioning solution, or even to detect outliers.

In summary, the railway environment is challenging to satellite navigation due to the obstructed environment a train has to travel through. However, the guided nature of railway transport, as well as its low dynamics, might be exploited to develop novel positioning techniques, requiring fewer measurements and relying more heavily on the position prediction.

3.3.2 Performance

Based on the sensors reviewed in section 3.1 and the performance of the systems presented in section 3.2.3, the following subsection analyses the capacity of each sensor to fulfil the positioning requirements defined in the previous chapter (section 2.3.3). Table 3.10 summarises this comparison.

Table 3.10 demonstrates that one system alone cannot fulfil the railway operations positioning requirements. An accuracy of 1 m (95%) can be achieved using RTK techniques. However, it may require the development of a dedicated augmentation network, as suggested by (Filip & Rispoli 2014b). Although EGNOS can be used to provide some integrity information, it cannot provide the levels required by the railway applications and additional integrity monitoring functions are needed at the receiver level.

Satellite-based positioning is limited by satellite visibility and is unable to provide a continuous positioning service over the required coverage area. While inertial sensors can be used to aid the continuity of the positioning service, the INS grade needs to be sufficient to maintain the positioning accuracy in areas with no satellite visibility, such as tunnels. Consider a train travelling at 45 km per hour through a 150 m long tunnel. To maintain the 1 m accuracy requirement, the INS drift should be less than 300 m per hour, which corresponds to a medium/high-grade system (see

Table 3.3).

Signals of opportunity offer an alternative positioning solution in urban areas. They can achieve an accuracy of 1 m or below (see Table 3.4) but have very limited integrity information, requiring the use of additional sensors to ensure the safety of operations.

Track circuits and axle counters do not provide the required level of positioning granularity and thus present a limited interest in the development of a novel positioning system. The onboard odometer, despite its limited accuracy, meets the railway safety requirements (safety integrity level (SIL)). Similarly, the balise offers discontinuous positioning information but with the required level of accuracy and integrity. Due to their certified performance, the integration of balise and odometer readings into the novel positioning system might be of interest.

Table 3.10 - Positioning performances and railway requirements

	Coverage Known routes	Accuracy 1m	Integrity risk 10^{-9}/h	Alert limit 3 m	Time to alert 2 sec	Continuity 10^{-6}	Availability 99.99%
SOOP	local	✓	×	×	×	×	×
INS	✓	depends on grade	×	×	×	×	×
PPP	×	✓	×	×	×	×	×
RTK	local	✓	×	×	×	×	×
GBAS	local	×	×	×	✓	×	×
SBAS	×	×	×	×	×	×	×
GNSS dual frequency	×	×	×	×	×	×	×
GNSS single frequency	×	×	×	×	×	×	×
Odometer	✓	×	✓*	×	×	×	×
Balise	×	✓	✓*	×	×	×	×
Track circuit Axle counter	✓	×	×	×	×	×	×

(*) as the integrity risk is derived from the railway tolerable hazard rate for which both the balise and the odometer are certified

In conclusion, the railway positioning requirements can only be achieved by using a combination of sensors. This sensor combination can follow two main schemes that are presented in the following sub-section.

3.3.3 Multi-sensors integration techniques

Integrating multiple sensors into one positioning system can be done following two main schemes, either loose or tight integration.

In loose integration, each sensor operates independently and produces its own positioning solution. The solutions are then blended to form the final solution that extracts the desirable attributes of each solution and suppress those that are undesirable. A feedback loop can be implemented to correct some measurements and to ensure filter convergence (Gebre-Egziabher 2007). Figure 3.7 illustrates a GPS/INS loose integration.

Loose integration can be implemented with any navigation equipment. However, measurements are available only when signals are both sufficient in number and of appropriate strength to generate a position solution (Groves 2013).

In tight integration, presented in Figure 3.8, basic measurements such as pseudoranges, accelerations and angular rates are blended to obtain a single solution. This approach has a greater computational requirement but enables gains in solution availability. Indeed, tightly coupled integration is able to incorporate terrestrial radio navigation measurements so when there are insufficient GNSS signals, a position solution can still be computed (Gebre-Egziabher 2007).

Tight integration can provide a feedback to the GNSS receiver to help the tracking loop and avoid loss of lock in a highly dynamic environment. The feedback loop can also be used to narrow the tracking loop, reducing the noise and making the system more robust to interference. Despite a gain in robustness, tight integration leads to an increased computational complexity compared to the loose integration scheme. Also, an error in the INS can affect the blended solution, which in turn can deteriorate the GNSS receiver performance through the feedback loop.

One major concern regarding the novel positioning system is the monitoring of the integrity of the solution. The sensors have to be combined with each other in such a way to enable efficient control of the integrity of the navigation system in its entirety. This means that the failure modes of each sensor have to be understood and also that the integration should be such that the impact of a single or combined failures can be assessed. Depending on the integration scheme, integrity monitoring is performed in the positioning domain or in the ranging domain.

Another interesting principle, context detection, is presented in (Groves et al. 2013). The idea is to design a context adaptive or cognitive positioning system, able to detect the environment it operates in (via camera and features recognition for instance) and to reconfigure its algorithms. This means that different solutions can be selected and combined on the basis of the environment detected. This technique can benefit railway applications as trains travel both through rural and dense urban areas. Context detection and adaptive algorithms are further discussed in section 5.1.3.

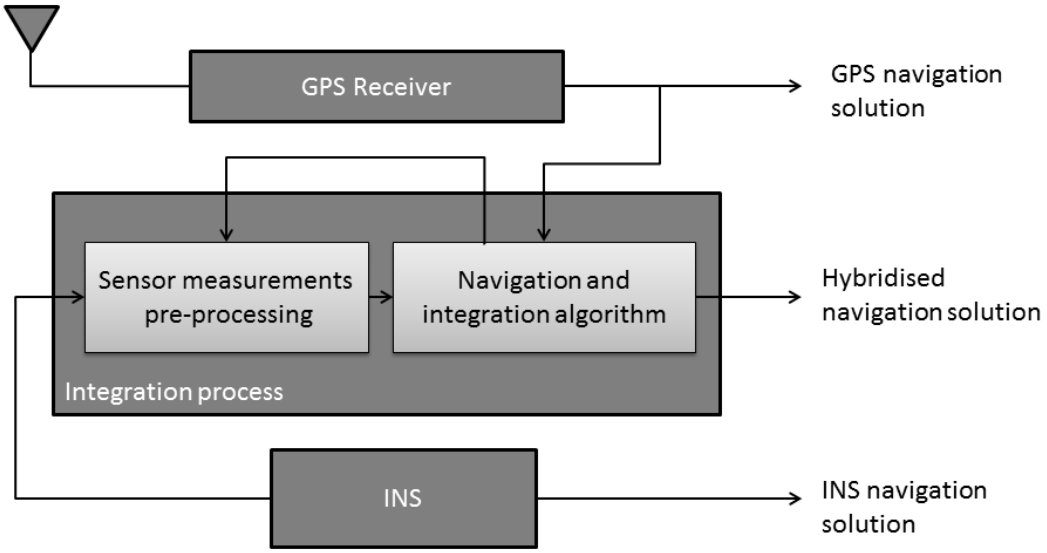


Figure 3.7 - GPS/INS loose integration, adapted from (Gebre-Egziabher 2007).

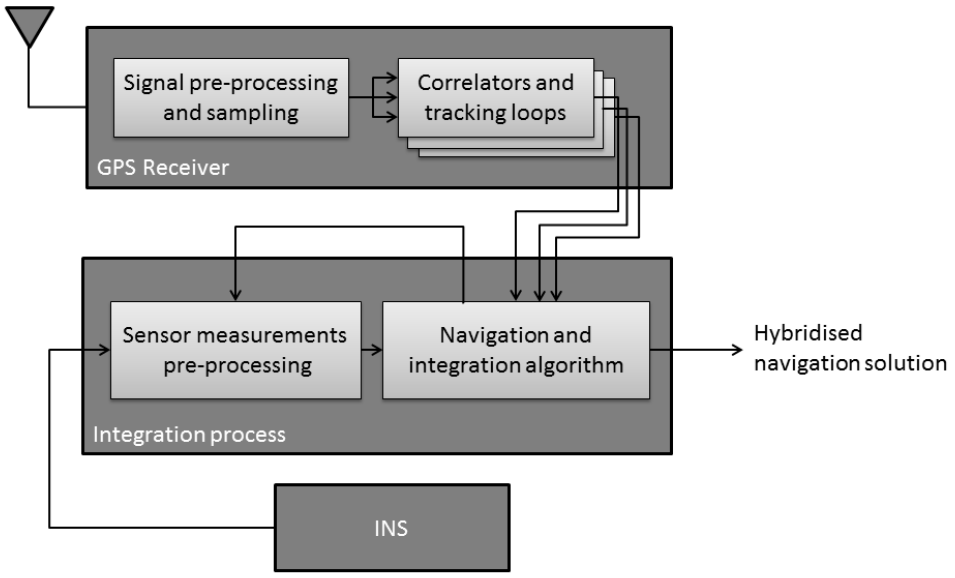


Figure 3.8 - GPS/INS Tight integration, adapted from (Gebre-Egziabher 2007).

3.3.4 Railway positioning system physical architecture

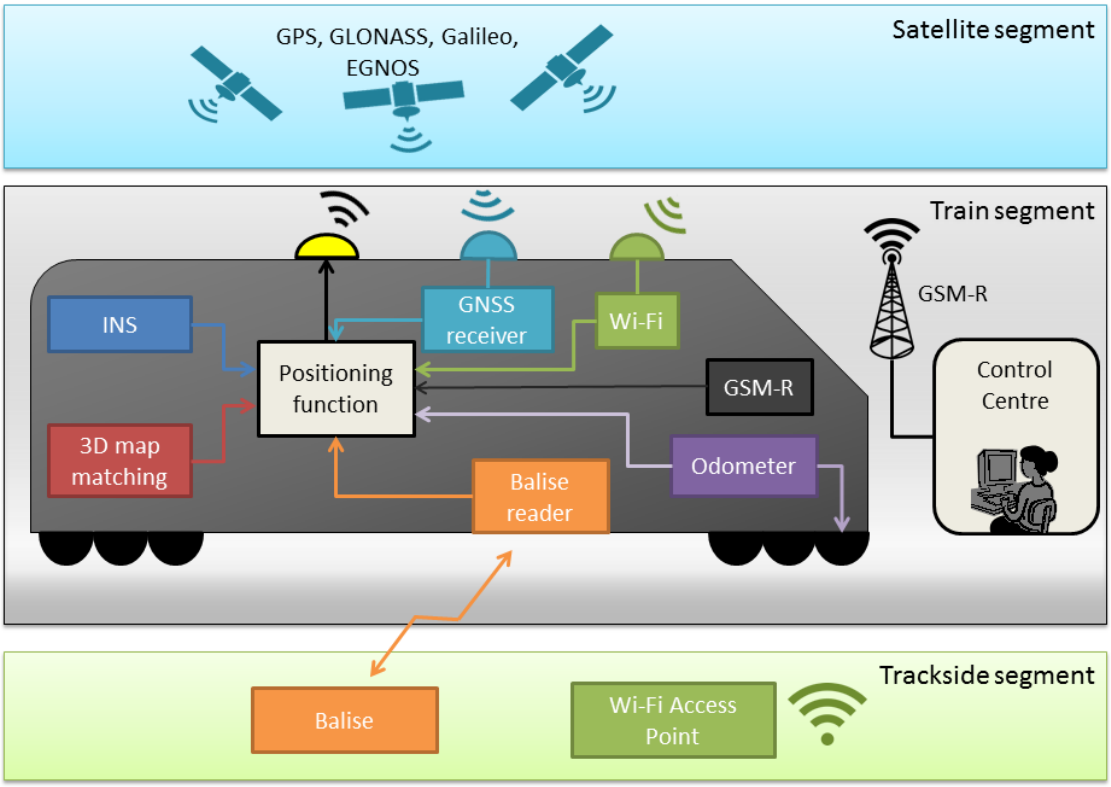


Figure 3.9 – Physical architecture of a novel GNSS-based positioning system to support railway location-based services.

Based on the characteristics of each sensor reviewed in section 3.1 and the combination of sensors currently used in the railway industry (section 3.2), three candidate physical architectures are discussed below.

One possible architecture consists in combining the GNSS receiver with signals of opportunity and three-dimension (3D) map matching techniques. In the case of GNSS signals outage, signals of opportunity such as Wi-Fi could be used to estimate the train position. The location of additional access points, for example in tunnels, will have to be determined in order to maintain the position accuracy. The 3D map matching will support the position estimate in urban areas and could help determine the track the train travels on as explained in section 3.1.3.2. The map could also be used as a positioning aid and provide the information of the train heading (based on the track location) to the positioning function. This system has the advantage of being independent of the train systems. However, it may not achieve the required level of integrity.

A second possible architecture consists in combining the previous sensors (GNSS, SOOP and 3D map matching) with the train odometer and balise reader. The use of additional sensors that are moreover certified for railway operations can bring a significant integrity gain. Balise readings can be used to

recalibrate the sensors and the positioning error while the odometer can be used to perform a consistency check.

Finally, the third possible architecture consists of combining all the previously mentioned sensors, as illustrated in Figure 3.9. Such a positioning system will benefit from the redundancy of the measurements. However, in order to ensure the integrity of the positioning solution, the failure mode of each sensor and of their combinations need to be characterised, thus a large number of sensors might significantly increase the system complexity.

Further research is required to investigate the positioning performance of these three candidate architectures. While acknowledging the need for a multi-sensor positioning system, this research concentrates on the GNSS receiver and investigates techniques to improve its positioning performance in the railway environment. The following chapter reviews the different error sources that can affect the performance of the GNSS receiver, focusing on the effect of multipath created by reflections of the signals on surrounding obstacles. Chapter 5 then characterises the impact of the railway environment on the GNSS signals and chapter 6 assesses the performance of different multipath mitigation techniques.

Chapter 4

GNSS Errors and Multipath Mitigation

As seen in chapter 3, GNSS has the ability to provide low-cost, global coverage, high accuracy positioning to support railway operations. However, GNSS is affected by several sources of errors that should be mitigated in order to improve the positioning system performance. This chapter reviews both the different error sources that affect GNSS signals and the existing mitigation techniques that can be used to reduce or remove their effect on the positioning solution. Code multipath errors, due to signal reflections on obstacles surrounding the receiver, are identified as a significant contributor to the positioning error in the railway domain and are reviewed in more detail in section 4.2. A particular interest is given to the use of weighting techniques to limit the effect of multipath affected measurement on the final positioning solution. Existing weighting techniques are reviewed in section 4.3 and their limitations are identified. These limitations are further discussed in section 4.4, where a preliminary analysis of the train GNSS data is performed.

4.1 Error sources and mitigation techniques

This first section presents the GNSS measurements used to estimate the user position and reviews the different sources of error that can affect these measurements and the existing mitigation techniques used to reduce or remove their effect. Finally, the impact of these errors on the positioning solution estimation is reviewed.

4.1.1 GNSS measurements

4.1.1.1 Code pseudorange

The basic GNSS observable is the transit time of the signal from the satellite to the receiver. This time is defined as the difference between the reception time as determined by the user receiver and the transmission time at the satellite as marked on the signal (Misra & Enge 2010). The measured apparent range R, called the pseudorange, can be determined by multiplying the transit time by the speed of light.

$$R(t) = c\big(t_u(t) - t^s(t-\tau)\big) \tag{4.1}$$

Where $t_u(t)$ is the arrival time measured by the receiver clock and $t^s(t-\tau)$ is the corresponding emission time at the satellite, τ being the transit time.

Accounting for the clock biases, the atmospheric delays and the measurement error, the measured pseudorange can be rewritten as:

$$R = \rho + c\left(\delta t_u(t) - \delta t^s(t-\tau)\right) + I_R + Tr_R + \varepsilon(t) \tag{4.2}$$

Where τ is the transit time, δt_u is the receiver clock bias, δt^s is the satellite clock bias, I_R and Tr_R are delays associated with the propagation of the signal through the ionosphere and troposphere respectively and ε represents the modelling errors and measurements errors.

The geometric range ρ between the receiver and the i^{th} satellite can be expressed as:

$$\rho_i = \sqrt{(x_i - x_u)^2 + (y_i - y_u)^2 + (z_i - z_u)^2} \tag{4.3}$$

Where (x_i, y_i, z_i) is the three-dimensional position of the satellite, (x_u, y_u, z_u) is the position of the user receiver.

The receiver only provides the user with the pseudorange R, which is a biased (i.e. affected by systematic errors) and noisy (i.e. affected by random errors) measurement of the true geometric range ρ required to compute the position. The accuracy of the position estimated using the pseudoranges depends therefore on the ability to compensate for, or if possible to eliminate, the bias and errors.

4.1.1.2 Carrier phase measurement

The carrier phase measurement is the difference of phases between the receiver-generated carrier signal and the carrier received from the satellite and is measured in units of cycles. As the pseudorange measurement, it can be used to estimate the geometric range between the user and the satellite. While the carrier phase can be measured with higher precision than the code measurement, it is, however, ambiguous. Indeed, the difference between the phases is composed of a number of whole cycles and a fractional cycle. While the fractional cycle can be measured precisely, no information is provided on the number of whole cycles, also called integer ambiguity, which makes the carrier phase measurement ambiguous, as illustrated in Figure 4.1.

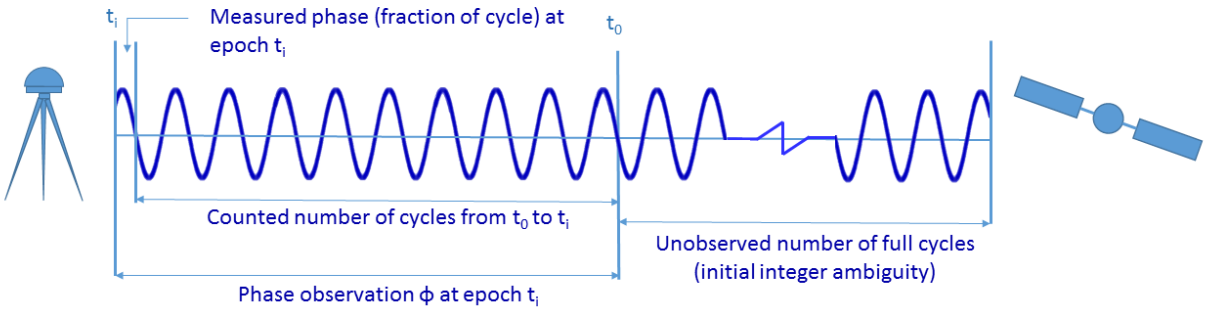


Figure 4.1 – Carrier phase measurement, adapted from (GIS Resources 2013)

The carrier phase measurement can be written:

$$\phi(t) = \phi_u(t) - \phi^s(t-\tau) + N_0 + \Delta N \quad (4.4)$$

Where ϕ_u is the phase of the receiver generated signal, ϕ^s is the phase of the satellite signal at the time of emission, τ is the transit time and N_0 is the initial integer ambiguity (unknown) and ΔN is the integer ambiguity measured by the receiver.

Accounting for the clock biases, the atmospheric delays and measurement error, the carrier phase measurement can be rewritten in a similar way to equation (4.2) as:

$$\Phi = \lambda\phi = \rho + c(\delta t_u - \delta t^s) + I_\phi + Tr_\phi + \lambda N + \epsilon \quad (4.5)$$

Where Φ is the carrier phase measurement in metres, ϕ is the carrier phase in cycles, λ is the wavelength of the signal, I_ϕ and Tr_ϕ are the delays associated with the propagation of the signal through the ionosphere and the troposphere, N is the integer ambiguity and ϵ represents the remaining errors.

In conclusion, two GNSS observables are available: the pseudorange measurement code that is noisy with a metre level accuracy and the carrier phase measurement that is very accurate (centimetre level accuracy) but ambiguous due to integer ambiguity.

4.1.2 Error sources and models

Both the code pseudorange and carrier phase measurements are affected by errors, traditionally grouped into three categories depending on their source:

- Satellite errors, which are due to errors in the parameter broadcast by the satellites in the navigation message
- Propagation medium errors, due to the effect of the atmosphere and environment on the signal propagation
- Receiver errors, which are due to the receiver clock and hardware

Additional errors such as site displacement created by solid Earth tide and ocean loading can also affect the measurements. However, their effect on the code measurement is relatively small compared to previously mentioned error sources (a few centimetres), and their effect is neglected in this thesis.

4.1.2.1 Satellite errors

Satellite errors include both clock and ephemeris errors. While both the ephemeris and clock parameters are computed by the control segment and broadcast by the satellites in the navigation message, their estimation and prediction are affected by errors. The magnitude of the error grows with the age of data (AoD), which corresponds to the time since the last data upload from the ground to the satellites. The control segment continuously monitors the clock and ephemeris errors and if they exceed a threshold a

contingency upload can be scheduled (Misra & Enge 2010). The range equivalent ephemeris error has a typical value of 3 m, while the clock error accounts for 0.5 to 1 m (Hegarty 2013).

4.1.2.2 Propagation medium errors

Depending on the satellite elevation angle, GNSS signals travel between 19000 and 26000 km to reach the receiver. During most of their travel, such signals propagate through a vacuum. However, at a height of 1000 km, the signals enter the atmosphere whose electrical content changes both the velocity and the transit time of the signals, thereby affecting the pseudorange and carrier phase measurements.

The ionosphere extends between 50 km and 1000 km above the earth. The electrical content of the ionosphere affects the speed of the signal, causing the code pseudorange to be delayed and the carrier phase to be advanced. As the ionosphere is a dispersive medium, this delay is a function of the signal frequency.

The pseudorange ionospheric delay is such that

$$I_R = \frac{40.3\ TEC}{f^2} \tag{4.6}$$

Where f is the signal frequency and TEC is the total electron content, which corresponds to the number of electrons in the path of the signals. The TEC is expressed in multiples of TEC units (TECU), defined such as TECU = 10^{16} electrons per m^2.

The ionospheric delays on the pseudorange and carrier phase are equal in magnitude but of opposite signs.

$$I_R = -I_\phi = \frac{40.3\ TEC}{f^2} \tag{4.7}$$

The ionospheric error varies with the time of the day and the seasons. Ionospheric delays on L1 range from 3 to 30 m at the zenith, and can be three times larger along oblique paths. Ionospheric delays on L2 are 1.6 times larger than the ones on L1 (Hegarty 2013).

After travelling through the ionosphere, the signal reaches the troposphere at a height of 50 km. In contrast to the ionosphere, the troposphere is non-dispersive and thus the tropospheric delay is the same for every frequency and the measurements experience the same delay.

$$Tr_\rho = Tr_\phi \tag{4.8}$$

The tropospheric error is approximately 2.5 m at the zenith and at sea level. There are two notable features of this error: it is smaller at higher altitudes and it is five times larger for satellites at an elevation angle of 10 degrees compare to a satellite at the zenith.

The propagation medium error also includes interference and multipath. The former are due to signals in a frequency band adjacent to GNSS signals and/or with a higher power, making the acquisition of signals erroneous or impossible. Interference can be either involuntary or voluntary and is a growing threat to GNSS. Concerns over individual privacy have driven the increased use of portable civilian jammers, also called "Personal Privacy Devices", used to prevent vehicle and people from being tracked (Pullen & Gao 2012). Such devices can have serious consequences on GNSS-based services. For instance, a portable jammer used onboard a truck was responsible for the loss of service of the GBAS station at Newark Airport, which had the potential to affect flight operations. Recent research has focussed on ways to detect interference and mitigate its effects (Pattinson et al. 2016).

Multipath refers to the phenomenon of a signal reaching an antenna via two or more paths (Misra & Enge 2010) due to reflections from nearby structures or from the ground, as illustrated in Figure 4.2.

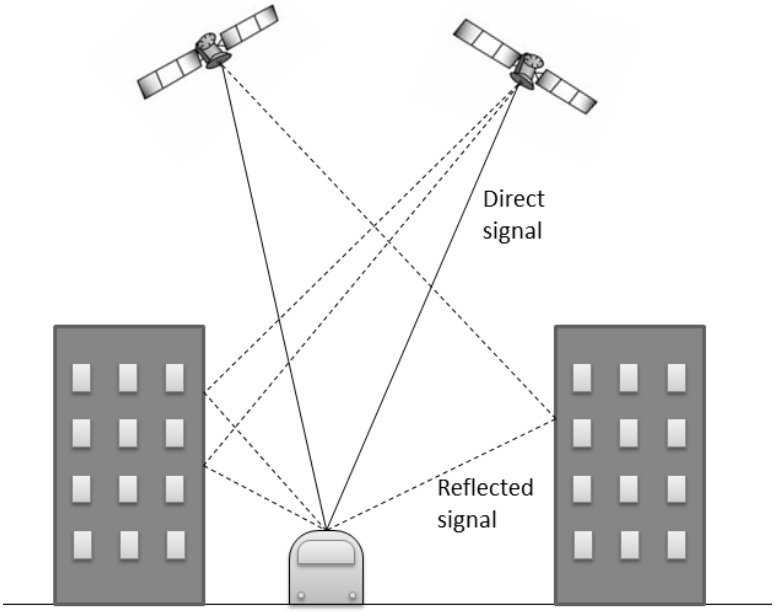


Figure 4.2 - Description of the multipath effect

Multipath affects both the pseudorange and carrier phase measurements but with a different magnitude. In theory, it can reach a value of 1.5 times the wavelength, representing 450 m for the pseudorange multipath. However, in practise values of up to 150 m have been observed. The carrier phase multipath is two orders of magnitude smaller, with a maximum value equal to a quarter of a wavelength. On L1, this corresponds to less than 5 cm. Further details on multipath are given in section 4.2.

4.1.2.3 Receiver errors

The receiver errors include the receiver clock error, the hardware delay and the receiver noise. The clock used in the user receiver is usually low cost and unstable, which means it is not synchronised with

either the satellite clock or the GNSS time. The difference between the receiver clock and the GNSS time needs to be estimated in order to compute the user's position.

The hardware delay is due to the different elements of the receiver, such as the antenna, filters, processing unit, that delay the signal. However, this delay only affects the receiver positioning accuracy by a few millimetres (Hegarty 2013).

The receiver noise covers the noise introduced by the different elements of the receiver such as the antenna, amplifiers and cables. For most receivers, the noise induced error is limited to the decimetre level for the pseudorange measurement and to the millimetre level for the carrier phase (Misra & Enge 2010).

4.1.3 Error mitigation

In order to improve positioning accuracy, several techniques have been developed to either eliminate or mitigate the effects of the different errors on GNSS measurement. They are reviewed in this section.

4.1.3.1 Satellite errors mitigation

The navigation message broadcast by the satellite contains orbit parameters and satellite clock error correction model parameters. However, higher quality ephemeris and clock products are available online from the International GPS Service (IGS). Different types of products are available depending on the application needs, such as final ephemerides with an accuracy of 2.5 cm for post-processing or ultra-rapid ephemerides with an accuracy of 5 cm for real-time applications (International GPS Service 2009).

4.1.3.2 Propagation error mitigation

Ionospheric delay is one of the main sources of measurement error, which can reach up to 90 metres for satellites at low elevation. In order to mitigate this error, a model based on an empirical approach is used by the single frequency user, named the Klobuchar model (Sanz Subirana et al. 2013). GNSS satellites broadcast the parameters of this model, which enables the ionospheric error to be reduced by about 50%.

If dual frequency measurements are available, they can be combined to form an ionospheric-free observable. Indeed, using equation (4.6), it can be seen that the ionospheric delays on L1 and L2 are such that:

$$f_1^2 I_1 - f_2^2 I_2 = 0 \tag{4.9}$$

where I_1 and I_2 are the ionospheric delays on L1 and L2 respectively.

By combining the measurements on L1 and L2 as follows, the ionosphere-free pseudorange and carrier phase measurements can be obtained.

$$R_C = \frac{(f_1^2 R_1 - f_2^2 R_2)}{f_1^2 - f_2^2} \tag{4.10}$$

$$\Phi_C = \frac{(f_1^2 \Phi_1 - f_2^2 \Phi_2)}{f_1^2 - f_2^2} \tag{4.11}$$

By replacing (4.2) and (4.5) in (4.10) and (4.11), it can be seen that the ionosphere delay is cancelled out.

$$R_C = \rho + c(\delta t_{rcv} - \delta t^{sat}) + Tr + \varepsilon_C \tag{4.12}$$

$$\Phi_C = \rho + c(\delta t_{rcv} - \delta t^{sat}) - Tr + B_C + \epsilon_C \tag{4.13}$$

where B_C is given by

$$B_C = \lambda_N \left(N_1 + \left(\frac{\lambda_W}{\lambda_2} \right) N_W \right) \tag{4.14}$$

and

$$\lambda_W = \frac{c}{f_1 - f_2}, \quad \lambda_N = \frac{c}{f_1 + f_2} \quad and \quad N_W = N_1 - N_2$$

The first order ionospheric error, which accounts for 99% of the total ionospheric error, can be removed by this ionosphere-free combination. However, this combination magnifies the measurement noise.

The tropospheric delay does not depend on the signal frequency and thus it cannot be removed by a combination of dual-frequency measurements. Instead, models and mapping functions are used. Two groups of models are used as alternatives: geodetic models (e.g. Saastamoinen model, Hopfield model) and navigation models (e.g. University of New Brunswick (UNB) model). Geodetic models are more accurate but more complex as they require surface meteorological data such as the air pressure, temperature and water vapour pressure. These values are then used to estimate the two components of the tropospheric delay: the hydrostatic component and the wet component (Saastamoinen 1972). These two delays are then projected in the direction of the satellite using a mapping function (Niell 1996). While navigation models are less accurate than geodetic models, they do not require meteorological data. Instead they use look-up tables containing the annual mean and amplitude for temperature, pressure and water vapour pressure, as a function of the receiver's latitude and height (Leandro et al. 2006). Geodetic models can reduce the tropospheric delay to 3 cm, while navigation models can achieve an accuracy between 6 and 20 cm (Mendes & Langley 1999).

Multipath and interference can be mitigated through antenna design, receiver design or measurement processing techniques. These techniques are reviewed in further details in section 4.2.2. Despite the

existing mitigation techniques, multipath remains a highly unpredictable source of errors, especially for dynamic users (i.e. moving receiver), where the environment around the antenna is constantly changing.

4.1.3.3 Receiver error mitigation

As the clock error needs to be estimated in order to compute the user position, a fourth measurement is required to perform position estimation. The receiver noise and hardware delay can be limited by design and usually they do not exceed 0.5 m for the code measurement and 2 mm for the phase measurement (Misra & Enge 2010).

4.1.3.4 Differential GNSS

In addition to the mitigation techniques presented above, differential GNSS can also be used to reduce the error in the measurements. The principal of differential GNSS is explained in 3.1.2.2, the technique uses a receiver at a known fixed location to compute corrections that are subsequently sent to the user receiver.

Given the satellite ephemeris and the known location of the reference station, the geometric distance $\rho_r^{\ i}$ between the reference station and the i^{th} satellite can be computed according to equation (4.3).

$$\rho_r^i = \sqrt{(x_i - x_r)^2 + (y_i - y_r)^2 + (z_i - z_r)^2} \tag{4.15}$$

The pseudorange measurement at the reference station can also be written as:

$$R_r^i = \rho_r^i + c(\delta t_r - c\delta t^i) + I_r + Tr_r + M_r + \varepsilon_r \tag{4.16}$$

By differencing the computed geometric range $\rho_r^{\ i}$ with the pseudorange measurement $R_r^{\ i}$, the pseudorange code correction can be computed, as given by equation (4.17).

$$\Delta R_r^i = \rho_r^i - R_r^i = -c(\delta t_r - \delta t^i) - I_r - Tr_r - M_r - \varepsilon_r \tag{4.17}$$

This correction is then broadcast to the user receiver where it is added to the user receiver pseudorange $R_u^{\ i}$ for the same i^{th} satellite.

$$R_u^i + \Delta R_r^i = \rho_u^i + c(\delta t_u - \delta t^i) + I_u + Tr_u + M_u + \varepsilon_u - c(\delta t_r - \delta t^i) - I_r - Tr_r - M_r - \varepsilon_r \tag{4.18}$$

The satellite clock error δt^i is cancelled out by the correction. Unless local atmospheric disturbances are present, the atmospheric errors are highly correlated over short distances (a few tens of kilometres). Thus, the ionospheric and tropospheric delays at the reference station and at the user receiver can be considered equal. However, non-correlated errors, such as the noise and multipath, are added up.

The corrected pseudorange at the user receiver can then be rewritten as:

$$R^i_{u,corr} = \rho^i_u + c\delta t_{um} + \varepsilon_{um} \tag{4.19}$$

where δt_{um} is the difference in user and reference station clock offset δt_u-δt_m. δt_m may be estimated at the reference station and removed from the pseudorange corrections (Hegarty 2013). ε_{um} contain the residual errors, including the uncorrelated multipath and noise errors.

A summary of the performance of the differential GNSS corrections and of the mitigation techniques reviewed in the previous subsection is presented in Table 4.1.

Table 4.1 - A summary of the errors in GPS measurements (estimated based on a 10 km baseline and a signal latency of 10 sec), adapted from (Misra & Enge 2010)

Source	Potential error size	Error mitigation and residual error
Satellite clock model	Clock modelling error: 2 m (RMS)	IGS Correction: product dependent DGPS: 0 m
Satellite ephemeris prediction	Component of the ephemeris prediction error along the line of sight: 2 m (RMS)	IGS Correction: product dependent DGPS: 0.1 m
Ionospheric delay	Delay in zenith direction = 2-10 m depending on user latitude, time of the day and solar activity Delay for a satellite at lower elevation angle can be up to three times larger	Single frequency receiver using broadcast model (Klobuchar): 1-5 m Dual-frequency receiver: 1m (RMS) DGPS: 0.2 m (RMS)
Tropospheric delay	Delay on zenith direction at sea level = 2.3-2.5 m, lower at higher altitudes Delay for a satellite at lower elevation can be up to 10 times larger	Models based on average meteorological conditions: 0.1-1 m DGPS: 0.2 m (RMS) plus altitude effect
Multipath	In an open-sky environment Code: 0.5-1 m Carrier: 0.5-1 cm	Uncorrelated between antennas Mitigation through antenna design and sitting, receiver design, and carrier-smoothing of code measurements
Receiver noise	Code: 0.25-0.5 m (RMS) Carrier phase: 1-2 mm (RMS)	Uncorrelated between receivers Mitigation through receiver design

In summary, most GNSS errors can be mitigated either by combining measurements or by using additional information such as models and corrections. However, errors highly dependent on the receiver's direct surrounding, such as interference and multipath, pose a greater challenge. The following section reviews the impact of the different error on the position estimation process.

4.1.4 Position estimation and positioning error

4.1.4.1 Least squares estimation

The pseudorange and carrier phase measurements produced by the user receiver can subsequently be used to estimate its position. To simplify the notation, only the pseudorange code measurements are considered in the following example. The pseudorange measurement from the i^{th} satellite R_i is composed of the geometric range ρ_i between the receiver and the satellite, and of an error term ε_i, including the errors reviewed in section 4.1.2

$$R_i = \rho_i + \varepsilon_i \tag{4.20}$$

In order to determine the three-dimensional position of the user (x_u, y_u, z_u) and the time offset between the receiver and the GNSS time t_u, a minimum of four pseudoranges is required.

After linearization of the equations (see Appendix I) around the user position and time (i.e. estimation of the receiver position and time from an initial known position), the equations can be written in a matrix version.

$$\Delta\rho = H\,\Delta x + \varepsilon \tag{4.21}$$

with

$$\Delta\rho = \begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \vdots \\ \Delta\rho_N \end{bmatrix},\quad H = \begin{bmatrix} a_{x1} & a_{y1} & a_{z1} & 1 \\ a_{x2} & a_{y2} & a_{z2} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{xN} & a_{yN} & a_{zN} & 1 \end{bmatrix},\quad \Delta x = \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ -\mathrm{c}\Delta t_u \end{bmatrix} \; and \; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

where N is the number of visible satellites, $\Delta\rho$ is a vector containing the terms of the linearized mathematical model, H is the design matrix, Δx contains the parameters that are being estimated, ε is the vector of residuals i.e. the amount by which the observations differed the model and $a_i = (ax_i, ay_i, az_i)$ are unit vectors pointing from the linearization point to the location of the i^{th} satellite.

Traditional least squares estimation assumes that the measurement noise ε is a white Gaussian noise such that:

$$\mathrm{E}[\varepsilon] = 0 \quad and \quad Cov(\varepsilon) = \sigma_\rho^2 I_N \tag{4.22}$$

This implies that measurements are uncorrelated from epoch to epoch and that they are equally noisy.

The least square solution to the user position and time error Δx is expressed as follows:

$$\Delta x = (H^T H)^{-1} H^T \Delta\rho \tag{4.23}$$

Details about the least square estimation can be found in Appendix II.

Assuming that each pseudorange is independent and has a standard deviation σ_ρ, the covariance of the user position error can be computed as:

$$cov(\Delta x) = \sigma_\rho^2 (H^T H)^{-1} \tag{4.24}$$

Equation (4.24) indicates that the positioning solution is affected by two factors: the error in the measurements through the term σ_ρ^2 and the geometry of the satellite, through the matrix H.

4.1.4.2 Kalman filter

The receiver position can also be estimated using an iterative algorithm, the Kalman filter (KF). This filter estimates the position at the time $n+1$ combining the measurements available at time $n+1$ and the prediction of the position based on the estimate at time n.

The Kalman filter model assumes that the state of a system x_n evolves according to:

$$x_{n+1} = F_n x_n + G_n u_n + w_n \tag{4.25}$$

where

- x_n is the state vector containing the terms of interest of the system
- u_n is the command vector, containing any control inputs
- F_n is the state transition matrix which applies the effect of each state at time n on the state at time $n+1$
- G_n is the command matrix, which applies the effect of each control input on the state vector
- w_n is the process noise vector. The process noise is assumed to be drawn from a zero mean multivariate normal distribution with a covariance given by the matrix Q_n

The measurements are also modelled as:

$$y_n = H_n x_n + v_n \tag{4.26}$$

where

- y_n is the vector of measurements
- H_n is the transformation matrix that maps the state vector parameters into the measurement domain
- v_n is the measurement noise vector containing the noise for each observation in the measurement vector. The measurement noise is also assumed to be zero mean Gaussian white noise with a covariance matrix R_n.

The estimator used in the Kalman filter is the minimum mean squared error, defined as:

$$\hat{x}_{n+1|n+1} = E[x_{n+1}|y_0, \dots, y_{n+1}] \tag{4.27}$$

At each epoch, the covariance matrix of the estimation error $\Sigma_{n+1|n+1}$ is computed:

$$\Sigma_{n+1|n+1} = Cov(x_{n+1}|y_0, \dots, y_n) \tag{4.28}$$

The Kalman filter consists of a set of equations which enables to compute at each epoch $\hat{x}_{n+1|n+1}$ and $\Sigma_{n+1|n+1}$.

The first equation expresses the a priori value of the state vector, as a function of the previous estimate:

$$\hat{x}_{n+1|n} = F_n \hat{x}_{n|n} + G_n u_n \tag{4.29}$$

$\hat{x}_{n+1|n}$ is the best prediction of the state vector x_{n+1}, knowing the measurement ($y_0, \dots, y_n$).

In a similar way, the a priori covariance is computed:

$$\Sigma_{n+1|n} = F_n \Sigma_{n|n} F_n^T + Q_n \tag{4.30}$$

A vector representing the a priori measurement error, the innovation vector $\hat{I}_{n+1|n}$ is computed as:

$$\hat{I}_{n+1|n} = y_{n+1} - H_{n+1} \hat{x}_{n+1|n} \tag{4.31}$$

The innovation can be used to control the quality of the Kalman filter, for instance by performing hypothesis testing on its mean value.

The covariance of the innovation vector can be computed as:

$$V_{n+1|n} = H_{n+1} \Sigma_{n+1|n} H_{n+1}^T + R_{n+1} \tag{4.32}$$

The Kalman gain is then computed as:

$$K_{n+1} = \Sigma_{n+1|n} H_{n+1}^T \left[H_{n+1} \Sigma_{n+1|n} H_{n+1}^T + R_{n+1} \right]^{-1} \tag{4.33}$$

The a posteriori estimate of the state vector and its covariance are then computed:

$$\hat{x}_{n+1|n+1} = \hat{x}_{n+1|n} + K_{n+1} [y_{n+1} - H_{n+1} \hat{x}_{n+1|n}] \tag{4.34}$$

$$\Sigma_{n+1|n+1} = \Sigma_{n+1|n} - K_{n+1} H_{n+1} \Sigma_{n+1|n} \tag{4.35}$$

The residuals, representing the a posteriori error can be computed:

$$Res_{n+1|n+1} = y_{n+1} - H_{n+1} \hat{x}_{n+1|n+1} \tag{4.36}$$

The covariance of the residuals is:

$$\mathrm{P}_{n+1|n+1} = H_{n+1} \Sigma_{n+1|n+1} H_{n+1}^T + R_{n+1} \tag{4.37}$$

Figure 4.3 summarises the principle of the Kalman filter and shows how the measurements and the predictions are combined.

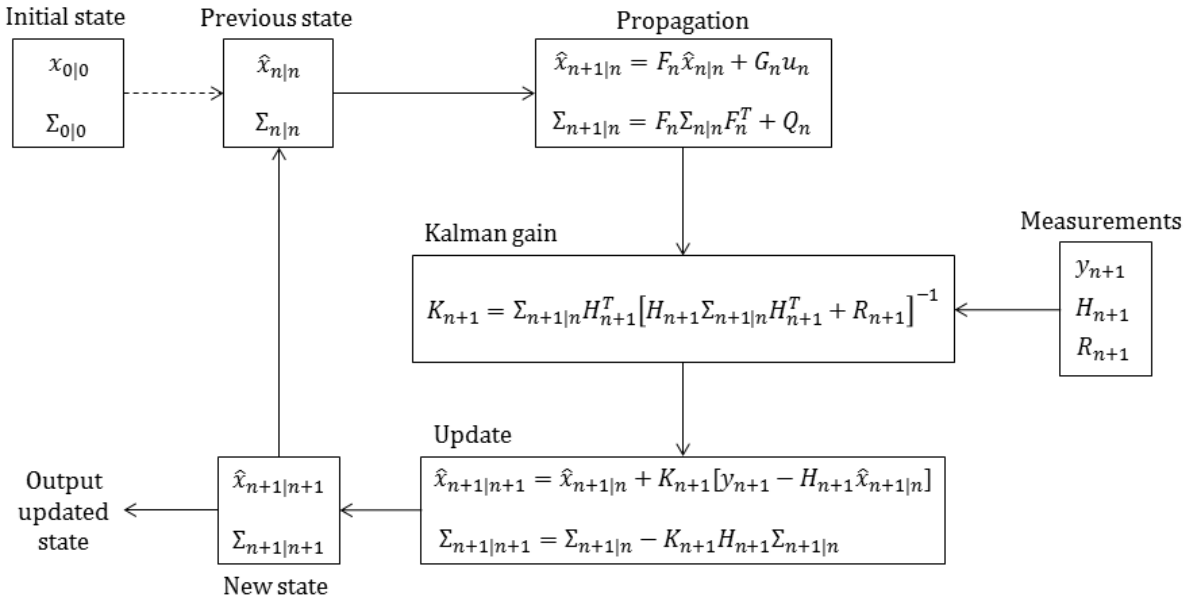


Figure 4.3 - Kalman filter data flow, adapted from (Groves 2013)

To use the Kalman filter, the system model given in equation (4.24) and the measurement model given in equation (4.25) have to be linear. If at least one of these models is non-linear, the extended Kalman filter (EKF) is used. It consists in the linearization of the models using partial derivatives evaluated at the estimated values of the state vector. The linearization is only used to compute the gain in equation (4.26). The estimated states are propagated using the nonlinear model and the predicted measurements are calculated using the nonlinear measurement model.

The difference between the least squares estimation presented in section 4.1.4.1 and the Kalman filter lies in the determination of the a priori information used in the estimation process. Least squares estimation relies on the previous estimate (Δx in equation (4.23) represents the position variation between the previous estimate and the current estimated position). In the Kalman filter, the a priori information is determined from the previous estimate and from the assumed system model (i.e. the knowledge of how the position changes with time). If the model is correct, the Kalman filter provides a more accurate solution. However, if it is incorrect, the filter still tries to fit results to the assumed model, providing a biased estimate of the position. Thus the Kalman filter should only be used when the system model can be correctly determined.

