# Digital Route Model Aided Integrated Satellite Navigation and Low-cost Inertial Sensors for High-Performance Positioning on the Railways

Ziyi Jiang

Department of Civil, Environmental and Geomatic Engineering

UCL

September 2010

Supervisors:
Professor Paul Cross
Professor Marek Ziebart

Thesis submitted for the degree of Doctor of Philosophy

I, Ziyi Jiang confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:.....

Date:.....

## Abstract

The basis of all railway signalling activities is the knowledge of the position and velocity of all trains in the system. The railways traditionally rely on train detection systems for this knowledge. However, the dependence of these systems on railway infrastructures limits their ability to cope with the advent of new high-speed lines and the development of freight networks across the Europe. Hence, there is a need for the introduction of modern positioning technologies into the railways. Unfortunately railways provide an unfriendly environment for satellite-based radio positioning systems (GNSS). For this reason it is common to integrate GNSS with low-cost inertial sensors (INS) but such systems cannot meet all railway positioning requirements. This thesis examines the potential of enhancing such an integrated GNSS/INS system with a digital route model (DRM).

The study is carried out through a series of simulations of typical railway positioning scenes. A simulated database of GNSS, inertial and DRM data is built from real GPS data collected on a rail line between Norwich and Lowestoft. Several tests are first performed to test the validity of the database. Simulations are then done with a number of traditional INS/GPS integration architectures to test the possible performance of each system in the railway environment using low-cost INS sensors. The DRM-aiding is then realised through an integration with the GNNS/INS system via an extended Kalman Filter.

Results from the study confirm the need for additional positioning information for an integrated system with low-cost inertial sensors to deal with difficult satellite signal situations such as tunnels, deep cuttings and covered stations. It is shown that a DRM leads to significant improvements in the overall system positioning performance. Also the optimal configuration, in terms of point spacing and accuracy, for a digital route model is selected from amongst simulated candidates.

# Contents

# List of Figures

# List of Tables

# Nomenclature

ABS Automatic Block Signalling

AC Alternating Current

AEP The Architecture Evolution Plan for GPS modernisation

AHRS Attitude and Heading Reference System

ARNS Aeronautical Radionavigation Service

ARS Automatics Route Setting

ATP Automatic Train Protection

AWS Automatic Warning System

C/A-codes Coarse and Acquisition codes for GPS signals

CDMA Code-Division Multiple Access

CS The control segment of GPS

DC Direct Current

DGPS Differential GPS

DMI Driver Machine Interface

DOP Dilution Of Precision

DR Dead Reckoning

DRM Digital Route Model

ECEF Earth-centred Earth-fixed

EGNOS European Geostationary Navigation Overlay Service

EKF Extended Kalman Filter

EOR European Operational Rules

ERTMS European Rail Traffic Management System

ESA The European Space Agency

ETCS European Train Control System

ETML European Traffic Management Layer

EU The European Union

EVC European Vital Computer

FDMA Frequency-Division Multiple Access

FEC Forward Error Correction

GBAS Ground Based Augmentation Systems

GBCC The Ground-Based Control Complex of GLONASS

GLONASS GLObal'naya Navigatsionnaya Sputnikovaya Sistema

GNSS Global Navigation Satellite System

GPS Global Positioning System

GSA GNSS Supervisory Authority

GSM–R Global System for Mobile communication – Railways

IGS The International GNSS Service

IMU Inertial Measurement Unit

INS Inertial Navigation System

ITRS The International Terrestrial Reference System

ITU International Telecommunications Union

JPO Joint Program Office

L-AII The Legacy Accuracy Improvement Initiative for GPS

L1 The Link 1 frequency of GPS signals

L2 The Link 2 frequency of GPS signals

L2C GPS Link 2 Civil signal

L5 The Link 5 frequency of GPS signals

LEU Lineside Electronics Unit

LOS Line-Of-Sight

MA Moving Authority

MCS The Master Control Station of GPS

MEMS Micro-machined Electromechanical System

OCX The Operational Control Segment of GPS

OCXO Oven-Controlled Crystal Oscillator

OS Galileo Open Service

PRN Pseudo-Random Noise

PRS Galileo Public Regulated Service

RAIM Receiver Autonomous Integrity Monitoring

RF The Radio Frequency

RNP Required Navigation Performance

RSSB Rail Safety and Standards Board

SBAS Satellite Based Augmentation Systems

SoL Safety-of-Life

STM Specific Transmission Module

TCXO Temperature-Compensated Crystal Oscillator

TEC Total Electron Content

TIU Train Interface Unit

TT&C Telemetry, Tracking and Control

UERE The user pseudorange error

UTC The Coordinated Universal Time

XO quartz crystal oscillators

## Acknowledgement

I am heartily thankful to my supervisors for their support over the course of my research, particularly Professor Paul Cross his guidance throughout. I would like to thank Professor Marek Ziebart for his advices and encouragements. It is an honour for me to finish this research under their supervision.

I am very grateful to the colleagues in the Space Geodesy and Navigation Laboratory for their numerous helps during my research. In particular, it is a pleasure to thank Dr Alexander Parkins, Mr Santosh Bhattarai and Mr Toby Webb for their time and efforts to proofread this thesis, and for their genuine suggestions to improve this thesis. It is always a delightful experience to recollect those discussions with Mr Peter Stacey and Mr Mojtaba Bahrami. I would also like to thank Dr Paul Groves for his help of my time management for the thesis writing.

The railway data were generously supplied by the Nottingham Scientific Limited and Rail Safety and Standards Board.

Memories for those who studied and worked in the department of geomatic engineering, Dr Zhenhong Li, Dr Lawrence Lau, Dr Andrew Stolagiewicz and Dr Joe Bradbury, will always be as warm as when I came to London four years ago. I am indebted to them for showing me the academia and life of doing research.

To my parents, without whom this thesis would not have been possible ...

To Xiao-wei, for supporting me through out this research ...

To Yi-lin, you are my biggest joy ...

仅以此文,

献给在我四年博士时光中给予我无限支持的父亲和母亲 ... ...

献给一直在身边默默支持和鼓励我的晓威 ... ...

献给带给我欢乐与感动的怿琳 ... ...

# 1. Introduction

## 1.1. Motivation

Global Navigation Satellite System (GNSS) has already been used in many non-safety-of-life applications on the railways, such as providing train arrival and departure information for the passenger. But safety-of-life applications involving train positioning for traffic control are still challenging for a GNSS stand-alone system. Although the accuracy of GNSS positioning in the railway environment is in general sufficient, it is its availability that limits the overall system performance. This is due to physical obstructions of the satellite signals along the rail route. Service outages can last for several minutes, especially in urban areas (RSSB, 2008).

Current research for high-performance positioning on the railways mainly focuses on the usage of integrated navigation systems. Dead-reckoning sensors such as odometers (Ernest *et al.*, 2004), inertial sensors (Genghi *et al.*, 2003) and other sensors such as Doppler radar (Mertens *et al.*, 2003) are commonly chosen to integrate with GNSS. However, the integrated positioning systems developed by current research are unsatisfactory from either performance or cost perspective. This is mainly because of two reasons: firstly, the integration level, i.e. the system integration architecture, of current systems is still kept at a high level using only processed positioning solutions from each sub-system (known as uncoupled system and loosely-coupled system); secondly, the fact that the movement of the train is constrained to the tracks is not considered during the system design.

This thesis therefore aims to investigate methods to integrate various sensors and data sources with GNSS at a lower level (known as tightly coupled system) to improve the system positioning performances (especially for availability) under the railway environment. This means the system integration proposed in this research will be mainly carried out at the measurement level instead of using positioning solution directly. Low-cost inertial sensors are used for the system integration research in this thesis to investigate the potential benefits they may bring to the safety-of-life positioning.

A Digital Route Model (DRM) is a database containing the position knowledge of surveyed points along the centre line of a rail route. The restricted movement of the train reduces the positioning task of solving its position in a three dimensional free space into a one dimensional problem which seeks the position along a predetermined trajectory. This predetermined trajectory is described by a DRM. The aiding information provided by a DRM can be expected to be helpful to maintain the availability of the positioning system during GNSS outages on the railways, but by how much is unknown.

A method of integrating DRM with GNSS and low-cost inertial sensors is developed and investigated in this thesis. The DRMs used by the new integrated system are built using different specifications, and how they affect the performance of DRM-aiding is hence studied.

## 1.2. Background

### 1.2.1. Train Detection and Signalling on the Railways

The railway is a guided transport system, which means the movement of trains is confined to the rail tracks. As a result, the relative position between two trains is more deterministic on the railways than other transport systems, i.e. a train can only be behind or ahead of other trains on the same track. Hence the only way for a traffic management to avoid train collisions is to make sure that each train arrives at the same point at a different time. A unique traffic control system has been developed on the railways to maintain a safe distance between trains on the same track. The train separation is maintained through a signalling mechanism that issues the authorisation for a train to precede on its journey on the track. The signalling mechanism on the railways is known as block signalling systems.

There are two types of block signalling system: fixed block and moving block (Health and Safety Executive, 2005). In a fixed block system, the rail route is separated into fixed small sections, and the train separation consists of several small sections. A moving block system, by contrast, is able to maintain a dynamic train separation based on the current position and velocity of trains. Deploying moving block systems is the future trend for railway signalling development (Department for Transport, 2007), and brings a need for high-performance positioning systems. A detailed system description of both signalling systems is given in Section 2.1.4.

The knowledge of accurate and up-to-date information on the position and movements of trains is the basis of signalling and traffic control operations. In modern railways, this information is provided by a train detection system. Various train detection sensors are integrated as parts of the railway infrastructure along the rail tracks. The presence of a train on a track section is automatically detected and input into the signalling system. Due to the limitation of along-track sensors, the detection results provided by the train detection system are always related to a certain part of the track and no higher resolution results are available.

Traditional train detection sensors are still widely used across the railways. There are two types of these traditional sensors that are usually deployed along the tracks for different train detection scenarios (Bailey, 1995). The first type of sensor detects the presence or absence of trains within the limits of a track section by using track circuits or axle counters. This type of sensor is commonly used along the route to indicate the section occupation status. The second type detects that a train has reached, is passing, or has passed a specific position on the tracks with wheel detectors and treadles. This type of sensor is commonly used at critical positions on the line, such as switch points, to provide train approach information. Detailed descriptions of traditional train detection sensors are presented in Section 2.1.2.

The main problems with traditional train detection systems are their maintenance and unfitness for future signalling development (Palmer, 2006). The maintenance problems for traditional train detection systems are both financial and safety based. The size of modern railway network makes it difficult and expensive to maintain the normal operation of sensors across the network infrastructure. Also, as a result of the complex environment of modern railways, it becomes difficult to maintain the appropriate safety and reliability standards for these traditional sensors with the pace of railway development.

An initiative led by European Union as an update of current railway systems among its members, the European Rail Traffic Management System, aims to eventually deploy a moving block signalling system across the European rail network (Bloomfield, 2006). The signalling and control component of the initiative, known as European Train Control System, proposes a progressive plan of updating

the current train detection systems. The new train control system takes the train as part of the new detection mechanism by deploying each train with onboard positioning functions and replaces the traditional sensors with new and more reliable sensors (RSSB, 2010). The ultimate goal of the new control system is to reduce the dependency of train position resolution on the railway infrastructure and to let the train determine its own position and velocity. A detailed description on the future train control system is given in Section 2.2.

The realisation of a train position determination mechanism that does not rely on track side infrastructure raises the need for modern positioning and navigation technologies. Since the railway is a public transportation system, positioning for trains becomes a safety-of-life application, which not only requires accuracy but also reliability for the positioning system performance. On the other hand, the railway is a difficult working environment for positioning systems, especially for GNSS. A detailed discussion on the railway positioning environment is shown in Section 2.3.1. As a result, an integrated positioning system is needed. However, for a large scale deployment of positioning systems across the railways network, the cost must also be considered. Hence an integrated system using low-cost low-performance inertial sensors aided by DRM for a safety-of-life application on the railways is proposed.

### 1.2.2. Integrated Navigation and Kalman Filtering

In integrated navigation, different sensors and data sources are combined to achieve an optimal positioning solution under various operation conditions. An integrated navigation system has the advantage of an improved positioning accuracy and greatly enhanced reliability, as a result of an optimal combination of measurements from sensors with complementary error characteristics. Although the choice of sensors and data sources used in an integrated system varies with the application, as a general principle for sensor selection, it is preferable to use sensors with complementary characteristics. It is usually chosen to integrate systems based on position fixing and systems based on dead reckoning.

Typically, for such an integrated system, the dead reckoning system is used as the reference system as it operates continuously at a higher rate than most position fixing systems, and its solutions are corrected using position-fixing measurements available in the system through the integration algorithm. Other sensors and data sources operate as aiding systems and provide additional information to be processed by an integration algorithm. The output of the integration algorithm contains corrections of the reference solution and estimated error characteristics of sub-systems. The final integrated solution is obtained by applying corrections to the reference solution. Error characteristics are fed back to relative sub-systems for calibration. The integration algorithm used by an integrated system is usually an optimal system state estimator, most commonly a variation of the Kalman filter (Crassidis & Junkins, 2004).

Some critical aspects for the design of an integrated system include sensor selection, system integration architecture and the adaption of the integration algorithm. Since the design of a system is often a trade-off among multiple factors, such as positioning accuracy, output update rate, system reliability, budget, size and mass, and whether vehicle attitude information is required in addition to position and velocity information, systems for various applications can have very distinctive designs.

The way that information from different sub-systems is combined depends on the integration architecture of an integrated system. Common integration architectures include cascaded integration, centralised integration and federated integration (Groves, 2008). For a cascaded integration, measurements from each sub-system are processed by their own navigation processor, and preliminary solutions of sub-systems are then used as inputs for the integration algorithm to produce an integ-

rated solution. The integration algorithm of a centralised integration, on the other hand, uses the measurements from sub-systems directly; no pre-processing of the measurements is needed. A federated integration firstly integrates the reference solution with measurements from each sub-system in its own navigation processor and then all integrated solutions are finally combined together with a weighting integration algorithm. Detailed descriptions of different integration architectures are given in Section 3.1.

The most commonly used integration algorithm in an integrated navigation system is known as the Kalman filter (Kalman, 1960). The Kalman filter is a mathematical algorithm used for optimal estimation. It estimates the states of a system using observed measurements that contains noise. In an integrated system, it serves as the key to obtain an optimal estimation from multiple sensors and data sources. The operation of a Kalman filter is a recursive process. Real-time estimates are maintained through a continuous propagation and update process of the system states. The knowledge of the deterministic and stochastic models of the system is used to propagate estimates of the states through time. Corrections of the propagated estimates are obtained by a weighted average with the new measurements from sensor outputs, known as measurement update. As a result, not only are measurements from the current epoch utilised for estimation, but also information from prior estimates. This is particularly efficient for a real-time application, as although old data may be discarded, prior information is still optimally stored for future estimation. A detailed description on Kalman filtering process is presented in Section 3.2.

For most personal positioning, road vehicle and asset-tracking applications, where the key factors may be cost, size, weight and power consumption, a GNSS alone system generally can be competent. Nevertheless, safety-of-life applications, such as civil aviation and the railways, have particular concerns about the system reliability, while still demanding a very high accuracy requirement. Systems designed for these applications usually take advantage of sensor integration. Although low-cost low-performance inertial sensors are not commonly used for safety-of-life applications alone, their application in an integrated system still needs to be explored. Moreover, similar to the way train detection systems use the physical track infrastructure to estimate train location, it is also possible to feed the position information of the track, e. g. a track database or DRM, into an integrated navigation system. Research has already shown accuracy and integrity improvements using track database with the GPS on railways (Zheng, 2008). However, no research on the use of a DRM to enhance a tightly coupled integrated system (integrating inertial sensors with GNSS at a lower level, i.e. using satellite ranging measurements) has been explored to data.

### 1.2.3. GNSS Positioning on the Railways

Global Navigation Satellite System (GNSS) is a term for satellite-based radio navigation systems that are capable of providing a 3D positioning solution for the user. Currently, the American GPS and the Russian GLONASS are the only two GNSSs in full operational capacity. More GNSSs are under development, such as the European Galileo system and the Chinese Compass. Detailed descriptions of different GNSSs are given in Section 4.1. The railways are shown as a difficult environment for modern positioning technologies, especially GNSS. The research for positioning applications of GNSS on the railways is actively carried out at present. For the research in this thesis, the GPS is assumed to be used for satellite positioning as it is the most widely used GNSS.

The most common way of using GPS for user position and velocity estimation is through the satellite ranging measurements produced by the receiver, or more specifically pseudoranges and pseudorange rates. Satellite signals carrying ranging information and navigation messages travel between each satellite and the receiver, and are processed by the ranging processor of the receiver to

produce pseudoranges and pseudorange rates. A pseudorange is the apparent range from a satellite at the signal transit time to the receiver at the signal receive time, whereas a pseudorange rate can be viewed as a projection of the relative velocity between the satellite and the receiver on the satellite to receiver line-of-sight direction.

The algorithm used by the receiver navigation processor to estimate its position and velocity can be based either on least-squares or Kalman filtering(Conley *et al.*, 2006). The least-squares based estimation algorithm works in a snap-shot mode. The ranging measurements of the same time epoch are processed together in an iterative least-squares process to produce positioning solutions for the current epoch. The Kalman filtering based estimation, on the other hand, uses a system model with pre-assumed receiver dynamics to propagate the positioning solution. Whenever satellite ranging measurements are available, a measurement update is performed to correct the predicted positioning solution. A detailed description of GPS ranging measurements and positioning solution estimation process are given in Section 4.2.

The railway operating environment contains physical obstructions for the satellite signals. Situations such as travelling through a tunnel or going into areas where the sky will be partly or completely obstructed causes either performance degradation or complete loss of service for GNSS working on the railway. The problems introduced by these situations are referred to as line-of-sight problems. An example of a typical train journey through an urban area is shown in Section 2.3.2 to illustrate the problem. For safety-of-life applications, the reliability of the positioning system is usually the first consideration. Thus integrating GNSS with other data sources become necessary.

Much research has been carried out (Mertens *et al.*, 2003; Simsky *et al.*, 2004; Zheng, 2007, 2008) to implement a modern positioning system using GNSS on the railways. The focus of the research is mainly on two aspects: utilising the fact of a guided train movement to aid the GPS receiver and bridging the satellite signal outage area with an integrated system. While research for the usage of guided train movement is currently limited to aid the GPS receiver only, research on developing an integrated system is focused on the implementation of a cascaded system based on the loosely-coupled principle and using high-performance inertial sensors. The main focus of this thesis is more detailed research on exploring different integration architectures and combining the usage of a digital route model and low-cost inertial sensors. A review of current positioning research on the railways is given in Section 2.3.3.

GNSS is under extensive development and modernisation. Modernised GNSS signals are broadcast with increased power and enhanced characteristics for multipath mitigation, which will significantly improve the signal-difficult situations on the railways as increased signal power increases the ability of signals to penetrate roofed stations and obstructions. Moreover, the presence of multiple constellations will potentially increase the overall availability along a rail line. The use of GNSS for safety-of-life applications increases the importance for system integrity monitoring, and future GNSSs will include system integrity information in the broadcast signals. All these developments of GNSS will be beneficial for safety-of-life applications.

### 1.2.4. Inertial Navigation Using MEMS Sensors

Inertial navigation is a dead reckoning navigation method where a navigation solution is obtained by continuously tracking changes of the vehicle's accelerations and their orientation (Chatfield, 1997). An inertial navigation system is often characterised by its continuous self-contained working mode and high output rate. Since inertial systems do not rely on the transmission of signals from the vehicle or any external source, they do not suffer from signal transmission difficulties as many radio positioning systems do. Consequently, the inertial system is usually chosen as the

reference system in an integrated system. However, the accuracy of an inertial system relies upon the knowledge of vehicle initial conditions. Initialisation and alignment process for an inertial system without external aiding usually involves intensive calculations and can be significantly longer than other positioning methods under difficult circumstances. In addition, as the inertial calculation is essentially a continuous integration process, the solution quality of an inertial system degrades as time progresses. A regular calibration of the system is therefore necessary to ensure the accuracy of the solution.

Three sets of inertial sensors, together commonly referred to as inertial measurement unit (IMU), are usually configured in three orthogonal axes so that each of them can measure the vehicle movement on a single direction. Each set of sensors in an IMU consists of an accelerometer, which measures the specific force, and a gyro, which tracks the turning of the axis. All inertial measurements are processed by the navigation processor, which contains an integrator to obtain the navigation solution. A detailed description on inertial navigation process is given in Section 5.4.

For an inertial system to be able to work independently, the inertial sensors must be capable of making very accurate measurement of the specific force and the angular rate of the vehicle. Micromachined electromechanical system (MEMS) based inertial sensors are a type of solid-state inertial sensor built with quartz and silicon. In comparison with conventional inertial sensors which usually contains a large number of parts requiring high-precision manufacturing and assembling techniques, MEMS inertial sensors feature small size, low weight, rugged construction, low power consumption, short start-up time, and reduced production cost. Nevertheless, their performance is limited at the current stage of development. As a result, independent inertial systems seldom use MEMS based sensors for navigation purposes. An integrated system, on the other hand, could still benefit by including these low-cost sensors, as they provide a better resolution of the dynamic status of the vehicle than other position fixing techniques.

## 1.3. Objectives

The objectives of this thesis are to explore the capability of a low-level integrated navigation system using low-cost inertial sensors on the railways and the potential of enhancing such systems with a digital route model to provide a high-performance positioning system for the railways. The detailed objectives of the thesis are summarised as following:

- What performance levels can low-cost MEMS/GNSS positioning systems achieve with various integration architectures in different GNSS signal situations in a railway environment?
- What improvement can the introduction of a DRM bring to an integrated low-cost MEMS/GNSS positioning system in an open railway environment?
- How can a tightly integrated low-cost MEMS/GNSS positioning system with aiding from a DRM perform under difficult GNSS conditions in a railway environment? What happens to this performance if the GNSS signal is completely lost?
- How can a DRM aid the initialisation process of a tightly integrated low-cost MEMS/GNSS positioning system when no, or only limited, GNSS signals are available?
- How should the information in a DRM be used so that it can be most beneficial to an integrated low-cost MEMS/GNSS positioning system?
- How does the accuracy of individual DRM points affect the performance of a DRM in aiding a low-cost MEMS/GNSS integrated positioning system?

- What is the optimal distance between adjacent DRM points in a DRM-aided integrated low-cost MEMS/GNSS positioning system?
- How would the use of a mismatched DRM (e.g. for the wrong section) affect the results of a DRM-aided integrated low-cost MEMS/GNSS positioning system?

## 1.4. Methodology

The methodology for this research comprises three parts: a simulation database providing both the truth and sensor measurements for the research, various integrated systems to be tested and designed testing scenarios for positioning on the railways study. A summary of the all three parts is given in Figure 1.1.

Terms commonly used to describe the performance of a positioning system include accuracy, integrity, continuity and availability. Accuracy is defined as the ability of a system to maintain a certain level of positioning error with a 95% confidence (Ober, 2003). Hence the definition of accuracy implies the degree of conformance between the estimated position and velocity and their true values. Integrity is a measurement of the trust in a positioning system. An integrity risk is the probability of an undetected failure of meeting the specified accuracy. Continuity is the ability of a system to maintain the required performance over a certain time window even if a fault occurs. Availability is the proportion of time that the system service is usable. The system must operate with accuracy, integrity and continuity requirements met in order to been considered available.

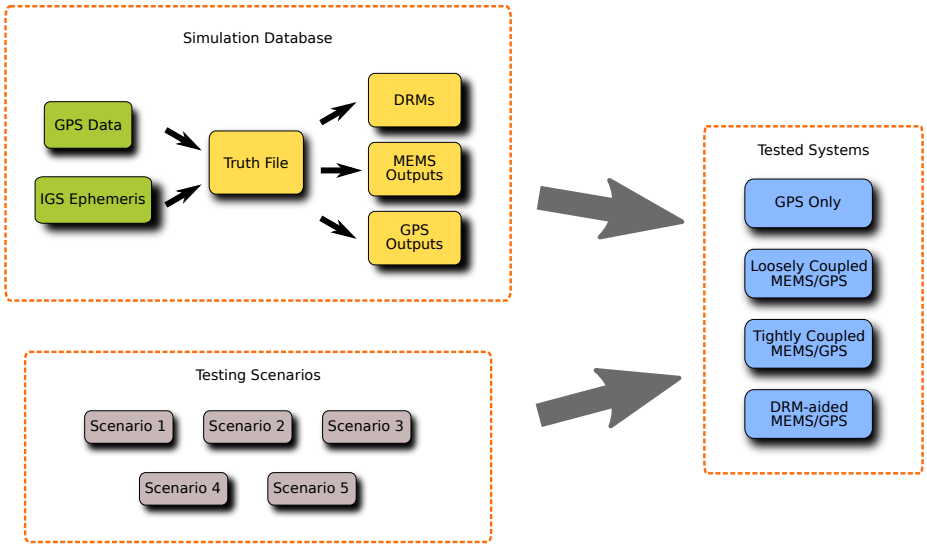



**Figure 1.1.:** Summary of the research methodology

### 1.4.1. Building a Simulation Database

A simulation database was built for this research in order to study the performance of an integrated system in the railway environment. Simulated sensor measurements were used to test various systems studied in this thesis. As a simulation environment provides the researcher full control of the positioning process, i.e. a known truth file and designed simulation scenarios, therefore it becomes possible to isolate each sensor and individually study their potential performance under a certain environment.

All simulated data produced in this research is based on real GPS data collected during a real train journey. A suburban and countryside rail route was selected for data collection, because it provides an environment that is almost free from severe physical obstructions, and GPS data collected on the route is thus more continuous and does not suffer from satellite signal outages. This enables of production of a high quality truth file.

The simulated train journey lasts for 32 min and the length of the journey is approximately 33 km. The train starts from being stationary and goes through a series of velocity changes, with a maximum speed of 169 km/h. Attitude changes of the carriage along the journey are simulated according to the actual physical and dynamic status of the train, calculated from movement of the train and changes of the track status through the journey.

The simulation database includes the status of the train carriage throughout the journey, information on the tracks along the route, and sensor outputs. The status of the carriage is expressed in terms of its movement and the changes of its attitude along the route. The train movement is shown by a truth file of its position, velocity and acceleration, whereas Euler angles (see Appendix B) are used to indicate the carriage attitude. Information on the rail tracks contains the position of the tracks, superelevation along the route, and simulated DRMs. The sensor outputs include simulated measurements from a MEMS inertial unit, satellite ranging measurements, and GPS position and velocity solutions.

A truth file of the train position, velocity and acceleration is generated from the collected GPS data. The changes of Euler angles are calculated based on the simulated track information throughout the journey. This truth file is used in the research as the main reference for system performance comparisons and testing results study. All system performances and results from proposed test scenarios are compared with the truth file to generate the real positioning errors.

The DRMs used in this research are generated based on the truth file. Different specifications are used to generate different DRMs in order to study the effect of using different DRMs for aiding. The DRM point interval values used in this research are 10 m, 50 m, 100 m and 500 m. Different levels of errors are also added to the DRM truth data to generate the simulated DRM data. The added errors are assumed to be zero-mean normally distributed, with different standard deviation levels at 0.1 m, 1 m and 10 m.

The simulated MEMS inertial measurement includes specific force from accelerometers and angular rate from gyros. The truth data for specific force and angular rate are generated based on the truth file. The IMU output is simulated by adding error characteristics to the truth data. The error characteristics simulated for this research are biases and random noise. The error characteristics used for the simulated inertial sensor output are based on available commercial IMUs.

Satellite ranging measurements are built upon an IGS final product ephemeris along with the truth file. Corrected pseudorange measurements and pseudorange rate measurements are simulated by adding errors to the true range and range rate data. Simulated ranging errors include the receiver clock error and the miscellaneous unmodeled range error for the individual satellite plus residuals after applying corrections from navigation data. GPS position and velocity solutions are generated by a Kalman-filter-based navigation processor using simulated ranging measurements.

### 1.4.2. General System Performance Study

The performance of several types of positioning systems is tested and studied in this thesis. Positioning systems studied in this thesis include a GPS only system using a Kalman-filter-based navigation processor, a loosely coupled MEMS/GPS integrated system, a tightly coupled MEMS/GPS integrated system and a DRM-aided MEMS/GPS integrated system using tightly coupled approach.

Note that apart from the GPS only system, all integrated systems use the solution from an inertial navigation processor as their reference solutions. Simulated MEMS measurements are processed by the inertial navigation processor to produce the reference inertial solution for each tested integrated system.

The GPS only system uses the simulated satellite ranging measurements. An extended Kalman filter is used as the system navigation processor to process the measurements and obtain train position and velocity solutions. The obtained position and velocity are also used as an GPS input for the loosely coupled MEMS/GPS integrated system. A detailed description is given in Section 4.2.4.

The loosely coupled MEMS/GPS integrated system adopts the cascaded integration architecture. GPS is used as an aiding system, whereas MEMS measurements with a higher update rate are used to produce a reference solution for the system integration. The system uses the position and velocity solution from the GPS only system as its external measurements. The integration algorithm estimates the errors in the reference solution and the inertial sensor error characteristics. A detailed description is given in Section 7.1.

The tightly coupled MEMS/GPS integrated system adopts the centralised integration architecture. GPS is still used as an aiding system, but the satellite ranging measurements are used as inputs instead of GPS solutions. MEMS measurements produce a reference solution for the system integration. The integration algorithm estimates the errors in the reference solution, the inertial sensor error characteristics and receiver clock errors. A detailed description is given in Section 7.2.

The proposed DRM-aided MEMS/GPS integrated system adopts the centralised integration architecture. Apart from GPS satellite ranging measurements, DRM position measurements information plus its constraining error model is also used as external measurements. All observed measurements are direct inputs of the integration algorithm. A detailed description is given in Section 8.1.4. DRM-aided MEMS/GPS integrated systems using DRMs with different specifications are also tested to study the effect of DRM specification as part of the research objectives.

The general performance of all tested systems is evaluated by comparing solutions from each system with the simulation truth file. The comparison between positioning errors is done by statistics such as root mean square errors, 95% error (the error at the 95% position from the first error value when arranging all errors in an ascending order). While the root mean square error shows the general scale of the positioning errors, the 95% error is used as an indication of the noise level on the positioning errors. In addition, the predicted $1\sigma$ system error (about 68% of all system errors assuming a normal distribution) information from each system is also used for system performance comparison. Apart from comparing with each other, statistics such maximum and median of the predicted $1\sigma$ system errors are used to compare with the actual positioning errors.

### 1.4.3. Test Scenarios

Five test scenarios are proposed and simulated in this thesis. These test scenarios cover typical situations on the railways where GNSS only positioning performance can be affected. Table 1.1 shows a summary of the test scenarios and the tested systems. The first three scenarios are tested for all systems, whereas the last two are designed for testing on DRM-aiding.

Scenario 1 simulates situations on the railways where the sky is completely blocked and GNSS signals are lost. This mostly happens when a train travels through a tunnel or sheltered areas. According to the general GNSS availability study performed by LOCASYS (see Section 2.3.2), the GNSS-unavailable situations on the railways usually only last for a short period, i.e. less than a minute. Therefore the length of the testing scenario is set as 30 s.

**Table 1.1.:** Summary of test scenarios and tested systems

| | | GPS only | Loosely coupled | Tightly coupled | DRM-aided |
|---|---|---|---|---|---|
| 1 | Loss of GPS signals | ▲ | ▲ | ▲ | ▲ |
| 2 | Difficult GNSS conditions | ▲ | ▲ | ▲ | ▲ |
| 3 | Constellation Geometry | ▲ | ▲ | ▲ | ▲ |
| 4 | Initialisation | | | ▲ | ▲ |
| 5 | Mismatched DRM | | | | ▲ |

Scenario 2 simulates situations on the railways where only a part of the sky is visible and the number of available satellites is less than four. This mostly happens when a train travels through a urban area, or stops at stations during the journey. The length of the testing scenario is set as 30 s.

Scenario 3 simulates situations on the railways where only half of the sky is visible but the number of available satellites is more than four. This mostly happens when a train travels through a urban area, or stops at stations during the journey. The length of the testing scenario is set as 30 s.

Scenario 4 simulates situations on the railways where the train stops at train station under a roof and its positioning system needs to be initialised. Under normal operation conditions, the initialisation process for a inertial unit can be aided by the GPS solution. When the GPS solution is not available, results of the initialisation process for a conventional integrated system fail to converge and cannot provide reliable information for navigation.

Scenario 5 simulates an integrity risk situation where the searching algorithm fails to find the correct DRM section. When a searching algorithm reports a wrong DRM section to the integrated system, the integration EKF processes the DRM aiding information with a incorrect error constraining model and a hazardous positioning solution may be produced by the system. Hence it is important to study the system behaviour under false DRM aiding information.

## 1.5. Contribution

A method of integrating DRM information with a tightly coupled integrated system is proposed and tested in this thesis. A DRM-aided MEMS/GPS positioning system, which carries out the system integration at the measurement level, is developed in order to prove the concept. The potential of an integrated positioning system using low-cost MEMS sensors at the measurement level with the aiding of a DRM to maintain high accuracy positioning performance on the railways under GNSS difficult situations is proved as feasible. Improvements on both system accuracy and reliability are achieved in comparison with conventional systems.

DRMs with different specifications are tested with the developed DRM-aiding system, and how they affect the performance of DRM-aiding are investigated. The specifications of DRM are important for both system performance improvement and the future construction and deployment of DRM-aiding. As a result, an optimal set of DRM specifications is selected among proposed candidates.

It is also found through the investigation that integrated positioning system with only MEMS sensors and GNSS is not enough to completely overcome various railway GNSS difficult situations no matter which integration architecture is used. The potential performance of low-cost MEMS/GNSS positioning systems with various integration architectures in different GNSS signal situations in a railway environment is studied and analysed. The conclusion serves as the basis of introducing

additional aiding information such as DRM for using low-cost MEMS sensors in a safety-of-life application.

## 1.6. Thesis Outline

This thesis consists of nine chapters and four appendices and is organised as following:

Chapter 1 starts with the background information of this thesis including the development of train detection systems, principles of integrated navigation, GNSS positioning on the railways and inertial navigation with MEMS sensors. The objectives of the research are listed to be answered by the end of thesis. The methodology of this research is introduced based on the simulation database, system performance testing and testing scenarios.

Chapter 2 reviews basic background information on signalling and control on the railways. Current train detection systems are described in detail. The development of future railway signalling and train positioning are introduced to demonstrate the need for a high-performance railway positioning system. A review of current research on railway positioning is given at the end of the chapter.

Chapter 3 introduces major aspects of integrated navigation. Different integration architectures are described. The basic principles for Kalman filtering are explained and the algorithm of a Kalman filter is given. Filter state selection and system observability are discussed at the end.

Chapter 4 reviews the status of current GNSSs. The basic positioning method using satellite pseudorange measurements is introduced as part of the background knowledge for this research. General GPS performance and various error sources during the satellite ranging process are also discussed as a part of the basis of building a simulation database.

Chapter 5 introduces basic working principles for inertial navigation and the current development status of MEMS inertial sensors. The computations of an inertial navigation processor are described in detail with discussion on various implementation issues.

Chapter 6 describes the process of constructing the simulation database used in this research. Preliminary tests are also performed to test the simulated sensor measurements.

Chapter 7 provides detailed descriptions of the two MEMS/GPS integrated systems tested in this research. General performances of both systems are studied along with a comparison with the performance of a GPS only system. Three test scenarios are performed to test system behaviours under GNSS-hostile environments on the railway.

Chapter 8 starts with a discussion of using DRM information to aid an integrated system. General principles of the proposed DRM-aided integration are described. The general performance of the system is studied by varying DRM point intervals and point accuracy levels. All five test scenarios are performed on the proposed DRM-aided system to study system behaviours.

Chapter 9 summarises the research process described in this thesis. Conclusions are drawn to address the objectives set at the beginning of the thesis. In addition, recommendations for further research on the subject are given.

Appendix A defines all reference frames involved in this research and gives the information on frame transformations and modelling the Earth surface.

Appendix B describes the basic kinematics used to represent the vehicle attitude and the calculation of relative direction cosine matrices.

Appendix C shows a brief derivation of the inertial error propagation equations in the local navigation frame. Sub-matrices for the integration Kalman filtering process are given out based on the derived error propagation equations.

Appendix D shows more results of the general performance study on the proposed DRM-aided system.

# 2. Signalling and Control on the Railways

The railways work in a characteristic way when compared to other transportation systems. On railways, a train is usually made up with several carriages. Theses heavy vehicles are constrained on fixed rails making movement only along the track. The traffic flow therefore only has one degree of freedom. While a train is moving, it is impossible for its driver to steer away to avoid obstruction and the braking capacity is small due to the minimal friction between steel wheels and rails. Meanwhile the average speed of trains has increased dramatically through years of development. All these lead to a potential risk of collision. As a result, the movement of a train is based on the authorisation of the signals. This makes the ability to identify positioning information (usually position, preferably velocity) for every vehicle in the system critical for a reliable signalling system.

Basics aspects for a railway signalling system and an description of various train detection systems are given in Section 2.1. ERTMS/ETCS as the future signalling and control system for European railways is introduced in Section 2.2. Section 2.3 gives a review of research that has been carried out for positioning on the railways.

## 2.1. Current Railway Signalling System

Current railways traffic control relies heavily on fixed block signalling systems and traditional train detection systems. Along with the rapid growth of traffic across the rail system, the proposal for a moving block system is on the agenda. A brief overview on various aspects of current railways signalling system is given through the following subsections with an emphasis on the current situation in the UK.

### 2.1.1. Principles of Railway Signalling System

According to Health and Safety Executive (2005), "the primary safety objective of any signalling system is to:

- prevent collision between trains;
- prevent derailment of trains at incorrectly set points or inadequately locked facing points;
- give an authority to proceed which does not conflict with the route set;
- protect level crossing."

Nevertheless, the objectives of a signalling system are not merely about safety. As the main function of the railways is to provide public transportation, objectives of a railways signalling system also include increasing the volume of traffic on the line. These two sides of the objective, however, cannot be achieved without each other. An increase of the traffic leads to an increase of the possibility for collisions and thus requires a high standard safety requirement for the signalling. On the other hand, a safe signalling achieved by sacrificing efficiency limits the ability of railways as a transport system.

The current railway signalling principle can be summarised as a three stages process: set, lock and protect a route for the train (Fenner, 2007). Under this principle, the path which the signalling centre wishes the train to take must firstly be clarified and indicated, known as "set". Secondly, by preventing activities such as unauthorised use of the set path and accidentally setting conflicting routes, a certain part of line is considered as "locked". The last part is to "protect" the train, which is usually done by holding the route until the train has passed or stopped and therefore minimising the risk of a driver failing to obey the authorisation.

### 2.1.2. Train Detection Systems

The basis of any safe railway signalling and traffic control operations is the knowledge of accurate and up-to-date information on the position and movements of trains. In modern railways this is achieved through a train detection system. The presence of a train on a track section is automatically detected and input into the signalling system.

Instead of indicating the coordinates and speed information of the vehicle like other modern positioning systems, the output from the train detection system is always presented as defined states (e. g. track section occupied or clear; vehicle present or no vehicle present at a specific location) and can therefore be naturally incorporated into block signalling based traffic control process.

There are two types of train detection systems that are currently deployed on UK railways according to Alston (2000). The two systems are deployed for different scenarios. The first type of systems detects the presence or absence of vehicles within the limits of a track by using track circuits or axle counters. While the second one detects that a train has reached, is passing or has passed a specific position with wheel detectors and treadles.

Requirements for the train detection system vary among different signalling systems. But in general the level of safety of the train detection system should be consistent with the overall safety of the signalling system. Similar to required navigation performance (RNP) parameters adopted by other transportation systems as a measurement of positioning system performance, the performance requirements for a train detection system is broken down into aspects like accuracy, integrity and continuity, as is discussed in Health and Safety Executive (2005) and Alston (2000). What should be noted, however, is that the concept of system availability is missing in these documents. Also detailed RNP figures for specific application scenarios are still under study and discussion. Therefore current requirements are more focused on general reliability and compatibility with current signalling systems.

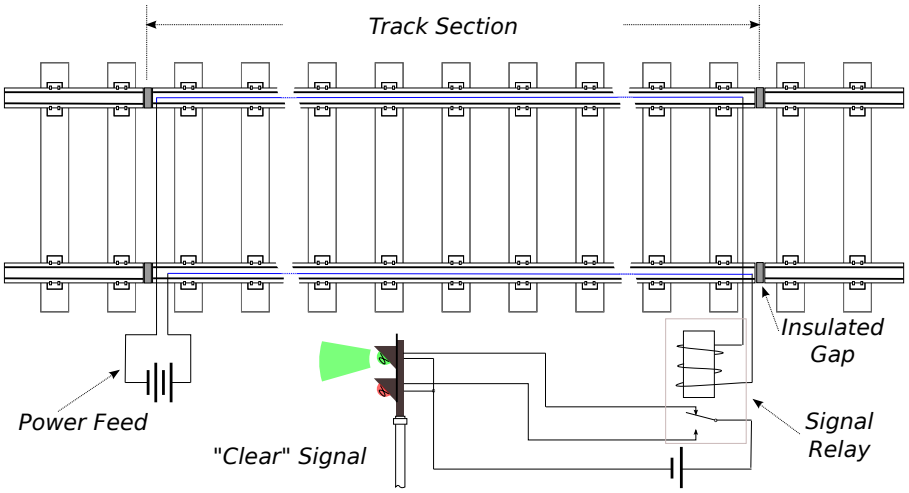
As a result of the nature of transportation vehicle, i. e. the train consists of a series of carriages, the railways ask more from a train detection system than merely delivering positioning information. Unless it can be shown that the risk of a train becoming divided is acceptably low, railways always requires the ability of the train detection system to identify and safeguard against the effects of train division as part of the integrity monitoring (Alston, 2000).

### 2.1.3. Sensors for Train Detection

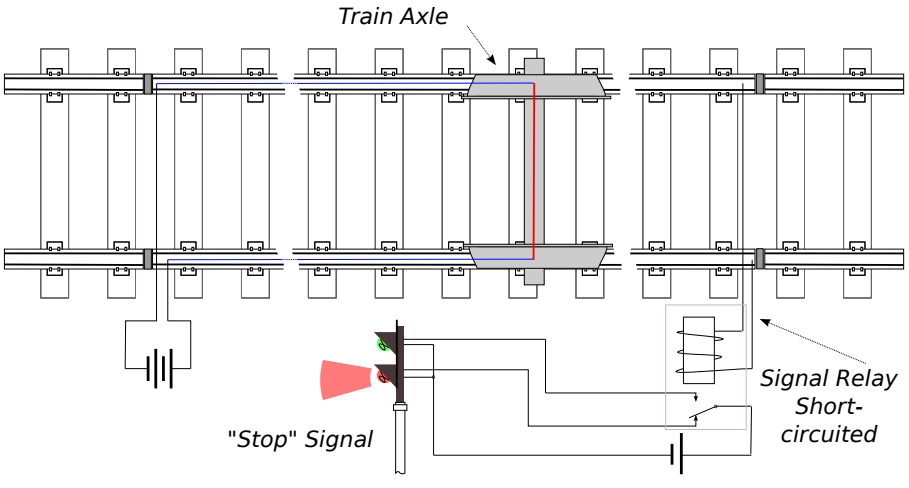
As mentioned in Section 2.1.2, there are two types of train detection system that currently deployed in UK. Each is designed for different application scenes and therefore uses different detection sensors. Three most common sensors are *track circuit*, *treadle* and *axle counter*.

#### 2.1.3.1. Track Circuits

*Track circuits* are the earliest yet still most widely adopted train detection sensors among those mentioned in Section 2.1.2. Their simplicity and utility is the main reason for it remaining a long lasting technology on the railways. In its simplest form, a track circuit is a track section with one end connected to a feed providing the voltage while the other end is connected with a relay. Each track section is insulated from adjacent sections. When the circuit is operational, an electric current circulates over both rails and the relay remains energised as long as the section is free. The signal box therefore reports a free track section, as shown in Figure 2.1a on this page. The axle of a train in the track section will conduct a path with lower resistance between rails. This causes a short circuit and the relay drops due to the lack of current indicating an occupied track section. The process is demonstrated in Figure 2.1b on the current page.




**(a)** Track section cleared




**(b)** Track section occupied

**Figure 2.1.:** Schematic drawing of track circuit

The length of a track section equipped with track circuit varies from 25 m to 1200 m (Goverde, 2005). However the minimum length of a section must exceed the longest distance between two axles in order to prevent situations that an occupied section is undetected and reported as clear, i. e. “loss of shunt”.

The major merit of the track circuit is its fail-safe design principle. Any failures caused by loss of power supply or broken wires would result in an apparently “occupied section” and prevent traffic

from using the section. This kind of failure mode is referred as a "right-side" failure.

Alas, wrong-side failure, situations where an occupied section is unreported, can happen too. This type of failure mode is more dangerous and could lead to disastrous accidents. Wrong-side failures for track circuits are mostly momentary, but occasionally there are persistent failures which are potentially more hazardous.

The majority of momentary wrong-side failures are caused by the lack of good electrical contact between the rails and the axle. This can be overcome by either making the speed of a signal relay responding to the electrical current cut-off slower or holding a route until the next section is reported occupied as is designed in modern Automatics Route Setting (ARS) system (Palmer, 2006).

Meanwhile the causes for persistent wrong-side failure are more complicated, but can be generally grouped into two categories. The chief cause for these persistent failures is contaminations on the rail-head. Contaminations can be from multiple sources. They can be either rust on the rail-head which is common on less busy rail lines or insulated objects lying between rail-heads and the axle such as common leaf mulch problems that happened in Britain during leaf-fall season (Palmer, 2006). The other potential situation for persistent track circuit failure is more serious and is caused by traction return current on electrified railways. As a railway design requirement, only AC track circuits tend to be used on DC electrified railways, and vice versa, in order to prevent the signal relay been accidentally energised by the massive current drawn by trains on the line. Unfortunately electro-magnetic compatibility is far more complicated than that. Although nowadays track circuits come with different operating frequencies, modern traction units packaged with various electrical devices like chopper circuits, inverters and variable frequency drives can easily generate interference right across the frequency spectrum. A good example of that would be the Eurostar's interference with current monitoring unit on London and Kent lines (Cross *et al.*, 2006).

#### 2.1.3.2. Treadles

*Treadles* are used on the railways to detect the movement of a train passing a particular point. As locations of trains indicated in the form of a section of track can be used directly for the traffic control, treadles are not as widely used as track circuits. But this additional information about the train position is still needed throughout railway systems. The most prominent usage is to provide exact train approaching information for critical positions on the line such as switch points (where the train changes track and different rails intercept with each other) and automatic level crossing points.

**Figure 2.2.:** An electro-mechanical treadle on the UK railway (Signalhead, 2009)

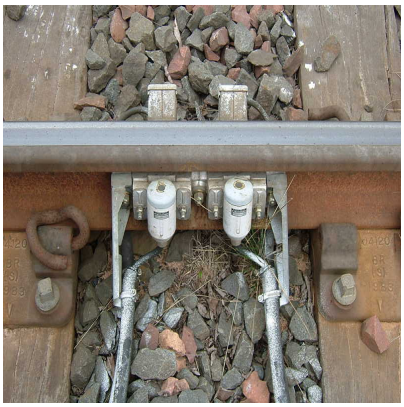


**Figure 2.3.:** An electronic treadle as part of an axle counter unit (Signalhead, 2007)

Depending on the design, there are different types of treadles deployed on the railways. In its simplest form, a treadle can be a mechanical bar connected to a locking bolt, which will then prevent signalman from changing a set of facing points while a train is over them. This is useful particularly when track circuits are unreliable due to rusty rails. A modern and more common type of treadle is an electro-mechanical device with a small arm that lies across the inner side of a rail. When the arm is pressed down, electrical unit connected with the arm changes output. Figure 2.2 shows an example of an electro-mechanical treadle. Some of these electro-mechanical treadles come with multiple arms. They are usually referred as "directional treadles". By sensing the sequence these arms are pressed, the movement direction of the train can be decided. Another more recent development of the treadle is a completely electronic device without mechanical interface known as "freddie". It exploits the disruption of an electro-magnetic field caused by a passing axle. More often these electronic sensors are fitted as part of an axle counter unit as shown in Figure 2.3.

#### 2.1.3.3. Axle Counters

*Axle counters* were introduced into the railways as a supplement for track circuits. They can provide similar information as track circuits about the occupation status of a specific track section. Compared with track circuits, no bonding and less cabling is required in an axle counter unit which implies less expensive installation and maintenance. In general, inherited problems such as rail-head contaminations do not affect the axle counter. So, it is possible for axle counter to work in situations where track circuits are unreliable, for example, wet tunnels. More importantly, the massive use of micro-processors and computer systems by the axle counters makes them easier to integrate into a modern centralised traffic control system (Lucic, 2005).

The whole axle counter unit is divided into two parts: detector heads and an evaluator. Detector heads are deployed on both ends of the track section. Each detector head usually includes a pair of freddie treadles (as in Figure 2.3). These detector heads are used to detect passing axle and decide whether the axle is moving in or moving out the section (using the same concept explained in Section 2.1.3.2). The count for the number of axles in the section is stored in the evaluator. Each axle entering the section detected increases the count while each leaving axle detected decreases the count. A zero count indicates a clear section. The overall process is shown in Figure 2.4. Because of the programmability and increasing processing power of micro-processors, multiple sections can share the same evaluator making a "multi-section axle counter" (Palmer, 2006).

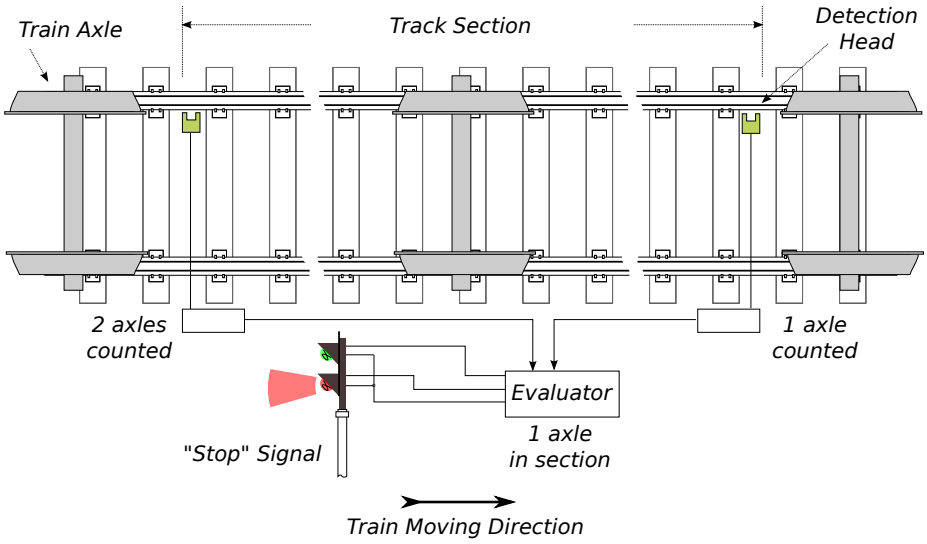



**Figure 2.4.:** Schematic drawing of axle counter

The main disadvantage of an axle counter over track circuits is its lack of an intrinsic safe failure mode. Most failure modes of axle counters are related to the fact that the evaluator is unable to detect the presence of a train in the section directly. All section status reporting is based on the work of detector heads. However, this indirect way of detection is not always reliable. Without further system fail-safe enhancement, a counter would have a problem maintaining correction count when the train wheel stop directly on the detector head or the train is switching rails at switch point. Also, the magnetic brakes equipped by most high speed trains might be present as another problem for ordinary detector heads. These brakes are physically large metal piece mounted on the bogie of the vehicle, only a few centimetres above the track. They can be accidentally recognised by the counter as an extra axle, alas this situation is not necessarily always happening on both sides of the section.