4.1.4.3 Neural networks

Traditional estimation techniques presented in sections 4.1.4.1 and 4.1.4.2 use iterative methods and linearization to attempt to solve the pseudorange equations. This linearization introduces a small error as the higher order terms are neglected. As opposed to the least square estimation or Kalman filtering, neural networks can be used to estimate the receiver position without linearization. Neural networks do

not require detailed knowledge of the sensors, such as the stochastic models of sensor random errors and a priori covariance information. They learn the relationship between the inputs and their respective errors during the training process and is able to account for non-linear errors.

Neural networks mimic the way the brain processes data and learn. As illustrated in Figure 4.4, they are composed of at least three main layers: the input layer, the hidden layer(s) and the output layer.

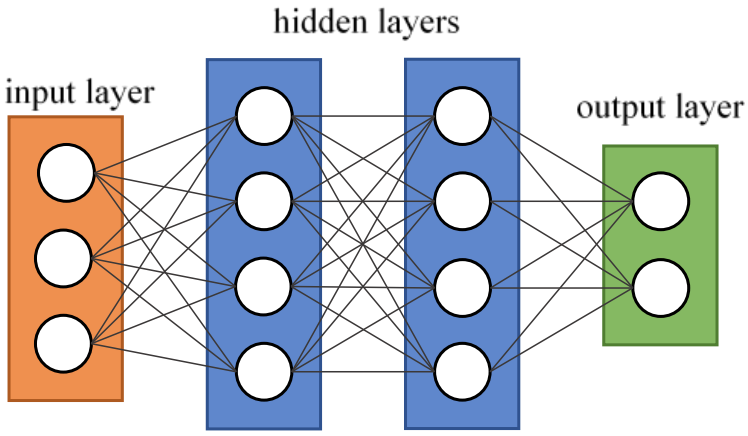


Figure 4.4 – A two-layer neural network

Each layer is composed of nodes, called neurones, between which the computation load is distributed. The interconnections between the nodes are given weight and, the higher the weight, the stronger the connection between the two neurones. These weights form the memory of the network and their values are set during a training phase. During this phase, the error between the neural network output and the known desired output is used to update the network weight. To ensure the generalisation of the neural network solution the training data set should be diverse and contain a large variety of scenario. In the case of navigation for example, data collected in diverse environment (e.g. open sky, urban canyon) should be used. The training data is processed several times until the error between the output and the expected solution reached a user set threshold or after a given number of epochs. The complexity of the neural network depends on the number of neurones needed. The larger this number, the greater the ability to predict the errors but the higher the computation complexity and the longer the training time.

Neural networks have been used in navigation to estimate the user position using GPS only (Chansarkar 2000) or using a combination of GPS and INS (Chiang & Chang 2010), enabling the development of an integrated low cost MEMS (micro-electromechanical sensor) and GPS system, 80% more accurate on average (RMS error) than classical integration schemes. Neural networks have also been implemented to mitigate the effect of multipath at both the receiver level and the observable level (Vigneau et al. 2006), leading to a reduction of the code tracking error by 50% and of the carrier tracking error by 40%. At the observable level, the pseudorange measurement error was reduced by up to 40%.

Neural networks present several advantages over traditional methods such as the least square estimation or Kalman filtering. For instance, they do not require statistical models of the sensors and they can implicitly detect non-linear relationships between variables (Tu 1996). However, they remain "black boxes" and cannot explicitly identify causal relationship. This makes it difficult to quantify the level of confidence of the solution and to perform integrity monitoring. The design of the neural network is also challenging as an insufficient number of nodes may cause the network to fail to converge, while too many nodes may lead to overfitting and poor predictive performance (Tu 1996).

Given these limitations and as GNSS measurement models are well established, neural networks are not further investigated in this thesis and the traditional least square estimation is used to compute the receiver position.

4.1.4.4 Satellite geometry

The matrix D is defined such that $D = (H^T H)^{-1}$ to reflect the geometry of the visible satellites. It is used to compute the dilution of precision (DOP) parameters, quantifying the repartition of the satellites.

$$D = (H^T H)^{-1} \equiv \begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{bmatrix} \tag{4.38}$$

The most general parameter is the geometric dilution of precision (GDOP), which represents the amplification of the standard deviation of the measurement errors due to the satellite geometry, as seen in equation (4.38).

$$GDOP = \sqrt{D_{11} + D_{22} + D_{33} + D_{44}} \tag{4.39}$$

Several dilutions of precision parameters are commonly used as outlined in equations (4.40): the position dilution of precision (PDOP), the horizontal dilution of precision (HDOP), the vertical dilution of precision (VDOP) and the time dilution of precision (TDOP).

$$\begin{aligned} PDOP &= \sqrt{D_{11} + D_{22} + D_{33}} \\ HDOP &= \sqrt{D_{11} + D_{22}} \\ VDOP &= \sqrt{D_{33}} \\ TDOP &= \sqrt{D_{44}}/c \end{aligned} \tag{4.40}$$

As shown in Figure 4.5, the relative position of the satellites can affect the position solution error. The dotted lines in Figure 4.5 represent the ranging measurement error bounds and the shaded areas are the associated position solution error bounds. Despite the ranging measurement error bounds being equal in both cases, the difference in satellite geometry leads to different error areas for the positioning

solution, and thus to different DOP values. Satellites that are close together produce high DOP values, typically over 5 while satellites that are further apart produce lower values of DOP, typically below 2.

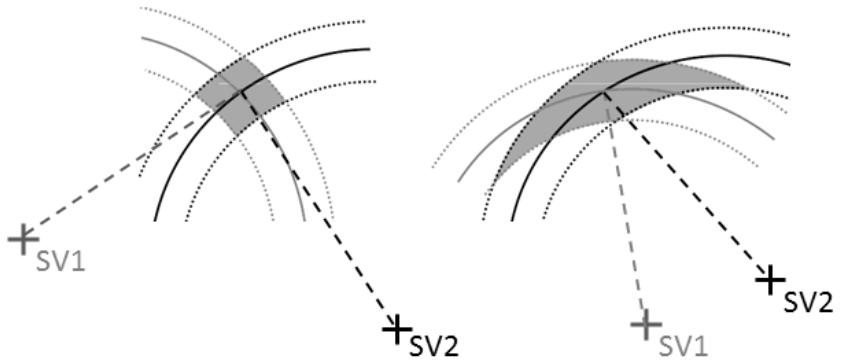


Figure 4.5- Relative geometry and dilution of precision: geometry with a low DOP (left), geometry with the high DOP (right), adapted from (Kouwenhoven 2011)

4.1.4.5 User equivalent range error

According to equation (4.24), the positioning error is affected not only by the satellite geometry but also by the pseudorange error σ_ρ. As seen in subsection 4.1.2, different elements contribute to the pseudorange error. The user equivalent range error (UERE) was developed to characterise the impact of each error source on the position accuracy and as such it represents the average range error standard deviation associated with each error source. Typical values of UERE can be found in Table 4.2.

Table 4.2 - Typical User Equivalent Range Error budget for single frequency GNSS, dual frequency GNSS and differential GNSS at one sigma, adapted from (Hegarty 2013; Misra & Enge 2010)

Error source	Single Frequency	Dual frequency	Differential GNSS
Ephemeris	0.8 m	0.8 m	0.1 m
Satellite clock	1.0 m	1.0 m	0.0 m
Ionosphere	7.0 m	0.1 m	0.2 m
Troposphere	0.2 m	0.2 m	0.2 m
Multipath	0.3 m	0.3 m	0.3 m
Receiver noise	0.1 m	0.1 m	0.1 m
Total	7.1 m	1.3 m	0.4 m

Based on equation (4.24), the positioning error can finally be estimated as the product of the UERE by the dilution of precision. For example, when using a single frequency receiver in a stand-alone mode,

if a VDOP of 1.7 and an HDOP of 1.0 are considered, then the expected error for the vertical position error at one sigma is 12.2 m and the horizontal position error is 7.4 m.

4.1.5 Conclusion

As seen in this section, the accuracy of the position estimated by using GNSS signals depends both on the satellites' geometry and on the different error sources that can affect the signals. In both these aspects, the railway environment poses considerable challenges. The large number of obstacles along the track (e.g. buildings, trees or bridges) limits satellite visibility, which can lead to poor satellite geometry. These obstacles also reflect GNSS signals, creating multipath that increases the measurement noise and decreases position accuracy, integrity, continuity and availability. In order to improve positioning performance in the railway domain, this thesis focuses on the mitigation of the multipath effect.

4.2 Multipath error and mitigation

GNSS modernization and augmentation have enabled the mitigation of numerous sources of error, leaving multipath as a significant contributor to the positioning error. This section discusses multipath and its effects, and the different mitigation techniques available, focusing on weighting methods.

4.2.1 Multipath model

Multipath signals are created by the reflection of satellite signals on objects surrounding the signal trajectory. Reflected multipath signals are divided into two groups: specular multipath and diffuse multipath. Specular multipath signals arise from reflections on large smooth surfaces and these reflections follow Descartes' laws. Diffuse multipath signals are created by reflections on scarred surface creating scattered signals that usually have a lower power than specular multipath. Although perfect specular reflections are rare in real life, most of the multipath analysis is based on this model (Moradi 2014). However, it has been shown, in practice, that such an assumption could be used to develop accurate propagation prediction models and improve positioning accuracy, for example in dense urban environment (Gowdayyanadoddi et al. 2015).

4.2.1.1 Geometric description

In terms of geometry, a single specular multipath can be presented in two dimensions as illustrated in Figure 4.6.

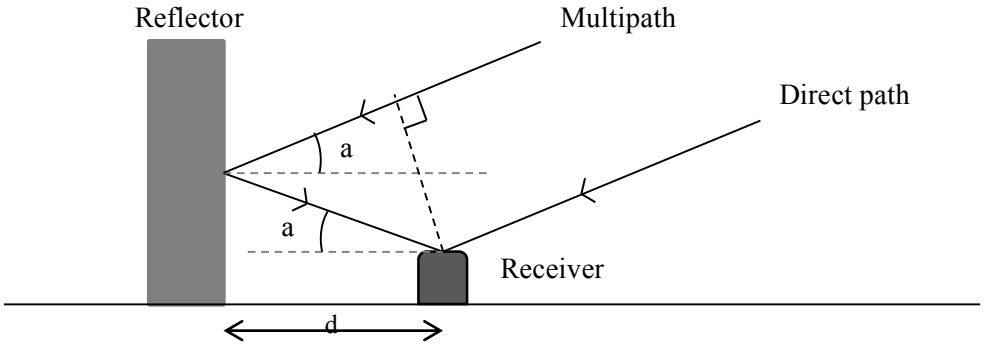


Figure 4.6 – Geometry of a single specular multipath, adapted from (Stolagiewicz 2009)

In this case, the extra distance travelled by the reflected signal can be computed as:

$$\delta_d = 2\, d \cos\,(a) \tag{4.41}$$

The equivalent phase shift in radians is given by:

$$\theta = \frac{2\pi\delta_d}{\lambda} \tag{4.42}$$

where λ is the wavelength of the signal.

Using a phasor diagram, the total carrier phase multipath can be expressed as (Axelrad et al. 1996):

$$\Phi_m = \tan^{-1}\left(\frac{\alpha \sin(\theta)}{1 + \alpha \cos(\theta)}\right) \tag{4.43}$$

Where α is the damping factor, i.e. the ratio between the amplitude of the reflected signal and the amplitude of the direct signal and θ is the phase shift as computed in equation (4.42).

These parameters can be estimated for static receivers using the repeatability of the GNSS constellation, as the satellite-receiver geometry repeats itself on an approximately daily basis. However, if the fixed geometry (static receiver) or the repeatability (observation over several days) cannot be exploited, as is the case for railway operations, then the estimation of the multipath parameters becomes considerably more difficult (Bisnath & Langley 2001).

4.2.1.2 Multipath impact of the GNSS receiver

The direct GNSS signal *s(t)* can be represented as a complex envelope of the transmitted signal *x(t)*.

$$s(t) = \alpha_0\, x(t - \tau_0) e^{-j\phi_0} e^{j2\pi f_c (t-\tau_0)} \tag{4.44}$$

where τ_0 is the signal propagation time between the receiver and the satellite, f_c is the carrier frequency, α_0 is the received amplitude of the signal and ϕ_0 is the received carrier phase.

If a signal is reflected by obstacles before reaching the antenna, N copies of $s(t)$ with different propagation times, carrier phases and amplitude are superimposed on the direct signal. The signal at the antenna can be represented by:

$$\mathrm{r(t)} = \alpha_0 e^{-j\phi_0} x(t-\tau_0) e^{-j2\pi f_c \tau_0} + \sum_{n=1}^{N} \alpha_n e^{-j\phi_n} x(t-\tau_n) e^{j2\pi f_n t} \qquad (4.45)$$

where α_n the received amplitudes of the multipath, τ_n the propagation delay of the multipath, ϕ_n the received carrier phases and f_n the received frequencies. Due to the satellite's motion, the parameters in equation (4.45) vary as a function of time.

The incoming signal at the receiver antenna is a combination of the direct signal and the multipath. In order to retrieve the information contained in the GNSS navigation message (e.g. time, ephemeris), the receiver needs to synchronise with the incoming signal. To do so, this signal is correlated with locally generated signal replicas. These replicas are offset by a certain value, defined in chips (similar to bits). The output of the correlation is then used to estimate the synchronisation error.

The combination of the direct signal and the multipath creates a distortion in the correlation function. If multipath is present, the correlation function becomes the sum of the direct signal correlation function and the reflected signal correlation function (see Figure 4.7). If the direct signal and the reflected signals are in phase, the interference is constructive and generates a positive ranging error. If the signals are out of phase, the interference is destructive and the ranging error generated is negative.

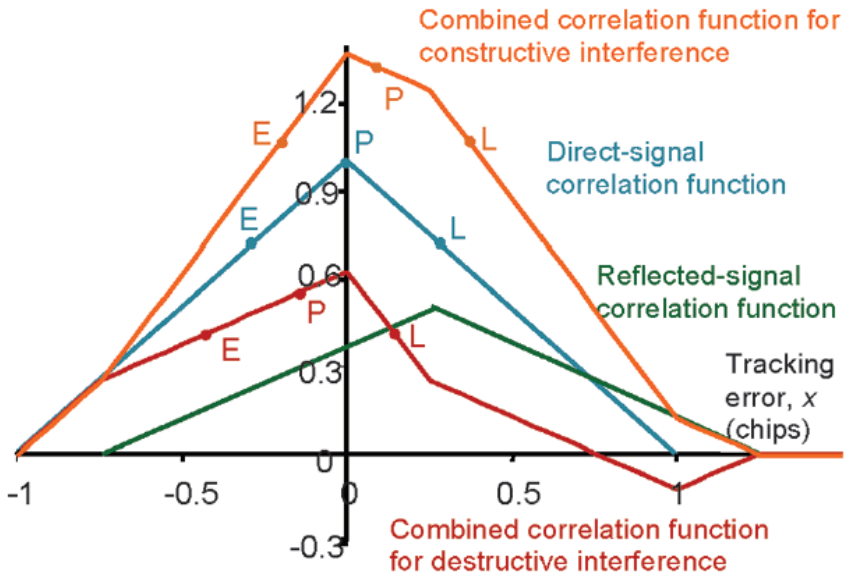


Figure 4.7 - Correlation functions of a signal subject to constructive interference (orange) and destructive interference (red), the direct signal being represented in blue and the reflected signal is green, from (P. Groves 2013)

If the multipath delay is large, for instance, more than a chip, the peak of the correlation function is not distorted and has little impact on the performance (Kaplan & Hegarty 2006). However, multipath signals created by nearby objects have short delays and distort the correlation function, introducing error in the pseudorange and carrier phase measurements.

The maximum pseudorange code measurement error that can be created by a multipath is half of a ranging code chip. This represents 150 meters for the GPS C/A code. The maximum carrier phase tracking error due to multipath is a quarter of a wavelength, which is approximately 5 cm on L1. This thesis focuses on mitigating the effect of code multipath for a dynamic receiver and the following sub-section presents a measurement combination that can be used to estimate the code multipath error.

4.2.1.3 Code multipath estimation

Braasch (1994) observed that a linear combination of the GNSS observables, the code minus carrier, can be used to produce a biased estimate of the pseudorange multipath. To form this novel observable, equations (4.2) and (4.5) are first rewritten to include the contribution of multipath errors

$$R_i = \rho + c(\delta t_u - \delta t^s) + I_i + Tr + M_i + \varepsilon_i \tag{4.46}$$

$$\Phi_i = \rho + c(\delta t_u - \delta t^s) - I_i + Tr + \lambda_i N_i + m_i + \epsilon_i \tag{4.47}$$

where the index i indicates the frequency, M is the code multipath error and m is the carrier multipath error.

The code minus carrier combination MP can then be formed as:

$$MP_i = R_i - \Phi_i \tag{4.48}$$

Substituting equations (4.46) and (4.47), the code minus carrier observable on L1 can be written as:

$$R_1 - \Phi_1 = 2\,I_1 - \lambda_1 N_1 + M_1 - m_1 + \varepsilon_1 - \epsilon_1 \tag{4.49}$$

The effect of the geometry (i.e. the range), the receiver and satellite clock errors and the tropospheric error are cancelled out in this differencing. As the carrier phase multipath and the carrier noise are approximately two orders of magnitude smaller than the code multipath and noise, their effect can be neglected. The code minus carrier observable can be rewritten as:

$$R_1 - \Phi_1 = 2\,I_1 + M_1 + \varepsilon_1 + b_1 \tag{4.50}$$

where b_1 is the bias due to the unknown integer ambiguity. This bias b_1 can be removed by averaging the code minus carrier observable. However, if the receiver is static, the averaging time has to be long enough to allow for low-frequency variations of the multipath. For a dynamic receiver, multipath estimates are more likely to have a Gaussian white noise characteristic, due to the constant change of the receiver environment, and the averaging process can be used to remove the bias created by the integer ambiguity (Stolagiewicz 2009). The novel code pseudorange observable can be written as:

$$MP_1 = R_1 - \Phi_1 - mean(R_1 - \Phi_1) \tag{4.51}$$

It can be seen in equation (4.50) that the ionospheric delay impact on the code minus carrier observable is twice as large as on the code or carrier measurement. Thus, to get an accurate estimate of the code multipath, the ionospheric delay needs to be estimated.

This can be done by either using models (see section 4.1.3) or dual-frequency measurements as the ionospheric delay can be estimated by differencing the carrier phase measurements at both frequencies:

$$\Phi_1 - \Phi_2 = -I_1 + I_2 + b' + m' + \epsilon' \tag{4.52}$$

Where *b'* is the bias due to the integer ambiguities, *m'* is the differenced carrier multipath and ϵ' is the noise.

The ionospheric delays difference can be rewritten using equation (4.9)

$$I_2 - I_1 = \frac{f_1^2}{f_2^2} I_1 - I_1 = \frac{f_1^2 - f_2^2}{f_2^2} I_1 \tag{4.53}$$

This difference of carrier phase can then be used to remove the ionospheric delay I_1 from the code minus carrier observable.

$$MP_1 = R_1 - \Phi_1 - \frac{2f_2^2}{f_1^2 - f_2^2}(\Phi_1 - \Phi_2) \tag{4.54}$$

$$MP_1 = M_1 + b'_1 + \varepsilon'_1 \tag{4.55}$$

The observable can then be averaged to remove the bias due to the integer ambiguities.

A similar estimate can be computed to estimate the code multipath on L_2.

$$MP_2 = R_2 - \Phi_2 - \frac{2f_1^2}{f_1^2 - f_2^2}(\Phi_1 - \Phi_2) \tag{4.56}$$

The code minus carrier observation is then used to estimate the multipath error on the pseudorange measurement. This indicator is used in chapter 5, where multipath signals are characterised in the railway environment (section 5.2.2.3).

4.2.2 Multipath mitigation

Different approaches have been developed to mitigate the effect of multipath on the position estimation process, including antenna-based techniques, signal processing techniques and measurement processing techniques. This section reviews these techniques and is based on (Moradi 2014).

4.2.2.1 Antenna-based techniques

Antenna based multipath mitigation techniques use specifically designed antennas that limit the reception of multipath signals.

The direct signals transmitted by the satellite have a right-hand circular polarisation (RHCP). However, reflections change the polarisation of the signal so that multipath signals have a left-hand circular polarisation (LHCP) or mixed RHCP-LHCP. Highly sensitive RHCP antennas are designed to associate a higher gain to RHCP signals and thus limit the reception of LHCP reflected signals. However, the antenna still receives some LHCP signals and is ineffective for signals arriving with low elevation angles.

Choke ring antennae, made of concentric rings, are designed to remove the antenna gain at low elevation angles. However, such antennae suppress multipath and direct signals alike, and as they are large and heavy, it may be impractical to install them on the train roof.

An antenna array can also be used to mitigate multipath. Composed of multiple antennae, the array can be configured to achieve a gain in a specific direction and suppress signals coming from other directions. The use of multiple antennas means that the array is quite large and additional processing is required to control the gain pattern.

In summary, antennas can be used to limit the reception of multipath though their effect is limited, particularly at low elevation angles, and their large size can be impractical.

4.2.2.2 Signal processing techniques

Receiver based multipath mitigation techniques make the tracking and acquisition stages of the receiver more robust to signals affected by multipath, for instance by narrowing the correlator. As seen in Figure 4.8, the narrower the correlator, the smaller the pseudorange error.

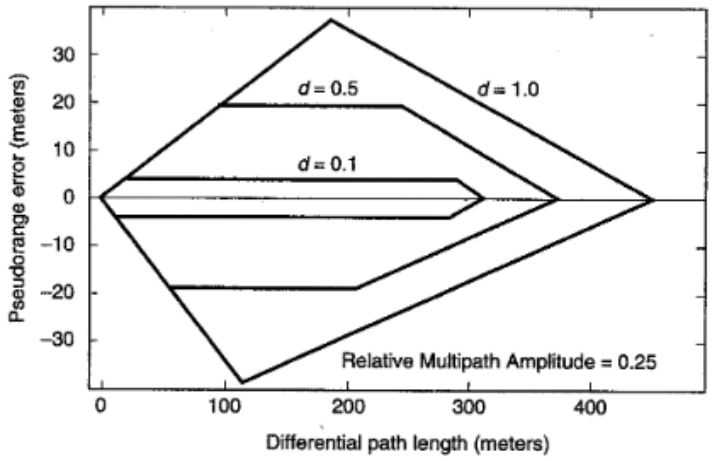


Figure 4.8 - Code multipath error envelope for different spacing of the correlator d, from (Misra & Enge, 2011, p423)

Multipath can also be estimated using a multipath estimation delay lock loop (MEDLL). However, these techniques cannot mitigate the short delay multipath errors which are the main source of multipath error.

4.2.2.3 Measurement processing techniques

Post receiver based multipath mitigation techniques use measurements, satellite and signal information to limit the impact of multipath on the position estimation process. This can be undertaken through the smoothing of measurements as well as their weighting.

Weighting methods associate different weights to different measurements, enabling higher quality measurements to contribute more to the final solution than measurements corrupted by multipath. The quality of the measurements can be assessed on the basis of different criteria such as the satellite elevation or the carrier to noise power density ratio. Weighting methods are reviewed in detail in section 4.3.

Smoothing techniques are based on the fact that carrier phase measurements are less sensitive to multipath signals than code measurements. As explained in section 4.2.1.2, the carrier phase multipath error is limited to a quarter of the wavelength, while the code multipath error can reach several tens of meters. Carrier-smoothing filters, also known as Hatch filters, use the more precise carrier phase measurements to reduce the noise and multipath in the code measurements (Sanz Subirana et al. 2013).

Considering the code and phase measurements for one satellite at epoch *k*, the smoothed code $\hat{R}$ can be computed as:

$$\widehat{R}(k) = \frac{1}{n}R(k) + \frac{n-1}{n}\left[\widehat{R}(k-1) + \left(\Phi(s,k) - \Phi(k-1)\right)\right] \tag{4.57}$$

Where $n = k$ as long as $k < N$, and $n = N$ when $k \geq N$. The constant N relates to the bandwidth of the filter, larger values of N giving more smoothing.

The subtraction of consecutive carrier phase measurements can be rewritten as:

$$\Phi(k) - \Phi(k-1) = \Delta\rho + \Delta Tr + \Delta I + \epsilon' \cong \Delta\rho + \epsilon' \tag{4.58}$$

Where $\Delta\rho$ is the range variation between epoch *k-1* and epoch *k*, and ΔTr and ΔI are the tropospheric and ionospheric error variations.

The residuals of temporally correlated errors such as the atmospheric delays (ΔI and ΔTr) can be neglected. As long as no cycle slip occurs, integer ambiguity is cancelled out by the differencing. This implies that the filter has to be reinitialised every time a cycle slip is detected. The phase variation can then be used to estimate the pseudorange variation and to compute the smoothed code measurement.

The transient duration (i.e. the time preceding the filter steady state) is approximately *2N*. The noise variance at the output of the filter is given by (Lo Presti & Visintin 2015) as:

$$\sigma_{\hat{R}}^2 = \frac{\sigma_R^2}{2N-1} + \sigma_\phi^2 \tag{4.59}$$

Where σ_R is the variance of the code noise and σ_ϕ is the variance of the carrier noise.

Therefore, as the value of *N* increases, the output noise variance decreases but the duration of the transient increases. A compromise must be found on the choice of *N* and Lo Presti & Visintin (2015) suggest that the selection of *N* is such that:

$$N = \frac{\sigma_R^2}{2\,\sigma_\phi^2} \tag{4.60}$$

This value leads to a noise variance at the output of the filter equal to $2\sigma_\phi^2$ while the transient duration is kept reasonably short.

Compared to antenna-based and signal-processing techniques, multipath mitigation using measurements processing does not require any additional hardware or modification to the receiver. This thesis focuses on measurement processing methods, particularly on measurement weighting techniques, which are presented in detail in the following section.

4.3 Weighting methods

GNSS-based positioning estimates the values of unknown parameters (the position and time offset of the receiver) based on measurements. In order to optimise the estimation process, the desired information has to be extracted from the measurements in an optimal way (Tiberius 1999). This can be done by the use of the least squares principle introduced in section 4.1.4.1. To apply this principle, both the measurement equations and their noise model must be specified. The measurement equations are specified as in equation (4.21), after correcting for the different deterministic error sources presented in section 4.1.3.

Measurement noise is mathematically formalised as a stochastic quantity represented by its covariance matrix. This matrix describes both the precision and the correlation of the measurements. For the least-square estimation to be optimal, the covariance matrix should contain the actual variance of each measurement, but these values are unknown a priori. Traditional least squares estimation assumes that measurements are equally noisy, assigning the same constant value to each measurement noise variance (equation (4.22)) and that the noise is uncorrelated between measurements. Weighting methods model the stochastic properties of each measurement in order to define a more realistic model for the noise covariance matrix and improve the accuracy of the position estimated.

4.3.1 Weighted least squares estimate

This section presents the weighted least squares estimation principle and relates the weights used in the weighting methods to the measurement noise. Consider the model presented in equation (4.21).

$$\Delta\rho = H\Delta x + \varepsilon \tag{4.61}$$

where ε is the error vector.

The weighted least square estimation assumes that ε is normally distributed but with a non-constant covariance matrix, i.e.

$$E[\varepsilon] = 0 \tag{4.62}$$

$$Cov(\varepsilon) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_N^2 \end{pmatrix} \tag{4.63}$$

where N is the number of measurements and σ_i^2 is the noise variance of the i^{th} measurement.

A weight w_i defined as the reciprocal of the standard error squared is assigned to each observation, forming the weight matrix W:

$$W = \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1/\sigma_N^2 \end{pmatrix} = \begin{pmatrix} \mathrm{w}_1 & 0 & \dots & 0 \\ 0 & \mathrm{w}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mathrm{w}_\mathrm{N} \end{pmatrix} \tag{4.64}$$

The weights are defined so that the larger the noise variance, the lower the weight and thereby adjust the contribution of each measurement to the final solution based on their noise level.

The weighted least square estimation leads to the following values of the parameter vector and residuals (see Appendix II):

$$\widehat{\Delta x} = (H^T W H)^{-1} H^T W\, \Delta\rho \tag{4.65}$$

$$\hat{\varepsilon} = H\widehat{\Delta x} - \Delta\rho \tag{4.66}$$

The a-posteriori covariance matrix of the parameters and residuals are also computed to assess the precision of the estimated position:

$$Cov\left(\widehat{\Delta x}\right) = (H^T W H)^{-1} \tag{4.67}$$

$$Cov(\hat{\varepsilon}) = \mathrm{W}^{-1} - H\left(\mathrm{H^T W H}\right)^{-1}\mathrm{H^T} \tag{4.68}$$

The precision of the weighted least squared (WLS) estimation process can be assessed by examining the residuals. It is expected that high precision observations (i.e. a small value σ_i) will have small residuals and vice versa. To verify that it is the case, the unit variance σ_0^2 is computed:

$$\hat{\sigma}_0^2 = \frac{\hat{\varepsilon}^T W \hat{\varepsilon}}{N - M} \tag{4.69}$$

Where *N* is the number of observation and *M* the number of parameters being estimated. (*N-M*) is called the degrees of freedom.

The unit variance should be such that $E[\sigma_0^2] = 1$. Therefore, if the unit variance estimated a-posteriori $\hat{\sigma}_0^2$ is significantly different from unity, the variance of the observations are on average underestimated by a factor $\hat{\sigma}_0^2$ [Cross 1994] and a new weight matrix W' is defined.

$$\mathrm{W}' = \frac{1}{\hat{\sigma}_0^2} W \tag{4.70}$$

This new formulation of the weight matrix changes neither the value of the estimated parameters nor the residuals (equations (4.65) and (4.66)), but the covariance matrices are now given by equations (4.71) and (4.72).

$$Cov(\widehat{\Delta x}) = \hat{\sigma}_0^2 (H^T W H)^{-1} \tag{4.71}$$

$$Cov(\hat{\varepsilon}) = \hat{\sigma}_0^2 \left[\mathrm{W}^{-1} - H\left(\mathrm{H^T W H}\right)^{-1} \mathrm{H^T} \right] \tag{4.72}$$

Testing the proximity of $\hat{\sigma}_0^2$ to unity only provides a factor by which the a priori covariance matrix is on average incorrect. This implies that the weight matrix can be wrong by a constant factor and that the precision of the position estimate is only related to the relative weighting between observations. Also, a covariance matrix that is incorrect by a certain factor does not impact the estimated parameters and residuals (see equations (4.65) and (4.66)) but rather impacts their covariance (see equations (4.67) and (4.68)), making the statistic $\hat{\sigma}_0^2$ important in assessing the quality of the estimation.

To test if $\hat{\sigma}_0^2$ is significantly different from unity, a ratio test is set up as follows.

$$\begin{aligned} H_0: &\quad \sigma_0^2 = 1 \\ H_1: &\quad \sigma_0^2 \neq 1 \end{aligned}$$

The statistic F is computed such as:

$$\begin{aligned} &if\ \hat{\sigma}_0^2 > 1, \quad F = \hat{\sigma}_0^2 \ and\ \nu_1 = N - M, \ \ \nu_2 = \infty \\ &if\ \ \hat{\sigma}_0^2 < 1, \quad F = \frac{1}{\hat{\sigma}_0^2}, \ \ and\ \nu_1 = \infty, \nu_2 = N - M \end{aligned} \tag{4.73}$$

The statistic follows a F-distribution with v_1 and v_2 degrees of freedom. A level of significance is selected i.e. the probability that the null hypothesis is rejected while it is true, and a F-distribution table is used to determine the percentile for the corresponding values of v_1 and v_2. The computed *F* value is then compared to the table percentile.

If F is smaller than the table percentile, the null hypothesis H_0 is accepted and the a priori weight matrix is assumed correct. The covariance matrices are computed according to equations (4.67) and (4.68).

If F is larger than the table percentile, the null hypothesis H_0 is rejected i.e. $\hat{\sigma}_0^2$ is significantly different from unity, meaning that the a priori weight matrix needs to be corrected by a factor $1/\hat{\sigma}_0^2$. The covariance matrices are computed using equations (4.71) and (4.72).

As seen in this section, it is necessary to specify the a priori covariance matrix of the observations in order to determine the weights used to compute the least square estimates of the parameters and the residuals. As the observation covariance is unknown a priori, the choice of its value remains quite subjective. However, several techniques have been developed to estimate the observation covariance matrix and their review is presented in the following section.

4.3.2 Computation of the weights

As seen in equation (4.64), the weights are inversely proportional to the variance of the measurement error and thus reflect information on this measurement. A measurement with a small error variance is given a large weight, as it contains relatively more information than such a measurement with a large error variance.

The variance of each measurement σ_i^2 is unknown a priori. However, several analytical and empirical models have been developed to estimate its value. These models express the noise variance as a function of the satellite elevation angle with respect to the receiver antenna, the carrier to noise power density ratio C/N0 or the multipath error estimate.

The computation of the variance of each measurement requires two things: a mapping function that defines the relative value of the variance as a function of the quality indicator (e.g. the satellite elevation angle) and a nominal value of the standard deviation of the error. Traditionally, the nominal error standard deviation is defined for a satellite at the zenith and determined to have a value of between 20 to 30 cm for the pseudorange and between 1 and 3 mm for the carrier phase (Elsobeiey & El-Rabbany 2010; Weiser & Brunner 2000). A review of existing mapping functions is presented below.

4.3.2.1 Elevation based weighting techniques

Elevation based weighting techniques assume that satellites with higher elevation angle (i.e. closer to the zenith) are less affected by errors, and thus should have a higher contribution to the final positioning solution. This assumption is based on the fact that signals coming from satellites around the zenith are less likely to encounter obstacles and, are also less affected by the atmosphere as their signal travels through a shorter layer of the atmosphere, as illustrated in Figure 4.9.

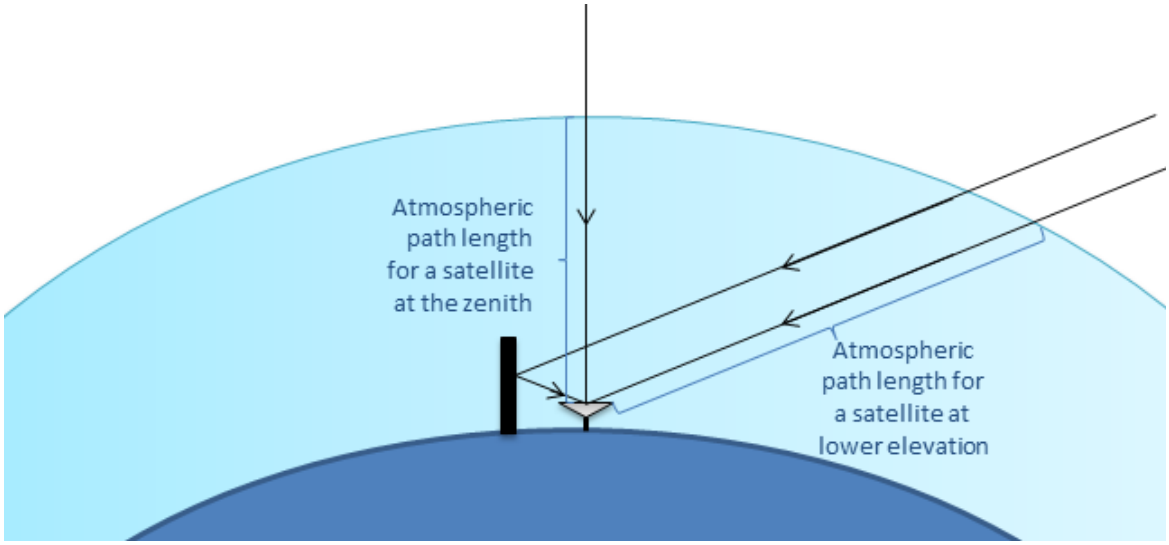


Figure 4.9 - The effect of satellite elevation on the atmospheric error

Conventional elevation based weights assume that the error standard deviation is proportional to the sine value of the elevation, as defined by (Hartinger & Brunner 1999):

$$\sigma_i = \frac{\sigma_0}{\sin(elevation)} \tag{4.74}$$

where σ_0 represents the nominal standard deviation of the measurements.

The issue with expression (4.74) is that small values of the elevation angle lead to excessively large standard deviation. In order to limit the value of the standard deviation, a cut-off angle can be used, e.g., all satellites with an elevation angle below 10° can be discarded from the positioning solution.

Another technique, based on the obliquity factor as computed by Black & Eisner (1984), can be used to estimate the variance of the measurement error as a function of the elevation. This technique removes the need to use a cut-off angle as constant values are added to the formula, limiting the divergence of the weight values for low elevation angles.

$$\sigma_i = \sigma_0 * \frac{1.001}{\sqrt{0.002001 + \sin(elevation_i)^2}} \tag{4.75}$$

This formula is used not only to model the effect of the satellite elevation on the tropospheric error (Black & Eisner 1984) but also to weight measurements.

The pseudorange error standard deviations have also been estimated using stochastic models based on empirical observations. Elsobeiey & El-Rabbany (2010) determined that the best fit model for the

pseudorange error standard deviation based on satellite elevation is an exponential decay function such as equation (4.76):

$$\sigma_{\rho_i} = (A + B\, e^{-C*elevation_i}) \tag{4.76}$$

where A, B and C are constant whose values depend on the observable and on the frequency.

Different values of the coefficients were determined for the different GPS observables, they are summarised in Table 4.3.

Table 4.3 – Values of the coefficient in the stochastic elevation model developed by Elsobeiey & El-Rabbany (2010)

Measurement	A	B	C
C1 or P1	0.1773	0.9232	0.0945
P2	0.1983	0.7220	0.1183
C2	0.2188	1.4488	0.1655

Figure 4.10 represents the standard deviation of the code measurement error as a function of the satellite elevation for the different methods described in (4.74), (4.75), and (4.76). As an example, a cut-off angle of 5° was used for the elevation model.

As expected, for the three weighting methods analysed the noise standard deviation increases as the elevation angle decreases. However, the rate of increase is different for each method, especially for angles below 50°. While the standard deviations are comparable for high elevations (>50°), a steep increase is observed for the elevation method and the Black and Eisner method for angle below 50°. These two methods assign very similar standard deviation values but the elevation method excludes satellites with an elevation lower than the cut-off angle. The stochastic elevation models consistently provide lower standard deviation values than both the elevation and Black and Eisner methods. This effect is even more pronounced for angles below 50°.

However, the noise standard deviation associated with each measurement in the stochastic elevation model is different from the values observed by (Al-Fanek et al. 2007), shown in Figure 4.11. According to these observations, the C/A code at L1 (C1) is expected to have a smaller standard deviation than the codeless pseudorange L2P (P2). The L2C code (C2) is expected to present a smaller standard deviation than C1 as it uses longer codes that present better autocorrelation and cross-correlation properties and thus should present better acquisition performance, especially for lower elevation angles.

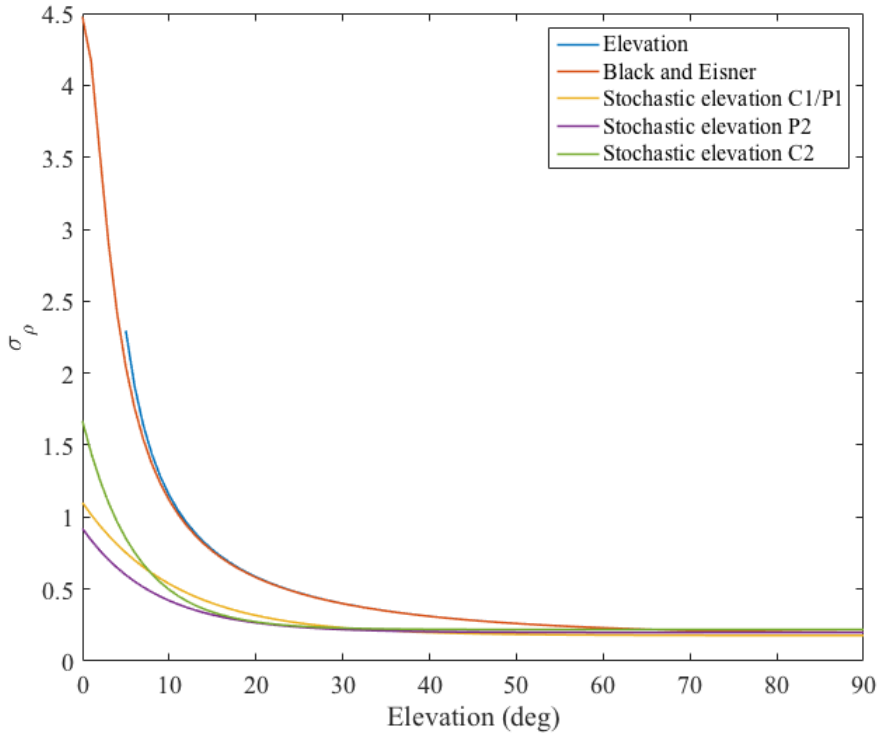


Figure 4.10 – Code measurement error standard deviation as a function of the satellite elevation, for $\sigma_{\rho 0}$= 0.20 m

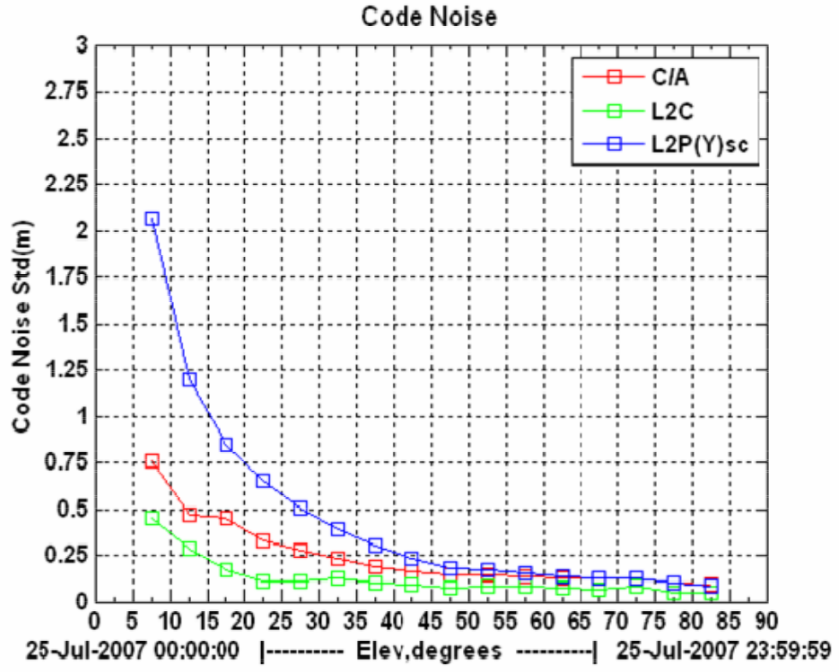


Figure 4.11 - Code noise standard deviation as a function of elevation angle, from (Al-Fanek et al. 2007)

In the stochastic elevation model formulated by Elsobeiey & El-Rabbany (2010), the P2 code consistently presents a lower noise standard deviation than both the C1 and C2 codes. The C2 code is assigned a lower standard deviation than the C1 code, except for elevation below 8°, while it is expected

that the C2 code correlation properties would improve the noise level at low elevation. Additional analysis of C1, C2, P1 and P2 signals are required to determine which model is more representative of the actual noise behaviour.

Both the elevation and the stochastic elevation weighting methods presented in this section are tested in Chapter 6, and their performance in terms of the positioning accuracy is assessed.

4.3.2.2 C/N0 based weighting techniques

Weighting techniques based on C/N0 relate the carrier to noise power density ratio to the quality of the signal. The C/N0 should provide a better reflection of the quality of the actual signal than the elevation as it is directly related to the signal received. It is computed at the receiver tracking stage as the ratio of the received signal power to the noise power and is commonly expressed in decibels. In the absence of interference, the C/N0 is expected to have a value higher than 45 dB-Hz (Groves 2013). The C/N0 is actually a measure of the carrier power, not of the code itself. However, as the code is carried by the carrier (i.e. the carrier wave is modulated to convey the code), it is assumed that the C/N0 can also be related to the power of the code measurement.

For the C/N0 based weighting techniques, signals with a higher C/N0 values are supposed to be of higher quality (i.e. signal level significantly higher than noise level) than those with smaller C/N0 values. In comparison to elevation-based weighting techniques, C/N0 weighting allows the use of noisier data from satellites with very low elevation angles, that in turn improves the satellite geometry and thus increase the attainable accuracy. However, the weights need to be correctly assigned so that noisy measurements are sufficiently weighted down. The sigma-ε model, proposed by (Hartinger & Brunner 1999) and described in equation (4.77), has been shown to achieve a good balance between down-weighting the noisy data and improving the satellite geometry:

$$\sigma_i = \sigma_0 \sqrt{C * 10^{-\frac{C/N0_i}{10}}} \tag{4.77}$$

where C is a constant whose value was estimated using empirical data at $1.6.10^4$ (Hartinger & Brunner 1999) and is sometimes fixed at 10^5 (POINT n.d.).