Another possible failure mode is caused by the reset process of the counter. As a result of no intrinsic safe fail mode, the equipment must continuously monitor the status of itself to become reliable. A very sophisticated hazard detection mechanism is implemented in the system as demonstrated in Lucic (2005). Any irregularity under this circumstance will lead the system reporting faulty and the count being needed to be manually reset by additional outer information. This requires a robust procedure to perform maintenance. Improper procedures lead to severe accidents like the 1991 Severn Tunnel accident, which was caused by improper axle counter reset timing, as stated by Palmer (2006).

### 2.1.4. Block Signalling Systems

In order to prevent head-tail collisions, the basic rule for traffic managements on the railways is to ensure safe separations between trains on a track. The following text gives a concise overview on aspects of a block signalling system based on single-track rail road examples. The emphasis is placed on current signalling and the evolution for future signalling. Actual signalling process for daily rail operation can be complicated by the increase of line number and intersections between tracks. More background on railway signalling can be found in Bailey (1995) and Goddard (2006).

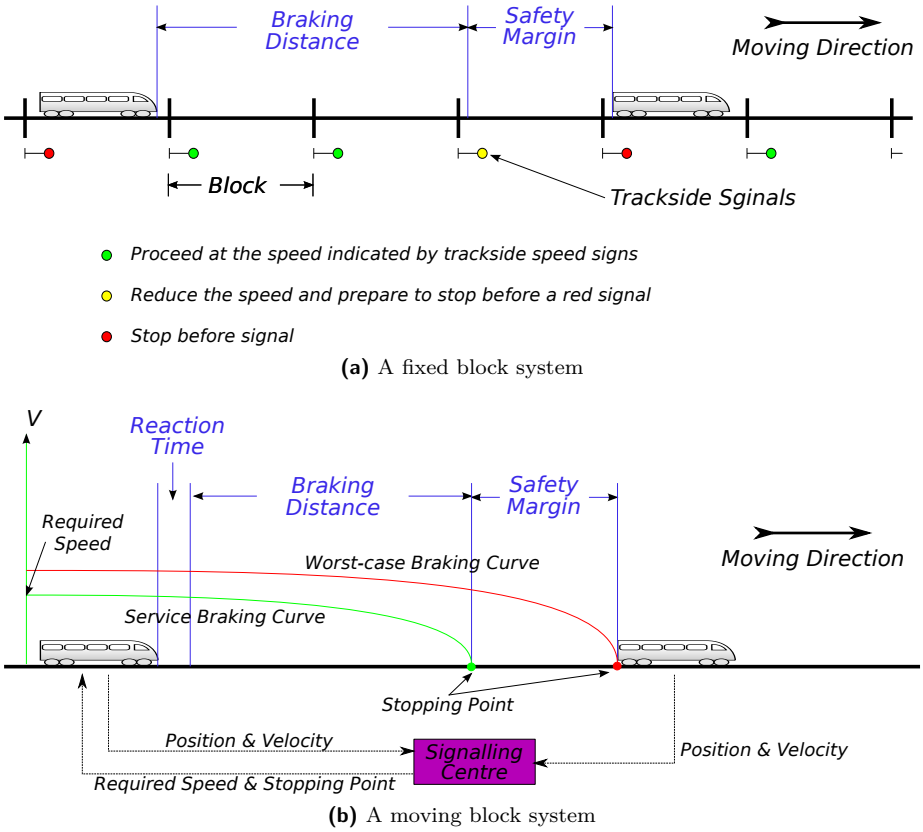
The separation between successive trains is calculated on the base of train speed and rated braking performance (Health and Safety Executive, 2005). A safe separation consists of a braking distance from the allowed maximum speed and a safety margin. The actual determination of a safe separation varies in different block signalling systems but shares the same basis, i.e. the reliable knowledge of locations of trains on the track.

Two types of block signalling system are available for the railway signalling. Based on the way they manage the separation between trains, there are *fixed block system* and *moving block system*.

In a *fixed block system*, the rail track is divided into sections or “blocks”. The length of a block was originally the distance between two signalmen, and is nowadays related to the distribution of train detection systems (normally the length of a track circuit unit). Signals installed at the start and end of each block form the Automatic Block Signalling (ABS) system. ABS issues automatic signals base on outputs form train detection systems. As a safety principle, only one train is allowed at one time to travel in a block. Each block is protected by a signal that issues a limited Moving Authority for the driver to enter a certain block.

The safe train separation in a fixed block system is maintained by protecting, or “locking”, of one or more unoccupied intervening block sections between occupied blocks. This mechanism is shown in Figure 2.5a on the following page. Additionally, the entry speed into a block of a train can be controlled by the delayed clearance of a signal.

In general, a fixed block system can provide a good level of safety as sufficient physical separation




**(a)** A fixed block system

**(b)** A moving block system

**Figure 2.5.:** Schematic drawing of block signalling systems

is guaranteed. But the problem occurs when there is a fast increase of traffic on the line, which is a matter of fact for all transports nowadays. As a result, it is becoming more and more difficult to keep this sufficient physical separation between trains without sacrificing the system efficiency. Because there are only finite track sections and almost two-thirds are being used as train separators.

A *moving block system*, on the other hand, compares the position of trains, their direction of travel and speed to determine the safe separation between trains (shown in Figure 2.5b). The major benefit of making the safe separation a variable is to maximise the line capacity of transportation on the line. Also, a moving block system reduces the reliance of train detection on track-side infrastructures. To implement a moving block system, the signalling system requires continuous and high resolution knowledge of the train position and velocity which can be beyond the capability of traditional train detection system for modern high speed railways. Pilot projects for a moving block system have reported successful for light railway (Lockyear, 1996). ETCS level 3, as mentioned in Section 2.2.3, is aiming at the realisation for moving block signalling on railways across Europe.

## 2.2. ERTMS/ETCS

The European Rail Traffic Management System (ERTMS) is an initiative led by European Union as an update of current railways systems among its members. The major driving force behind ERTMS is the demand for increasing cross border traffic as a result of the advent of new high-speed lines and development of freight networks across Europe. In addition, the increase of the rail traffic itself has brought more and more pressure on current railway infrastructures. This led railways to rethink their strategies on infrastructure modernisation.

RSSB (2010) states “three main outcomes for ERTMS to achieve as:

- Facilitate, improve and develop international rail transport services within European Union
- Contribute to the progressive creation of the internal market in equipment and services for the construction, renewal, upgrading and operation of the trans-European rail system
- Contribute to the interoperability of the trans-European rail system."

ERTMS consists of four elements. Among them, European Train Control System (ETCS) provides the train controlling and signalling for ERTMS with different application levels and operating modes including Automatic Train Protection (ATP). Global System for Mobile communication – Railways (GSM-R) is the telecommunication element of ERTMS which not only provides voice communication but also carries ETCS data between trains and track-side infrastructures. European Traffic Management Layer (ETML) is a concept to optimise railway operations through improved management. Lastly, European Operational Rules (EOR) is a set of agreed rules to standardise certain aspects of rail operation across Europe.

ERTMS is currently under active construction among both European countries and non-European countries (such as China, South Korea, etc.). Pilot lines fitted with ETCS level 2 (as described in Section 2.2.3.2) have already been reported across Europe either in trail or partially into commercial services (Bloomfield, 2006). According to Department for Transport (2007), the first UK rail line with ERTMS will be in operation by 2014. Meanwhile, the first version of Operational Concept for ERTMS in UK has been published for information (RSSB, 2009). However, what should be noted among these rapid progresses though is the lack of operational concept for ERTMS/ETCS level 3 (as described in Section 2.2.3.3).

### 2.2.1. ETCS Overview

ETCS forms the crucial part for signalling under ERTMS. Along with other supporting equipment such as other conventional signalling equipments and GSM-R data network, they consist of the next generation signalling system for railways across Europe. As conventional signalling, ETCS provides the supervision to ensure the train does not travel further than it has permission to do so. In normal operation, information on maximum authorised travel distance, speed profile of the track ahead and the route that has been set is delivered to the driver who is responsible for the movement of the train. Should there be any irregularities on the train operations such as exceeding the distance or speed for which the permission to move has been granted, ECTS will be able to provide train protection functions as current ATP systems. In this case, a warning is firstly issued to the train driver and interventions will be then performed by ETCS to either slow the train down or bring it to a stand.

More detailed ETCS behaviours depend on specific application scenarios. The overall ETCS application is grouped into five levels in order to provide a step-by-step introduction of the new technology while maintaining the high safety standard of railway signalling and ensuring the interoperability with current systems across the Europe.

The overall ETCS is divided into two parts: the track-side equipment and the onboard equipment. ETCS track-side equipment mainly covers the track-side functions for train detection and radio communication. While communication equipment varies through application levels, train detection system (as described in Section 2.2.2) is always included as described by RSSB (2010). The responsibility of ETCS onboard equipment includes train position and speed determination, communication with central traffic management, interfaces for both driver and other train systems and data recording. Figure 2.6 shows an overview of ETCS onboard equipment. All onboard equipments are controlled under a centralised computing unit called European Vital Computer (EVC).

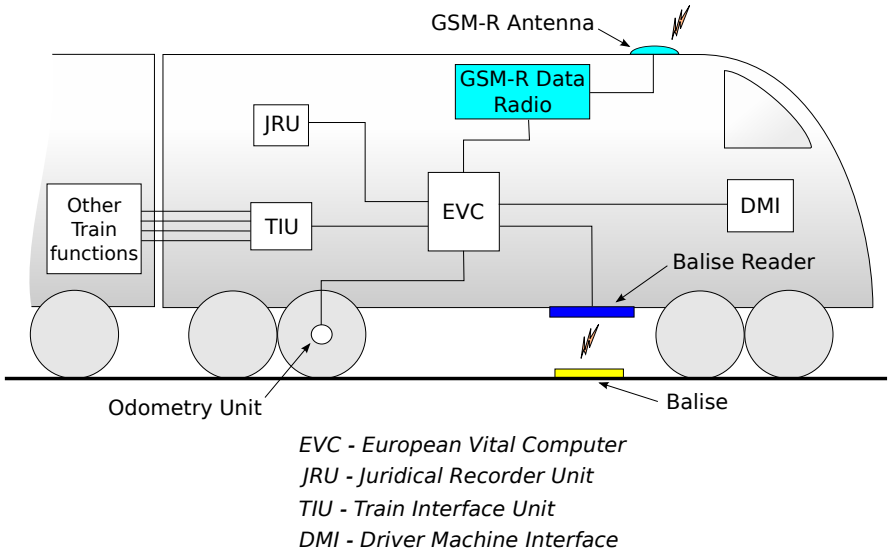



**Figure 2.6.:** An overview of ETCS onboard equipment

### 2.2.2. Train Detection Equipments for ETCS Level 1/2

As mentioned in Section 2.1.2, knowledge of train location is critical for a railway signalling and control system. In orders to cope with problems on current train detection system as described in Section 2.1.3, ETCS adopted new sensors for train detection. The main equipment used by ETCS level 1 and level 2 to determine train location and speed is track-side *Eurobalise* and onboard *Odometry*.

*Eurobalise* as a train detection system is mainly composed by two parts: the actual balise installed on the track and a balise reader installed onboard.

A balise is a radio-frequency identification (RFID) electronic transponder mounted on or between sleepers in the centre line of the track as shown in Figure 2.7. Unlike treadles, balises need no power supply and will only be energised by passing train equipped with balise reader. Once the balise is energised, it transmits an electronic telegram back to the onboard balise reader. The content of telegrams sent by balises can include information on the location of the balise, geometry of the line (such as curves, gradients and speed limits) and the position of any signals. The actual content will vary depending on the purpose of the balise, and can be easily programmed as shown in Figure 2.8. In addition, the balise does not necessarily always transmit the same data. In a real application, it is possible to connect the balise with a local switching unit to enable different transmissions according to the input received (Siemens AG, 2008a).

Balises are typically deployed in groups consisting of two to eight balises. A balise group not only enables the train to determine its travelling direction from the order in which it passes over the group but also allows a larger amount of data to be transmitted to the train.

All trains equipped with ETCS have a balise reader onboard. The balise reader energises the balise when passing through it, which enables the balise to transmit the telegram to the train. The telegram is then received by the reader and passed to EVC.

The *Odometry* system on board a ETCS-fitted train is used to estimate speed and distance travelled information. As the railways require high integrity so the odometry system is usually a combination for more than one mechanism. Typically, EVC does calculation based on combined information from a tachometer and a speed radar. An industrial example is shown in Siemens AG (2008b). The ETCS standards require the odometry system to maintain an accuracy no worse than $\pm$(5 m +5% of the distance travelled) since the passing of a balise group (RSSB, 2010).

**Figure 2.7.:** A Siemens Eurobalise in Germany (Gemaakt, 2004)

**Figure 2.8.:** Programming for a Eurobalise (Siemens AG, 2008a)

### 2.2.3. ETCS Application Levels

As mentioned in Section 2.2.1, ETCS can be deployed in different configurations and therefore provide different levels of operation. There are currently five application levels available for ETCS including level 0 to 3 and a Specific Transmission Module (STM) level. Among all the levels, *level 0* and *level STM* are aiming to deliverer a smooth transaction of signalling and control from current systems across Europe. *Level 1* and *2*, on the other hand, is focused on the progressive introduction of new signalling system and train detection equipment. *Level 3* is currently in its conceptual phase and allows for the introduction of a "moving block" signalling system (Davies, 2009).

#### 2.2.3.1. Level 0 and Level STM

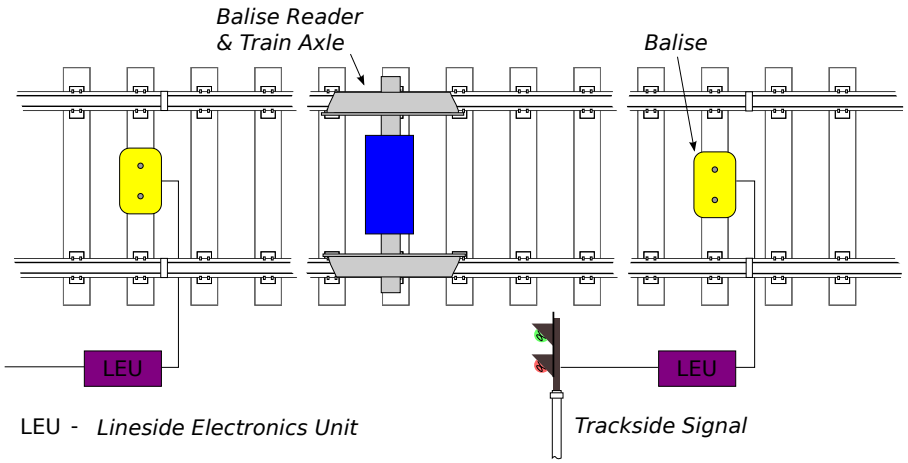
*Level 0* of ETCS operation covers the situation of ETCS-fitted trains operating in lines that are either not equipped with ETCS track-side equipment or in commissioning (i. e. ETCS infrastructure exists but is ignored). ETCS only provides supervision on the current train speed and against the maximum speed permitted in the area. While the driver retains the full responsibility to ensure the train stays within its moving authorisation and does not exceed the maximum allowed speed, any national train protection systems existed are to operate normally.

*Level STM* of ETCS operation allows onboard ETCS equipment to provide the normal functionality of a national train protection system such as Automatic Warning System (AWS) and ATP system. When operating at this level, ETCS DMI provides the driver interface for the national system and TIU provides the interfaces required by the national system (see Figure 2.6 on the previous page for an overview of ETCS onboard equipment). It is possible for ETCS onboard equipments to read messages from any existing national track-side devices under STM operating level. The main benefit of having a STM in ETCS is to avoid the need for a separate national train protection installation, which simplifies the train system structure and allows the switching between ETCS and national train protection system to be realised under software control.

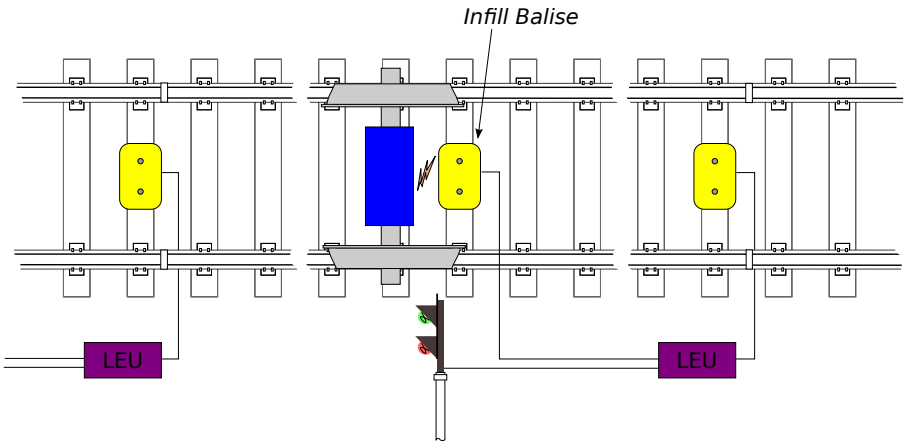
#### 2.2.3.2. Level 1 and Level 2

*Level 1* of ETCS operation is designed as an overlay of current national signalling systems and Eurobalise system is introduced to work with current track-side infrastructure as the first step of a progressive update process. A balise group (as described in Section 2.2.2) is deployed on approach to each signal and is connected to the signal via a Line-side Electronics Unit (LEU) so that the information provided to the train is linked to the signal, as shown in Figure 2.9a. The data

transmitted include Moving Authority (MA) which is related to the signal and information for train speed supervision.




**(a)** ETCS level 1 (basic)




**(b)** ETCS level 1 with infill balise group

**Figure 2.9.:** ETCS level 1 overview

The major limitation of this spot-based signalling mechanism is that the information update rate depends on the density of balise groups along the track, while the signalling information held by the train is only as good as the last time the train passed a balise group. As a result of this, situations when a yellow signal is followed by a green signal in the next section might lead the train to an unwanted stand because the EVC is expecting incoming red signal after a yellow one (see Figure 2.5a for signalling rules) and has to stop the train if the updated status of next signal does not arrive in time due to insufficient balise density. This can be partly overcome by installing infill balise groups in the same section (see Figure 2.9b). Alas, the installation and maintenance cost will increase. The other way to tackle this problem is to introduce Euroloop or Radio in-fill Unit. Both of them are equipment that can transmit information to the train over a length of the track and are only available for ETCS level 1 operation (RSSB, 2010).

*Level 2* of ETCS operation features the inclusion of radio data link using GSM-R and the ability to work without track-side signals. In theory, level 2 is capable of signalling with or without track-side signals. But level 2 with track-side signals is considered a migratory step towards the final removal of track-side signals. An illustration of ETCS level 2 operation without track-side signals is shown in Figure 2.10 on the following page. Under level 2 operation, signalling information is transmitted to trains via a radio data link. Because of this, a centralised traffic management is possible to be

established directly between trains and control centre and allows "all associated external hardware to be dispensed with" (Fenner, 2007). Unlike ETCS level 1, the information provided by balise groups no longer contains information on MA and tracks. Instead, balise groups become position references along the track. There is a list of balise groups on the line and distances between groups stored in ETCS onboard equipment. By reading real time information from balise groups and comparing them against the stored database, the train is able to validate the expected travelling route, supervise the performance of the onboard odometry system and reset the accumulated error of the odometry system.

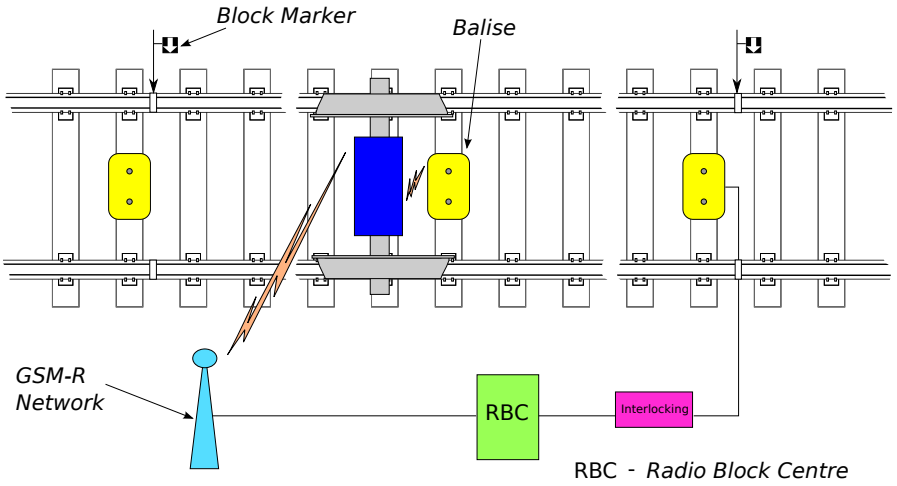



**Figure 2.10.:** ETCS level 2 overview

In comparison with traditional fixed block signalling system, ETCS level 2 adopted a new functionality called "co-operative shortening of MA" in order to increase the line capacity. This is essentially a more flexible means to shorten a route ahead of a train which leads to a considerable reduction on time taken for train route revocation. The request for shortening the MA can be sent through ETCS track-side equipment to ETCS onboard equipment. ETCS onboard equipment calculates whether the new MA is going to automatically cause the brake to be applied because of the trains approaching towards the end of a route. Depending on the outcome of this calculation, either the new MA will be applied and informed to track-side equipment, as a result the extra route can be released, or the original MA will be maintained. Extra route means more capacity, "more capacity means more trains moving, thus more benefits" (Davies, 2009).

#### 2.2.3.3. Level 3

*Level 3* of ETCS operation is similar to level 2 in most aspects only that the requirement for track-side train detection equipment is no longer crucial. The ETCS onboard equipment estimates the train position and speed and is also responsible for checking the train integrity and reporting completeness to traffic control centre. Nevertheless, the deployment of train detection equipment over switches and crossings is permitted to increase overall signalling performance. The ultimate goal for implementing ETCS level 3 is to realise a moving block signalling system (see Section 2.1.4 on page 34 for detailed description).

In terms of methods to acquire knowledge of train location (i. e. the position and speed of a train), ETCS level 3 introduces modern positioning technology into railways signalling and proposes a signalling system with access to more accurate, reliable and consistent knowledge of train location. This requires the combination of positioning and detection as two fundamentally different methods to deliver knowledge of train location.

A train detection system implies a passive way to estimate train location which heavily relies on railway infrastructures. The train location is delivered only when a sensor is triggered by the train passing by. The attained knowledge is always related to a certain part of the track. Meanwhile, the output quality depends on both the sensor been used (such as track circuits and balise groups) and the infrastructure (for example, the density of balise groups on the line). A signalling system solely relying on train detection systems increases expenses for infrastructure construction and maintenance and yet could not promise a consistent knowledge of train location.

On the other hand, positioning provides an active way to estimate train location that is independent of railway infrastructures. By using available positioning and navigation technologies such as dead reckoning and satellite positioning, train location information can be delivered in real time by the individual train. The quality of location knowledge is independent from infrastructures and can be consistent throughout the overall traffic system provided the positioning system functions normally.

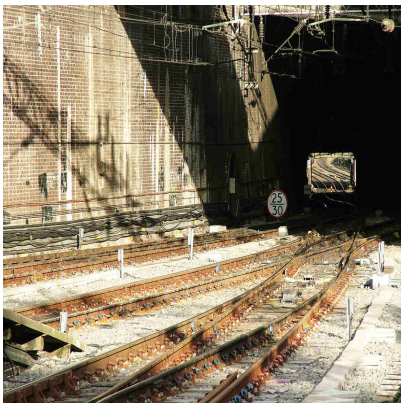
## 2.3. Positioning on the Railways

### 2.3.1. The Working Environment on the Railways

#### 2.3.1.1. A Guided Transport System

Compared with other transportation methods, a train cab moves on a rail track with less freedom and its motion is hence simpler to some extent. Firstly, as the train cab is designed to operate on a track, the movement of a train is always guided along the track. Alas, this means that the train does not have opportunities to take manoeuvring actions like vehicles do in other transports to avoid collisions. Thus "collision can only be avoided by arriving at a different time" (Fenner, 2007) on the railways. Secondly, straight lines and long flat curves are preferred in rail track design conventions. It is almost impossible to see significant changes in the trajectory of a train movement. A train always accelerates and decelerates along a smooth curve. Meanwhile a single rail track is a much smaller object in contrast with most high ways.

**(a)** Approaching New Street Station, Birmingham (Thryduulf, 2006)



**(b)** Looking east from New Street Station, Birmingham(Ashton, 2006)

**Figure 2.11.:** Typical railway environment

In general, rail tracks are most commonly scattered in the open area before entering the station. However this simplicity soon disappears and is replaced by a complexity of parallel tracks and switching points when a train approaches the station and dozens of tracks start to converge together

in a high density. Figure 2.11b on the preceding page shows a typical image of track complexity when approaching a station.

As a result of these characteristics, the railways developed their own train detection methods based on the block signalling architecture as described in Section 2.1.4. One advantage of these methods is that their location solutions do not scatter in 3-D space as the results of modern positioning systems do. This is particularly useful for situations when there are many closely placed tracks which are either intersected or parallel to each other. Train detection methods can guarantee its location solution is always related to the correct track, which is crucial to a guided transport system of which basic signalling assumption is that all trains are on the right track. Modern satellite positioning systems are capable of a result with errors small enough that enables it to be fixed on the correct track, whereas the overall availability and continuity are still to be studied. An integrated navigation approach opens the possibility to take advantage of the railway characteristics in a similar way to train detection methods. While train detection uses the physical track infrastructure to estimate train location, it is also possible to feed the position information of the track, e. g. a track database or Digital Route Model (DRM), into an integrated navigation system. Research has already shown accuracy and integrity improvements using track database with the GPS on railways (Zheng, 2008). However, more researches for adopting a DRM into an integrated system is needed.

#### 2.3.1.2. The Line-of-sight Problem for Radio Signals

The main difficulty for positioning in the railway environment is caused by physical obstructions. Situations such as travelling through a tunnel or going into areas where the sky will be partly blocked or completely invisible, as shown in Figure 2.11b are common throughout a typical train journey. Modern positioning technologies which are based on radio signals such as GNSS experience line-of-sight (LOS) problems under these situations.

The LOS problem can cause either performance degradation or complete loss of service for GNSS working on the railway. The occasional degradation of positioning performance is mainly a result of obstructions along the track, for example dense foliage and high platform gantries or roof structures. Because of the blocking of satellite signals, there is a decreased satellite visibility accompanied with a poorer geometry of the visible constellation. In addition, multipath effects can be stronger in a railway canyon, which is another cause of performance degradation. On the other hand, shelters above the track leads to a complete loss of all satellite signals. Structures like tunnels and completely roofed stations are common across the railway, and they usually cause a few minutes loss of GNSS services. The loss of services, however, will continue for longer as a result of the time taken to re-acquire the satellite signals. The actual time that is needed to regain the normal services mostly depends on receivers.

Thus additional information from different sources, especially from technologies that are based on different principles, becomes crucial to provide high performance positioning on the railways. An INS/GPS integrated system is one good example for this approach.

### 2.3.2. On the Use of GNSS on the Railways

The performance of GNSS positioning on the railways varies from different routes and can be improved with the presence of an augmentation service.

According to RSSB (2008), GPS positioning without augmentations in general is able to achieve 13 m accuracy at 95% confidence, while no integrity is available. With the implementation of Receiver Autonomous Integrity Monitoring (RAIM), a better than 10 m accuracy performance is

available at 95% confidence, whereas the integrity is subject to sufficient satellites in view (more than 5 satellites required).

In comparison with the GPS performance without augmentations, 1 to 2 m accuracy at 95% confidence can be expected with currently available augmentation services including Terrestrial differential GPS (DGPS), European Geostationary Navigation Overlay Service (EGNOS) and other commercial Satellite Based Augmentation Systems (SBAS). However this number is subject to many other factors. At present, it is difficult to receive information directly from geostationary satellites. The data can alternatively be transferred through terrestrial communications. However, the lack of standard communication arrangements to the onboard train devices limits the usage of all augmentation services. Lastly, terrestrial communication especially DGPS signals could be subject to interference on the railway due the complexity of electro-magnetic environment on the railways.

As mentioned in Section 2.1.3, electro-magnetic compatibility is critical for a safe railway operation. Therefore any GNSS equipment must be compliance tested and certificated to meet railway group standards. This is to ensure no interference will occur on normal rail operation including both train and track-side infrastructure. On the other hand, electro-magnetic interference generated by onboard train equipment is capable of affecting the acquisition and tracking ability of hardware GPS receivers. Bertran & Delgado-Penin (2004) concluded that some old train carriages may exhibit excessive radiation level and that for this reason modern software enhanced GPS receivers can be preferable on the railways.

A national-wide data collection project carried out and initiated by Rail Safety and Standards Board (RSSB), LOCASYS, has studied "GNSS forecasting within the rail industry of the UK" (Thomas *et al.*, 2007). One of the lines used in the experiment is a suburban rail line across Birmingham city. The line includes a wide variety of urban, suburban and rural environments. The result on GNSS availability during the journey is summarised in Figure 2.12[1]. As can be seen from Figure 2.12, the overall GNSS availability throughout the journey is about 89.36% of the travel time. GNSS is capable of providing positioning services for the majority of the journey. Nonetheless positioning difficulties and longer service outages tend to happen at areas around the station where traffic is most busy. Alas, these areas can be critical sections for the signalling. As the traffic become busy, the need for a reliable signalling increases and hence the requirement for a high performance train positioning system. In other words, the highest positioning performance is required at the most unfriendly and busy part of the railway environment.

Use of GNSS on the railways is not necessarily limited to safety-critical train positioning application only. Other ranges of application include time synchronisation, passenger information systems, automatic passenger assistant, electronic ticketing and track monitoring, etc. Many of these non-positioning applications are already in service.

## 2.3.3. Positioning and Navigation Research on the Railways

Much research has been carried out to implement a modern positioning system on the railways. This section presents a brief review of this research based on the different system architectures and sensor information that has been applied.

### 2.3.3.1. Researches for GNSS Positioning

LOCOPROL is a research project that is aiming at delivering a low-cost positioning system for European "low density" lines (Mertens *et al.*, 2003). The developed system is purely based on GPS.

[1] Courtesy of Nottingham Scientific Ltd.

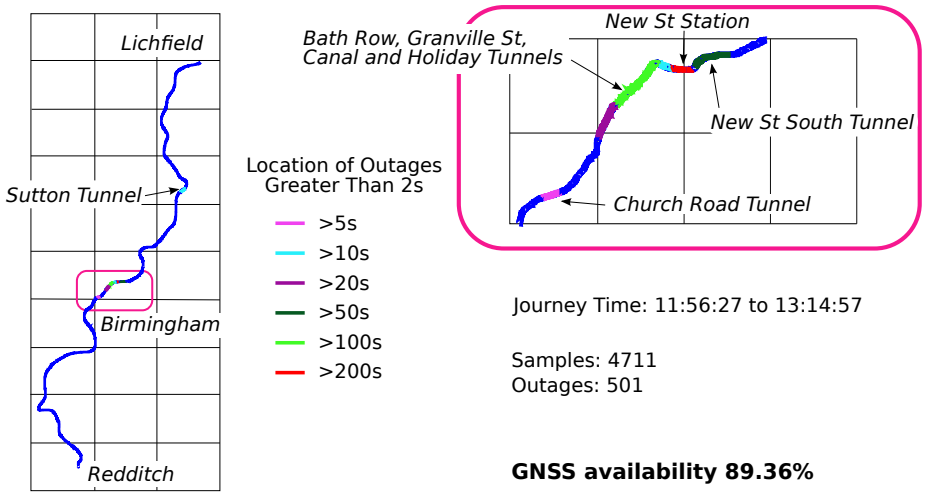



**Figure 2.12.:** An example of GNSS availability on the railways

A "hyperbolic 1-D positioning algorithm" is developed as the main approach to tackle the LOS problem. The track is utilised as an additional information source and is broken into consecutive sections. Pseudoranges from 2 satellites are used to determine a hyperboloid of revolution as loci of all possible locations of the antenna. The intersection of the obtained hyperboloid and the track line is the position of the train. Outputs of the system are in the form of the distance travelled along the track. As a result, the overall GPS availability can be improved and the redundancy of satellites visibility is increased. The demonstration trial (Simsky *et al.*, 2004) showed that the average "precision for position is 300 m along the track" and the average "precision for velocity is 4 m/s along the track" both associated with "a wrong side failure rate of $10^{-11}$/Hr".

Zheng (2007) developed an algorithm that is capable of integrating a track database (i. e. position knowledge of points along the track) directly into the least squares process of GPS code positioning. Sections between track database points are treated as curves and are projected into 2-D planes. The coordinates of both the train position and two track points can be related in the 2-D plane. Because only two 2-D projections are need to fix a curve in 3-D space, two known track points can be used to reduce the minimum number of required satellites number for positioning solution. Meanwhile a searching strategy for the occupied track section was also suggested. Results from Zheng (2008) show both accuracy and integrity improvements with the aid of a track database. Availability is also increased as the required number of visible satellites is reduced.

#### 2.3.3.2. Research for Integrated Navigation

The RUNE project (Genghi *et al.*, 2003 and Albanese & Marradi, 2005) presented a prototype system with advanced integrity check algorithm and fallback strategies for a cascaded system. The system is designed to be incorporated with ETCS architecture. "The primary objective is to demonstrate the improvement of the train self-capability in determining its own position and velocity, with a limited or no support from the track side." Sensors available in the system include a GPS/EGNOS receiver, IMU and train odometer unit. A balise reader is also used with a "virtual balise map" to provide information from balises on the route. A GPS solution is provided to the main data fusion unit in the form of position and velocity and hence the integration is loosely-coupled in nature (see Section 3.1.1 for information on system integration). The evaluation is focused on a study of individual sensor performances and possible integrity risks. The project concluded that "availability

and integrity requirements are the most challenging obstacles towards an ERTMS/ETCS compliant GNSS aided railway navigation equipment."

A similar study carried out by Mirabadi *et al.* (1998), (1999) and (2003) discusses a fault detection and isolation scheme for a cascaded system and experiments different combinations of sensors. While most sensors available for this study resemble those in the RUNE project, Doppler radar is also included. Again, among all experimented sensor combinations, only position and velocity solutions from GPS are used to form loosely-coupled systems.

Other prototype systems based on the loose-coupled principle of GPS with other information source include "RadioCompass" and INTEGRAIL from Kayser-Threde (Bedrich & Müncheberg, 2004) and GLLS from Seagull Technology (Mueller *et al.*, 2003). Both "RadioCompass" and INTEGRAIL are deployed with tactical grade inertial sensors to improve performance. While "RadioCompass" comes with an adaptive Kalman filter, INTEGRAIL integrated similar sensors as RUNE plus map-matching ability. The GLLS features "a GPS heading system using 3 low-cost GPS receivers" and developed in US.

Instead of focusing on the train positioning problem, there are some studies focused on the improvement of the odometry unit performance. Ernest *et al.* (2004) and Mázl & Přeučil (2003) uses onboard accelerometer to detect and correct odometer slippage. The system positioning task is mainly carried out by GPS without data fusion. The enhanced odometry takes over according to a switch strategy as soon as the GPS is unavailable. Although odometer calibration aided by the accelerometer shows improvement on slippage detection and positioning accuracy, the output from the odometer is unable to continuously contribute to the system because of the absence of a data fusion module. Rome (2003), on the other hand, exploits the connection between GPS position solution and odometry unit outputs through the modelling on the measurement model of a Kalman filter. Odometer slippage is detected and corrected by continuous updates from GPS solution.

In addition, research has also been done to introduce new sensors for railway positioning. Böehringer (2003) showed a fusion framework with a new type of device referred as "Eddy current sensor". The new device "detects inhomogeneities in the magnetic resistance along the track caused by rail clamps, switch blades or by irregularities of the rail".

#### 2.3.3.3. Summary of Current Research

Current positioning research on the railway is mainly focused on the implementation of a cascaded system based on the loosely-coupled principle. The systems usually rely on GPS for high quality positioning solutions. IMU and other sensors are primarily useful for "GPS-dark areas". Advantages and disadvantages for the cascaded way of integration are discussed in Section 3.1.1. The cascaded architecture makes the system easy to adopt extra sensors. But it limits the system performance under situations where satellite signals availability is limited, which is actually the most common case for a GNSS positioning difficulty on the railway. The GNSS availability issue has been discussed in Section 2.3.2. Meanwhile, research on an integration system based tightly-coupled principle may help to improve system performance in the railway environment, but the studies are yet to be done.

Although applications using high performance INS come with promising results, tactical grade inertial sensors are still rarely used for train positioning research. Due to expense issues and the rapid development of micro-machine technologies, the future IMU onboard is more likely to consist of low-cost MEMS sensors. The problem with the application of MEMS is its lack of promising independent performance. Without external aiding, MEMS IMU is only applicable up to a limited short period as mentioned in Section 5.2 and hence cannot be expected to bridge "GPS-dark areas" alone.

The most often used dead reckoning sensor is the odometer. While odometers are in general reliable, they may fail if bad adhesion between train wheel and rail occurs due to rain, ice or leaves on the track. Also when the train is braking or accelerating, adhesion is impaired. Thus a real time calibration is required for an odometer to provide satisfactory outputs.

Research utilising track point position information, i. e. track database or DRM, has been constraint to GNSS-alone systems due to modelling complexities for integrated systems. Balises on the track provide similar functions as a DRM. However, according the prospective plan described by ETCS level 3, the future railway train positioning will reduce the need for track-side infrastructures to a minimum amount. This means the density of balises is going to decrease and therefore limit their contribution as a DRM.

# 3. Integrated Navigation

A multi-sensor integrated navigation system is a system that depends on a combined use of different sensors and data sources to achieve its positioning solutions. The idea lying behind an integrated system is that with the help of different sensors and data sources, not only the accuracy but also the reliability of a system can be greatly enhanced. Sensors and data sources that can be used in an integrated system vary according to specific applications. Some common options are GNSS, INS, terrestrial radio navigation system, feature matching techniques and other dead reckoning sensors. As a general principle for sensor selection, sensors with complementary characteristics are preferable. Consequently, position fixing systems and dead reckoning systems are usually chosen to integrate. Typically for such an integrated system, the dead reckoning system is responsible for providing the reference solution as it operates continuously at a higher rate than most position fixing systems, whereas other sub-systems operate as aiding systems to provide additional information to be processed by an estimation algorithm to produce corrections to the final integrated solution. The Kalman filter is most commonly used as the estimation algorithm for the system.

The design of an integrated system is a trade-off and must take many factors into account, such as positioning accuracy, output update rate, system reliability, budget, size and mass, and whether vehicle attitude information is required in addition to position and velocity information. Systems for various applications can therefore have very distinctive designs in terms of sensor selection and system integration architecture. For most personal positioning, road vehicle and asset-tracking applications, where the key factors may be cost, size, weight and power consumption, a GNSS only system generally is often sufficient. Nevertheless, safety-of-life applications such as civil aviation and the railways, focus more on the system reliability while still demanding a very high accuracy requirement in certain situations. Systems designed for these applications usually take advantages of an integrated system.

The major aspects of the integrated navigation are briefly introduced in this chapter. Section 3.1 gives a review of various system integration architectures. A summarised description of the Kalman filter is introduced in Section 3.2. A short discussion on the topic of system state selection and system observability is presented in Section 3.3.

## 3.1. Integration Architectures

The integration architecture of a positioning system is defined by how it combines information from different sub-systems. The choice for an integration architecture is a balancing act between maximising the accuracy and robustness of the navigation solution, minimising the complexity, and optimising the processing efficiency (Groves, 2008). The characteristics of each sensor must be taken into account in the design process of the system. Inertial or dead reckoning solutions are usually provided in a higher frequency than other sensors, but their positioning errors increase along with time. These sensors may be calibrated by including other position fixing methods in the system. GNSS and terrestrial radio navigation systems require a minimum number of signals to form a position solution, so their results may not be always available. Positions are usually delivered at a lower rate than inertial sensors. Another factor that needs to be considered during the design is

that often some sub-systems work as a "black box" due to manufacturing issues. Such "black box" systems not only offer limited information on their error characteristics but also may not accept integrated feedback corrections. In addition, if an integrated system is capable of calibrating its sub-systems during its operation, there is always a risk that the failure of one sub-system could contaminate the feedback corrections. Fault detection and integrity monitoring can be implemented in the system to prevent this situation. However, the introduction of these mechanisms can introduce an extra calculation load. In summary, through a properly designed integrated system, not only is the potential of individual sensors improved but integrity checks can be introduced, and resulting in an overall improved high-performance system.

Based on the information included in the system estimates, usually residing in the state vector of a Kalman filter, the implementations of an integrated systems can be classified into two categories: total-state implementation and error-state implementation. The estimates for a total-state implementation are the positioning solutions plus some error parameters. In contrary, an error-state implementation estimates the error contained in the positioning solutions. Hence a reference system is usually needed for an error-state implementation.

Depending on whether the integration algorithm is divided into local and central levels (i.e. whether individual sub-systems delivers raw measurements instead of final positioning solution) and whether measurements from the reference system (usually INS) are used in the local level, the integration architecture can generally be divided into three categories: cascaded, centralised and federated integration.

### 3.1.1. Cascaded Integration

In a cascaded integration, measurements from each sensor are first processed locally producing preliminary solutions, and Kalman filters may be used for local processors. Then an integration Kalman filter processes all preliminary solutions to generate the final solution. A total-state system does not distinguish roles for its sub-systems. The solution from a total-state system, whose measurement vector is composed of all the measurements from each sub-system, is an integrated positioning solution based on the weighted average of all its inputs.

An error-state system, however, is deployed with a reference system, and other sub-systems are referred to as aiding systems. The reference system is usually an INS or dead reckoning system, and GNSS is often used as an aiding system. Since the measurement vector of an error-state system consists of the differences between solutions from the reference system and aiding systems, outputs from the integration Kalman filter are corrections for the reference system. The corrections are fed back to the reference system at a regular interval as part of a complementary filtering process. Figure 3.1, adapted from Groves (2008), shows an example of an error-state complementary cascaded integration. A typical cascaded system is a loosely coupled INS/GNSS system. One thing to notice is that the use of a feedback in the reference system calibration introduces difficulties for the fault detection as each sub system is less independent.

A potential weakness of a cascaded architecture lies in the multiple filtering processes in both the local and system levels. An assumption of Kalman filtering is the positioning errors are composed of systematic errors, estimated as states, and white noise, modelled as system noise and measurement noise (Grewal *et al.*, 2007). However, the existence of Kalman filter and other filtering techniques in preliminary navigation calculation introduces time-correlated noises. The integration Kalman filter must therefore be carefully tuned to prevent instability, especially when integrating a "black box" system in cascaded architecture where the assumed measurement noise matrix must account for all operational conditions. Otherwise, the cascaded architecture should be avoided.

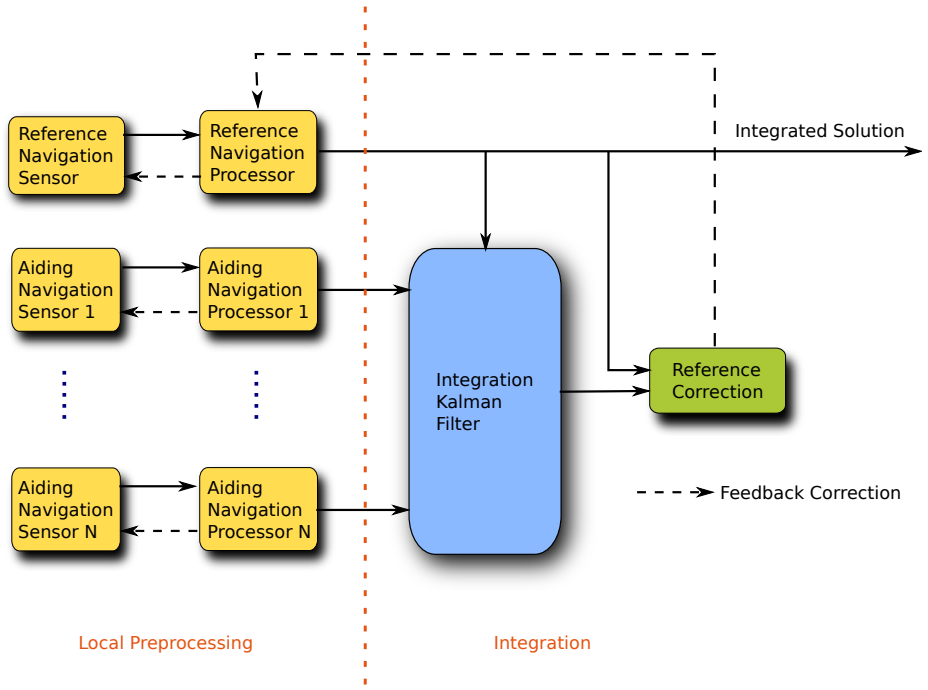



**Figure 3.1.:** An error-state complementary cascaded integration (adapted from Groves, 2008)

### 3.1.2. Centralised Integration

The centralised integration uses a single Kalman filter to replace both local and central level filters in cascaded integration. Similar to cascaded integration, the system implementation can be either total-state or error-state based. It should be noted that different systems are suited for different implementations. Systems, whose sub-systems consist only of positioning systems, are only suitable for a total-state implementation, whereas the integration of INS or other DR systems with positioning systems is usually realised with an error-state implementation.

Figure 3.2 shows an example of an error-state complementary centralised integration. As can be seen from the figure, raw measurements rather than preliminary positioning solutions of the aiding sensors are fed into the integration filter. Meanwhile the reference system is still providing positioning solutions to the filter as the baseline for corrections. The filter can generate corrections for both the reference system and aiding systems. These corrections can be fedback to individual sensors in a closed loop mode. A tightly-coupled INS/GNSS system is a typical centralised integration. The centralised integration architecture enables radio positioning systems such as GNSS to contribute to the integrated solution even when insufficient number of satellite are available to compute a stand-alone solution.

The major advantage of a centralised architecture is that both systematic errors and navigation sensors noise can be modelled in a single Kalman filter. In this way all error correlations, which are not available for a cascaded integration, are accounted for. Subsequently all measurements are optimally weighted, and maximum information can be used to calibrate each error. Also, a higher gain can be used before there is an instability risk in the filter since there is no assumption on measurement noises. However as this architecture requires raw sensor measurements and their error characteristic information, it is not compatible with “black box” navigation systems. A centralised architecture generally is able to provide an optimal positioning solution in terms of accuracy and robustness. However, with all the modelling process in one place, the downside for this architecture is a higher computation load for the integration processor. What is more, the integrity monitoring

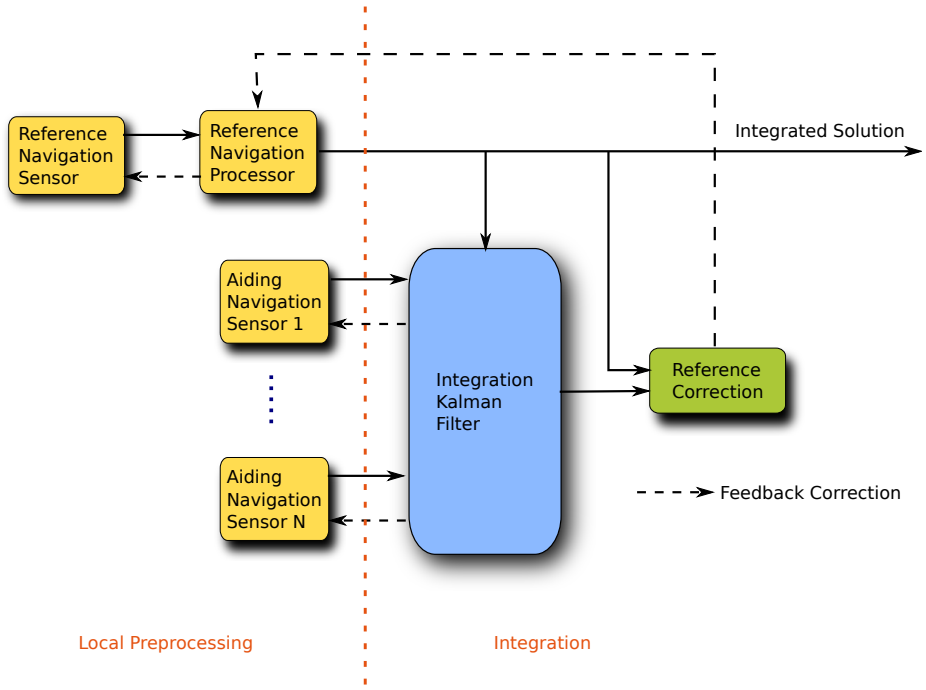



**Figure 3.2.:** An error-state complementary centralised integration (adapted from Groves, 2008)

under such an architecture becomes more difficult due to the unavailability of sub-system positioning solutions. For applications having high integrity requirements, processor-intensive parallel filters are often needed.

### 3.1.3. Federated Integration

Instead of feeding solutions from the reference system directly into the integration Kalman filter, as in cascaded or centralised integration, they can also be integrated locally with each aiding system in a set of local Kalman filters, as shown in Figure 3.3. This method of integration is called a federated integration. Each locally integrated solution is then combined using a centralised integration algorithm. The output from the integration algorithm is usually the final integrated solution, and complementary corrections can also be fed back to the reference system as in other integration architectures. Some system designs choose to input the reference solution to the centralised integration algorithm as well.

The integration in local filters can be either cascaded or centralised. With different ways to combine results from local filters, there are no-reset, fusion-reset, zero-reset and cascaded versions of a federated integration.

Both no-reset and fusion-reset implementations use a least-squares based integration algorithm (Carlson & Berarducci, 1994). The difference between the two methods is that the fusion-reset implementation feeds estimated states and error covariance back to each local filter. While the no-reset implementation suffers from overoptimistic error covariance, the fusion-reset implementation has a problem of how to model the correlations between the common and local states in each local filter.