Weiser & Brunner (2000) used empirical data to determine the carrier phase error standard deviation as a function of the signal to noise ratio in the extended sigma-ε model, presented in equation (4.78).

$$\sigma_{\Phi_i} = \sqrt{V + C * 10^{-\frac{C/N0_i}{10}}} \tag{4.78}$$

Where the constants C and V were estimated to be for L1, $C = 0.244$ and $V = 0$, and for L2, $C = 0.77.10^{-3}$ and $V = 0.88.10^{-6}$.

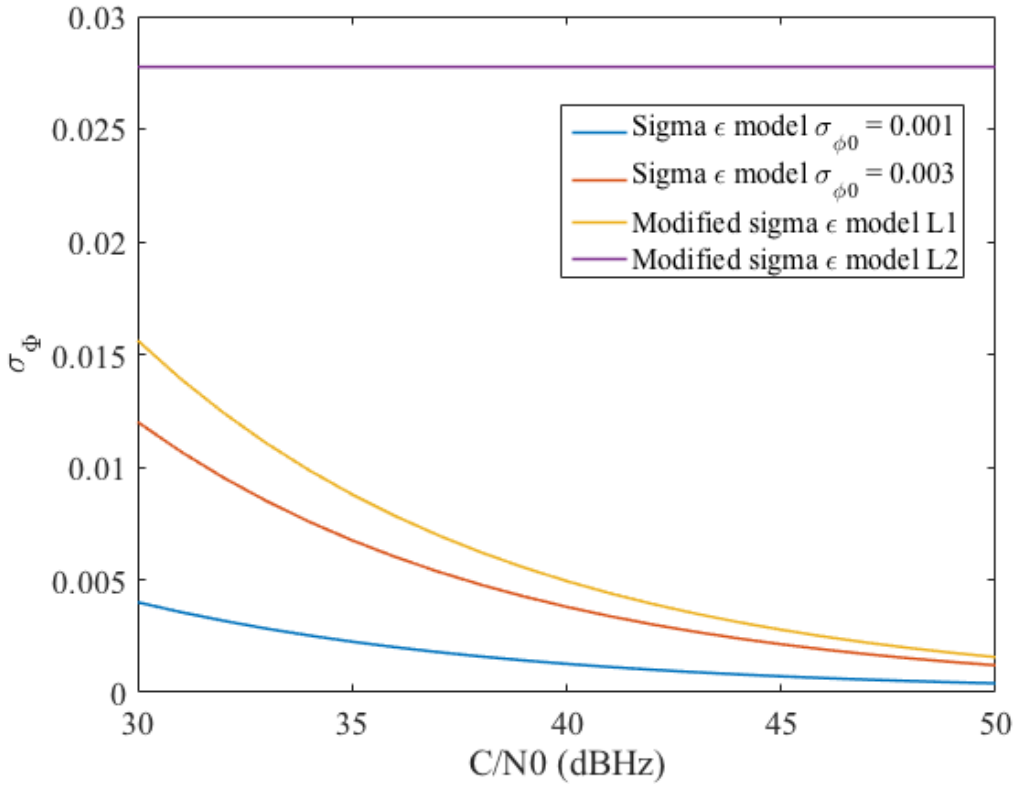


Figure 4.12 – Phase error standard deviation as a function of the carrier to noise power density ratio (C/N0)

Figure 4.12 represents the standard deviation of the carrier phase error as a function of the signal C/N0 for the sigma-ε model described by equation (4.77) and its modified version described by equation (4.78). As expected the noise standard deviation increases when the C/N0 decreases and therefore, signals with a smaller C/N0 value are assigned a larger noise variance. The modified sigma-ε model for the phase on L1 is comparable to the sigma-ε model with a nominal carrier phase standard deviation of 3 mm. However, in the case of the modified sigma-ε model for the phase on L2, the variation is of the order of 10^{-7} for a C/N0 varying from 50 dBHz to 30dBHz. Thus, this model assigns a significantly larger standard deviation to the phase measurements on L2, almost independently of the C/N0 value.

The performance of both the sigma-ε and the modified sigma-ε methods on the estimated position are tested in chapter 6.

4.3.2.3 Multipath based weighting

The elevation and C/N0 may not provide a good indication of the measurement quality (i.e. noise variance) in an environment with multipath. Instead, the code multipath observable proposed in equation (4.54) can be used to estimate the variance of the multipath error for each satellite (Bisnath & Langley 2001). However, no formal formulation of a multipath based noise variance is offered by

Bisnath & Langley (2001). A possible formulation of the multipath based weight is further discussed in section 5.2.2.4.

4.3.2.4 Elevation and C/N0 based weighting techniques

Some weighting techniques have combined several quality indicators such as the C/N0 and the satellite elevation angle. Marais et al. (2015) proposed to weight the measurements according to equation (4.79):

$$\sigma_i = k * \frac{\sqrt{10^{-\frac{C/N0}{10}}}}{\sin(elevation)} \tag{4.79}$$

Where k is a constant which was fixed to $k = 2$ after testing (Marais et al. 2015).

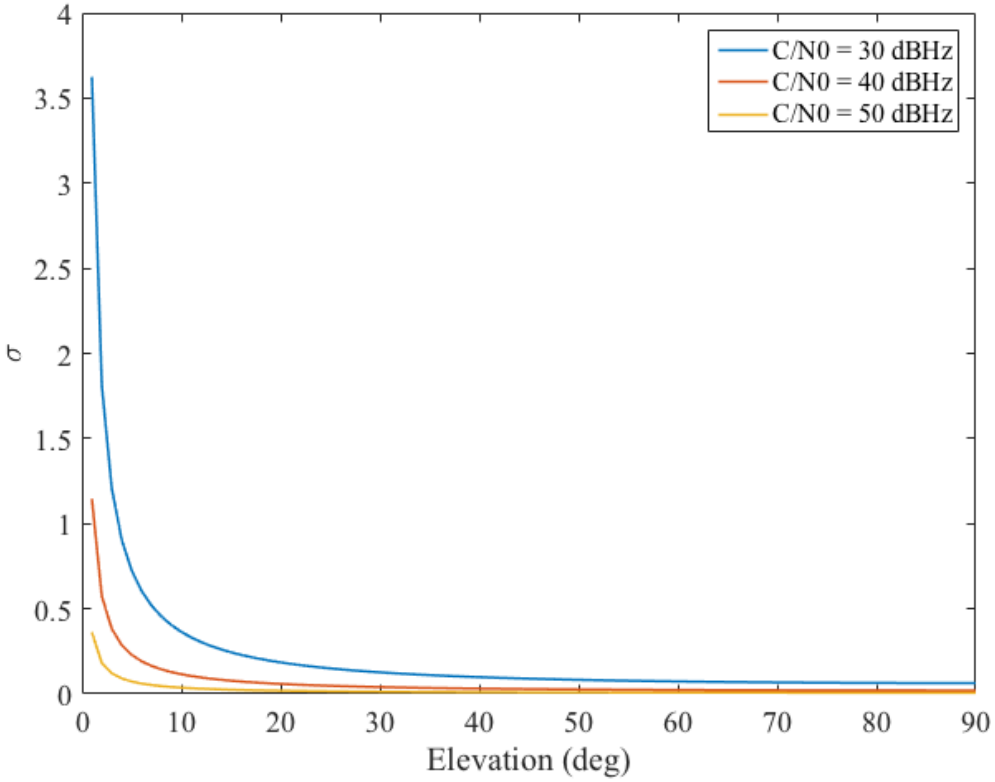


Figure 4.13 - Error variance as a function of the carrier to noise power density ratio (C/N0) and the satellite elevation

Figure 4.13 represents the measurement noise variance as a function of the satellite elevation for different values of C/N0, as defined in equation (4.79) (Marais et al. 2015). As expected, the variance decreases with the elevation angle and the lower the C/N0, the steeper the decrease. This model combines information from both the satellite elevation and the signal C/N0 and enables measurements from low elevation satellites but with a sufficient C/N0 to contribute more to the position estimation than measurements from higher elevation satellites but with a lower C/N0 value. For example, a satellite

with a 20° elevation and a C/N0 equal to 40 dBHz is assigned a smaller standard deviation than a satellite with a 40° elevation but a C/N0 of 30 dBHz.

Also, this combined model produces standard deviation values ranging from 0.006 m to 3.6 m but does not clearly differentiate between the code and the carrier phase standard deviation. Thus, the code standard deviation may be underestimated while the carrier standard deviation is overestimated.

4.3.3 Limitations

To achieve reliable positioning results, a realistic noise variance model has to be specified, which requires a good understanding of the quality of the measurements. The methods presented above rely on several assumptions and their limitations are discussed below.

The elevation based weighting has been proved to be a very approximate model (Bisnath & Langley 2001). The use of a cut-off angle in the traditional elevation based weighting method reduces the number of available satellite. Moreover, by giving more weights to satellites above the receiver, the geometric diversity of measurements (i.e. DOP) is reduced, and as seen in equation (4.24), this decreases the positioning accuracy. Also, the assumption that high elevation satellites are less likely to be affected by errors may not be valid in areas with high buildings where satellites with a high elevation may be subject to multipath.

Using static data, Satirapod & Wang (2000) proved that C/N0 values reflect a more realistic trend than the elevation. However, C/N0 values are approximate and sensitive to the antenna and the receiver characteristics. Different receivers and antennae produce different values of C/N0 under the same conditions (Satirapod & Wang 2000). This can be an issue when performing differential positioning though it is a good indication of the relative signal strengths between different satellites tracked by the same receiver (Collins & Stewart 1999).

Finally, the multipath observable provides a good indicator of the presence of multipath if all other sources of errors, such as ionospheric and tropospheric errors have already been mitigated.

In summary, each quality indicator has its limitations and they should be taken into account when designing the weight matrix.

4.4 Analysis

In order to assess the correctness of the relationships presented in section 4.3.2 between the satellite elevation angle, C/N0 and code multipath and noise, a preliminary analysis is carried out. The three

datasets presented in section 1.4 are used to test these relationships under different conditions (i.e. static, dynamic and high-speed dynamic).

4.4.1 Static case

First, a dataset collected by a static receiver is analysed, see section 1.4.1 for further details. The receiver was static for approximately eight hours and installed in a relatively open area, but the presence of trees and a sloping terrain below antenna level can cause signal reflections, especially for signals coming from satellites with a low elevation angle.

Figure 4.14 represents the value of C/N0 as a function of the satellite elevation angle. The mean and the standard deviation of the C/N0 values were computed for 5° beams.

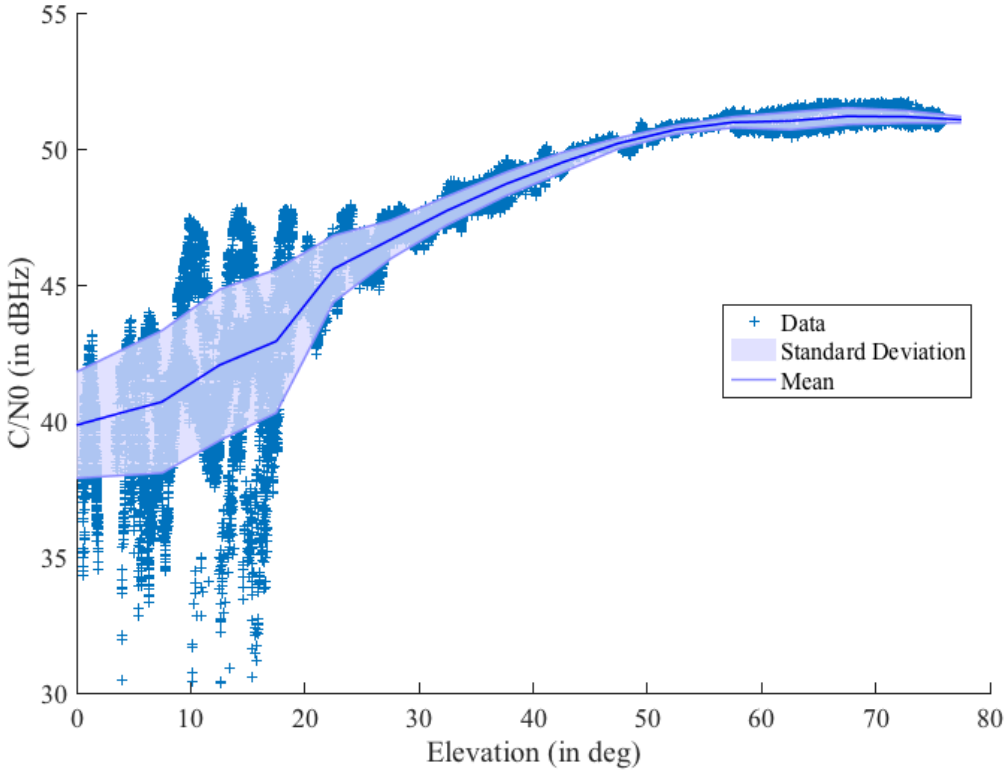


Figure 4.14- Carrier to noise power density ratio as a function of the satellite elevation

The dependence of the C/N0 on the satellite elevation can clearly be observed. For an elevation angle above 50°, the C/N0 is almost constant and equal to 50 dBHz. The standard deviation for high elevation angles is relatively small but its value increases as the elevation angles become smaller. Thus, the elevation-based weighting methods and the C/N0 based methods may lead to comparable results for high elevation satellites. However, for low elevation satellites, the C/N0 values present a larger variability. For example, for an elevation angle of 15°, the C/N0 can have values ranging from

approximately 30 to 47 dBHz. This variability may be due to the fact that signals coming from low elevation satellites are more prone to be reflected by the ground or by surrounding obstacles.

Thus, elevation based weighting is representative of the signal to noise ratio for elevation angles larger than 20°, in the case the receiver is static. However, it is not representative of the signal to noise ratio variability for low elevation angles. The use of a cut-off angle can limit this effect but, as discussed previously, it also reduces the satellite geometry diversity.

Figure 4.15 represents the code noise and multipath on L1 as a function of the satellite elevation angle. The code multipath appears to decrease as the satellite elevation increases. For elevation angles above 30°, the code multipath estimate is smaller than 0.2 m and the standard deviation of the multipath is relatively small. For angles below 30°, the code multipath has a high variability and reaches values of up to 0.6 m. As observed in Figure 4.14, the satellite elevation is representative of the code multipath error for elevation angles above 30°. However, the relationship between the code multipath and the elevation is limited for angles below 30° due to the large multipath variability.

Figure 4.16 represents the code noise and multipath on L1 as a function of the C/N0. The relation between both can clearly be seen: the code multipath and noise mean and standard deviation increase as the C/N0 value decreases. Thus it appears that the C/N0 provides an indication of the level of code multipath in the case the receiver is static.

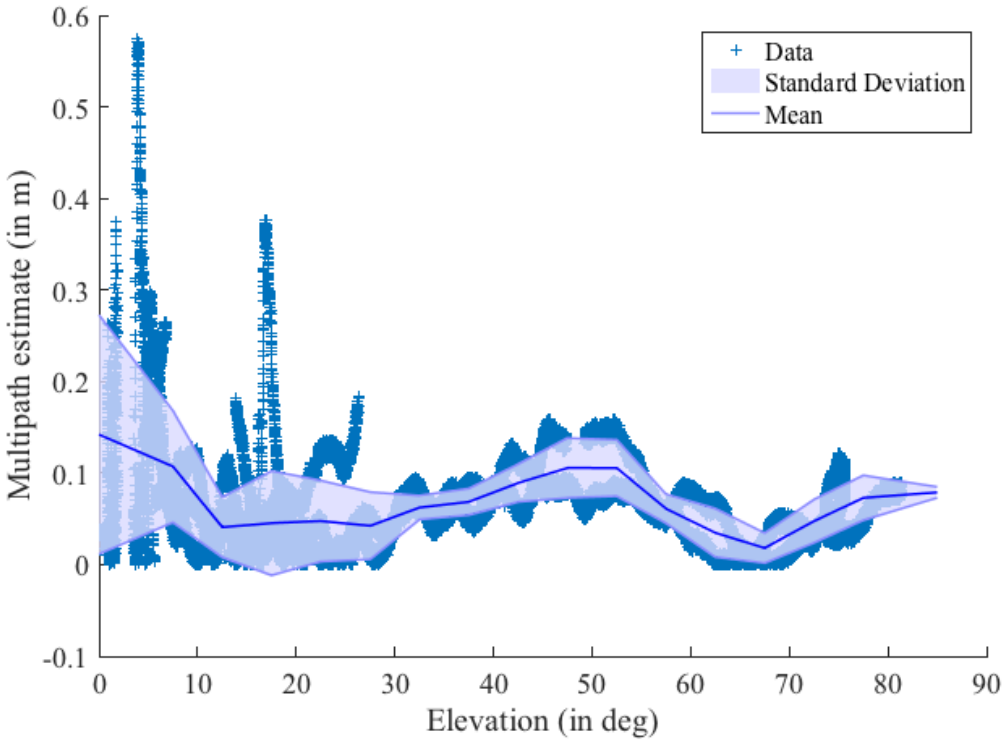


Figure 4.15 – C/A code noise and multipath on L1 as a function of the satellite elevation angle

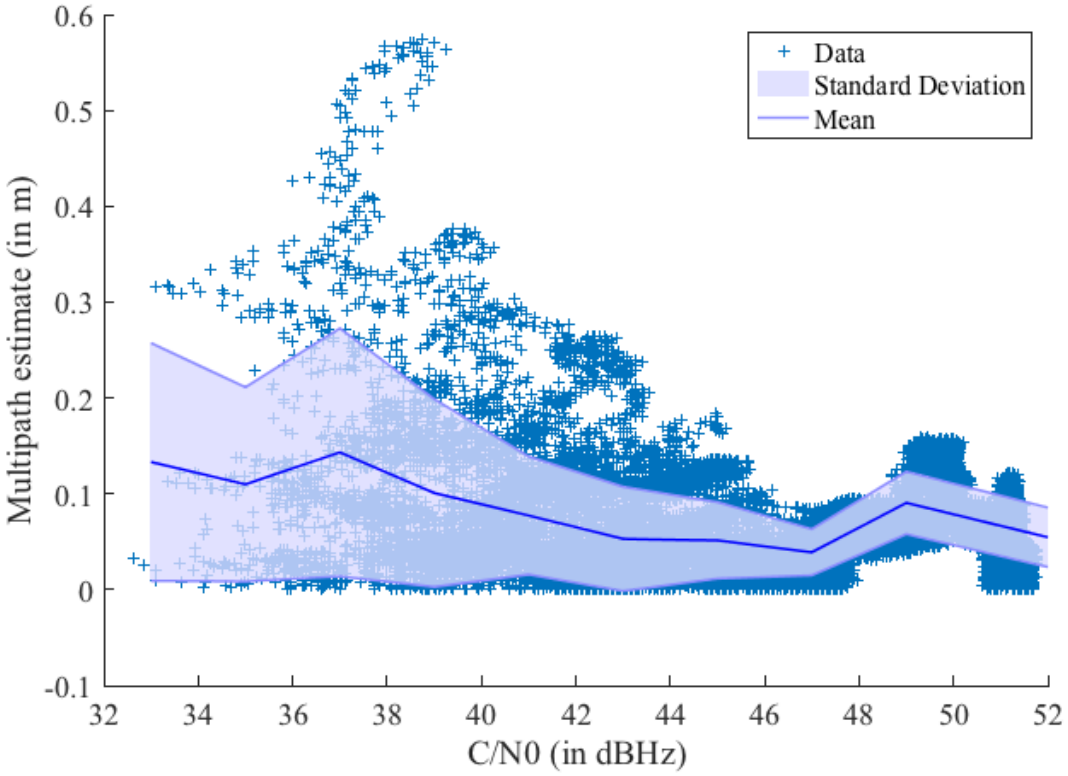


Figure 4.16 – C/A code noise and multipath on L1 as a function of the C/N0

4.4.2 High-speed train

The dataset analysed in the case of the high-speed train contained approximately 38,000 epochs, corresponding to over ten hours of data. It was collected on board a train travelling mostly through rural areas, with very few obstacles on each side of the track. A detailed description of this dataset can be found in section 1.4.2.

Figure 4.17 represents the carrier to noise ratio as a function of the satellite elevation angle. As for the static case, the C/N0 values tend to decrease with the elevation. However, the C/N0 presents a large variability, especially for angles below 35°, with C/N0 values between approximately 27 and 50 dBHz. Thus, in the dynamic case, the satellite elevation is not always representative of the signal to noise ratio, especially at a low elevation angle. However, it provides a good indication of the signal strength. Further analysis is required to assess the impact of both elevation-based and C/N0 based weighting in the case of a dynamic scenario. Such an analysis is presented in chapter 6.

Figure 4.18 represents the value of the code multipath estimate as a function of the satellite elevation. The mean value of the code multipath estimate remains relatively constant for all angles between 5° and 90°, with a value of approximately 0.4 m. However, the code multipath variability increases as the elevation angle decreases (i.e. the standard deviation becomes larger as the angle decreases). The trend modelled by the formulas in section 4.3.2.1 (elevation based weighting) as an exponential decay term

is not observed in Figure 4.18. Thus, the traditional elevation-based weighting techniques will systematically assign lower weights to measurements coming from low elevation satellites, even though these satellites may actually be affected by a small error. This could potentially limit the accuracy of the position but will, however, prevent large multipath errors (as observed below 40°) from reducing the positioning accuracy.

Figure 4.19 represents the value of the code multipath estimate as a function of the C/N0. According to the mean value and standard deviation, the code multipath estimate increases as the C/N0 decreases. However, high values of C/N0 are also associated with large code multipath estimates (e.g. 7 m code multipath for a C/N0 of 45 dBHz), limiting the correlation between the C/N0 value and the code multipath.

In conclusion, the relationships assumed by the weighting models presented in section 4.3 appear to be relatively correct in terms of trends: the code multipath decreases as the elevation angle increases and as the C/N0 decreases. However, in the railway environment, the code multipath estimate is more variable, limiting the applicability of the previously mentioned models.

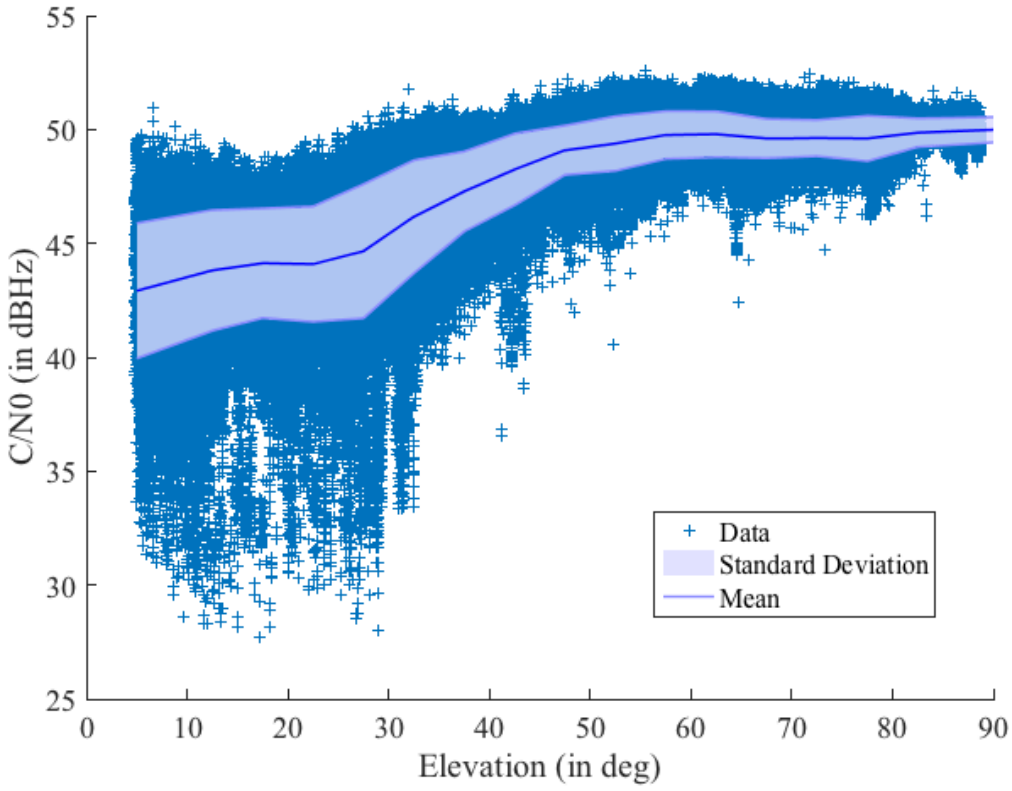


Figure 4.17- Carrier to noise power density ratio as function of the satellite elevation

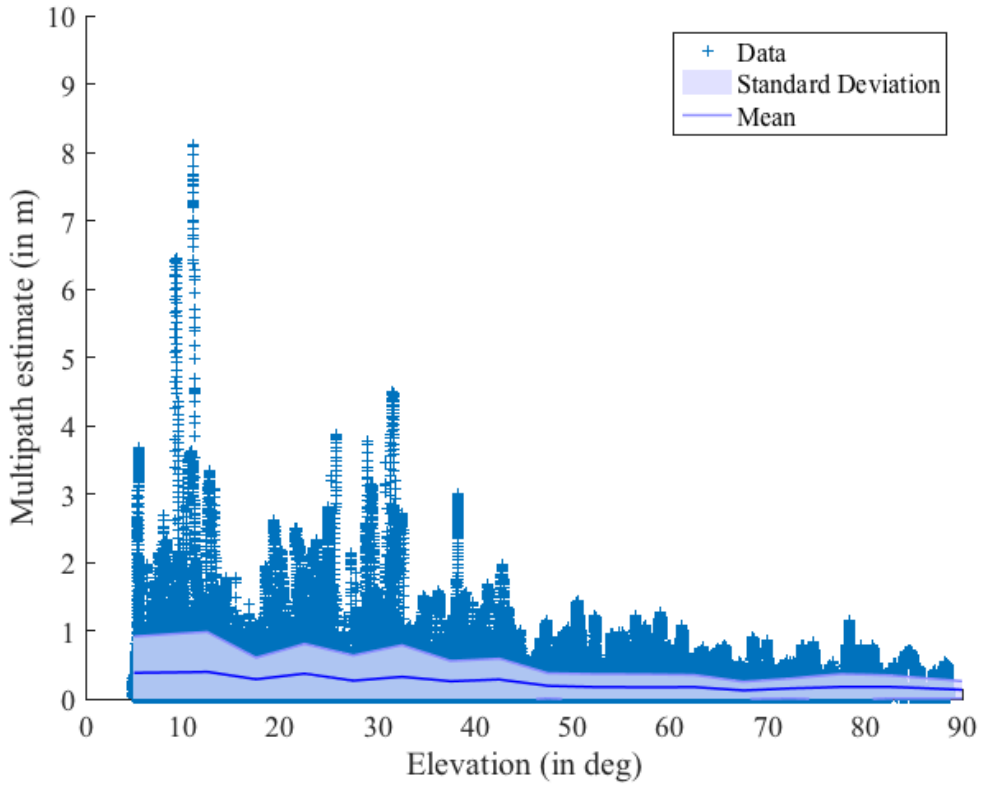


Figure 4.18 - C/A code multipath estimate on L1 as a function of the satellite elevation angle

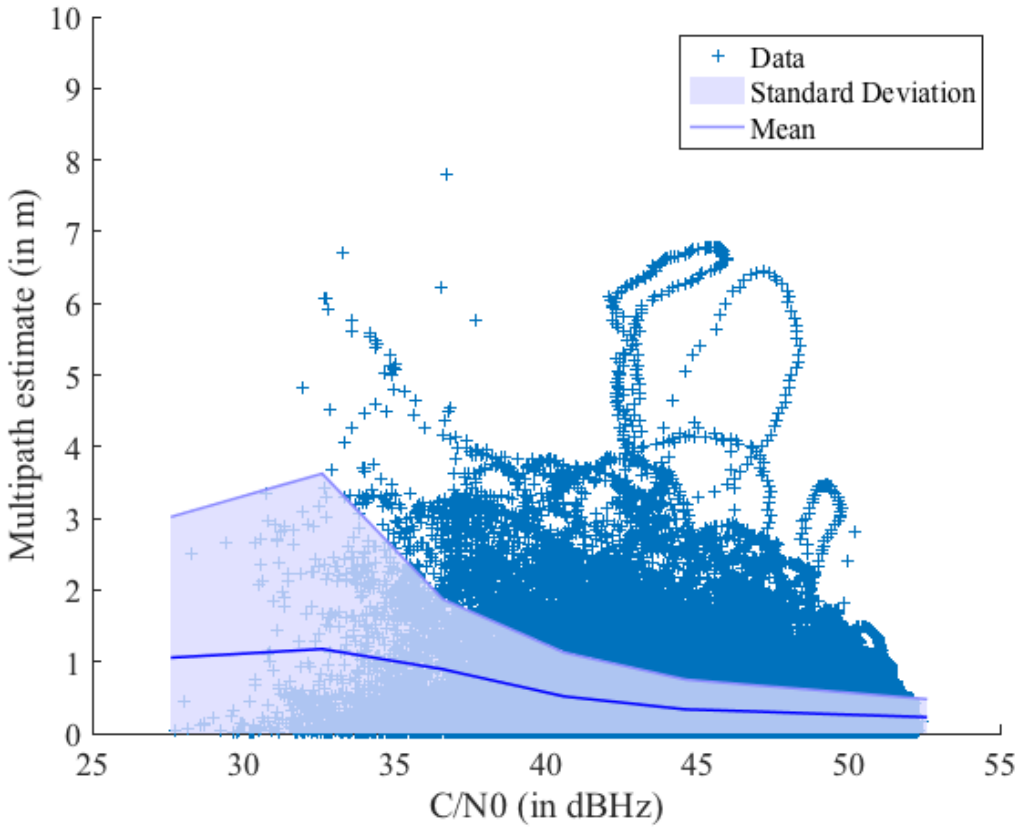


Figure 4.19 - C/A code multipath estimate on L1 function of the carrier to noise power density ratio

4.4.3 Mainline railway

The third dataset analysed in this chapter was collected on board a mainline train in the suburban area of Birmingham (see description in section 1.4.3). Unlike the high-speed train, which travels mainly through rural areas, the mainline train faces many obstacles are present along the track including trees, buildings and overhead bridges. The duration of the dataset is 25 minutes and, as illustrated in Figure 4.20, does not cover all elevation angles.

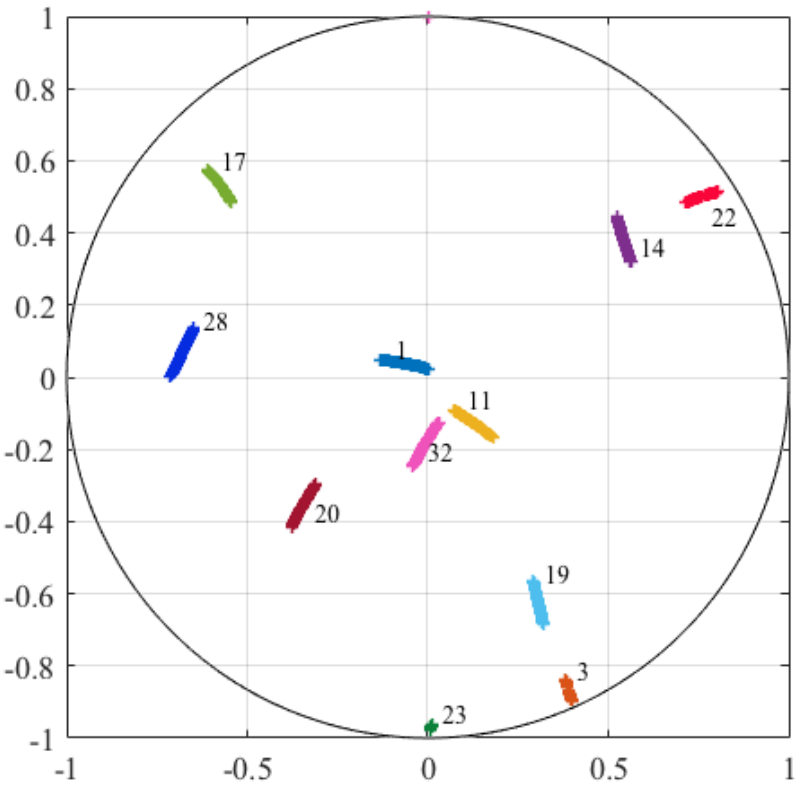


Figure 4.20- Skyplot representing the satellite elevation and azimuth angles.

Figure 4.21 represents the carrier to noise ratio value for the different satellites as a function of their elevation angle. As for the previous cases, the C/N0 appears to decrease with the elevation angle. However, the effect is less pronounced with an average ranging from approximately from 43 to 47 dBHz. Also, the C/N0 standard deviation is larger, which can be related to the greater variability of the environment through which the train travels. The outliers observed on high elevation satellites can be related to overhead obstacles. For example, the C/N0 values of 34.6 dBHz for satellite 1 and of 37.11 dBHz for satellite 11 correspond to the epoch 476986, during which the train is travelling below a pedestrian bridge as can be seen in Figure 4.23. The signals from the overhead satellites are still received, but the carrier to noise ratio is relatively low compared to the satellite elevation.

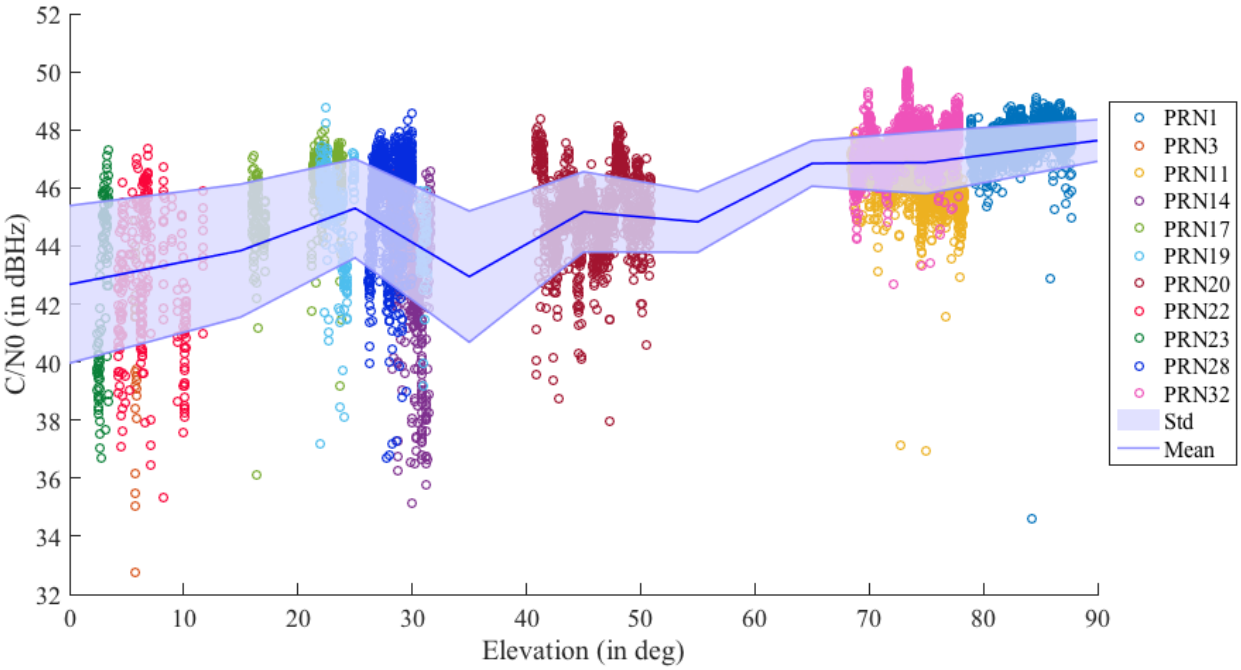


Figure 4.21- Carrier to noise power density ratio as a function of the satellite elevation

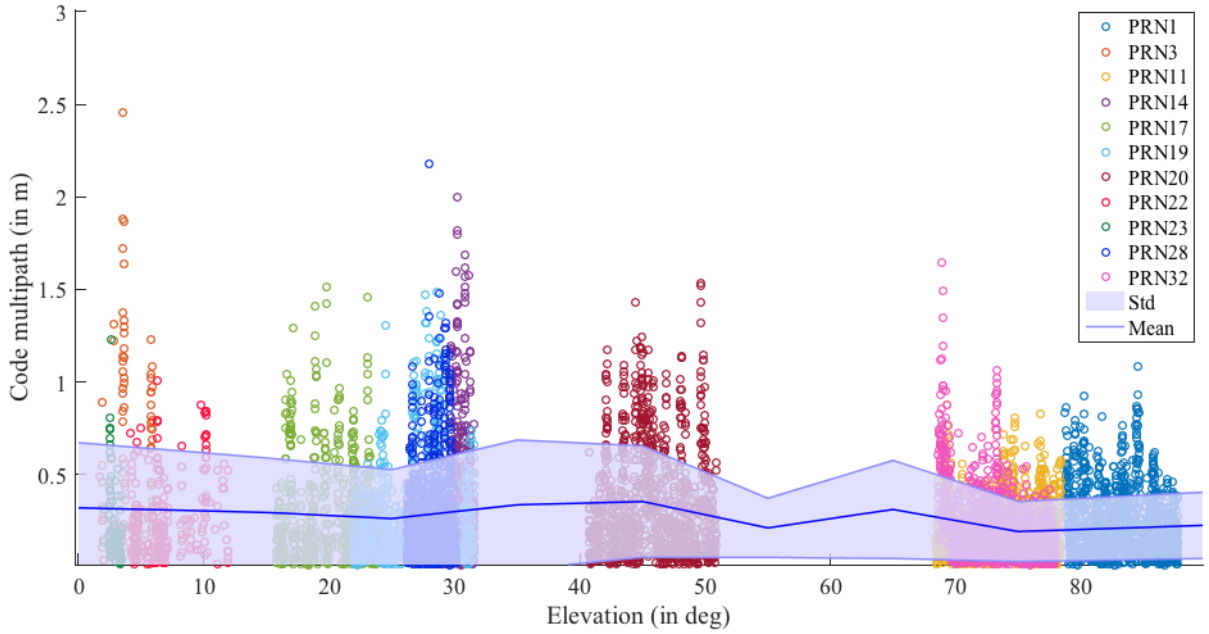


Figure 4.22- C/A code multipath estimate as a function of the satellite elevation



Figure 4.23 – Train position at epoch 476986, (Google Earth 2012)

The C/A code multipath and noise on L1 was estimated for each satellite by computing the multipath estimate defined in equation (4.54) and by averaging its value. Figure 4.22 and Figure 4.25 represent the C/A code multipath and noise on L1 as a function of the satellite elevation and of the C/N0 values.

For angles above 40°, the code multipath and noise appear to slightly decrease as the satellite elevation increases. However, angles below 30° seem to produce smaller code multipath, which may be related to the limited amount of data available for low elevation satellites. In any case, the code multipath and noise variation as a function of the elevation presents quite a large variability and the elevation-based weighting models presented in section 4.3.2.1 are unable to represent such cases. They are, however, representative of the general trend of the noise to increase as the elevation decreases.

Figure 4.24 represents the code multipath and noise of satellites at different elevation angles, compared with the variation in elevation. For the high elevation satellite, the code noise appears to decrease as the elevation increases. For the medium and low elevation satellites, the correlation between the elevation changes and the code multipath and noise level is less visible.

Finally, Figure 4.25 represents the code noise and multipath on L1 as a function of the C/N0. The code multipath and noise slightly increases as the C/N0 decreases but the correlation between the C/N0 and the code noise seems to be limited. The C/N0 based weighting methods presented in section 4.3.2.2 may then have a limited impact on the estimation of the position.

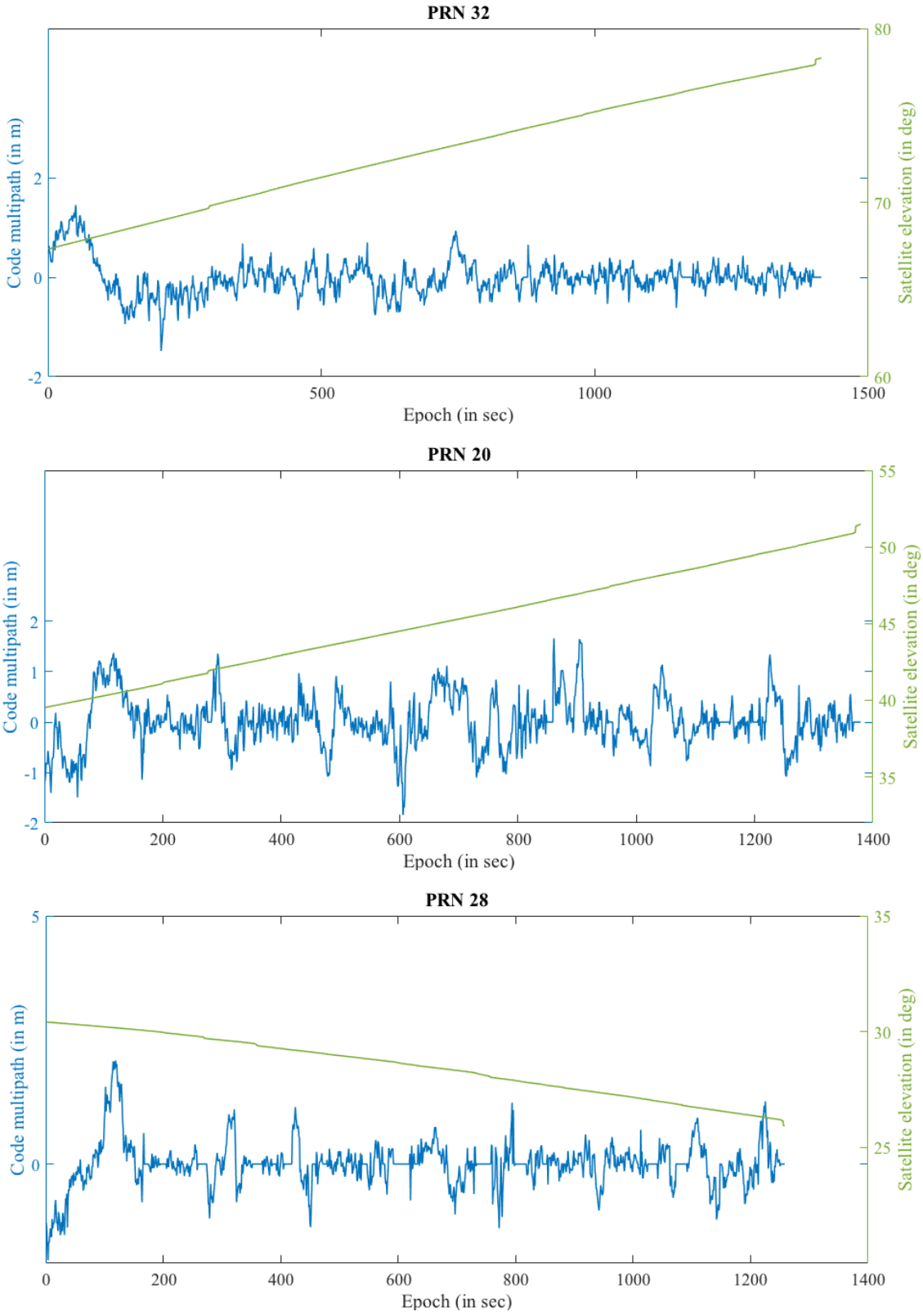


Figure 4.24 – C/A code noise and multipath on L1 and satellite elevation as a function of time for (1) a high elevation satellite PRN 32, (2) a medium elevation satellite PRN 20 and (3) a low elevation satellite PRN 28.

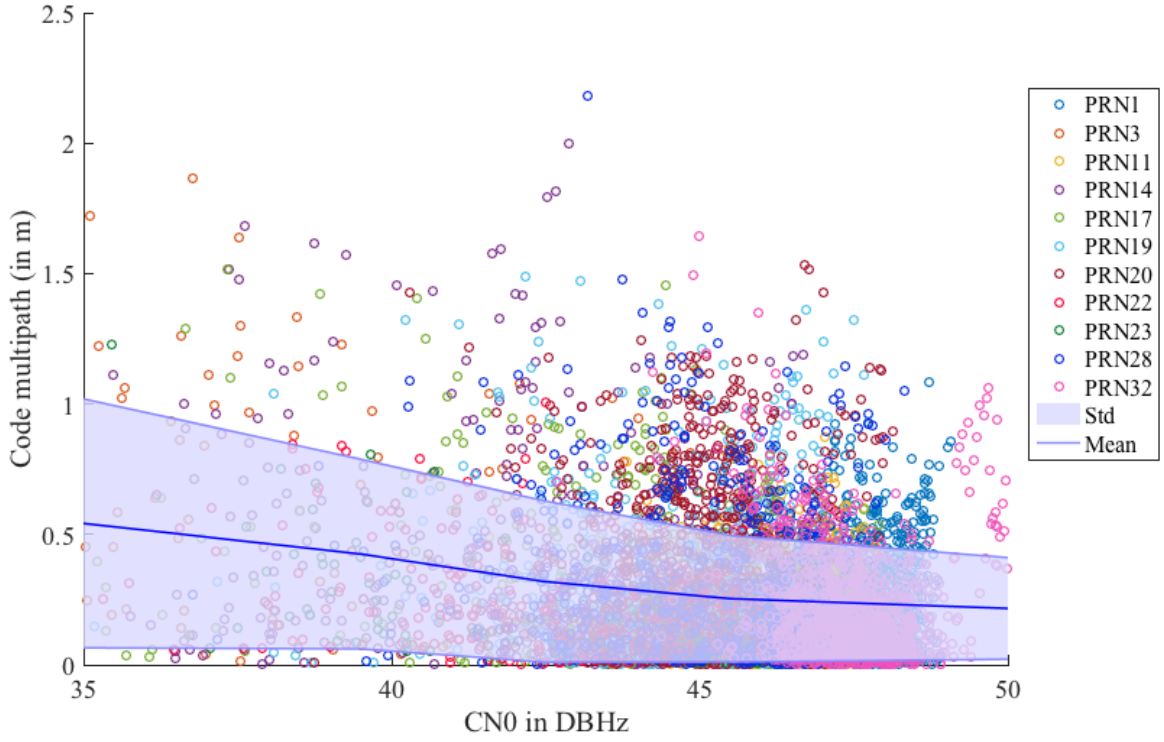


Figure 4.25- C/A code multipath estimate as a function of the carrier to noise power density ratio

In conclusion, the satellite elevation angle and C/N0 value appear to be correlated in static cases, especially for elevation angles above 30°. In dynamic cases, the C/N0 trend to decrease with the satellite elevation can still be observed but presents a higher variability. Thus, the models presented in section 4.3.2 may not perform as well with a dynamic dataset than with a static dataset. However, they are representative of the general trends observed in the data and should improve to some extent the positioning accuracy.

4.5 Conclusion

This chapter has reviewed the main sources of error affecting GNSS signals. The deterministic nature of some errors, as well as their spatial and temporal correlations, can be exploited to mitigate or remove the effect they have on the measurements. However, due to their high variability in dynamic contexts, the effect of multipath cannot easily be removed and multipath has been identified as a significant contributor to the positioning error in the railway environment, where numerous trackside obstacles can reflect the GNSS signals. A description of the multipath effect on the GNSS measurements and of existing mitigation techniques was given in section 4.2. Among these mitigation techniques, this thesis focuses on weighting techniques that do not require any additional hardware or software modification. Existing weighting techniques have been reviewed and their limitations identified in section 4.3. A preliminary analysis carried out in section 4.4 demonstrated that the relationships assumed by the

weighting techniques between the satellite elevation angle, C/N0 and code multipath and noise are valid for a static receiver. However, the C/N0 and code multipath and noise present a larger variability in the case of a dynamic receiver, making any relation between these quantities approximate or even wrong for low elevation satellites or low C/N0 values. Based on the limitations of existing weighting methods, an enhanced method developed to weight the measurement in the context of railway operations is presented in the following chapter along with the modifications to the software made to better mitigate the effect of multipath by weighting.

Chapter 5

A Novel Heading based Weighting Method

Based on the limitations of existing weighting methods reviewed in the previous chapter, this chapter presents a novel weighting method based on the train heading (i.e. the angle between the North and the direction of travel), which addresses the challenges posed by multipath in the railway domain. Section 5.1 presents this method, which introduces the orthogonal distance in order to quantify the alignment of a satellite with the train track. In section 5.1.2, measurement weights are defined based on this distance and a possible parameterisation of the novel method, enabling an adaptive weight calculation, is presented in section 5.1.3.

Section 5.2 presents the software used to implement this weighting technique. Different functions added to the software in order to perform the analysis are presented, including the detection of cycle slips (section 5.2.1), the estimation of the code multipath (section 5.2.2) and the computation of the carrier-smoothed code using the Hatch filter (section 5.2.3). Cycle slip detection is required to both estimate the code multipath and smooth the code. The estimation of the code multipath is performed to characterise the measurement error associated with multipath signals in the railway environment. This error is used to weight the measurements as discussed in section 5.2.2.4. Finally, the Hatch filter is added to compare its performance to the weighting methods, in terms of multipath mitigation.

The aim of this chapter is to quantify the effect of these different methods on the GNSS measurements, before assessing their impact on the positioning solution in chapter 6.

5.1 Novel weighting method

As seen in chapter 4, the relationships between the code multipath and the satellite elevation angles and signal C/N0 values, are more approximate for a dynamic receiver in the railway environment, than they are for a static receiver. The weighting method presented in this section is based on the development of a new indicator incorporating the satellite elevation, C/N0 and the multipath estimate, as well as the geometry of the railway.

This indicator is introduced in section 5.1.1, the associated weights are defined in 5.1.2 and a possible parameterisation of the model is discussed in section 5.1.3.

5.1.1 Orthogonal distance

As discussed in section 3.3.1, railway tracks are composed of straight sections connected by curves with a large radius. This implies that the areas ahead and behind a train are more likely to be free of obstacles and that satellites situated along the track are more likely to be visible and less affected by multipath than satellites across the track, in the case where no overhead obstacle is present. A new measurement, named the orthogonal distance, is introduced to quantify the satellite alignment with the train track. It corresponds to the orthogonal projection of the satellite position onto the train direction as shown in Figure 5.1. The satellite azimuth is represented by the angle made by the North and the satellite (SV). The satellite elevation is proportional to the distance from the centre of the graph to the satellite. If the satellite is at the centre of the graph, it is at the zenith and the elevation angle is equal to 90°. If the satellite is on the edge of the graph, its elevation is equal to 0°. Due to the guided nature of the railway, the track direction is equal to the train's heading.

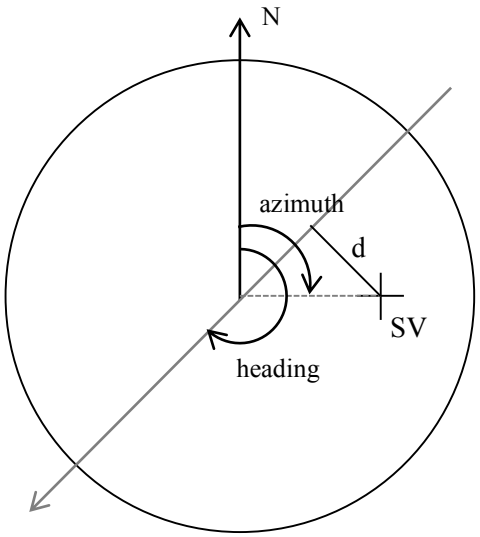


Figure 5.1 - Definition of the orthogonal distance between the satellite and the train heading, as viewed on a skyplot.

The orthogonal distance d is a function of the satellite elevation and azimuth angles, and of the train heading. It is computed using equation (5.1).

$$d = \left(1 - \frac{2}{\pi} * elevation\right) * |\sin\left(heading - azimuth\right)| \tag{5.1}$$

The value of the orthogonal distance d ranges between zero and one and is dimensionless. Figure 5.2 shows the value of the orthogonal distance for different satellite elevation and azimuth angles, assuming the train heading is equal to 45°.

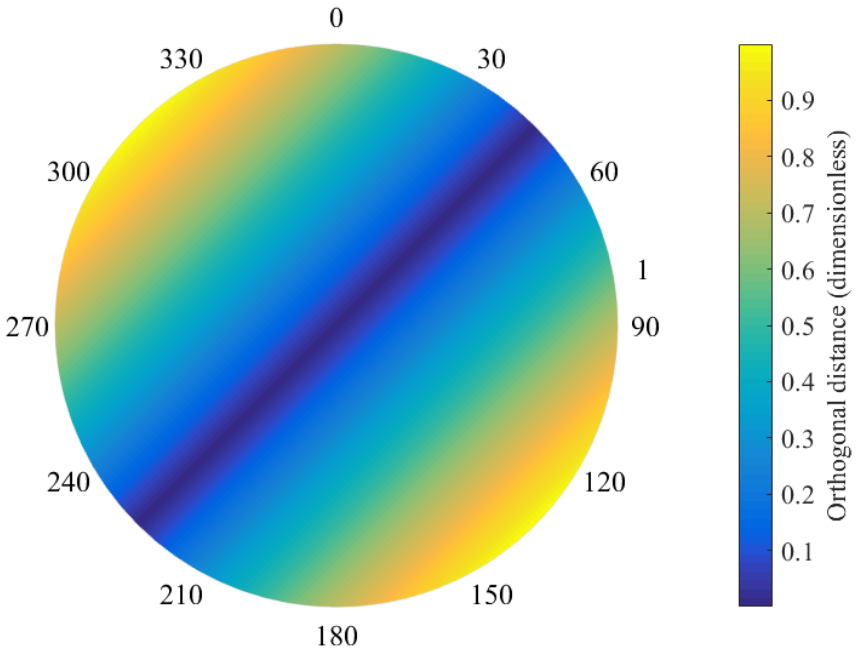


Figure 5.2 - Orthogonal distance as a function of the satellite azimuth and elevation angles for a heading of 45°.

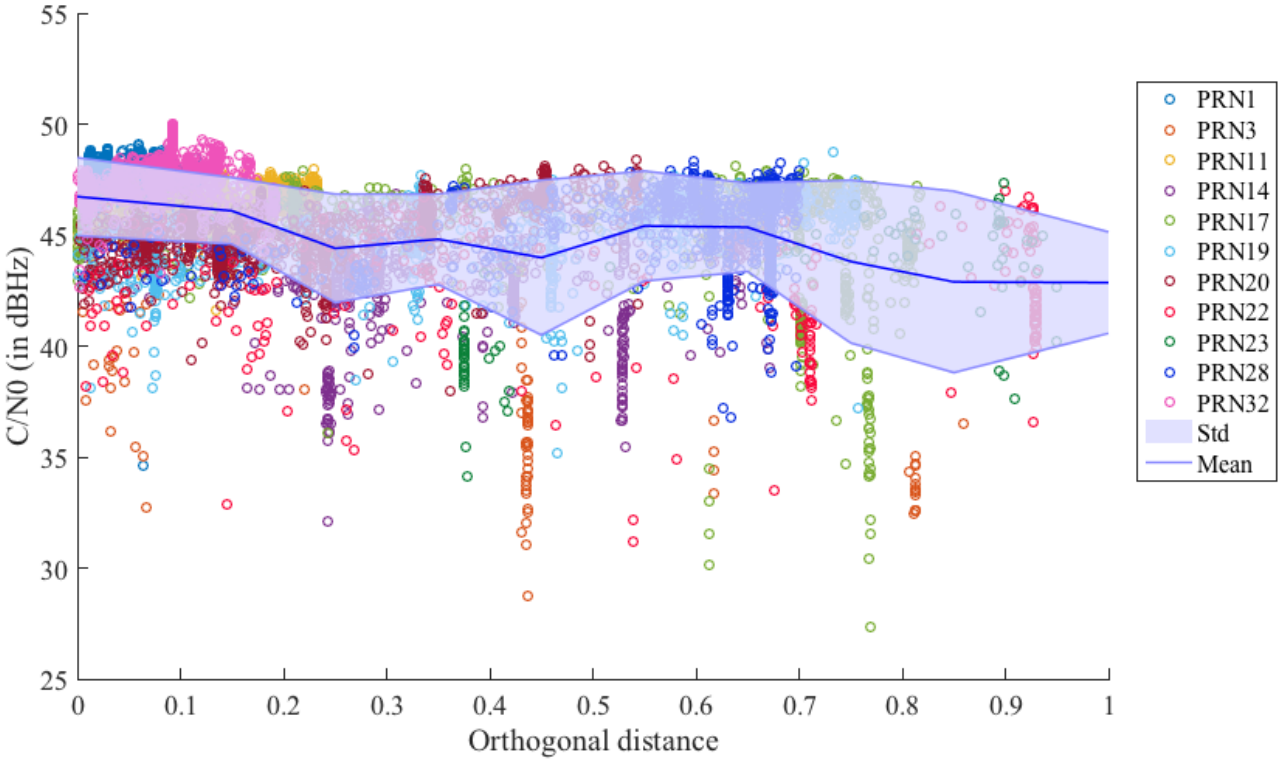


Figure 5.3 – Carrier to noise power density ratio as a function of the orthogonal distance

Figure 5.2 shows that for satellites with an azimuth angle equal to the heading (modulo 180°), the distance *d* is equal to zero, regardless of the satellite elevation angle. This distance *d* decreases as the difference between the train heading and the satellite azimuth increases.

Figure 5.3 shows the C/N0 value as a function of the orthogonal distance. Overall, the C/N0 tends to decrease as the orthogonal distance increases. This means that satellites aligned with the track tend to have higher C/N0 values than satellites that are further from the track. However, certain satellites produce low values of C/N0, regardless of the distance. These values of C/N0 are generated when the train is static, e.g. satellite 3 provides C/N0 values between 30 and 40 dBHz (for an orthogonal distance of approximately 0.45) when the train stops at Wylde Green Station. As illustrated in Figure 5.4, the receiver has a good visibility in the satellite direction and is thus able to acquire its signal. The orthogonal distance associated with satellite 3 is also relatively low (d = 0.43) as the satellite azimuth angle (156°) is close to the train direction. However, this satellite has an elevation of 3°, which explains its low C/N0 values. Thus the orthogonal distance may not be representative on low elevation satellites aligned with the track.



Figure 5.4 – Position of the train at Wylde Green Station (orange pin) with the azimuth of satellite 3, (Google Earth 2012)

Figure 5.5 represents the code multipath estimate function of the satellite orthogonal distance. The code multipath value and its variance appear to increase with the orthogonal distance. However, the limited amount of data available results in a qualitative rather than quantitative observation, especially for orthogonal distances greater than 0.8.

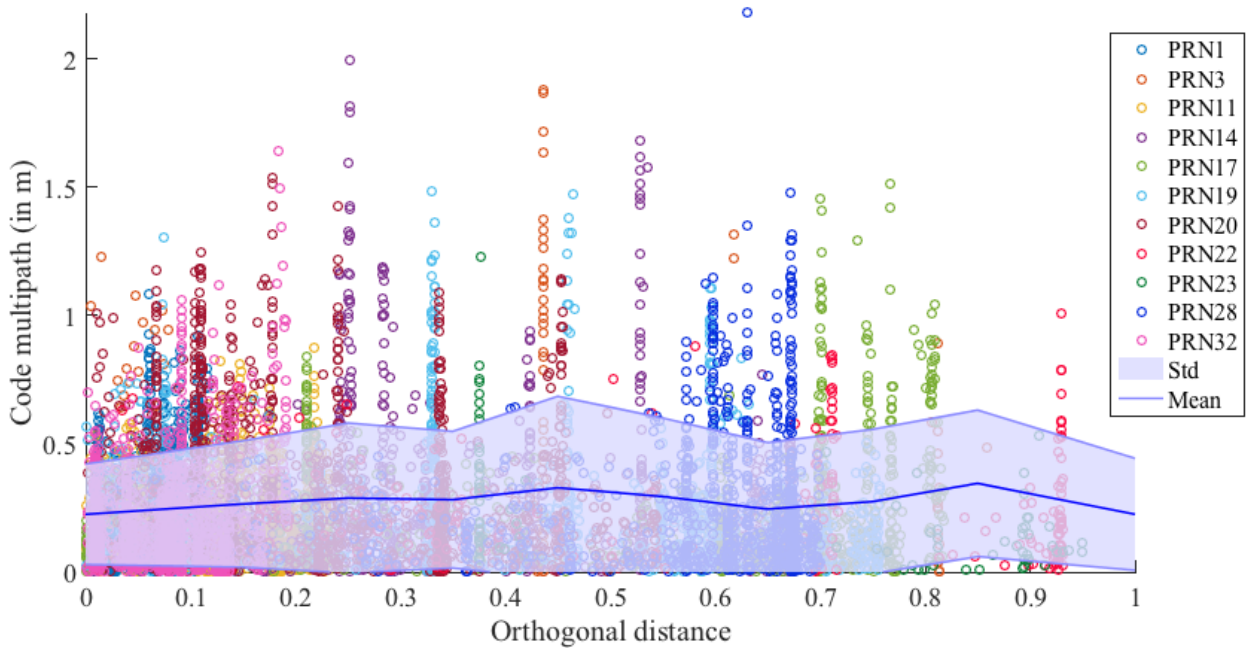


Figure 5.5 – C/A code multipath estimate as a function of the satellite orthogonal distance

5.1.2 Definition of weights

The orthogonal distance *d* is used to weight the measurements and to do this, it is assumed that satellites closer to the track (i.e. with a small value of *d*) are less likely to be affected by multipath. These satellites should then be assigned a lower noise variance. An initial variance definition is:

$$\sigma_i = \sigma_0 * (A + B * d) \tag{5.2}$$

The constants are selected such that the standard deviation is comparable to the traditional weighting methods, specifically the Black and Eisner method, which has values ranging from approximately $1\times\sigma_0$ to $22\times\sigma_0$ (see equation (4.75)). Thus, A is chosen to equal 1 and B to equal 21.

This formulation of the variance is compared to the weighting methods, previously reviewed, based on the satellite elevation angle: the Black and Eisner method and the stochastic elevation method (see section 4.3).

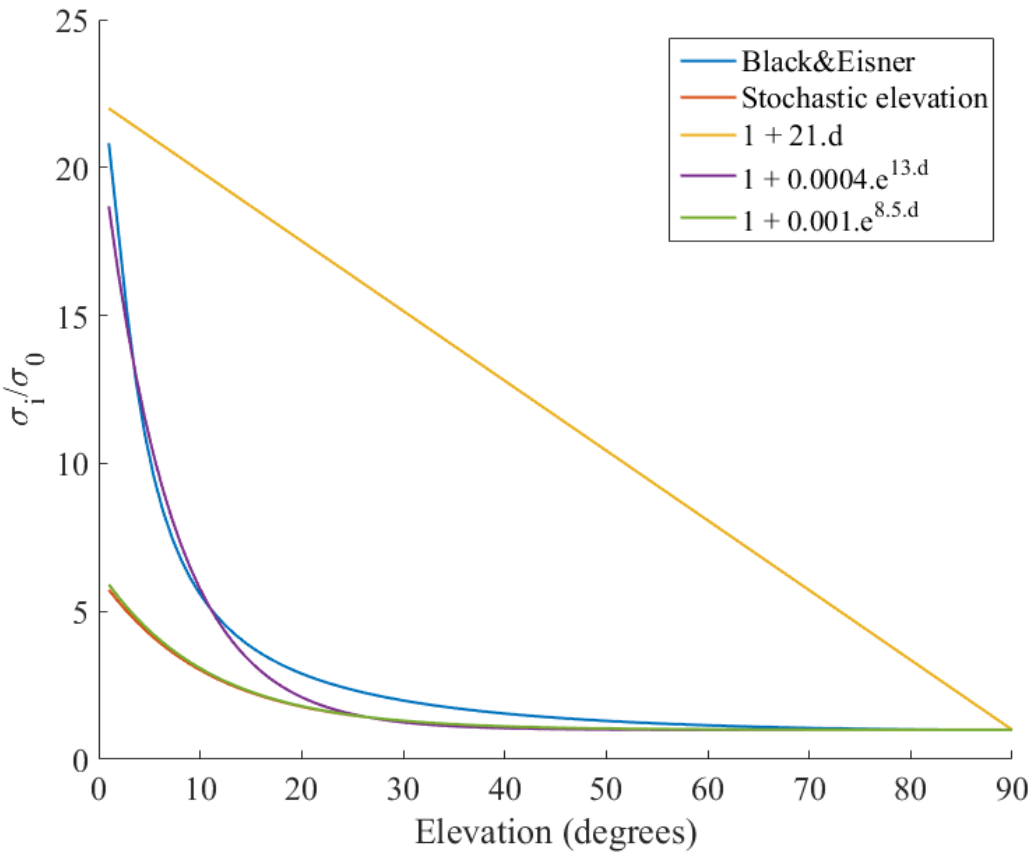


Figure 5.6 - Measurement noise variance as a function of the satellite elevation angle for different weighting methods

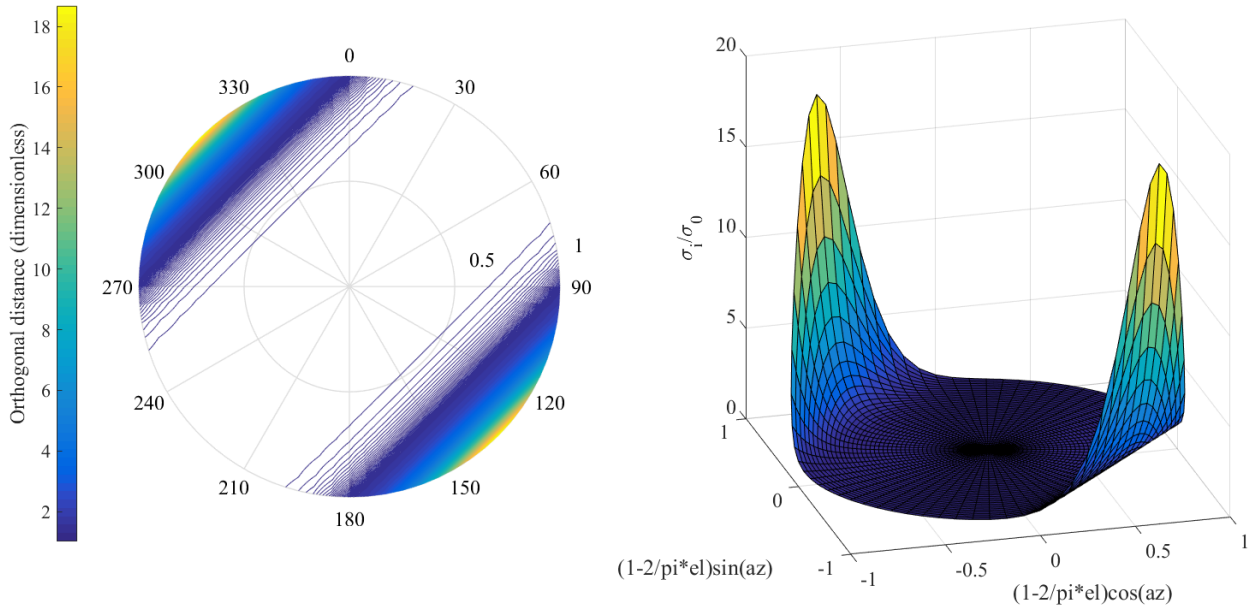


Figure 5.7 - Measurement noise variance based on the squared orthogonal distance exponential decay term as a function of the satellite elevation and azimuth, for a heading of 45°, as viewed on a skyplot (right) and as view in 3D (left).

As seen in Figure 5.6, compared to the Eisner & Black and the stochastic elevation methods, the novel method defined by equation (5.2) consistently assigns a larger variance to the measurements, both for satellites at low and high elevation angle. Alternatively, an exponential decay term can be used, as in (Elsobeiey & El-Rabbany 2010) in the stochastic elevation model (equation (4.76)).

$$\sigma_i = \sigma_0(1 + A * e^{B*d}) \tag{5.3}$$

For the variance to be similar to the stochastic model, *A* is fixed at 0.001 and *B* at 8.5. For a variance similar to the Eisner & Black model, *A* is fixed to 0.00004 and *B* to 13. As can be seen in Figure 5.6, this formulation assigns a noise variance more closely related to the traditional methods and the distribution of the noise variance as a function of the satellite azimuth and elevation in the case where A is equal to 0.00004 and B to 13 is presented in Figure 5.7.

In Figure 5.7, the effect of the distance *d* on the variance can clearly be seen. The variance increases for satellites further from the track heading (i.e. large values of *d*) while it remains constant for satellites close to the track. The effect is particularly pronounced for values of *d* above 0.8.

However, satellites along the track but with a low elevation angle are assigned the same noise variance as satellites directly above the train. Yet, these satellites are more likely to be affected by multipath or by larger residual atmospheric errors than the satellites with a high elevation angle. In order to increase the noise variance assigned to these satellites, an additional term based on the Eisner and Black method is used. This term assigns the noise variance solely based on the satellite elevation angle, as seen in Figure 5.8.

The new formulation of the measurement noise variance is given by equation (5.4).

$$\sigma_i = \sigma_0 \left[\frac{1}{C} * (1 + A * e^{B*d}) + \left(1 - \frac{1}{C}\right) * \left(\frac{1.001}{\sqrt{0.002001 + \sin(elevation)^2}}\right)\right] \tag{5.4}$$

By adjusting the value of the coefficient *C*, the novel variance distribution can either depend solely on the satellite elevation angle (for a large value of *C*) or on the orthogonal distance *d* (for a small value of *C*), as illustrated in Figure 5.9. By combining the effect of both the satellite elevation angle and the orthogonal distance, the novel weighting method can increase the noise variance assigned to satellites with a low elevation angle, and further increase this value for satellites that are also further from the track. For example in Figure 5.9, for C equal to 10 (bottom-right), the variance depends solely on the satellite elevation. When C is equal to 1 (top-left), the variance depends only on the orthogonal distance and satellites aligned with the track are assigned low variance regardless of their elevation angle. Finally, for C equal to 2 and 3 (respectively top-right and bottom-left), the variance distribution takes into account both the satellite elevation and the orthogonal distance. Satellites with low elevation angles are assigned larger variance, and this value is further increased if the satellite is far from the track (i.e. large orthogonal distance).

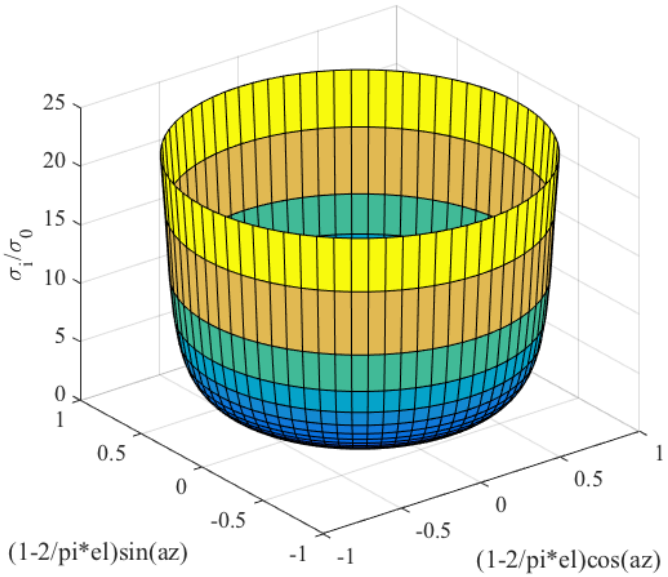


Figure 5.8 - Measurement noise variance based on the Eisner and Black method as a function of the satellite elevation and azimuth, as view in 3D.

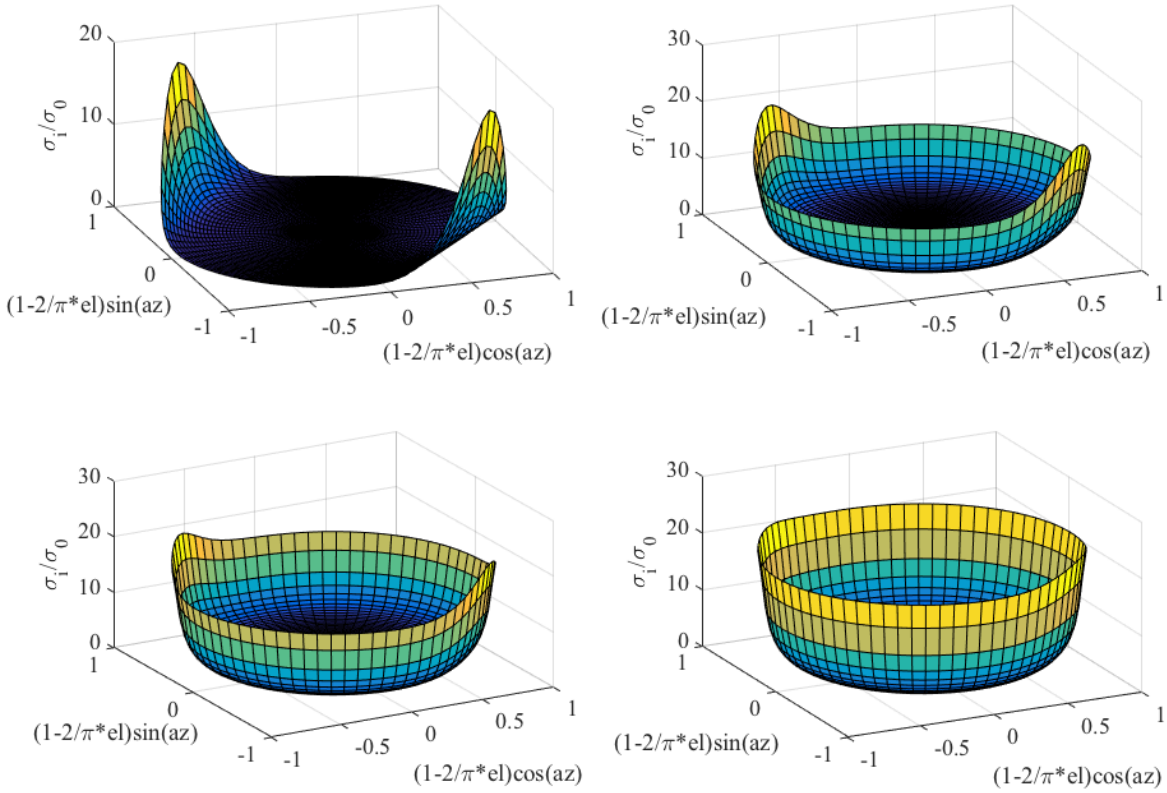


Figure 5.9 - Measurement noise variance of the novel weighting method as a function of the satellite elevation and azimuth, as view in 3D for: (top-left) C =1, (top-right) C = 2, (bottom left) C = 3 and (bottom-right) C = 10.

5.1.3 An adaptive model

It has been seen in the previous section that the values of the parameters *A*, *B* and *C* defined in equations (5.3) and (5.4) influence the noise standard deviation assigned to a measurement coming from a satellite at a particular elevation and azimuth. These parameters can be linked to signal quality indicators such as the signal to noise power density ratio (C/N0) or the pseudorange multipath estimate, as outlined below. In this manner, the noise standard deviation model can be adapted to different railway environments, as explained in section 3.3.3.

First, the parameter *A*, which can be related to the noise standard deviation decrease rate as a function of the orthogonal distance, is expressed as a function of the C/N0 in a similar way to (Hartinger & Brunner 1999).

$$A = k * 10^{-\frac{C/N0}{10}} \tag{5.5}$$

The measurement noise standard deviation for the novel heading based weighting method is represented in Figure 5.10 for different values of C/N0, leading to different values of the parameter *A* through equation (5.5). In this figure, the angle between the track and the satellite azimuth is assumed to be equal to 90°.The parameter *k* is fixed to 1/18 in order to generate standard deviation values comparable to the traditional methods such as the Black and Eisner model. The parameter *A* is such that the higher the C/N0, the smaller is the noise standard deviation. The reciprocal is also true, the lower the C/N0, the higher the noise standard deviation. Figure 5.10 compares the novel technique to the Eisner and Black method. For orthogonal distances below 0.6, the new model assigns an approximately constant variance equals to 1. The variance then increases with the orthogonal distance, and the higher the C/N0, the steeper the increase. In comparison, the Black and Eisner model continuously increases the noise standard deviation as a function of the elevation. In the new model, C/N0 values below 35 dBHz are assigned larger variances than with the Eisner and Black model. Higher values of C/N0 produce lower variances than the Eisner and Black model. The influence of the parameter A on the measurement noise variance distribution is further shown in Figure 5.12.

The parameter *C* adjusts the respective contribution of the elevation only based weighting and the novel heading based weighting methods, according to equation (5.4). This parameter is linked to the pseudorange multipath in equation (5.6). The logarithmic function is selected as it provides relatively large values for small multipath estimates and decreases at a slow rate when the multipath estimate increases. Therefore the elevation based weighting will be predominant when multipath estimate is small, and the contribution of the satellites of the side of the track will decrease as the multipath estimate increases.

$$C = 1.7 - \log(|multipath_i|) \tag{5.6}$$

where the code multipath estimate $multipath_i$ is computed according to equation (4.54).

The value taken by *C* for multipath estimates between 0.001 m and 1.8 m is represented in Figure 5.11. The multipath values are generated based on the code multipath distribution defined in section 5.2.2.3.

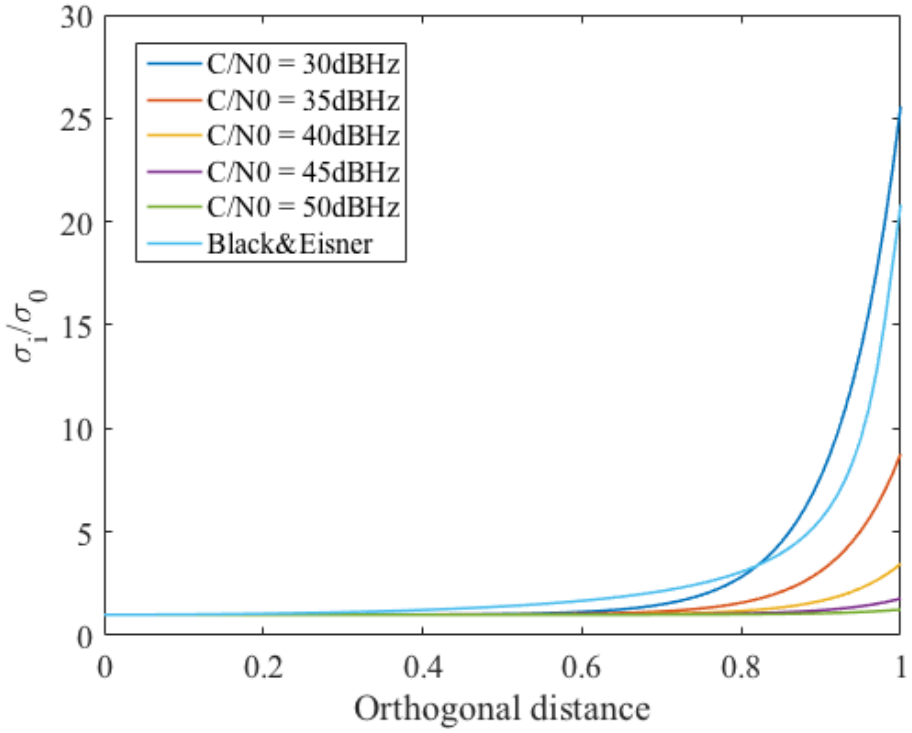


Figure 5.10 - Measurement noise variance for the novel heading based weighting method and the Eisner & Black method, for different values of C/N0

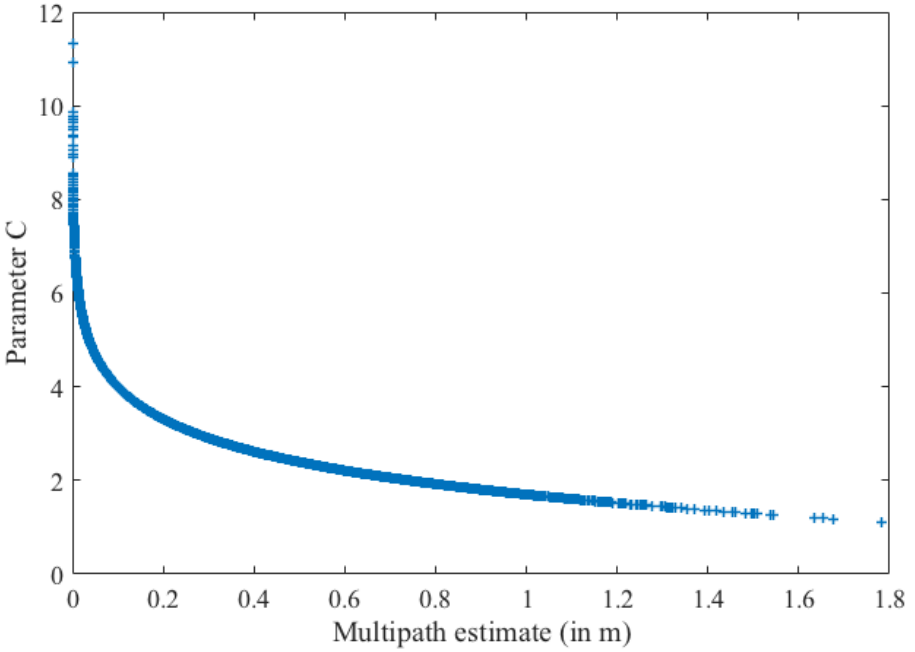


Figure 5.11- Value of parameter D as a function of the pseudorange multipath estimate

If the multipath estimate is less than 0.7 m, the parameter *C* is greater than 2, implying that the noise variance value is driven mainly by the satellite elevation. Thus, in an environment with low multipath signals, the measurements are weighted as a function of the satellite elevation. Conversely, for a multipath estimate greater than 0.7 m, the value of *C* is less than 2 with a noise variance driven mainly by the satellite orthogonal distance. This implies that in the presence of multipath, satellites further from the track are assigned larger noise variance. This relies on the assumption that multipath signals are most likely created by trackside obstacles, and therefore weighting down measurements coming from satellites on the side of the track may improve the positioning accuracy.

Figure 5.12 represents the influence of both the C/N0 and the multipath estimate on the measurement noise variance assigned by the novel heading based weighting techniques as a function of the satellite azimuth and elevation angles.

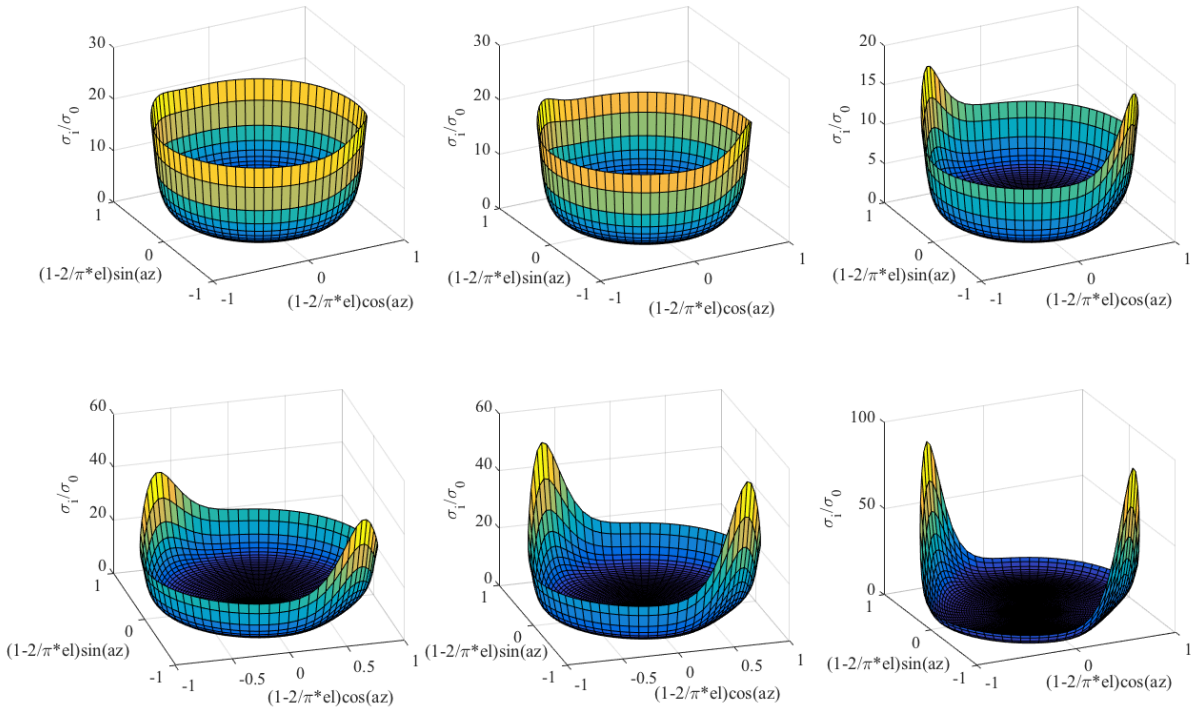


Figure 5.12- Measurement noise variance assigned the heading weighting method as a function of the satellite elevation and azimuth, as view in 3D for (C/N0, multipath estimate) = top-left (45 dBHz, 0.01m), bottom-left (35 dBHz, 0.01m), top-middle (45dBHz, 0.1m), bottom-middle (35dBHz, 0.1m), top-right (45dBHz,1m), bottom right (35dBhz, 1m)

In the best-case scenario, where the C/N0 is high and the multipath estimate is low, the measurements are weighted on the basis of the satellite elevation. The same applies if the multipath estimate remains low while the C/N0 is low. However, satellites further from the track are assigned smaller weights (i.e. larger noise variance). If the C/N0 is high but the multipath estimate is low, satellites further from the track are weighted down while satellites close to the track are assigned a low noise variance. These two last cases are, however, unlikely to happen as the presence of multipath is likely to affect the value of

the C/N0. Finally, if the C/N0 is low and the multipath estimate is high, satellites further from the track are weighted down, however, as discussed in section 5.1.2, the value associated with the noise variance might then be too large.

The novel heading weighting technique combines traditional measurement quality indicators such as the elevation angle, the multipath estimate and the C/N0, with an additional indicator, the orthogonal distance, which reflect the geometry of the railway environment. The technique is made adaptive by parameterising the contribution of each indicator to the final solution.

The effect of this novel weighting technique and of its parameterisation on the positioning solution is discussed in Chapter 6.

5.1.4 Heading determination

The computation of the novel weights required the knowledge of the train heading and three methods by which it can be measured at presented below. One method by which the heading can be determined is by the use of map-matching techniques, as explained in section 3.1.3.2, which compare the estimated train position to a track map. The track alignment with respect to the North at the estimated position can then be extracted from the map and used as an estimate of the heading.

The heading can also be determined using inertial sensors, as seen in section 3.1.3.1. As the train body is approximately parallel to the track, a single z-axis gyroscope, measuring the train rotation about the vertical axis, directly provides a measure of the heading. However, the sensor can be affected by slopes or banks on the track that generate an error. When the vehicle is stationary, the heading can be assumed constant in order to prevent degradation.

Finally, the train heading can be estimated on the basis of the train velocity solution, as given by the GNSS-receiver, using equation (5.7):

$$heading = arctan_2\left(\frac{v_E}{v_N}\right) \tag{5.7}$$

where $arctan_2$ is the four-quadrant arctangent function, v_E is the East velocity and v_N is the North velocity.

The GNSS velocity can be estimated by differencing consecutive positions or using the Doppler shift measured in the carrier tracking loop of the receiver. This Doppler shift corresponds to the change in the apparent frequency of the signal received, due to the relative motion between the receiver and the satellite. Given the satellite velocity, the Doppler shift can be used to determine the receiver velocity (Misra & Enge 2010). While the former method only has an accuracy of a few meter per seconds, the Doppler measurements enable to estimate the velocity with an accuracy of a few centimetres per second.

The uncertainty associated with the heading is then given by:

$$\sigma_{heading} = \frac{\sqrt{v_E^2 \sigma_N^2 + v_N^2 \sigma_E^2}}{v_N^2 + v_E^2} \tag{5.8}$$

where σ_N and σ_E are the North and East velocity uncertainties (Groves 2013).

Thus, the accuracy of the GNSS-based heading estimate is a function of the train velocity, and the lower the velocity, the larger the error. Due to the noisiness of the velocity-derived heading at low speed, the GNSS-based heading determination should not be used at low speed, for example below 2 m per second (Groves et al. 2009). In such cases, the heading can be assumed constant. However, such assumption can introduce errors, especially if a train is travelling in a curve at a very low speed (e.g. below 1 m per second).

Figure 5.13 represents the train heading as measured by the INS on board the mainline suburban train (in red) together with the heading estimated on the basis of GNSS measurements (in blue). Both the INS-based and GNSS-based estimates were smoothed at low velocity, i.e. for speeds below 1 m per second, the heading was assumed equal to its previous value. As the GNSS receiver starts from a static position without an initial heading, it gives a heading equal to zero for the first 165 epochs (not represented in the figure), producing a large error. If these first epochs are not taken into account, the root mean square error between the INS-based and GNSS-based heading is equal to 4°. However, GPS outliers can generate an error of more than 50°, as seen at epoch 300. This error is due to the presence of a tunnel and a cutting, shown in Figure 5.14. As the train exits the tunnel, the GNSS-based position estimates are biased, possibly due to the signal reflections on the cutting walls. Thus the heading estimated from these positions is very different from the actual train heading.