Both zero-reset and cascaded implementations use a master Kalman filter to integrate results from local filters. Under a zero-reset implementation, all local filter states are zeroed and corresponding elements of the error covariance matrix are reset to their initial value after master filter receives inputs (Carlson, 2002). This implementation can be useful to process measurements at a faster

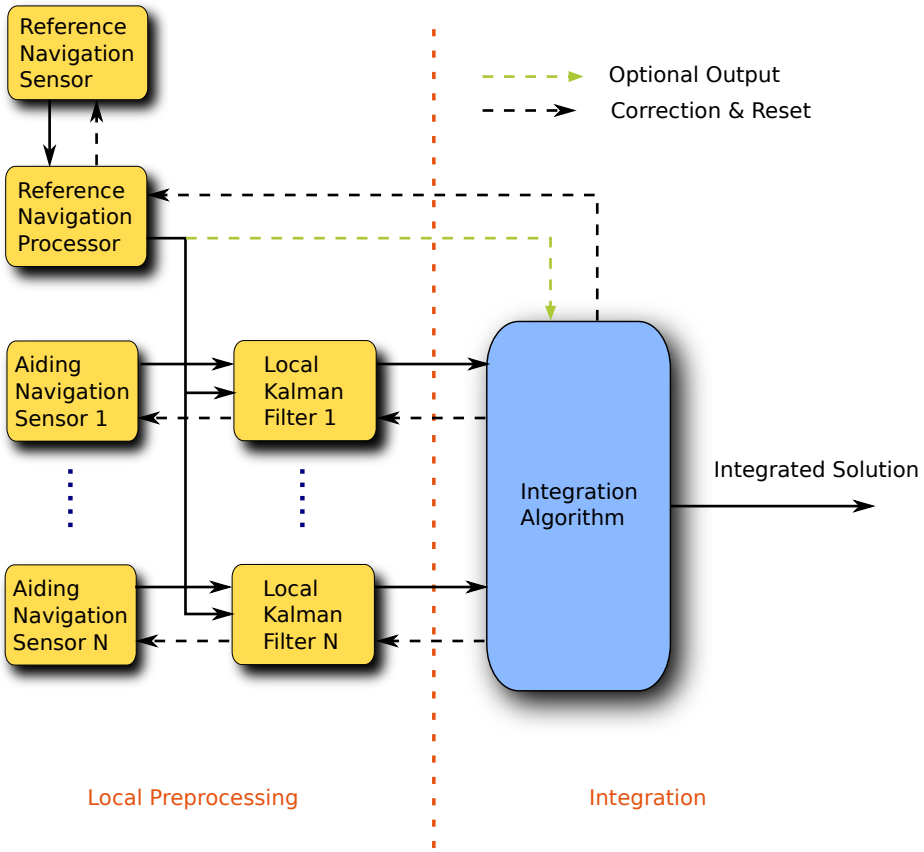



**Figure 3.3.:** A complementary federated integration (adapted from Groves, 2008)

rate (Groves, 2008). The cascaded implementation, on the other hand, has similar advantages and disadvantages from a cascaded integration as described in Section 3.1.1.

## 3.2. Kalman Filtering

The Kalman filter is a mathematical algorithm that is used for optimal estimation. It estimates the states of a system, usually crucial information for specific applications, using observed measurements that contain noise. Since its invention by R. E. Kalman in 1960, the Kalman filter has formed the bases of many navigation and positioning applications. Especially in an integrated system, it serves as the key to obtain an optimal estimation from multiple sensors and data sources.

Unlike other estimation algorithms such as least-squares, the operation of a Kalman filter is a recursive process. Therefore, real-time estimates can be maintained through a continuous propagation and update process of the system states. An initial set of estimates is required to start the recursion. During the filtering, estimates of the states are propagated through time with knowledge of the deterministic and statistical models of the system. In order to account for effects which are statistically random in the system models, a set of uncertainties in its estimation is also maintained. Whenever there are observed measurements, a weighted average of the old estimates predicted by the filter and the new ones derived from measurements is performed. As a result, new averaged estimates are produced and used for future propagation. This is referred as the state update. In this recursive manner, a Kalman filter utilises not only measurements of the current epoch but also information from prior estimates. This is particularly efficient for a real-time application, as although old data may be discarded, prior information is still optimally stored for future estimation.

A brief description of the essential aspects of the Kalman filtering process is given in this section. Section 3.2.1 gives a summary of main applications of the Kalman filter in the navigation and positioning field. Section 3.2.2 introduces basic elements of a Kalman filter, and a concise review of the filtering process is presented in Section 3.2.3. Algorithms and models for a Kalman filter are listed in Section 3.2.4. Section 3.2.5 describes an extension of the basic Kalman filter, the Extended Kalman Filter (EKF). A short discussion of implementation issues that have been encountered during this research is shown in Section 3.2.6. Information on other integration algorithms is also provided in Section 3.2.7. More detailed description on the Kalman filter and optimal estimation theory can be found at Grewal & Andrews (2008) and Crassidis & Junkins (2004).

### 3.2.1. Applications of a Kalman Filter

The Kalman filter and its various extensions have become the main tools for optimal estimation of dynamic systems since its first publication.

The main applications of the Kalman filtering techniques for navigation and positioning can be categorised based on the systems they work with. While working with an INS, a Kalman filter is usually used for alignment of the sensors and calibration for the solution. In a GNSS receiver, it is usual to implement a filter in the navigation processor as the main estimation method for positioning solutions. For GNSS signal monitoring, with the knowledge of receiver position and velocity and the access to high performance clocks, it is possible to estimate the time-correlated range errors with a Kalman filter. Lastly, a Kalman filter commonly serves as the core of an integration system, combing data from different sensors.

Other applications of the Kalman filter includes orbit determination from range and line-of-sight (angle) observations, target tracking of aircraft, attitude determination for space vehicles, etc. Moreover, the current realisation of GPS system time is also done through a Kalman filter with the concept of a composite clock (Brown Jr, 1991).

### 3.2.2. Elements of a Kalman Filter

A Kalman filter consists of four core elements: the state vector and its error covariance matrix, system model and system noise, measurement vector and measurement noise, and measurement model. Note that bold lowercase symbols are used through the chapter to represent vectors, and bold uppercase symbols represent matrices.

The state vector, $\mathbf{x}$, contains a set of parameters which describe system status, normally referred as the "states" of a system. These states are the estimates of a Kalman filter, and may be either constant or time-varying. For a real-time navigation application, some of the states can be streamed as outputs of the system. In a positioning system, position and velocity are always estimated as fundamental information. Other information such as acceleration, attitude and angular rate is commonly included for high dynamical GNSS applications and INS related systems. In addition, various error parameters, for instance sensor error characteristics, GNSS receiver clock errors and signal propagation errors, can be also estimated as states in specific applications for improved performance. However, the choice of states for an integrated system must also take the system observability issue into account. A more detailed discussion is given in Section 3.3.

The state error covariance matrix, $\mathbf{P}$, is an indication of uncertainties in states and the correlation between states. The error covariance is propagated and updated at each epoch associated with the state vector during the filtering. The uncertainties in the system come from statistically random factors of the system, either as measurements noise or system process noise. These noises are modelled as zero-mean Gaussian white-noise processes. The correlation information is also important

as information from measurements is not always enough to estimate each state independently due to observability issues. By maintaining such an information of uncertainty, estimates can be produced with an optimal combination of various sensor measurements. On the other hand, information provided by the error covariance is also useful for performance analyses during a Kalman filter design process. The Kalman filter algorithm is a recursive process. The initial values for both the state vector and error covariance matrix must be manually designated. The choice of initial values affects the speed with which the filter converges.

The system model in Kalman filter describes how states and error covariance propagate through time. The propagation for system states is based on known properties of the system such as the integral relationship between position and velocity in a positioning system. The system model is described by a system matrix, $\mathbf{F}$. A standard Kalman filter assumes a linear system model which is not true for some navigation applications. A solution to the problem is to perform a linearisation on the system model, which results in a family of Kalman filter extensions, such as the extended Kalman filter. While the dynamics for states propagation is deterministic, propagation for error covariance relies on taking account of unknown changes in the system. These changes, usually referred to as system noise or process noise ($\mathbf{w}_s$), can be either unmeasured dynamics or random instrument noises and their statistical properties are usually estimated in advance by the system designer. The system noise covariance matrix, $\mathbf{Q}$, quantifies the noise associated with the system model. System model and system noise together consist the basis of prediction or propagation phase for a Kalman filter in a single time epoch.

In the case of an integrated system, the measurement vector, $\mathbf{z}$, may contain measurements from various sources. Different integration architecture may use different measurements from the same sensor. For example, a loosely-coupled INS/GNSS system uses position and velocity solutions from the GNSS receiver, while a tightly-coupled system uses pseudoranges from all available satellites. The measurement vector is a function of the system states and contains latest information on the system status. In order to account for uncertainties in the measurements, a measurement noise covariance matrix ($\mathbf{R}$), which contains information for measurement noise (denoted as $\mathbf{w}_m$), is also fed into the filter. The uncertainty information is weighted against the uncertainties from states propagation during the update phase. For a real-time application, measurements do not always arrive simultaneously or at regular intervals as sensors come with different output rate.

The measurement model is the link between system states and available measurements. It is the function that describes how the measurement vector varies as the true state vector changes in the absence of measurement noise. The measurement model is also deterministic, similar to the system state model. In the Kalman filter algorithm, the model is described by a measurement matrix ($\mathbf{H}$), which contributes to the calculation of Kalman gain and the update of the state error covariance matrix. Note that a standard Kalman filter assumes a linear relation between states and measurements; if this is not the case then other variations of the Kalman filter such as an extended Kalman filter can be defined with a differentiable vector function.

### 3.2.3. Phases of Kalman Filtering

Kalman filtering is a recursive process, and each epoch can be divided into two phases: system propagation and measurement update. Calculations at each epoch start with system propagation, followed by a measurement update, provided sensor information is available. The two phases do not necessarily need to be performed at the same frequency. An illustration of the elements and phases of Kalman filtering is given in Figure 3.4. Note that symbols with a hat on top are used in Kalman filtering to represent estimates, and symbols with a tilde on top are measurements. Also, a minus

sign on the top right corner of a symbol means values before measurement update and a plus sign represents values after a measurement update.

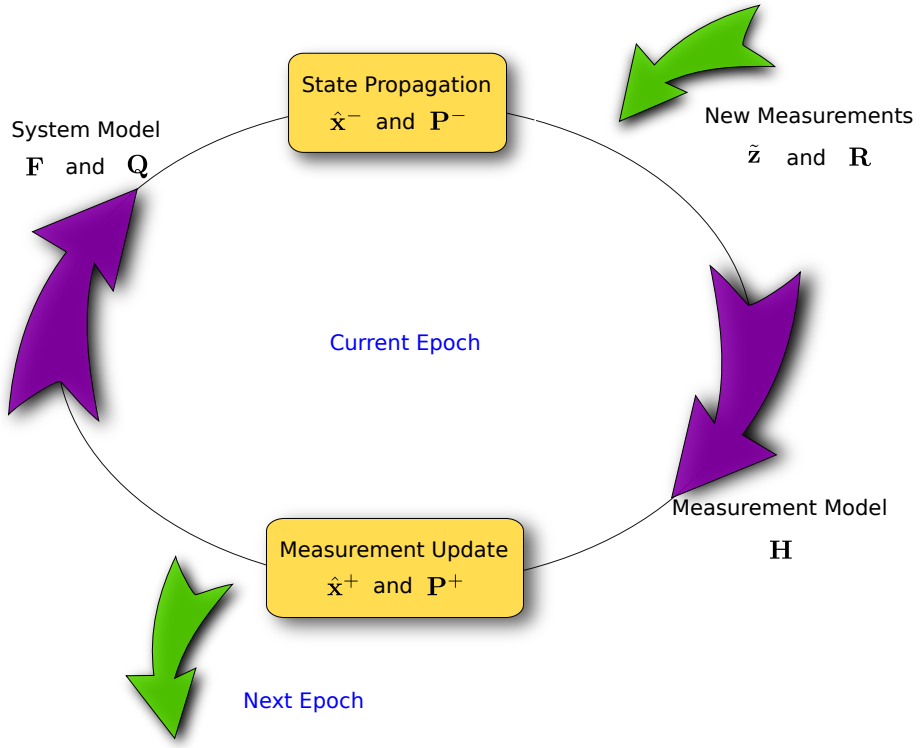



**Figure 3.4.:** A summary of Kalman filtering

During system propagation, both the system state vector and error covariance matrix are predicted forward using the system model. This provides the best estimation of the states from a filter at the current epoch without new measurement information, $\hat{\mathbf{x}}^-$ and $\mathbf{P}^-$. At the start of system propagation both the deterministic system model and noise model are firstly calculated. As most system models are based on continuous time modelling, while sensors work in discrete time mode, the calculated system matrix must be transformed into its discrete counterpart, known as the transition matrix, $\mathbf{\Phi}$. The transition matrix is a discrete representation of the system model and is a function of the time interval between Kalman filter iterations ($\tau_s$) and the system matrix. Similarly, a discrete representation of the system noise model can also be obtained with the time interval. The system states are then propagated into the current epoch using the deterministic system model, while the system noise error covariance matrix is used to compute the noise of the predicted state.

When there are measurements available, a measurement update is performed to incorporate new information into current estimates. Firstly, the measurement vector and its associated measurement noise covariance matrix are formed. A new measurement matrix may also need to be computed based on the types of measurements in the measurement vector. The core of the update process is an optimal weighted average scheme between filter prediction and estimates obtained through the measurement model. A gain is calculated by comparing uncertainties of both measurements and states predictions, referred as the Kalman gain matrix ($\mathbf{K}$). Finally measurement data is weighted using the Kalman gain and is incorporated into the state estimates, $\hat{\mathbf{x}}^+$. Meanwhile, the state error covariance is updated as well with information from measurement noise, $\mathbf{P}^+$.

## 3.2.4. Models and Algorithms

### 3.2.4.1. System Model

The system model describes how system states vary over time. In order to propagate the state vector into current epoch during calculation, the deterministic system dynamics and system uncertainties is required.

The assumption that a standard Kalman filter holds towards its system model is a linear relationship between the time derivative of each state and between white system noises. Therefore, the true state vector at time $t$, $\mathbf{x}(t)$, can be expressed using the following model,

$$\dot{\mathbf{x}}(t) = \mathbf{F}(t)\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}_s(t) \tag{3.1}$$

where $\mathbf{F}(t)$ and $\mathbf{w}_s(t)$ are respectively the system matrix and the system noise vector as covered in Section 3.2.2, and $\mathbf{G}(t)$ is the system noise distribution matrix. Both $\mathbf{F}(t)$ and $\mathbf{G}(t)$ are determined by known system properties and can be either constant or time varying.

In a positioning application, sensors provide output in discrete time. Thus a discrete Kalman filter, whose system model and measurement model are in a discrete form, is commonly used. In a discrete Kalman filter, the current states estimates are modelled as a linear function of their previous value and can be expressed as

$$\hat{\mathbf{x}}_k^- = \mathbf{\Phi}_{k-1}\hat{\mathbf{x}}_{k-1}^+ \tag{3.2}$$

where both $k$ and $k-1$ denote a single epoch during the operation, and are equivalent to times $t$ and $t-\tau_s$ when the measurements arrive. The discrete system model is a digitised representation of its continuous form. With $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}(t)$ and $\hat{\mathbf{x}}_{k-1} = \hat{\mathbf{x}}(t-\tau_s)$ , the transition matrix, $\mathbf{\Phi}_{k-1}$, at epoch $k-1$, can be expressed with $\mathbf{F}_{k-1}$ by approximations (Brown & Hwang, 1997),

$$\mathbf{\Phi}_{k-1} \approx \exp(\mathbf{F}_{k-1}\tau_s) \tag{3.3}$$

In general there are many numerical methods to compute (3.3). One common approach is to compute $\mathbf{\Phi}_{k-1}$ as a power-series expansion of $\mathbf{F}_{k-1}\tau_s$:

$$\mathbf{\Phi}_{k-1} = \sum_{r=0}^{\infty} \frac{\mathbf{F}_{k-1}^r \tau_s^r}{r!} = \mathbf{I} + \mathbf{F}_{k-1}\tau_s + \frac{1}{2}\mathbf{F}_{k-1}^2\tau_s^2 + \cdots \tag{3.4}$$

For most applications, a first order solution of (3.4) would work.

The propagation of the error covariance matrix in a discrete filter can be written in many forms. The most typical form is given by

$$\mathbf{P}_k^- = \mathbf{\Phi}_{k-1}\mathbf{P}_{k-1}^+\mathbf{\Phi}_{k-1}^T + \mathbf{Q}_{k-1} \tag{3.5}$$

and the system noise covariance matrix, $\mathbf{Q}_{k-1}$, is defined by

$$\mathbf{Q}_{k-1} = \mathbf{G}_{k-1} E\left(\int_{t-\tau_s}^{t}\int_{t-\tau_s}^{t} \mathbf{w}_s(t')\mathbf{w}_s(t'')dt'd''\right)\mathbf{G}_{k-1}^T \tag{3.6}$$

### 3.2.4.2. Measurement Model

The measurement model connects state estimates with measurements and accommodates new information to update predicted estimates. As mentioned in Section 3.2.2, a linear relationship between measurements and states is assumed by the standard Kalman filter, so the measurement

model in continuous time form can be modelled as

$$\mathbf{z}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{w}_m(t) \tag{3.7}$$

Note that measurement matrix $\mathbf{H}$ can be both time-varying or constant. For example, the measurement matrix of a loosely-coupled error-state INS/GNSS system, whose state vector contains errors in dead reckoning solution and measurement vector consists of differences between solutions from two systems, can simply be an identity matrix. Whereas the measurement matrix for a Kalman filter in the navigation processor of a GNSS receiver is a function of the satellite-to-user line-of-sight vector and needs to be recalculated for each update.

Similar to the system model, a discrete form of the measurement model needs to be formed for a discrete filter. As $\mathbf{z}$ and $\mathbf{x}$ have a linear relationship, a discrete measurement model can be written directly as

$$\tilde{\mathbf{z}}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{w}_{mk} \tag{3.8}$$

The core of a measurement process is to compute the Kalman gain, $\mathbf{K}_k$, for the epoch. $\mathbf{K}_k$ is essentially a weighted scale factor to be put on the difference between predicted and the observed measurements. The Kalman gain can be written in many different forms, but they are equivalent in theory. Although some of them may bring numerical benefits to a operational system, the most common expression is

$$\mathbf{K}_k = \mathbf{P}_k^-\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_k^-\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{3.9}$$

The measurement covariance matrix, $\mathbf{R}_k$, is usually a diagonal matrix and provides information on the measurement quality.

The update for system states is achieved by adding a correction to the prediction. This correction will be the scaled difference of observed and predicted measurements. Therefore, the update equation for the state vector is

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\tilde{\mathbf{z}}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^-) \tag{3.10}$$

The Joseph form of the error covariance update is given here as it provides better numerical stability (Brown & Hwang, 1997) and is presented as

$$\mathbf{P}_k^+ = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^-(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \tag{3.11}$$

#### 3.2.4.3. Algorithm Summary for Discrete Kalman Filter

A summary of the discrete Kalman filter algorithm is listed in this section with a reference to the equation involved at each step. A flow chart of a typical Kalman filtering steps is shown in Figure 3.5.

At the start of each epoch, inputs of the filter are estimates, $\hat{\mathbf{x}}_{k-1}^+$, and error covariance, $\mathbf{P}_{k-1}^+$, from previous epoch. Then the calculation goes through steps as following:

1. Calculate system matrix $\mathbf{F}_{k-1}$ and then transform it into transition matrix $\mathbf{\Phi}_{k-1}$ using (3.4).
2. Form the system noise covariance matrix, $\mathbf{Q}_{k-1}$.
3. Predict estimates for current epoch, $\hat{\mathbf{x}}_k^-$ , with $\hat{\mathbf{x}}_{k-1}^+$using (3.2).
4. Predict the error covariance for current epoch, $\mathbf{P}_k^-$, with $\mathbf{P}_{k-1}^+$ using (3.5).
5. If there are new observed measurements, construct the measurement vector $\tilde{\mathbf{z}}_k$.
6. Construct the associated measurement noise covariance matrix, $\mathbf{R}_k$.

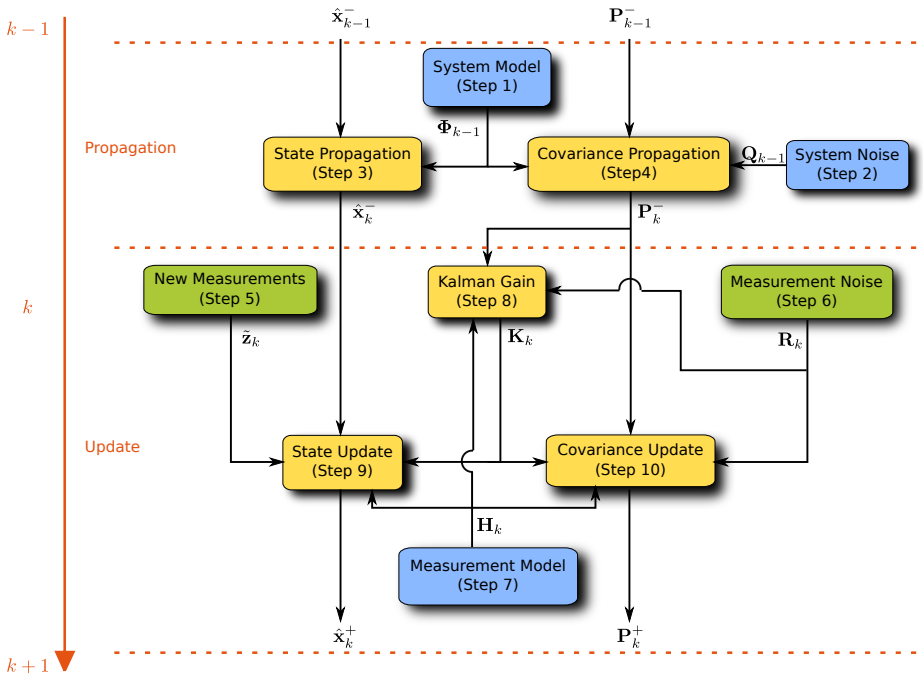



**Figure 3.5.:** Data flow of a Kalman filtering epoch

7. Calculate the new measurement matrix, $\mathbf{H}_k$.
8. Kalman gain computation using (3.9) to get $\mathbf{K}_k$.
9. Update the state vector with (3.10) to obtain $\hat{\mathbf{x}}_k^+$.
10. Update the error covariance with (3.11) to obtain $\mathbf{P}_k^+$.

Note that the system noise matrix $\mathbf{Q}$ and the new measurement matrix $\mathbf{H}$ can be constant for some systems. Furthermore, in order to reduce the computation load, the propagation of the error covariance $\mathbf{P}$ can be performed at a lower frequency than the state propagation.

### 3.2.5. Extended Kalman Filter

As mentioned in Section 3.2.2, one of the assumptions in a standard Kalman filter is linear system and measurement models. However, a large class of estimation problems involve nonlinear models. For instance, the system model of a total-state INS/GNSS system and the measurement model of a GNSS navigation filter are both highly nonlinear. These nonlinear systems can be described using

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{G}(t)\mathbf{w}_s(t) \tag{3.12}$$

$$\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t)) + \mathbf{w}_m(t) \tag{3.13}$$

where $\mathbf{f}$ and $\mathbf{h}$ are nonlinear functions of the state vector in respectively the system model and the measurement model.

There are many possible ways to produce a linearised version of Kalman filter to deal with systems as described by (3.12) and (3.13), and one of them is known as the extended Kalman filter (EKF). The fundamental concept of the EKF is that the error dynamics can be approximated by a first order Taylor series expansion, as long as the error in the state vector is sufficiently small. With the state vector residual, $\delta\mathbf{x}$, defined as $\delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$, a linear model can be applied to the state vector

residual:

$$\delta\dot{\mathbf{x}}(t) = \mathbf{F}(t)\delta\mathbf{x}(t) + \mathbf{G}(t)\mathbf{w}_s(t) \tag{3.14}$$

$$\delta\tilde{\mathbf{z}}(t) = \mathbf{H}(t)\delta\mathbf{x}^-(t) + \mathbf{w}_m(t) \tag{3.15}$$

where $\delta\tilde{\mathbf{z}}$ is the measurement innovation and is defined as $\delta\tilde{\mathbf{z}} = \tilde{\mathbf{z}} - \mathbf{h}(\hat{\mathbf{x}}^-)$. The system matrix, $\mathbf{F}(t-\tau_s)$, and the measurement matrix, $\mathbf{H}(t)$, are defined according to the first order term of a Taylor series expansion,

$$\mathbf{F}(t-\tau_s) = \mathbf{F}_{k-1} = \left.\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{k-1}^+} \tag{3.16}$$

$$\mathbf{H}(t) = \mathbf{H}_k = \left.\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_k^-} \tag{3.17}$$

The discrete form of the linearised model can therefore be given as follow:

$$\delta\hat{\mathbf{x}}_k^- = \mathbf{\Phi}_{k-1}\delta\hat{\mathbf{x}}_{k-1}^+ \tag{3.18}$$

$$\delta\tilde{\mathbf{z}}_k = \mathbf{H}_k\delta\hat{\mathbf{x}}_k^- + \mathbf{w}_{mk} \tag{3.19}$$

The transition matrix $\mathbf{\Phi}_{k-1}$ can be calculated using (3.4).

During the implementation of an error-state Kalman filter, the estimated error in the system solution can be fed back to correct the system itself either at every iteration or at regular intervals. Filtering processes with this structure are called complementary filtering, or closed-loop filtering in comparison to open-loop filtering where the estimates generally grow as time progresses. For an EKF, the complementary structure is desirable, because it keeps the state estimates small and therefore minimises the effect of neglecting higher order term of the Taylor series during the linearisation process. This, to some extent, keeps the Kalman filter assumption of linear models satisfied to the first order.

If the error estimates are applied to correct the system solution immediately after the measurement update, the state estimates become zero at the beginning of each state propagation phase and (3.18) can therefore be omitted; whereas the error covariance propagation stays the same. The propagation of the state vector and the calculation of predicted measurement is achieved through the original nonlinear functions $\mathbf{f}$ and $\mathbf{h}$. The algorithm for complementary EKF, which is modified from Section 3.2.4.3, is listed as follows:

1. Calculate system matrix $\mathbf{F}_{k-1}$ using (3.16) and then transform it into transition matrix $\mathbf{\Phi}_{k-1}$ using (3.4).
2. Form the system noise covariance matrix, $\mathbf{Q}_{k-1}$.
3. Predict estimates for current epoch, $\hat{\mathbf{x}}_k^-$ , with $\hat{\mathbf{x}}_{k-1}^+$using

$$\hat{\mathbf{x}}_k^- = \hat{\mathbf{x}}_{k-1}^+ + \int_{t-\tau_s}^{t} \mathbf{f}(\hat{\mathbf{x}}, t')dt' \tag{3.20}$$

4. Predict the error covariance for current epoch, $\mathbf{P}_k^-$, with $\mathbf{P}_{k-1}^+$ using (3.5).
5. If there are new observed measurements, construct measurement vector $\tilde{\mathbf{z}}_k$.
6. Construct the associated measurement noise covariance matrix, $\mathbf{R}_k$.
7. Calculate the new measurement matrix, $\mathbf{H}_k$, using (3.17).

8. For measurement innovation, $\delta\tilde{\mathbf{z}}_k$ , with $\tilde{\mathbf{z}}_k$ and $\hat{\mathbf{x}}_k^-$ using

$$\delta\tilde{\mathbf{z}}_k = \tilde{\mathbf{z}}_k - \mathbf{h}(\hat{\mathbf{x}}_k^-) \tag{3.21}$$

9. Compute the Kalman gain matrix using (3.9) to get $\mathbf{K}_k$.

10. Update the state vector with

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k\delta\tilde{\mathbf{z}}_k \tag{3.22}$$

to obtain $\hat{\mathbf{x}}_k^+$.

11. Update the error covariance with (3.11) to obtain $\mathbf{P}_k^+$.

### 3.2.6. Implementation Issues

During the process of implementing a practical Kalman filter, there are usually various numerical issues involved. Most are related to the rounding errors due to precision limitation of computers and the large number of matrix operations (especially matrix inversion) at each epoch. Also, the processing capacity of the platform where the filtering takes place occasionally can cause an increase in algorithm complexity. A robust design of a Kalman filter algorithm must take these factors into account. This section gives a brief review of some numerical issues that were encountered during this research.

#### 3.2.6.1. Error Covariance Matrix Symmetry and Positive Definiteness

The error covariance matrix should always be symmetric and positive semi-definite by its definition. However, as the digit precision of the computer is limited, when rounding errors start to accumulate through iterations the error covariance matrix may become non-symmetric and lose its positive semi-definite property. This affects the Kalman gain calculation and may therefore result in temporary divergence of the solution. Under such circumstances, the subsequent performance of the system is compromised as the information carried by the error covariance is no longer accurate. There are few techniques that can be applied to deal with the problem:

1. As mentioned in Section 3.2.4.2, the update equation of the error covariance matrix can be written in various forms. The Joseph form, given in (3.11), is a symmetric form and can be applied.

2. Since the error covariance matrix is symmetric according to its definition, either the upper or the lower triangular part of the matrix can be redundant. That is, only the diagonal and half of the triangular part of the matrix are stored during the calculation. For applications with very limited processing power, this method also substantially reduces memory and computational requirements.

3. By taking the average of the error covariance and its transposed matrix using

$$\mathbf{P} = \frac{1}{2}(\mathbf{P} + \mathbf{P}^T) \tag{3.23}$$

at each iteration or at regular intervals, the error covariance can be forced to maintain its symmetry. Note in case any non-numerical errors occur during the calculation and cause $\mathbf{P}$ to become non-symmetric, (3.23) would worsen the situation.

4. A factorisation method may be used. Any symmetric matrices can be factorised into the production of several matrices with special forms. For Kalman filtering, one common factorisation method is the UD factorisation. For a UD implementation of the Kalman filter, the error covariance matrix is factorised as a triangular matrix and a diagonal matrix. Factorised matrices are propagated and updated during the filtering instead of the state error covariance. However, a factorised Kalman filter is very complex to implement and increases the computational requirements.

#### 3.2.6.2. Scaled State Estimates

The matrix inversion operation in a computer is done in various numerical ways, commonly using Gaussian elimination, without actually forming the inverse. The advantage of doing this is mainly for higher computational efficiency, i.e. less computation time and less residual errors. The drawback of these numerical methods is an increased sensitivity to the condition of the matrix. In theory, any non-singular square matrix is invertible. But matrices that are very close to singular or bad-conditioned may also be problematic to numerical matrix inversion.

For some integration applications, values of estimates in the state vector could contain large differences in scale. For example, the estimated position expressed in local navigation in the form of latitude, longitude in radians is significantly smaller than other components in the state vector such as the velocity. This results in a badly scaled error covariance matrix. In such cases, the matrix inversion presented in Kalman gain calculation with (3.9) may face difficulties. The results of (3.9) may be either a less accurate result or a report of faults because of non-invertible matrices.

The best solution to this problem is to choose appropriate units for all the estimates in the state vector, or even manually apply a scale to some states so that the error covariance matrix maintains a reasonable range.

#### 3.2.6.3. Algorithm Design Issues

For the design of a practical Kalman filter and eventually an integrated positioning system, the processing capacity of the platform must always be considered. Since the filtering process involves lots of matrix operations, for which the memory capacity requirement depends on the size of the matrices and vectors, it is very common to adapt the basic algorithms in Section 3.2.4.3 and Section 3.2.5 in order to accommodate the limitations of the hardware.

Depending on the system properties, some systems do not have a time-varying system model or measurement model, i.e. $\mathbf{F}$ or $\mathbf{H}$ could be constant during the filtering. Furthermore, the system noise covariance and the measurement noise covariance could also be constant. All constants in the filtering can be processed off-line, and subsequently reducing the processing load. In addition, all major matrices in the filtering process contain a large amount of zeros. It would be more efficient to applied sparse matrix algorithms for these applications.

The calculation for the error covariance matrix requires significantly more computational effort. This is especially stressful for the processor, especially when there are high rate sensors in an integrated system. On the other hand, the accuracy required for the state vector is higher than the error covariance. In order to balance the requirements for computational power against calculation accuracy, different steps of a Kalman filter algorithm can be performed at a different rate. The state propagation is usually performed at the highest rate, whereas the error covariance can be propagated at a lower rate.

### 3.2.7. Other Integration Algorithms

Apart from the standard Kalman filter and the EKF described in previous sections, there are other algorithms that can be applied for optimal estimation and for an integrated positioning system. A concise introduction to some of them is given in this section. The application of different integration algorithms has effects on the system performance. The focus of this thesis is, however, to investigate the performance of integrated positioning systems with DRM aiding, so the effects of different integration algorithms is not investigated; this may be an interesting area for further research.

#### 3.2.7.1. Weighted Least Squares

Weighted least squares is a modified version of the standard linear least squares estimation. All least squares methods follow Gauss's principle of least squares, which selects an optimum set of unknown parameters that minimises the sum of the squares of the estimation residual errors. The standard least squares assumes all observed measurements are of the same quality. This is however not true for most applications. Placing equal emphasis on measurements from different sensors for an integrated system is logically unsound. Weighted least squares therefore incorporates a weighting scheme during the calculation.

Simplicity and low processor load are the main advantages of using weighted least squares for system integration. The weighted least square algorithm works in a snap-shot mode in comparison with a Kalman filter, i.e. running iterations in a single epoch until the result converges. Thus it is not suitable for dead reckoning based sensors and inertial sensors, due to it lacking means to calibrate the continuous drift of the sensors. The errors in the sensors are simply assigned with a weight with no consideration of the drift.

#### 3.2.7.2. Other Extensions of the Kalman Filter

Apart from the EKF described in Section 3.2.5, there are other extensions of the standard Kalman filter that have been developed. There are many practical issues in applying a Kalman filter to various system parameter estimation problems. Each of these extensions has been developed to deal with certain aspects of these issues.

The linearised Kalman filter is based on the same principle as an EKF in order deal with the application of standard Kalman filter for nonlinear models. For both the linearised Kalman filter and the EKF, the nonlinear models involved are linearised with respect to a nominal trajectory of the state vector. The nominal trajectory for the EKF is defined to be equal to the estimated trajectory, or previous estimates. On the other hand, for applications such as orbiting satellites or guidance weapons, the vehicle is expected to follow a predetermined trajectory that can be computed in advance. A linearised Kalman filter takes advantage of this knowledge and the estimates of the filter are usually the error between the actual and expected trajectory (Brown & Hwang, 1997).

The unscented filter is developed specifically to deal with nonlinear model situations where the first-order Taylor series approximation is no longer adequate. The premise behind an unscented filter is that a Gaussian distribution should be easier to approximate than a nonlinear function with a fixed number of parameters (Julier & Uhlmann, 1997). A series of samples known as sigma points is taken at each epoch to start the estimation. Comparing with the EKF, the unscented filter comes with some advantages including the ability to be applied to non-differentiable functions and avoiding the derivation of Jacobian matrices. However the prices for these advantages are more complex algorithms and heavier system load.

The adaptive filter is useful for situations where accurate knowledge of the noise covariance matrices is not available (Hide *et al.*, 2003). For a normal Kalman filtering process, both the system

noise covariance matrix, $\mathbf{Q}$, and the measurement noise covariance matrix, $\mathbf{R}$, are predetermined during the system design. If these matrices do not actually represent the noise, the subsequent performance of a standard filter is not optimal. The concept of an adaptive filter is to estimate $\mathbf{R}$ and/or $\mathbf{Q}$ during the operation. This can be achieved through either innovation-based methods or multiple-model based methods.

## 3.3. Observability and State Selection

From the point view of modern control theory, a system may have many internal variables and fewer outputs. For a normal control application, a controller is designed to optimise the outputs. In the case of a positioning system, the outputs can be measured (i.e. available measurements from sensors), and the knowledge of the state vector is desired but is not directly measured. The problem is referred as 'state estimation' in control theory. The integration algorithm, usually a Kalman filter, thus is essentially a state estimator. When designing a state estimator, it is important to determine whether the states can be estimated from available measurements. In other words, the observability of the system should be considered.

The observability of a system is indicated by its observability matrix. For a system described by (3.2) and (3.8) with $n$ states, the observability matrix is defined as

$$\boldsymbol{\mathcal{O}}^T = \left[\mathbf{H}^T,\ \boldsymbol{\Phi}^T\mathbf{H}^T,\ \cdots,\ \left(\boldsymbol{\Phi}^T\right)^{n-1}\mathbf{H}^T\right] \tag{3.24}$$

The system is said to be observable if $\boldsymbol{\mathcal{O}}$ has a full rank, i.e. its rank equals $n$. The implication behind this test is that if $n$ rows are linearly independent, then each of the $n$ states is viewable through linear combinations of the output variables $\mathbf{z}$.

In terms of the state selection for an integrated system, some common states that are usually included for an INS/GNSS integration are discussed here. The position and velocity of the system is always included, either in the form of actual position and velocity in a total-state implementation, or in the form of errors in the position and velocity in an error-state implementation. The more specific reference frame that the estimates are resolved in depends on the system design. For the local navigation frame, the position is in the form of latitude, longitude and height.

For INS, attitude error is almost always helpful unless the highest grade sensors are used. The attitude error is expressed in the form of small angles and resolved in the body frame. The biases for sensors in the IMU, such as the accelerometer biases and the gyro biases, are estimated in most INS/GNSS systems. Whether the biases should be modelled as the sum of a static component and a dynamic component or as a single parameter depends on the individual system. Including of other error characters, such scale factors and cross-coupling errors, is usually on the borderline. A further discussion can be found in (Groves, 2008).

The choice for GNSS states, on the other hand, varies between integration architectures. For a loosely-coupled system, there is no need to include any GNSS states, whereas for tightly-coupled and deep integration, the receiver clock offset and drift must be estimated in order to compensate for receiver clock errors in the pseudorange measurements.

# 4. Global Navigation Satellite Systems

Global Navigation Satellite System (GNSS) is a term for satellite-based radio navigation systems that are capable of providing 3D positioning solution for the user. Using GNSS for safety-of-life applications has been proved feasible and is now standardised for other public transportation systems such as aviation. The railways are shown as a characteristic environment for modern positioning technologies, especially GNSS. The research for application of GNSS on the railways is actively carried out at present.

This chapter presents a brief description of current GNSS status and the basic positioning method using satellite pseudorange measurements. A review of major GNSSs is given in Section 4.1 covering basic information on their system architecture and signal characteristics. Section 4.2 describes methods of estimating user position and velocity with pseudorange measurements. Main error sources for ranging are introduced in Section 4.3 and typical positioning performance figures are summarised for reference in Section 4.4. Finally, Section 4.5 includes a concise discussion on the future development of GNSS and its impact on GNSS applications on the railways.

## 4.1. Current GNSSs

### 4.1.1. GPS

Global Positioning System (GPS) was first development by the US government for military usages and is controlled by Joint Program Office (JPO) under the auspices of the Department of Defence (DoD). The full operational capacity is achieved at the end of 1994 and the system is now under active modernisation process. Also, plans for the future GPS III are under discussion.

#### 4.1.1.1. The Space and Control Segments

GPS system consists of three parts: space, control, and user equipment segments. The space segment is a satellite constellation in orbit and its satellites provide ranging signals and data messages with navigation information to system users. The control segment (CS) is various ground stations whose responsibility are to monitor, command and control the satellite constellation. Lastly, the user equipment segment is devices that perform the positioning and navigation, timing, or other related applications.

The GPS satellite constellation operates with a baseline constellation of 24 satellites. The satellite orbits are nearly circular with an approximate orbital radius of $26,600\,\text{km}$. GPS satellites are deployed in six orbital planes. These orbital planes are equally distributed around the equator at a $60^\circ$ separation, and have a nominal inclination relative to the equatorial plane of $55^\circ$. The nominal orbital period of a GPS satellite is approximately half a sidereal day (11 hours, 58 minutes). There are at least four Block II satellites in each orbital plane which are not evenly spaced. This is designed to ensure satellite visibility for most places on the earth. As a result, 5 to 14 satellites were planned to be available at most times, assuming a $5^\circ$ elevation mask angle and a clear line of sight (Spilker, 1996).

GPS satellites have been through generations of improvements. The Block I satellites were prototypes and the Block II satellites were the first full operational satellites. Both types are no longer in service. The Block IIA satellites are similar to Block II but with system enhancements to attain a longer 180 days of operation without contact from the CS (Dorsey *et al.*, 2006). The Block IIR are operational replenishment satellites features with satellites incorporate automatic navigation (Autonav). This feature enables Block IIR to update navigation data from intersatellite ranging measurements. The modernised replenishment satellites Block IIR-M introduce new L-band signals with increased power. The next model of GPS satellites that is beginning to be deployed is Block IIF (follow-on sustainment satellites). The new model will introduce a third signal frequency L5 and the first one is launched in mid-may 2010. According to Madden (2009), there are currently 30 satellites been set healthy on orbits including 11 Block IIA satellites, 12 Block IIR satellites and 7 Block IIR-M satellites. Furthermore, there are one more Block IIR-M satellite waiting to be set healthy and 3 additional satellites in residual status. A summarised GPS constellation status is listed in Table 4.1.

**Table 4.1.:** Generations of GPS satellites

| GPS Satellites | Years in Commission | Number of Satellites |
|---|---|---|
| Block I | 1978-1985 | 10 |
| Block II | 1989-1990 | 9 |
| Block IIA | 1990-1997 | 19 |
| Block IIR | 1997-2004 | 12 |
| Block IIR-M | 2005-2009 | 8 |
| Block IIF | From 2010 | 1 |
| Block III | From 2014 | - |

The main elements of the CS include the Master Control Station (MCS), monitor stations and ground uplink antennas. The major responsibility for the MCS is to maintain satellites constellation, generate navigation messages and maintain GPS timing services. Monitor stations are under the control of the MCS and form a network across the world to continuously collect satellites ranging data, satellites status data, and local meteorological data for the MCS to process. Ground uplink antennas forms another global network to perform satellites commanding and navigation message uploading. Current CS configuration comprised dual MCSs, six monitor stations and four ground antennas (Dorsey *et al.*, 2006). Update plans for the CS such as the Legacy Accuracy Improvement Initiative (L-AII) and the Architecture Evolution Plan (AEP) is currently in progress aiming to provide more monitor stations and new MCS mainframe. Also the planning for the next generation control segment, Operational Control Segment (OCX), is under discussion (Crews, 2008).

#### 4.1.1.2. Signals

Signals from GPS satellites contain three components: the radio frequency (RF) carrier, ranging codes, and navigation data. Among these components, ranging codes are a family of pseudo-random noise (PRN) codes that allow precise ranging measurements. Navigation data is a binary-coded message containing information about the satellite status and ephemeris. Both ranging codes and navigation data are modulated on the RF sinusoidal carrier to be broadcast by satellites.

There are currently two frequencies for GPS satellites to broadcast signals. Both frequencies are

located in the L-band and are refereed as Link 1 (L1) and Link 2 (L2). The centre frequencies of L1 and L2 are as below:

$$\text{L1}: f_{\text{L1}} = 1575.42\,\text{MHz}, \text{ and } \quad \text{L2}: f_{\text{L2}} = 1227.60\,\text{MHz}.$$

A new signal frequency had been proposed by modernised GPS, called L5. The L5 is available for civil applications and is centred at $f_{\text{L5}} = 1176.45\,\text{MHz}$. A L5 demo payload is included on the Block IIR-M7 satellite (SVN 49, lunched on March 2009) broadcasting L5 signal with no data modulated. The L5 signal is officially available along with the first launch of Block IIF satellite in 2010.

There are two types of ranging codes available on both frequencies, one for civil users (SPS), and the other for DoD-authorised users (PPS). The SPS codes include Coarse and Acquisition codes (C/A-codes) on L1 and the new Link 2 Civil (L2C) codes on L2. The PPS codes include precision (encrypted) codes, or P(Y)-codes, and the new Military (M)-codes on both L1 and L2. L2C and M-codes were introduced since the first launch of Block IIR-M on September 2005. A summary of GPS signals is listed in Table 4.2. Planned signals for modernised GPS are also included in Table 4.2.

**Table 4.2.:** A summary of GPS signals

| Codes | Carrier Band | Service | Minimum Received Signal Power (dBW) | Frequency Protection? | Satellite Blocks |
|---|---|---|---|---|---|
| C/A | L1 | SPS | -158.5 | Yes (ARNS band) | All |
| P(Y) | L1 | PPS | -161.5 | Yes (ARNS band) | All |
| | L2 | PPS | -164.5 | Not assured | All |
| L1C | L1 | SPS | -163 & -158.3 | Yes (ARNS band) | From III |
| L2C | L2 | SPS | -160 | Not assured | From IIR-M |
| M | L1 | PPS | - | Yes (ARNS band) | From IIR-M |
| | L2 | PPS | - | Not assured | From IIR-M |
| L5 | L5 | SPS | -158 | Yes (ARNS band) | From II-F |

One thing needs to be noted for safety-of-life railway applications is that L2 does not enjoy the same institutional protection from International Telecommunications Union (ITU) against interferences as L1 and L5. Unlike L1 and L5 which are allocated in Aeronautical Radionavigation Service (ARNS) band, L2 is in a shared band with other applications as well as amateur radios and is thus facing more potential interferences than L1 and L5. Therefore, L2C is unlikely to be used for safety-of-life applications (Misra & Enge, 2006 and Groves, 2008).

The broadcast navigation message consists of three parts including information on GPS time and satellites health status, ephemeris data containing 16 Keplerian orbital parameters of the satellite, and almanac with approximate orbital information of the entire GPS constellation. The ephemeris is used to precisely determine the satellite orbital position and thus is updated more frequently than other information, currently 2 hours. The almanac data, with a longer validated period, is intended to aid user equipment in selecting satellites and acquiring signals.

Currently, there are three formats available for the broadcast navigation message. The legacy navigation message is used by both the C/A-code and P(Y)-codes. CNAV messages are broadcast

on the L2C and MNAV messages are on the M-codes. The information on legacy messages is organised in a fixed-frame format with 5 subframes. The data rate is 50 bit/s and the full message lasts 12.5 minutes. CNAV and MNAV, however, offer a more modernised information organisation and require less bandwidth. Instead of using a frame/subframe architecture, CNAV and MNAV feature a new pseudo-packetized format. Packets are headed by a message type indicator and can be transmitted in any order. Each packet uses Forward Error Correction (FEC) and thus has more advanced error detection than the legacy messages. In addition to information provided by the legacy format, extra information on integrity, GNSS interoperability, and augmentation can also be included in CNAV and MNAV. MNAV packets are longer than CNAV packets and come with higher precision ephemeris.

#### 4.1.1.3. Augmentation Systems

Augmentation systems for GPS are mainly divided into two types based on the area the system is design to serve and the way to transmit augmentation data. While the service of Ground Based Augmentation Systems (GBAS) is in general localised and the broadcasting of GBAS messages is through ground based transmitters, Satellite Based Augmentation Systems (SBAS) provide services to a wide area through satellites with a lower precision. Additional ranging signals, differential correction service, and integrity alerts is delivered to user by augmentation systems. Examples of GBAS applications include local area DGPS services for airfields or harbours. On the other hand, there are currently five SBAS systems at varying stages of development, as summarised in Table 4.3 (Hein *et al.*, 2007 and Walter *et al.*, 2010).

**Table 4.3.:** A summary of current SBAS

| SBAS | Service Coverage | Constellation | Geostationary Satellite Longitudes | Current Status |
|---|---|---|---|---|
| WAAS | North America | GEO (2) | 107° W and 133° W | Full performance |
| EGNOS | Europe and surrounding countries | GEO (3) | 21.5° E, 25° E and 15.5° W | Initial operation |
| MSAS | Japan | GEO (2) | 140° E and 145° E | Initial operation |
| GAGAN | India | GEO (3) | 34° E, 83° E and 132° E | Under development |
| SDCM | Russia | GEO (3) | 95° E, 167° E and 16° W | Under development |

EGNOS is the European SBAS made up of three geostationary satellites and a network of ground stations. Until the end of 2009, EGNOS is only certificated by the European Commission for Non-Safety-of-Life use and is expected to be certified for Safety-of-Life applications in 2010. However, in terms of providing augmentation for positioning on UK railways, EGNOS suffers from severe visibility issues, as mentioned in Section 2.3.2 on page 42, because of low elevation angle for geostationary satellites. To address visibility problems for EGNOS satellites, ESA released an internet service for continuous delivery of EGNOS signals, called SISNet. Nevertheless, communication arrangements within the railways signalling system and the compatibility with future ERTMS is yet to be standardised.

#### 4.1.1.4. GPS Modernisation

GPS modernisation process is gradually starting to take place and will eventually lead to the future GPS III system and beyond. The plan for GPS modernisation focuses on both hardware and satellite signals.

The hardware upgrade includes the launch of new satellites and the construction of new control station. The Block IIF satellites are able to broadcast a new civil signal frequency. Meanwhile, Block III satellites are currently under design and will mark the coming of GPS III system. This new type of satellites will feature increased signal power, the introduction of signal integrity and another new civil signal on L1. On the other hand, the CS modernisation is focusing on the building of the next generation control segment OCX. Upgrades will enable the CS to incorporate with modernised satellite signals and navigation messages. In addition, the enhanced OCX will be capable of not only monitoring a larger modernised GPS constellation but also providing information on system integrity.

The new signal frequency L5 is available along with the launch of Block IIF satellites. There are two PRN codes available on L5 known as I5 and Q5. The new L5 frequency is located in ARNS band and is intended for civil safety-of-life application. Further introduction of civil signals will be realised by Block III satellites and the new signal is referred as L1C. New civil signals in ARNS band will allow safety-of-life applications, such as positioning on the railways, to take the advantage of positioning with multiple frequencies. Meanwhile, the plan for GPS III also includes the proposal for an increased signal broadcasting power. The increase of signal power can improve the availability of GPS services under signal difficulty situations. Railway line-of-sight problems will substantially benefit from an increased signal power.

### 4.1.2. GLONASS

GLONASS, GLObal'naya Navigatsionnaya Sputnikovaya Sistema, is the radio satellite navigation system developed by Soviet Union and is now operated by the Russian government. The GLONASS constellation was once allowed to decay due to financial problems after its completion in 1995. Since 2001, a modernisation program has been carried out including the launch of new satellites, introducing new signals, and updating control complex.

The GLONASS constellation consists of 21 operational satellites till the time of writing (Russian Space Agency, 2010). The satellites are deployed in 3 orbital planes, each with a 120° interval to other planes in longitude and an inclination of 64.8° to the equator. The GLONASS orbits has a radius of 25,600 km and each satellite completes the orbit with a period of 11 hours, 15 minutes. Most of the constellation consists of modernised GLONASS-M satellites. According to Revnivykh (2008), the development of new GLONASS-K satellite will finish flight test by 2011.

At present, GLONASS signals are broadcast on two frequencies, known as GLONASS L1 and L2. Each of these frequencies lies just above their GPS counterparts. PRN codes that are modulated on both frequencies are C/A codes for civil applications (available on L2 from GLONASS-M) and P code for military applications. Each PRN code is broadcast with a different navigation message. While specifics for the P code navigation message are not publicly published, the C/A code navigation message contains information on satellite ephemeris, almanac of the satellite constellation. Future plans to include differences between GLONASS and other GNSSs, such as system time and reference frame, in navigation messages are under consideration.

There will be a third frequency L3 available when the new satellites GLONASS-K are ready. At the moment, GLONASS signals use Frequency-Division Multiple Access (FDMA), which introduces extra complexities in GLONASS receivers and causes problems for GNSS interoperability. But the

use of Code-Division Multiple Access (CDMA) will be available along with the launch of GLONASS-K (Revnivykh, 2008).

Unlike GPS, the ground-based control complex (GBCC) of GLONASS is mostly restrained within Russian territory. The current GBCC is made up of a system control centre and two monitor stations near Moscow, four telemetry, tracking and control (TT&C) stations (Feairheller & Clark, 2006). The modernisation plan for the GBCC includes the extension of ground control network, improvements on system time and orbit, and new monitoring network outside Russia.

The specification of system positioning accuracy for civil applications is "100 m (95% probability) in horizontal, 150 m (2 sigma) in the vertical, and 15 cm/sec (2 sigma) in velocity" (Feairheller & Clark, 2006), whereas the specification of military applications is not published. However the actual system performance is better than the specification when the full constellation is available. The modernisation program should enable GLONASS to achieve a similar positioning accuracy to GPS.

### 4.1.3. Galileo

Galileo is the European satellite navigation system initiated by the European Union (EU) and the European Space Agency (ESA). Unlike GPS and GLONASS, Galileo is a purely civil navigation system, which is funded by the EU and will be managed by GNSS Supervisory Authority (GSA). The System is currently in Testbed v2 phase. There are two experimental satellites, GIOVE-A and GIOVE-B, in orbits. The in-orbit validation for the system will soon start with the plan to launch another 4 operational satellites and test ground infrastructure. The full operational capacity is expected to be achieved by 2014.

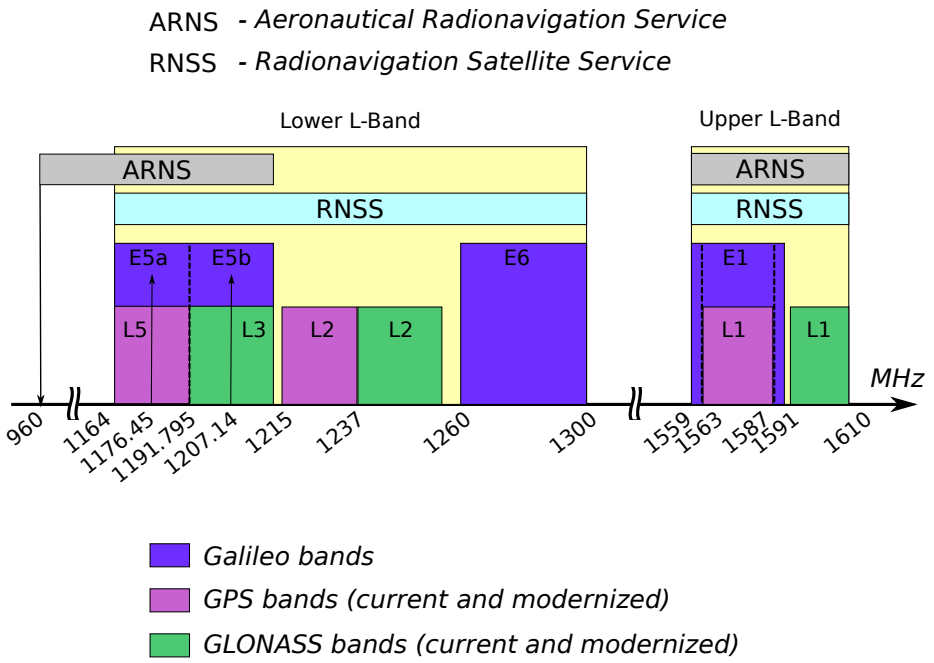



**Figure 4.1.:** Frequency distribution for GPS, GLONASS and Galileo

The Galileo system is designed to offer four services: an open service (OS), a safety-of-life (SoL) service, commercial services and a public regulated service (PRS). The OS and SoL service are for public civil applications and use same signals, but integrity data is added to SoL service. Commercial cervices with improved accuracy will be provided to users willing to pay subscription charges through additional encrypted signal. The PRS is designed to provide high integrity and continuity, interference resistant and encrypted positioning and navigation to government-designated users.

The planned Galileo constellation will consist of 30 satellites. The orbital radius of Galileo satellites is 29, 600 km with a period of 14 hours, 5 minutes. There will be three orbital plane for satellite deployment. Each orbital plane is separated by 120° in longitude and inclined at 56° to the equator (Groves, 2008).

Galileo broadcast signals are located across three frequency bands: E5, E6 and E2-L1-E1. The E5 band is 51.15 MHz wide and centred at 1, 191.795 MHz. There are two codes modulated onto a single E5 carrier, known as E5a and E5b. While the E5a signal supports the OS, the E5b signal will be available for the OS, commercial services, and SoL services. Two codes, E6C and E6P, modulated on the E6 band (40.92 MHz wide, centred at 1, 278.75 MHz) are both encrypted for commercial services and PRS usages. Meanwhile, E2-L1-E1 band, which is also 24.552 MHz wide and is centred at 1, 575.42 MHz, will transmit two signals: L1F for the OS, commercial services and the SoL services; and L1P for the PRS. A summary of the Galileo signals is shown in Figure 4.1 in comparison with both GPS and GLONASS signals. A more detailed description of Galileo signals can be found in Falcone *et al.* (2006) and European Space Agency (2010).

The ground segment of Galileo system is currently under active construction. The completed ground segment will be composed of a global network of sensor stations and TT&C stations, a global real-time oriented network of uplink stations, geographically redundant control centres and a high-performance communication network. The Current plan for the ground segment architecture will contain 34 sensor stations, 3 control centres, 9 Mission Uplink stations and 5 TT&C stations (Lisi, 2009).

### 4.1.4. GNSS Interoperability

GNSS interoperability issue is mainly inspired by the potential benefits of using multiple GNSS constellations for positioning. In general, there are two fundamental elements in building a GNSS: a reference frame and a time scale. The reference frame is used by a GNSS to indicate the positions of satellites and users, whereas the time scale is used to define the offsets of all clocks in the system. All current GNSSs define their own reference frames and time scales. Consequently, the prerequisite for GPS, GLONASS and Galileo to be interoperable is the development of transformations among different reference frames and time scales.

**Table 4.4.:** A summary the reference frames and time scales of current GNSSs

| GNSSs | Reference Frame | Time Scale Realisation |
|---|---|---|
| GPS | WGS84 | GPS System Time (GPST) steered to UTC(USNO) |
| GLONASS | PZ-90 (previously known as SGS85) | GLONASS System Time steered to UTC(SU) |
| Galileo | Galileo Terrestrial Reference Frame (GTRF) | Galileo System Time (GST) steered to a set of EU UTC |

A summary of reference frames and time scales for GPS, GLONASS and Galileo are listed in Table 4.4. Note that both GTRF and WGS84 are an independent realisation of the International Terrestrial Reference System (ITRS). The transformation between PZ-90 and WGS84 are estimated empirically. Both GPS time and GLONASS time are defined using different realisation of the Coordinated Universal Time (UTC). GPS+GLONASS receiver usually sacrifices a pseudorange measurement in order to estimate the bias between time scales of the two GNSSs (Misra & Enge, 2006). The transformation between Galileo and other systems are currently under study.

# 4.2. PVT Estimation with GPS

## 4.2.1. Receiver Position and Clock Bias Estimation

### 4.2.1.1. Pseudorange Measurements

Pseudorange is the apparent range between the satellite position at the signal transit time and the receiver position at the signal receive time. Two time epochs involved are signal transit time based on the satellite clock and the signal receive time corresponding to the receiver clock. Because both the satellite clock and the receiver clock are biased with respect to the GPS system time, the difference between the two epochs, i. e. the signal propagation time, is equal to the difference between time offsets of the two clocks with respect to the system time. Each GPS satellite transmits different PRN codes as part of the broadcast signal, as described in Section 4.1.1.2. The PRN code is generated according to the individual satellite clock. When the transmitted PRN code reaches the receiver, a replica of the code is generated according to the receiver clock. The amount of time shift required to align the code with its replica (known as the correlation) is measured as the signal propagation time. A range measurement, that is, the pseudorange, can be obtained by multiplying the time difference with the speed of light.

The estimation of position using pseudorange measurements is adopted by most single-frequency receivers for navigation applications. A pseudorange can be modelled by the following equation from Misra & Enge (2006)

$$\rho^{(k)}(t) = r^{(k)}(t, t-\tau) + c\left[\delta t_u(t) - \delta t^{(k)}(t-\tau)\right] + I_\rho^{(k)}(t) + T_\rho^{(k)}(t) + \varepsilon_\rho^{(k)}(t) \tag{4.1}$$

where:

| | |
|---|---|
| $\rho^{(k)}(t)$ | is the pseudorange measurement for the $k$th satellite, in meters. |
| $t$ | is the signal reception time in GPS time, in seconds. |
| $\tau$ | is the signal transit time, in seconds. |
| $r^{(k)}(t, t-\tau)$ | is the true geometric range between the $k$th satellite at the signal emission time $t-\tau$ and the receiver at the signal reception time $t$, in meters. |
| $c$ | is the speed of light in a vacuum, in meters/second. |
| $\delta t_u(t)$ | is the receiver clock bias relative to GPS time at the signal reception time $t$, in seconds. |
| $\delta t^{(k)}(t-\tau)$ | is the $k$th satellite clock bias relative to GPS time at the signal emission time $t-\tau$, in seconds. |
| $I_\rho^{(k)}(t)$ | is the propagation delay on the pseudorange measurement of the $k$th satellite caused by the Earth's ionosphere, in meters. |
| $T_\rho^{(k)}(t)$ | is the propagation delay on the pseudorange measurement of the $k$th satellite caused by the Earth's troposphere, in meters. |
| $\varepsilon_\rho^{(k)}(t)$ | is the miscellaneous unmodeled range error for the individual satellite. The unmodeled error usually consists of receiver noise and multipath, as well as the error in satellite position. |

As a part of the ranging signal, GPS satellites broadcast navigation data messages which enable the receiver to estimate the satellite clock bias $\delta t^{(k)}(t-\tau)$, ionospheric delay $I_\rho^{(k)}(t)$, and tropospheric

delay $T_\rho^{(k)}(t)$ for the individual satellite. Subsequently, these terms can be removed from (4.1), and the equation for a corrected pseudorange measurement $\tilde{\rho}^{(k)}$ can be derived

$$\tilde{\rho}^{(k)} = r^{(k)} + c\delta t_u + \tilde{\varepsilon}_\rho^{(k)} \tag{4.2}$$

Note that the $\tilde{\varepsilon}_\rho^{(k)}$ now consists of residuals after applying corrections from navigation data and other unmodeled errors.

By expressing the position of the $k$th satellite at the signal emission time in Cartesian coordinates, $\mathbf{p}^{(k)} = \left[x^{(k)},\, y^{(k)},\, z^{(k)}\right]^T$ , and the receiver position at the signal reception time as $\mathbf{p} = [x,\, y,\, z]^T$, the true geometric range $r^{(k)}$ can be written as following

$$r^{(k)} = \sqrt{(x^{(k)} - x)^2 + (y^{(k)} - y)^2 + (z^{(k)} - z)^2} = \parallel \mathbf{p}^{(k)} - \mathbf{p} \parallel \tag{4.3}$$

The bold, lower-case characters are used in this thesis to denote vectors and bold, upper-case ones are used to denote matrices; $\parallel \cdots \parallel$ represents the magnitude of a vector. The Cartesian coordinates are expressed in *Earth-centred Earth-fixed* (ECEF) frame, or Earth frame. The conventional superscript notation for vectors in Earth frame, $\mathbf{p}^e$, is dropped in this chapter for the simplicity unless specifically pointed out. Whereas vectors in other reference frames will be denoted as the convention. Bringing (4.3) into (4.2), we can rewrite $\tilde{\rho}^{(k)}$ as

$$\tilde{\rho}^{(k)} = \parallel \mathbf{p}^{(k)} - \mathbf{p} \parallel + b_{rc} + \tilde{\varepsilon}_\rho^{(k)} \tag{4.4}$$

where the receiver clock bias $c\delta t_u$ term has been replaced with a simpler term $b_{rc}$ in the unit of meters.

#### 4.2.1.2. Single-Point Position Solution with Pseudoranges

There are four unknowns in (4.4): three components of $\mathbf{p}$, and $b_{rc}$. Assuming there are $K$ satellites available ($K > 4$), the standard approach to estimate these four unknowns is through a least-squares iteration processes starting with an initial value of the receiver position $\mathbf{p}_0$ and clock bias $b_0$.

A predicted pseudorange $\rho_0^{(k)}$ is obtained based on the initial value as

$$\rho_0^{(k)} = \parallel \mathbf{p}^{(k)} - \mathbf{p}_0 \parallel + b_0 \tag{4.5}$$

The true receiver position $\mathbf{p}$ and clock bias $b_{rc}$ can be expressed using the initial value as $\mathbf{p} = \mathbf{p}_0 + \delta\mathbf{p}$ and $b_{rc} = b_0 + \delta b$, where $\delta\mathbf{p}$ and $\delta b$ are corrections to be applied to the initial value. Thus the residual between the predicted and the measurement $\delta\rho^{(k)}$ is

$$\begin{aligned} \delta\rho^{(k)} &= \tilde{\rho}^{(k)} - \rho_0^{(k)} \\ &= \parallel \mathbf{p}^{(k)} - \mathbf{p}_0 - \delta\mathbf{p} \parallel - \parallel \mathbf{p}^{(k)} - \mathbf{p}_0 \parallel + (b_{rc} - b_0) + \tilde{\varepsilon}_\rho^{(k)} \end{aligned} \tag{4.6}$$

Equation (4.6) must be linearised by performing a Taylor series approximation in order to form a least-squares equation. The residual after linearisation becomes

$$\begin{aligned} \delta\rho^{(k)} &\approx -\frac{(\mathbf{p}^{(k)} - \mathbf{p}_0)}{\parallel \mathbf{p}^{(k)} - \mathbf{p}_0 \parallel} \cdot \delta\mathbf{p} + \delta b + \tilde{\varepsilon}_\rho^{(k)} \\ &= -\mathbf{1}^{(k)} \cdot \delta\mathbf{p} + \delta b + \tilde{\varepsilon}_\rho^{(k)} \end{aligned} \tag{4.7}$$

$\mathbf{1}^{(k)}$ in (4.7) is the estimated line-of-sight vector from the estimated receiver position to the $k$th

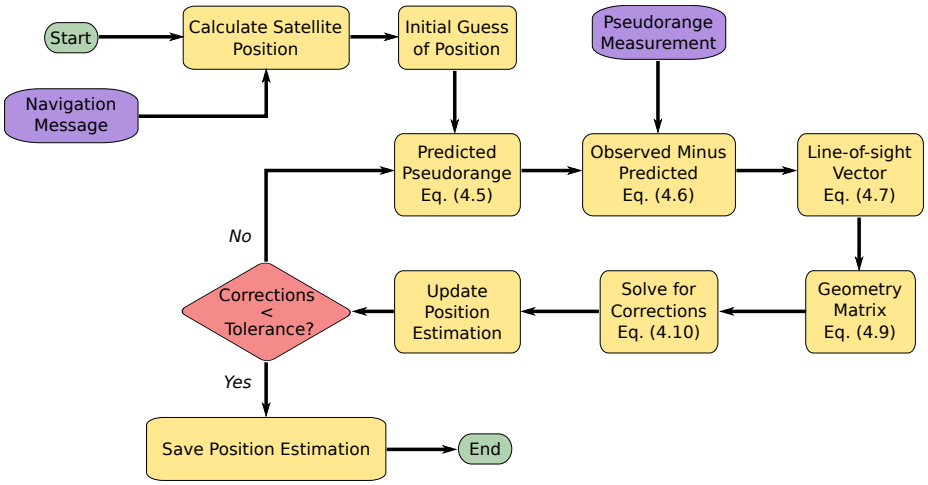



**Figure 4.2.:** Illustration of single-point position solution

satellite.

When $K$ satellites are tracked, a set of (4.7)s can be arranged in matrix form, as

$$\mathbf{z}_\rho = \mathbf{G} \begin{bmatrix} \delta\mathbf{p} \\ \delta b \end{bmatrix} + \mathbf{w}_\rho \tag{4.8}$$

where the measurement vector, $\mathbf{z}_\rho = \left[\delta\rho^{(1)}, \delta\rho^{(2)}, \cdots, \delta\rho^{(K)}\right]^T$, the measurement matrix, $\mathbf{G}$, or the user-satellite geometry matrix, is

$$\mathbf{G} = \begin{bmatrix} (-\mathbf{1}^{(1)})^T & 1 \\ (-\mathbf{1}^{(2)})^T & 1 \\ \vdots & \vdots \\ (-\mathbf{1}^{(K)})^T & 1 \end{bmatrix} \tag{4.9}$$

and the measurement noise , $\mathbf{w}_\rho = \left[\tilde{\varepsilon}_\rho^{(1)}, \tilde{\varepsilon}_\rho^{(2)}, \cdots, \tilde{\varepsilon}_\rho^{(K)}\right]^T$.

The least-squares corrections to the initial value can be computed with

$$\begin{bmatrix} \delta\mathbf{p} \\ \delta b \end{bmatrix} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{z}_\rho \tag{4.10}$$

This process is repeated until the corrections are towards zero using the corrected estimation as the initial value at each iteration. Steps for the process are summarised in Figure 4.2. Because pseudorange measurements do not usually all have the same quality, a weight matrix is often used with (4.10) during the estimation. One straight forward way to construct the weight matrix may be based on satellite elevation angle. More detailed discussion on weighting measurements can be found in Misra & Enge (2006) and Conley *et al.* (2006).

### 4.2.2. Receiver Velocity and Clock Drift Estimation

The relative motion between a satellite and the receiver causes changes on the observed frequency of satellite signals, known as Doppler shift. The Doppler shift measurement is taken from the carrier

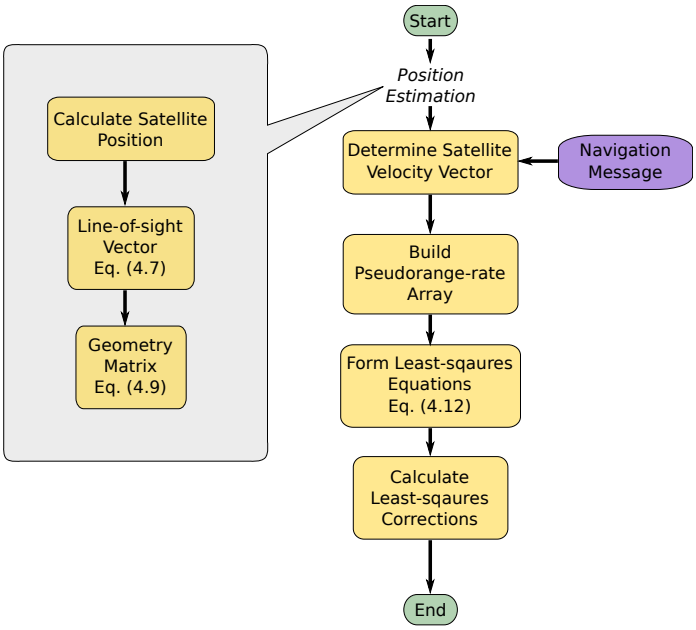



**Figure 4.3.:** Illustration of single-point velocity solution

tacking loop of the receiver and transformed into a pseudorange rate measurement. The pseudorange rate can be view as a projection of the relative velocity vector on the satellite to receiver line-of-sight direction. The receiver velocity and clock drift can then be obtained using pseudorange rates through a least-squares process similar to the single point position solution.

The model for the pseudorange rate can be derived by differentiating (4.2) and is written here as

$$\tilde{\dot{\rho}}^{(k)} = (\mathbf{v}^{(k)} - \mathbf{v}) \cdot \mathbf{1}^{(k)} + \dot{b}_{rc} + \varepsilon_{\dot{\rho}}^{(k)} \tag{4.11}$$

where:

$\tilde{\dot{\rho}}^{(k)}$ is the pseudorange rate of the $k$th satellite, in meters/seconds.

$\mathbf{v}^{(k)}$ the velocity vector of the $k$th satellite, normally obtained from the navigation data, in meters/seconds.

$\mathbf{v}$ is the velocity vector of the receiver, in meters/seconds.

$\mathbf{1}^{(k)}$ is the line-of-sight vector from the estimated receiver position to the $k$th satellite, as defined in Section 4.2.1.2.

$\dot{b}_{rc}$ is the receiver clock drift, i. e. the change rate of the receiver clock bias, in meters/seconds.

$\varepsilon_{\dot{\rho}}^{(k)}$ is the combined term for unknown pseudorange rate errors, in meters/seconds.

Note that all vectors are expressed in the Earth frame. Four unknowns can be seen in (4.11): three components of $\mathbf{v}$ and $\dot{b}_{rc}$. (4.11) is linear for the unknowns, so the least-squares equation is obtained by rearranging the equation

$$(\tilde{\dot{\rho}}^{(k)} - \mathbf{v}^{(k)} \cdot \mathbf{1}^{(k)}) = -\mathbf{1}^{(k)} \cdot \mathbf{v} + \dot{b}_{rc} + \varepsilon_{\dot{\rho}}^{(k)} \tag{4.12}$$

when more than four satellites are tracked, (4.12)s for each satellite are stacked in matrix form

$$\mathbf{z}_{\dot{\rho}} = \mathbf{G} \begin{bmatrix} \mathbf{v} \\ \dot{b}_{rc} \end{bmatrix} + \mathbf{w}_{\dot{\rho}} \tag{4.13}$$

Measurement vector $\mathbf{z}_{\dot{\rho}}$ represents a set of $(\tilde{\dot{\rho}}^{(k)} - \mathbf{v}^{(k)} \cdot \mathbf{1}^{(k)})$ for every tracked satellites, measurement matrix $\mathbf{G}$ is the same as defined in (4.9), and $\mathbf{w}_{\dot{\rho}} = \left[\varepsilon_{\dot{\rho}}^{(1)}, \varepsilon_{\dot{\rho}}^{(2)}, \cdots, \varepsilon_{\dot{\rho}}^{(K)}\right]^T$. A least-squares solution can then be calculated as the single point position solution in (4.10). Steps for the process are summarised in Figure 4.3

### 4.2.3. Impact of Visible Satellite Constellation Geometry

The user-satellites geometry of the visible satellite constellation has a substantial impact on the positioning accuracy of GPS. Under the same circumstances, a good user-satellite geometry leads to a higher accuracy positioning solution. The effect of the user-satellite geometry is firstly demonstrated with a classical 2D ranging example shown in Figure 4.4. Each graph in Figure 4.4 contains two radio stations sending ranging signals in analogy to GPS satellites. Ideally, two circles can be drawn from error-free ranging measurements, and one of their intersection points is the receiver position, as shown by the solid lines in all graphs. However, the inevitable errors increase uncertainties in measurements resulting an uncertainty belt on each side of the circle, seen as regions circled by dashed lines. The intersection of the uncertainty regions for each measurement is the uncertainty of the positioning solution, i.e. the grey areas in each graph.

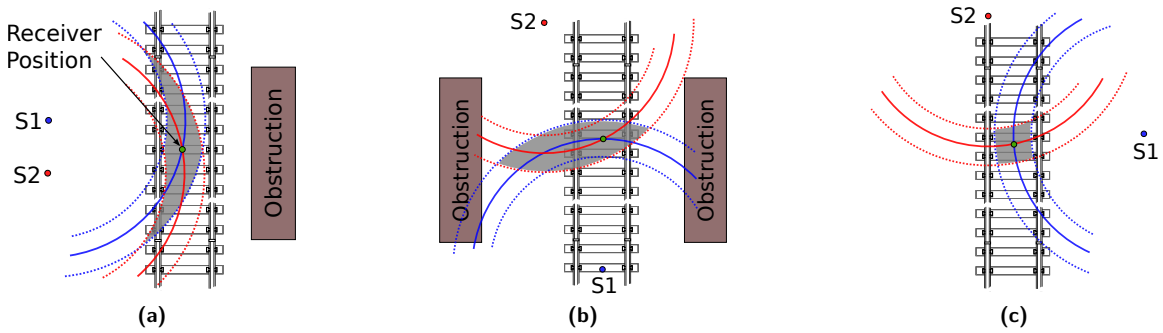



**Figure 4.4.:** A 2D ranging example on the railways

Figure 4.4a shows the situation where both stations are on the same side of the receiver. This resembles real scenes on the railways where half of the sky is blocked by physical obstructions. The along-track uncertainty is increased under such scenes, which can be translated into an increased uncertainty on train separation determination. On the other hand, Figure 4.4b shows the situation where both stations are on a line to the receiver. This resembles railway scenes where the "urban canyon" only reveals a line of the sky, as shown in Figure 2.11b in Section 2.3.1.1 when the train is approaching the station. In this case, a larger cross-track uncertainty can be demonstrated, presenting problems for reliable rail track identification. Only an evenly distributed set-up of stations as in Figure 4.4c ensures minimum uncertainties on both along-track and cross-track directions.

The metric used to describe the impact of user-satellites geometry is called dilution of precision

(DOP). The DOP matrix can be computed from the user-satellite geometry matrix, $\mathbf{G}$,

$$(\mathbf{G}^T\mathbf{G})^{-1} = \begin{bmatrix} D_{11} & - & - & - \\ - & D_{22} & - & - \\ - & - & D_{33} & - \\ - & - & - & D_{44} \end{bmatrix} \tag{4.14}$$

where the most common metric geometric dilution of precision (GDOP) is obtained from DOP matrix elements using

$$\text{GDOP} = \sqrt{D_{11} + D_{22} + D_{33} + D_{44}} \tag{4.15}$$

The covariance matrix of the positioning estimations is a function of both the user pseudorange error (UERE), $\sigma_{UERE}$, and the DOP matrix

$$\text{cov}\begin{bmatrix} \hat{\mathbf{p}} \\ \hat{b} \end{bmatrix} = (\mathbf{G}^T\mathbf{G})^{-1}\sigma^2_{UERE} \tag{4.16}$$

As can be seen from (4.16), DOP matrix interprets the error on ranging measurements geometrically into the estimation error. Note that DOP matrix in (4.14) is expressed in the Earth frame. DOP representations in other reference frames can be calculated from (4.14) using frame transformation matrices.

### 4.2.4. PVT Estimation with an Extended Kalman Filter

The least-squares based PVT estimation described in Section 4.2.1 and 4.2.2 only utilises pseudorange and pseudorange-rate measurements at a single epoch and recursively solves for the solution with an initial estimation for that instant. However, previous solutions also contain useful information which is helpful to obtain a more accurate estimation for the current epoch. Therefore most GPS receivers use a Kalman filter for PVT estimation.

A Kalman filter, described in Section 3.2, can perform an optimum weighting of current ranging measurements against previous estimations. In order to achieve this, a receiver dynamics model is maintained along with the weighting process. This can be achieved by either taking a simplified assumption of the receiver dynamics or integrating dynamics information from other sensors. The first approach is described in this section, assuming a low-dynamics application, such as on a train. The latter approach is described in Section 7.2 as the part of the integrated navigation process. Since the receiver dynamics are not linear, an extended Kalman filter (EKF) is usually used based a linearisation of the receiver dynamics (Crassidis & Junkins, 2004).

#### 4.2.4.1. State Selection

GPS navigation solutions form the basic part of a Kalman filter state vector. The reference frame chosen for the system dynamics described in this section is the local-navigation frame. This means the position estimates are in the form of latitude, $L$, longitude, $\lambda$ and height, $h$. For a low-dynamics application, the velocity, $\mathbf{v}^n$, must be included in the sate vector to accommodate the movement of the receiver, whereas accelerations can be modelled as system noise on the velocity estimation. The receiver clock bias, $b_{rc}$ and drift, $\dot{b}_{rc}$ are always included.

The inclusion of other correlated ranging errors, caused by the combination the ephemeris and the residual satellites clock, ionospheric and tropospheric error, are optional because of their observability issues when the visible number of satellites are low. However, estimating them is beneficial for improving the accuracy, provided a good number of satellites are in view.

Therefore eight elements are included in the state vector for the system described in this section: components of position and velocity in the local-navigation frame, receiver clock bias and drift. The state vector, $\mathbf{x}^n_{GPS}$, is written as

$$\mathbf{x}^n_{GPS} = \left[L,\ \lambda,\ h,\ \mathbf{v}^n,\ b_{rc},\ \dot{b}_{rc}\right]^T \tag{4.17}$$

#### 4.2.4.2. System Model

The system model describes the propagation of the estimated states and their uncertainties in order to account for the receiver dynamics and the clock behaviours between successive ranging measurements update. As the system model is expressed in the local-navigation frame, curvilinear position is used, and the time derivatives of the position is given by (5.22), (5.23) and (5.24). Also the relationship between the receiver clock bias and drift is

$$\frac{\partial}{\partial t} b_{rc} = \dot{b}_{rc} \tag{4.18}$$

Although (5.22) and (5.23) do not strictly fit the linearity assumption of a Kalman filter, the denominators can be treated as constant over the propagation interval because of their slow variation. The system matrix is therefore written as (Groves, 2008)

$$\mathbf{F}^n_{GPS} \approx \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{F}^n_{12} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & 1 \\ \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & 0 \end{bmatrix} \tag{4.19}$$

$$\mathbf{F}^n_{12} = \begin{bmatrix} \frac{1}{R_N(\hat{L})+\hat{h}} & 0 & 0 \\ 0 & \frac{1}{(R_N(\hat{L})+\hat{h})\cos\hat{L}} & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{4.20}$$

Note that Kalman filter matrices here are expressed in terms of sub-matrices corresponding to components of the state vector. Symbols with a hat on top, such as $\hat{L}$, are used to denote estimations. The transition matrix for a discrete system is given as the first order power-series approximation

$$\mathbf{\Phi}_k = \mathbf{I}_{8\times8} + \mathbf{F}_k \tau_s \tag{4.21}$$

where $\tau_s$ is the state propagation interval.

The increase of state uncertainties is mainly due to user motion and the receiver clock drift. This

increase process is described by the system noise covariance matrix

$$\mathbf{Q}^n_{GPS} = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{Q}^n_{22} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & 0 \\ \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & n^2_{rc}\tau_s \end{bmatrix} \tag{4.22}$$

$$\mathbf{Q}^n_{22} = \begin{bmatrix} n^2_N & 0 & 0 \\ 0 & n^2_E & 0 \\ 0 & 0 & n^2_D \end{bmatrix} \tau_s \tag{4.23}$$

where $n^2_N$, $n^2_E$, and $n^2_D$ are respectively the acceleration PSD in north, east and down directions, and $n^2_{rc}$ is the receiver clock drift PSD. For low-dynamics applications, $n^2_N = n^2_E = 10\,\mathrm{m}^2/\mathrm{s}^3$ is reasonable (Groves, 2008) and the clock drift PSD can vary up to $0.04\,\mathrm{m}^2/\mathrm{s}^3$ (Axelrad & Brown, 1996).

#### 4.2.4.3. Measurement Model

The measurement model is used by a Kalman filter to update the state using ranging measurements. The measurement vector contains pseudorange and pseudorange-rate measurements from the ranging processor. Thus, for $K$ satellites tracked,

$$\mathbf{z}_{GPS} = \left[\tilde{\rho}^{(1)},\, \tilde{\rho}^{(2)},\, \cdots,\, \tilde{\rho}^{(K)},|\, \tilde{\dot{\rho}}^{(1)},\, \tilde{\dot{\rho}}^{(2)},\, \cdots,\, \tilde{\dot{\rho}}^{(K)}\right]^T \tag{4.24}$$

where the tilde on top denotes a measurement. The estimated states at each update epoch, $\hat{\mathbf{x}}^-_k$, are used to produce a predicted measurement vector and is written as

$$\mathbf{h}_{GPS}(\hat{\mathbf{x}}^-_k) = \left[\hat{\rho}^{(1)},\, \hat{\rho}^{(2)},\, \cdots,\, \hat{\rho}^{(K)},|\, \hat{\dot{\rho}}^{(1)},\, \hat{\dot{\rho}}^{(2)},\, \cdots,\, \hat{\dot{\rho}}^{(K)}\right]^T \tag{4.25}$$

The predicted pseudorange, $\hat{\rho}^{(k)}$, and pseudorange-rate, $\hat{\dot{\rho}}^{(k)}$, can be calculated using

$$\hat{\rho}^{(k)} = \parallel \hat{\mathbf{p}}^{(k)} - \hat{\mathbf{p}}^- \parallel + \hat{b}^-_{rc} \tag{4.26}$$

$$\hat{\dot{\rho}}^{(k)} = (\hat{\mathbf{v}}^{(k)} - \hat{\mathbf{v}}^-) \cdot \hat{\mathbf{1}}^{(k)} + \hat{\dot{b}}^-_{rc} \tag{4.27}$$

Note that all vectors in (4.26) and (4.27) are expressed in the Earth frame. The estimated receiver position $\hat{\mathbf{p}}^-$ and velocity $\hat{\mathbf{v}}^-$ for local-navigation frame implementation are first transformed into the Earth frame before using (4.26) and (4.27). The measurement innovation vector is therefore given by

$$\delta\mathbf{z}_{GPS} = \mathbf{z}_{GPS} - \mathbf{h}_{GPS}(\hat{\mathbf{x}}^-_k) \tag{4.28}$$

The estimated line-of-sight vector in local-navigation frame consists of three components, $\hat{\mathbf{1}}^{(k),n} =$

$\left[u_N^{(k),n},\, u_E^{(k),n},\, u_D^{(k),n}\right]^T$. Hence, following Groves (2008), the measurement matrix is

$$\mathbf{H}_{GPS}^n \approx \left[\begin{array}{cccccccc} h_L u_N^{(1),n} & h_\lambda u_E^{(1),n} & u_D^{(1),n} & 0 & 0 & 0 & 1 & 0 \\ h_L u_N^{(2),n} & h_\lambda u_E^{(2),n} & u_D^{(2),n} & 0 & 0 & 0 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ h_L u_N^{(K),n} & h_\lambda u_E^{(K),n} & u_D^{(K),n} & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 0 & -u_N^{(1),n} & -u_E^{(1),n} & -u_D^{(1),n} & 0 & 1 \\ 0 & 0 & 0 & -u_N^{(2),n} & -u_E^{(2),n} & -u_D^{(2),n} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -u_N^{(K),n} & -u_E^{(K),n} & -u_D^{(K),n} & 0 & 1 \end{array}\right] \tag{4.29}$$

where

$$h_L = -\left[R_N(\hat{L}) + \hat{h}\right] \tag{4.30}$$

$$h_\lambda = -(R_N(\hat{L}) + \hat{h})\cos\hat{L} \tag{4.31}$$

Noise-like errors on pseudorange and pseudorange-rate measurements are modelled in the measurement noise covariance matrix, $\mathbf{R}_{GPS}$. $\mathbf{R}_{GPS}$ is formed as a diagonal matrix assuming no correlations between pseudorange and pseudorange-rate measurements during the simulation process in this thesis.

### 4.2.5. Receiver Oscillators

Most GPS receivers use a reference oscillator to control the timing. The reference oscillator provides a frequency standard for oscillators used in the receivers front-end and baseband processor and the receiver clock. The long-term errors and drift in the receiver frequency standard are estimated as part of the positioning solution and are compensated using the positioning solution. Therefore, GPS receivers do not usually require a high performance reference oscillator as those atomic clocks on board GPS satellites. Instead, various quartz crystal oscillators (XO) are often used in GPS receivers. However, the new development of portable high performance oscillators, such as chip-scale atomic clocks, is still highly appreciated especially for receivers that have to work under challenging signal environments (Misra & Enge, 2006).

As the dominant error source for a XO is the variation of frequency standard caused by temperature changing, high performance XOs commonly come with a temperature compensating mechanism. A temperature-compensated crystal oscillator (TCXO) adopts a temperature sensor to control the oscillator voltage according to the temperature variation and can stabilise "the frequency to within one part in $10^8$ over a one-second interval" (Groves, 2008). The cost of a TCXO is tens of dollars or Euros. A more expensive oven-controlled crystal oscillator (OCXO), usually costs a thousand dollars or Euros, is able to maintain the "a frequency variation of about one part in $10^{11}$ over a second, with a frequency bias of one part in $10^8$" (Groves, 2008).

With a basic XO, the range-rate bias due to the receiver clock drift can be up to $3,000\,\mathrm{m/s}$. The magnitude of this bias can be reduced to about $200\,\mathrm{m/s}$ by using a TXCO, while an OCXO can maintain the bias on the order of $2-3\,\mathrm{m/s}$ (Misra & Enge, 2006). More information on high

performance oscillators and GPS timekeeping can be found at Allan *et al.* (1997).

## 4.3. Error Sources for Ranging

### 4.3.1. Satellite Clock Error

The atomic clock on board a satellite controls all its timing operations including the generation of broadcast signal. Although an atomic clock is highly stable, the satellite clock error arises as a result of oscillator noise accumulation. The offset between satellite time and GPS system time can be as large as 1 ms (equivalent to a 300 km pseudorange error) as shown by the clock correction fields in the navigation message (Conley *et al.*, 2006).

The correction for this offset is broadcast through three calibration coefficients (known as clock bias, clock drift and frequency drift) together with a reference time and a relativistic correction. The GPS master control station is responsible for the determination of these parameters and the transmission of them to the satellites. The total satellite clock correction can be worked out by receivers using a second-order polynomial.

As these parameters are calculated by a curve-fit to the prediction of the actual clock errors, there is a residual error after applying the correction and this residual error increases along with the ageing of the broadcast data. The size of pseudorange error as a result of this residual error can be ranged from 0.8 m at zero age of data up to 1-4 m at 24 hours after an upload (Conley *et al.*, 2006). The rubidium clocks that are used in new satellites are more stable and will lead to these values dropping.

As the navigation data is updated every 2 hours at present (Groves, 2008), the grow of the residual error after correction is expected to be small. On the other hand, the continuous update of the constellation decreases the residual error as well.

### 4.3.2. Ephemeris Prediction Error

The broadcast ephemeris contains information on satellite orbital position. Similar to the generation process of the satellite clock correction, the orbit information is produced through a curve-fit to the orbit prediction of the satellite. The residual satellite position vector is made up of components in three directions, known as along-track, cross-track and radial direction in a satellite coordinate system (Vallado & McClain, 2001). The error in radial direction is usually the smallest as it is the most observable one to the ground monitoring network and has most effect on pseudorange measurements because it projects significantly on the line-of-sight towards the Earth.

The effective pseudorange error caused by the residual satellites position error is on the order of 0.8 m ($1\sigma$) (Conley *et al.*, 2006). This value will decrease with the employing of a larger monitoring network.

### 4.3.3. Ionosphere Propagation Error

The ionosphere is a dispersive part of the atmosphere which is about 50 km to 1000 km above the surface of the Earth. The gas in this region is ionised by the sun's radiation and is therefore full of free electrons and ions. The transmission of satellite broadcast signals are affected by these free electrons. As a result, the modulated information on signals, i. e. the PRN code and navigation data, is delayed whereas the carrier phase experiences an advance. This phenomenon is referred to as *ionospheric divergence* or *code-carrier divergence*. The consequence of this phenomenon is that

same amount of errors are added to both pseudorange and carrier-phase measurements, only with different sign.

The density of free electrons on the propagation path of a signal decides its speed. This density is defined as total electron content (TEC), known as "the number of electrons in a tube of $1\,\mathrm{m}^2$ cross section extending from the receiver to the satellite."(Misra & Enge, 2006). The ionospheric delay on pseudorange measurements at zenith position ($I_z$) can be expressed in terms of TEC value and the signal frequency $f$,

$$I_z = \frac{40.3 \cdot TEC}{f^2} \tag{4.32}$$

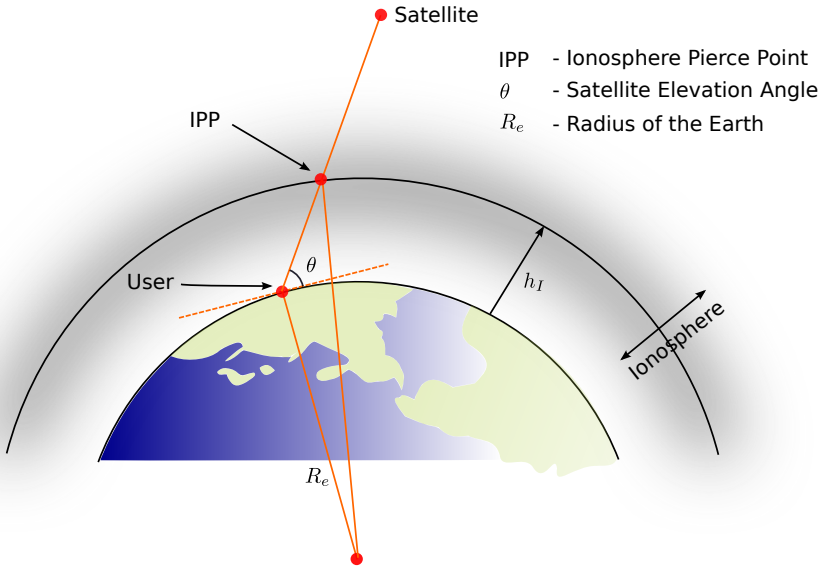



**Figure 4.5.:** Geometrical modelling of ionospheric delay

In general, the ionosphere can be thought as a thin shell surrounding the Earth. Following this geometrical modelling, the path length of a signal through the ionosphere becomes a function of the satellite elevation angle. The TEC value is smallest when the satellite is vertically above the receiver and increases as the satellite elevation angle decreases because of the increased signal path. Therefore, an obliquity factor ($OF_I$) is usually applied to relate the zenith TEC value to TEC values at other satellite positions. Typically, the obliquity factor is a function of the satellite elevation angle $\theta$. One example of the obliquity factor, from Conley *et al.* (2006), is

$$OF_I = \left[1 - \left(\frac{R_e \cos\theta}{R_e + h_I}\right)^2\right]^{-1/2} \tag{4.33}$$

where terms are defined in Figure 4.5 on this page. The general ionospheric delay on pseudorange measurements for all satellite positions, $I_\rho$, can be written as:

$$I_\rho = I_z \cdot OF_I \tag{4.34}$$

Two common methods to cope with ionospheric delay are forming ionospheric-free pseudorange measurements from a dual-frequency receiver and using the Klobuchar model broadcast in the navigation data for a single-frequency receiver. The pseudorange error caused by the residual ionospheric delay after applying Klobuchar is typically 7 m ($1\sigma$) for the average of the globe and over elevation angles (Conley *et al.*, 2006).

### 4.3.4. Troposphere Propagation Error

The troposphere is a layer of the atmosphere which is lower than the ionosphere. The height of the troposphere is up to about 50 km above the earth surface. The troposphere is nondispersive to GPS signals. Because of this, delays on both modulated signal information and the carrier are the same. The amount of tropospheric delay depends on the temperature, pressure, and relative humidity of the local atmosphere. Hence the tropospheric delay shows a low variability compared to the ionospheric delay.

The modelling of a tropospheric delay often includes both a dry (hydrostatic) and wet (nonhydrostatic) component. The dry component, which is caused by dry air of the troposphere, produces about 90% of the delay and can be predicted very accurately. Whereas the wet component presents more difficulties to be predicted as a result of the atmospheric uncertainty.

Corrections for the tropospheric delay can be obtained either from meteorological observations or using tropospheric models. While meteorological measurements are impractical for most navigation applications, tropospheric models make estimations of the zenith tropospheric delay. A mapping function, usually modelled as a function the satellite elevation angle $\theta$, is applied to account for satellite elevation angles other than 90°. As an example, the Black and Eisner's mapping function, $m(\theta)$, is given here,

$$m(\theta) = \frac{1.001}{\sqrt{0.002001 + \sin^2 \theta}} \tag{4.35}$$

The zenith tropospheric delay is usually $2.3 - 2.6$ m at sea level (Misra & Enge, 2006). The residual tropospheric delay after the correction without meteorological data is typically 0.2 m ($1\sigma$) (Conley *et al.*, 2006).

### 4.3.5. Multipath

The multipath effect refers to the reception of diffused or reflected satellite signals in addition to the direct signal. The diffused multipath is caused by groups of electrically small objects on the signal path, which reflect and diffract the signal. As the diffused multipath effect is generally uncorrelated over time and the sum of the effect is on the order of receiver noise values (Braasch, 1996), the error caused by it is usually merged into the receiver noise term in the observation equations. The reflected signal can come from multiple sources, such as the ground and buildings around the location of the receiver.

In general, there can be three scenarios for the reception of reflected signals: there can be no reflected signals at all as the receiver stands in an open field with no reflective objects in view; or there can be only reflected signals as the direct signal is blocked or attenuated which is very likely for an indoor or sheltered environment; but for most situations the received signals is always a combination of reflected signals and the direct signal. Because of the greater chance of encountering reflective objects, satellites with low elevation angles commonly suffer from the greatest multipath error.

The multipath error arises since the travel path of a reflected signal is different to the direct signal. Both the signal modulation and the carrier phase of the combined received signal are thus distorted. The code and carrier tracking accuracies of the receiver is reduced due to the existence of reflected signals. As a result, both code- and carrier- phase based positioning accuracies are degraded. The magnitude of multipath error caused by a reflected signal with given amplitude and lag depends on the signal type and the receiver. In general, the code-phase error caused by the multipath is at the meter-level, and varies greatly as the location of the receiver, whereas the carrier-phase error caused by multipath is mostly on centimetre-level (Braasch, 1996). For navigation applications, the

code-phase multipath error is therefore the main study object.

The impact of a multipath signal is usually characterised relative to the direct signal by four parameters including its relative amplitude, time delay, relative phase and the change rate of the relative phase. The mathematical modelling of multipath therefore usually comprises both the direct signal and a composite of multiple multipath signals.

The spread spectrum signalling which is used by GPS signals in its nature is helpful in multipath environment. For multipath signals that have delays longer than a chip width of the PRN code are commonly referred as long-delay multipath, and others known as short-delay multipath. Long-delay multipath signals do not cause any pseudorange errors as they are mostly suppressed by the auto-correlation properties of the PRN code (Misra & Enge, 2006). Applying a narrow correlator spacing in the delay lock loop of the receiver can reduce pseudorange errors caused by short delay multipath signals that has a larger differential path length (Van Dierendonck *et al.*, 1992). On the other hand, carrier-phase multipath errors exist for both long- and short-delay multipath, but are small compared to code-phase errors due the property of carrier phase lock loop.

Studies carried out by Counselman & Gourevitch (1981) concluded that the effects of multipath can be reduced to the centimeter-level over short baselines by averaging the signal over a period of an hour or more in a differential carrier-phase tracking system. However, for navigation applications this is usually not applicable. A receiver can readily solve multipath delay "greater than twice the spreading code period for a BPSK-R modulation" (Conley *et al.*, 2006). Other multipath signals can be mitigated through various mitigation techniques. Multipath mitigation techniques usually focus on either the receiver signal reception or the receiver processing.

Special designs or novel usages of antennas aim to reduce the multipath during the signal reception process. Specially designed multipath-limiting antennas have the ability of reducing the gain for signals from lower the horizontal level in comparison to normal patch antennas (Misra & Enge, 2006). RF absorbing ground planes can also be used with patch antennas, as most choke-ring antennas, to reduce the effect of multipath.

On the other hand, techniques for multipath mitigation during the receiver processing are divided into non-parametric and parametric processing. Non-parametric techniques mainly rely on the deployment of discriminator designs, which are less sensitive to multipath errors. Meanwhile, parametric techniques try to model multipath signals and estimate parameters for the model in order to apply multipath corrections.

A harsh environment full of physical obstructions, such as the railways, not only increases the number of reflected signals but also includes occasional shadowing and attenuation of the direct signal. Multipath errors in the satellite positioning results, especially code-phase errors, can increase greatly under such situations. Although the modelling for multipath errors in the simulated satellite ranging measurements in this thesis is not carried out at a detailed level due to the lack of real time collected data, studies for multipath modelling and mitigation under the railway environment is proposed as part of the future work for this research.

## 4.4. Standalone GPS Performance

As indicated in Section 4.2.3, the position error is a function of both pseudorange errors (UERE) and user-satellites geometry. An error budget for pseudorange errors is given in this section to serve as a guideline for position error analyses. Main error sources for ranging as discussed in Section 4.3 and their contributions into the UERE are listed according to the segment of their occurrence.

Table 4.5 shows a typical contemporary UERE budgets for a single-frequency C/A code based on data presented in Section 4.3. Note that position errors are far less than predicted using DOP due

to the high correlation of the residual ionospheric errors among satellites.

**Table 4.5.:** Typical SPS UERE Budget for GPS (Conley *et al.*, 2006)

| Segment | Error Source | $1\sigma$ Error (m) |
|---|---|---|
| Space & Control | Satellite Clock | 1.1 |
| | Broadcast Ephemeris | 0.8 |
| User | Ionospheric Delay | 7.0 |
| | Tropospheric Delay | 0.2 |
| | Multipath | 0.2 |
| System UERE | Total | 7.1 |

## 4.5. Improvements of Future GNSS from a Railway Perspective

Positioning on the railways as a safety-of-life application for GNSS spontaneously require a higher performance especially in terms of availability, continuity, integrity and accuracy. The current GPS system is unable provide a satisfactory positioning service alone for the railways as discussed in Section 2.3.2. However, the recent development of GNSS proves to be inspiring for safety-of-life applications.

As mentioned in Section 2.3.1.2, the LOS problem presents the most challenge for GNSS positioning applications on the railways because of their effects on decreasing the system availability and continuity. Modernised GNSS signals are broadcast with increased power and enhanced characteristics for multipath mitigation, which will significantly improve the signal-difficult situations on the railways. Weaker signals may therefore still be tracked by the receiver for less sheltered situations such as roofed stations. Moreover, the presence of multiple constellations will potentially increase the overall availability along a rail line. But the GNSS interoperability issue as discussed in Section 4.1.4 will need to be tackled first, plus the increase of available satellites does not necessarily guarantee an improved user-satellites geometry due to the similar orbital arrangement of most GNSS constellations. Thus for future high performance positioning services on the railways, an integrated system with the access to multiple modernised GNSS constellations is more feasible.

Integrity is another key feature for the improvements of future GNSS and is also another main concern for most safety-of-life kinematic positioning applications. Currently there is no integrity information available from GPS. Thus, most SBASs provide additional information on integrity as one of their supplemental functions. For an integrated system, integrity is usually assured through the existence of different sensors and parallelled integration processors. Built-in integrity monitoring function is also available for high performance receivers. The stress of signal-in-space integrity will be included in GPS III and modernised GLONASS. More recently designed systems such as Galileo intrinsically contain signal-in-space integrity information.

The present GNSS positioning accuracy is in general competent for positioning on the railways, and the future improvement will bring a better performance. The introduction of new satellites atomic clocks and larger monitoring network will enable an improved broadcast ephemeris. Meanwhile a better understanding of atmospheric activities through years of studies employing GPS signals provides high quality models for both ionosphere and troposphere. All these factors affect the accuracy performance of a single-frequency GNSS receiver. According to Madden (2008), the signal-in-space user range error till June 2008 is 0.92 m, which is much lower than the original

specification. On the other hand, new civil GPS signals, especially L5, will allow safety-of-life applications to take advantage of multiple frequencies positioning. Positioning with centimetre-level accuracy will no longer be too fragile for the railways.

# 5. Inertial Navigation and MEMS Sensors

The basic working principles of inertial navigation and the current development status of MEMS inertial sensors are introduced in this chapter. Section 5.1 provides a concise overview of various inertial sensors, their working principles and performance. A more detailed review on the status of MEMS inertial sensors is given in Section 5.2, with a summary of typical sensor performance at the end. Section 5.3 describes the different error characteristics of inertial sensors and gives a general error model based on described error characteristics. Some basic aspects of inertial navigation are then introduced in Section 5.4, followed by the implementation of inertial navigation equations for the local navigation frame.

Detailed discussion on the working principles of various inertial sensors can be found in Titterton & Weston (2004) and Barbour (2004). Information on high accuracy implementation of inertial navigation systems is available in Chatfield (1997).

## 5.1. An Overview of Inertial Sensors

Since the dawn of inertial sensing in the early 1920s, the development of inertial sensors has advanced far more beyond the original imagination of people. The range of applications in which inertial sensors can be used is extensive. Inertial systems have been used to navigate ships, aircraft, guided weapon and spacecraft. Some more novel applications include robotics, active suspension in racing motor cars and high precision surveying. The diversity of applications implies a broad selection of sensors, in terms of not only the underlying principles but also factors such as performance, dimensions and cost.

Conventionally, mechanical sensors used to dominate the field of guidance, navigation and control. However, the introduction of new technologies has enabled unconventional sensors based on different physical phenomena to challenge the situation. Initially, the driving force for introducing new technologies was a desire to improve sensor performance and reliability. Recently, the emphasis has been redirected to achieve equivalent performances at a lower cost.