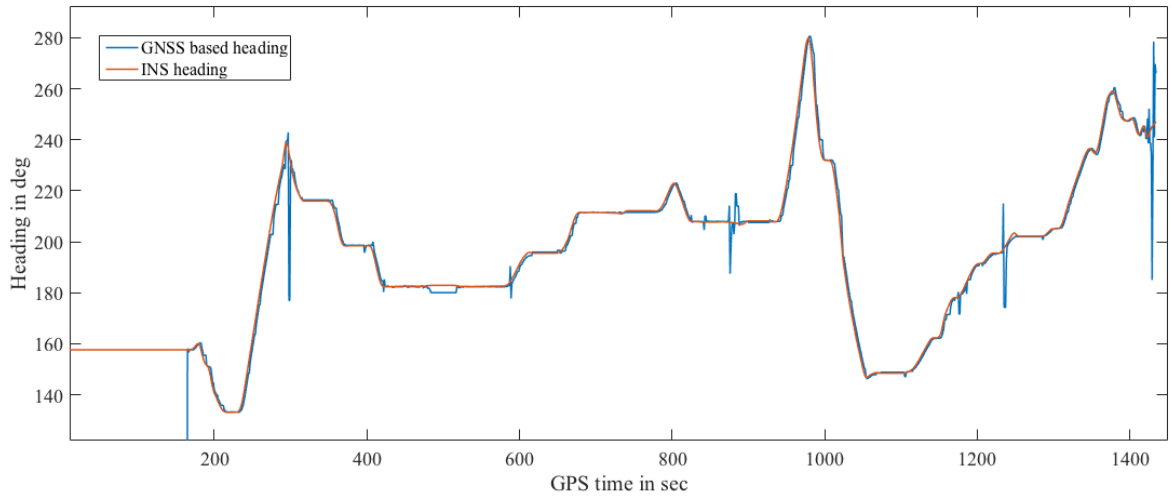


Figure 5.13 – Comparison of the train heading measured by the INS (after smoothing) with the GNSS-based heading estimate.

The GNSS-based heading estimation is thus vulnerable to GNSS outliers. However, given the guided nature of railway, it is expected that any change in the train heading will be relatively smooth, and then previous values could be used to smooth the data. The difference in consecutive heading values can also be used to detect outliers.



Figure 5.14 – Mainline suburban train trajectory at epoch 300 (indicated by the white arrow), as the train exits the Sutton Coldfield tunnel (indicated in red), (Google Earth 2012).

Due to the limited availability of GNSS measurements, for example in tunnels, the heading value should be estimated from a heading sensor (e.g. a gyroscope) if available. If such sensor is unavailable, the GNSS-based estimate still provides a good indication of the heading but it might be affected by some large errors.

In all three cases, the heading estimate is only used to compute the orthogonal distance, which is then used to weight the measurement. Thus, a heading estimation error only has a limited impact on the final positioning solution, as shown in equations (5.11) and (5.12). For example, a heading error of 5° creates a maximum orthogonal distance error of 0.09 (if the satellite elevation is equal to 0° and the heading is equal to the azimuth). In turn, this orthogonal error can create an error in the standard deviation estimate of approximately $2\sigma_0$, corresponding to 40 cm for the code measurement and 6 mm for the carrier phase. Thus, an error on the heading estimation only has a limited impact on the final positioning solution.

$$d + \Delta d = \left(1 - \frac{2}{\pi} * el\right) * |\sin(h + \Delta h - az)| \cong d + \left(1 - \frac{2}{\pi} * el\right) * \cos(h - az) * \sin(\Delta h) \quad (5.9)$$

$$\sigma_i + \Delta\sigma_i = \sigma_0\left(1 + A\, e^{B(d+\Delta d)}\right) \cong \sigma_i + \sigma_0 * 2.222 * A\, e^{Bd} \quad (5.10)$$

5.2 POINT Software

In order to test the novel heading-based weighting method presented in section 5.1 and the different multipath mitigation techniques presented in section 4.2.2.3, the POINT software was used. This software was developed as part of the innovation Navigation using new GNSS signals with hybridised technologies (iNsight) project (iNsight 2011). It uses C++ language and runs in Microsoft Windows (Jokinen 2014).

A simplified architecture of the POINT software is presented in Figure 5.15.

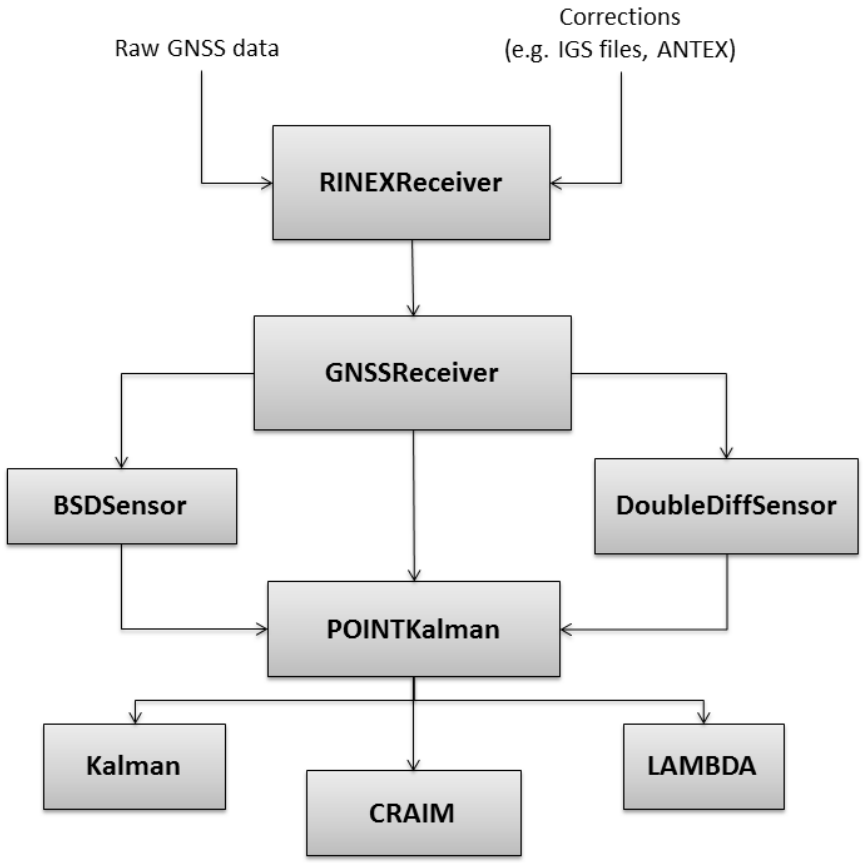


Figure 5.15 – Architecture of the POINT software, adapted from (Jokinen 2014)

Raw GNSS data in RINEX format are processed by the RINEXReceiver module. The GNSSReceiver module then applies the relevant corrections to this data and computes the relevant measurement combinations such as the ionosphere-free combination. The noise standard deviation is assigned to each measurement based on its type (pseudorange or carrier phase), its frequency, the nominal noise standard deviation defined by the user and the model used for weighting. The least squares solution is also computed.

The corrected measurements and their assigned noise standard deviations are then forwarded to the POINTKalman module which estimates the position using an extended Kalman filter, implemented in the Kalman module. POINT has the capability to perform RTK and PPP positioning (section 3.1.2.3), in which cases the measurements are differenced before being sent to the POINTKalman module. Measurements differencing can be done between satellites in the BSDSensor or between satellites between receivers (double differencing) in the DoubleDiffSensor. For RTK and PPP, the carrier phase ambiguity resolution and validation are performed in the LAMBDA and CRAIM modules (Jokinen 2014).

The main contribution of this thesis to the POINT software is the implementation of the novel heading-based weighting method (section 5.1). In addition, several functions were implemented to compare the performance of this method with existing multipath mitigation techniques, including the weighted least square algorithm (section 4.3.1), cycle slip detection, heading computation, carrier-smoothed code (section 4.2.2.3), multipath estimation (section 4.2.1.3) and existing weighting techniques (section 4.3). As explained in Figure 5.16, these functions are required to compute the positioning solution.

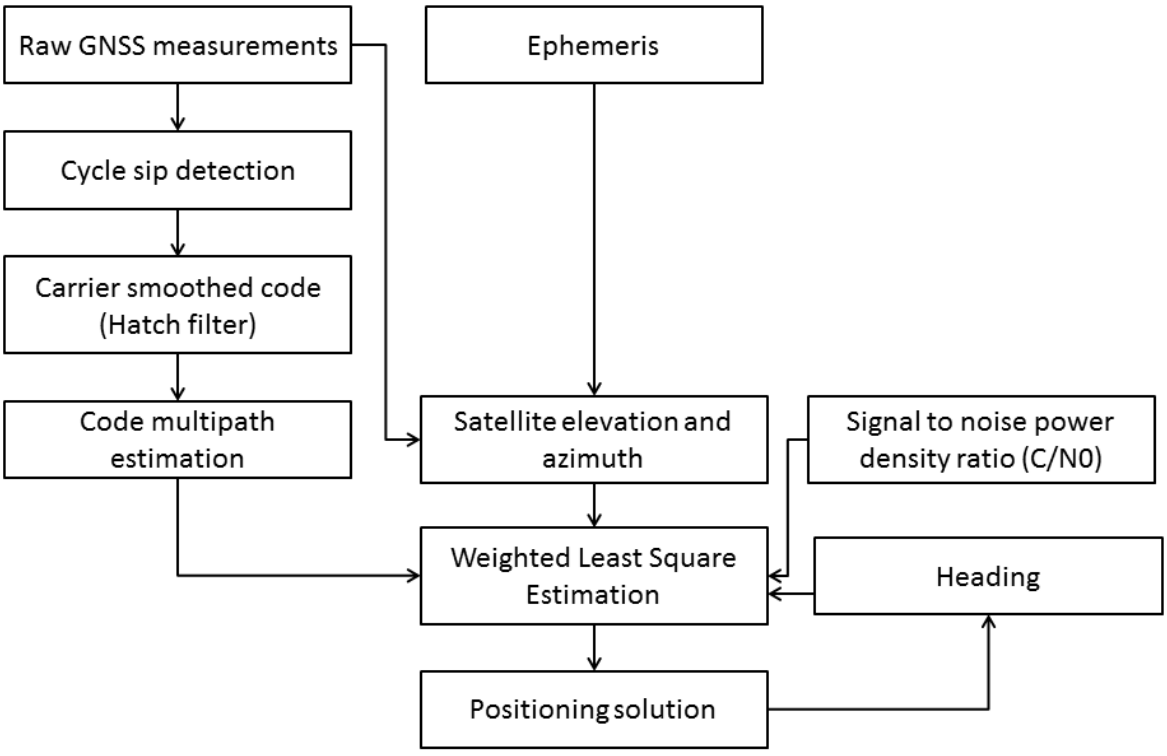


Figure 5.16 – Position estimation flowchart

The implementation of some these functions is discussed in the following sections.

5.2.1 Cycle slip detection

Cycle slips are caused by the receiver loss of lock when tracking the signal, which in turn creates discontinuities in the phase measurements. As discussed in sections 5.2.2 and 5.2.3, these discontinuities affect the position estimation process and must be detected.

Cycle slips can be seen in the carrier phase measurement as jumps proportional to the signal wavelength. However, as the range between the satellite and the receiver can change up to hundreds of metres in one second, cycle slips are hard to detect directly on the carrier phase (Sanz Subirana et al. 2013), as can be seen in Figure 5.17 and Figure 5.18. Instead, the geometry-free phase measurement is formed by differencing the carrier phase measurements on L1 and L2, as seen in equation (5.13).

$$\Phi_I = \Phi_1 - \Phi_2 \tag{5.11}$$

This difference removes the geometry (i.e. range), the clock error and the tropospheric delay, providing a precise test signal, where cycle slip can be detected.

The cycle slip detection process is implemented by computing the difference in time of consecutive geometry-free phase measurements. The n^{th} order differences can be used to amplify the jump. However, this process also increases the noise and thus the probability of false detection. For this thesis, the 2^{nd} order difference, shown in equation(5.12), was selected based on its increased probability to detect consecutive cycle slips (Sanz Subirana et al. 2013).

$$D(k) = \big(\Phi_I(k) - \Phi_I(k-1)\big) - \big(\Phi_{\mathrm{I}}(\mathrm{k}-1) - \Phi_{\mathrm{I}}(\mathrm{k}-2)\big) \tag{5.12}$$

A threshold has to be set to decide if cycle slip occurred. After analysing the 2^{nd} order difference value for each satellite in the dataset, this threshold was set to four centimetres and therefore, if the second order difference is greater than this value, a cycle slip is detected.

Figure 5.17 and Figure 5.18 represent the effect of cycle slips on the carrier phase measurement on L1, the geometry-free carrier phase (equation (5.11)), the 1^{st} and 2^{nd} order geometry-free carrier differences (equation (5.12)), respectively for high elevation satellite (PRN32) and a low elevation satellite (PRN17). As previously discussed, the rapid change in range makes cycle slips difficult to detect directly on the carrier phase but they appear clearly on the geometry-free observable. Finally, comparing the differencing value to a threshold, cycle slips can be detected.

As expected, the low elevation satellite is affected by more cycle slips than the high elevation satellite. Indeed, cycle slips occur more frequently when the signal is weak (i.e. low C/N0), as for low elevation satellites.

The performance of the cycle slip detection is assessed by visual inspection of the measurements and appears to detect cycle slips visible on the measurements. Residual cycle slips may still be present but they are relatively small and have a limited impact on the Hatch filter and the multipath estimate.

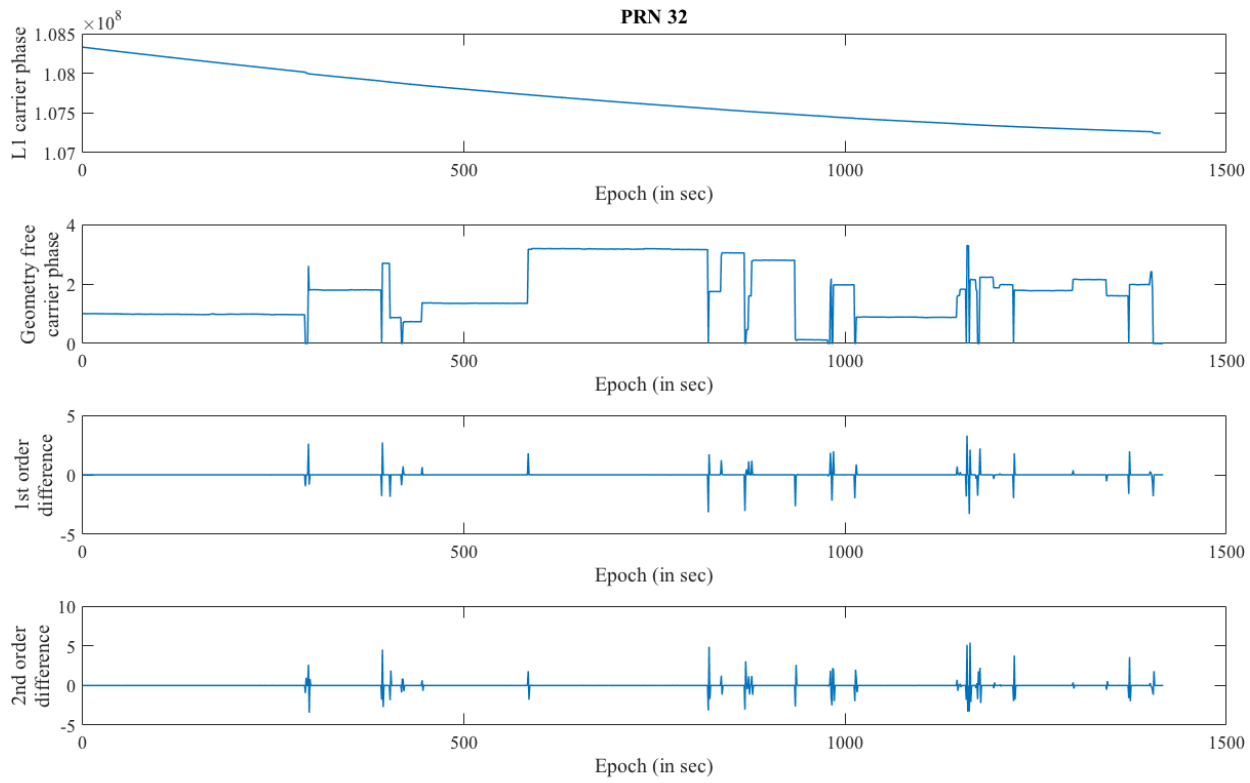


Figure 5.17 – Cycle slip effect on (1) the carrier phase measurement, (2) the geometry-free carrier, (3) 1st order and (4) 2nd order geometry-free carrier differences, for a high elevation satellite.

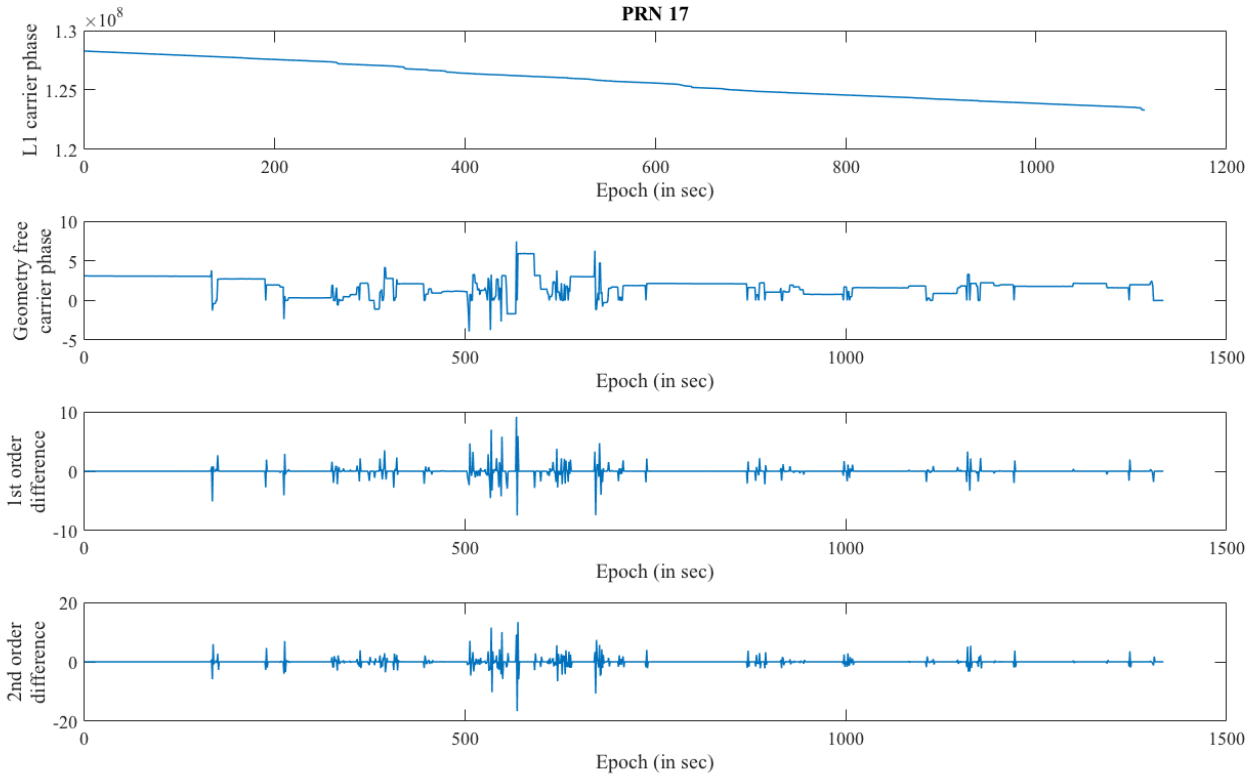


Figure 5.18 - Cycle slip effect on (1) the carrier phase measurement, (2) the geometry-free carrier, (3) 1st order and (4) 2nd order geometry-free carrier differences, for a low elevation satellite.

5.2.2 Multipath estimation

The multipath estimate is used in the novel algorithm to weight measurements (equation (5.6)), thus a real time estimate is required. It has been seen in section 4.2.1.3 that the C/A code multipath on L1 can be computed as:

$$MP_1 = R_1 - \Phi_1 - \frac{2f_2^2}{f_1^2 - f_2^2}(\Phi_1 - \Phi_2) \tag{5.13}$$

Similarly, the P(Y) code multipath on L2 can be estimated as:

$$MP_2 = R_2 - \Phi_2 - \frac{2f_1^2}{f_1^2 - f_2^2}(\Phi_1 - \Phi_2) \tag{5.14}$$

Figure 5.19 represents the C/A code multipath estimates on L1 and the P(Y) code multipath estimate on L2, as defined in equations (5.13) and (5.14). As discussed in section 4.2.1.3, the multipath estimate is affected by an offset due to the unknown number of integer ambiguity contained in the estimate. The effect of cycle slips on the carrier phase measurements can also clearly be seen as they create jumps in the estimate. Thus, cycle slips need to be detected and the multipath estimate needs to be averaged between cycle slips in order to assess the actual multipath and noise estimate.

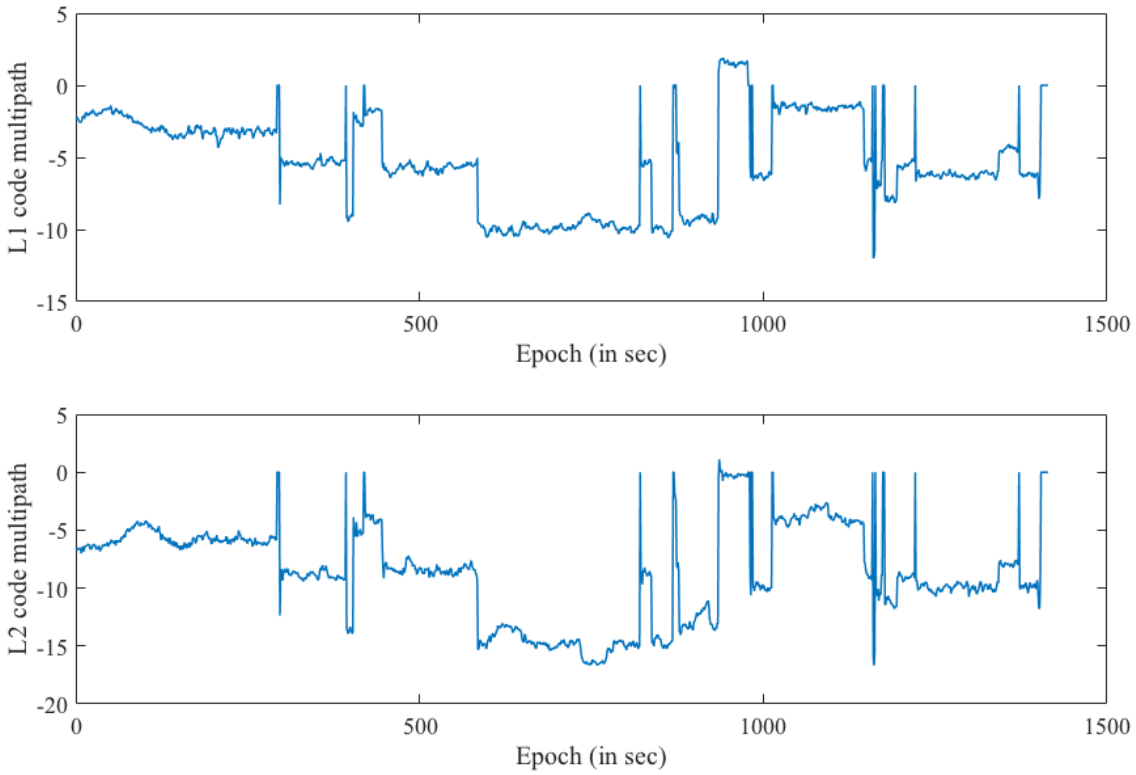


Figure 5.19 – L1 C/A code and L2 P(Y) code multipath estimates before averaging for PRN32

5.2.2.1 Multipath estimate average

To average the multipath estimate in real time, two types of moving average can be used: the simple moving average and the cumulative moving average. These are considered in turns below.

The simple moving average consists of averaging the measurements over a fixed number of epochs *M*, sometimes called a window, as seen in equation (5.15):

$$\overline{MP_1}(\mathrm{n}) = \frac{MP_1(\mathrm{n}) + MP_1(n-1) + \cdots + MP_1(n-(M-1))}{\mathrm{M}} \tag{5.15}$$

The cumulative moving average computes the average at the current epoch based on all the previous values:

$$\overline{MP_1}(\mathrm{n}) = \frac{MP_1(1) + MP_1(2) + \cdots + MP_1(n)}{\mathrm{n}} \tag{5.16}$$

In both cases, the average has to be reset every time measurements are unavailable (e.g. no satellite visibility) and when a cycle slip occurs.

Figure 5.20 compares the effect of different moving averages on the code multipath estimate. The values obtained using simple moving averages with windows of 5, 10, 30 and 60 seconds are compared to the value obtained using a cumulative moving average and to the actual mean value of the estimate. The data was selected free of cycle slips, between epochs 590 to 820 in Figure 5.19 and the mean value was computed based on the whole sample. The moving average using a fixed window only produces an average value once it has sufficient data to do so.

Figure 5.21 represents the error between the multipath averaged using the actual mean value and the multipath averaged using the different moving averages.

Figure 5.21 shows that as the averaging window size increases, the error decreases. Averaging using short windows (5 or 10 seconds) produces large errors. In fact, such averages are following the high-frequency variations of the multipath estimate due to the noise and multipath. By subtracting these average values to the multipath estimate, the variation of the multipath will be smoothed. As the multipath estimate is used in this thesis to quantify the level of noise and multipath, the use of such small windows will be counterproductive.

Longer window sizes (30 and 60 seconds) produce a smaller error than short windows, once they have enough data to compute the averaged value. Also, they follow the low-frequency variations of the multipath estimate that can be observed on a static receiver and thus mitigate this effect.

The cumulative average produces the smallest error and tends to the real value of the multipath estimate (i.e. error equals to zero) and conserves both the high and low-frequency variations observed on the multipath estimate.

Figure 5.22 represents the multipath estimate affected by a cycle slip. The mean value of the multipath was computed before and after the cycle slip, but it was assumed that the cycle slip was undetected and thus the averages were not reset.

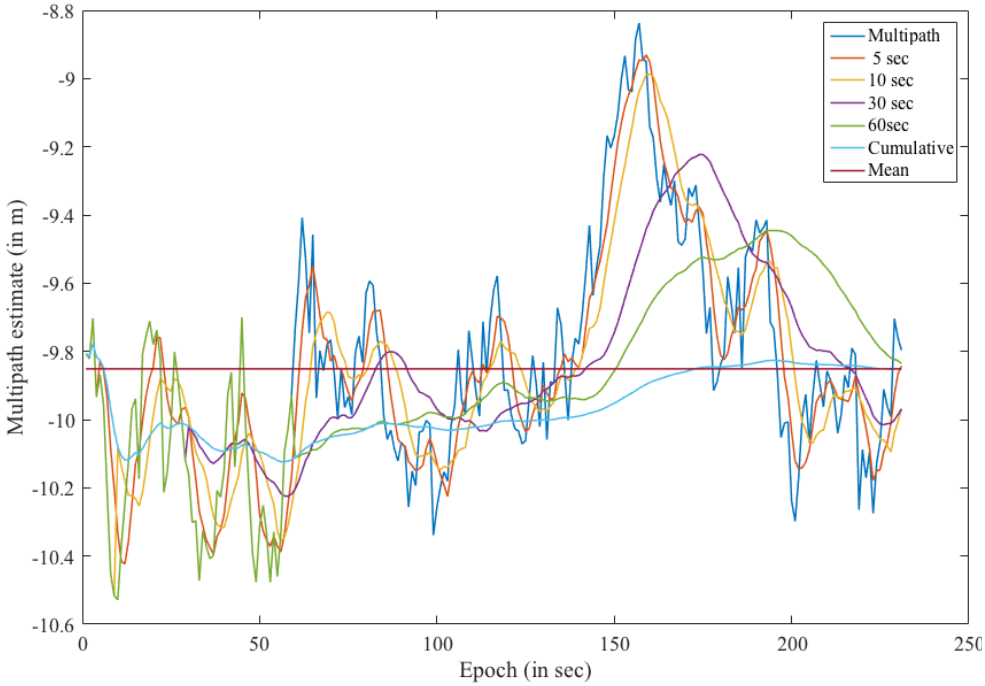


Figure 5.20 - Comparison of the code multipath mean value with the values of the moving averages for different averaging windows and for the cumulative moving average

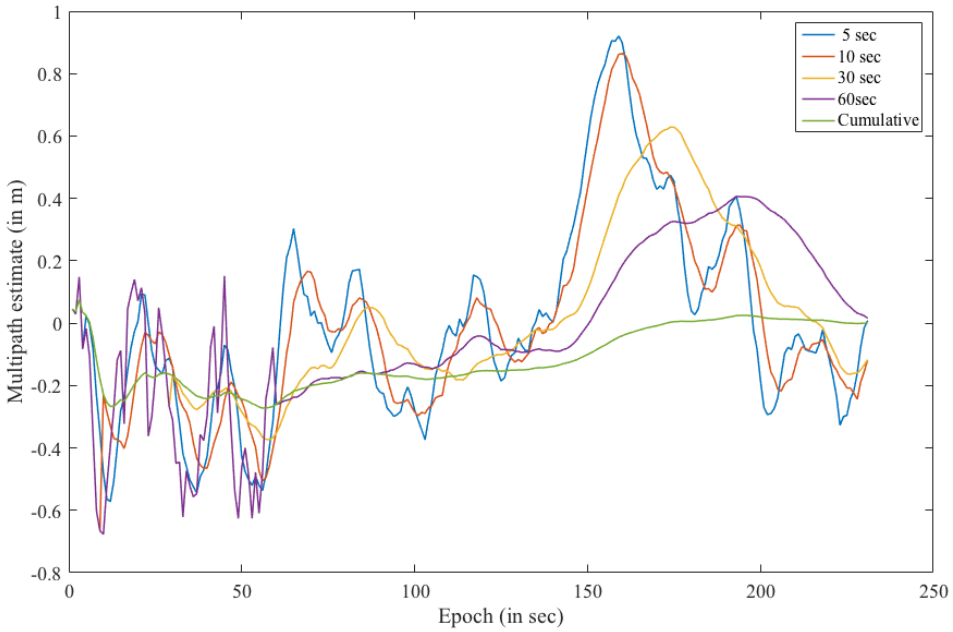


Figure 5.21 – Error between the actual averaged multipath value and the estimated averaged multipath using moving averages

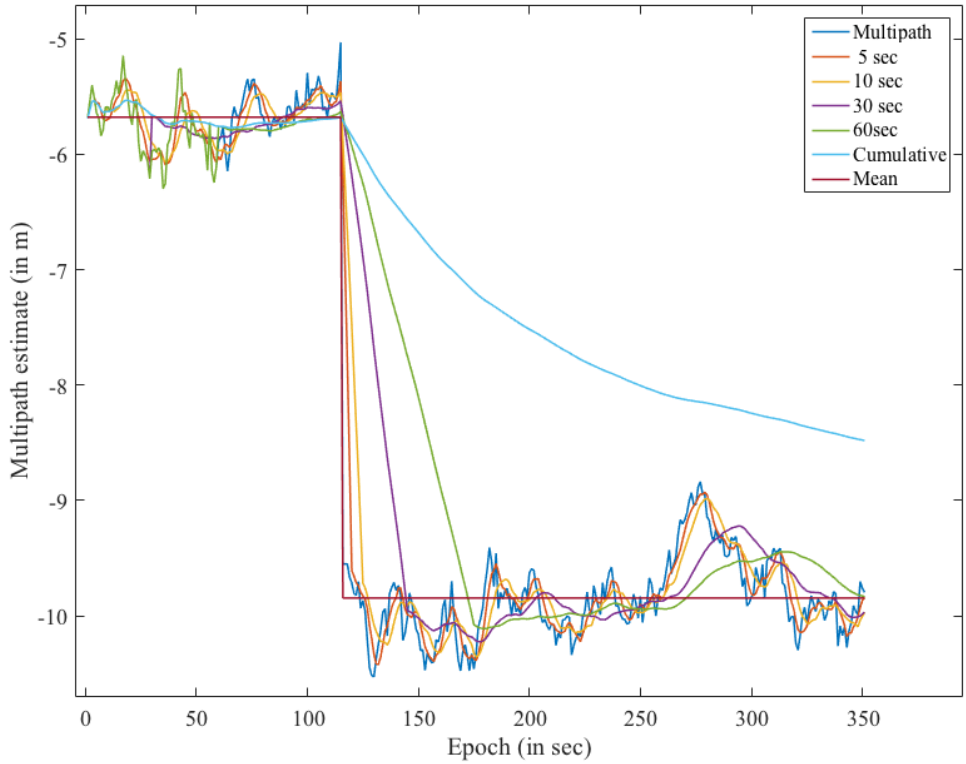


Figure 5.22 - Comparison of the code multipath mean value with the values of the moving averages for different averaging windows and for the cumulative moving average, in the presence of a cycle slip.

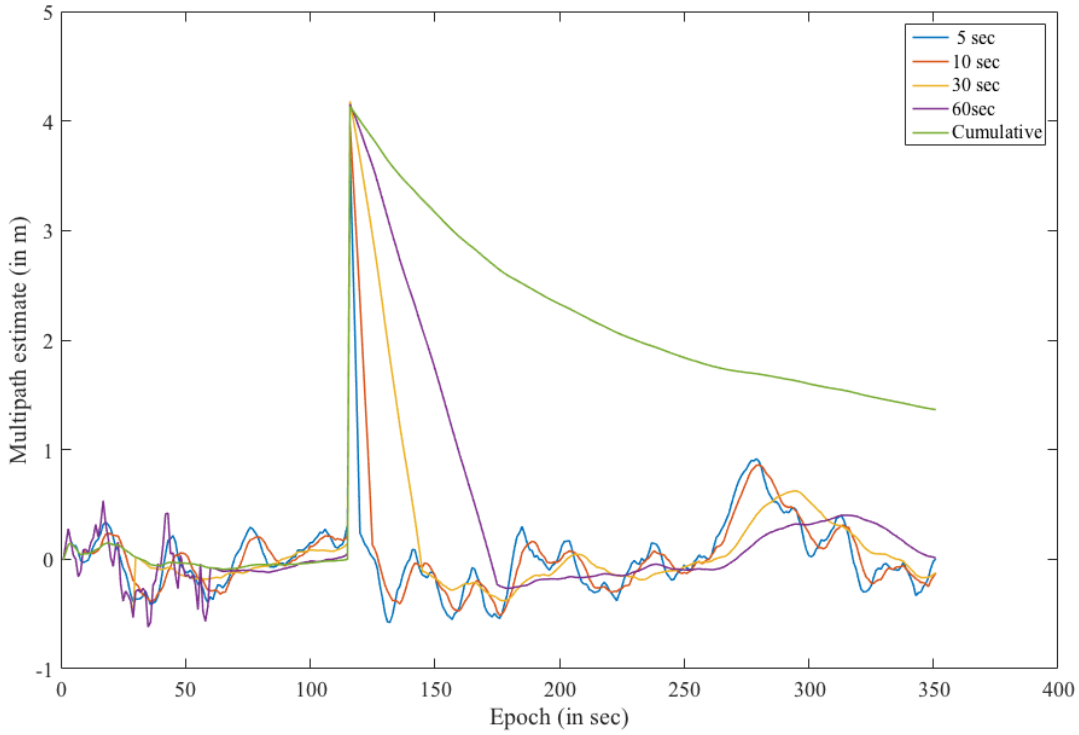


Figure 5.23 - Error between the actual averaged multipath value and the estimated averaged multipath using moving averages, in the presence of a cycle slip

It can be seen in Figure 5.22 that the moving averages using a fixed window recover faster from the missed cycle slip than the cumulative average. The shorter the window, the faster the average converges. This can be explained by the fact that the moving averages only use a limited amount of data and thus, after the duration of the window size, none of the data biased by the cycle slip is used to compute the average. In the case of the cumulative average, all the previous data are used and the cycle slip creates a large bias, which decreases slowly compared to the other moving averages.

Figure 5.23 represents the error in the averaged multipath. The effect of the cycle slip can clearly be seen as it creates a jump, corresponding to an error of approximately four metres. The shorter the window size, the faster is the decrease of the error. However, in the case of the cumulative average, the error decreases at a low rate.

In conclusion, moving averages using a fixed window size produces more accurate averages as the window size increases. However, as the average needs to be reset every time data is missing or when a cycle slip occurs, the window size is limited. Figure 5.24 was obtained by computing the duration of uninterrupted data (i.e. consecutive epochs with measurements from more than three satellites). It can be seen that less than 20% of the data is uninterrupted for 100 seconds, and thus using a moving average with a window size of 100 seconds may not provide a significant gain compared to a shorter window size.

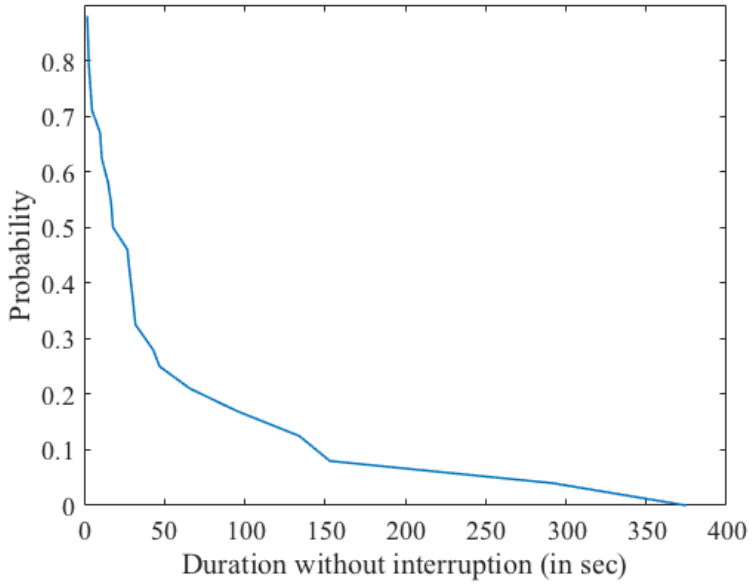


Figure 5.24 – Duration of uninterrupted data

The cumulative average has a higher availability than the moving average and can deliver an estimate over both short and long durations. However, it is very sensitive to cycle slips. Due to its greater availability and its capacity to preserve the variations of the multipath estimate, the cumulative average

was implemented in the POINT software requiring the reliable detection of cycle slips as discussed in section 5.2.1.

Figure 5.25 represents the multipath estimates on L1 and L2 (shown on Figure 5.19) after averaging. When the averaging was unavailable (i.e. right after cycle slip) or measurements were not available, the multipath estimate was set equal to zero.

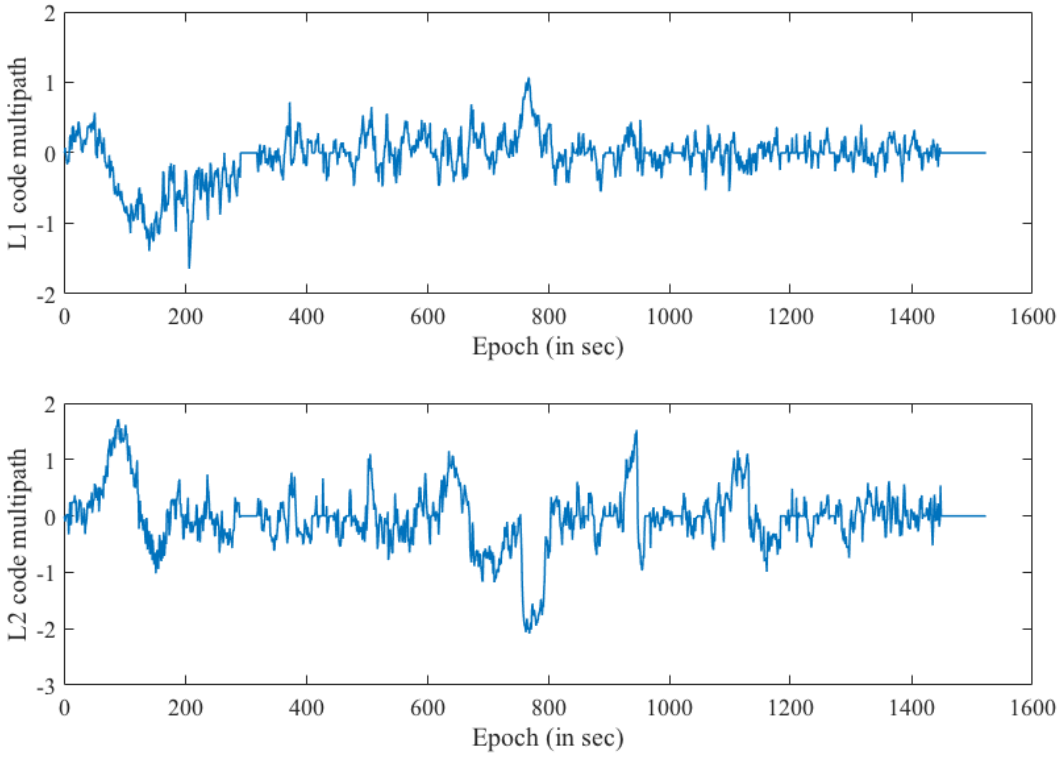


Figure 5.25 - L1 C/A code and L2 P(Y) code multipath estimates after averaging for PRN32

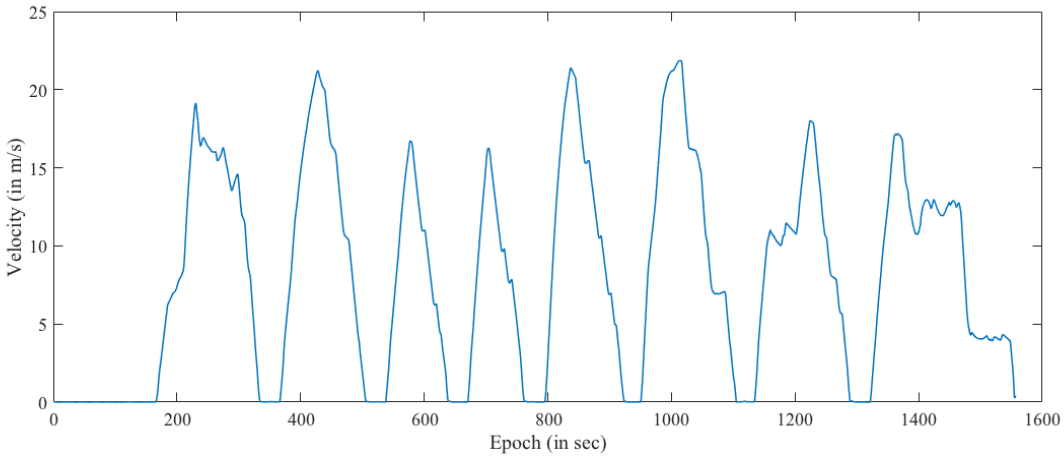


Figure 5.26 – Train speed

The code multipath and errors in Figure 5.25 can be correlated to the train speed in Figure 5.26. The code multipath and error tend to be higher when the train is static. In fact, while the train is moving,

only the stochastic part of the multipath error is visible. However, when the train stops, the deterministic part of the multipath error, which changes at low frequency, is visible.

Figure 5.25 and Figure 5.26 also imply that the locations of the stops are prone to signal reflections. This effect is illustrated in Figure 5.27 and Figure 5.28. The train journey starts from Four Oaks station from which the train departs at epoch 180. As seen in Figure 5.27, the large station canopy may reflect the GNSS signals and create the code multipath seen in Figure 5.25. Around epoch 950, the train stops at Sutton Coldfield station. As illustrated in Figure 5.28, this station is surrounded by trees and banks on each side of the track. It also has an overhead pedestrian bridge. These obstacles can be responsible for the code multipath seen in Figure 5.25 as the GNSS signals may be reflected on the banks or on the bridge before reaching the antenna.



Figure 5.27 – 3D view and picture of Four Oaks station, from (Google Earth 2012) and (Optimist on the run 2008)



Figure 5.28 – 3D view of Gravelly Hill station from (Google Earth 2012) and (Optimist on the run 2007)

5.2.2.2 Availability of the code multipath estimate

As discussed in the previous section, the code multipath estimate availability is affected by gaps in the data and cycle slips. As seen in equation (5.15), the code measurement and the carrier measurements on both L1 and L2 frequencies are required to compute the code multipath estimate. Thus if measurements on L2 are unavailable, the estimate cannot be computed and the multipath average has to be reset. The presence of cycle slips also required the average to be reset. To compute a new value of the average, at least two data points are needed, reducing further the estimate availability.