Inertial measurements consist of specific force and angular rate. Specific force is the non-gravitational force per unit mass measured with respect to an inertial frame. Accelerometers are used to sense the specific force. Angular rate is the rate of rotation of the frame axes measured with respect to an inertial frame. Gyroscopes are used to sense the angular rate.

This section gives a concise review on the current development and future trend of inertial sensors. Practical accelerometers, in Section 5.1.1, and gyroscopes, in Section 5.1.2, are introduced separately. A summary of the sensor development and performance grades is given in Section 5.1.3. The development of sensors based on micro-machined electromechanical system (MEMS) is discussed separately in Section 5.2.

### 5.1.1. Accelerometers

Most accelerometers are designed based on the basic model of a suspended proof mass. When a specific force is applied, the displacement of the proof mass indicates the magnitude of the specific

force. Practical accelerometers used in a strapdown inertial system (see Section 5.4.2.2) use either a pendulous or a vibrating-beam design.

The force-feedback pendulous accelerometer is a typical mechanical closed-loop accelerometer. An electromagnetic device is used to produce a torque on the pendulum so that it can maintain a constant position with respect to the case. Any displacement of the pendulum, which occurs in the presence of an applied specific force, is measured by the pick-off and transformed into a specific force measurement. The torque is adjusted accordingly to produce additional electromagnetic force to offset the pendulum displacement and maintain the pick-off output at zero.

Vibrating-beam accelerometers use quartz crystal technology. A common configuration of this class of sensors uses a pair of quartz crystal beams, mounted symmetrically back-to-back. Each beam supports a proof mass pendulum on an opposite end. Both beams vibrate at the same frequency in their static status. The sensitive axis of the accelerometer is in the direction along the beams. Once a specific force is applied, because two proof masses are attached at the opposite end of each beam, one beam experiences compression whilst the other one is stretched. The result is that the two beams vibrate at different frequencies, with one decreasing and one increasing. The difference between the two frequencies is measured and transformed into a specific force measurement.

### 5.1.2. Gyroscopes

The term gyroscope is not longer limited to the original spinning-mass type of sensor, but refers to any angular-rate sensors that do not require an external reference. There are three main types of gyroscopes that are widely used at present: conventional, optical and vibratory.

Conventional gyroscopes rely on Newton's second law of dynamics regarding the conservation of angular momentum. A spinning wheel, or rotor, tends to maintain the direction of its spin axis with respect to inertial space by the virtue of its angular momentum and a reference direction is defined. The spinning mass is mounted in an instrument case in such a way that it is free to rotate about the two axes perpendicular to its spin axis. When the case rotates at a certain angular rate, the spinning mass maintains its orientation in inertial space and appears to rotate with respect to the case. A torque, which is a function of the angular rate, is applied by the mounting device to maintain the initial position of the rotor.

Optical gyroscopes use the principle of a constant light speed in a given medium to sense rotation. Given a closed-loop waveguide made up of mirrors where the light can be sent in both directions, the path length is the same for both light beams if the waveguide is not rotating. In case a rotation perpendicular to the waveguide plane is applied to the waveguide, the light path which is in the same direction of the rotation is increased from the prospective of an inertial frame, whereas the light path in the opposite direction is decreased. This is known as the Sagnac effect. The angular rate of the rotation with respect to the inertial space can therefore be measured by comparing the changes of light paths in both directions. Two most developed sensors in this class are ring laser gyroscope and interferometric fibre optic gyroscope. Generally, all the types of optical gyroscopes are suitable for various strapdown applications.

Vibratory gyroscopes contain a part, usually made from quartz or silicon, which is undergoing harmonic motion. The shape of the vibrating part may be various including a tuning fork, disk, hemisphere, etc. This results in different types of vibratory gyroscopes. The basic principle of these sensors is that the vibratory motion of the part creates an oscillatory linear velocity. An applied rotation about the axis orthogonal to this velocity introduces a Coriolis acceleration. Provided this acceleration can be detected, it can indicate the magnitude of the angular rate.

### 5.1.3. Summary

IMUs and inertial sensors can be grouped into five categories including marine, aviation, intermediate, tactical and automotive based on the performance of sensors. The highest performance inertial sensors are usually used in ships, submarines, and some spacecraft. Generally, marine-grade inertial sensors are designed to navigate independently for a long period offering navigation solutions that drift less than 1.8 km in a day. Aviation-grade sensors are commonly used in both commercial and military aircraft. US standard navigation unit 84 standard specifies a maximum horizontal position drift of $\sim 1.5$ km in the first hour of operation. An intermediate-grade sensor can be used in small aircraft and helicopters, but the performance is an order of magnitude lower than aviation-grade sensors. Tactical-grade sensors can only provide useful stand-alone navigation solutions for a few minutes and are therefore typically used in an integrated system. Guided weapons and unmanned air vehicles often are equipped with tactical-grade sensors. Lastly, applications other than navigation use inertial sensors, performances of which are likely to be too poor for inertial navigation, for various purposes. The costs of inertial sensors go up as their performance increases.

Most commercial accelerometers currently available, either mechanical or vibratory, are based on the same principles of measuring the movement of a proof mass. Table 5.1 shows a comparison of typical performances of the two types of sensors described in Section 5.1.1.

**Table 5.1.:** Typical performance of accelerometers (adapted from Titterton & Weston, 2004)

| Characteristic | Force-feedback pendulous | Vibrating quartz |
|---|---|---|
| Input range (g) | $\pm 100$ | $\pm 200$ |
| Turn-on bias (mg) | $0.1 \sim 10$ | $0.1 \sim 1$ |
| Bandwidth (Hz) | 400 | 400 |

Other types of accelerometers have also been developed or are under research. Very high precision could be achieved with a pendulous integrating gyro accelerometer, but the sensor is only suitable for a platform inertial system. Novel research on new types of accelerometer is exploring the possibility of using optical sensors (Plaza *et al.*, 2004), or even atom interferometry techniques (McGuirk *et al.*, 2002).

Currently, the very high precision rotation measurement (bias stability $10^{-4} \sim 10^{-5}\,^{\circ}$/hr) remains in the regime of the mechanical gyroscopes, and is usually available in platform inertial system for marine applications. Ring laser gyroscopes are usually the first choice for mid-range applications because of their high scale-factor stability. Other common choices include mechanical gyroscopes designed to balance performance with cost and size such as the dynamically tuned gyroscope, fibre optic gyroscope and vibratory sensors such as the hemispherical resonator gyroscope. Table 5.2 shows a comparison of typical performances of different types of gyroscopes. Note that very high performance mechanical gyroscopes are not listed here, and vibratory gyros do not include the hemispherical resonator gyroscope.

There are also a number of novel types of gyroscopes that are currently being researched, including nuclear magnetic resonance, flueric sensors, and atom interferometry techniques.

## 5.2. MEMS Sensors

The name MEMS does not refer to a particular product, but rather a method of producing miniature sensors. The chemical etching and batch processing techniques of the electronics integrated circuit

**Table 5.2.:** Typical performance of gyroscopes (adapted from Titterton & Weston, 2004)

| Characteristic | Ring laser gyro | Fibre optic gyro | Dynamically tuned gyro | Vibratory gyros |
|---|---|---|---|---|
| g-Independent bias (°/hr) | $0.001 \sim 10$ | $0.5 \sim 50$ | $0.05 \sim 10$ | $360 \sim 1800$ |
| g-Dependent bias (°/hr/g) | 0 | $< 1$ | $0.01 \sim 10$ | $30 \sim 180$ |
| Anisoelastic bias (°/hr/g$^2$) | 0 | $< 0.1$ | $0.1 \sim 0.5$ | 18 |
| Scale-factor non-linearity (%) | $5 \sim 100$ (ppm) | $0.05 \sim 0.5$ | $0.01 \sim 0.1$ | $0.2 \sim 0.3$ |
| Maximum input rate (°/s) | $> 1000$ | $> 1000$ | 1000 | $> 1000$ |
| Bandwidth (Hz) | $> 200$ | $> 100$ | 100 | 500 |
| Shock resistance | Good | Good | Moderate | $> 25,000\,\mathrm{g}$ |

industry is used by MEMS sensor technology to produce solid-state inertial sensors with quartz and silicon. In comparison with conventional inertial sensors which usually contain a large number of parts requiring high-precision manufacturing and assembling techniques, MEMS inertial sensors feature smaller size, low weight, rugged construction, low power consumption, short start-up time, and reduced production cost. As a result, MEMS sensors have become the focus of current inertial sensor development. However, their relatively poor performance is currently limiting their wider application, especially for MEMS gyros.

This section gives a review of current MEMS inertial technologies with Section 5.2.1 focusing on MEMS accelerometers and Section 5.2.2 focusing on MEMS gyroscopes. A summary of current MEMS sensor performance and future development is given in Section 5.2.3.

## 5.2.1. MEMS Accelerometer

Similar to the conventional accelerometers, the working principles of MEMS accelerometers mostly fall into two classes, depending on the way in which acceleration is sensed: either through the displacement of a proof mass, or the change in vibrating frequency of a element. Due to limitation on the size of the sensor, vibrating devices tend to have a potentially higher performance than pendulous type sensors.

### 5.2.1.1. Pendulous mass MEMS accelerometers

There are two types of pendulous mass MEMS accelerometer, known as out-of-plane (sometimes called z-plane) and in-plane pendulous devices, divided by the direction of their sensitive axis. Both two sensors are now in quantity production.

An out-of-plane sensor is similar to conventional pendulous accelerometers in its working principle, and can be designed to work in either open-loop or closed-loop mode. With a miniature pendulous proof mass, the out-of-plane sensor can measure specific force perpendicular to the plane of the chip. One major attraction of this type of sensor is the versatility of the packaging, which enables planar mounting of the sensor. Careful characterisation of the sensor is required, because the scale factor tends to decrease with increasing temperature (Barbour, 2004).

For an in-plane accelerometer, the proof masses are usually arranged in a comb finger structure. Any displacement of the masses is then measured through changes in the capacitance across the comb fingers. This class of sensor is more sensitive to specific force applied in the horizontal plane of the chip than in the orthogonal direction.

With careful design and packaging, the combined use of these two types of sensors can produce an accelerometer set on a single chip. For example, a set of accelerometers containing one out-of-plane sensor and two in-plane sensors can measure specific force along three axes.

#### 5.2.1.2. Resonant MEMS accelerometers

Resonant MEMS accelerometers rely on a vibrating beam structure to sense specific force. The sensitive axis can be configured either in-plane or out-of-plane. The specific force is sensed as a result of resonant frequency change of the beam oscillators under the inertial loading of a proof mass. Sensors fabricated with both silicon and quartz are available.

The out-of-plane resonant accelerometer is achieved by micro-machining a piezoelectric resonator formed by at least one flexure member. As the flexure member is bent under proof mass motion when the specific force is applied, the resonant frequency changes accordingly.

One possible configuration to realise an in-plane sensor is a monolithic vibrating tuning fork structure with a large silicon proof mass, which is driven electrostatically. A specific force applied in the wafer plane causes the beams to load axially and change resonant frequency.

#### 5.2.1.3. Tunnelling MEMS accelerometers

This is a technology under development which promises to offer a significant performance enhancement over the current sensors using read-out methods based on changes in capacitance. The basic idea of this technology is that a servo electrode mechanism maintains a constant gap between a tunnelling tip and a cantilevered proof mass, and hence holds constant a tunnelling current ($\sim 1\,\text{nA}$). The output signal is the change in voltage at the electrode when subjected to a specific force. Sensors of this technology can provide a better resolution, smaller size and higher bandwidth than capacitive sensors.

#### 5.2.1.4. Electrostatically levitated MEMS accelerometers

Electrostatically levitated MEMS accelerometers inherit the idea of removing constraints imposed by the elastic restraint and non-linear response of supporting mechanisms. The aim of the development of this technology is to produce a high performance accelerometer with high sensitivity, very accurate resolution and an easily adaptable response bandwidth, without modification to the structure. This new sensor contains a small sphere with 1 mm diameter and a mass of 1.2 mg suspended electrostatically. The position change of the sphere is sensed capacitively and a closed-loop electrostatic force is used to maintain its position.

### 5.2.2. MEMS Gyroscope

Almost all reported MEMS gyroscopes are non-rotating devices using the Coriolis acceleration effect on a vibrating proof mass to detect angular rotation with respect to the inertial space. The underlying working principle is the same as vibratory gyroscopes described in Section 5.1.2. The main difference between various sensors lies in the shape of the vibrating mass used. Whilst there can be various shapes of oscillator in the gyroscope, the recent focus of developments has been based upon tuning fork gyroscopes and resonant ring gyroscopes.

#### 5.2.2.1. Tuning fork MEMS gyroscopes

Sensors in this class usually have a balanced oscillator part with two identical proof masses working as a pair in order to overcome the sensitivity may accompany with only one proof mass. These gyroscopes are often fabricated out of a single piece of silicon or quartz.

One possible configuration of the silicon based tuning fork gyroscope consists of two silicon proof mass plates suspended over a glass substrate using folded beams. The masses are made to vibrate in-plane but are 180° out of phase. The out-plane motion caused by Coriolis force is detected by changes in capacitance between the proof mass and the substrata.

There are also a number of tuning fork MEMS gyroscopes using quartz as the base material for the sensing element. The use of piezoelectric quartz material simplifies the sensing element, leading to a reliable and durable sensor that is stable over both temperature and time. A pair of coupled H-shaped tuning forks is used in the sensor.

#### 5.2.2.2. Resonant ring MEMS gyroscopes

A resonant ring MEMS gyroscope is able to detect angular rate applied about the axis perpendicular to its ring plane. The advantage of using a ring resonant part is that the structure can maintain the drive and sense vibrational energy in one plane. However, such devices do suffer from the drawback of having a relatively low vibrating mass, and hence exhibit a low scale factor. More recently, researches have explored the possibilities for detecting angular rates applied about three mutually orthogonal axes using the Coriolis coupling between in-plane and out-of-plane displacements with sensor of this class.

### 5.2.3. Summary

MEMS sensors are the most active developing area of inertial sensors, progressing substantially and rapidly in many respects. New manufacturing approaches have proved successful and subsequently leading to improvement in performance. A summary of the typical performance of tactical-grade MEMS inertial sensors is given in Table 5.3, based on information from Honeywell Aerospace (2007), Atlantic Inertial Systems (2008), Rockwell Collins (2008) and Mirobotics Inc. (2009).

**Table 5.3.:** Typical performance of MEMS inertial sensors

| Characteristic | MEMS accelerometer | MEMS gyro |
|---|---|---|
| Turn-on bias | $5 \sim 10\,\mathrm{mg}$ | $20 \sim 100\,^\circ/\mathrm{hr}$ |
| In-run bias | $0.2 \sim 0.5\,\mathrm{mg}$ | $3 \sim 30\,^\circ/\mathrm{hr}$ |
| Scale factor (ppm) | $300 \sim 1500$ | $350 \sim 500$ |
| Operating range | $5 \sim 100\,\mathrm{g}$ | $\pm 1000\,^\circ/\mathrm{s}$ |
| Random Walk | $150\,\mu\mathrm{g}/\sqrt{\mathrm{hr}}$ | $0.3 \sim 0.5\,^\circ/\sqrt{\mathrm{hr}}$ |

Novel technologies that are currently under active research include micro-optical electro-mechanical systems. The new systems could provide a true solid-state sensor by replacing the current capacitive pick-offs with optical readings.

## 5.3. Error Modelling for Inertial Sensors

The systematic error exhibited by all inertial sensors contains four components including a fixed contribution, a temperature-dependent variation, a run-to-run variation, and an in-run variation. Both the fixed contribution and the temperature-dependent variation are calibrated by the IMU processor using laboratory calibration data. This is usually done before the sensors have been put into operation. The run-to-run variation is a constant contribution to the systematic error within the scope of a single operation, but it varies each time the sensor is used. The IMU processor cannot calibrate this type of error, and it is usually corrected by the INS alignment and/or the integration algorithm of an integrated system. The in-run variation is the error source that changes during the operation. It cannot be calibrated by the IMU or alignment and is usually difficult for an integration system to correct.

For discussion on the error modelling of inertial sensors in this section, the fixed contribution and the temperature-dependent variation are neglected as they are considered as compensated for in an operational inertial system. It is the post-calibration performance of the inertial sensors that is the most relevant to the design and performance of an integrated system, but it should be borne in mind that all calibration processes have residual errors left.

Common systematic errors that affect all inertial sensors include biases, random noise, scale factor and cross-coupling errors. These are discussed in the following sections. Further error sources, which are either higher order or dependent on the sensor type, are also briefly discussed. A generalised error model of inertial sensors is given at the end.

### 5.3.1. Biases

The bias, or more precisely the g-independent bias, refers to the sensor output which is present even in the absence of an applied input, such as a rotation or specific force. It is a constant contribution to the systematic error and is exhibited by all gyros and accelerometers. The size of the bias is independent of any motion to which the sensor may be subject to. Hence the bias is usually known as g-independent bias in manufacture specifications in contrast to the g-dependent bias (discussed in Section 5.3.4).

The cause of the bias may be a combination of a number of effects, and is therefore dependent on the sensor type and operation environment. For instance, the causes of the bias for a rate-integrating gyro may include residual torques, and fluid flow around the float assembly as a result of uneven thermal distribution. Under most circumstances, the bias is the dominant source of the systematic error of an inertial sensor.

A more detailed modelling of the bias divides the bias into two components, a static one and a dynamic one. The static component, known as the turn-on bias, consists of the run-to-run variation of the sensor bias plus a residual fixed contribution after the IMU calibration, whereas the dynamic component, known as in-run bias, contains the in-run variation of the sensor bias plus residual temperature-dependent variation. Typically, the magnitude of the dynamic component is about 10 percent of the static component.

Biases for the inertial sensors are usually denoted by vectors, and are defined for accelerometers and gyros respectively as

$$\mathbf{b}_a = [b_{a,x},\ b_{a,y},\ b_{a,z}]^T \tag{5.1}$$

$$\mathbf{b}_g = [b_{g,x},\ b_{g,y},\ b_{g,z}]^T \tag{5.2}$$

where $x$, $y$ and $z$ are axes for the body frame. The subscript for vectors in the body frame is omitted

here as the biases are always expressed in the body frame. It should be noted that (5.1) and (5.2) assume an orthogonal configuration of the inertial instrument cluster (see descriptions in Section 5.4.2.1). In case the system uses the skewed configuration, the biases for sensors may still be denoted by three-component vectors, but individual components no longer correspond to a specific sensor.

**Table 5.4.:** Typical accelerometer and gyro biases (adapted from Groves, 2008)

| IMU Grade | Accelerometer Bias ($m/s^2$) | Gyro Bias (rad/s) |
|---|---|---|
| Marine | $10^{-4}$ | $5 \times 10^{-9}$ |
| Aviation | $3 \times 10^{-4} \sim 10^{-3}$ | $5 \times 10^{-8}$ |
| Intermediate | $10^{-3} \sim 10^{-2}$ | $5 \times 10^{-7}$ |
| Tactical | $10^{-2} \sim 10^{-1}$ | $5 \times 10^{-6} \sim 5 \times 10^{-5}$ |
| Automotive | $> 10^{-1}$ | $> 5 \times 10^{-4}$ |

The SI unit for an accelerometer bias is $m/s^2$. However manufacturer specifications commonly quote the accelerometer bias in terms of the acceleration due to gravity (g), expressed as milli-gal (mgal) or micro-gal ($\mu$gal). The unit conversion between the two follows $1\,g = 9.80665\,m/s^2$ (Tennent, 1971). The unit for the gyro bias are usually quoted as degree per hour ($^\circ$/hr) in sensor specifications or degree per second ($^\circ$/s) for low performance sensors, and can be converted into radians per second (rad/s) using $1\,^\circ/hr = 4.848 \times 10^{-6}$ rad/s. The typical accelerometer and gyro bias figures for various grades of IMU in Table 5.4 are expressed in using SI units.

### 5.3.2. Random Noise

All inertial sensor outputs are subjected to some level of random noise. Random noise in the sensor output with frequency lower than 1 Hz can be approximated as zero-mean white noise (Groves, 2008). The causes of the random noise in the output are mostly sensor dependent. For example, whilst the mechanical instability of a pendulous accelerometer and a dynamically tuned gyro is the main source for random noise, the noise for vibratory sensors could come from high-frequency resonances. The noise affects the resolution of an inertial sensor, particularly for MEMS sensors.

The effect of random noise exhibited in the sensor output is a random walk error in the navigation solution, and it is the random walk error that occasionally gets quoted in manufacturer documents. Random noise on the specific force produces a random walk error in the velocity solution after integration. Similarly, random noise on the angular rate results in a random walk error on the attitude.

MEMS sensors can exhibit significant high-frequency noise (Groves, 2008). This can produce problems when the IMU is rotating because the noise does not average out to the same extent in the reference navigation frame as it used to when the IMU is stationary. As a result, highly dynamic applications usually prove to be challenge for MEMS based sensors. On the other hand, because most commercial MEMS gyro currently available are vibratory gyros, additional problems can arise when applying the sensor to a vibration environment where the external vibration frequency is close to the resonant frequency of the sensor. Under such circumstances, the resulting noise in the sensor output increases dramatically.

White random noise cannot be calibrated or compensated. In order to mitigate such noise, pre-filtering techniques can be applied on the measurements, such as wavelet and neural network methods (El-Rabbany & El-Diasty, 2004). But the bandwidth of the sensor is limited when pre-filtering techniques are used.

The random noise on an inertial sensor cluster is denoted by three-component vectors, defined as

$$\mathbf{w}_a = [w_{a,x},\ w_{a,y},\ w_{a,z}]^T \tag{5.3}$$

$$\mathbf{w}_g = [w_{g,x},\ w_{g,y},\ w_{g,z}]^T \tag{5.4}$$

respectively for accelerometers and gyros.

**Table 5.5.:** Typical accelerometer and gyro random noises

| IMU Grade | Accelerometer ($\mu g/\sqrt{Hz}$) | Gyro ($^\circ/\sqrt{hr}$) |
|---|---|---|
| Aviation | 20 | 0.002 (Spinning mass gyro)<br>0.01 (Ring laser gyro) |
| Tactical | 100 | $0.03 \sim 0.1$<br>(Interferometric fibre optical gyro<br>or Quartz vibratory gyro) |
| Automotive | 1,000 | 1 (Silicon vibratory gyro) |

Random noises are usually described in terms of their root power spectral density (PSD), $n$. The standard deviation of the random noises can be obtained by dividing their root PSD by the root of the sampling interval. The units for the root PSD of random noises on inertial sensors varies from manufacturer to manufacturer. A common one for accelerometers is $\mu g/\sqrt{Hz}$, where $1\,\mu g/\sqrt{Hz} = 9.80665 \times 10^{-5}\,m/s^{1.5}$ (Crassidis & Junkins, 2004). Common units for gyro random noise are $^\circ/\sqrt{hr}$ and $^\circ/hr/\sqrt{Hz}$, where $1\,^\circ/\sqrt{hr} = 2.909 \times 10^{-4}\,rad/s^{0.5}$, $1\,^\circ/hr/\sqrt{Hz} = 4.848 \times 10^{-6}\,rad/s^{0.5}$.

Typical random noise figures for some inertial sensors are listed in Table 5.5 using information provided in Titterton & Weston (2004) and El-Sheimy & Niu (2007).

### 5.3.3. Scale Factor and Cross-Coupling Errors

Ideally, a sensor should output exactly what it measures from the input, i.e. the input-output ratio of the sensor should be unity. However, this is not true in reality and there is always a departure of the sensor input-output ratio from unity. The error caused by this departure is referred to as the scale factor error. The actual output from sensors is proportional to the true quantity.

The cross-coupling error refers to the error caused by the misalignment of the sensor sensitive axes and the orthogonal axes of the body frame. This is usually due to manufacturing limitations. The result of this error is that accelerometers are only sensitive to specific forces that are not orthogonal to their sensitive axes, and rotations measured by gyros are not actually about the body frame.

A normally orthogonal matrix is used to denote the scale factor ($s$) and cross-coupling ($m$) errors, and can be defined as

$$\mathbf{M}_a = \begin{bmatrix} s_{a,x} & m_{a,xy} & m_{a,xz} \\ m_{a,yx} & s_{a,y} & m_{a,yz} \\ m_{a,zx} & m_{a,zy} & s_{a,z} \end{bmatrix}, \quad \mathbf{M}_g = \begin{bmatrix} s_{g,x} & m_{g,xy} & m_{g,xz} \\ m_{g,yx} & s_{g,y} & m_{g,yz} \\ m_{g,zx} & m_{g,zy} & s_{g,z} \end{bmatrix} \tag{5.5}$$

respectively for accelerometers and gyros. Note that $x$, $y$, $z$ represents three axes of the body frame, and the combination for two of them denotes the plane formed by the axes.

The scale factor and cross-coupling errors are often expressed in parts per million (ppm) or as a percentage. Most inertial sensors have the scale factor and cross-coupling errors on the magnitude

of 100 $\sim$ 1,000 ppm (Groves, 2008). Whereas MEMS sensors could exhibit significant error as high as 0.1% (El-Sheimy & Niu, 2007).

For an integrated system, scale factor and cross-coupling errors are usually on the borderline for state vector inclusion. Apart from some high dynamic cases, such as guided weapons, these errors are not observable for most air, land, sea and space applications (Groves, 2008).

### 5.3.4. Further Error Sources

Some gyros exhibit biases which are sensitive to the specific force been applied. This type of biases is called g-dependent bias. A common cause for the g-dependent bias is the mass unbalance of key mechanical parts in the sensor. Thus spinning mass and vibratory gyros usually suffer from g-dependent biases. The g-dependent bias for a gyro triad is denoted by a $3 \times 3$ matrix $\mathbf{G}_g$.

Additional sensor errors related to the existence of scale factor errors are scale-factor non-linearity and scale-factor asymmetry. The scale-factor non-linearity is an indication on the systematic deviation from the least-squares fitted function of input and output, and is usually quoted in percentages. The scale-factor asymmetry describes the difference between outputs with inputs having the same quantity but opposite directions. Low-cost sensors usually suffer from significant scale-factor asymmetry.

There are also higher order errors that exist for spinning mass gyros, such as the anisoelastic ($g^2$-dependent) bias and anisoinertia bias. Anisoelastic bias is proportional to the product of acceleration along orthogonal pairs of axes, whereas anisoinertia bias is due to inequalities in the gyroscope moments of inertia about different axes. These higher order errors are mostly sensor dependent, and should be treated based on specific system design.

### 5.3.5. General Error Model

A general error model for inertial sensors based on the main error sources discussed in previous sections is given here. The outputs from an accelerometer and a gyro are written respectively as following

$$\tilde{\mathbf{f}}_{ib}^b = (\mathbf{I}_{3\times3} + \mathbf{M}_a)\mathbf{f}_{ib}^b + \mathbf{b}_a + \mathbf{w}_a \tag{5.6}$$

$$\tilde{\boldsymbol{\omega}}_{ib}^b = (\mathbf{I}_{3\times3} + \mathbf{M}_g)\boldsymbol{\omega}_{ib}^b + \mathbf{b}_g + \mathbf{G}_g\mathbf{f}_{ib}^b + \mathbf{w}_a \tag{5.7}$$

where the $\tilde{\mathbf{f}}_{ib}^b$ and $\tilde{\boldsymbol{\omega}}_{ib}^b$ are the specific force and the angular rate measurements, and other terms are defined through previous sections. The superscript b represents the body frame and the subscript ib represents the body frame with respect to the inertial frame. In case the estimates about the biases, scale factor and cross-coupling errors, and gyro g-dependent biases can be made, they can be applied to the measurements producing the corrected measurements

$$\hat{\mathbf{f}}_{ib}^b = (\mathbf{I}_{3\times3} + \hat{\mathbf{M}}_a)^{-1}\tilde{\mathbf{f}}_{ib}^b - \hat{\mathbf{b}}_a \tag{5.8}$$

$$\hat{\boldsymbol{\omega}}_{ib}^b = (\mathbf{I}_{3\times3} + \hat{\mathbf{M}}_g)^{-1}\tilde{\boldsymbol{\omega}}_{ib}^b - \hat{\mathbf{b}}_g - \hat{\mathbf{G}}_g\hat{\mathbf{f}}_{ib}^b \tag{5.9}$$

## 5.4. Inertial Navigation

This section introduces the basic concepts for inertial navigation. Section 5.4.1 gives an overview of inertial navigation. Different implementations of inertial systems and sensor configurations are described in Section 5.4.2. Navigation equations for the local navigation frame are listed in Section 5.4.3. Lastly, some equation implementation issues are discussed in Section 5.4.4.

### 5.4.1. Overview

Inertial navigation is a dead reckoning navigation method in which the solution is obtained by continuously measuring the vehicle's accelerations and changes in its pointing directions. The operation of an inertial navigation system (INS) therefore relies on the laws of classical mechanics. Three sets of inertial sensors, together commonly referred to as an inertial measurement unit (IMU), are usually configured in three orthogonal axes so that each of them can measure the vehicle movement in a single direction not observed by the others. These three axes define the body reference frame of the vehicle, and the rotation between the body frame and the navigation reference frame is the attitude of the vehicle. Each set of sensors in an IMU consists of an accelerometer, which measures the specific force, and a gyro, which tracks the turning of the axis. All inertial measurements are processed by the navigation processor, which contains an integrator, to obtain the navigation solution.

Whilst the underlying principles of all inertial systems are the same, the implementation of specific systems may take a variety of forms. Through the development of inertial navigation technology, which can be dated back as early as the German V-series rockets at the end of World War II, the implementation of inertial systems roughly can be divided into two categories: stable platform systems and strapdown systems. Different system implementations use the gyro sensor differently. Stable platform systems usually come with very sophisticated mechanical structures, whereas strapdown systems need a higher processing power and rely on a fine alignment with the mounted vehicle platform (see Section 5.4.2.2).

No matter which implementation an inertial system adopts, a very accurate knowledge of the vehicle attitude is always critical for the final solution. Apart from using high performance gyros to ensure the measurement quality, the calculation of attitude kinematics forms another important part of attitude processing. There are several ways to represent the attitude of a vehicle including Euler angles, direction cosine matrix (or frame transformation matrix) and quaternions. A brief description of these attitude representations is given in Appendix B. More information on attitude kinematics can be found in Kuipers (1999) and Wie (2008).

The solution of position and velocity from an inertial system may be resolved in several reference frames. Common choices of the reference frame include the inertial frame, the earth frame and the local navigation frame. Definitions of various reference frames involved in the navigation calculation in this thesis is given in Appendix A, along with frame transformation matrices between them. Because of the relative movement between reference frames, additional apparent forces, which are functions of the frame motion, act upon the vehicle as well as the actual driving specific force. As a result, navigation equations used in inertial navigation are always derived corresponding to different choices of reference frames. The local navigation frame is chosen as the reference frame for inertial navigation during this research.

The usages of inertial technology do not always require a complete suite of INS. In fact, a typical construction of an inertial system may come with different building blocks and is used at different levels. Figure 5.1 shows an illustration of the building blocks for a strapdown system and terminologies for various usages. Note that an attitude and heading reference system (AHRS) is a combination of instruments that are capable of maintaining an accurate estimate of the vehicle attitude as the vehicle manoeuvres. More details on AHRS design and integration can be found in Farrell (2008).

An inertial navigation system is often characterised by its continuous high output rate and self-contained working mode. In an integrated system, the output rate of an inertial system is usually much higher than other aiding sensors. The output rate for a GNSS receiver is usually less than 10 Hz, whereas the output rate of an inertial system is commonly higher than 50 Hz (Mirobotics Inc., 2009). The inertial system is thus usually chosen as the reference system in an integrated system.

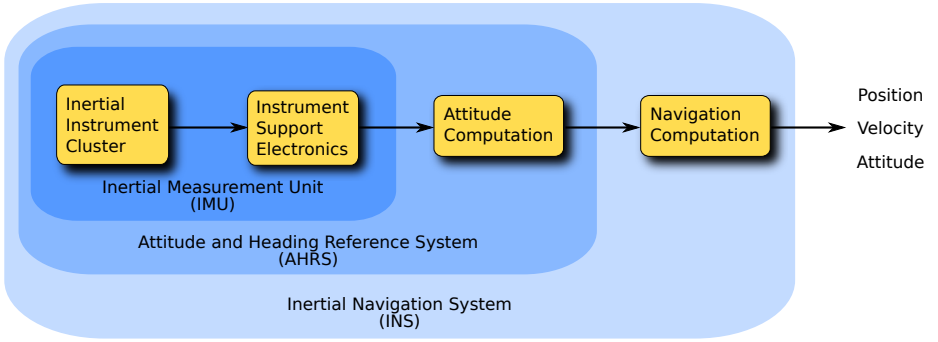



**Figure 5.1.:** Strapdown inertial navigation building blocks (adapted from Titterton & Weston, 2004)

On the other hand, inertial systems are not dependent on the transmission of signals from the vehicle or any external source. Consequently, inertial systems do not suffer from signal transmission difficulties as many radio positioning systems do.

However, the accuracy of an inertial system relies upon the knowledge of vehicle initial conditions. The initialisation and alignment process for an inertial system without external aiding usually involves intensive calculations and can be significantly longer than other positioning methods under difficult circumstances. In addition, as the inertial calculation is essentially a continuous integration process, the solution quality of an inertial system degrades as time progresses. A regular calibration of the system is therefore necessary to ensure the accuracy of the solution.

## 5.4.2. Sensor Configurations and System Implementations

### 5.4.2.1. Sensor Configurations

The sensor configuration of an inertial system refers to the way inertial sensors are installed in the instrument cluster and output measurements are represented with respect to the body frame. The instruments contained in the cluster usually include a number of gyros and accelerometers. While gyros are used to measure the angular rate, accelerometers are used to measure the specific force. The inertial sensors installed are not necessarily single-axis and can be dual-axis (so they can sense the movement in two directions).

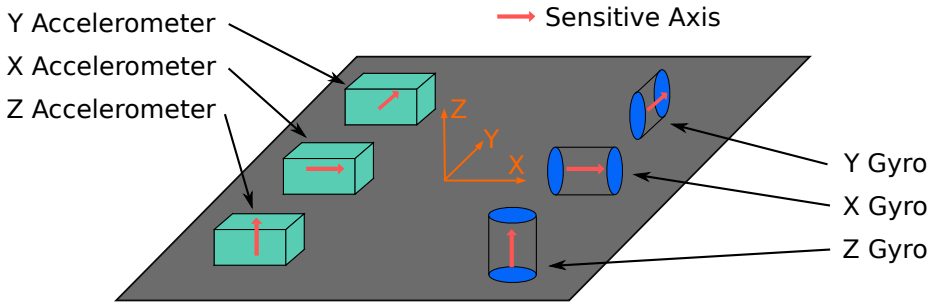



**Figure 5.2.:** Orthogonal sensor configuration for the instrument cluster

The simplest sensor configuration is an orthogonal configuration. Figure 5.2 shows an illustration of an instrument cluster under the orthogonal configuration assuming all sensors are single-axis. As can be seen from the figure, sensors are divided into three groups with one accelerometer and one gyro in each group. The sensitive axes of sensors in the same group are configured pointing the same direction. Three groups of sensors are arranged in such a way that a Cartesian reference frame could be defined with the sensitive axes of three groups. As pointed out earlier, dual-axis sensors

can be used in the cluster as well. The use of dual-axis sensors can be carefully configured in such a way that not only fewer sensors are needed but also a redundant measurement can be provided as a built-in integrity monitoring function.

Instead of using the sensitive axis of a sensor directly, it is possible to mount the sensors in orientations other than the body frame axes. This is usually referred as a skewed configuration. The measurements provided by this type of configuration are expressed as independent linear combinations of the orthogonal components of angular rate and specific force. The main reason for using this configuration is because of high integrity requirement applications. Also, for situations where the turn rate of a single axis is higher than current instrument range, a skewed configuration may be used. However, the processing load for this configuration is higher as the extraction of angular rate and specific force becomes part of the navigation processing.

In the simulations used in this research, it is assumed that only single-axis sensors are available and the instrument cluster is arranged in the orthogonal configuration.

#### 5.4.2.2. System Implementations

As mentioned in Section 5.4.1, an inertial system may be realised in different forms. This section gives a brief review on both stable platform systems and strapdown systems.

The stable platform systems are the initial realisation of the practical inertial systems. The core of this implementation is a very sophisticated mechanical mechanism containing a platform and a number of gimbals. The platform is where the instrument cluster is mounted. The gimbals connect the platform to the vehicle so that at least three degrees of rotational freedom is available for the platform. Any rotational movements are sensed by the gyros, the output of which is fed back to a torque motor to rotate the gimbals in the opposite direction. As a result, the platform in 'isolated' from the rotational motion of the vehicle, and the accelerometer triad on the platform is always held in alignment with the chosen reference frame. Figure 5.3 shows a functional block diagram of the stable platform system.

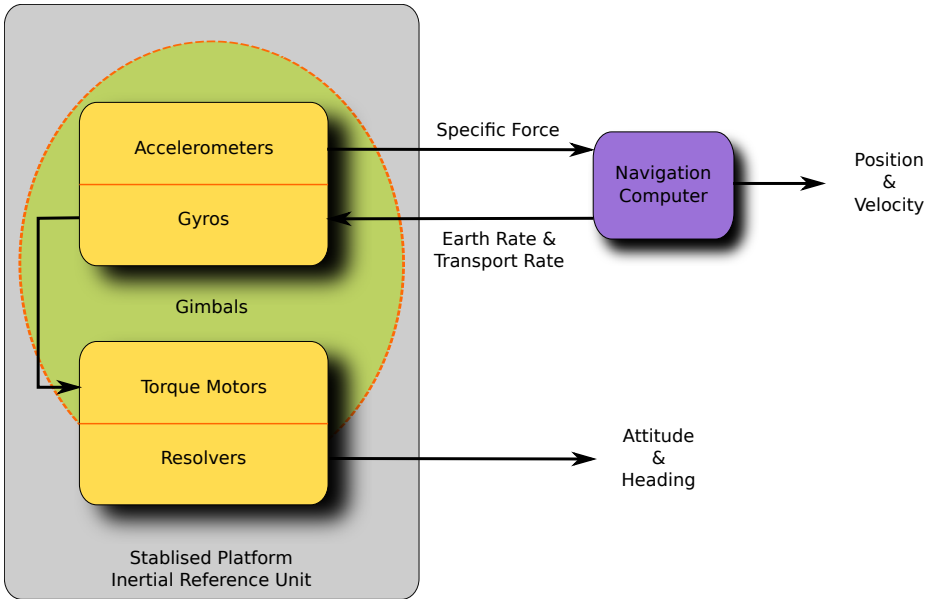



**Figure 5.3.:** Stable platform inertial system

The stable platform systems were first developed because neither sensors with proper dynamic range nor computers with sufficient processing power were available for the strapdown system to be produced. Since the platform is always kept stable, the navigation calculation process for a stable

platform system is much simpler than a strapdown system. It should be noted that the stable platform system is still in wide use today, especially for applications requiring a very long period (unaided) of accurate navigation results, such as navigating ships and submarines.

The strapdown system, on the other hand, avoids the gimbal mechanism by rigidly mounting the instrument cluster to the vehicle. Inertial sensors in a strapdown system are therefore subjected to the vehicle rotational motion. Figure 5.4 shows a functional block diagram of the strapdown system. As shown by the figure, the mechanical complexity of the platform system is replaced by additional computations. These additional computations process the signals from inertial sensors prior to the navigation computation. Unlike the stable platform system, gyro measurements are used for attitude calculation rather than torque feedback.

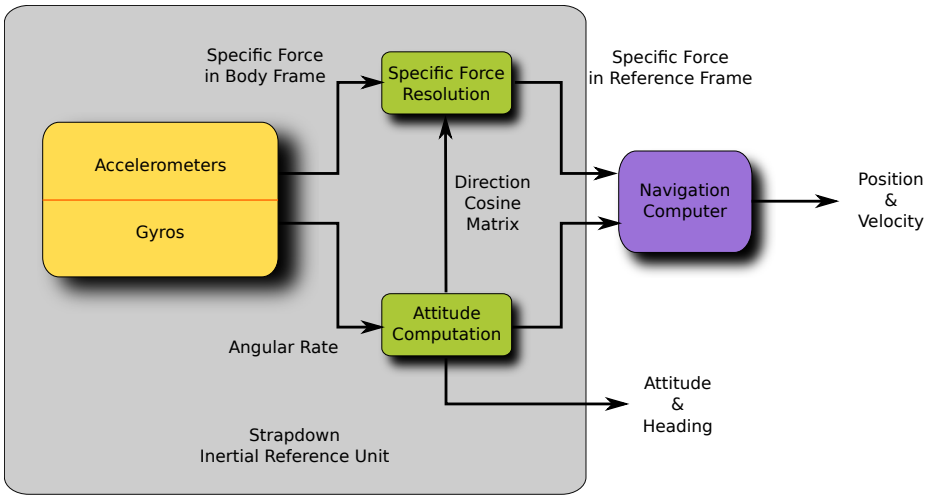



**Figure 5.4.:** Strapdown inertial system

Because of the use of a computer to resolve inertial data, the cost and size of a strapdown system are subsequently reduced. For IMUs that are based on MEMS sensors, the system implementation is usually strapdown. Therefore the simulation of inertial sensor data and calculation in this research are all based on the assumption of a strapdown system.

### 5.4.3. Navigation Equations in the Local Navigation Frame

The assumption for the inertial system simulation in this thesis is established in Section 5.4.1 and Section 5.4.2, that is the simulated system is a strapdown system with orthogonal sensor configuration and the chosen navigation reference frame is the local navigation frame. The navigation equations for various reference frames are derived differently. This section covers the navigation equations in the local navigation frame and the basic calculation procedure. A more detailed derivation process of these equations can be found in Titterton & Weston (2004). Figure 5.5 shows a block diagram of the basic calculation procedure.

Note that epochs $k$ and $k-1$ are generalised as $(+)$ and $(-)$ in order to simplify the notations. The symbols used in Figure 5.5 is defined as:

$\mathbf{C}_b^n$ is the direction cosine matrix of the frame transformation from the body frame to the local navigation frame.

$\mathbf{v}_{eb}^n$ is the velocity vector in local navigation frame, in meters/second.

$L$, $\lambda$, $h$ are respectively the geodetic latitude, longitude, and height, in radians and meters.

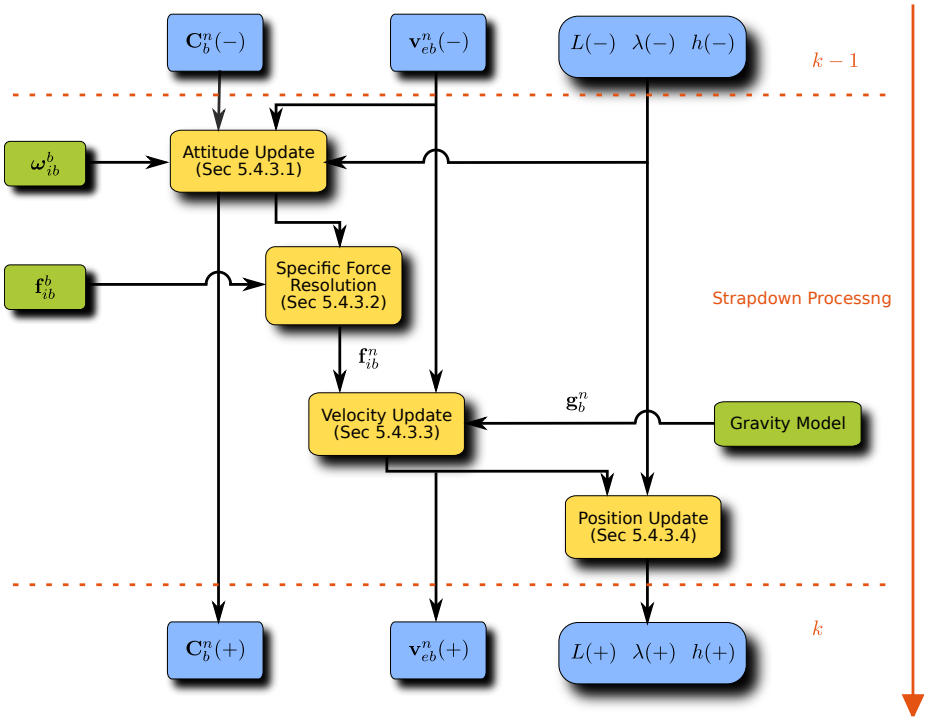



**Figure 5.5.:** Navigation equations for the local navigation frame

$\boldsymbol{\omega}_{ib}^b$ is the angular rate sensed by gyros, in radians/second.

$\mathbf{f}_{ib}^b$ is the specific force sensed by accelerometers, in meters/second$^2$.

$\mathbf{f}_{ib}^n$ is the specific force after frame transformation resolved in local navigation frame, in meters/second$^2$.

$\mathbf{g}_b^n$ is the gravity model resolved in local navigation frame.

The notation in this chapter follows the convention of most inertial navigation literature. Reference frames are denoted by lower-case characters: the inertial frame is $i$, the Earth frame is $e$, the local navigation frame is $n$, and the body frame is $b$. For most vector and matrices, the superscript on the top right corner of a symbol represents the reference frame in which the vector or matrix is resolved. The subscript on the bottom right corner of a symbol describes the relative motion between two reference frames which the symbol is used to describe. On the other hand, the subscript of a direction cosine matrix means the reference frame that a frame transformation starts with, whereas the superscript denotes the destination frame of the transformation.

#### 5.4.3.1. Attitude Update

As shown by Figure 5.5, the attitude update process for the local navigation frame uses the previous solution as well as angular rate measurements from the gyros. The time derivative of the direct cosine matrix, $\mathbf{C}_b^n$, is given by

$$\dot{\mathbf{C}}_b^n = \mathbf{C}_b^n \boldsymbol{\Omega}_{ib}^b - (\boldsymbol{\Omega}_{ie}^n + \boldsymbol{\Omega}_{en}^n)\mathbf{C}_b^n \tag{5.10}$$

where $\boldsymbol{\Omega}$ denotes the skew-symmetric matrix of an angular rate term. There are three rotational rate terms in (5.10). The first term, $\boldsymbol{\Omega}_{ib}^b$, is the inertial reference angular rate, sensed by the gyros. The second term, $\boldsymbol{\Omega}_{ie}^n$, is the rotation of the Earth frame with respect to the inertial frame, referred

to as the Earth rotation rate. The last term, $\boldsymbol{\Omega}_{en}^{n}$, occurs as the local navigation frame moves with respect to the Earth and is usually known as the transport rate. The last two terms appear in the equation as a result of the relative motions between different frames.

The Earth rotation rate vector, $\boldsymbol{\omega}_{ie}^{n}$, is defined by

$$\boldsymbol{\omega}_{ie}^{n} = \begin{bmatrix} \omega_{ie}\cos L \\ 0 \\ -\omega_{ie}\sin L \end{bmatrix} \tag{5.11}$$

where $\omega_{ie}$ the Earth rotation constant. For WGS-84, $\omega_{ie} = 7.292115 \times 10^{-5}$ rad/s (NGA, 1997). Its skew-symmetric matrix is

$$\boldsymbol{\Omega}_{ie}^{n} = \omega_{ie} \begin{bmatrix} 0 & \sin L & 0 \\ -\sin L & 0 & -\cos L \\ 0 & \cos L & 0 \end{bmatrix} \tag{5.12}$$

The transport rate vector, $\boldsymbol{\omega}_{en}^{n}$, can be obtained using previous navigation solution, and is written as

$$\boldsymbol{\omega}_{en}^{n} = \begin{bmatrix} \frac{v_{eb,E}^{n}}{R_E(L)+h} \\ \frac{-v_{eb,N}^{n}}{R_N(L)+h} \\ \frac{-v_{eb,E}^{n}\tan L}{R_E(L)+h} \end{bmatrix} \tag{5.13}$$

where subscripts $E$ and $N$ are used to denote the easting and northing components of a vector in the local navigation frame. Its skew-symmetric matrix is

$$\boldsymbol{\Omega}_{en}^{n} = \begin{bmatrix} 0 & -\omega_{en,z}^{n} & \omega_{en,y}^{n} \\ \omega_{en,z}^{n} & 0 & -\omega_{en,x}^{n} \\ -\omega_{en,y}^{n} & \omega_{en,x}^{n} & 0 \end{bmatrix} \tag{5.14}$$

where subscripts $x$, $y$ and $z$ denote the three components of $\boldsymbol{\omega}_{en}^{n}$ defined by (5.13).

Many numerical methods can be used to solve (5.10). For most applications, a first-order Euler method can be suffice, and the solution of (5.10) is given as

$$\mathbf{C}_{b}^{n}(+) \approx \mathbf{C}_{b}^{n}(-)(\mathbf{I}_{3\times 3} + \boldsymbol{\Omega}_{ib}^{b}\tau_i) - (\boldsymbol{\Omega}_{ie}^{n}(-) + \boldsymbol{\Omega}_{en}^{n}(-))\mathbf{C}_{b}^{n}(-)\tau_i \tag{5.15}$$

where $\tau_i$ is the update interval for the inertial system and $\mathbf{I}_{3\times 3}$ is a 3 by 3 identity matrix. Note that $\boldsymbol{\Omega}_{ie}^{n}(-)$ and $\boldsymbol{\Omega}_{en}^{n}(-)$ are calculated using solutions from last epoch (i.e. $L(-)$, $h(-)$ and $\mathbf{v}_{eb}^{n}(-)$). A higher precision numerical process is discussed in Section 5.4.4.3.

#### 5.4.3.2. Frame Transformation for Specific Force

After obtaining attitude information, the specific force sensed by the accelerometers which is resolved in the body frame can be transformed into the local navigation frame using

$$\mathbf{f}_{ib}^{n}(t) = \mathbf{C}_{b}^{n}(t)\mathbf{f}_{ib}^{b}(t) \tag{5.16}$$

For most tactical grade or lower grade applications, a simple approximation can be applied to (5.16),

$$\mathbf{f}_{ib}^{n} \approx \frac{1}{2}(\mathbf{C}_{b}^{n}(-) + \mathbf{C}_{b}^{n}(+))\mathbf{f}_{ib}^{b} \tag{5.17}$$

#### 5.4.3.3. Velocity Update

The time derivative of the velocity can be expressed in terms of the specific force, gravity, and centrifugal acceleration, which gives

$$\dot{\mathbf{v}}_{eb}^{n} = \mathbf{f}_{ib}^{n} + \mathbf{g}_{b}^{n}(L,h) - (\mathbf{\Omega}_{en}^{n} + 2\mathbf{\Omega}_{ie}^{n})\mathbf{v}_{eb}^{n} \tag{5.18}$$

The gravity model used here follows Crassidis (2006), and is defined as following

$$\mathbf{g}_{b}^{n}(L,h) = \begin{bmatrix} 0 \\ 0 \\ g_{b,D}^{n} \end{bmatrix} \tag{5.19}$$

where $g_{b,D}^{n}$ is the down component of the gravity and is

$$\begin{aligned} g_{b,D}^{n} = {} & 9.780327(1 + 5.3024 \times 10^{-3}\sin^2 L - 5.8 \times 10^{-6}\sin^2 2L) \\ & - (3.0877 \times 10^{-6} - 4.4 \times 10^{-9}\sin^2 L)h + 7.2 \times 10^{-14}h^2 \end{aligned} \tag{5.20}$$

The third part of (5.18), the centrifugal part, is a combination of the Coriolis acceleration and transport rate terms. The Coriolis force is an apparent force caused by the movement of the local navigation frame in a rotating reference, or the Earth frame.