Table 5.1 presents the percentage of time that the code multipath estimates are available per satellite for the mainline train dataset. This percentage is equal to the number of epochs the estimate is computed divided by the number of epochs the code measurement on L1 is available. The availability of measurements on L2 is also presented and it corresponds to the number of epoch measurements on L2 are available over the number of epochs measurements on L1 are available.

Table 5.1 – Code multipath estimate availability per satellite

Satellite PRN	Code multipath	L2 availability	Elevation	Orthogonal distance
1	93%	98%	77° - 88°	0.01 – 0.13
3	67%	86%	1.6° - 7°	0.38 – 0.46
11	92%	98%	68° - 80°	0.07 – 0.19
14	75%	95%	28° - 32°	0.49 – 0.56
17	75%	94%	15° - 24°	0.58 – 0.63
19	77%	93%	22° - 33°	0.29 – 0.37
20	89%	97%	40° - 51°	0.28 – 0.38
22	47%	80%	4° - 12°	0.69 – 0.76
23	60%	83%	2° - 3°	0.06 – 0.09
28	77%	96%	26° - 30°	0.65 – 0.71
32	92%	98%	67° - 78°	0 – 0.04

Despite a slightly lower availability for low elevation satellites, L2 measurements are available over 80% of the time when L1 is available. Thus, cycle slips and gaps in the data are the main factors affecting the code multipath availability, especially for low elevation satellites. The estimate availability is higher for high elevation satellites (PRN 1, 11 and 32) than for low elevation satellites (PRN 3, 22, 23). It reaches as low as 47% for satellite 22 which has an elevation angle between 4° and 12° and is additionally far from the track (i.e. large orthogonal distance).

Thus, due to its limited availability, the multipath estimate may not be a good quality indicator for low elevation satellites and the measurements from such satellites should be weighted down based on their elevation, as discussed in section 5.1.2.

5.2.2.3 Characterization of multipath

In order to characterise the code multipath, the estimates computed for each satellite of the suburban mainline dataset are aggregated and their distributions analysed. The level of multipath, computed as the Root Mean Square Error (RMSE), is of 0.45 m for the C/A code on L1 and 0.68 m for the P(Y) code on L2.

Figure 5.29 represents the Probability Density Function (PDF) of the code multipath estimate on L1 and Figure 5.30, the probability density function of the estimate on L2. Both estimates appear to have normal-like distribution.

Figure 5.29 and Figure 5.30 show that a t-location-scale model better represents the code multipath estimate than the normal distribution. The t-location-scale distribution, also called Student's t-distribution (probability distribution given in equation (5.17)), is symmetric and bell-shaped like the normal distribution but has heavier tails, meaning it is more prone to outliers (i.e. multipath with large values are more likely to occur).

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{\left(\frac{x-\mu}{\sigma}\right)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \tag{5.17}$$

where $\Gamma(.)$ is the gamma function, μ is the location parameter, σ is the scale parameter and ν is the shape parameter.

The code multipath estimate large values, due to reflections especially when the receiver is static, are therefore better represented by the t-location-scale distribution. This distribution is described by three parameters: the degree of freedom, the location parameter and the scale parameter. However, in order to compromise between the goodness-of-fit of the model and its complexity, the normal distribution is selected to represent the multipath. The normal distribution is described using only two parameters, its mean μ and standard deviation σ. Figure 5.31 represents the code multipath estimate on the L1 distribution and is used to assess the goodness of fit of the normal distribution.

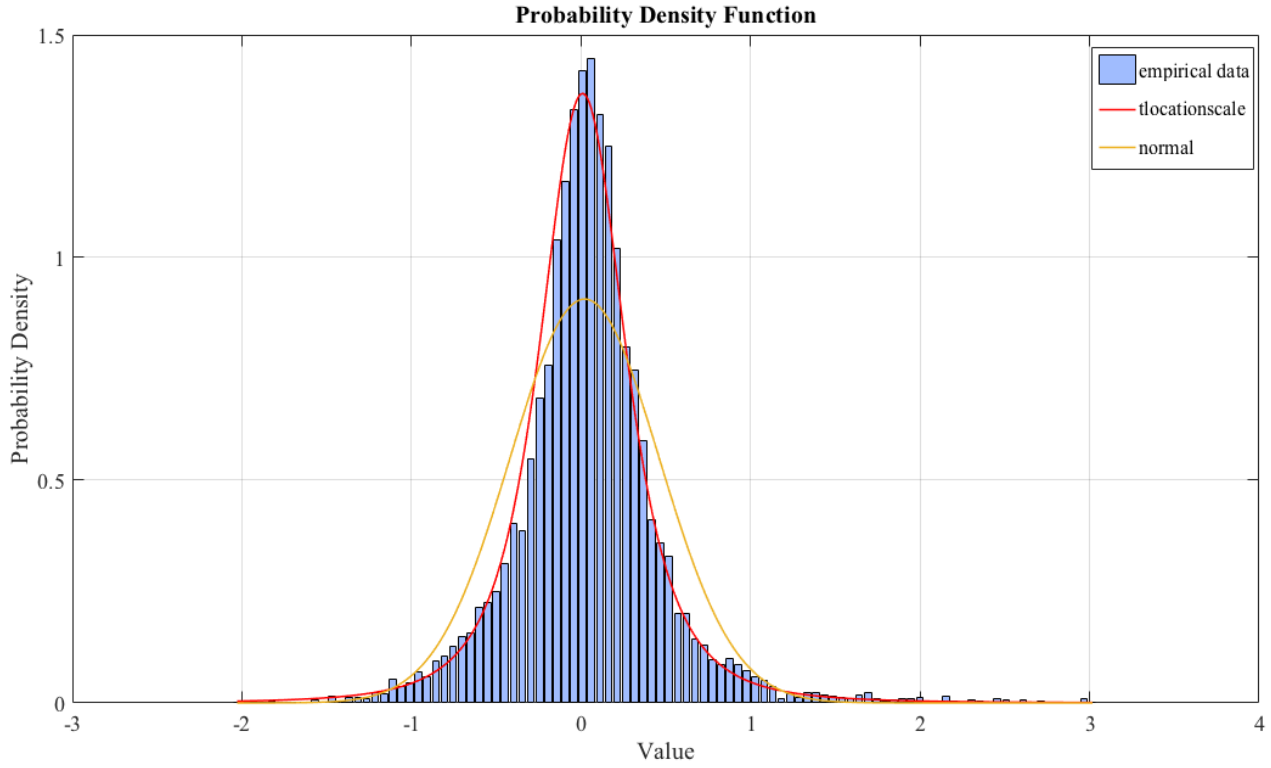


Figure 5.29 – C/A code multipath estimate on L1 probability density function

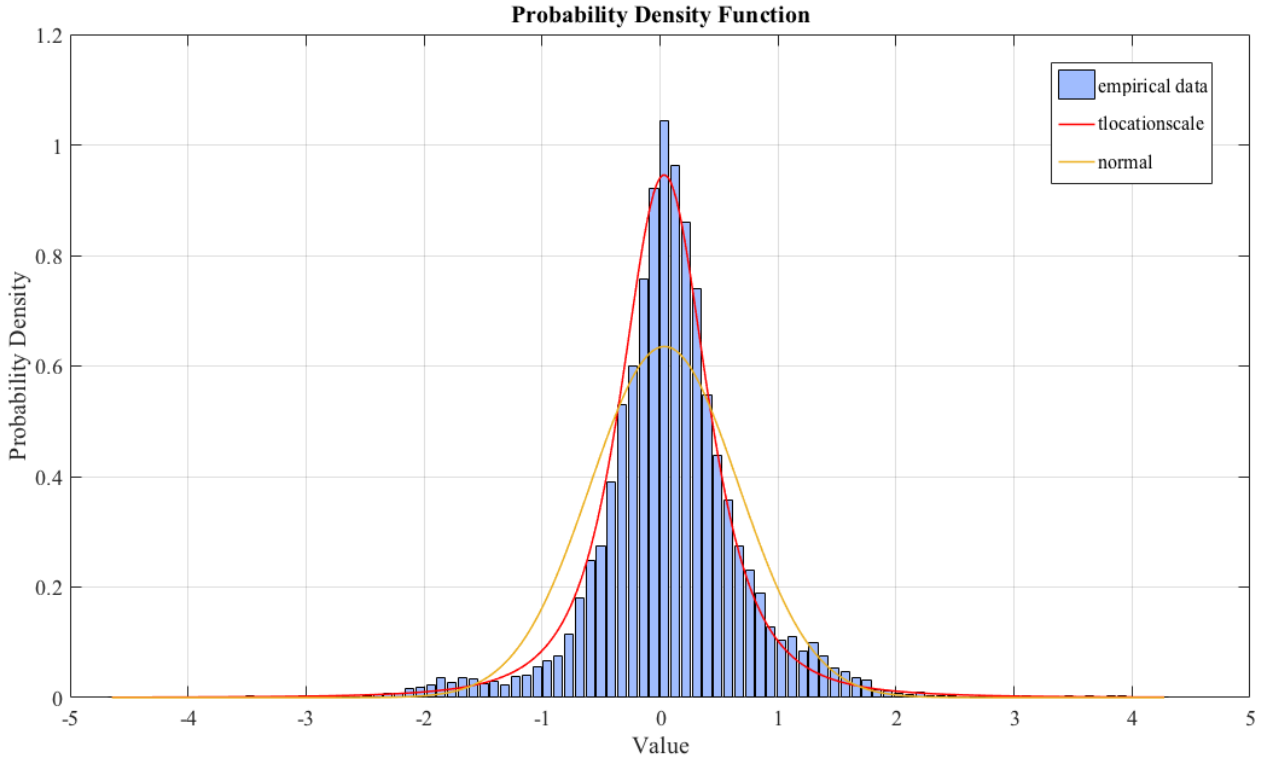


Figure 5.30 – P(Y) code multipath estimate on L2 probability density function

The Q-Q plot, in the top-right corner, is used to compare both the empirical data to the theoretical distribution. If the two distributions are identical, the Q-Q plot is expected to follow the line y = x. It can be seen that this is true for multipath estimate between ± 0.5 m. However, larger multipath estimates differ from the straight line y = x, which implies that the empirical data has heavier tails than the normal distributions and that these tails are not represented by the normal distributions. However, multipath estimates larger than 0.5 m in absolute value represent less than 10% of the dataset, and it is assumed that the normal distribution is a good representation of the multipath estimate.

The Cumulative Distribution Functions (CDF) of both the model and the data are represented on the bottom-left corner of the figure. These distributions are very similar, showing that the normal distribution is an appropriate model for the multipath estimate. The P-P plot, in the bottom-right corner, compares the two cumulative distribution functions. It can be seen that the data passes through the point (0.5, 0.5), which implies that both distributions have the same mean value.

Normal distributions are then fitted to multipath estimate for each satellite in order to validate the hypothesis that the multipath estimate follows a normal distribution. The means and standard deviations of these distributions are presented in Table 5.2. The root-mean-squared error (RMSE) between the empirical data and theoretical cumulative distribution is given as an indication of the model goodness of fit.

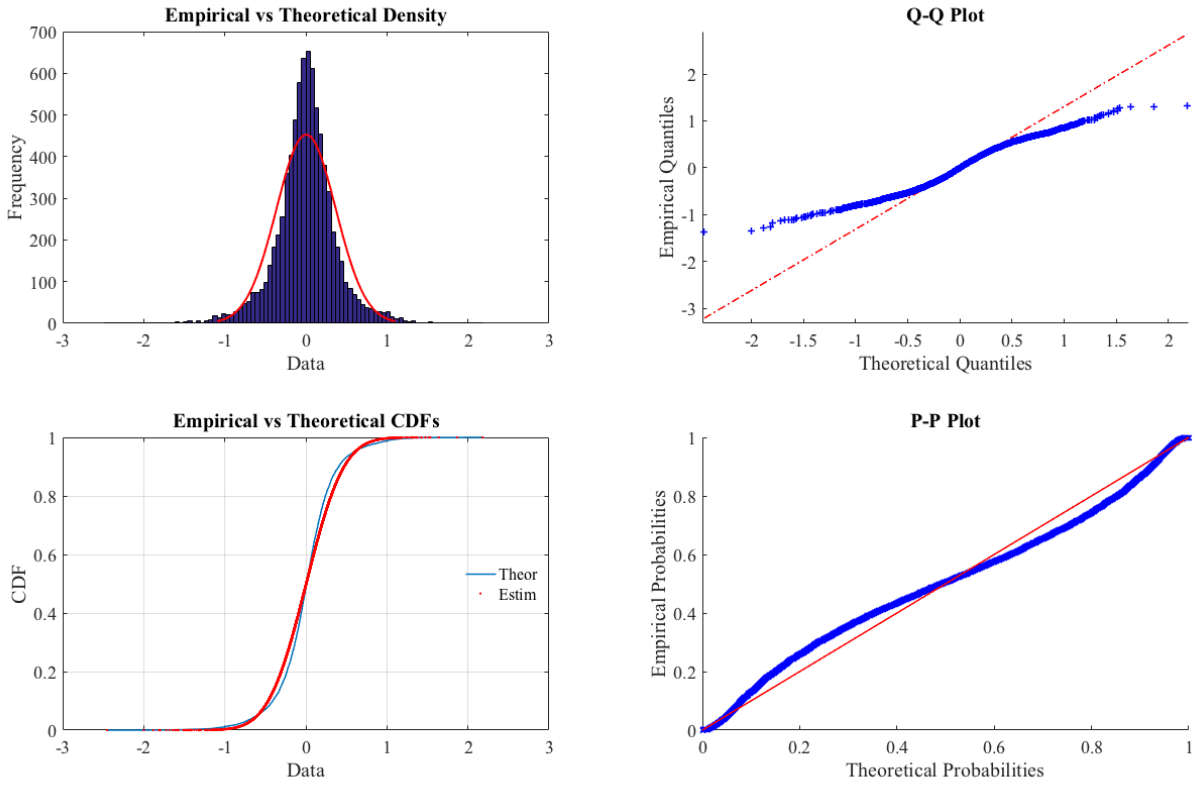


Figure 5.31 – Characterization of the normal distribution fit to the code multipath estimate on L1 for all satellites

Table 5.2 – Mean value μ and standard deviation σ of the normal distribution used to represent the code multipath estimate distribution per satellite and for all satellites

Observable	MP1				MP2			
Satellite	RMSE	μ	σ	GOF	RMSE	μ	σ	GOF
All	0.446	0.028	0.444	0.052	0.685	0.040	0.684	0.059
1	0.334	-0.042	0.331	0.010	0.580	0.179	0.552	0.041
3	0.621	0.145	0.605	0.040	1.285	-0.114	1.283	0.135
11	0.284	-0.035	0.281	0.026	0.586	0.026	0.585	0.0411
14	0.408	0.003	0.408	0.060	0.833	0.054	0.832	0.081
17	0.690	0.196	0.662	0.107	0.776	0.135	0.764	0.107
19	0.436	0.036	0.435	0.038	0.488	0.062	0.485	0.041
20	0.510	0.059	0.507	0.036	0.665	-0.103	0.658	0.044
22	0.325	0.004	0.326	0.025	0.565	0.011	0.566	0.018
23	0.318	-0.080	0.309	0.053	0.785	0.070	0.786	0.050
28	0.515	0.147	0.493	0.067	0.791	0.080	0.787	0.058
32	0.360	-0.070	0.353	0.053	0.563	-0.036	0.562	0.044

It can be seen that, on average, the mean value obtained is close to 0 and the standard deviation varies between approximately 0.3 and 0.7 for the C/A code on L1, and 0.5 and 1.3 for the P(Y) code on L2. Based on the cumulative density functions RMSE, the normal distribution appears to be a good representation of the code multipath estimate for most satellites. The goodness of fit is lower for satellite 17 which is related to the presence of large multipath estimates experienced while the train is waiting to depart from Four Oaks. These estimates may be caused by the station roof (see Figure 5.27) located between satellite 17 and the antenna.

In conclusion, based on the results presented in Table 5.2 for all the satellites, the L1 C/A code and L2 P(Y) code multipath estimates are assumed to follow normal distributions with the following parameters:

$$MP1 \sim N(0, 0.45) \tag{5.18}$$

$$MP2 \sim N(0,0.68) \tag{5.19}$$

These definitions of the code multipath estimate are then used to weight the measurements in the novel heading-based method, as explained in section 5.1.3.

5.2.2.4 Multipath based weighting

Based on the previously determined code multipath distribution, the multipath weighting method mentioned in section 4.3.2.3 is further refined. It is proposed to weight the measurements based on the

code multipath estimate using an exponential term as done in the stochastic elevation model developed by (Elsobeiey & El-Rabbany 2010) in equation (4.76).

$$\sigma_i = \sigma_0\left(1 + A * e^{B*|multipath_i|}\right) \tag{5.20}$$

The values of the constants A and B are selected to generate σ_i comparable to the other weighting methods presented in section 4.3.2, i.e. ranging from approximately $1\times\sigma_0$ and $22\times\sigma_0$. It is decided to assign a standard deviation equal to $22\times\sigma_0$ to code multipath equals to 2 m. This choice is made on the basis that the probability to have a code multipath estimate larger than 2 m is 0.45% (from the code multipath cumulative distribution), representing extreme cases. A standard deviation of $2\times\sigma_0$ is assigned to code multipath equals to 0.5 m, as it was observed that multipath estimates over 0.5 m do not follow a normal distribution. These assumptions lead to A being equal to 0.36 and B equal to 2.03. The associated weights are presented in Figure 5.32.

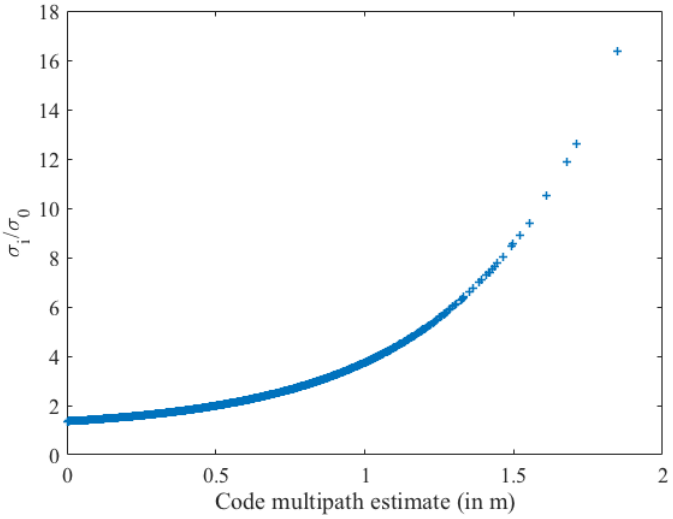


Figure 5.32 - Measurement error standard deviation as a function of the code multipath estimate

The impact of this new multipath-based weighting model on the estimated position is assessed in chapter 6.

5.2.3 Hatch filter

The hatch filter, presented in section 4.2.2.3, was added in the GNSSReceiver file. This filter is used to smooth the code measurements using the accurate but ambiguous carrier measurements. The hatch filter was implemented according to equation (5.21) with the possibility of adjusting the filter time constant N.

$$\widehat{R}(k) = \frac{1}{n}R(k) + \frac{n-1}{n}\left[\widehat{R}(k-1) + \left(\Phi(s,k) - \Phi(k-1)\right)\right] \tag{5.21}$$

Where $n = k$ as long as $k < N$, and $n = N$ when $k \geq N$. The constant N relates to the bandwidth of the filter, larger values of N giving more smoothing.

In order to compute the carrier smoothed code $\hat{R}$, the current code and carrier phase measurements are required as well as the previous carrier phase measurement and carrier smoothed code. Also, as discussed in section 4.2.2.3, cycle slips need to be detected, otherwise the difference between the phase measurements is equal to the range variation plus an integer number of cycle.

5.2.3.1 Filter time constant

As stated in section 4.2.2.3, as the value of N increases, the output noise variance decreases but the duration of the transient increases. The carrier-smoothed code is computed for time constants ranging from 5 seconds to 100 seconds. In order to assess the effect of the smoothing, the code minus carrier observable is then computed using the carrier-smoothed code. This difference removes the range and mitigates the clock error and the tropospheric delay, leaving the code multipath and the ionospheric delay. Figure 5.33 and Figure 5.34 represent the code minus carrier observable for different values of the filter time constant, respectively for a static receiver and a dynamic receiver.

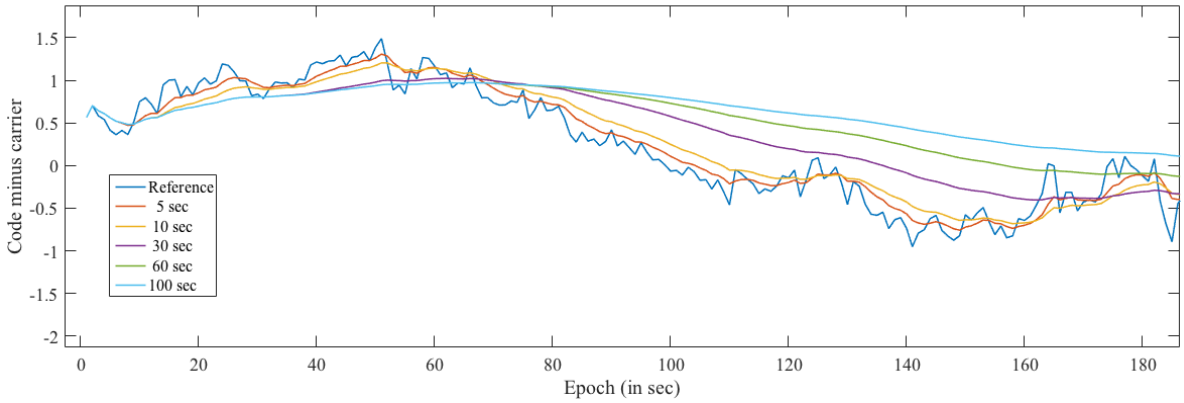


Figure 5.33 – Carrier-smoothed code minus carrier observable, for different value of the Hatch filter time constant, when the receiver is static

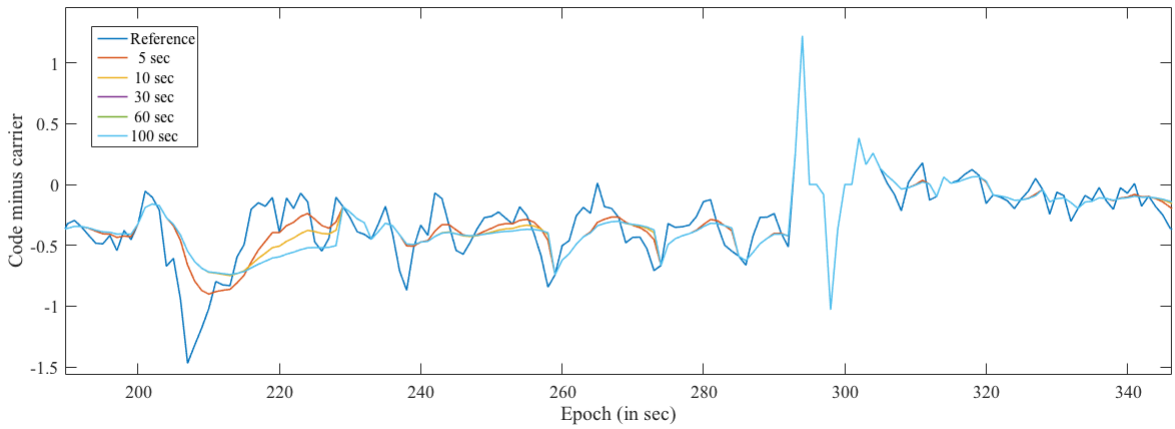


Figure 5.34 – Carrier-smoothed code minus carrier observable, for different value of the Hatch filter time constant, when the receiver is dynamic

Time constants of 30 seconds and longer have a significant effect on the code measurements when the receiver is static. However, when the receiver is moving, their effect is limited, as illustrated in Figure 5.34. The values produced by the time constants equal to 30 seconds, 60 seconds and 100 seconds overlap, which means that the filter was reset before the index k in equation (5.21) reached the time constant. This effect can be related to the frequency of data gaps illustrated in Figure 5.24, as only 50% of the data are uninterrupted for more than 30 seconds (without accounting for cycle slips). As the filter reset does not influence the smoothed code estimate, a large time constant is chosen as it reduces the measurement noise especially when the receiver is static.

The carrier-smoothed code is then used to estimate the code multipath, as defined in equation (5.13). Figure 5.35 represents the code multipath estimate distributions for different Hatch filter time constants and the fitted normal distribution with its associated mean and standard deviation.

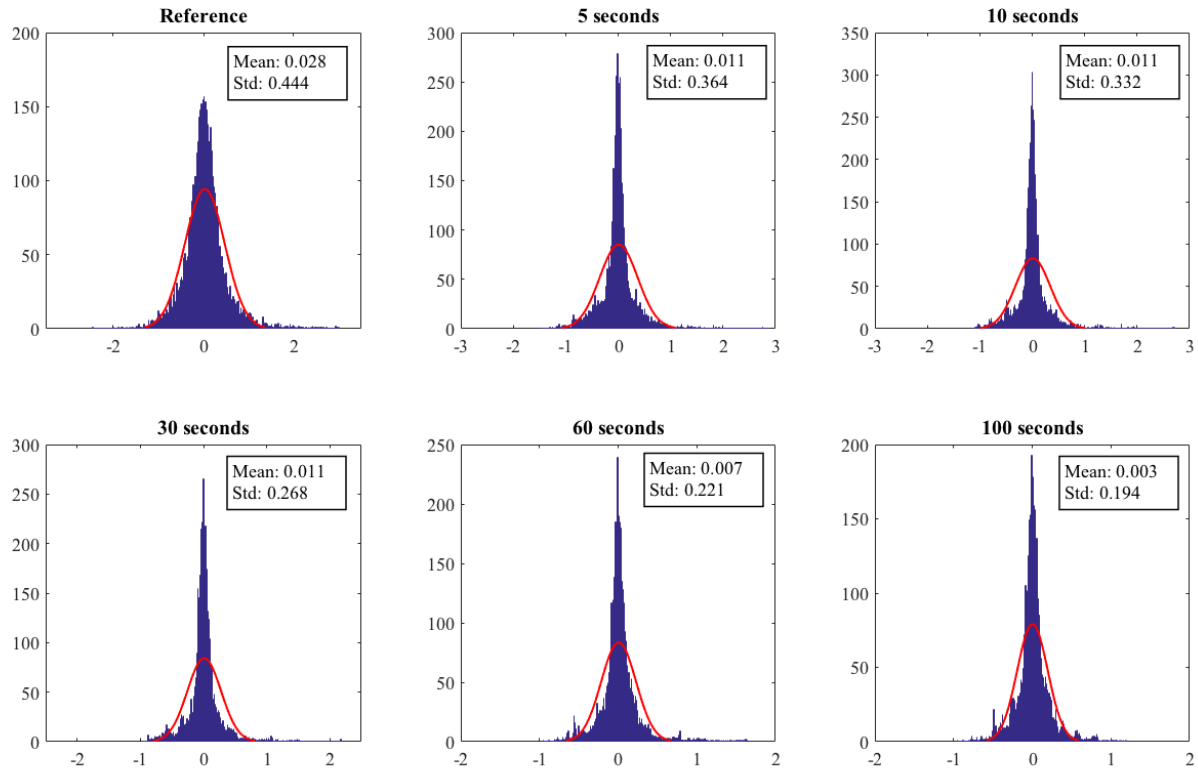


Figure 5.35 – C/A code multipath estimate on L1 distribution for different values of the Hatch filter time constant

The mean values remain relatively small and close to zero in all five cases. However, as the time constant increases, the multipath estimate standard deviation significantly decreases. A time constant of 5 seconds already reduces the standard deviation by 18%. Extended the time constant to 60 seconds reduces the standard deviation by an additional 32%. These values are however lower than the theoretical gain predicted by (Lo Presti & Visintin 2015) in equation (4.59), as seen in Table 5.3, which

can be explained by the fact that the filter has to be frequently reset due to gaps in the data and cycle slips.

Table 5.3 – Effect of the Hatch filter time constant on the code multipath estimate standard deviation

Hatch filter time constant	Standard deviation	Standard deviation gain	Theoretical standard deviation	Theoretical gain
5	0.364	18%	0.151	66%
10	0.332	25%	0.105	76%
30	0.268	40%	0.061	86%
60	0.221	50%	0.044	90%
100	0.194	56%	0.035	92%

5.2.3.2 Availability

In order to compute the carrier-smoothed code, the code measurement, the previous code measurement, the carrier phase and the previous carrier phase values are required. These values are available as long as there are no gaps in the data and no cycle slips are detected. The availability of the carrier-smoothed code was computed for each satellite and it represents the number of epochs the smoothed code is computed over the number of epochs where the C/A code measurement on L1 is available.

Table 5.4 – Carrier-smoothed code availability per satellite

Satellite PRN	Carrier smoothed code	Elevation	Orthogonal distance
1	94%	77° - 88°	0.01 – 0.13
3	70%	1.6° - 7°	0.38 – 0.46
11	93%	68° - 80°	0.07 – 0.19
14	79%	28° - 32°	0.49 – 0.56
17	78%	15° - 24°	0.58 – 0.63
19	79%	22° - 33°	0.29 – 0.37
20	90%	40° - 51°	0.28 – 0.38
22	51%	4° - 12°	0.69 – 0.76
23	61%	2° - 3°	0.06 – 0.09
28	82%	26° - 30°	0.65 – 0.71
32	93%	67° - 78°	0 – 0.04

The carrier-smoothed code availability decreases with the satellite elevation and the orthogonal distance. As seen previously in section 0, this effect is related to low elevation satellites being more prone to data gaps and cycle slips due to their higher likelihood of attenuation or blockage by obstacles.

In conclusion, the carrier-smoothed code may reduce the noise on the measurements coming from high elevation satellites but its effect is limited on low elevation satellites. The effect of low elevation measurements on the positioning solution can, however, be limited using weighting techniques. The combination of the carrier-smoothed code with the different weighting techniques presented in sections 4.3 and 5.1 is reviewed in chapter 6.

5.3 Summary

In this chapter, a novel weighting method based on the geometry of railway tracks was developed. It assumes that satellites aligned with the track are less likely to be affected by noise and multipath than satellites perpendicular to the tracks, due to the presence of trackside obstacles, such as buildings and trees. This new model was parameterised using the carrier to noise ratio and the multipath estimate, making the method adaptive to its environment.

Additional functions implemented in the POINT software in order to assess the positioning performance of the novel weighting method were presented. The importance of cycle slips detection was highlighted. Code multipath was characterised as having a normal distribution with a mean of zero and a standard deviation of 0.45 m and 0.68 m for L1 C/A code and for L2 P(Y) code respectively. The challenge posed by averaging the multipath estimate in real time was discussed and a cumulative moving average was chosen. Finally, the effect of the carrier smoothed code on the multipath estimate was assessed and proved to significantly reduce the code multipath standard deviation.

While this chapter focused on the measurement domain by characterising the code multipath error and the effect of the Hatch filter on this error, the following chapter evaluates the impact of multipath mitigation techniques on the position itself.

Chapter 6

Characterization of the Novel Heading-based Weighting Method Positioning Performance

Following the characterization of the effect of multipath on the code measurements (Chapter 5) and the review of the different techniques that can be used to reduce its impact on the positioning solution (Chapter 4), this chapter firstly quantifies the improvement of positioning accuracy from the current code multipath mitigation techniques. These techniques include the use of the carrier-smoothed code along with weighting techniques based on the satellite elevation angle, C/N0 value or multipath estimate. Secondly, this chapter characterises the performance of the novel heading-based weighting method developed in section 5.1 and compares to the current methods.

The results presented in this chapter are based on the suburban mainline dataset. The accuracy statistics are based on a reference trajectory generated as explained in section 6.1.1. Furthermore, this section presents the positioning solution obtained with the traditional least squares (i.e. non-weighted). Section 6.2 presents and discusses the results from the application of carrier-based smoothing. The performances of the elevation-based and C/N0 based weighting methods, as well as the multipath–based weighting, are discussed in section 6.3. Finally, section 6.4 presents the positioning performance of the novel heading based weighting method.

6.1 Preliminary analysis

6.1.1 Reference trajectory

A reference trajectory is required to compare the estimated position with the actual position of the train. In order to compute this reference trajectory, raw dual-frequency GNSS measurements are combined with the INS data collected by the Novatel SPAN receiver (section 1.4.3.1) using Inertial Explorer software (Novatel 2014). This software integrates data from INS sensors with a GNSS post-processed solution, using either a loose coupled or tightly coupled processing.

The GNSS measurements are processed in differential mode (section 3.1.2.3) using measurements from an Ordnance Survey Net reference station in Lichfield. The distance between the station and the receiver

varies from approximately 15 km at Four Oaks to 27 km at Birmingham New Street. The integration was performed in the measurement domain (tight coupling) in order to ensure a maximum performance in terms of availability. Figure 6.1 shows the estimated standard deviations ranging from 3 cm up to 50 cm, reflecting the uncertainty of the positioning solution.

The higher values of the standard deviation in Figure 6.1 (e.g. around time 476950) can be correlated to epochs with low or no satellite visibility in Figure 6.2. During these periods, the position is estimated only using the INS, and the standard deviation increases with the time due to the sensor drift error, as explained in section 3.1.3.1. The positioning solution is then smoothed by Inertial Explorer using both forward (i.e. sequential order) and reverse filter, which explains the bell-shaped peaks seen in Figure 6.1.

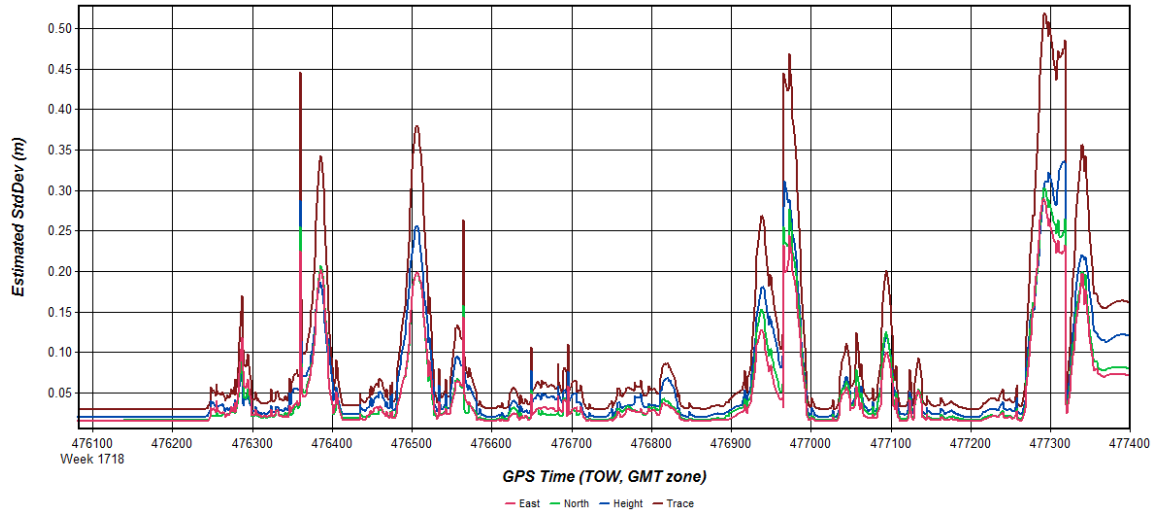


Figure 6.1 – Tightly coupled positioning solution standard deviation as estimated by Inertial Explorer

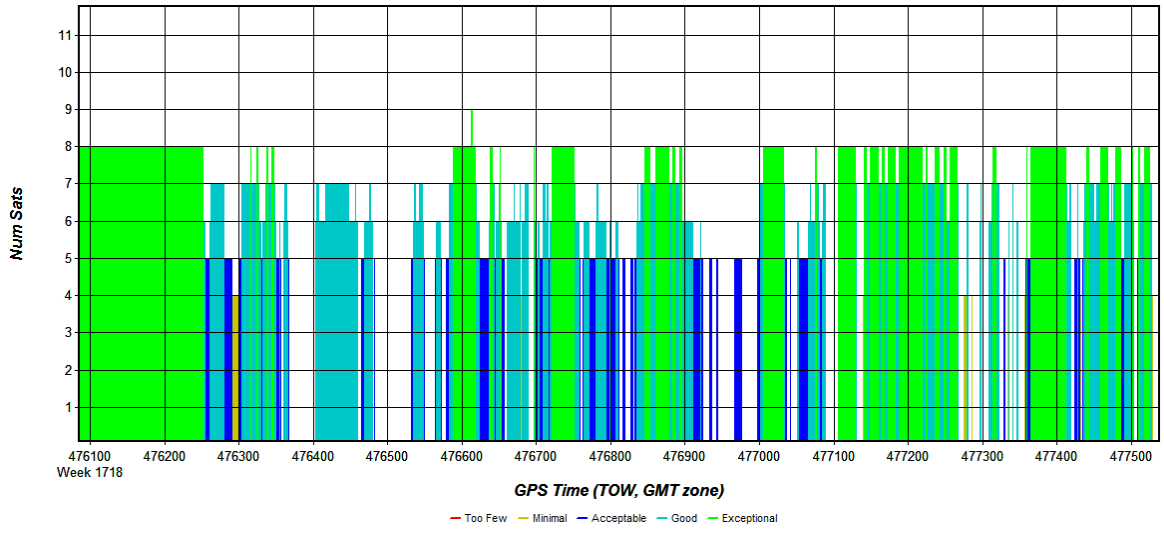


Figure 6.2 – Number of visible satellites as a function of the epoch

The results presented in this chapter are based on the comparison of positioning solutions obtained using single point, dual frequency GPS positioning, which are expected to have an accuracy ranging from 3 and 5 m in good geometry (section 3.1.2). Thus, based on the standard deviations in Figure 6.1, the reference trajectory obtained by the tight integration of the GNSS and INS measurements is assumed accurate enough to quantify the positioning accuracy from the different weighting methods.

6.1.2 Positioning solution

6.1.2.1 Error mitigation

As explained in section 4.1, GNSS measurements are affected by different error sources which need to be mitigated in order to limit their impact on the positioning solution. Table 6.1 summarises how the different error sources were mitigated in the subsequent analysis.

Table 6.1 – Error sources mitigation

Error source	Mitigation technique
Ionosphere	If dual frequency measurements are available, the ionospheric-free combination is used (equation (4.10)). Otherwise, the Klobuchar model (section 4.1.3.2) is used to estimate the ionospheric delay.
Troposphere	The troposphere delay is computed using EGNOS tropospheric correction model (Penna et al. 2016)
Satellite clock	The satellite clock error is obtained from the navigation message
Receiver clock	The receiver clock error is estimated in the least squares algorithm

6.1.2.2 Least squares positioning solution

To analyse the effect of the weighting methods, a non-weighted solution is computed first using the traditional least squares algorithm presented in section 4.1.4.1.

Figure 6.3 represents the Northing, Easting and Up positioning errors of the non-weighted solution. These errors reach values as high as 60 m, due mainly to poor satellite geometry. As shown in equation (4.24), the positioning solution accuracy depends on both the measurement errors and the satellite geometry, represented by the dilution of precision (DOP). The position (PDOP), horizontal (HDOP) and vertical (VDOP) values of the DOP are presented in Figure 6.4. They reach values as high as 40 m and are highly correlated to the number of visible satellite shown in Figure 6.5. In summary, a weak geometry leads to a high DOP value and in turn, to a large positioning error.

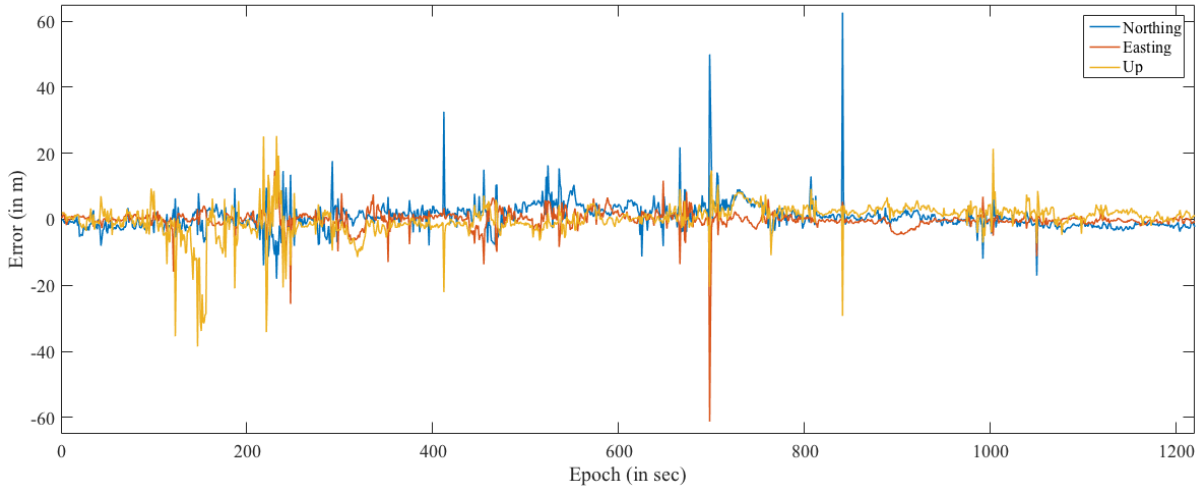


Figure 6.3 – Traditional least square solution North, East and Up positioning errors

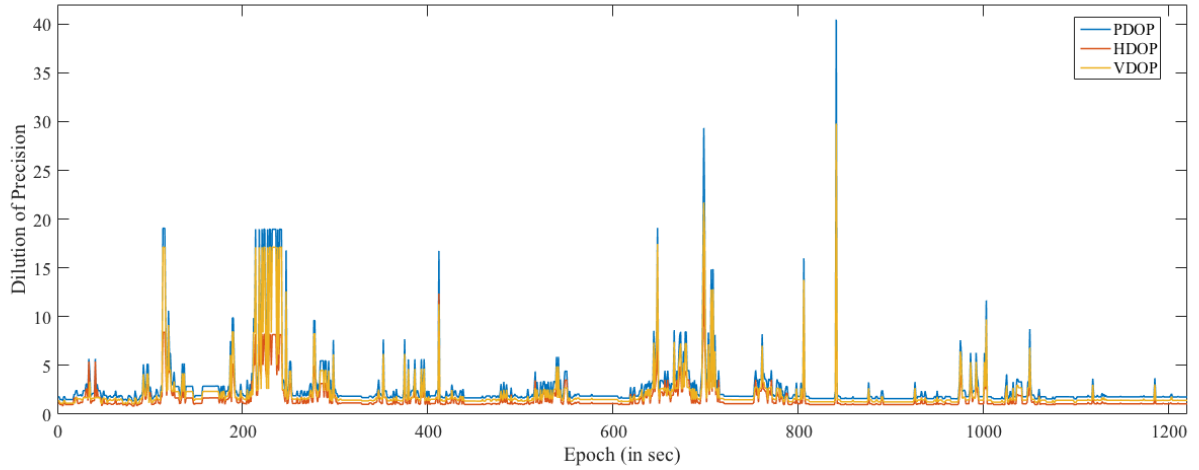


Figure 6.4 – Position (PDOP), Horizontal (HDOP) and Vertical (VDOP) dilution of precision values

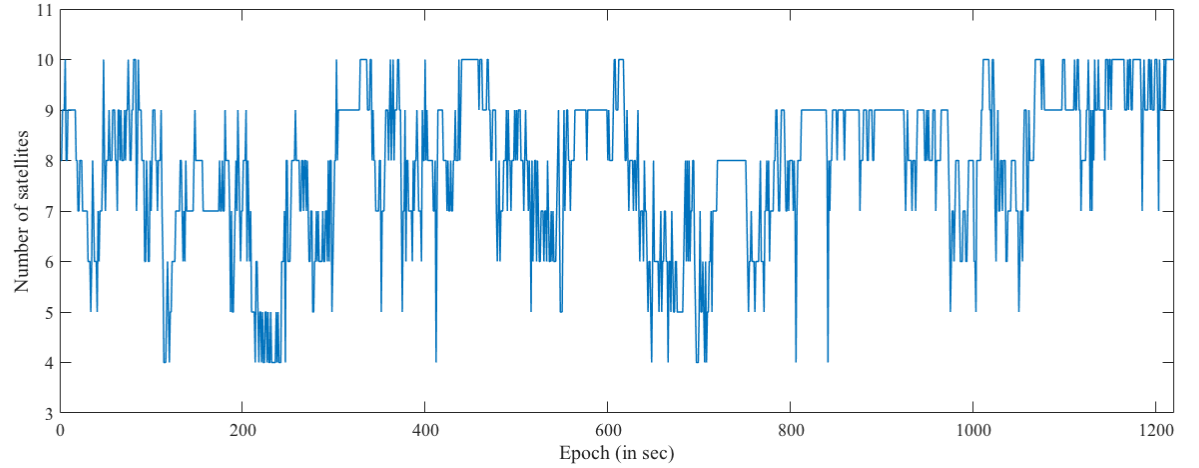


Figure 6.5 – Satellite visibility

As seen in Figure 6.5, the number of visible satellites along the track varies significantly. The large values of DOPs are related to epochs where only 4 or 5 satellites are visible (e.g. around epoch 200 or 700). The limited visibility is due to the presence of obstacles on the track, such as cuttings or overhead bridges, as shown in Figure 6.6, Figure 6.7 and Figure 6.8.