It is very difficult to obtain a full analytical solution to (5.18). As a result, numerical methods with some approximations are usually used. The Coriolis and transport rate terms are smaller than other terms, and their variations over the update interval in general are neglected. Moreover, as the variation of the gravity is slow over the interval, it is also considered constant over the update interval. Therefore, (5.18) can be solved with first-order Euler approximation, written as

$$\mathbf{v}_{eb}^{n}(+) = \mathbf{v}_{eb}^{n}(-) + [\mathbf{f}_{ib}^{n} + \mathbf{g}_{b}^{n}(L(-),h(-)) - (\mathbf{\Omega}_{en}^{n}(-) + 2\mathbf{\Omega}_{ie}^{n}(-))\mathbf{v}_{eb}^{n}(-)]\,\tau_i \tag{5.21}$$

#### 5.4.3.4. Position Update

The time derivative of latitude, longitude and height is

$$\dot{L} = \frac{v_{eb,N}^{n}}{R_N(L) + h} \tag{5.22}$$

$$\dot{\lambda} = \frac{v_{eb,E}^{n}}{(R_E(L) + h)\cos L} \tag{5.23}$$

$$\dot{h} = -v_{eb,D}^{n} \tag{5.24}$$

The variation of the meridian and transverse radii of curvature, $R_N$ and $R_E$, with the geodetic latitude $L$ is weak, and the radii are consequently considered as constant over the update interval. Following Groves (2008), a suitable approximation for the position update is given by

$$h(+) = h(-) - \frac{\tau_i}{2}\left(v^n_{eb,D}(-) + v^n_{eb,D}(+)\right) \tag{5.25}$$

$$L(+) = L(-) + \frac{\tau_i}{2}\left(\frac{v^n_{eb,N}(-)}{R_N(L(-)) + h(-)} + \frac{v^n_{eb,N}(+)}{R_N(L(-)) + h(+)}\right) \tag{5.26}$$

$$\lambda(+) = \lambda(-) + \frac{\tau_i}{2}\left(\frac{v^n_{eb,E}(-)}{(R_E(L(-)) + h(-))\cos L(-)} + \frac{v^n_{eb,E}(+)}{(R_E(L(+)) + h(+))\cos L(+)}\right) \tag{5.27}$$

assuming a linear variation of velocity over the update interval. Note that (5.25) to (5.27) must be calculated in the given order.

## 5.4.4. Equation Implementation Issues

Section 5.4.3 gives the navigation equations in the local navigation frame. In order to achieve a practical inertial navigation processor, there are several issues related with the use of numerical methods that need to be addressed. These issues are listed and discussed in this section.

### 5.4.4.1. Numerical Integration and Iteration Rates

The navigation equations in Section 5.4.3 are mainly in the form of time derivatives. In general, it is difficult obtain a full analytical solution of those equations. Therefore, approximations are usually made to simplify the numerical calculation of the equations. There are two types of approximations made during the calculation, both based on the assumption of a high update rate. One is that terms which varies slowly during the update interval, such as the transport rate and Coriolis terms in the Earth frame and the local navigation frame, are considered to be constant over the update interval and are consequently calculated with the values from previous epoch. The other one is that the specific force and the angular rate are constant over the averaging interval. Because of these approximations, additional errors could arise if the navigation equations are processed with a low iteration rate.

An iteration rate which is higher than the IMU output rate is usually helpful to increase the calculation precision unless exact equations are used. Moreover, in order to gain full benefits of iterating an equation at a higher rate, the inputs from the previous stage must be at the same rate or higher (Groves, 2008).

Apart from increasing the iteration rate, implementing a higher order numerical integration algorithm may also be helpful. Equations using the first-order Euler method are given in Section 5.4.3 as preliminary solutions. Nevertheless, for low-grade sensor outputs and sensitive steps in the calculation (e.g. attitude update and specific force frame transformation), higher order algorithms should be used. A fourth-order Runge-Kutta algorithm is given here, following Kreyszig (2007). The integration of the time derivatives in Section 5.4.3 is generalised as

$$\mathbf{x}(t + \tau_i) = \mathbf{x}(t) + \int_t^{t+\tau_i} \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t))dt \tag{5.28}$$

where $\mathbf{f}$ is the function of the time derivative $\mathbf{x}$ and input $\mathbf{u}$. Note that $t + \tau_i$ and $t$ equal epochs $k$ and $k-1$, which are simplified as $(+)$ and $(-)$ in Section 5.4.3. (5.28) can be integrated in five

steps:

$$\begin{aligned}
\mathbf{k}_1 &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) \\
\mathbf{k}_2 &= \mathbf{f}\left(\left(\mathbf{x}(t) + \frac{1}{2}\tau_i \mathbf{k}_1\right), \mathbf{u}(t + \frac{1}{2}\tau_i)\right) \\
\mathbf{k}_3 &= \mathbf{f}\left(\left(\mathbf{x}(t) + \frac{1}{2}\tau_i \mathbf{k}_2\right), \mathbf{u}(t + \frac{1}{2}\tau_i)\right) \\
\mathbf{k}_4 &= \mathbf{f}\left(\left(\mathbf{x}(t) + \tau_i \mathbf{k}_3\right), \mathbf{u}(t + \tau_i)\right) \\
\mathbf{x}(t + \tau_i) &= \mathbf{x}(t) + \frac{1}{6}\tau_i(\mathbf{k}_1 + 2\mathbf{k}_2 + 3\mathbf{k}_3 + \mathbf{k}_4)
\end{aligned} \tag{5.29}$$

The Runge-Kutta algorithm can either been used at individual steps or process the navigation equations as a whole.

#### 5.4.4.2. Orthonormality of the Direction Cosine Matrix

The direction cosine matrix, $\mathbf{C}$, is orthonormal by the definition, i.e. the dot product of any pairs of its row or column vectors is zero. For exact equations, the orthonormality of the direction cosine matrix is not affected by the update process. However, as approximations are made and rounding errors could accumulate during the numerical integration process, the orthonormality is likely to be lost between the iterations. Therefore a self-consistency check should always be implemented to check whether the sum of the squares of the elements in each row or column equals unity. In case the check fails, a reorthogonalisation and renormalisation algorithm should be applied.

Following Titterton & Weston (2004), a direction cosine matrix consists of three row vectors and can be written as

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \mathbf{c}_3^T \end{bmatrix} \tag{5.30}$$

The dot product of the $i$th and $j$th row vector is defined as

$$\Delta_{ij} = \mathbf{c}_i^T \mathbf{c}_j \tag{5.31}$$

When $\Delta_{ij}$ does not equal zero, it represent an angle error defined about the axis perpendicular to $\mathbf{c}_i$ and $\mathbf{c}_j$, and the reorthogonalisation is achieved by applying it as a correction. Since there is no ways to identify the row in which the error belongs, or either row is equally likely to be in error, the correction is equally distributed between the two rows. Applying the correction process through each pair of vectors, the new row vector is given by

$$\begin{aligned}
\mathbf{c}_1(+) &= \mathbf{c}_1(-) - \frac{1}{2}\Delta_{12}\mathbf{c}_2(-) - \frac{1}{2}\Delta_{13}\mathbf{c}_3(-) \\
\mathbf{c}_2(+) &= \mathbf{c}_2(-) - \frac{1}{2}\Delta_{12}\mathbf{c}_1(-) - \frac{1}{2}\Delta_{23}\mathbf{c}_3(-) \\
\mathbf{c}_3(+) &= \mathbf{c}_3(-) - \frac{1}{2}\Delta_{13}\mathbf{c}_1(-) - \frac{1}{2}\Delta_{23}\mathbf{c}_2(-)
\end{aligned} \tag{5.32}$$

The normalisation error, $\Delta_{ii}$, is obtained by comparing the sum of the squares of the elements in a row against unity:

$$\Delta_{ii} = 1 - \mathbf{c}_i^T \mathbf{c}_i \tag{5.33}$$

The renormalisation of the direction cosine matrix can be achieved at each row by using

$$\mathbf{c}_i(+) = \mathbf{c}_i(-) - \frac{1}{2}\Delta_{ii}\mathbf{c}_i(-) \tag{5.34}$$

Instead of rows, this process may also be performed on the columns.

#### 5.4.4.3. A Precise Implementation of Navigation Equations

The calculation process shown by Figure 5.5 is sufficient for systems with high performance sensors and high rate updates. Nevertheless, the approximations used in Section 5.4.3 are over simplified for systems with sensors having both lower performance and output rate, such as a MEMS based sensor cluster. A more precise implementation of the navigation equations is introduced in this section following Chatfield (1997) and Groves (2008).

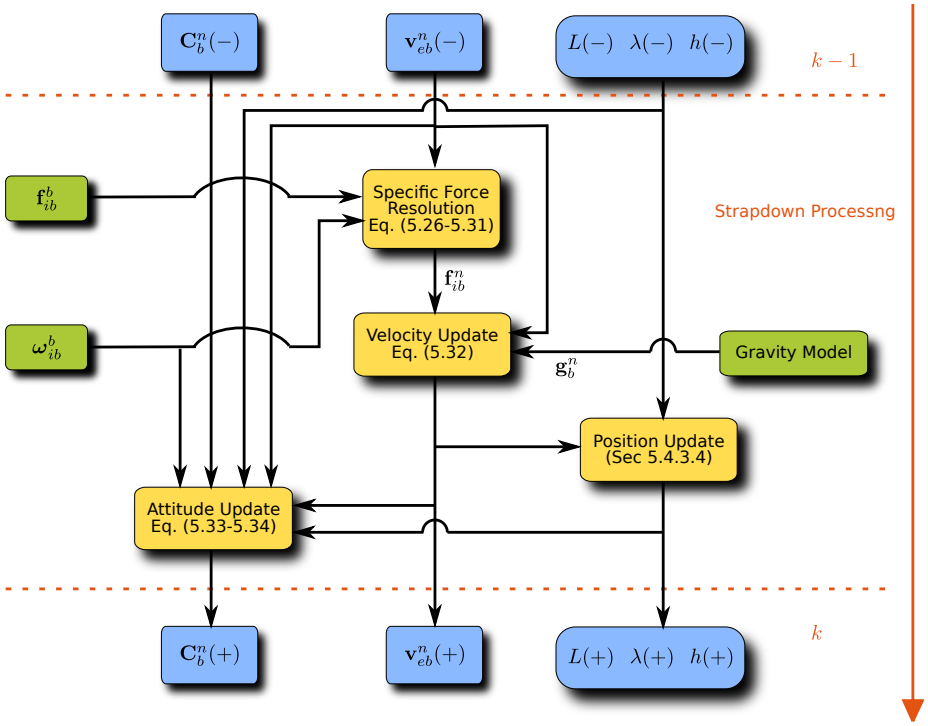



**Figure 5.6.:** A more precise implementation of the navigation equations

Figure 5.6 shows the calculation procedure for this alternative implementation. As can be seen from the figure, the attitude update phase, which is the first phase in Figure 5.5, is now performed as the last step of the epoch. By using the updated information, the precision of the attitude update can be increased and consequently the precision of the strapdown calculation.

Although the update for the direction cosine matrix is performed as the last step, the frame transformation for the specific force still needs attitude information. Thus the attitude increment, $\boldsymbol{\alpha}_{ib}^b$, is calculated first. $\boldsymbol{\alpha}_{ib}^b$ is a rotation angle vector, the direction and magnitude of which can define the rotation of the body frame from epoch $k-1$ to $k$. The attitude increment is defined as:

$$\boldsymbol{\alpha}_{ib}^b = \boldsymbol{\alpha} + \delta\boldsymbol{\alpha} \tag{5.35}$$

in which $\boldsymbol{\alpha}$ is the inertially measurable angular motion sensed by gyros, and is calculated by:

$$\boldsymbol{\alpha} = \int_t^{t+\tau_i} \boldsymbol{\omega}_{ib}^b dt \tag{5.36}$$

and $\delta\boldsymbol{\alpha}$ is a component of $\boldsymbol{\alpha}_{ib}^{b}$ that is non-inertially measurable, modelled by

$$\delta\boldsymbol{\alpha} = \int_{t}^{t+\tau_i} \boldsymbol{\alpha} \times \boldsymbol{\omega}_{ib}^{b} dt \tag{5.37}$$

The presence of $\delta\boldsymbol{\alpha}$ is because of the rotation of $\boldsymbol{\omega}_{ib}^{b}$ which makes $\boldsymbol{\alpha}_{ib}^{b}$ not longer a simple sum of the incremental angle measurements.

In order to transform the specific force into the local navigation frame, an average direction cosine matrix over the time interval is used. The average direction cosine matrix, $\bar{\mathbf{C}}_{b}^{n}$, is approximated by

$$\bar{\mathbf{C}}_{b}^{n} = \mathbf{C}_{b}^{n}(-)\mathbf{C}_{\bar{b}}^{b-} - \frac{1}{2}(\boldsymbol{\Omega}_{ie}^{n}(-) + \boldsymbol{\Omega}_{en}^{n}(-))\mathbf{C}_{b}^{n}(-)\tau_i \tag{5.38}$$

where $\mathbf{C}_{\bar{b}}^{b-}$ is a intermediate attitude update matrix, which is essentially a direction cosine matrix that transforms the body frame from the average update epoch to the beginning of the update, and is defined as

$$\mathbf{C}_{\bar{b}}^{b-} = \mathbf{I}_{3\times 3} + \frac{1-\cos|\boldsymbol{\alpha}_{ib}^{b}|}{|\boldsymbol{\alpha}_{ib}^{b}|^2}\left[\boldsymbol{\alpha}_{ib}^{b}\times\right] + \frac{1}{|\boldsymbol{\alpha}_{ib}^{b}|^2}\left(1 - \frac{\sin|\boldsymbol{\alpha}_{ib}^{b}|}{|\boldsymbol{\alpha}_{ib}^{b}|}\right)\left[\boldsymbol{\alpha}_{ib}^{b}\times\right]^2 \tag{5.39}$$

Note that $\left[\boldsymbol{\alpha}_{ib}^{b}\times\right]$ denotes the skew-symmetric matrix of $\boldsymbol{\alpha}_{ib}^{b}$. Therefore, the specific force can be transformed using

$$\mathbf{f}_{ib}^{n} = \bar{\mathbf{C}}_{b}^{n}\mathbf{f}_{ib}^{b} \tag{5.40}$$

The update for velocity is achieved through a two-step recursive methods, written as

$$\begin{aligned}
\bar{\mathbf{v}}_{eb}^{n} &= \mathbf{v}_{eb}^{n}(-) + \left[\mathbf{f}_{ib}^{n} + \mathbf{g}_{b}^{n}(L(-),h(-)) - (\boldsymbol{\Omega}_{en}^{n}(-) + 2\boldsymbol{\Omega}_{ie}^{n}(-))\mathbf{v}_{eb}^{n}(-)\right]\tau_i \\
\mathbf{v}_{eb}^{n}(+) &= \mathbf{v}_{eb}^{n}(-) + \left\{ \begin{array}{c} \mathbf{f}_{ib}^{n} + \mathbf{g}_{b}^{n}(L(-),h(-)) - \frac{1}{2}[\boldsymbol{\Omega}_{en}^{n}(-) + 2\boldsymbol{\Omega}_{ie}^{n}(-)]\mathbf{v}_{eb}^{n}(-) \\ -\frac{1}{2}[\boldsymbol{\Omega}_{en}^{n}(L(-),h(-),\bar{\mathbf{v}}_{eb}^{n}) + 2\boldsymbol{\Omega}_{ie}^{n}(-)]\mathbf{v}_{eb}^{n}(-) \end{array} \right\}\tau_i
\end{aligned} \tag{5.41}$$

Note that the intermediate velocity, $\bar{\mathbf{v}}_{eb}^{n}$, is used to obtain an updated Coriolis term. The position update uses the same equations given in (5.25) to(5.27).

The update for the direct cosine matrix uses another attitude update matrix, $\mathbf{C}_{b+}^{b-}$, which transform the body frame from the end of attitude update to the beginning. $\mathbf{C}_{b+}^{b-}$ is defined as

$$\mathbf{C}_{b+}^{b-} = \mathbf{I}_{3\times 3} + \frac{\sin|\boldsymbol{\alpha}_{ib}^{b}|}{|\boldsymbol{\alpha}_{ib}^{b}|}\left[\boldsymbol{\alpha}_{ib}^{b}\times\right] + \frac{1-\cos|\boldsymbol{\alpha}_{ib}^{b}|}{|\boldsymbol{\alpha}_{ib}^{b}|^2}\left[\boldsymbol{\alpha}_{ib}^{b}\times\right]^2 \tag{5.42}$$

Thus the updated direction cosine matrix can be obtained with

$$\mathbf{C}_{b}^{n}(+) = \mathbf{C}_{b}^{n}(-)\mathbf{C}_{b+}^{b-} - (\boldsymbol{\Omega}_{ie}^{n}(-) + \frac{1}{2}\boldsymbol{\Omega}_{en}^{n}(-) + \frac{1}{2}\boldsymbol{\Omega}_{en}^{n}(+))\mathbf{C}_{b}^{n}(-)\tau_i \tag{5.43}$$

where $\boldsymbol{\Omega}_{en}^{n}(+)$ is calculated using $L(+)$, $h(+)$ and $\mathbf{v}_{eb}^{n}(+)$. It should be noted that this implementation is very processor intensive.

# 6. Simulation Database Construction

In order study the possible performance of an integrated system in railway environments, simulated sensor measurements are used in this research. This is because a simulation environment provides the researcher with full control of the positioning process, i.e. a known truth file and designed simulation scenarios, therefore it becomes possible to isolate each sensor and study their potential performance under a certain environment individually.

The simulation process described in this chapter starts with real GPS data collected on a rail route. MEMS sensor outputs and GPS ranging measurements are simulated based on the collected data. Truth file of the simulation process is also generated along with sensor outputs. A strapdown inertial calculation process and GPS filter based PVT estimation process is then carried out to test the simulated measurements and possible performances.

Section 6.1 provides justifications for the simulation approach. Section 6.2 gives an brief overview of the simulation process. Information on the collection of GPS data used in this thesis is described in Section 6.3. The process of generating a truth file of the train's position, velocity and acceleration during the simulation is shown in Section 6.4. The DRM used in this research is generated from the position truth file in Section 6.5. Rail tracks and superelevation along the route are simulated in Section 6.6, which are then used to generate the attitude information of the carriage. The truth file of the specific force and angular rate measurements are simulated respectively in Section 6.8 and Section 6.7. The error characteristics of MEMS sensors are added to the generated true sensor measurements to produce realistic system outputs, details are given in Section 6.9. Section 6.10 describes the process to simulate corrected GPS pseudorange and pseudorange rate measurements. Finally, the strapdown INS calculation process and the GPS PVT estimation process are simulated respectively in Section 6.11 and Section 6.12.

## 6.1. Justification for the Simulation Approach

In order to study the possible performance of an integrated system in railway environments, simulated sensor measurements are used in this research. The simulated data includes DRMs with different point intervals and accuracy, outputs from simulated MEMS sensor and outputs from a simulated single frequency GPS receiver. All simulated measurements are based on a truth model of a real train's movement and the track status through the journey. The truth model is constructed from the GPS data collected on a suburban railway. The justification for both using the simulation approach and the concern of multiple data streams generated from a single truth file is firstly provided in this section before the description of the simulation database construction process.

A simulation approach provides the researcher with full control of the positioning process, i.e. a known truth file and designed simulation scenarios. A known truth file provides confidence that the simulated system performance and can be useful to validate the system integrity performance by comparing it against the error information calculated by the system. Different simulation scenarios enable the study of the potential behaviour of a positioning system under particular situations that are designed to test a certain aspect of the system performance. For example, the test of a system

operating under faulty situations is usually difficult to carry out in real time but can be simulated either using recorded data or simulated data.

The sensor data used in a simulation approach based study generally comes from two different sources: either recorded data is used or the data is generated from a known truth file with specified errors added.

Using recorded data has the advantage of an increased confidence on simulation results. As the real data is used, the study of system performance under various simulated scenarios is more realistic than using purely simulated sensor data. However, the use of recorded data also suffers from the limitation of sensor performance variation and the availability of different types of sensors. In the case of studying possible designs of an integrated positioning system as in this thesis, using only recorded data limits the extent of exploring potential system designs and their performances. Nevertheless once proposed system designs are selected, recorded data is useful to validate previous simulated system performance and extending the scope of studying realistic system behaviour by using simulation scenarios including different possible system failure modes. In practice, recorded data must be used in real time system testing.

The main disadvantage of using simulated sensor data is the risk of over simplifying the original process and consequently leading an over-optimistic conclusion for system performances. In terms of the simulated sensor data used in this research, the risk is minimised by controlling various error characteristics that are added to the truth sensor data. Error characteristics used in this research are referenced from major industrial manufacturers and published research literature, which are the best knowledge source of understanding them when no direct access to various sensors is available. In addition, simulated sensor data is respectively used for the position calculation for all sub-systems. Obtained positioning results of each sub-system are then checked against nominal system performance available in published literatures to reduce the risk of over simplification.

For all simulation processes, simplifications are always involved to some extent. The simplifications involved in the simulation process described in this chapter are mainly aspects of the journey that are difficult to simulate and are better examined in a real time test, such as sudden jerks on inertial measurements caused by the change of rail track sections. Most of these events happen randomly in nature.

Another concern with using simulated sensor data is the argument that multiple data streams, i.e. simulated outputs from different sensors used in the integrated system, are generated from the same truth file and may result in different data streams become dependent on each other.

The justification for this should be firstly explained individually for each simulated sensor data. For the inertial systems, the measurements of carriage velocity and attitude are different aspects of the same physical movement. Their truths are therefore correlated in nature. But the final sensor measurements comprise both the truth and error characteristics. Different sensors have different error characteristics and their figures independent to each other. The satellite ranging measurements are produced using both the train position truth file and the real satellite ephemeris file. These two data sources are independent and come from different data streams. Although DRM truth points are generated from the train trajectory of the simulated, they do not carry the same relationship between time and position as the actual position truth file does. In addition, the errors that added to the DRM truth points are generated independently from the original truth file.

Whilst the same train position truth file is used in the process of generating different simulation data used in this research, it is not the only information been incorporated to produce that final simulated sensor data. Different sensors outputs are generated using the position truth file and other source of information. The fact that they all use the same position truth increases ensures that all simulated outputs realistically reflected the same train movement.

Another factor that must be taken into account in choosing a simulation approach is the process of getting real time experiment approved on the railways. Various considerations including route selection, test bed design and equipment safety certification, which are not research related, are involved. These issues make it difficult to start the research with a desired extensive data collection. By taking a simulation approach, a proven improvement by including DRM-aiding will ease the process of designing real time experiments and obtaining recorded data. As a result, real time experiments are considered a necessary part of future work to carry out in order to further research in this area.

## 6.2. An Overview of the Simulation Process

The data generated through the simulation process includes the status of the simulated train throughout the journey, information on the tracks along the route, and sensor outputs. The status of the train is expressed in terms of its movement and the changes of its attitude along the route. The train movement is shown by a truth file of its position, velocity and acceleration, whereas Euler angles are used to indicate the attitude. Information on the rail tracks contains the position of the tracks, superelevation along the route, and DRMs with various specifications. The sensor outputs include simulated measurements from a MEMS inertial unit, satellite ranging measurements, and GPS position and velocity solutions.

It should be noted that the DRMs simulated in this research are matched with the progress of the train movement, i.e. for each position measurement during the simulation process, the exact knowledge of the DRM points behind and ahead of the train current position is known. In practice, the use of DRM aiding may require a reliable mechanism of search the relevant DRM points at each time epoch. One example of the searching algorithm is demonstrated by Zheng (2008). Since the main objective of this research is focused on studying the effect of DRM aiding on an integrated navigation system, the knowledge of relevant DRM points at each time epoch is needed, and the effect of having a searching mechanism can also be simulated based on this truth knowledge.

A flow chart of the simulation process is shown in Figure 6.1. A truth file of the position, velocity and acceleration of the train carriage is firstly generated from the collected GPS data. DRMs are built with different point intervals based on the generated position truth file and are added with different level of errors. The truth DRM data is used to simulate the track status along the route including the rail track position and superelevation. In order to generate superelevation, several sections of the tracks are fitted with a circular curve, and transition curves are identified on both end of each section. The attitude information is built with both track status and the position truth file, and is expressed in the form of Euler angles. Direction cosine matrices between the local navigation frame and the body frame are derived from obtained Euler angles.

The truth file for angular rate and specific force is calculated based on the generated information on train movement and attitude. The error characteristics of a MEMS inertial unit are simulated based on available information of current commercial IMUs. The final simulated sensor outputs are then generated by adding these error characteristics to the available truth files.

The GPS ranging measurements are simulated by using additional information from the IGS precise ephemerides. The visible constellation is selected along with the process of calculating the truth file of range and range rate measurements. The corrected ranging measurements are obtained by adding receiver clock errors and correction residuals.

Finally, a strapdown INS calculation process is performed using the simulated inertial sensor outputs to test the validity of the simulated MEMS error characteristics. A GPS filtered positioning process is also performed to test the validity of simulated ranging measurements, as well as providing

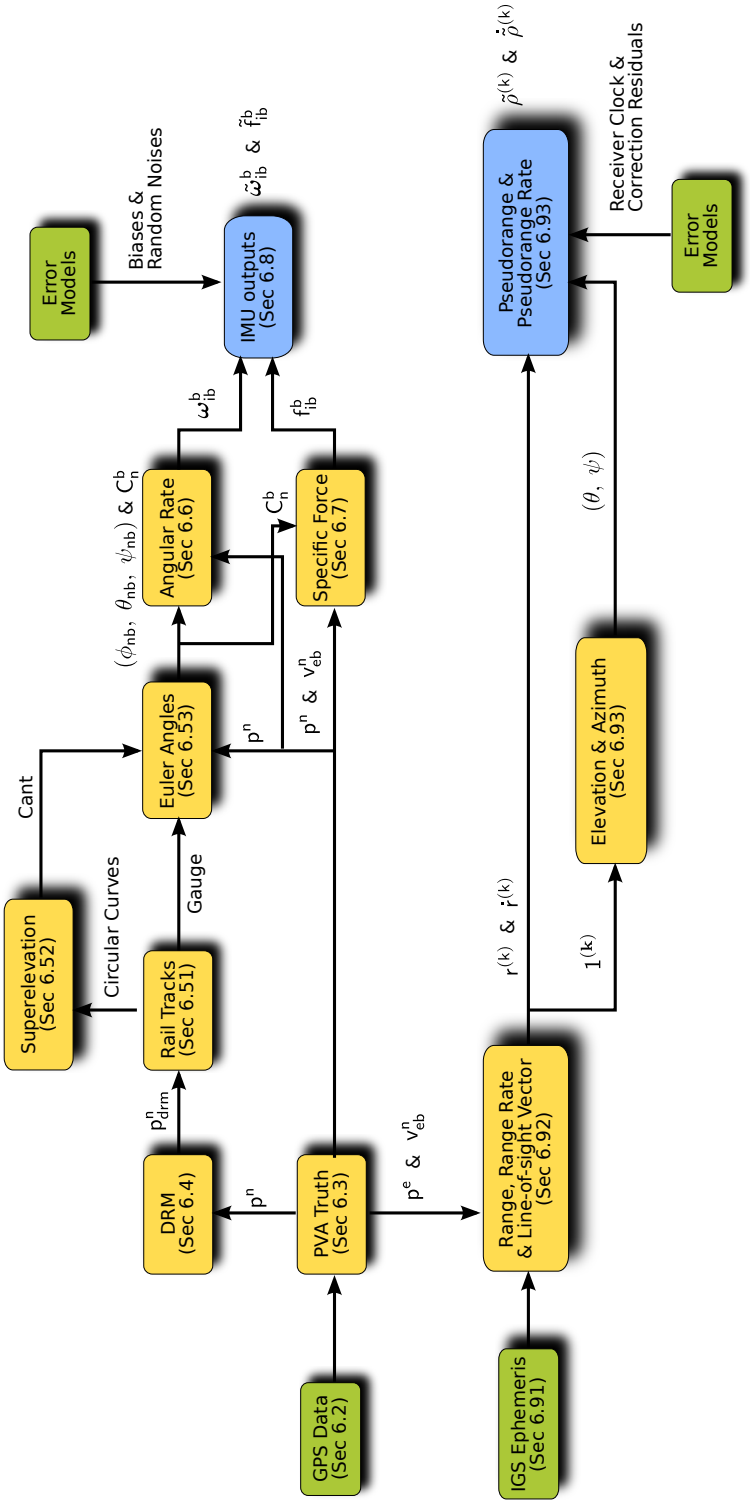



**Figure 6.1.:** An overview of generating simulation data

the simulated position and velocity solutions from a GPS receiver.

## 6.3. GPS Data Collection

The GPS data collection of this research was performed as part of the LOCASYS project (see Section 2.3.2). The GPS data were collected on the rail route between Norwich and Lowestoft by data collection unit fitted on the train. Figure 6.2 shows an overview of the experiment rail route. Markers in the figure are train stations along the route at which the train stopped during its journey, and the blue line indicates the route.

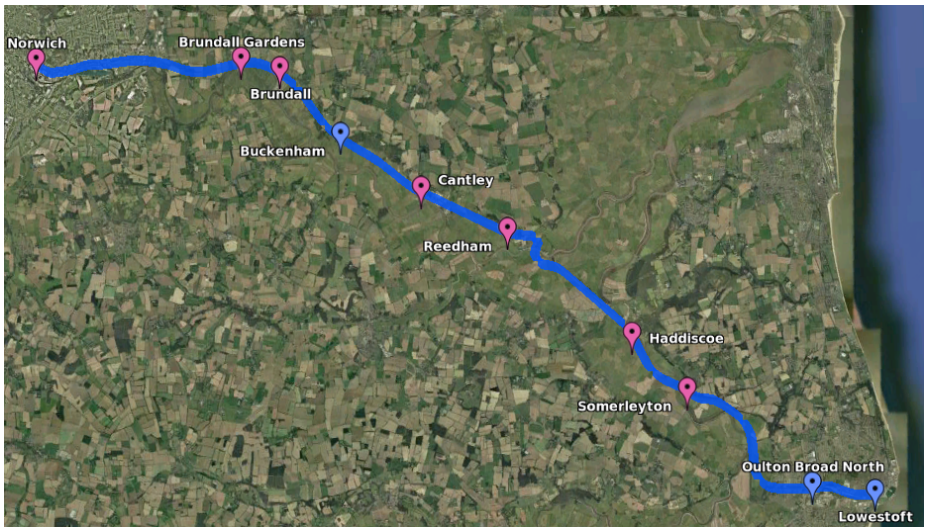



**Figure 6.2.:** Overview of rail route between Norwich and Lowestoft

As can be seen from the figure, the environment along the rail route mainly consists of suburban and countryside areas. The main reason for choosing this rail route is that it can provide an environment which is almost free from severe physical obstructions as discussed in Section 2.3.1.2, and GPS data collected on the route is thus more consistent, and continuous. This enables a truth model of the rail track position to be built from collected GPS data, and the simulation database derived from the truth model.

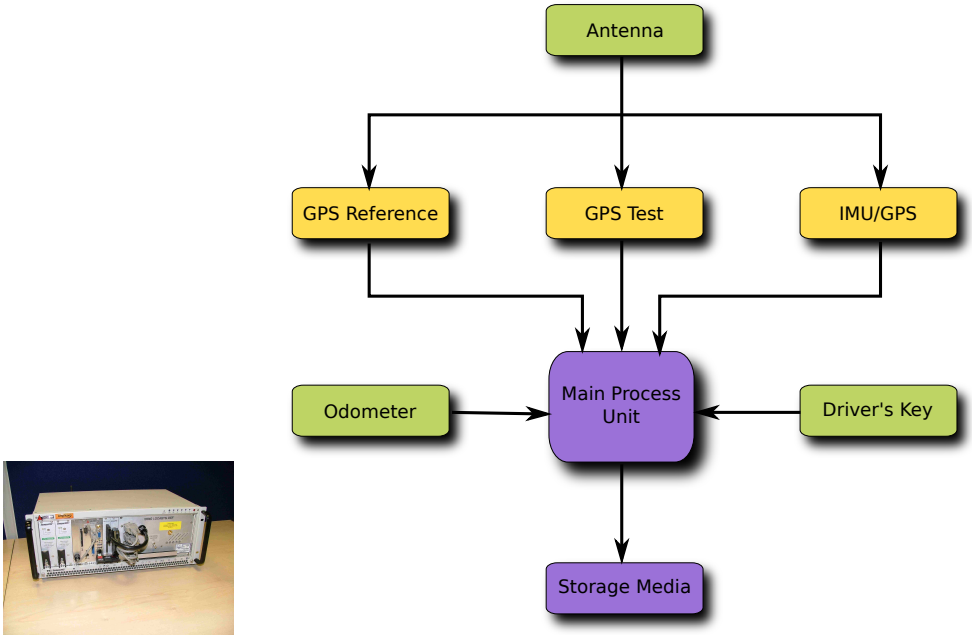
The equipment used to collect the GPS data is comprised of three main components: a dual-frequency geodetic-grade GPS receiver, a single-frequency GPS receiver and an IMU with access to its own single-frequency GPS receiver. Figure 6.3 shows the GPS data collection unit and a functional block figure of its structure[1]. The data used for this research were collected by the dual-frequency receiver. The equipment mounted in the train carriage was installed as shown in Figure 6.4a, and the antenna was located on top of the carriage as shown in Figure 6.4b[1].

The collected data is a position file of the train in the form of latitude, longitude and height in WGS84. The data update interval is 0.5 s (2 Hz). The overall journey took about 40 min from Norwich to Lowestoft.

## 6.4. Position, Velocity and Acceleration Simulation

Truth data of the position, velocity and acceleration were generated from the collected GPS data, to serve as the basis of further simulation. Only the first 32 min of the overall collected data (i.e. the

[1] Courtesy of Nottingham Scientific Ltd.




**(a)** Main sensor unit

**(b)** Structure of the data collection unit

**Figure 6.3.:** GPS data collection unit

**(a)** Receiver box in the carriage

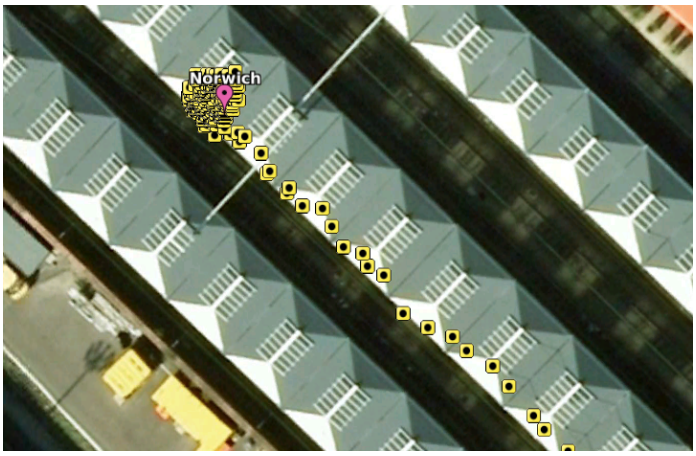
**(b)** Antenna installation

**Figure 6.4.:** Installation of the data collection unit

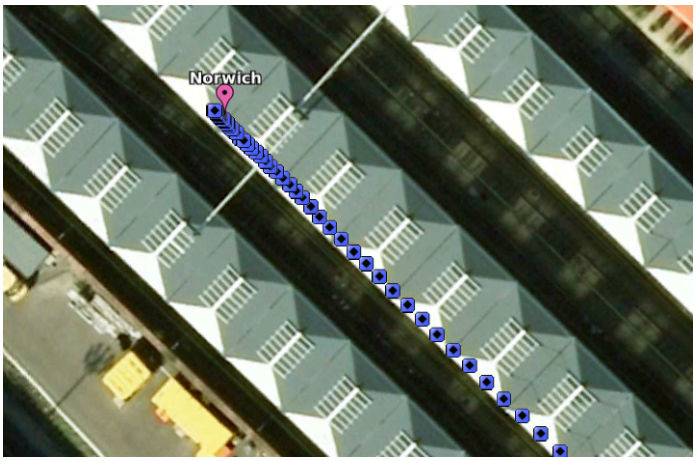
whole journey excluding the area around Oulton Board North and Lowestoft) is used to construct the simulation database. This is because continuous signal outages and time tag corruptions presented in the last section of collected data. The time tag of data is regulated according to the specified data update rate, because occasional irregular time jumps in the original data cause an unrealistic and erroneous velocity and acceleration data profile.

Outliers are then filtered by checking standard deviations and means of the position data with a moving window of 20 points. The filtered data is used as the position truth file for this research. Figure 6.5 shows an example of the collected data and the generated position truth data around Norwich station. An overview of the generated position truth is provided in Figure 6.6. The top figure shows a truth trajectory in the form of latitude and longitude in degrees, and the progress of time is marked on the trajectory. As can be seen from the figure, the train is mainly moving approximately along the east direction with a s-shape turning between 20 min and 25 min. The bottom figure shows the height information plotted against time.

The time derivatives of the geodetic coordinates ($\dot{L}$, $\dot{\lambda}$ and $\dot{h}$) are used to produce the velocity file,




**(a)** GPS data collected around Norwich station




**(b)** Generated position truth data around Norwich station

**Figure 6.5.:** An example of collected data and generated position truth

as given by (5.22) to (5.24). The calculated velocity is the velocity truth in the local navigation frame, and consists of velocity in the north, east and down directions. First-order numerical differentiation is performed on the position truth file to calculate the derivatives. The velocity file can therefore be obtained by rewriting (5.22) to (5.24) as

$$v^n_{eb,N} = \dot{L}(R_N(L) + h) \tag{6.1}$$

$$v^n_{eb,E} = \dot{\lambda}(R_E(L) + h)\cos L \tag{6.2}$$

$$v^n_{eb,D} = -\dot{h} \tag{6.3}$$

The generated velocity truth is shown in Figure 6.7, where the velocities in the directions of north, east and down are respectively plotted against time.

The truth of the acceleration in the local navigation frame can then be generated by performing first-order numerical differentiation on the velocity truth data. The generated acceleration truth is shown in Figure 6.8, where the accelerations in the directions of north, east and down are respectively plotted against time.

In order to simulate the inertial measurements, the position truth is interpolated into the simulated IMU update rate, 100 Hz as suggested in (Rockwell Collins, 2008), to produce high update rate position truth data. The high-update-rate velocity and acceleration truth data are then calculated using the same procedure as described in previous paragraphs.

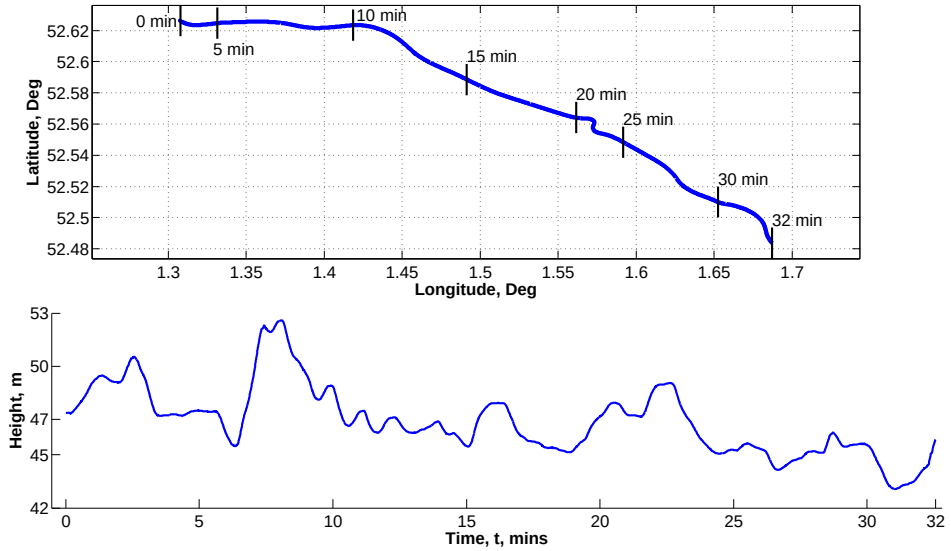



**Figure 6.6.:** The generated position truth

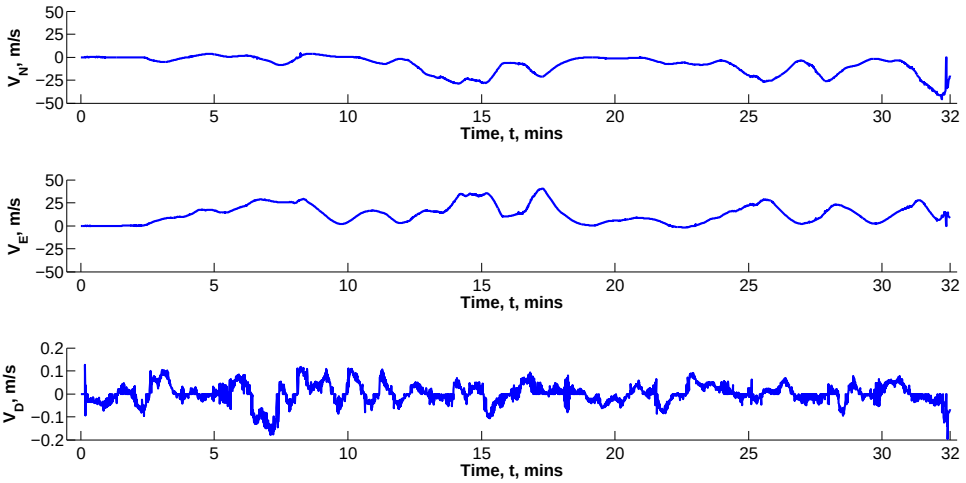



**Figure 6.7.:** The generated velocity truth in local navigation frame

Finally, all position data are transformed to the earth frame in WGS84, and eastings and northings in OSGB36.

## 6.5. DRM Construction

The digital route mode (DRM) is a series of points along the rail route with a fixed distance between each other, the position information of which is known. The DRM used for this research is built based on the position truth file obtained in Section 6.4. A time-tagged profile associating each individual DRM point with the progress of the simulation is also generated during the construction of the DRM. The time-tagged DRM profile contains information on DRM points that are in front of and behind each position truth point at the specific time. As a result, the need for a DRM searching algorithm is avoided.

The building of the DRM is essentially an iterative searching algorithm performed along a fitted

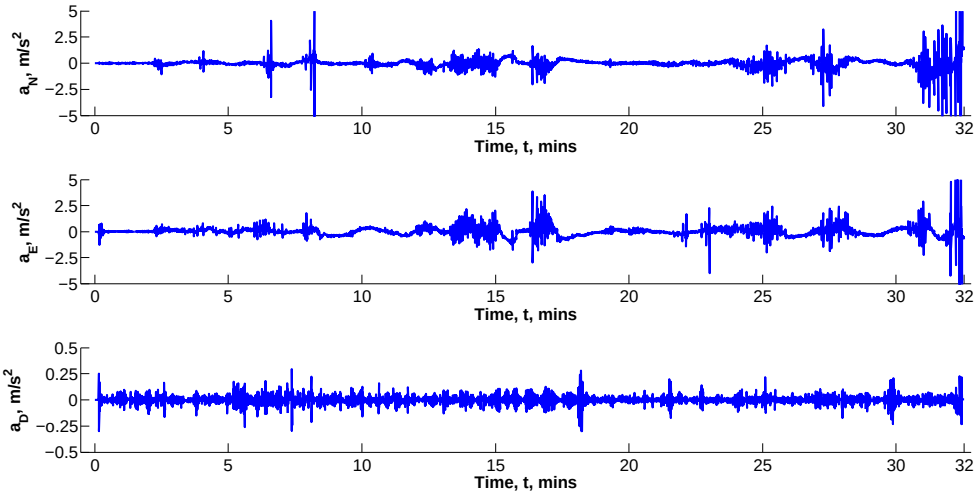
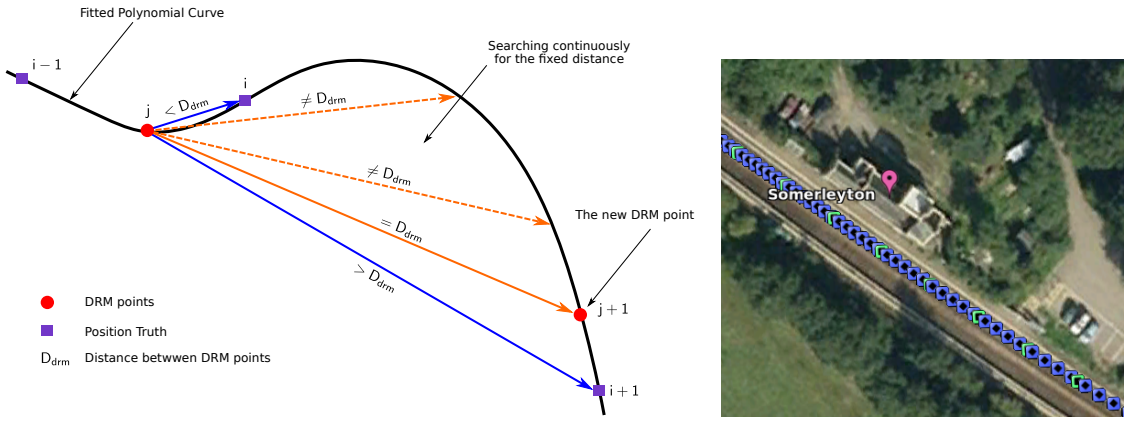



**Figure 6.8.:** The generated acceleration truth in the local navigation frame




**(a)** An illustration of searching for the next DRM point

**(b)** An example of the DRM point around Somerleyton station

**Figure 6.9.:** The construction of DRM with the position truth data

polynomial curve. Figure 6.9a is an illustration of the searching process using the generated position truth data as a reference. The position truth data is firstly divided into sections. A 2D tenth-order polynomial curve is fitted to the eastings and northings of the position truth data in each section. The reason for using a tenth-order polynomial is that polynomial fitting for a set of discrete data can simulate the smooth changes in curvature when the tracks change from straight lines to circular curves and tenth-order polynomials presented the minimum root mean square fitting errors for this particular data set. The search begins with the last known DRM point ($j$) and the last position truth point that is behind the known DRM ($i-1$). The fixed distance between two DRM points is referred as $D_{drm}$. The distances between $j$ and the position truth points from $i$ onwards are calculated and compared with $D_{drm}$. When the distance found is longer than $D_{drm}$ at point $i+1$, the search for the next DRM point ($j+1$) starts with $i$ and moves along the fitted polynomial curve at small steps. The search stops when the distance is calculated equal to $D_{drm}$ and point $j+1$ is found. The height information of $j+1$ is interpolated from height information of $i$ and $i+1$.

Figure 6.9b shows an example of the DRM data and position truth data, where the green-square marks represents the DRM points with $D_{drm} = 10\,\text{m}$ and the blue-square marks are the position truth points. The first position truth point is used as the first known DRM point in order to start

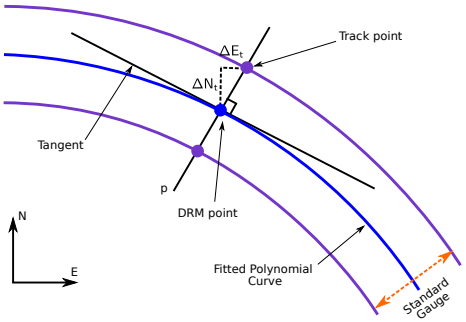
the iteration. This process generates a truth data for the simulated DRM.

Several DRM truth data sets are generated using different $D_{drm}$. The $D_{drm}$ values used in this research are 10 m, 50 m, 100 m and 500 m. A practical DRM is most likely to be constructed by surveying at a specified accuracy level. Different levels of errors are thus added to the DRM truth data to generate the simulated DRM data. The added errors are assumed to be zero-mean normally distributed, with different standard deviation levels at 0.1 m, 1 m and 10 m.

## 6.6. Attitude Information

### 6.6.1. Rail Tracks

The rail tracks are simulated using the polynomial curves and DRM truth data obtained in Section 6.5. The gauge of the simulated tracks ($D_{gauge}$) is assumed to be the standard gauge, which is the UK railways standard and $D_{gauge} = 1,435$ mm (Jenkins, 1960).




**(a)** An illustration of simulating rail tracks


**(b)** An example of simulated rail tracks around Haddiscoe station

**Figure 6.10.:** The simulation of rail tracks

The basic algorithm used in the simulation of rail tracks is illustrated in Figure 6.10a. The slope of the tangent line ($k_t$) is the first-order derivative of the fitted polynomial curves at the section. The purple curves in the figure represent the simulated tracks. The line $p$ is perpendicular to the tangent line and the two lines intersect at a specific DRM point. The intersecting points of line $p$ and the simulated tracks are the track points to be calculated. The slope of the line $p$ , $k_p$, can

therefore be calculated with

$$k_p = -\frac{1}{k_t} \tag{6.4}$$

The difference between the track point coordinates and the DRM point coordinate in eastings and northings is obtained using

$$\Delta E_t = \frac{1}{2}\cos(\arctan(k_p))D_{gauge} \tag{6.5}$$

$$\Delta N_t = \frac{1}{2}\sin(\arctan(k_p))D_{gauge} \tag{6.6}$$

and the tracks points ($E_t$, $N_t$) can be found by applying the difference to the coordinate of the DRM point ($E_{drm}$, $N_{drm}$)

$$E_t = E_{drm} \pm \Delta E_t,\ N_t = N_{drm} \pm \Delta N_t \tag{6.7}$$

Figure 6.10b shows an example of the generated track points with corresponding DRM points near Haddiscoe station. The yellow square marks in the figure represent the generated tracks points, whereas the green square marks are the DRM points. The area circled in the figure is enlarged to the points more clearly.

### 6.6.2. Superelevation

It is a common practice on the railways to raise the outer rail above the inner one when the route is a curve, and the raised height is called superelevation or 'cant'. The cant is constant on a curve which is the arc of a true circle. As the rails are at the same level on a straight route, a length of track over which the outer rail is gradually raised from zero to the height of the cant needs to be set up to connect the two, known as transition curve.

The superelevation is the main cause for the roll movement (see Appendix B for descriptions on roll) of the train carriage. The simulation of the superelevation is based on DRM truth points obtained in Section 6.5. The first step is to find the location of all circular curves along the route. Each section of the circular curve and its transition curves are manually identified along the route. A circle is then fitted to each section using DRM points within the section, and its radius ($R_c$) is recorded. This process is shown in Figure 6.11. The line consisting of red crosses is the DRM truth data. Black square marks along the line are the start and/or end point of each circular curve and transition curve. As can be seen from the figure, circles drawn with blue lines are fitted to the DRM data along the route. The area marked with a purple square is enlarged to show a clearer example.

The design of the superelevation is usually related to the maximum speed ($V_{max}$) allowed on a route. For this research, the maximum speed of the velocity profile obtained in Section 6.4 is used. The superelevation is calculated as following (Jenkins, 1960),

$$\text{cant} = \frac{D_{gauge}V_{max}^2}{gR_c} \tag{6.8}$$

where $g$ is the acceleration due to gravity. An interpolation is performed between the superelevation and the height of the levelled tracks to produce the height increase of a transition curve. The calculated superelevation information is then assigned to the relative track.