Figure 6.6 – Presence of a cutting between epoch 214 and 242, from (Google Earth 2012)

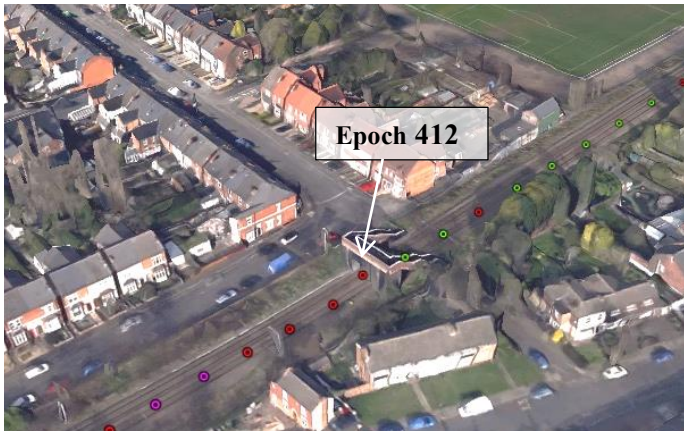


Figure 6.7 – Pedestrian bridge above the track at epoch 412, from (Google Earth 2012)



Figure 6.8 – Large highway bridge at epoch 841, from (Google Earth 2012)

6.1.2.3 Characterization of the positioning error

In order to characterise the positioning error, its distribution is presented in Figure 6.9. For a better visibility of the results, the graph was truncated for errors larger than 20 m, which are caused by poor satellite visibility and only represent 2.4% of the data. The cumulative distribution function shows that 50% of the errors are smaller than 3.6 m and 95th percentile is equal to 11.9 m. The root mean square error of the position estimated with the traditional least squares is equal to 7.45 m over the entire track. These values are relatively large compared to the 3 to 5 m accuracy (95%) reported by (RSSB 2012a). This is related to the poor satellite visibility and high values of DOP which biased the positioning error, as 7.6% of the data has a PDOP larger than 3.

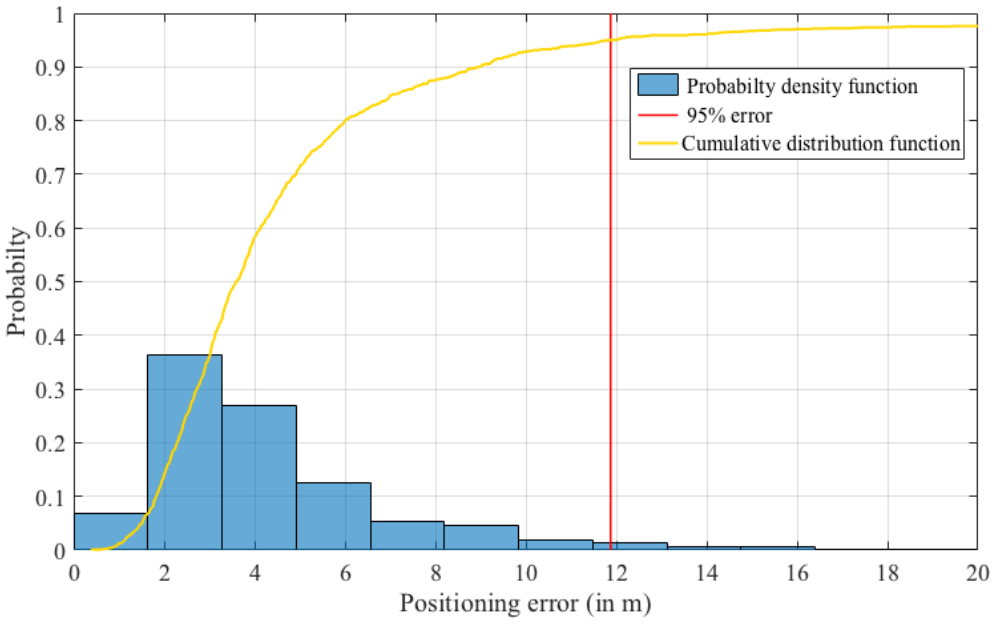


Figure 6.9 – Positioning error probability density function, cumulative distribution function and 95th percentile

In order to analyse the positioning performance independently of the effect of the satellite geometry, a subset of the data was selected, corresponding to epochs with a limited change in the DOP values (i.e. DOP values below 5). The characteristic of the positioning error over these epochs is presented in Figure 6.10. The cumulative distribution function shows that 50% of the errors are smaller than 3.1 m, the 95th percentile is equal to 5.8 m and RMS error is equal to 3.6 m. These values reflect the expected positioning performance reported by (RSSB 2012a) and show the high impact of the satellite geometry has on this specific dataset.

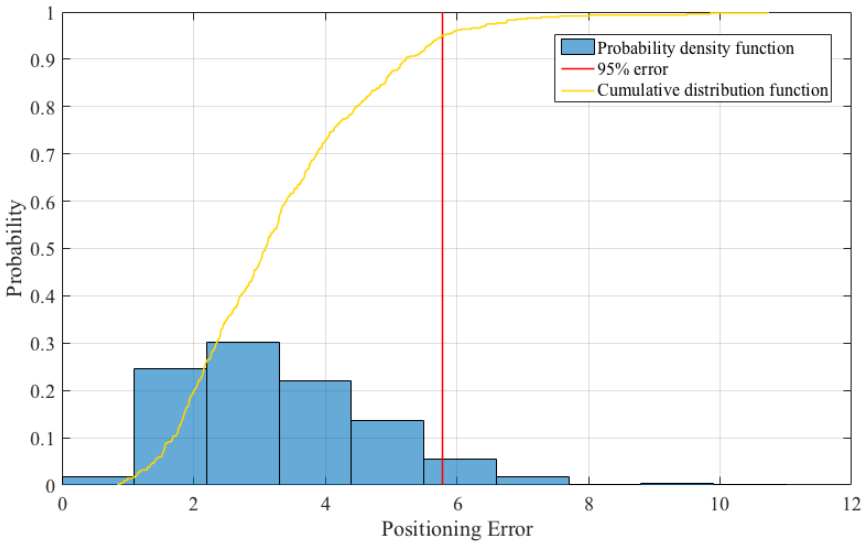


Figure 6.10 - Positioning error probability density function (blue), cumulative distribution function (yellow) and 95th percentile (red) over sections of track with a good satellite visibility

The non-weighted least squares solution presented in this section is used as the baseline to compare the performance of the code multipath mitigation techniques presented in the following sections.

6.2 Carrier-smoothed code

In order to test the effect of the Hatch filter on the positioning solution, the code measurements were smoothed using the carrier phase measurements before being used in the least squares algorithm. The positioning errors associated with different values of the Hatch filter constant are presented in Figure 6.11. It can be seen in this figure that the Hatch filter does not remove or smooth the large error as they are related to the satellite geometry.

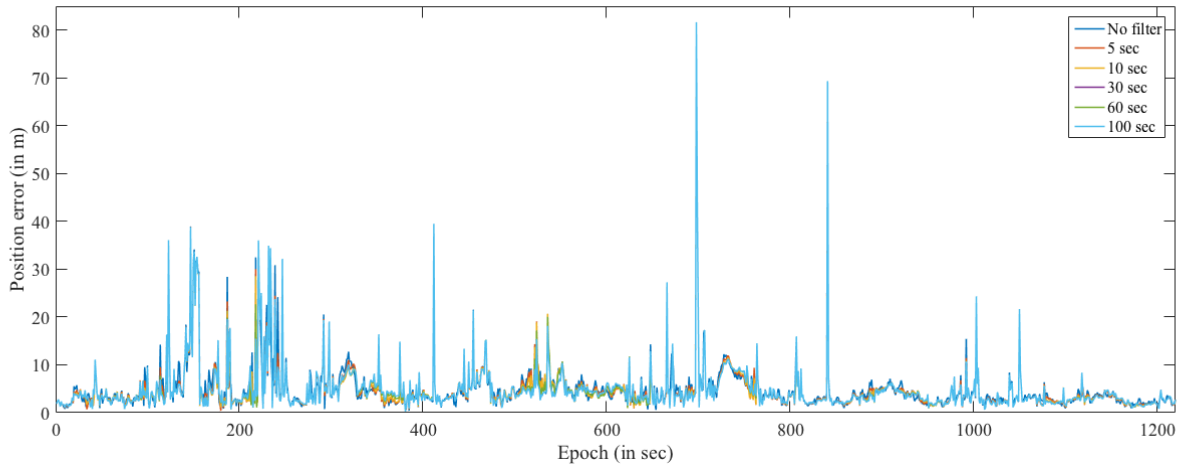


Figure 6.11 – Positioning error for different values of the Hatch filter time constant

Figure 6.12 shows the positioning error for the different value of the time constant between epoch 290 and 670. In the majority of cases, the use of the carrier smoothed code reduces the positioning error. The longer the hatch filter constant, the larger the reduction. For example, at epoch 320, the positioning error is reduced from 13 m for the least squares estimation without code smoothing, to 11 m for a 5 seconds smoother and to 10 m for time constants of 10 seconds and longer.

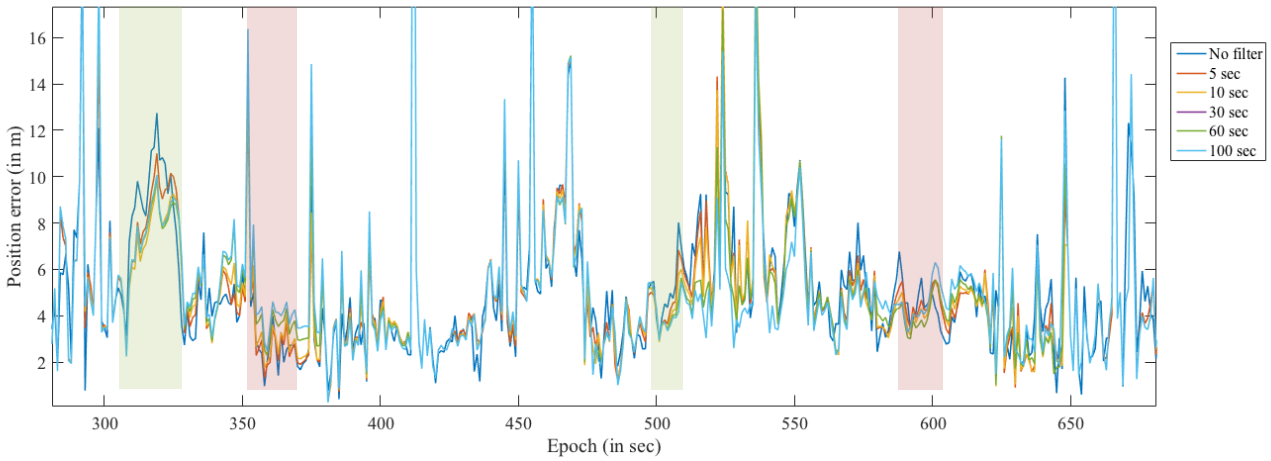


Figure 6.12 - Positioning error for different values of the Hatch filter time constant, between epoch 290 and 670

However, the use of the carrier-smoothed code also introduces some larger errors, such as from epochs 353 and 600, represented in red in Figure 6.12. In these cases, the smoothed code generates a positioning error larger than the non-smoothed code. These epochs correspond to cases when the train is moving from a relatively obstructed area to an open sky, as represented in Figure 6.13. When the train departs the station at epoch 353, the code measurements can be affected by a multipath error due to the presence of the station roof (in white on Figure 6.13). These measurements are then used to compute the consecutive smoothed code as shown in equation (5.21). The duration of the effect of this biased first measurement on the position error depends on the time constant N used in the Hatch filter. Therefore, this first code measurement is only used to smooth the N consecutive measurements, thus the smaller the value N, the fewer measurements are affected. At epoch 374, the carrier-smoothed code with a time constant of 100 sec is still generating an error 1 m larger than the non-smoothed code. Thus, a large time constant may not bring as much gain as expected and should be chosen carefully in order to minimise the positioning error.

From Figure 6.14, it can be seen that the use of the Hatch filter on average reduces the number of epochs with a positioning error larger than 5 meters but has a limited effect on errors smaller 2 m.

In order to analyse performance in detail, each filter time constant, the RMS error, 95th percentile and average positioning error reduction brought by the used of the carrier-smoothed code was computed. The results are shown in Table 6.2 and Table 6.3.



Figure 6.13 – Train trajectory between epoch 353 and 374, from (Google Earth 2012)

The effect of the smoothed code on the distribution of the positioning error is shown in Figure 6.14 for different values of the filter time constant.

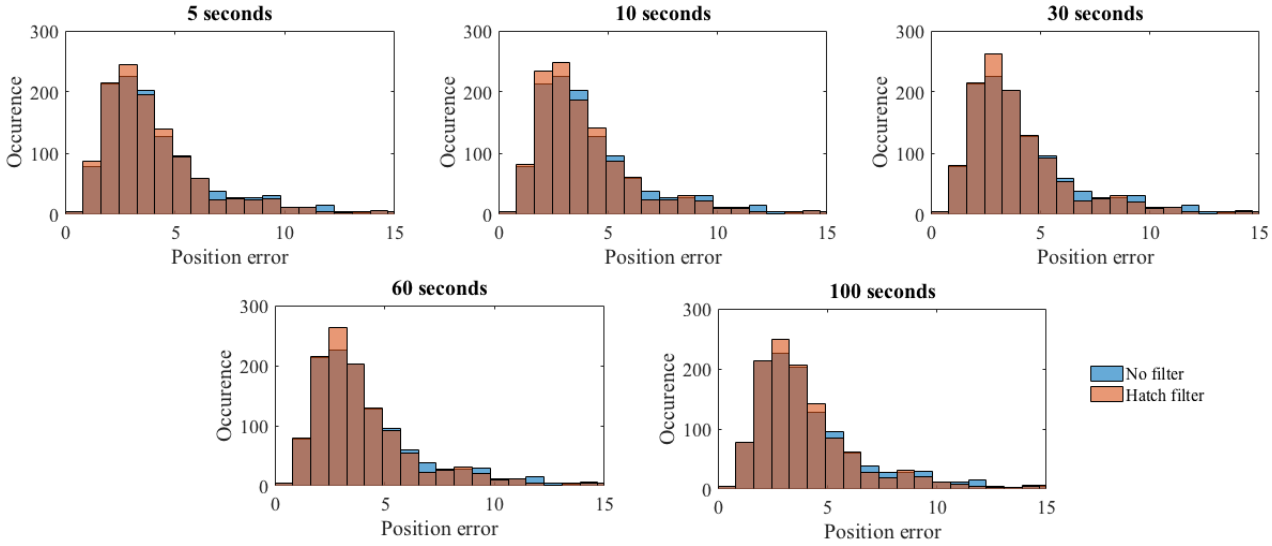


Figure 6.14 – Positioning error distribution with and without carrier-smoothed code, for different values of the filter time constant

Table 6.2 – Characterization of the Hatch filter time constant on the positioning error over the entire track

Filter time constant	No filter	5 sec	10 sec	30 sec	60 sec	100 sec
Positioning error (RMS)	7.45 m	7.32 m	7.27 m	7.27 m	7.27 m	7.27 m
95% percentile positioning error	11.88 m	11.41 m	11.19 m	11.35 m	11.35 m	11.37 m
Average positioning error reduction	0.0 m	0.15 m	0.20 m	0.19 m	0.19 m	0.18 m

Table 6.3 - Characterization of the Hatch filter time constant on the positioning error for over part of the track with a good satellite visibility (DOP<5)

Filter time constant	No filter	5 sec	10 sec	30 sec	60 sec	100 sec
Positioning error (RMS)	3.62 m	3.46 m	3.41 m	3.40 m	3.43 m	3.45 m
95% percentile positioning error	5.78 m	5.48 m	5.31 m	5.36 m	5.56 m	5.59 m
Average positioning error reduction	0.0 m	0.10 m	0.14 m	0.16 m	0.13 m	0.12 m

In summary, the error reduction is relatively small compared to the positioning error. It should be noted also that a time constant of 30 seconds or higher has less impact on the positioning error, potentially due to the effect of biased code measurements described in previously. Thus a filter time constant equal to 10 seconds appears to provide the highest benefit, in position accuracy.

In order to analyse in detail the benefits of using the carrier-smoothed code in the railway environment, Figure 6.15 represent the observable (i.e. code or smooth-code with a different time constant) generating the smallest positioning error at each point on the train's trajectory. This enables to determine if the carrier-smoothed is particularly adapted to certain environments.

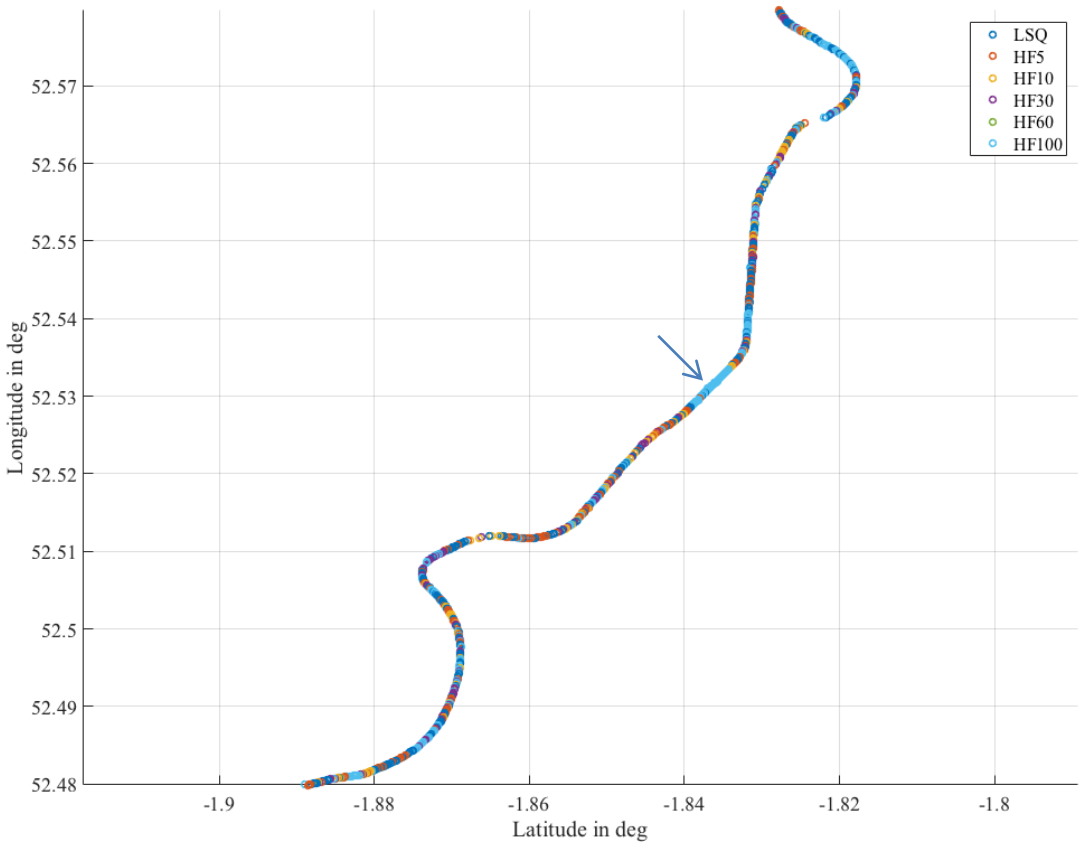


Figure 6.15 – Hatch filter time constant producing the minimum positioning error as a function of the train position.

The carrier-smoothed code seems to be particularly beneficial in the area indicated by an arrow, which corresponds to a straight section of track between Chester Road and Erdington stations, with trees on each side, as illustrated in Figure 6.16. This implies that the Hatch filter has the potential to improve the positioning solution in an area with limited visibility on the side of the track.

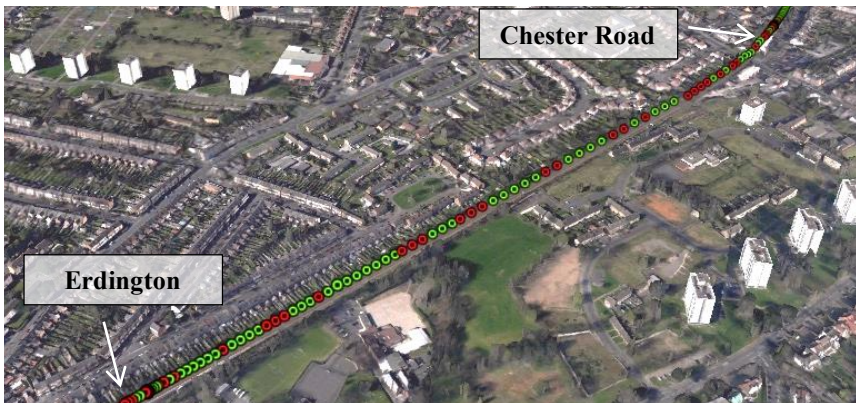


Figure 6.16 – Train trajectory between Chester Road and Erdington station, from (Google Earth 2012)

In conclusion, the use of carrier smoothed code provides a reduction of the positioning error. However, this improvement is marginal compared to the amplitude of the error. This limited improvement of the positioning solution can be related to the rapidly changing satellite visibility, which causes the filter to frequently reset and thus limiting its performance. Furthermore, the presence of multipath at the initialization of the filter could create a positioning bias. Hence, the longer the filter time constant, the larger the impact of this bias on the positioning solution is. Nevertheless, the use of the carrier-smoothed code in open areas demonstrates potential in enhancing the position accuracy.

6.3 Existing weighting techniques

In order to analyse the effect of the weighting methods reviewed in section 4.3.2, the positioning solution is estimated using the weighted least squares algorithm described in section 4.3.1. The following section describes the effect of the elevation-based, stochastic elevation-based, C/N0-based and multipath-based weighting methods on the positioning solution, and compares the relative performance of these methods.

Table 6.4 and Table 6.5 compare the average positioning performance of each of the four methods over the entire track and for selected sections of track with good satellite visibility (DOP<5), respectively.

In terms of average performance over the entire track, the sigma-ε method provides the largest improvement, with a reduction of the RMS error by 9.4%. The elevation and stochastic elevation

methods only marginally improve the position accuracy (3.5% and 2.7% respectively). On average, the multipath-based weighting increases the positioning error by 3%.

Table 6.4 - Characterization of the impact of the different weighting methods on the positioning error over the entire track

Weighting method	No weight	Elevation	Stochastic elevation	Sigma-ε	Multipath
Positioning error (RMS)	7.45 m	7.19 m	7.25 m	6.75 m	7.66 m
95% percentile positioning error	11.9 m	12.07 m	11.56 m	10.50 m	11.61 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.31 m	0.20 m	0.43 m	-0.08 m

Table 6.5 - Characterization of the impact of the different weighting methods on the positioning error over selected sections of track with a good satellite visibility (DOP<5)

Weighting method	No weight	Elevation	Stochastic elevation	Sigma-ε	Multipath
Positioning error (RMS)	3.61 m	3.29 m	3.44 m	3.41 m	3.63 m
95% percentile positioning error	5.78 m	5.56 m	5.79 m	5.47 m	5.90 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.31 m	0.14 m	0.18 m	-0.01 m

However, in sections of track with a good satellite visibility (DOP<5), the elevation-based method outperforms the other methods and reduces the positioning error by 8.7%. The stochastic elevation and sigma-ε methods respectively provide an improvement of 4.7% and 5.5%, whereas the multipath method slightly increases the positioning error (by 0.5%).

In summary, except for the multipath-based method, the weighting techniques have comparable performance. Furthermore, the sigma-ε method demonstrates better performance in low satellite visibility environment while the elevation method appears to provide higher accuracy in areas with sufficient satellite visibility (DOP<5). This indicates that different methods might be more suitable for different environments.

Therefore, it is important to compare each method on an epoch by epoch basis, along the entire trajectory. Figure 6.17 represents the weighting method generating the smallest positioning error at each point on the train's trajectory, thus enabling to determine if a single method could consistently enhance positioning accuracy or if changes in the environment require the use of different methods. It can be seen clearly from Figure 6.17 that not a single method outperforms the others over the entire track. The

percentage of epochs each weighting method produces the smallest positioning error is presented in Table 6.6. The elevation method provides the most accurate position in 43% of cases, followed by the multipath method for 23% of cases. The sigma-ε method produces the highest accuracy in 19.5% of the cases, corresponding to epochs where the train is stopped at stations.

While on average the multipath method increases the positioning error (see Table 6.4 and Table 6.5), on an epoch by epoch basis it provides the best accuracy in 23% of the total epochs considered. One of the main areas where the multipath-based weighting outperforms the other methods is represented by a green arrow in Figure 6.17 and corresponds to the straight portion of track between Chester Road and Erdington station represented in Figure 6.16. Thus the multipath-based weighting appears to improve positioning accuracy when the visibility on the side on the track is limited.

Table 6.6 - Percentage of epochs each weighting method produces the smallest positioning error

Time constant	No weight	Elevation	Stochastic elevation	Sigma-ε	Multipath
Percentage	9.3%	43%	5.2%	19.5%	23%

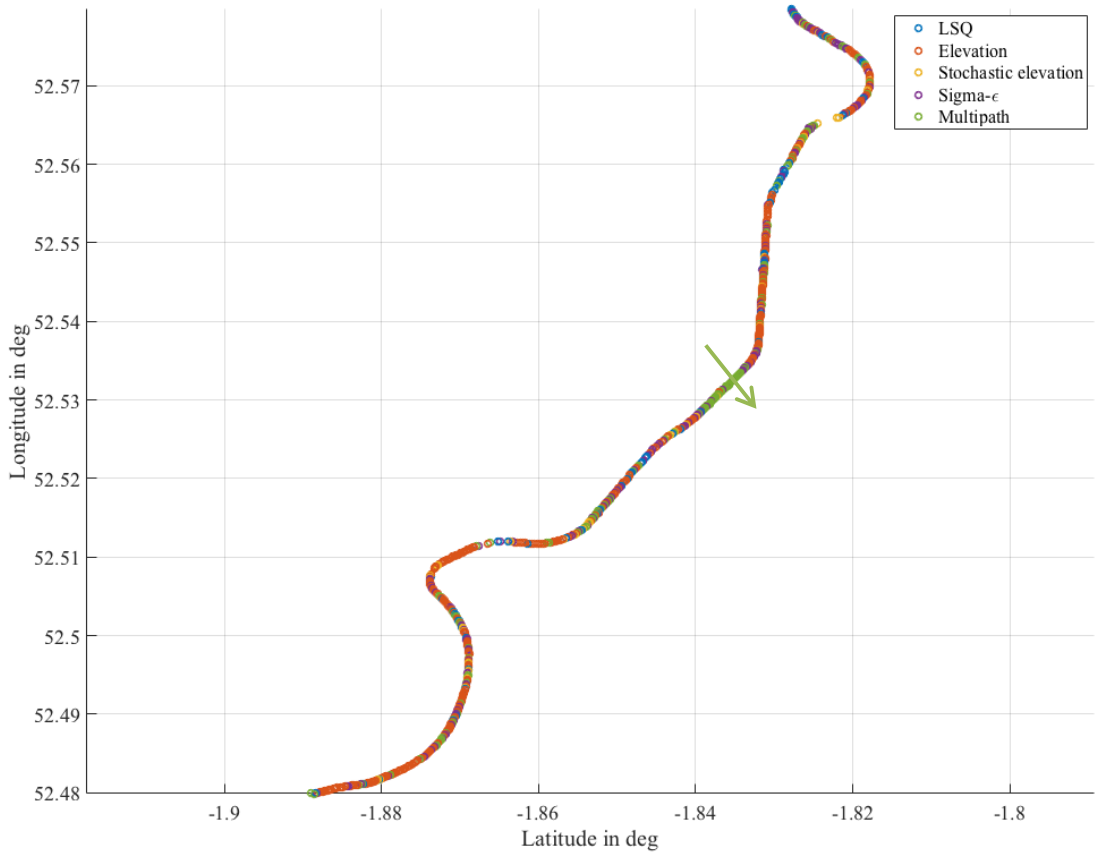


Figure 6.17 – Weighting method producing the minimum positioning error as a function of the train position.

From the results in this section, it can be observed that the correct combination of the elevation-based weighting, the C/N0 weighting (i.e. sigma-ε model) and the multipath-based weighting could potentially improve the positioning solution accuracy.

6.4 Novel heading based weighting

This section presents the positioning performances of the novel heading-based weighting method introduced in section 5.1. Initially, the measurements are weighted based on the orthogonal distance formulated in section 5.1.1. Then, the measurements are weighted using the heading and elevation based weights defined in section 5.1.2. Finally, the adaptive formulation of the model offered in section 5.1.3 is tested.

6.4.1 Orthogonal distance-based weighting method

Firstly, the measurements are weighted based on the satellite alignment with the track, represented by the orthogonal distance *d* presented in section 5.1.1. In that case, the weights are defined as:

$$\sigma_i = \sigma_0(1 + A * e^{B*d}) \tag{6.1}$$

Where *A* is fixed to 0.00004 and *B* to 13.

The effect of the elevation is captured in the determination of the distance *d* and the orthogonal-based weighting formulation can be assimilated to the elevation-based weighting method reduced to one along the train heading.

Table 6.7 and Table 6.8 compare the average positioning performance of the orthogonal distance weighting method to the positioning performance of the current weighting methods analysed in section 6.3. The orthogonal distance weighting method reduces the average positioning error (RMS) of the non-weighted solution by less than 1% over the entire track and even increases by 1% the positioning error (RMS) in sections of track with a good satellite visibility (DOP<5).

However, as discussed in section 6.2, the average positioning performance over the entire track does not reflect the effect of the weights on the solution and the specific conditions in which this method may improve the positioning accuracy. To analyse the performance of this novel weighting method epoch by epoch, its positioning error is compared to the other weighting methods position error epoch by epoch. The percentage of epochs the novel method presents a smaller position error than the current methods is presented in Table 6.9.

Table 6.7 - Characterization of the different weighting methods impact on the positioning error over the entire track

Weighting method	No weight	Distance only	Elevation	Stochastic elevation	Sigma-ε	Multipath
Positioning error (RMS)	7.45 m	7.39 m	7.19 m	7.25 m	6.75 m	7.66 m
95% percentile positioning error	11.9 m	11.97 m	12.07 m	11.56 m	10.50 m	11.61 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.03 m	0.31 m	0.20 m	0.43 m	-0.08 m

Table 6.8 - Characterization of the different weighting methods impact on the positioning error over selected sections of track with a good satellite visibility (DOP<5)

Weighting method	No weight	Distance only	Elevation	Stochastic elevation	Sigma-ε	Multipath
Positioning error (RMS)	3.61 m	3.64 m	3.29 m	3.44 m	3.41 m	3.63 m
95% percentile positioning error	5.78 m	6.20 m	5.56 m	5.79 m	5.47 m	5.90 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.01 m	0.31 m	0.14 m	0.18 m	-0.01 m

Table 6.9 – Comparison of the orthogonal distance-based weighting method with the current weighting methods

Weighting method	No weight	Elevation	Stochastic elevation	Sigma-ε	Multipath
Percentage of epoch with an smaller positioning error	54%	33%	36%	38%	51%
Average positioning error reduction (with respect to each weighting method)	0.03 m	0.07 m	0.01 m	0.10 m	0.02 m

The orthogonal distance provides a more accurate positioning solution than the elevation method in 33% of cases. These cases are represented in green in Figure 6.18 and can be related to locations with side-track obstacles such as buildings at stations or trees along the line.

The distance-based weighting method also provides a more accurate positioning solution than the sigma-ε model in 38% of cases, represented in green in Figure 6.19. Most of these cases are similar to the ones identified in Figure 6.18. Furthermore, the distance based method performs best at the end of

the trajectory, when the train travels through a suburban area with industrial buildings on one the side of the track.

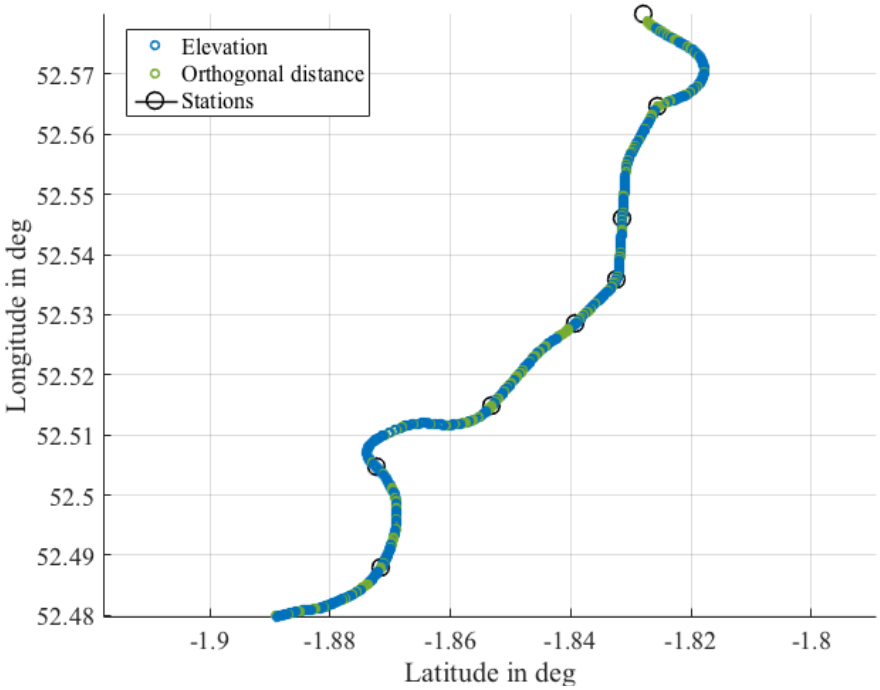


Figure 6.18 – Weighting method providing the smallest positioning error at each epoch, between the elevation method (blue) and the orthogonal distance (green)

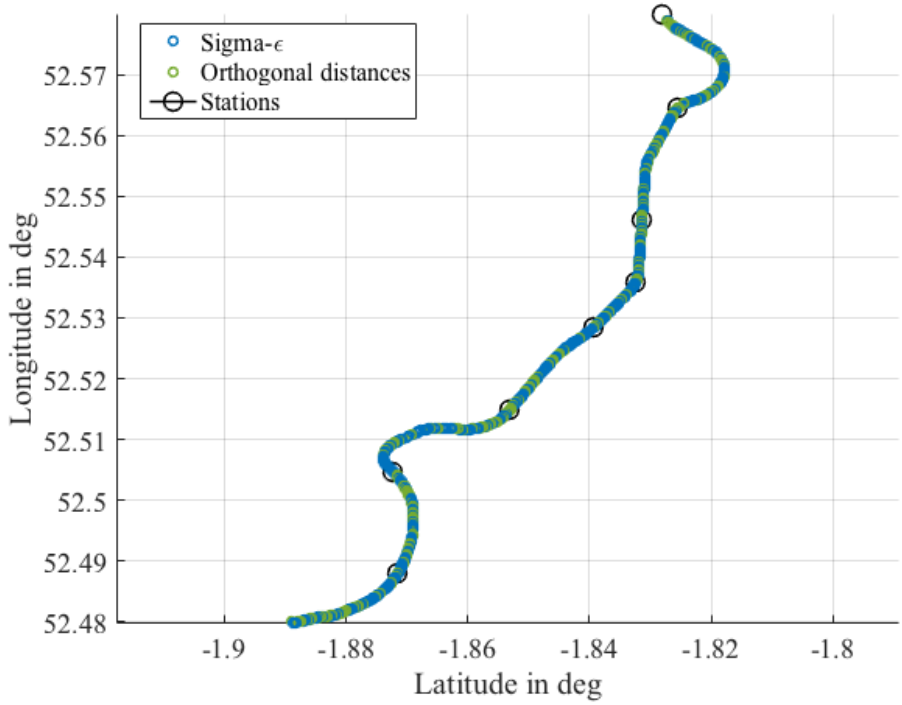


Figure 6.19 - Weighting method providing the smallest positioning error at each epoch, between the sigma-ε method (blue) and the orthogonal distance (green)

In summary, the orthogonal distance-based weighting method does on average degrade the positioning accuracy (Table 6.7 and Table 6.8). However, it has the potential to provide a more accurate solution than current weighting methods in the presence of obstacles such as trees or buildings, on the side of the track.

6.4.2 Heading based weighting method

As discussed in section 5.1.2, the weighting of the GNSS measurements on the basis of the orthogonal distance is limited as satellites located above the track and with a low elevation angle are assigned a high weight despite their likeliness to be affected by larger atmospheric residual errors. To counter this effect, an alternative weight definition was introduced in section 5.1.2, such as:

$$\sigma_i = \sigma_0 \left[\frac{1}{C} * (1 + A * e^{B*d}) + \left(1 - \frac{1}{C}\right) * \left(\frac{1.001}{\sqrt{0.002001 + \sin(elevation)^2}} \right) \right] \quad (6.2)$$

Where A is fixed to 0.00004, B to 13 and C is an adjustable parameter.

As seen in Figure 5.9, the parameter C is used to adjust the contribution of satellites along the tracks as a function of their elevation.

Table 6.10 and Table 6.11 compare the average positioning performance of the heading based weighting method for different values of the parameter C. As the value of C increases, the RMS positioning error decreases. The heading based method provides average positioning performance comparable to the current methods presented in Table 6.7 and Table 6.8. For values of C equal to 5 or 10, the performance is similar to the elevation based method, which can be related to the heading method similarity to the elevation method for high values of C, as seen in Figure 5.9. Based on these tables, the optimal value for C over the entire track is 2.0 as it minimises the 95th percentile error while reducing the RMS positioning error. Over sections of tracks with a good satellite visibility (DOP<5), a value of C equal to 10 reduces both the RMS positioning error and the 95th percentile error. This implies that is these cases, elevation-based weighting performs better than the heading based method.

As demonstrated in section 6.2, the average positioning performance does not entirely reflect the effect of the weights in the positioning solution, as each method may perform better at specific locations along the track. To analyse the positioning performance of the heading method relative to the current methods, the percentage of epoch the heading weight provides a more accurate solution is presented in Table 6.12.

Table 6.10 - Characterization of the impact of the different weighting methods on the positioning error over the entire track

Weighting method	No weight	C=1	C=1.5	C=2.0	C=3.0	C=5.0	C=10.0
Positioning error (RMS)	7.45 m	7.39 m	7.23 m	7.21 m	7.19 m	7.19 m	7.19 m
95% percentile positioning error	11.9 m	11.97 m	11.47 m	11.24 m	11.49 m	11.53 m	11.86 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.03 m	0.22 m	0.25 m	0.27 m	0.29 m	0.30 m

Table 6.11 - Characterization of the impact of the different weighting methods on the positioning error over selected sections of track with a good satellite visibility (DOP<5)

Weighting method	No weight	C=1	C=1.5	C=2.0	C=3.0	C=5.0	C=10.0
Positioning error (RMS)	3.61 m	3.64 m	3.45 m	3.40 m	3.36 m	3.33 m	3.31 m
95% percentile positioning error	5.78 m	6.20 m	5.88 m	5.74 m	5.67 m	5.53 m	5.52 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.01 m	0.15 m	0.20 m	0.21 m	0.27 m	0.29 m

Table 6.12 - Comparison of the heading based weighting method with the current weighting methods

Percentage of epochs with a smaller positioning error	No weight	Elevation	Stochastic elevation	Sigma-ε	Multipath
C =1.0	53.7%	33.1%	35.7%	38.0%	51.5%
C = 1.5	64.2%	37.3%	52.2%	47.1%	58.0%
C = 2.0	63.6%	38.0%	61.9%	50.8%	59.5%
C = 3.0	63.9%	38.2%	61.7%	53.5%	60.7%
C = 5.0	63.5%	39.1%	61.2%	53.7%	60.3%
C = 10.0	63.6%	39.9%	60.6%	53.8%	60.7%

As seen from Table 6.12, the heading based weighting methods provide a more accurate positioning solution than the current weighting methods in a large number of cases. For example, the heading based solution is more accurate than the sigma-ε model in 47% of cases for *C* equal to 1.5, represented in green in Figure 6.20.

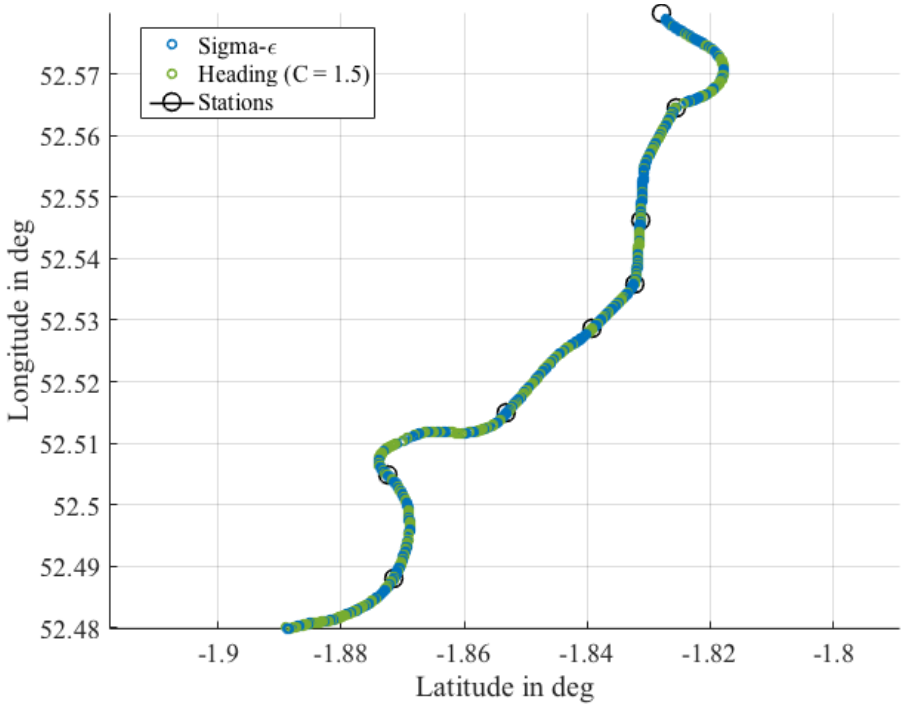


Figure 6.20 - Weighting method providing the smallest positioning error at each epoch, between the sigma-ε method (blue) and the heading based method (green)

The heading based weighting method also provides a more accurate solution than the elevation based weighting in 37% of cases for *C* equal to 1.5, as represented in Figure 6.21.

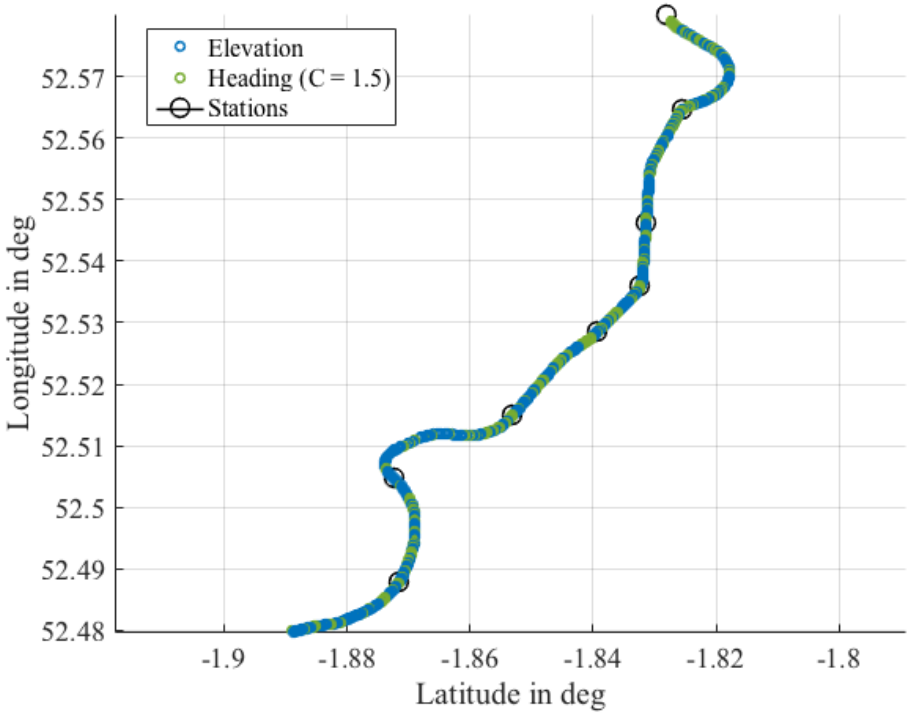


Figure 6.21 - Weighting method providing the smallest positioning error at each epoch, between the elevation method (blue) and the heading based method (green)

In summary, the heading based weighting method positioning accuracy is similar to the current weighting methods. However, it outperforms the stochastic-elevation, sigma-ε and multipath methods over 53% of the time when the comparison is performed on an epoch by epoch basis. It also produces a more accurate position than the elevation-based method in approximately 40% of cases.

6.4.3 Adaptive weighting method

Finally, the positioning performance of the adaptive weighting model presented in section 5.1.3 is analysed. Table 6.13 and Table 6.14 compare the average positioning performance of the adaptive model to the current weighting methods. Over the entire track, the adaptive model reduces the RMS positioning error by 4%, to a level comparable to the elevation weighting method. This reduction is relatively less important for epochs with a good satellite visibility (DOP<5), which implies that the adaptive model reduces some of the errors associated with poor geometry.

Table 6.13 - Characterization of the impact of the different weighting methods on the positioning error over the entire track

Weighting method	No weight	Heading	Elevation	Stochastic elevation	Sigma-ε	Multipath
Positioning error (RMS)	7.45 m	7.18 m	7.19 m	7.25 m	6.75 m	7.66 m
95% percentile positioning error	11.9 m	11.59 m	12.07 m	11.56 m	10.50 m	11.61 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.26 m	0.31 m	0.20 m	0.43 m	-0.08 m

Table 6.14 – Characterization of the impact of the different weighting methods on the positioning error over selected sections of track with a good satellite visibility (DOP<5)

Weighting method	No weight	Heading	Elevation	Stochastic elevation	Sigma-ε	Multipath
Positioning error (RMS)	3.61 m	3.39 m	3.29 m	3.44 m	3.41 m	3.63 m
95% percentile positioning error	5.78 m	5.67 m	5.56 m	5.79 m	5.47 m	5.90 m
Average positioning error reduction (with respect to the non-weighted solution)	0.0 m	0.21 m	0.31 m	0.14 m	0.18 m	-0.01 m

To further assess the performance of the adaptive weighting model, the percentage of epoch this method produces a positioning error smaller than the current methods is presented in Table 6.15.

Table 6.15 - Comparison of the adaptive based weighting method with the current weighting methods

Weighting method	No weight	Elevation	Stochastic elevation	Sigma-ε	Multipath
Percentage of epoch with an smaller positioning error	62.4%	34.0%	58.5%	51.6%	59.0%
Average positioning error reduction	-0.01 m	0.03 m	-0.02 m	0.06 m	-0.02 m

The adaptive weighting method provides a significant improvement over the stochastic elevation, the sigma-ε and multipath weighting methods as it provides a more accurate positioning estimate in more than 50% of cases. Furthermore, the adaptive weighting method provides a more accurate solution than the elevation based method in 34% of cases, as shown in Figure 6.23.

In conclusion, the adaptive weighting model developed in this thesis present positioning performance comparable on average to the current elevation-based and sigma-ε methods. By analysing the positioning error epoch by epoch, this adaptive method appears to mitigate the positioning error in places where trackside obstacles such as buildings, trees or cuttings are present.

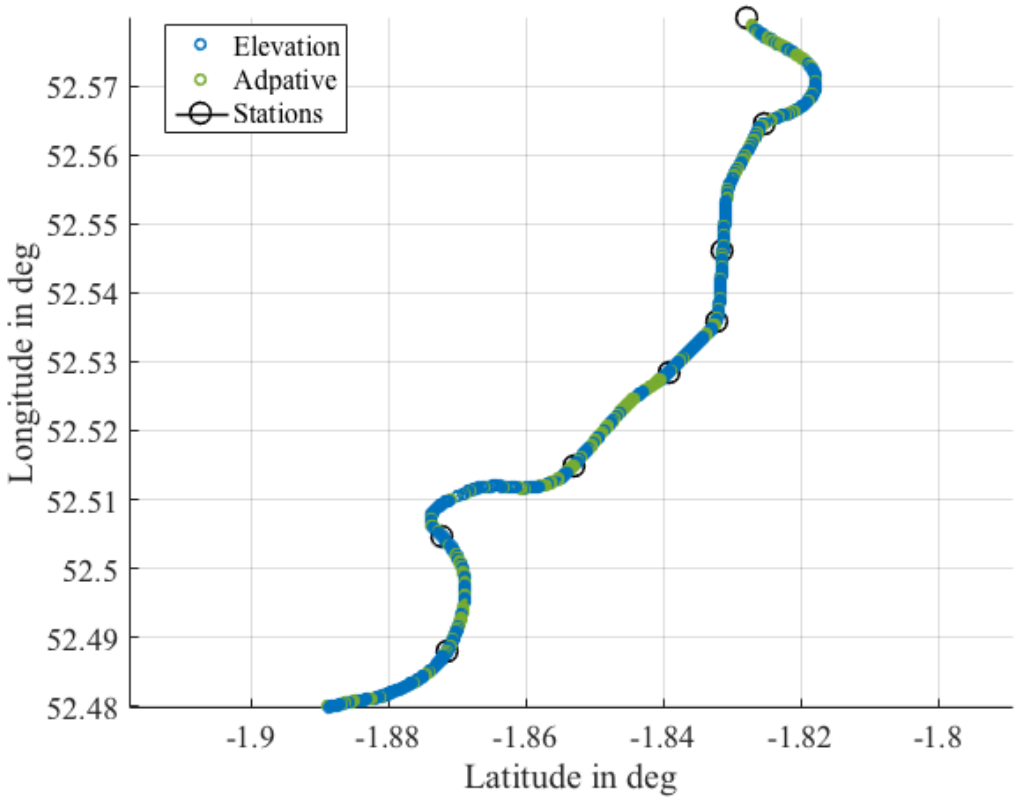


Figure 6.22 - Weighting method providing the smallest positioning error at each epoch, between the elevation method (blue) and the adaptive method (green)

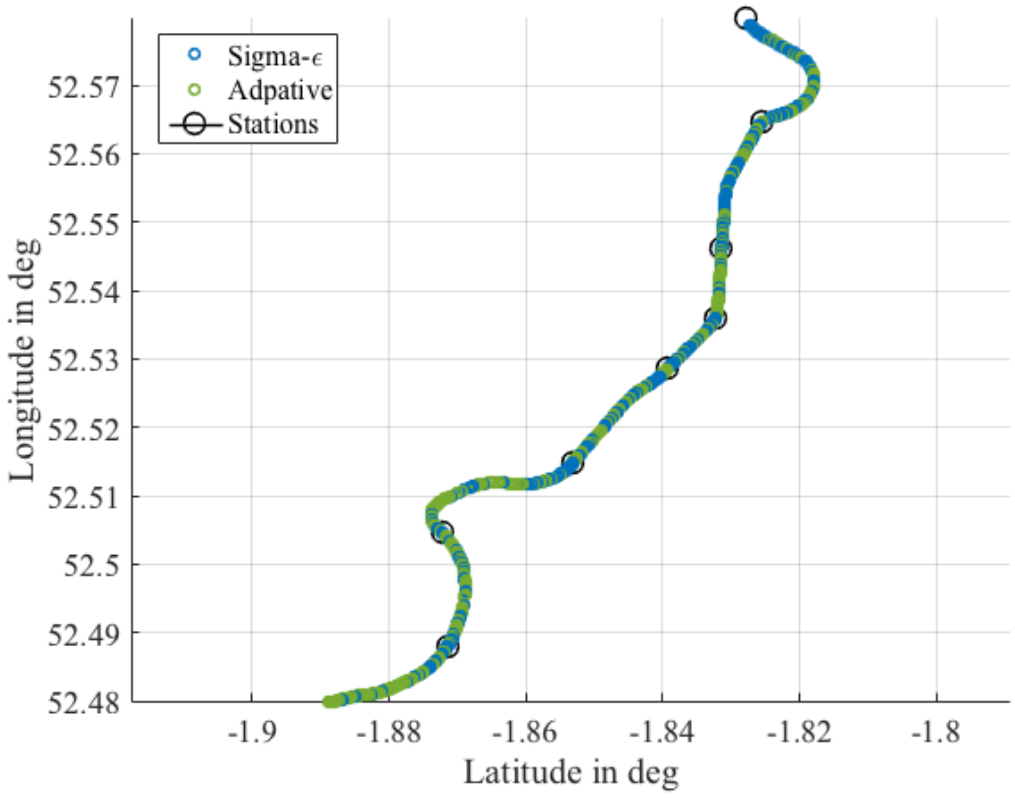


Figure 6.23 - Weighting method providing the smallest positioning error at each epoch, between the sigma-ε method (blue) and the adaptive method (green)

6.5 Conclusion

In this chapter, the performance of the different multipath mitigation techniques (presented in chapters 4 and 5) has been reviewed in the railway environment. This enabled to demonstrate that no single method consistently outperforms all the others, probably due to the variability of the railway environment.

Nevertheless, the elevation method consistently outperformed the other methods, especially in cases of good satellite visibility (DOP>5). On the contrary, the novel multipath and heading-based adaptive methods demonstrated potential in enhancing positioning accuracy in cases of low satellite visibility.

Therefore, it is apparent from the result that a careful combination of the elevation and heading based methods has the potential to mitigate code multipath in the railway environment. The practical application of these methods will require that a relevant parameter such as the DOP is used to select when to use each approach. Further work is required to fully characterise the performance of the methods based on an extensive representative dataset.

Chapter 7

Conclusions and Recommendations

This final chapter presents the main findings of the research carried out in this thesis on code multipath mitigation in the railway environment in section 7.1, with reference to the research objectives set in section 1.2. Section 7.2 proposes a set of recommendations and suggestions for further research.

7.1 Conclusions

The aim of this thesis was to improve GNSS-based positioning in the railway domain by developing a novel code multipath mitigation technique based on the current measurement weighting methods, but taking into account the unique features of the railway, in particular, the track geometry. To achieve this aim, six objectives were defined in section 1.2. The findings related to each objective are discussed in turn in the following section.

1. Critical review of the railway operations and identification of location-based railway services

A critical review of the railway operations was carried out in chapter 2, leading to the identification of a large number of railway location-based services, together with their required navigation performance. A set of requirements reflecting the most stringent values found in the literature was selected with the purpose of developing a positioning system that can address the widest range of railway location-based applications.

2&3. Review of current positioning systems and specification of an architecture for a railway positioning system

The review of railway positioning systems has shown that existing systems cannot fulfil the requirements identified in objective 1, leading to the review of alternative positioning systems and techniques. GNSS was identified as an attractive solution due to its provision of an autonomous positioning service with a global coverage. However, GNSS-based positioning is sensitive to the loss of signals and must be combined with additional sensors to offer the required continuity of services. Several alternative sensors were identified as having the capability to complement GNSS in the railway environment, including inertial sensors, signals of opportunity, map matching and also traditional

railway systems such as the odometer and balise. Three positioning system architectures with GNSS as the core were discussed as having the potential to meet the required navigation performance.

4. Review of GNSS error sources and associated mitigation techniques

With GNSS at the core of the positioning system and following a detailed review of the code measurement error sources, code multipath is identified as the main contributor to the GNSS positioning error due to its variability in dynamic contexts. Existing multipath mitigation techniques were reviewed with a focus on weighting techniques that do not require any additional hardware or software modification.

Using data, it was demonstrated that the relationships assumed by the weighting techniques between the satellite elevation angle, C/N0 and code multipath and noise are valid for a static receiver, but that they are very approximate and sometimes wrong in the case of a dynamic receiver, and thus have the potential to decrease the positioning accuracy in some instances.

5. Development of an improved weighting technique to mitigate the effect of code multipath in the railway environment

Accounting for the limitations of the current techniques, a method was developed to weight the measurements in the context of railway operations. This method assumes that satellites aligned with the track are less likely to be affected by noise and multipath than satellites perpendicular to the tracks, due to the presence of trackside obstacles, such as buildings and trees. A novel quality indicator, the orthogonal distance, is introduced to quantify the satellite alignment with the train track. This is an enhanced combination of the measurements quality indicators such as the satellite elevation angle, the carrier to noise ratio and the code multipath estimate.

6. Testing and validation of the developed improved weighting technique

The effect of the heading-based method on the position accuracy was compared to current weighting methods. In over 50% of the cases analysed, the heading-based method provides a more accurate solution than the stochastic elevation, sigma-ε and multipath weighting methods. It also provides a more accurate solution than the elevation-based method in 34% of cases. The heading-based method performs best in locations with weak satellite geometry, such as in the presence of trees or trackside buildings. From the results of the analysis, it can be observed that a correct combination of the elevation and the heading-based methods has the potential to mitigate the effect of code multipath and thus to improve the positioning solution accuracy in the railway environment.

7.2 Recommendations for future works

The research presented in this thesis has shown the limitations of current weighting methods in dynamic contexts, demonstrating that not a single method could consistently enhance the positioning accuracy. The heading-based algorithm developed in this thesis accounts for the scenario where trackside obstacles are present, and with a careful dynamic adjustment, could also function as an elevation-based weighting where appropriate (e.g. open sky environment). Therefore further exhaustive analysis and validation using a more representative dataset are required to fully characterise the heading-based method performance.

As stated previously, each weighting method appears to improve the positioning accuracy in specific scenarios, making the case for the implementation of an adaptive weighting method. Further work is then required to relate parameters such as the DOP, SNR values or multipath estimates to specific environments. Such parameters could then be used to detect typical scenario (e.g. urban canyon, trackside trees) and to dynamically apply the optimal weighting technique, leading to the smallest positioning error. Such work will require the access to a larger dataset, covering a large variety of railway environments.

In addition to the developed heading-based weighting method, a formulation for a multipath-based weighting method was offered in section 5.2.2.4. This methods has shown promising results, as for the dataset analysed, it provides the best accuracy for 23% of the epochs (section 6.3). However, the formulation of the multipath-based weight is purely theoretical, as it is defined to be consistent with existing weighting models (section 5.2.2.4). This model could then be further refined, for example by comparing observed pseudoranges with their correct values (using the reference solution). This is also applicable to the determination of the coefficients in the heading-based model (section 5.1.2).

Additionally, the distribution of the code multipath error in the railway domain was derived for the dataset used in this thesis. While these results give an indication of the level of code multipath error and of its variation with the environment, a larger amount of data should be analysed to fully characterise the code multipath error in the railway domain. Such analysis could be exploited to compute the code multipath error bound, which could be then used for developing integrity monitoring solutions. It is also important to notice that the influence of the GNSS receiver on the multipath estimates should be further investigated. As explained in section 4.2.2.2, several techniques can be used at the receiver level to mitigate the effect of multipath signals. It is thus expected that different receivers will produce different level of multipath error in the measurement domain.

While the heading-based weighting method seems adapted to the railway environment where the track design favours the satellite visibility along the train heading, the performance of this novel method

for road vehicles in an urban environment could be investigated. Indeed, a car travelling in an urban canyon may benefit from using satellites along the axis of the street.

Finally, at a higher level, and as summarised by (Marais et al. 2017), several issues need to be addressed to support the introduction of GNSS-based positioning systems in the railway domain. First, standard positioning requirements are needed in order to drive the design of the GNSS-based positioning function. Such requirements will help to standardisation and certification of GNSS-based positioning solutions for railway operations. Additionally, GNSS errors in the railway environment need to be better modelled. This include the multipath error previously mentioned, but also other error sources such as atmospheric errors. These models are required to develop integrity monitoring principles and ensure that the GNSS-based positioning system is resilient to order and can be used to support safety-critical railway applications.

References

Al-Fanek, O. et al., 2007. Evaluation of L2C Observations and Limitations. In *Proceedings of the ION GNSS 2007, Forth Worth, Texas, September 25-28*.

Albanese, A. & Marradi, L., 2005. The RUNE project: the integrity performances of GNSS-based railway user navigation equipment. In *Proceedings of the 2005 ASME/IEEE Joint Rail Conference, 2005.* IEEE, pp. 211–218. Available at: http://ieeexplore.ieee.org/document/1460846/ [Accessed October 6, 2016].

ALSTOM Signaling Inc., 2004. *A Centennial History of ALSTOM Signaling Inc. 1904-1004*,

Anon, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems*,

APOLO, 1999. *User Needs and Technology Assesment*,

Axelrad, P., Comp, C.J. & Macdoran, P.F., 1996. SNR-based multipath error correction for GPS differential phase. *IEEE Transactions on Aerospace and Electronic Systems*, 32(2), pp.650–660.

Barbu, G., 2001. *Accuracy Performance Test Methodology for Satellite Locators on Board of Trains - Developments and results from the EU Project APOLO*,

Barbu, G. & Marais, J., 2014. The SATLOC project. In *Transport Research Arena, Apr 2014*. Paris, p. 10.

Beugin, J. et al., 2010. Galileo for railway operations: question about the positioning performances analogy with the RAMS requirements allocated to safety applications. *European Transport Research Review*, 2(2), pp.93–102. Available at: http://dx.doi.org/10.1007/s12544-010-0032-3.

Bisnath, S.B. & Langley, B.R., 2001. Pseudorange Multipath Mitigation By Means of Multipath Monitoring and De-Weighting. In *Kinematic Systems in Geodesy, Geomatics and Navigation 2001, 5-8 June*. Banff, Alberta.

Black, H.D. & Eisner, A., 1984. Correcting satellite doppler data for tropospheric effects. *Journal of Geophysical Research*, 89(D2), pp.2616–2626.

Blanchard, D., 2012. Galileo Programme Status Update. In *ION GNSS 2012*. Nashville TN, USA, pp. 553–587.

Bloomfield, R. (Network R., 2006. Fundamentals of European Rail Traffic Management System - ERTMS. In *The 11th IET Professional Development Course on Railway Signelling and Control Systems, 2006*. pp. 165–184.

Braasch, M.S., 1994. Isolation of GPS Multipath and Receiver Tracking Errors. *Journal of the Institute of Navigation*, 41(4), pp.415–435.

Chansarkar, M., 2000. Neural Networks in GPS Navigation. *GPS Solutions*, 4(2), pp.14–18.

Chiang, K.-W. & Chang, H.-W., 2010. Intelligent sensor positioning and orientation through constructive neural network-embedded INS/GPS integration algorithms. *Sensors*, 10, pp.9252–9285.

Collins, P. & Stewart, P., 1999. *GPS SNR Observations*,

Ding, X. & Yang, J., 2012. BeiDou Navigation Satellite System of China, for the World. In *ION GNSS 2012*. Nashville TN, USA.

Dumville, M. & Hutchinson, M., 2012. *Data analysis for cost-effective GPS-based locator with simple augmentations - Position-dependent applications in the rail domain*,

Elsobeiey, M. & El-Rabbany, A., 2010. On stochastic modeling of the modernized global positioning system (GPS) L2C signal. *Measurement Science and Technology*, 21(5), p.55105. Available at: http://stacks.iop.org/0957-0233/21/i=5/a=055105?key=crossref.a3a7e956567d32ea93d77aa65def028a [Accessed September 23, 2016].

ERTMS Users Group, 1998. *ODOMETER FFFIS*,

Eurocontrol, 2013. Current GBAS Situation. Available at: http://www.ecacnav.com/Precision/Current_GBAS_Situation [Accessed June 25, 2013].

European Comission, 2011. *White Paper on Transport*,

European Commission, 2011a. *Commission Staff Working Document - Accompanying the White Paper - Roadmap to a Single European Transport Area – Towards a competitive and resou rce efficient transport system*,

European Commission, 2012. *Energy Roadmap 2050*,

European Commission, 2016. ERTMS - FAQ on ERTMS - Transport. Available at: http://ec.europa.eu/transport/modes/rail/ertms/general-information/faq_ertms_en.htm [Accessed October 6, 2016].

European Commission, 2011b. *EU transport in figures*,

European Commission, 2010. *iTREN - 2030: Experiences and results for integrated technology, energy and transport policy assessment*,

European Commission et al., 2012. *Memorandum of Understanding (MoU) between the European Commmission, the European Railway Agency and the European Rail secter Associations (CER - UIC - UNIFE - EIM - GSM-R Industry Group - ERFA) concerning the strengthening of cooperation for the managemen*, Available at: http://www.era.europa.eu/Document-

Register/Documents/MoU-betweenEC-ERA-and-Sector-Associations-on-ERTMS.pdf.

European Commssion, 2004. *Reclaiming city streets for people - Chaos or quality of life?*,

European GNSS Agency, 2016. E-GNSS enabled railway signalling - from vision to action. Available at: https://www.gsc-europa.eu/news/e-gnss-enabled-railway-signalling-from-vision-to-action [Accessed August 1, 2016].

European GNSS Agency, 2012. GNSS Introduction In The Rail Sector. Available at: http://www.gsa.europa.eu/gnss-introduction-rail-sector [Accessed February 4, 2015].

European Space Agency, 2005. ECORAIL - successful demonstration of EGNOS for railway control applications. Available at: http://www.esa.int/Our_Activities/Navigation/ECORAIL_-_successful_demonstration_of_EGNOS_for_railway_control_applications.

Faragher, R., 2011. Snap and No Snap. Available at: http://www.cl.cam.ac.uk/~rmf25/videos/snap_and_nosnap.wmv [Accessed August 31, 2016].

Filip, A. & Rispoli, F., 2014a. Safety concept of GNSS based train location determination system SIL 4 compliant for ERTMS/ETCS. In *ENC-2014, 14-17 April 2014*. Rotterdam, The Netherlands.

Filip, A. & Rispoli, F., 2014b. SIL 4 Compliant Train Location Determination System Based on Dual-Constellation EGNOS-R Interface for ERTMS/ETCS. In *CERGAL 2014 International Symposium on Certification of GNSS Systems & Services*. pp. 109–114.

Fontgalland, B. de., 1984. *The world railway system*, Cambridge University Press.

Freight Transport Association, 2008. *The Importance of Rail Freight*,

GAARDIAN, 2009. *GAARDIAN project*,

GaLoROI, 2012. *Galileo Localisation for Railway Operation Innovation*, Available at: http://www.galoroi.eu/upload/GaLoROI-Flyer-en.pdf.

Gebre-Egziabher, D., 2007. What is the difference between “loose”, “tight”, “ultra-tight” and “deep” integration strategies for INS and GNSS? *Inside GNSS*, p.28. Available at: http://www.insidegnss.com/auto/JanFeb07GNSSSolutions (secured).pdf.

Geistler, A., 2002. Train location with eddy current sensors. In *CompRail 2002*. WIT Press, pp. 12–14.

GEORAIL, 2008. *GEORAIL - Railway Geodesy - Guidelines for use of absolute coordinates in railway geo-referenced applications*,

GIS Resources, 2013. Fundamentals of GPS signal and data. *GISResources.com*. Available at: http://www.gisresources.com/fundamentals-of-gps-signal-and-data_2/ [Accessed April 4, 2017].

Glover, J., 2013. *Principles of Railway Operation*, Ian Allan Publishing.

GNSS Rail User Forum, 2000. *Requirements of Rail Applications*,

Google Earth, 2012. No Title.

Gowdayyanadoddi, N.S. et al., 2015. A ray-tracing technique to characterize GPS multipath in the frequency domain. *International Journal of Navigation and Observation*, 2015, p.16.

GRAIL, 2006a. *Application Survey*,

GRAIL, 2008. GNSS Subsystem Requirements Specification for Enhanced Odometry Application. , (30).

GRAIL, 2006b. *GRAIL: GNSS Introduction in the RAIL sector - Application Survey*,

Groves, 2013. *Principles of GNSS, inertial, and multisensor integrated navigation systems* 2nd ed.,

Groves, P., 2013. How Does Non- Line-of-Sight Reception Differ From Multipath Interference? *Inside GNSS*, (November/December), pp.40–44. Available at: www.insidegnss.com [Accessed September 12, 2016].

Groves, P., Handley, R. & Parker, S., 2009. Vehicle Heading Determination Using Only Single-antenna GPS and a Single Gyro. *Proceedings of the 22nd International Technical Meeting of the satellite division of the Institute of Navigation (ION GNSS 2009). (pp. 1775 - 1784)*.

Groves, P.D. et al., 2013. Context Detection, Categorization and Connectivity for Advanced Adaptive Integrated Navigation. In *Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2013)*. Nashville, pp. 1039–1056. Available at: http://discovery.ucl.ac.uk/1394967/1/ION_GNSS13_F2_1_Groves_et_al_1_0%28Context%29.pdf [Accessed October 6, 2016].

Groves, P.D. et al., 2012. Intelligent Urban Positioning using Multi- Constellation GNSS with 3D Mapping and NLOS Signal Detection.

GSA, 2005. *Galileo Integrated Receiver for Advanced Safety of Life Equipment*,

Gurnik, P. & Trégl, O., 2005. *Satellite based train location*,

Hall, S., 2005. *Modern Signalling Handbook*, Ian Allan Publishing.

Hartinger, H. & Brunner, K.F., 1999. Variances of GPS Phase Observations: The SIGMA-ε Model. *GPS Solutions*, 2(4), pp.35–43. Available at: http://dx.doi.org/10.1007/PL00012765.

Hegarty, C., 2013. GNSS Measurements and Error Sources. In *Workshop on GNSS Data Application to Low Lattitude Ionospheric Research*.

ICAO, 2006. *Annexe 10: Aeronautical Telecommunications, Volume I, Radio Navigation Aids, Sixth Edition*,

iNsight, 2011. Innovative navigation using new GNSS signals with hybridised technologies. *www.insight-gnss.org*. Available at: http://www.insight-gnss.org/index.html [Accessed April 4, 2017].

InteGRail, 2010. *InteGRail: Intelligent Integration of Railway Systems - Publishable Final Activity Report*,

International GPS Service, 2009. IGS Products. Available at: https://igscb.jpl.nasa.gov/components/prods.html.

Jokinen, A.S., 2014. *Enhanced ambiguity resolution and integrity monitoring methods for Precise Point Positioning*.

Julien, O., 2011. Concepts Avances du GNSS.

Kaplan, E.D. & Hegarty, C.J., 2006. *Understanding GPS Principles and Applications* 2nd ed.,

Kouwenhoven, T., 2011. DOP effect in positioning. *Navipedia*. Available at: http://www.navipedia.net/index.php/File:DOP_effect_in_positioning.png [Accessed October 6, 2016].

Kyriakidis, M., 2013. A framework to identify and quantify the most significant Performance Shaping Factors in Railway Operations.

Leandro, R.F., Santos, M.C. & Langley, R.B., 2006. UNB neutral atmosphere models: development and performance. In *Proceedings of ION NTM 2006, the National Technical Meeting of the Institute of Navigation, Monterey, California, 18-20 January 2006*. pp. 564–573.

LOCOPROL, 2005. LOCOPROL - Train Protection, Control and Command - Final Report. , p.74.

Marais, J. et al., 2012. Accurate Localisation Based on GNSS and Propagation Knowledge for Safe Applications in Guided Transport. *Procedia - Social and Behavioral Sciences*, 48, pp.796–805.

Marais, J., 2014. ERTMS, applications sécuritaires et applications satellitaires, quelles avancées de la recherche ? In *Recherche & Innovation Ferroviaires - Les bénéfices de la navigation par satellites européenne*. p. 25.

Marais, J. et al., 2015. Weighting with the pre-knowledge of GNSS signal state of reception in urban areas. In *European Navigation Conference 2015*. Bordeaux.

Marais, J., Beugin, J. & Berbineau, M., 2017. A Survey of GNSS-Based Research and Developments for the European Railway Signaling. *IEEE Transactions on Intelligent Transportation Systems*, (99), pp.1–17.

Mendes, V.B. & Langley, R.B., 1999. TROPOSPHERIC ZENITH DELAY PREDICTION ACCURACY FOR HIGH-PRECISION GPS POSITIONING AND NAVIGATION. *Journal of*

the Institute of Navigation, 46(1), pp.25–34.

Misra, P. & Enge, P., 2010. *Global positioning system : signals, measurements, and performance*, Ganga-Jamuna Press.

Moghtadaiee, V., Dempster, A.G. & Lim, S., 2011. Indoor Localization Using FM Radio Signals: A Fingerprinting Approach. In *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*.

Moradi, R., 2014. Carrier multipath mitigation in linear combinations of Global Navigation Satellite Systems measurements.

NASA, 2004. SOFTWARE SAFETY STANDARD. , (September 1997).

National Audit Office, 2015. *A Short Guide to Network Rail*,

Neri, A. et al., 2012. An Analytical Evaluation for Hazardous Failure Rate in a Satellite-based Train Positioning System with reference to the ERTMS Train Control Systems. , (May 2009), pp.1–15.

NGTC, 2016. Results & Publications. Available at: http://www.ngtc.eu/results-publications/.

Niell, A.E., 1996. Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research: Solid Earth*, 101(B2), pp.3227–3246. Available at: http://dx.doi.org/10.1029/95JB03048.

Novatel, 2014. *Inertial Explorer® - User Guide*,

Novatel, 2016a. *OEM6™ Series OEMStar™ Receivers*,

Novatel, 2011. *OEVM-1™ Series*,

Novatel, 2016b. *SPAN-CPT ™ SINGLE ENCLOSURE GNSS+INS RECEIVER DELIVERS 3D POSITION, VELOCITY AND ATTITUDE*,

Optimist on the run, 2008. Four Oaks Railway station. Available at: https://en.wikipedia.org/wiki/Four_Oaks_railway_station#/media/File:Four_Oaks_station_-_2008-01-26.jpg [Accessed October 31, 2016].

Optimist on the run, 2007. Gravelly Hill railway station. Available at: https://commons.wikimedia.org/wiki/File:Gravelly_Hill_-_2007-09-07.jpg.

Pascoe, R.D. & Eichorn, T.N., 2009. What is communication-based train control? *IEEE Vehicular Technology Magazine*, 4(4), pp.16–21.

Pattinson, M. et al., 2016. Monitor, Detect, Characterise, Mitigate and Protect – Introducing STRIKE3. *Proceedings of the Institute of Navigation (ION) GNSS+ 2016*. Available at: https://www.ion.org/gnss/abstracts.cfm?paperID=4030.

Penna, N., Dodson, A. & Chen, W., 2016. Assessment of EGNOS Tropospheric Correction Model.

POINT, POINT Software.

Lo Presti, L. & Visintin, M., 2015. Can you list all the properties of the carrier-smoothing filter? *Inside GNSS*.

Profillidis, V.A., 2013. *Railway Management and Engineering*,

Pullen, S. & Gao, G., 2012. GNSS Jamming in the Name of Privacy. *Inside GNSS*, (2), pp.34–43. Available at: http://www.insidegnss.com/node/2976.

Quddus, M., Ochieng, W. & Noland, R., 2005. Integrity of Map Matching Algorithms.

Rousseau, M. (ALSTOM) et al., 2004. LOCOLOC - Final presentation. , (December).

RSSB, 2012a. *Data analysis for a cost-effective GPS-based locator with simple augmentations - Locator performance report*,

RSSB, 2012b. Data analysis for a cost-effective GPS-based locator with simple augmentations - Requirements analysis for locator technology in GB Railways.

RSSB, 2010. *ETCS System Description*,

RSSB, 2007. *Global Navigation Satellite Systems - Data Coverage Analysis for Railway Operations*,

RSSB, 2003. *Interlocking Principles*,

Saastamoinen, J., 1972. Contributions to the theory of atmospheric refraction. *Bulletin Géodésique (1946-1975)*, 105(1), pp.279–298. Available at: http://dx.doi.org/10.1007/BF02521844.

Sánchez, J.S. et al., 2016. Use of a FishEye Camera for GNSS NLOS Exclusion and Characterization in Urban Environments. In *ION ITM 2016, International Technical Meeting*. Monterey, United States.

Sanz Subirana, J., Juan Zornoza, J.M. & Hernandez-Pajares, M., 2013. *GNSS Data Processing - Volume I: Fundamentals and Algorithms* European Space Agency, ed.,

SaPPART, 2015. *SaPPART White paper - Better use of Global Navigation Satellite Systems for safer and greener transport*,

Satirapod, C. & Wang, J., 2000. Comparing the quality indicators of GPS carrier phase observations. *Geomatics Research Australasia*, 73(2), pp.75–92.

Schon, W. et al., 2013. *Railway Signalling and Automation - Volume 1*,

Scot, G., 2014. Etat des lieux des programmes GNSS européens EGNOS et Galileo. In *Recherche & Innovation Ferroviaires - Les bénéfices de la navigation par satellites européenne*. p. 13.

Senesi, F., 2016. *ERSAT - ERTMS + Satellite*, Available at: http://www.fsitaliane.it/fsi-en/Media-and-

events/Press-releases-and-news/RFI-ERSAT-project-satellite-technologies-to-manage-regional-railway-traffic.

Stolagiewicz, A.A., 2009. *Contributions to the foundations of a safety case for the use of GNSS in railway environments*. University College London.

Tiberius, C.C.J.., 1999. The GPS data weight matrix: what are the issues? In *Proceedings of the 1999 National Technical Meeting of The Institute of Navigation*. Sab Diego, pp. 219–227.

TSLG, 2012. *The Future Railway - The Industry's Rail Technical Strategy 2012 Supporting Railway Business*,

Tu, J. V., 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11), pp.1225–1231. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0895435696000029 [Accessed April 4, 2017].

UK Parliament, 2012. Rail 2020. Available at: http://www.publications.parliament.uk/pa/cm201213/cmselect/cmtran/writev/rail2020/ror3a.htm

UNIFE, 2015. *Anuual Report 2014*,

UNIFE, 2012a. *ERTMS Factsheet 10 - Increasing infrastructure capacity - how ERTMS improve railway performance?*,

UNIFE, 2012b. *UNIFE welcomes rail sector commitment to ERTMS*, Copenhagen. Available at: www.unife.org [Accessed October 12, 2016].

UNISIG, 2012. *ERTMS/ETCS - FFFIS for Eurobalise, SUBSET-036*,

Urech, A., 2002. *GADEROS - Galileo Demonstrator for Railway Operation System*,

USA Department of Defense, 2008a. Global Positioning System Standard Positioning Service Performance Standard. , (September).

USA Department of Defense, 2008b. Global Positioning System Wide Area Augmentation System (WAAS) Performance Standard. , (October).

Vigneau, W. et al., 2006. Neural networks algorithms prototyping to mitigate GNSS multipath for LEO positioning applications. In *Proceedings of the ION GNSS 19th International Technical Meeting of the Satellite Division, 26-29 September 2006, Fort Worth, TX*. pp. 1752–1762.

Wang, L., Groves, P.D. & Ziebart, M.K., 2013. Urban Positioning on a Smartphone: Real-time Shadow Matching Using GNSS and 3D City Models. , pp.1606–1619.

Weiser, A. & Brunner, F.K., 2000. An extended weight model for GPS phase observations. *Earth Planets Space*, 52, pp.777–782.

Appendix I – Linearization of the pseudorange equations

The geometric range between the receiver and the i^{th} satellite can be expressed as:

$$\rho_i = \sqrt{(x_i - x_u)^2 + (y_i - y_u)^2 + (z_i - z_u)^2} + c.t_u \tag{I.1}$$

$$\rho_i = f(x_u, y_y, z_u, t_u) \tag{I.2}$$

Where (x_i, y_i, z_i) is the three dimensional position of the satellite, (x_u, y_u, z_u) is the position of the user receiver and t_u is the time offset between the receiver and the satellite.

As seen in expression (I.1), the relation between the range and the user position is non-linear. In order to estimate the user position, this expression is linearized with respect to the approximate position of the receiver.

Considering the user approximate position and receiver clock offset ($\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u$), the approximate pseudorange can be computed as:

$$\hat{\rho}_i = \sqrt{(x_i - \hat{x}_u)^2 + (y_i - \hat{y}_u)^2 + (z_i - \hat{z}_u)^2} + c.\hat{t}_u \tag{I.3}$$

$$\hat{\rho}_i = f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u) \tag{I.4}$$

An offset (Δx_u, Δy_u, Δz_u) between the true receiver position and the approximated one can be defined such as:

$$x_u = \hat{x}_u + \Delta x_u \tag{I.5}$$

$$y_u = \hat{y}_u + \Delta y_u \tag{I.6}$$

$$z_u = \hat{z}_u + \Delta z_u \tag{I.7}$$

$$t_u = \hat{t}_u + \Delta t_u \tag{I.8}$$

Therefore:

$$f(x_u, y_y, z_u, t_u) = f(\hat{x}_u + \Delta x_u, \hat{y}_u + \Delta y_u, \hat{z}_u + \Delta z_u, \hat{t}_u + \Delta t_u) \tag{I.9}$$

This function can be expanded about the approximate position and receiver clock offset using a Taylor series truncated after the first order partial derivatives.

$$f(\hat{x}_u + \Delta x_u, \hat{y}_u + \Delta y_u, \hat{z}_u + \Delta z_u, \hat{t}_u + \Delta t_u)$$
$$= f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u) + \frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{x}_u}\Delta x_u + \frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{y}_u}\Delta y_u \quad \text{(I.10)}$$
$$+ \frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{z}_u}\Delta z_u + \frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{t}_u}\Delta t_u$$

Replacing *f* by its expression in equation (I.1), the partial derivative terms can be computed as:

$$\frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{x}_u} = -\frac{x_i - \hat{x}_u}{\hat{r}_i} \quad \text{(I.11)}$$

$$\frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{y}_u} = -\frac{y_i - \hat{y}_u}{\hat{r}_i} \quad \text{(I.12)}$$

$$\frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{z}_u} = -\frac{z_i - \hat{z}_u}{\hat{r}_i} \quad \text{(I.13)}$$

$$\frac{\partial f(\hat{x}_u, \hat{y}_u, \hat{z}_u, \hat{t}_u)}{\partial \hat{t}_u} = c \quad \text{(I.14)}$$

where:

$$\hat{r}_i = \sqrt{(x_i - \hat{x}_u)^2 + (y_i - \hat{y}_u)^2 + (z_i - \hat{z}_u)^2} \quad \text{(I.15)}$$

By replacing equations (I.11), (I.12), (I.13) and (I.14) in (I.10), the pseudorange ρ_i can be computed as:

$$\rho_i = \hat{\rho}_i - \frac{x_i - \hat{x}_u}{\hat{r}_i}\Delta x_u - \frac{y_i - \hat{y}_u}{\hat{r}_i}\Delta y_u - \frac{z_i - \hat{z}_u}{\hat{r}_i}\Delta z_u + c\Delta t_u \quad \text{(I.16)}$$

The equation can be rearranged by introducing the following new variables:

$$\Delta\rho_i = \hat{\rho}_i - \rho_i \quad \text{(I.17)}$$

$$a_{xi} = -\frac{x_i - \hat{x}_u}{\hat{r}_i} \quad \text{(I.18)}$$

$$a_{yi} = -\frac{y_i - \hat{y}_u}{\hat{r}_i} \quad \text{(I.19)}$$

$$a_{zi} = -\frac{z_i - \hat{z}_u}{\hat{r}_i} \quad \text{(I.20)}$$

The vector $a_i = (a_{xi}, a_{yi}, a_{zi})$ represents the direction cosines of the unit vector pointing for the approximate user position to the i^{th} satellite.

The equation (I.16) can be rewritten:

$$\Delta\rho_i = a_{xi}\Delta x_u + a_{yi}\Delta y_u + a_{zi}\Delta z_u - c\Delta t_u \tag{I.21}$$

If the receiver acquire ranges from n satellites, the following set of equations are available:

$$\begin{aligned}
\Delta\rho_1 &= a_{x1}\Delta x_u + a_{y1}\Delta y_u + a_{z1}\Delta z_u - c\Delta t_u \\
\Delta\rho_2 &= a_{x2}\Delta x_u + a_{y2}\Delta y_u + a_{z2}\Delta z_u - c\Delta t_u \\
&\vdots \\
\Delta\rho_n &= a_{xn}\Delta x_u + a_{yn}\Delta y_u + a_{zn}\Delta z_u - c\Delta t_u
\end{aligned} \tag{I.22}$$

These equations can be written in matrices:

$$\Delta\rho = \begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \vdots \\ \Delta\rho_n \end{bmatrix}, \quad H = \begin{bmatrix} a_{x1} & a_{y1} & a_{z1} & 1 \\ a_{x2} & a_{y2} & a_{z2} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ a_{xN} & a_{yN} & a_{zN} & 1 \end{bmatrix} \text{ and } \Delta x = \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ -c\Delta t_u \end{bmatrix}$$

Leading finally to:

$$\Delta\rho = H\,\Delta x \tag{I.23}$$

Appendix II – Least squares estimation

After linearization of the pseudorange equations (see Appendix I) around the user position and time (i.e. estimation of the receiver position and time from an initial known position), the equations can be written in a matrix version:

$$\Delta\rho = H\,\Delta x + \varepsilon \tag{II.1}$$

where $\Delta\rho$ is a vector containing the terms of the linearized mathematical model, H is the design matrix, Δx contains the parameters that are being estimated, and ε represents the measurement noise.

The measurement noise ε is assumed to be Gaussian multivariate, such that:

$$E[\varepsilon] = 0 \quad and \quad cov(\varepsilon) = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_N^2 \end{pmatrix} = W^{-1} \tag{II.2}$$

where W is the inverse of the covariance matrix of the measurements

The least square position estimate $\widehat{\Delta x}$ is defined as the one which minimises the quadratic form of the residuals, subject to the constraints defined by equation (II.1).

$$\widehat{\Delta x} = \min_{\Delta x} \varepsilon^T W \varepsilon \tag{II.3}$$

An objective function J can then be defined as:

$$J(\Delta x) = \frac{1}{2}\varepsilon^T W \varepsilon = (\Delta\rho - H\Delta x)^T W (\Delta\rho - H\Delta x) \tag{II.4}$$

The derivative with respect to Δx is

$$\frac{\partial J(\Delta x)}{\partial \Delta x} = -2H^T W \Delta\rho + 2H^T W H \Delta x \tag{II.5}$$

When the columns of H are linearly independent, the second order derivative $\partial^2 J/\partial\Delta x \partial\Delta x^T = X^T W X$ is positive definite. Therefore the function J has a minimum. The least squares estimate for Δx, noted $\widehat{\Delta x}$, is found by setting $\partial J/\partial\Delta x = 0$.

$$\widehat{\Delta x} = \ (H^T W H)^{-1} H^T W \Delta\rho \tag{II.6}$$

The estimate for $\Delta\rho$, termed $\widehat{\Delta\rho}$ is:

$$\widehat{\Delta\rho} = \ H\widehat{\Delta x} = H(H^T W H)^{-1} H^T W \Delta\rho = M\Delta\rho \tag{II.7}$$

where $M = H(H^T W H)^{-1} H^T W$

Using equation (II.7), the vector of least squares residual $\hat{\varepsilon}$ is computed as:

$$\hat{\varepsilon} = \Delta\rho - \widehat{\Delta\rho} = \Delta\rho - M\Delta\rho = (I - M)\Delta\rho \quad \text{(II.8)}$$

The covariance matrices of the estimates $\widehat{\Delta x}$, $\widehat{\Delta\rho}$ and $\hat{\varepsilon}$ are then computed as:

$$Cov\left(\widehat{\Delta x}\right) = \hat{\sigma}^2 (H^T H)^{-1} \quad \text{(II.9)}$$

$$Cov\left(\widehat{\Delta\rho}\right) = \hat{\sigma}^2 M \quad \text{(II.10)}$$

$$Cov(\hat{\varepsilon}) = \hat{\sigma}^2 (I - M) \quad \text{(II.11)}$$

Appendix III – Presentations and Papers

Damy, Sophie, Majumdar, Arnab & Ochieng, Washington Y. (2016), GNSS-based High Accuracy Positioning for Railway Applications, *Proceedings of the 2016 International Technical Meeting of The Institute of Navigation, Monterey, California, January 2016, pp. 1003-1014*.

Damy, Sophie (2016), Bringing Train Travel into the Space Age, *Imperial College London Graduate School Three Minute Thesis Competition, London, April 2016*

Damy, Sophie (2015), Navigating through the space industry, *2nd Annual Space Industry Careers Fair, Leicester, February 2015*

Damy, Sophie (2014), A Satellite-based Optimisation of the European Freight Transport Network, *The European Interparliamentary Space Conference, Paris, October 2014*

Damy, Sophie, Ochieng, Washington & Majumdar, Arnab (2014), Space-based technologies to support railway operations, *Imperial SpaceLab 2014 , London, September 2014*

Damy, Sophie (2014), SPoT – Smart Positioning of Trains, *Transport Research Arena Visions, Paris, France, April 2014*

Damy, Sophie (2013), Novel GNSS-based positioning system to support railway applications, *GNSS PhD Summit, Munich, November 2013*