Figure 6.12a shows the height change between rails along the route during the simulation, where zero means the route is straight. Figure 6.12b shows an example of track points plotted with simulated superelevation information. The area shown in the picture is marked by purple squares in both the route overview and the superelevation plot in Figure 6.12a. Yellow square markers are

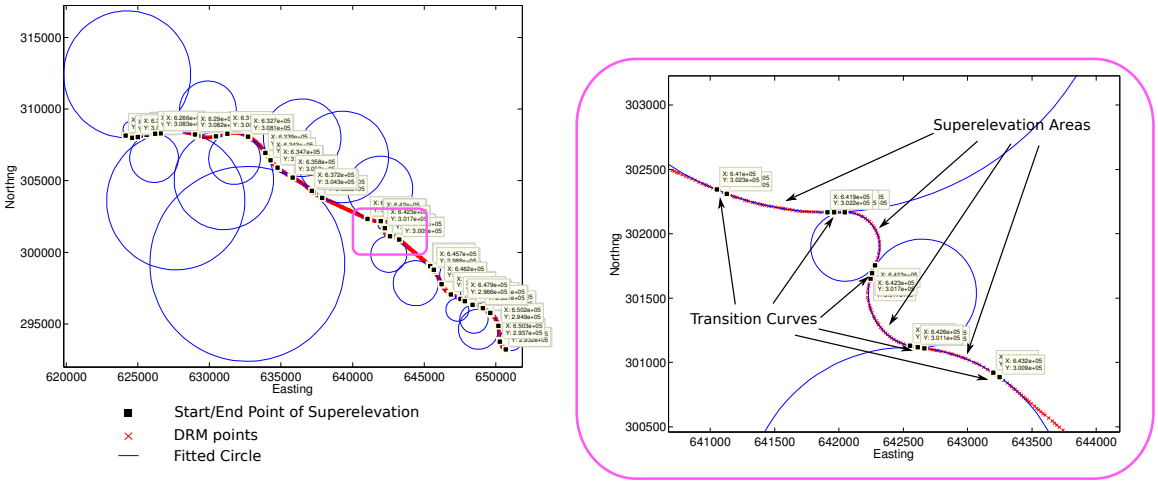
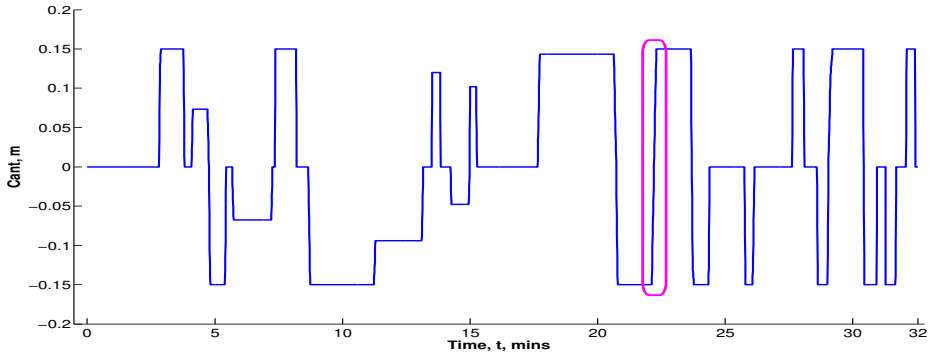
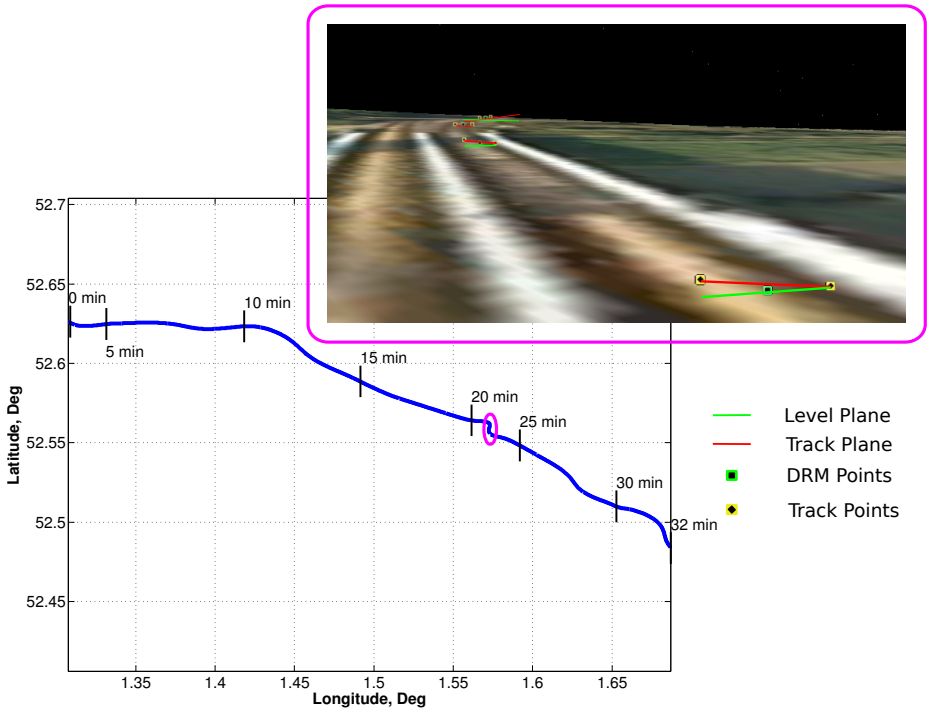



**Figure 6.11.:** The fitted circular curves along the rail route




**(a)** The simulated superelevation along the rail route




**(b)** An example of simulated superelevation

**Figure 6.12.:** The simulation of the superelevation along the route

the track points, whereas the green-square marks are the DRM truth points. The red lines in the picture mark out the track plane, and the green lines indicate the levelled plane.

### 6.6.3. Euler Angles and Direction Cosine Matrices

The attitude information is obtained firstly in the form of Euler Angles (see Appendix B for definitions of Euler Angles). The three components of the Euler Angles, roll, pitch and yaw, are calculated using information gained in previous sections.

The roll angle, $\phi_{nb}$, represents the tilting movement of the carriage about the $x$-axis of its body frame. The roll rotation of the carriage is mainly because of the superelevation along the route, which raises the outer rail and cause the tilting of the carriage. Therefore, the simulation for the roll angle of the carriage is based on the superelevation information obtained in Section 6.6.2. The calculation is based on the trigonometric relations of a right angle triangle, and the equation is written as

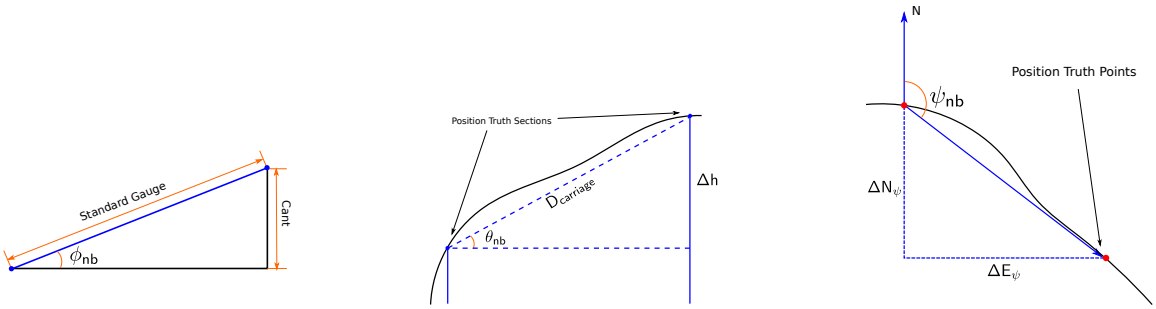
$$\phi_{nb} = \arcsin\left(\frac{\text{cant}}{D_{gauge}}\right) \tag{6.9}$$

Figure 6.13a illustrates the triangle of the tracks and the superelevation.

The pitch angle, $\theta_{nb}$, represents the change of the carriage attitude about the $y$-axis of its body frame, i.e. the changes of its elevation. The pitch rotation is mainly caused by the height change along the route. The carriage is a rigid body with fixed length and is assumed to move on a straight line between sections of tracks with different height. The route is firstly divided into small sections and the height changes in each section are then calculated. Thus the pitch angle data is calculated from the height change between adjacent position truth sections, as shown in Figure 6.13b. The equation used is written as

$$\theta_{nb} = \arcsin\left(\frac{\Delta h}{D_{carriage}}\right) \tag{6.10}$$

where $\Delta h$ is the height change in a section and $D_{carriage} = 20\,\text{m}$ (First Great Western, 2010) is the fixed length of the carriage.




**(a)** Roll **(b)** Pitch **(c)** Yaw

**Figure 6.13.:** Simulating Euler angles

The yaw angle, $\psi_{nb}$, is also known as heading in the local navigation frame. It represents the rotation about the $z$-axis of the carriage body frame. The carriage is assumed to move on a straight line between adjacent position truth points, and the angle between the movement vector and the North direction is the yaw angle. Figure 6.13c illustrates the calculation process. The coordinate difference of adjacent points, $\Delta E_\psi$ and $\Delta N_\psi$, is calculated along the route. The yaw angle is then calculated using

$$\psi_{nb} = \arctan 2(\Delta E_\psi,\ \Delta N_\psi) \tag{6.11}$$

where arctan 2 is the four-quadrant inverse tangent.

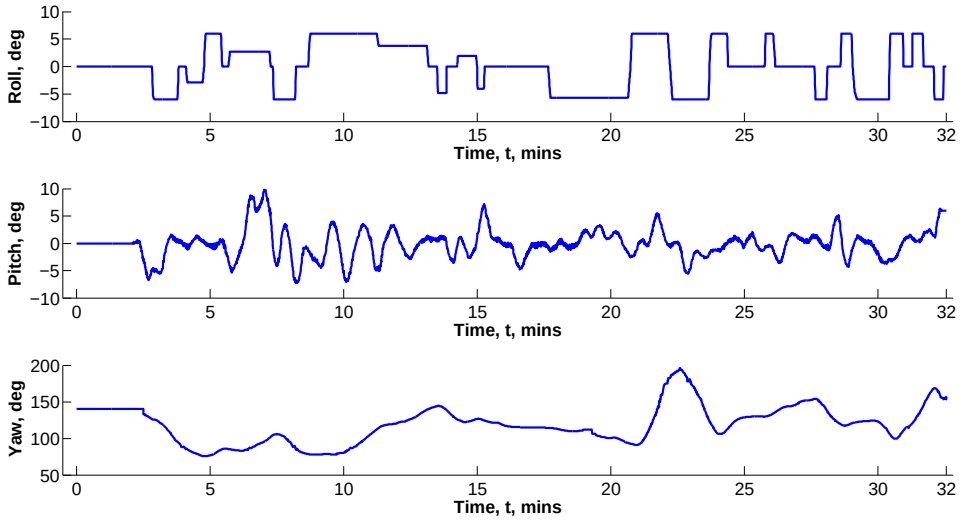



**Figure 6.14.:** The generated Euler Angle truth data

The update rate of the Euler angle truth data is 2 Hz. In order to simulate the high-update-rate IMU output, the truth data is interpolated to 100 Hz. Figure 6.14 shows the simulated Euler angle truth, which is plotted against time as the simulation progresses. From the top graph to the bottom one, roll, pitch and yaw are respectively plotted. The effect of the superelevation change can be seen clearly from the roll angle, whereas the changes of yaw correspond to changes of the trajectory.

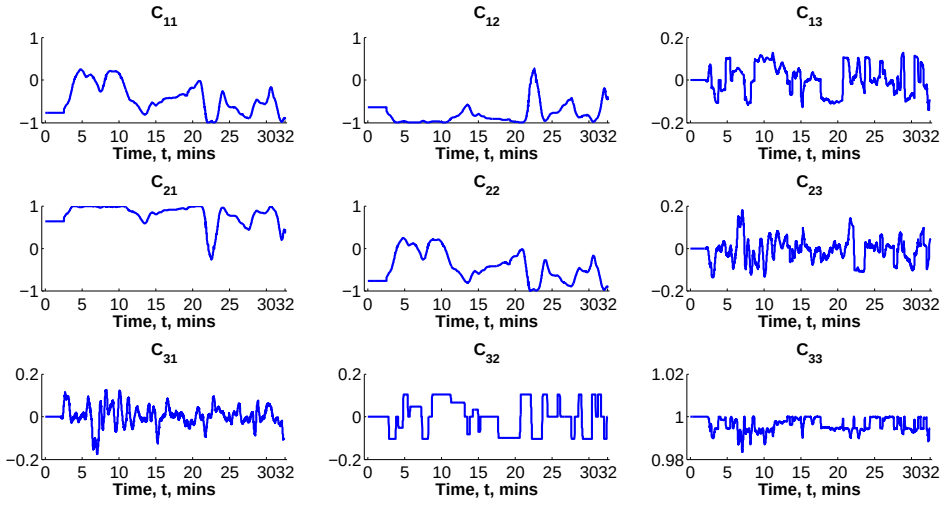



**Figure 6.15.:** The generated direction cosine matrix $\mathbf{C}_b^n$ data

Obtained Euler Angle data is also transformed into direction cosine matrices, $\mathbf{C}_n^b$ and $\mathbf{C}_b^n$, with Equation (B.10) and Equation (B.6) for the use of other simulations. Figure 6.15 shows the direction cosine matrix $\mathbf{C}_b^n$ data plotted against time. $C_{ij}$ denotes the $j$th elements of the $i$th row of the direction cosine matrix $\mathbf{C}_b^n$.

## 6.7. Angular rate

The angular rate measured by the gyro is the rotation of the body frame with respect to the inertial frame (see Section 5.1). The Euler angle simulated in Section 6.6.3, however, only represents rotations from the body frame to the local navigation frame. Angular rates caused by the relative rotations among the local navigation frame, the Earth frame and the inertial frame need to be calculated in order to get the gyro measurement. The simulation process of the angular rate measurement vector $\boldsymbol{\omega}_{ib}^{b}$ thus contains two steps.

The angular rate caused by the rotation between the body frame and the local navigation frame, $\boldsymbol{\omega}_{nb}^{b}$, is calculated in the first step. Although the Euler angle represents the frame transformation between the body frame and the local navigation frame, its time derivative cannot be related to $\boldsymbol{\omega}_{nb}^{b}$ directly. This is because the frame transformation performed through Euler angle is a series of rotations with each component of the Euler angle defining a different intermediate reference frame. The Euler derivative, $(\dot{\phi}_{nb},\, \dot{\theta}_{nb},\, \dot{\psi}_{nb})$, is related to $\boldsymbol{\omega}_{nb}^{b}$ with the equation following Farrell (2008),

$$\boldsymbol{\omega}_{nb}^{b} = \begin{bmatrix} 1 & 0 & -\sin\theta_{nb} \\ 0 & \cos\phi_{nb} & \sin\phi_{nb}\cos\theta_{nb} \\ 0 & -\sin\phi_{nb} & \cos\phi_{nb}\cos\theta_{nb} \end{bmatrix} \begin{bmatrix} \dot{\phi}_{nb} \\ \dot{\theta}_{nb} \\ \dot{\psi}_{nb} \end{bmatrix} \tag{6.12}$$

The first-order numerical differentiation is performed on the Euler angle data to obtain its time derivative, $(\ddot{\phi}_{nb},\, \ddot{\theta}_{nb},\, \ddot{\psi}_{nb})$. $\boldsymbol{\omega}_{nb}^{b}$ is calculated with (6.12).

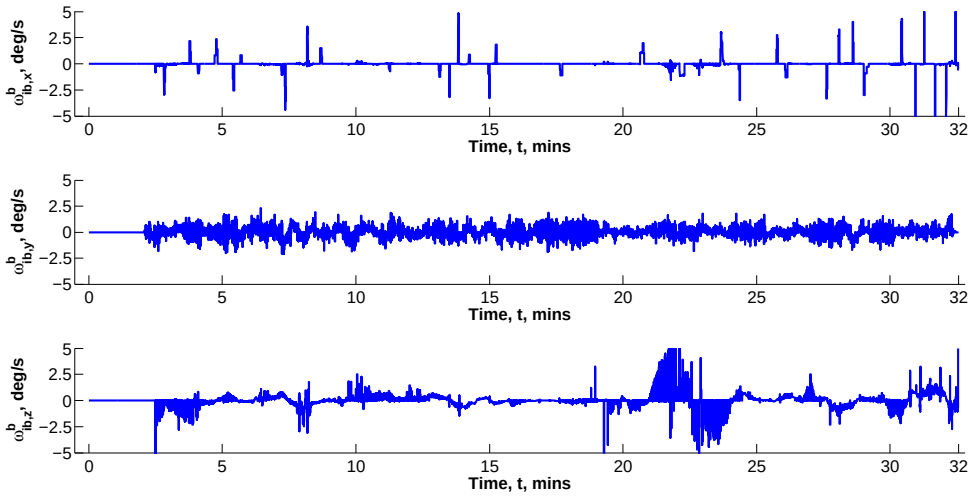



**Figure 6.16.:** The generated angular rate truth data

The second step is to add components in $\boldsymbol{\omega}_{ib}^{b}$ that are caused by relative movement between the Earth frame, the local navigation frame and the inertial frame. These components include the Earth rotation rate, $\boldsymbol{\omega}_{ie}^{n}$, defined by (5.11) and the transport rate, $\boldsymbol{\omega}_{en}^{n}$, defined by (5.13). The calculation of $\boldsymbol{\omega}_{ie}^{n}$ and $\boldsymbol{\omega}_{en}^{n}$ requires information on the position and velocity of the train, which is provided through simulations in Section 5.2. The angular rate measurement $\boldsymbol{\omega}_{ib}^{b}$ can therefore be calculated by rearranging the equation given in Titterton & Weston (2004) as

$$\boldsymbol{\omega}_{ib}^{b} = \boldsymbol{\omega}_{nb}^{b} + \mathbf{C}_{n}^{b}(\boldsymbol{\omega}_{ie}^{n} + \boldsymbol{\omega}_{en}^{n}) \tag{6.13}$$

where the direction cosine matrix $\mathbf{C}_n^b$ is generated in Section 6.6.3.

The generated angular rate data is plotted against time shown in Figure 6.16. The components of $\boldsymbol{\omega}_{ib}^b$ plotted in the figure are the angular rates about three axes of the body frame respectively. The update rate for the generated angular rate data is 100 Hz.

## 6.8. Specific Force

Similar to the simulation of the angular rate, the specific force simulation is also divided into two steps.

The first step is to calculate the specific force in the local navigation frame, $\mathbf{f}_{ib}^n$. The acceleration of the carriage resolved in the local navigation frame, $\dot{\mathbf{v}}_{eb}^n$ , is defined by (5.18). As can be seen from (5.18), $\mathbf{f}_{ib}^n$ is only part of $\dot{\mathbf{v}}_{eb}^n$. Other parts include the gravity and the centrifugal part, information on the position, velocity and acceleration of the carriage is needed to remove them. Thus, in order to obtain $\mathbf{f}_{ib}^n$, (5.18) is rearranged as

$$\mathbf{f}_{ib}^n = \dot{\mathbf{v}}_{eb}^n - \mathbf{g}_b^n(L,h) + (\boldsymbol{\Omega}_{en}^n + 2\boldsymbol{\Omega}_{ie}^n)\mathbf{v}_{eb}^n \tag{6.14}$$

where the gravity, $\mathbf{g}_b^n(L,h)$, the Earth rotation rate, $\boldsymbol{\Omega}_{ie}^n$, and the transport rate, $\boldsymbol{\Omega}_{en}^n$, are respectively defined by (5.19), (5.12) and (5.14).

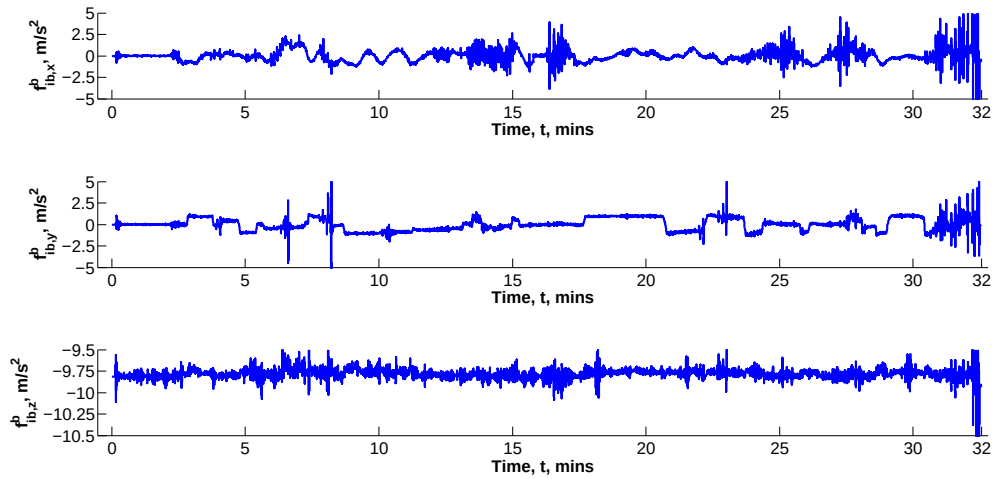



**Figure 6.17.:** The generated specific force truth in the body frame

The second step is to transform the reference frame of $\mathbf{f}_{ib}^n$ into the carriage body frame. This can be achieved using

$$\mathbf{f}_{ib}^b = \mathbf{C}_n^b \mathbf{f}_{ib}^n \tag{6.15}$$

where the direction cosine matrix for transformations from the local navigation frame to the body frame, $\mathbf{C}_n^b$, is calculated using obtained attitude information, as described in Section 6.6.3.

The generated specific force data comes with an update rate of 100 Hz. Figure 6.17 shows three components of $\mathbf{f}_{ib}^b$ plotted against time.

## 6.9. IMU Measurements

The generation of the truth data for specific force and angular rate was described respectively in Section 6.7 and Section 6.8. The evaluation of the IMU output is described in this section, based on adding error characteristics to the truth data. The error characteristics and general error model for inertial sensors are discussed in Section 5.3. The error characteristics simulated for this research are biases and random noises. This is because they exhibit the most significant effect on the sensor output and are the most observable for most integrated systems (see the discussion in Section 3.3 for state selection and observability of an integrated system).

Table 6.1 shows the error characteristics used for the simulated inertial sensor output. The turn-on bias and in-run bias figures are based on the commercial IMU SiIMU02 (Atlantic Inertial Systems, 2008), whereas the random noise figure is based on another commercial IMU MIDG IIC (Mirobotics Inc., 2009). All figures are listed in SI units, and the original quotations are given in parentheses. The transformation of units follows information given in Section 5.3.

**Table 6.1.:** Simulated MEMS IMU performance characteristics

| | Turn-on bias ($1\sigma$) | In-run bias ($1\sigma$) | Random noise ($1\sigma$) |
|---|---|---|---|
| Accelerometer | $9.8 \times 10^{-2}\,\mathrm{m/s^2}$ ($10\,\mathrm{mg}$) | $4.9 \times 10^{-3}\,\mathrm{m/s^2}$ ($0.5\,\mathrm{mg}$) | $1.4 \times 10^{-2}\,\mathrm{m/s^{1.5}}$ ($150\,\mu\mathrm{g}/\sqrt{\mathrm{Hz}}$) |
| Gyro | $4.8 \times 10^{-4}\,\mathrm{rad/s}$ ($100\,^\circ/\mathrm{hr}$) | $3.15 \times 10^{-5}\,\mathrm{rad/s}$ ($6.5\,^\circ/\mathrm{hr}$) | $1.35 \times 10^{-10}\,\mathrm{rad/s^{0.5}}$ ($0.1\,^\circ/\mathrm{s}/\sqrt{\mathrm{Hz}}$) |

The general error model given in Section 5.3.5 is therefore simplified, and is written here as

$$\tilde{\mathbf{f}}_{ib}^{b} = \mathbf{f}_{ib}^{b} + \mathbf{b}_a + \mathbf{w}_a \tag{6.16}$$

$$\tilde{\boldsymbol{\omega}}_{ib}^{b} = \boldsymbol{\omega}_{ib}^{b} + \mathbf{b}_g + \mathbf{w}_g \tag{6.17}$$

where terms in equations are defined as in Section 5.3.

Because of the existence of in-run biases, the biases for both accelerometers and gyros are modelled as

$$\dot{\mathbf{b}}_a = \mathbf{w}_{ad} \tag{6.18}$$

$$\dot{\mathbf{b}}_g = \mathbf{w}_{gd} \tag{6.19}$$

where $\mathbf{w}_{ad}$ and $\mathbf{w}_{gd}$ are zero-mean Gaussian white-noise processes caused by in-run biases. A discrete model of the bias is derived based on Crassidis (2006), and is written as

$$\mathbf{b}_a(+) = \mathbf{b}_a(-) + n_{ad}\mathbf{N}_{ad} \tag{6.20}$$

$$\mathbf{b}_g(+) = \mathbf{b}_g(-) + n_{gd}\mathbf{N}_{gd} \tag{6.21}$$

where $n_{ad}$ and $n_{gd}$ are the square root PSD for in-run biases listed out in Table 6.1, $\mathbf{N}_{ad}$ and $\mathbf{N}_{gd}$ are zero-mean random variables with unit variances. Note that $(+)$ and $(-)$ are generalised to represent epoch $k$ and $k-1$ respectively. The simulation process starts with the known fixed turn-on biases as listed in Table 6.1.

The simulated biases are shown in Figure 6.18 for accelerometers and Figure 6.19 for gyros. Biases shown in the figures are all plotted against time.

The discrete model derived here for simulated sensor output follows notes in Crassidis (2006), and

is written as

$$\tilde{\mathbf{f}}_{ib}^{b}(+) = \mathbf{f}_{ib}^{b}(+) + \frac{1}{2}\left(\mathbf{b}_a(+) + \mathbf{b}_a(-)\right) + \left[\frac{n_a^2}{\tau_i} + \frac{1}{12}n_{ad}^2\right]^{\frac{1}{2}} \mathbf{N}_a \tag{6.22}$$

$$\tilde{\boldsymbol{\omega}}_{ib}^{b}(+) = \boldsymbol{\omega}_{ib}^{b}(+) + \frac{1}{2}\left(\mathbf{b}_g(+) + \mathbf{b}_g(-)\right) + \left[\frac{n_g^2}{\tau_i} + \frac{1}{12}\Phi n_{gd}^2\right]^{\frac{1}{2}} \mathbf{N}_g \tag{6.23}$$

where $n_a$ and $n_g$ are the root PSD for random noise on the accelerometers and gyros respectively, as listed in Table 6.1. The generated IMU measurements are shown in Figure 6.20 for specific force and Figure 6.21 for angular rate respectively. Measurement data is drawn in red lines, and truth data are plotted in blue for comparison.

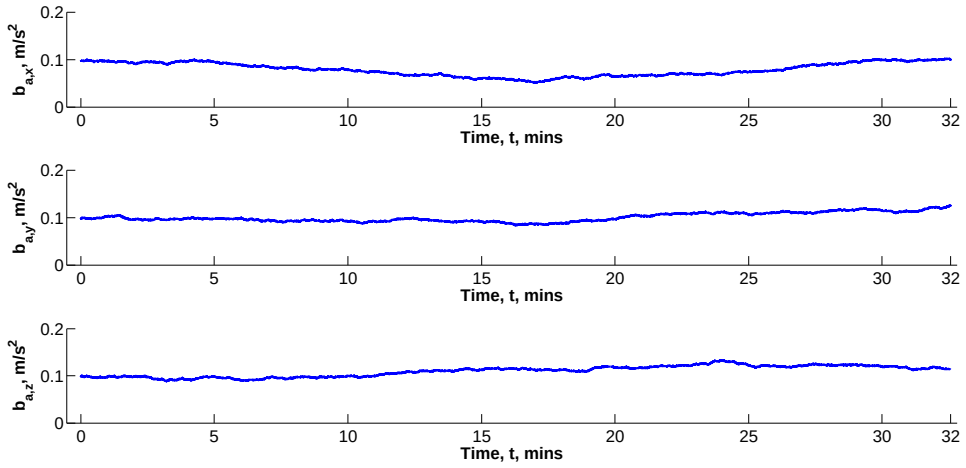



**Figure 6.18.:** The generated bias truth data for accelerometers

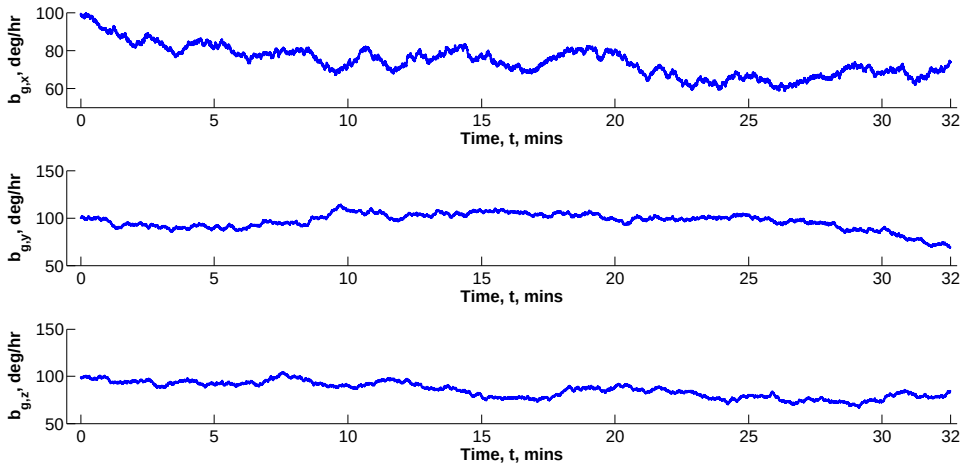



**Figure 6.19.:** The generated bias truth data for gyros

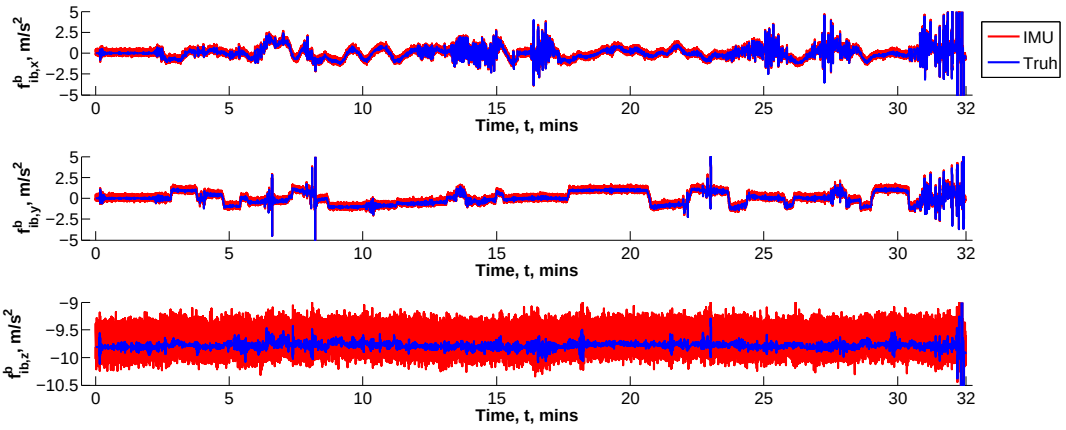



**Figure 6.20.:** The generated specific force measurement

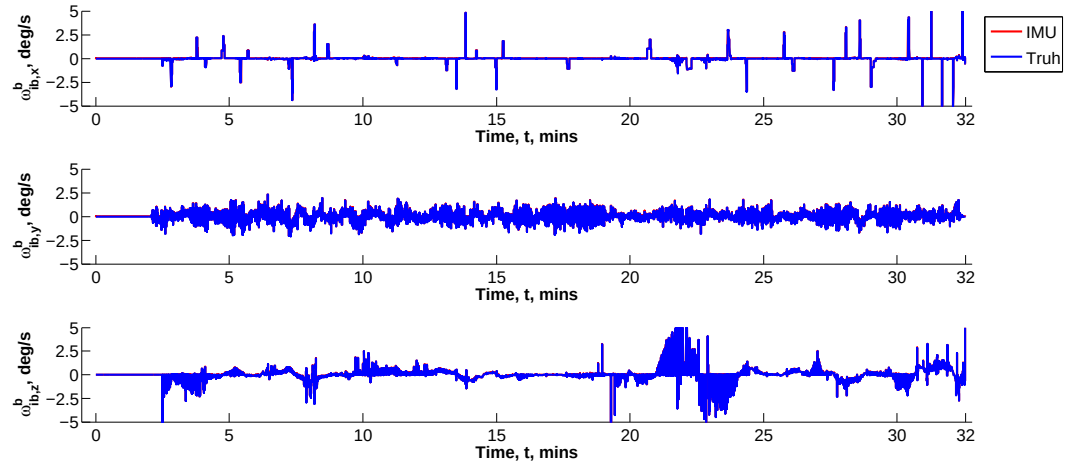



**Figure 6.21.:** The generated angular rate measurement

## 6.10. Corrected GPS Pseudorange

### 6.10.1. Ephemeris

In order to simulate the pseudorange measurements, satellite ephemerides are required for realistic satellite orbit positions. The IGS final product ephemerides are used in this research for GPS constellation information during the simulation. The ephemeris file is downloaded from the IGS data products server.

The downloaded ephemerides are in the form of the national geodetic survey standard GPS format SP3. The information in the ephemerides includes constellation status and Cartesian coordinates for satellites. The satellite coordinates are expressed in the ITRF2005/IGS05 reference frame. As there are no official transformation parameters between new realisations of WGS84 and the ITRF2005, the coordinates are considered to be expressed in WGS84 at 10 cm level (ITRF, 2007; NGA, 1997), and a frame transformation is not performed.

The downloaded ephemeris covers the time period from 00:00 to 23:45 on 21st February 2009, with an update interval of 15 min. Since the simulation lasts 32 min, only part of the ephemeris file is used. The satellite positions are then interpolated to the simulated GPS update rate, 0.5 s

(2 Hz). The satellite velocities are obtained by performing a first-order numerical differentiation on the position data. The generated satellite position and velocity data are considered to be the truth file for this simulation.

### 6.10.2. Range, Range Rate and Line-of-Sight Vector

The time tag of the generated satellite position and velocity data is firstly synchronised with the train movement data generated in Section 6.4.

The true geometric range between the $k$th satellite and the receiver, $r^{(k)}$, is calculated using (4.3). The carriage position data is used as the receiver position. The generated range is plotted against time in Figure 6.22. Note that only satellites visible from the carriage position are plotted. The actual selection of visible satellites needs information on satellite elevation, which is calculated as stated in Section 6.10.3.

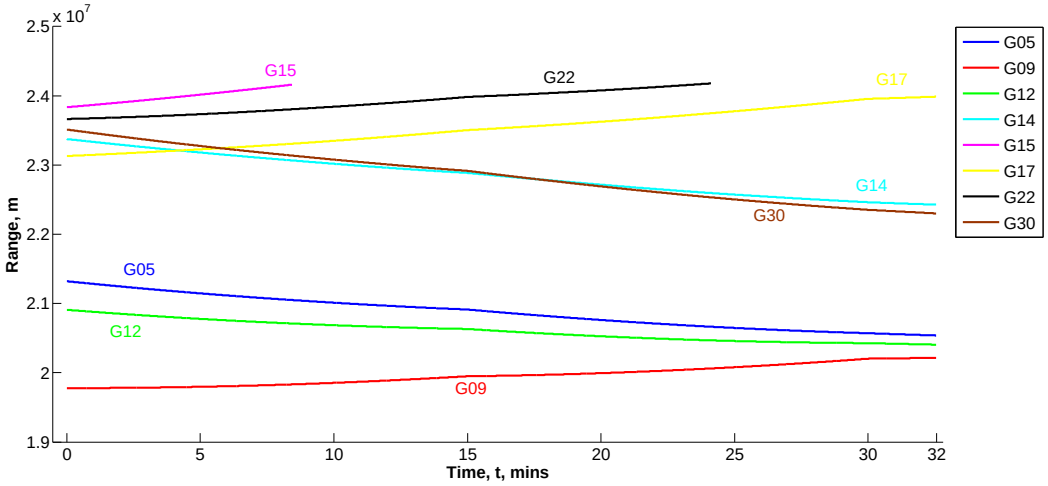



**Figure 6.22.:** The generated satellite-to-receiver range data

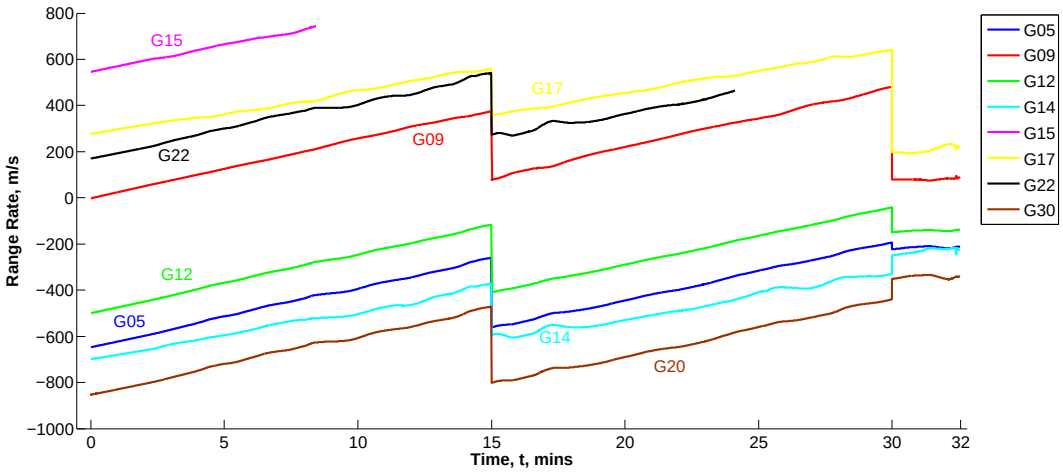



**Figure 6.23.:** The generated satellite-to-receiver range rate data

Velocity data obtained in Section 6.4 is in the local navigation frame. In order to calculate the range rate, the velocity truth data is firstly transformed from the local navigation frame into the

Earth frame using

$$\mathbf{v}_{eb}^{e} = \mathbf{C}_{n}^{e}\mathbf{v}_{eb}^{n} \tag{6.24}$$

where $\mathbf{v}_{eb}^{e}$ is the velocity vector resolved in the Earth frame, and $\mathbf{C}_{n}^{e}$ is the direction cosine matrix defined by (A.2).

The line-of-sight vector for the $k$th satellite in the Earth frame, $\mathbf{1}^{(k),e}$, is defined in (4.7), and is rewritten here as

$$\mathbf{1}^{(k),e} \approx \frac{(\mathbf{p}^{(k),e} - \mathbf{p}^{e})}{\| \mathbf{p}^{(k),e} - \mathbf{p}^{e} \|} \tag{6.25}$$

where the subscript $e$ represents the Earth frame.

The range rate is then calculated using

$$\dot{r}^{(k)} = (\mathbf{v}^{(k),e} - \mathbf{v}_{eb}^{e}) \cdot \mathbf{1}^{(k),e} \tag{6.26}$$

The generated range rate is plotted against time in Figure 6.23. Note that the jumps in the generated data are caused by the linear interpolation process of the satellite orbit position, and are not related to any physical changes of the status of both satellites and the train carriage. A more effective interpolation method for the satellite orbit position and velocity is to use a high order polynomial.

### 6.10.3. Elevation and Azimuth

The direction of a satellite with respect to the user antenna is defined by its elevation and azimuth. The definition of satellite elevation and azimuth is illustrated in Figure 6.24. The orientation of the satellite line-of-sight vector with respect to the local navigation frame corresponds to the elevation and azimuth angles.

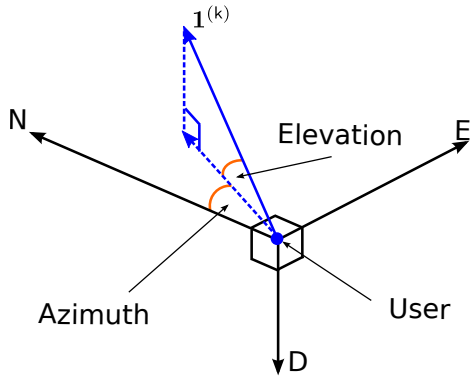



**Figure 6.24.:** Satellite elevation and azimuth

In order to select the visible constellation, elevation and azimuth information of each satellite with respect to the user antenna is necessary. The calculation of satellite elevation and azimuth requires the line-of-sight vector expressed in the local navigation frame, $\mathbf{1}^{n}$. The line-of-sight vector in the Earth frame, $\mathbf{1}^{e}$, is calculated in Section 6.10.2. $\mathbf{1}^{n}$ of each satellite can therefore be obtained by performing a frame transformation,

$$\mathbf{1}^{n} = \mathbf{C}_{e}^{n}\mathbf{1}^{e} \tag{6.27}$$

where $\mathbf{C}_{e}^{n}$ is the direction cosine matrix from the Earth frame to the local navigation frame.

The transformed line-of-sight vector consists of three components, $\mathbf{1}^{n} = [u_N^n,\ u_E^n,\ u_D^n]^T$. The

satellite elevation, $\theta$, and azimuth, $\psi$, can then be calculated using

$$\theta = -\arcsin(u_D^n) \tag{6.28}$$

$$\psi = \arctan 2(u_E^n,\ u_N^n) \tag{6.29}$$

where the four-quadrant inverse tangent must be used for the azimuth.

For the simulation in this research, the cut-off angle of satellite elevation is set to be 15°. Satellites with elevation lower than 15° are therefore not visible for the simulated train carriage. The elevations for the visible constellation during the simulate are plotted against time in Figure 6.25 with the cut-off angle marked by a dashed line. Whilst the rising and falling of each visible satellite can be seen clearly from the figure, two satellites (G15 and G22) fall below the cut-off angle angle during the simulation. The changing of visible satellite azimuth is also plotted in Figure 6.26.

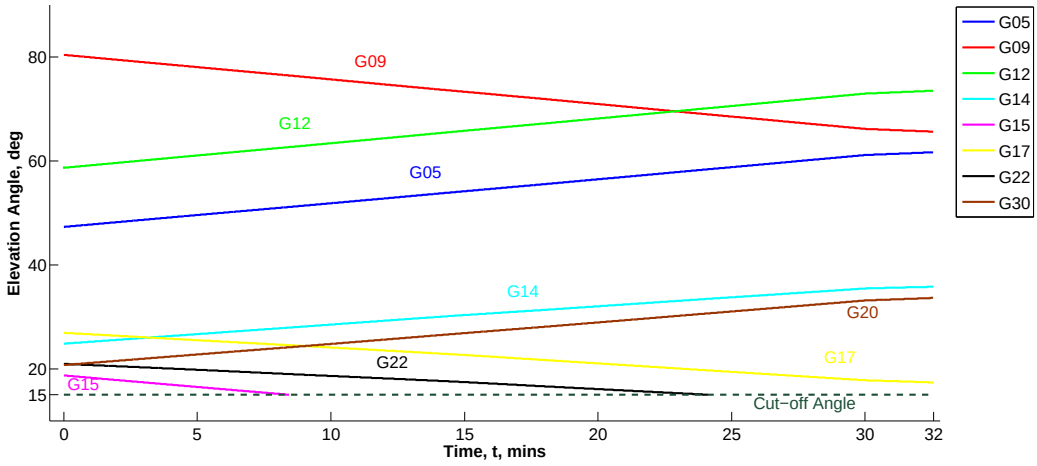



**Figure 6.25.:** The generated satellite elevation angle

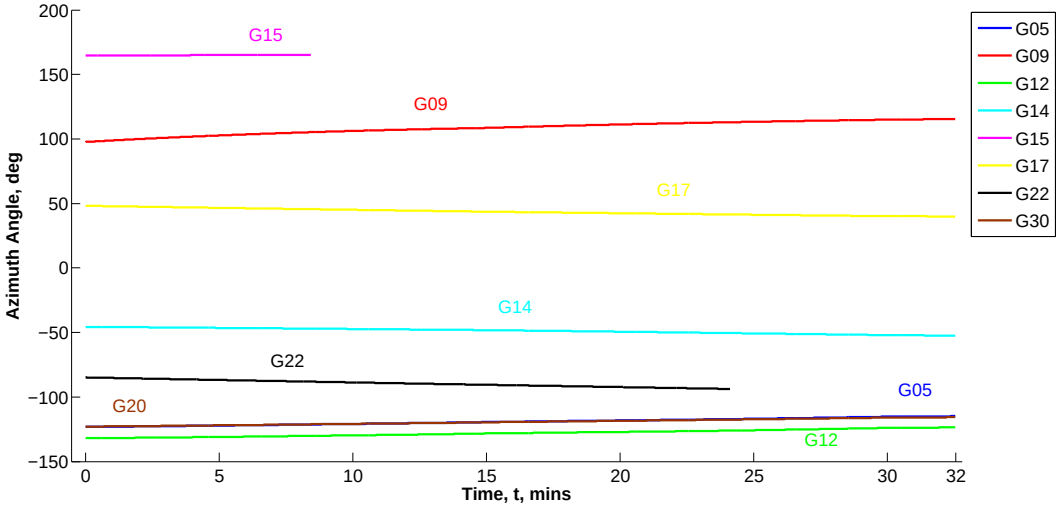



**Figure 6.26.:** The generated satellite azimuth angle

### 6.10.4. Receiver Clock Errors and Correction Residual Errors

A corrected pseudorange measurement is defined in (4.4). As can be seen from this equation, it consists of three main components: the true geometric range, $r^{(k)}$, the receiver clock bias, $b_{rc}$, and the miscellaneous unmodeled range error for the individual satellite plus residuals after applying corrections from the navigation data, $\tilde{\varepsilon}_{\rho}^{(k)}$. Similar to the pseudorange measurement, a pseudorange rate measurement as defined by (4.11) consists of the true range rate, $\dot{r}^{(k)}$, the receiver clock drift, $\dot{b}_{rc}$, and unmodeled errors, $\varepsilon_{\dot{\rho}}^{(k)}$. While the true geometric range and range rate are calculated as in Section 6.10.2, the receiver clock errors and correction residuals need to be simulated to obtain a realistic satellite ranging measurement.

The receiver clock drift is modelled as a random walk process starting with a fixed drift rate, $\dot{b}_{rc}(0)$, following the dynamic model derived in Farrell (2008). The fixed drift rate is chosen as the typical value of an OCXO following Misra & Enge (2006) (see discussion in Section 4.2.5), and $\dot{b}_{rc}(0) = 3\,\mathrm{m/s}$. The relationship between the receiver clock bias and drift is defined in (4.18). Therefore, the model for the receiver clock errors can be written here as

$$\frac{\partial}{\partial t} b_{rc} = \dot{b}_{rc} \tag{6.30}$$

$$\frac{\partial}{\partial t} \dot{b}_{rc} = w_{rc} \tag{6.31}$$

where $w_{rc}$ is a zero-mean Gaussian white-noise process with root PSD $\Phi_{rc} = 0.2\,\mathrm{m/s}^{-1.5}$ (Groves, 2008). A discrete model for the calculation is given as

$$b_{rc}(+) = b_{rc}(-) + \frac{1}{2}\left(\dot{b}_{rc}(+) + \dot{b}_{rc}(-)\right)\tau_i$$

$$\dot{b}_{rc}(+) = \dot{b}_{rc}(-) + \Phi_{rc} N_{rc} \sqrt{\tau_i}$$

where $N_{rc}$ is a zero-mean random variable with unit variance. The generated receiver clock error data is plotted against time and shown in Figure 6.27. The top graph in Figure 6.27 shows the generated receiver clock bias, whereas the bottom one shows the generated receiver clock drift. Note that the artificial reset of the receiver clock, which is common practice for commercial receiver design, is not simulated in this research as it is not part of the natural clock behaviour.

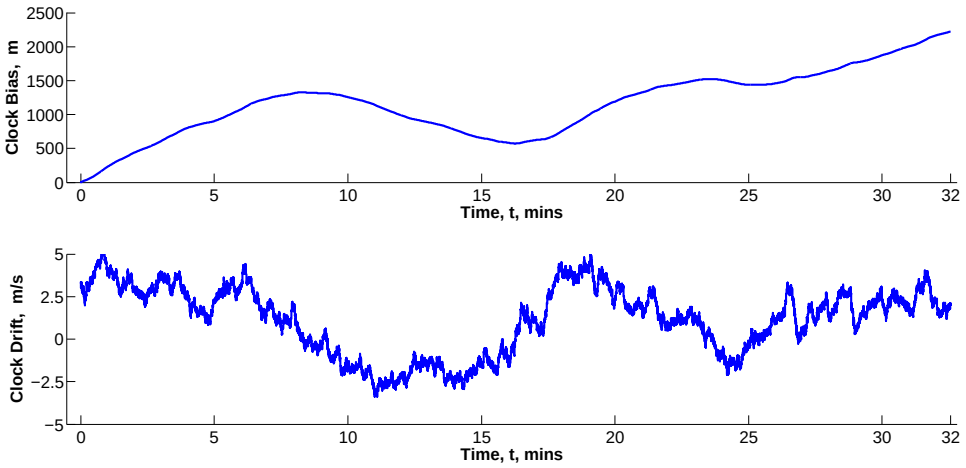



**Figure 6.27.:** The generated receiver clock error data

The main error sources for satellite ranging are discussed in Section 4.3. As can be seen from

Table 4.5, the atmospheric delay including the tropospheric delay and ionospheric delay is the largest contribution of UERE. As the magnitude of the atmospheric delay is related to the satellite elevation, a mapping function is usually applied. A simple mapping function is used in this research following the pattern of (4.33) and (4.35), and is written as

$$M(\theta) = \frac{1}{\sin\theta} \tag{6.32}$$

The zenith residuals for the pseudorange ($\varepsilon_{\rho,z}$) and the pseudorange rate ($\varepsilon_{\dot{\rho},z}$) are assumed to be $\varepsilon_{\rho,z} = 7\,\text{m}$ and $\varepsilon_{\dot{\rho},z} = 10\,\text{m/s}$, as given in Table 4.5. The simulated residual error for each visible satellite is therefore written

$$\varepsilon_\rho = \varepsilon_{\rho,z} M(\theta) N_\rho \tag{6.33}$$

$$\varepsilon_{\dot{\rho}} = \varepsilon_{\dot{\rho},z} M(\theta) N_{\dot{\rho}} \tag{6.34}$$

where $N_\rho$ and $N_{\dot{\rho}}$ are zero-mean random variables with unit variances.

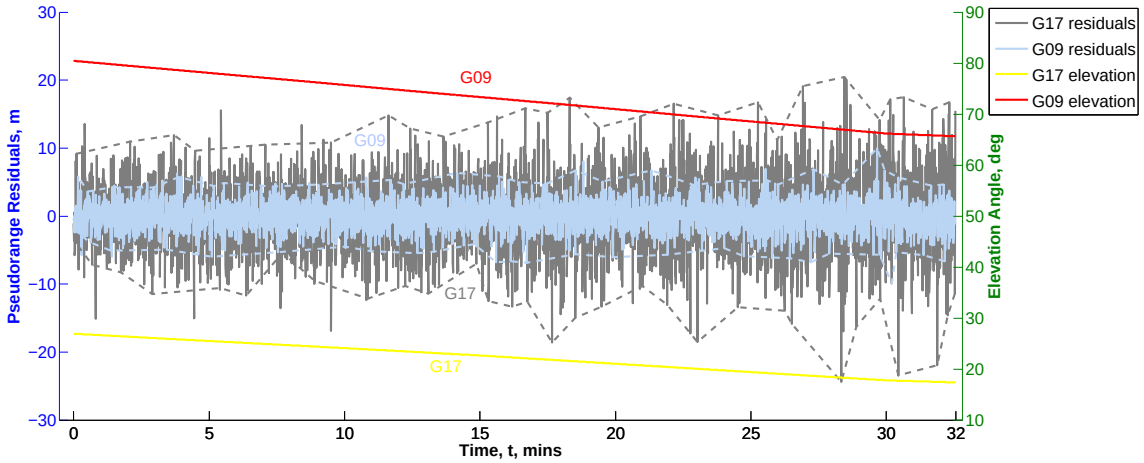



**Figure 6.28.:** An example of simulated residual error against satellite elevation angle

An example of the generated residual errors is shown in Figure 6.28, with the residual errors of satellite G09 and G17 plotted against time together with envelopes. Their elevation angles are also plotted in the figure. As can be seen from the graph, the residual error for satellite G17, with an elevation angle between 15° and 30°, is larger than satellite G09, with an elevation angle between 65° and 80°. In addition, the increase of residual errors as both satellites are setting can be seen from the envelope of the data.

The simulated receiver clock and correction residual errors are then added to the true range and range rate data to produce the generated pseudorange and pseudorange rate measurements.

## 6.11. INS Simulation

A strapdown inertial calculation process is simulated by using the inertial sensor outputs generated in previous sections. Through the simulation, the inertial calculation functions are tested before further research and the performance of the simulated inertial sensors is also evaluated against realistic figures. A detailed introduction to the strapdown inertial calculation process is given in Section 5.4. The calculation mechanism used for the simulation in this research follows the process

discussed in Section 5.4.4.3. A flow chart of the calculation procedure is shown by Figure 5.6.

Since the simulated inertial sensor is assumed to be a low-performance MEMS sensor, only 30 s calculation is simulated. The simulation starts 5 min from the beginning of the overall simulation. The true status of the simulated train is used as the initial solution for the inertial calculation.

The true states of the train attitude, position and velocity during the 30 s are shown in Figure 6.29. As can been seen from the vehicle trajectory in Figure 6.29b, the train is mainly moving in easterly direction. The speed of the train is approximately 16 m/s and it is decelerating, as shown by Figure 6.29c. The change of roll angle in Figure 6.29a indicates that the train just came out of a section of circular curve, with the superelevation on the rails.

Before using the simulated sensor outputs, the simulated true specific force and angular rate are first used as the sensor measurements to test the validity of the inertial processor, and also serves as a performance indication. The simulation result is shown in Figure 6.30. The three plots in Figure 6.30 respectively show the attitude, velocity and position error from the top to the bottom. All errors are plotted against time.

Although the truth data is used, the reason for the errors is mainly numerical, i.e. because of the approximations used in the navigation processor. For instance, position and velocity information from the previous epoch is used for the calculation of the gravity and the earth radius at the current epoch assuming their slow variations over the update interval. This can be shown by the oscillating pattern of errors on the attitude and velocity, where no drifting can be seen. The growth of position error is because the calculation for position using (5.26) to (5.25) is a pure integration process of the velocity, and therefore any velocity errors cause the position error to grow continuously.

The error in the roll angle increases around 25 s and is caused by the process of train going through a transaction curve as shown by the roll angle truth in Figure 6.29a. The velocity error in the easting direction is larger than in other directions. This is due to the fact that the train is mainly moving along the easting direction, and the velocity component in the easting direction includes the most of the velocity vector. The result shown in Figure 6.30 proves the validity of the simulated navigation processor.

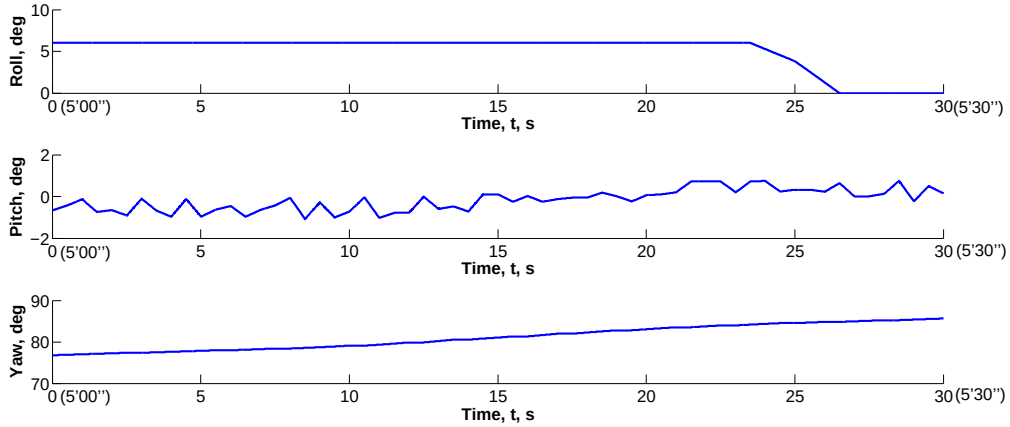
The simulated MEMS sensor measurements are then used for the simulation. The fixed biases as listed in Table 6.1 are applied to the sensor measurement at each epoch by rewritten (5.8) and (5.9) as

$$\hat{\mathbf{f}}_{ib}^{b} = \tilde{\mathbf{f}}_{ib}^{b} - \hat{\mathbf{b}}_{a} \tag{6.35}$$

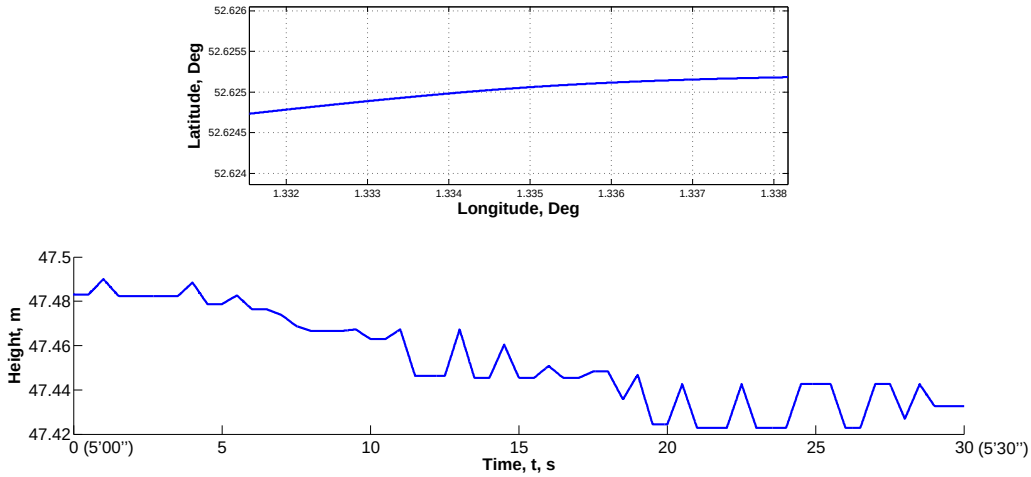
$$\hat{\boldsymbol{\omega}}_{ib}^{b} = \tilde{\boldsymbol{\omega}}_{ib}^{b} - \hat{\mathbf{b}}_{g} \tag{6.36}$$

before being used for calculation. The results are shown in Figure 6.31. The three plots in Figure 6.31 respectively shows the attitude, velocity and position error from the top to the bottom. All errors are plotted against time.

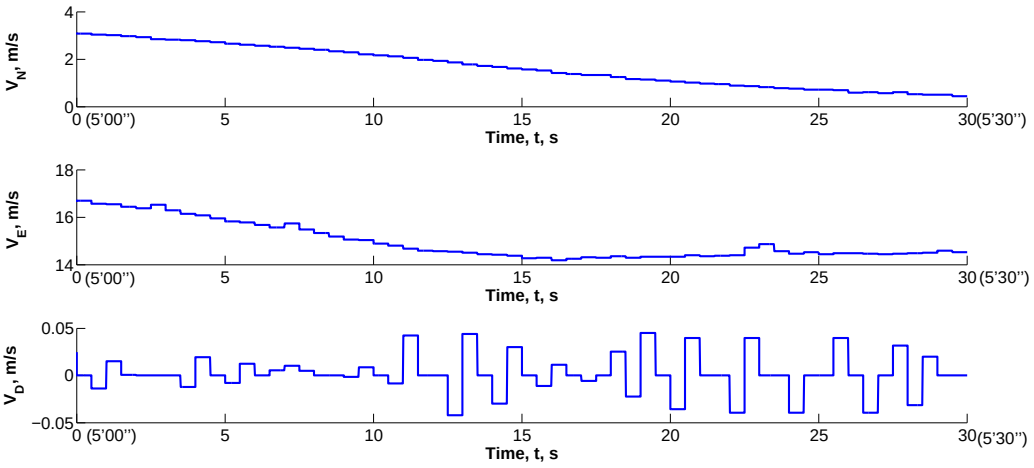
The drifting of the calculation results can be seen clearly from Figure 6.31. The drifting is mainly caused by the residual biases and random noise. As the MEMS sensors tend to suffer from higher biases variation and noises, in comparison with high-performance sensors, compensation with lab tested biases, as been done with (6.35) and (6.36) is usually not enough. The scale of the position error grows to approximately 30 m by the end of the 30 s simulation. El-Sheimy & Niu (2007) reports an example of 30 s MEMS inertial calculation with real-time sensor measurements, sensor specifications of which are similar to this research. The position error is reported as high as $30 - 50$ m after 30 s. Through the comparison with both results from El-Sheimy & Niu (2007) and the calculation errors using the simulated truth data, the MEMS sensor simulation are close to real sensors in terms of performance.




**(a)** Attitude




**(b)** Position




**(c)** Velocity

**Figure 6.29.:** Truth information of the 30 s INS simulation

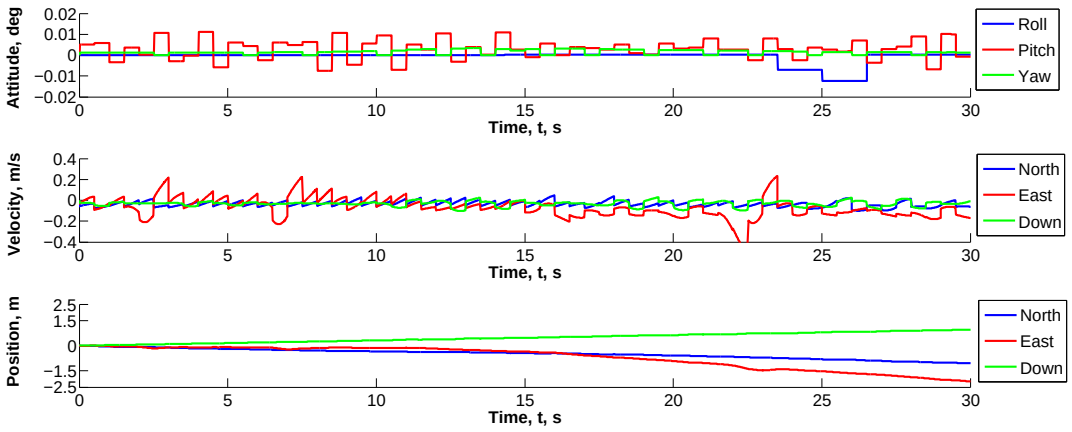



**Figure 6.30.:** Inertial calculation error with truth data

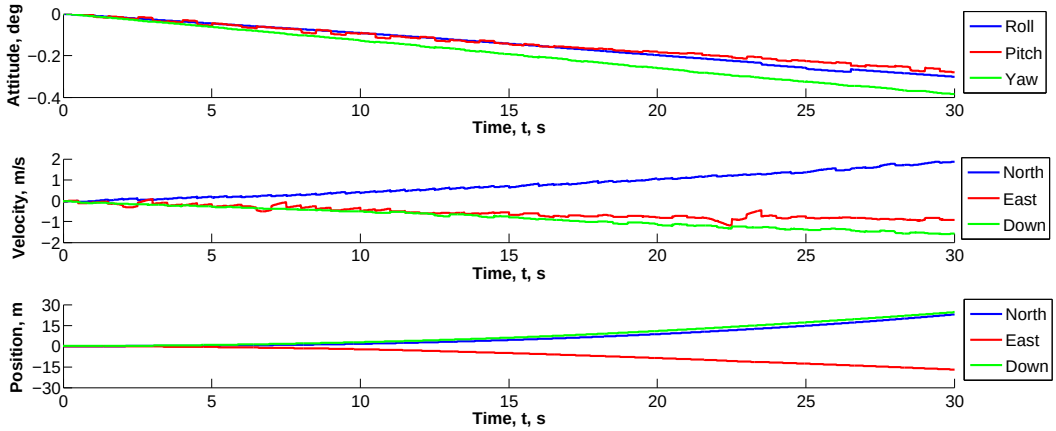



**Figure 6.31.:** Inertial calculation error with MEMS data

## 6.12. GPS Position and Velocity Solution Simulation

A GPS PVT estimation process using an EKF is simulated by using the pseudorange and pseudorange rate measurements generated in Section 6.10. Through the simulation, the validity of the generated GPS measurements is tested, and position and velocity solutions are provided for further simulation in this research. The summary of EKF algorithm is described in Section 3.2.5, whereas the filter design for GPS PVT estimation is discussed in Section 4.2.4.

The simulation lasts 32 min, going through the whole simulation process generated through previous sections. The truth trajectory is shown in Figure 6.6. The visible satellite constellation is selected in Section 6.10.3, and a summary of the constellation status is given in Figure 6.32.

The left half of Figure 6.32 is a skyplot of the visible constellation during the simulation. The moving direction of each satellite is marked with an arrow on its moving track. The skyplot indicates there are four satellites (G05, G12, G14 and G30) ascending during the simulation, and others (G09, G15, G17 and G22) are descending.

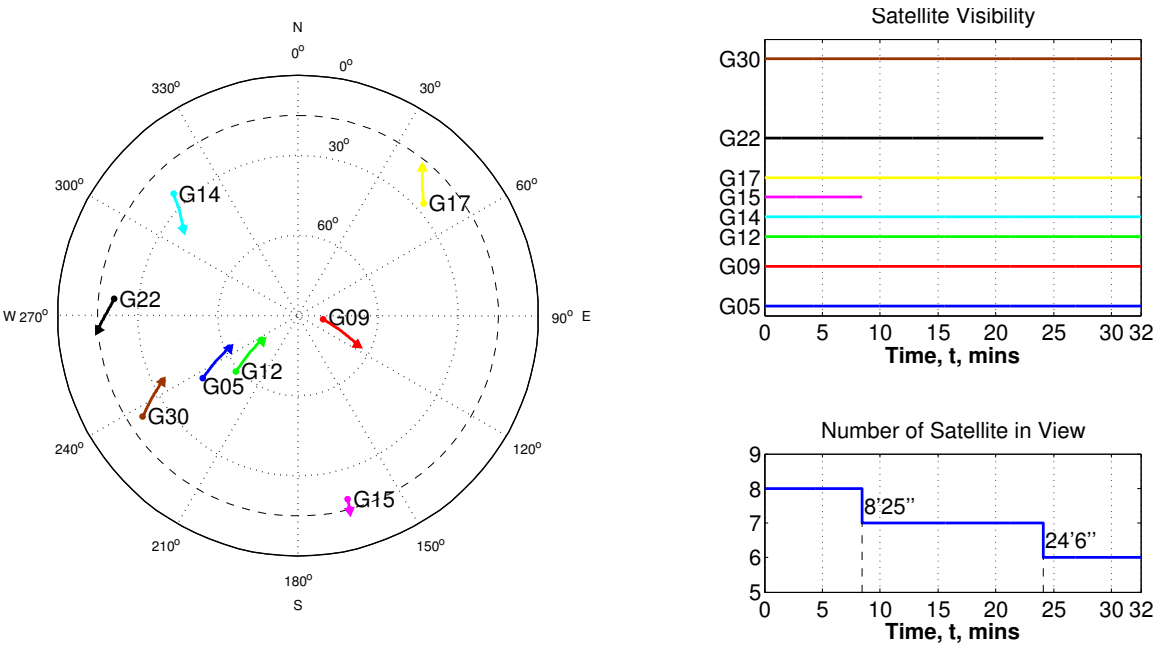



**Figure 6.32.:** Constellation status during the simulation

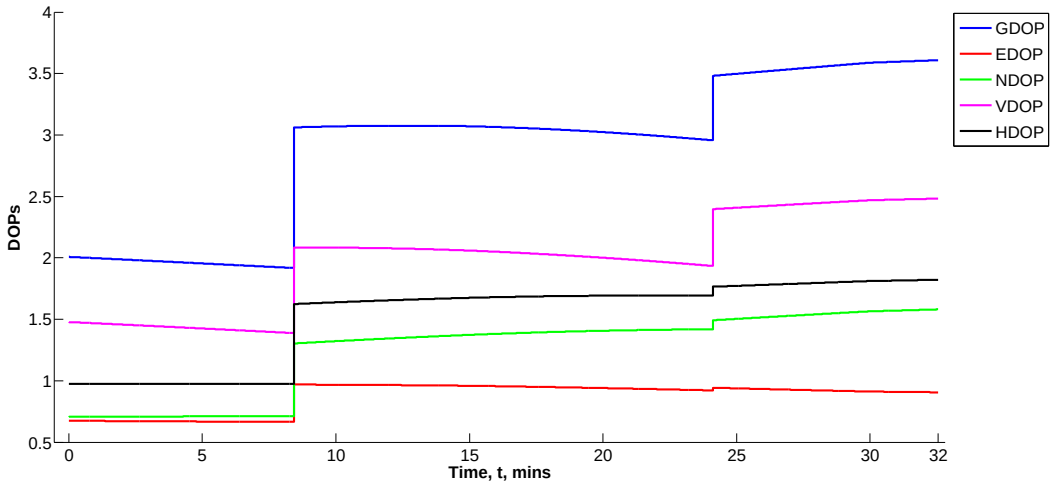


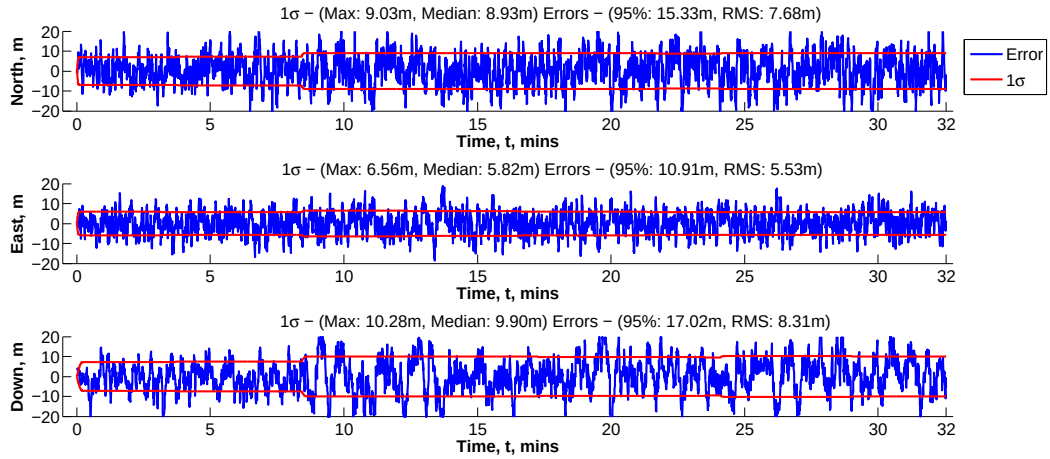

**Figure 6.33.:** DOPs during the simulation

Whilst there are 8 satellites available at the beginning of the simulation, two of them (G15, G22) become invisible with a cut-off angle of 15° as shown by the satellite visibility plots in the right half of Figure 6.32. Satellite G15 and G22 become invisible respectively after 8 min 25 s and 24 min 6 s, leaving only 6 satellite available by the end of the simulation.

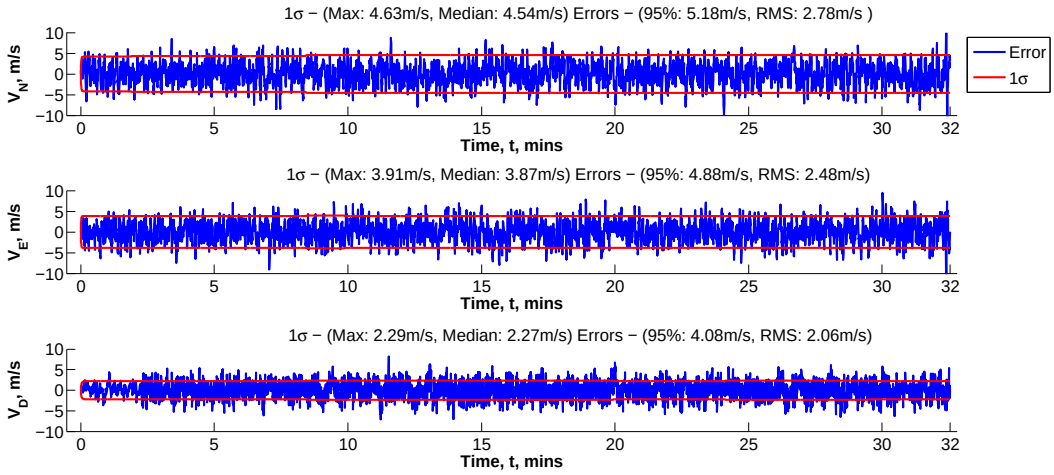
Figure 6.33 shows the DOPs for the simulation process. A detailed discussion of the satellite geometry and the calculation of DOPs is given in Section 4.2.3. The values of GDOP, EDOP, NDOP, VDOP and HDOP are calculated during the simulation. The skyplot indicates a reasonable user-satellites geometry, with the northern part of the sky (the area around 0° N) having fewer satellites than other directions as a result of the GPS constellation design. The effect of the lack of satellites around 0° N is indicated by the DOP values during the simulation, as shown in Figure 6.33. The NDOP values are in general higher than EDOP, implying a poorer geometric spread-out of the satellites. In addition, the impact of losing a satellite can be seen clearly from the steps in

the DOPs, as they increase along with the decreasing of the number of available satellites.

The simulated PVT estimation results are shown in Figure 6.34 with the position and velocity estimation errors expressed in the local navigation frame north, east and down directions. The red lines in all figures indicate the $1\sigma$ boundaries predicted by the EKF estimation covariance matrix. The effect of loosing satellite can be seen from the steps around 8 min 25 s and 24 min 6 s in all $1\sigma$ boundaries.




**(a)** Position errors

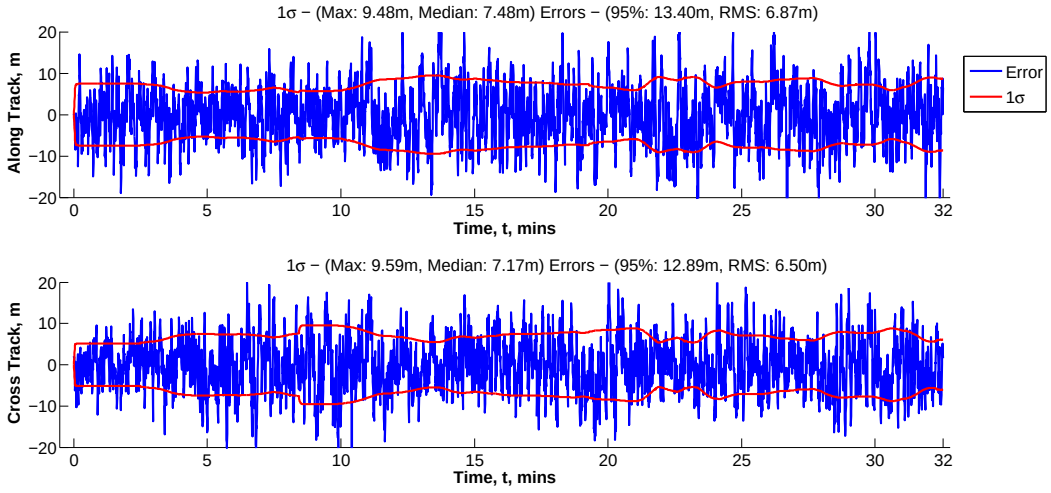



**(b)** Velocity errors

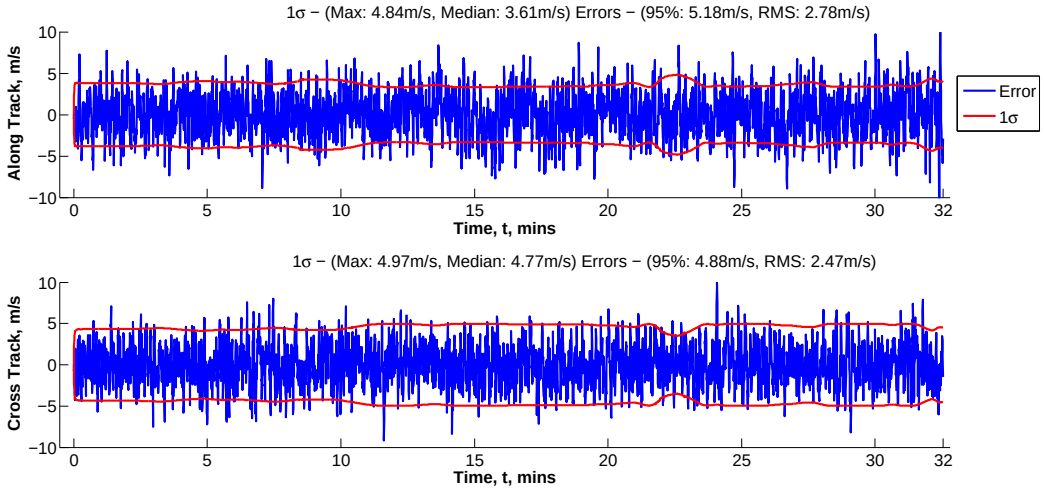
**Figure 6.34.:** Estimation errors of the simulation process in NED

The $1\sigma$ boundary for the position error in the north direction has a maximum value of 9.03 m, and a median value of 8.93 m. Meanwhile the maximum and median for the boundary of the east direction are respectively 6.56 m and 5.82 m. The maximum and median for the down direction are 10.28 m and 9.90 m. In addition, without assuming the distribution of the positioning error, the errors are sorted by ascending order and the 95% of the ascending value is used as another metric for performance indication, along with the root mean square (RMS) of the errors. The 95% values for the position errors in north, east and down are respectively 15.33 m, 10.91 m and 17.02 m, whereas the RMSs are 7.68 m, 5.53 m and 8.31 m. These values fit the general performance characteristics for using GPS positioning in the UK. The performance difference between the north

and east directions are explained as a result of the lack of satellites around 0° N, which is a result of GPS constellation design. The down direction carries the worst performance is because there is no symmetrical distribution of satellites under the vehicle. The same case can be applied to the velocity errors, as shown in Figure 6.34b.




**(a)** Position errors




**(b)** Velocity errors

**Figure 6.35.:** Estimation errors of the simulation process in along track and cross track

The PVT estimation errors are also transformed into the along track and cross track directions, as shown in Figure 6.35. The variation of the $1\sigma$ boundaries is mainly caused by the fact that the vehicle changes its heading during the simulation and the positioning performance in the east direction is better than the north direction.

The RMSs for positioning errors along the track and cross the track are respectively 6.87 m and 6.50 m. By comparing the statistics of the along track and cross track positioning results as shown in the graph, there is no indication of significant positioning performance difference between the along track and the cross track direction. This also confirms a reasonable user-satellites geometry during the simulation, as discussed through the skyplot and DOPs.

The simulation results prove the validity of the simulated satellite ranging measurements, and that a close to real operation positioning performance is achieved by using the simulated measurements.

The positioning solutions are also used as the outputs from a GPS receiver for further system integration research.

# 7. Investigations into MEMS/GPS System Integration

The combination of inertial sensors and satellite positioning enables an integrated positioning system to achieve a better performance from the variety of system characteristics. The conventional system integration requires high performance inertial sensors, which are capable of providing independent positioning solutions without external calibration for a certain period. Low-cost inertial sensors, such as sensors based on MEMS, are usually not used for independent inertial navigation tasks because of their poor performance. However, based on the general GNSS availability study on the railways performed in UK (see Section 2.3.2), the GNSS difficulty situations on the railways are usually for a short period (less than half a minute). This raises the potential opportunity of using MEMS sensor for positioning tasks on the railways.

The core part of an integrated system is the integration algorithm. For the research in this thesis, an EKF is used. A detailed description of the EKF algorithm and its implementation is given in Section 3.2.5. As for the case of most integrated systems, an inertial calculation process is maintained to provide a reference solution for the integration corrections. The implementation of the inertial calculation process is described in Section 5.4. The GPS outputs used by the integrated systems in this chapter are produced using the satellite ranging measurements generated in Section 6.10 through a filtered GPS navigation calculating process, as described in Section 6.12. The derivation of the inertial error states dynamic equations and the design of inertial matrices for the EKF algorithm is presented in Appendix C.

Two common integration architectures used in the inertial sensor and GPS integration are the loosely coupled and tightly coupled approaches. A loosely coupled integrated system uses the GPS position and velocity solution, and the overall system architecture is essentially a cascaded system. A detailed description of the cascaded systems is presented in Section 3.1.1. On the other hand, a tightly coupled system incorporates the satellite ranging measurements directly, and the overall system architecture becomes a centralised system. A description of the centralised integration are presented in Section 3.1.2. All integrated systems investigated in this chapter are implemented in the error-state form.

This chapter investigates both loosely coupled and tightly coupled approaches for integrating MEMS and GPS. The loosely coupled approach is described in Section 7.1, along with an overall system performance analysis using sensor measurements generated in Chapter 6. The tightly coupled approach is described in Section 7.2, along with an overall system performance analysis. Three testing scenarios are proposed in Section 7.3 covering the main problems for using satellite positioning on the railways, as discussed in Section 2.3.1.2. Both systems are tested for all three scenarios, and the results are discussed.

## 7.1. Loosely Coupled Integration Approach

A loosely coupled MEMS/GPS integration system adopts the cascaded architecture as described in Section 3.1.1. Under such an architecture, GPS is used as an aiding system, and MEMS measure-

ments with a higher update rate are used to produce a reference solution for the system integration. The function blocks of the system are illustrated in Figure 7.1.

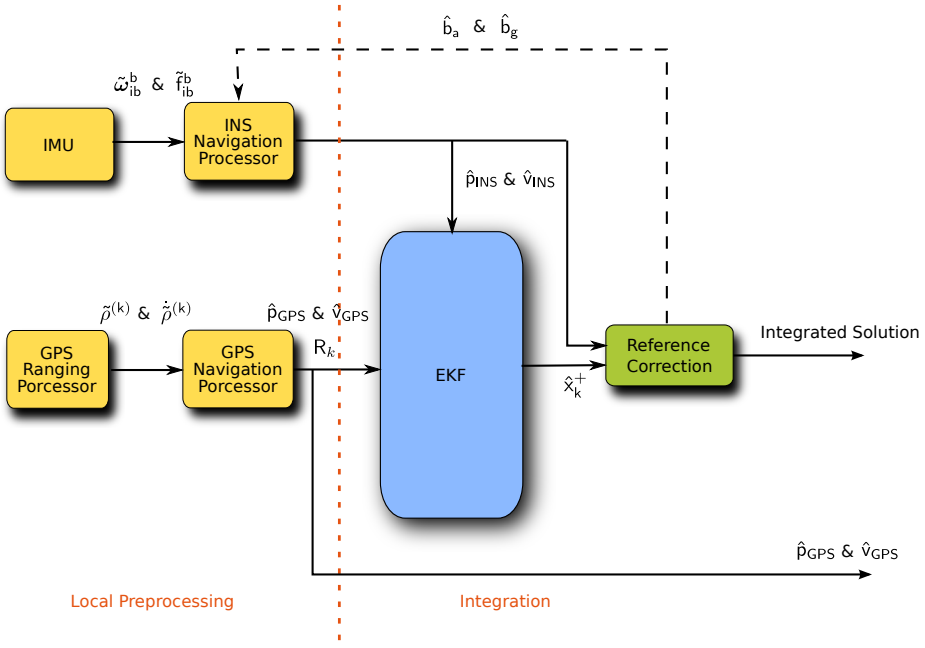



**Figure 7.1.:** The loosely coupled integration approach

The IMU measurements used in the system are generated in Section 6.9. The specification for simulated MEMS sensors are listed in Table 6.1. The satellite ranging measurements used are generated in Section 6.10. The simulated GPS navigation processor estimates the position and velocity based on a filtered approach. No further augmentations are used for the GPS navigation processing. The overall system update rate, $\tau_s$, is aligned with the inertial sensor update rate, $\tau_s = \tau_i = 0.01\,\text{s}$. The overall system operation time is 32 min.

### 7.1.1. State Selection and System Model

The state vector for the loosely coupled system used in this research contains only the inertial states. An error-state implementation is chosen for the loosely coupled system. As a result, the state vector and the system model are defined as

$$\mathbf{x} = \mathbf{x}_{INS} \tag{7.1}$$

$$\mathbf{F} = \mathbf{F}_{INS},\ \mathbf{\Phi} = \mathbf{\Phi}_{INS} \tag{7.2}$$

$$\mathbf{Q} = \mathbf{Q}_{INS} \tag{7.3}$$

where the definitions of $\mathbf{x}_{INS}$, $\mathbf{F}_{INS}$, $\mathbf{\Phi}_{INS}$ and $\mathbf{Q}_{INS}$ is shown in Appendix C.

### 7.1.2. Measurement Model

The measurement vector consists of the differences between the GPS solutions and the integrated system solution. Therefore the measurement vector and the measurement matrix are given as

$$\delta\mathbf{z}_k = \begin{bmatrix} \hat{\mathbf{p}}_{GPS} - \hat{\mathbf{p}}_{INS} \\ \hat{\mathbf{v}}^n_{GPS} - \hat{\mathbf{v}}^n_{eb} \end{bmatrix}_k \tag{7.4}$$

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \tag{7.5}$$

Note that the lever arm, the distance between the IMU and the GPS antenna, is not accounted for in the simulation; it is assumed that both parts of the system share the same position.

The measurement noise covariance matrix, $\mathbf{R}_k$, is formed at each GPS update epoch based on the estimation quality given by the GPS navigation processor.

### 7.1.3. System Performance Analysis

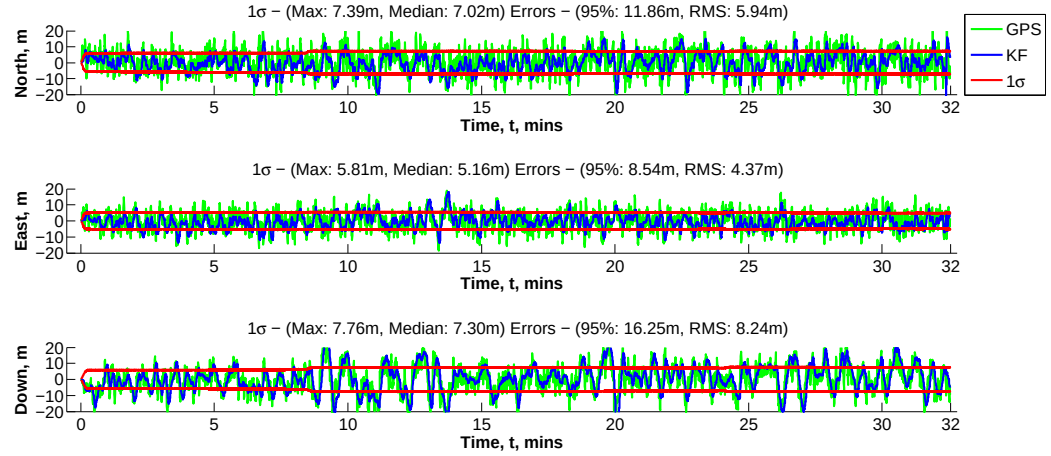
The performance of the loosely coupled integrated system is plotted in Figure 7.2, with the access to the full simulated constellation and no physical obstructions simulated. The three graphs in the figure respectively show the position, velocity and attitude error in the integrated solution. All error information is plotted against the simulation time. The error in the GPS solution is plotted in each graph in green line as a comparison. In addition, the predicted $1\sigma$ boundary of the system positioning solution error, provided by the state error covariance matrix, is plotted in each graph as the red lines. Statistics for the integrated positioning error and the predicted $1\sigma$ are also calculated and marked in each graph.

The position errors, shown in Figure 7.2a, are expressed in the direction of north, east and down in the local navigation frame in meters. The RMS errors for these three directions are respectively 5.94 m, 4.37 m and 8.24 m. In comparison with the RMS errors of GPS solutions (listed in Section 6.12) of 7.68 m, 5.53 m and 8.31 m, the accuracy improvement of using loosely coupled approach can be observed. The error pattern for the loosely coupled solution is seen from the figure roughly following the error pattern of the GPS only solution. This is because GPS serves as the main contributor for the overall system performance in the integration process. More importantly, as can be seen from the figure, position errors of the loosely coupled system are less noisy than the GPS only solution, and the performance difference between the north direction and the east direction is also improved. This can be shown by the decrease of the 95% error on all three directions from 15.33 m, 10.91 m and 17.02 m for the GPS only solution, to 11.86 m, 8.54 m and 16.25 m for the loosely coupled solution.

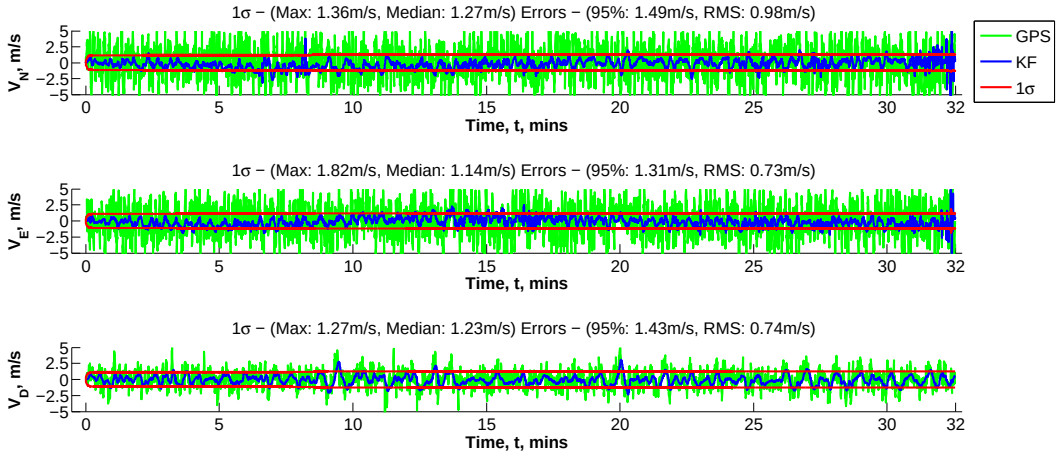
The velocity errors, shown in Figure 7.2b, are also expressed in the direction of north, east and down in the local navigation frame, in meters per second. With the RMS error decreases from 2.78 m/s, 2.48 m/s and 2.06 m/s for the GPS only solution, as listed in Section 6.12, to 0.98 m/s, 0.73 m/s and 0.74 m/s for the loosely coupled solution, the accuracy improvement for the velocity estimation using an integrated system can be seen clearly from the figure. The error pattern for the loosely coupled solution does not exactly follow the pattern of the GPS only solution. In addition, the loosely coupled velocity solution shows less variation than the GPS solution, as indicated by the closeness of the 95% errors and the RMS errors for all directions. The reason for this is that the inertial calculation process provides a better resolution than the GPS process in the case of the velocity estimation.

The attitude errors, shown in Figure 7.2c, are expressed in the form of Euler angles. The errors shown in the roll and pitch directions have the approximately the same level of accuracy performance, and their RMS errors are respectively 0.31° and 0.34°. The largest error is in the yaw direction, with a RMS error of 2.49°. Compared with the drifting pattern shown by the attitude errors of the 30 s inertial calculation in Section 6.11, all attitude errors during the 32 min operation appear to be well bounded.

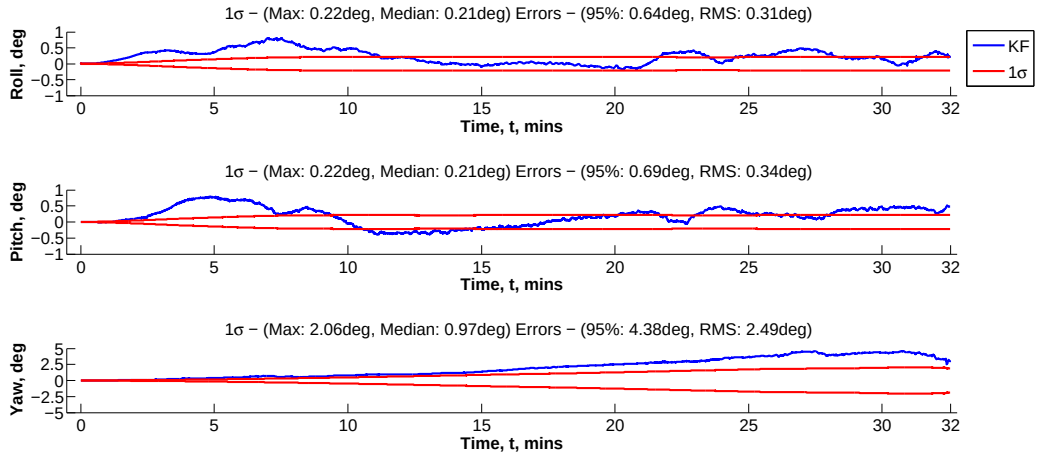
It should be noted that the results shown in Figure 7.2 are the best results after careful tuning




**(a)** Position errors




**(b)** Velocity errors

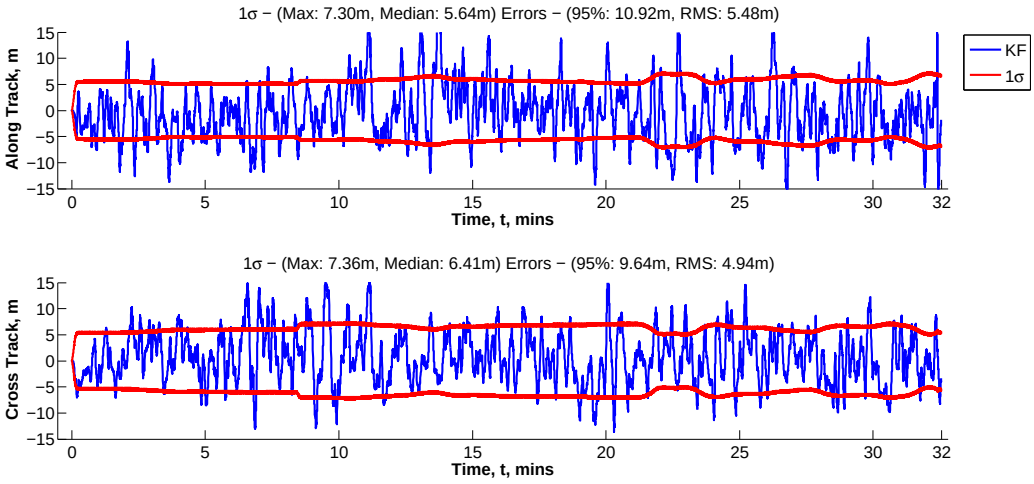



**(c)** Attitude errors

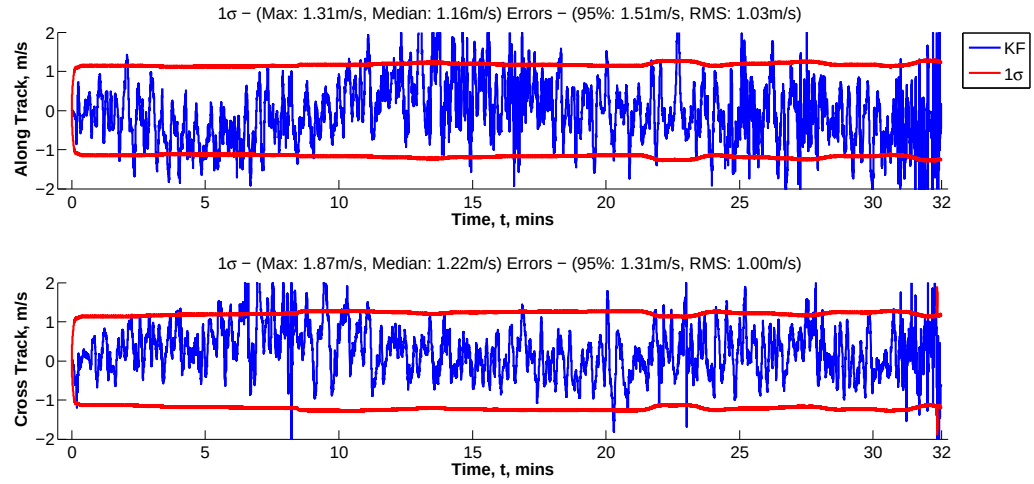
**Figure 7.2.:** Estimation errors of the loosely coupled system

of the integration EKF parameters, such as initial state error covariance matrix ($\mathbf{P}_0$), the system noise covariance matrix ($\mathbf{Q}$) and the measurement noise covariance matrix ($\mathbf{R}_k$). As discussed in Section 3.1.1, because of the existence of multiple estimation algorithms, i.e. Kalman filters, in the system, the estimated outputs error information provided by the GPS navigation processor is no longer optimal for the integration EKF. Without proper tuning, the system shows over-optimistic $1\sigma$ boundaries and increased RMS errors. For an example of untuned system performance see Table 7.1 on page 147, where the position error is reported with a RMS error of 7.66 m in the north direction, whereas the predicted median $1\sigma$ is only 4.05 m.

The estimation errors are also transformed into the along track and cross track directions, as shown in Figure 7.3. Accuracy improvements are shown through the RMS errors, 5.46 m and 4.94 m respectively for along track and cross track position errors, for example. Similar to the along track and cross track performance of the GPS only solution, there is no strong indication of significant positioning performance difference between the along track and the cross track direction.




**(a)** Position errors




**(b)** Velocity errors

**Figure 7.3.:** Estimation errors of the loosely coupled system in along track and cross track

## 7.2. Tightly Coupled Integration Approach

A tightly coupled MEMS/GPS integration system adopts the centralised architecture as described in Section 3.1.2. Under such an architecture, although GPS is still used as an aiding system, the satellite ranging measurements are used as inputs for the integration EKF instead of the position and velocity estimates. As a result, no generic GPS only positioning solution is available. However, a GPS navigation processor can be maintained in parallel at the cost of increasing system complexity and computation load. Meanwhile, MEMS measurements are still used to produce a reference solution for the system integration. The function blocks of the system are illustrated in Figure 7.4.

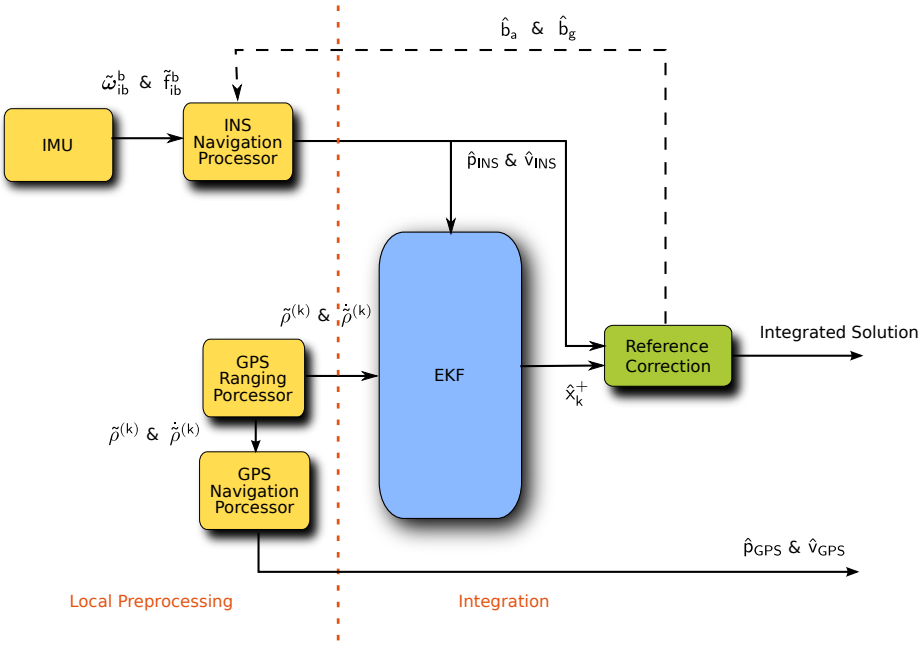



**Figure 7.4.:** The tightly coupled integration approach

The IMU measurements used in the system are generated as in Section 6.9. The specification for simulated MEMS sensors is listed in Table 6.1. The satellite ranging measurements used are generated as in Section 6.10. The specification for the receiver clock error and residual correction errors are listed in Section 6.10.4. No further augmentations are used for the GPS processing. The overall system update rate, $\tau_s$, is aligned with the inertial sensor update rate, $\tau_s = \tau_i = 0.01\,\text{s}$. The overall system operation time is 32 min.

### 7.2.1. State Selection and System Model

Because the GPS navigation processor is no longer necessary under a tightly coupled approach, the estimation of the receiver clock errors must be performed in the EKF in order to correct the satellite ranging measurements from the ranging processor. Thus it is necessary to define the EKF matrices for the GPS states. The receiver clock error parameters include the receiver clock bias, $b_{rc}$, and the receiver clock drift, $\dot{b}_{rc}$. The GPS state vector can therefore be written as

$$\mathbf{x}_{GPS} = [b_{rc},\ \dot{b}_{rc}]^T \tag{7.6}$$

the model for the receiver clock errors is given by (6.30) and (6.31). Based on the receiver clock error model, the system matrix for the GPS states can be defined as

$$\mathbf{F}_{GPS} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{7.7}$$

The transition matrix, $\mathbf{\Phi}_{GPS}$, for the discrete propagation model can be computed with $\mathbf{F}_{GPS}$ using a first order power-series expansions of (3.4), and is written here as

$$\mathbf{\Phi}_{GPS} = \mathbf{I}_{2\times 2} + \mathbf{F}_{GPS}\tau_s \tag{7.8}$$

The system noise covariance matrix for the GPS states is defined as

$$\mathbf{Q}_{GPS} = \begin{bmatrix} 0 & 0 \\ 0 & n_{rc}^2 \end{bmatrix} \tau_s \tag{7.9}$$

where $n_{rc}^2$ is the power spectral density for the random noise on $\dot{b}_{rc}$.

The state vector for the tightly coupled system consists of both inertial states and GPS states, and is defined as

$$\mathbf{x} = [\mathbf{x}_{INS},\ \mathbf{x}_{GPS}]^T \tag{7.10}$$

where the definition of $\mathbf{x}_{INS}$ is given in Appendix C. The system matrix and the transition matrix accordingly become

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{INS} & \mathbf{0}_{15\times 2} \\ \mathbf{0}_{2\times 15} & \mathbf{F}_{GPS} \end{bmatrix},\ \mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{INS} & \mathbf{0}_{15\times 2} \\ \mathbf{0}_{2\times 15} & \mathbf{\Phi}_{GPS} \end{bmatrix} \tag{7.11}$$

and the overall system noise covariance matrix is

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{INS} & \mathbf{0}_{15\times 2} \\ \mathbf{0}_{2\times 15} & \mathbf{Q}_{GPS} \end{bmatrix} \tag{7.12}$$

All inertial sub-matrices are given in Appendix C.

### 7.2.2. Measurement Model

The measurement vector consists of the differences between the ranging measurements from the ranging processor and the predicted ranging measurements, which are calculated from the estimated states. Therefore the measurement vector for epoch $k$, assuming there are $K$ satellites in view, is given as

$$\delta\mathbf{z}_k = \begin{bmatrix} \delta\mathbf{z}_{\rho,k} \\ \delta\mathbf{z}_{\dot{\rho},k} \end{bmatrix} \tag{7.13}$$

where $\delta\mathbf{z}_{\rho,k}$ and $\delta\mathbf{z}_{\dot{\rho},k}$ are respectively defined as following

$$\delta\mathbf{z}_{\rho,k} = \left[\tilde{\rho}^{(1)} - \hat{\rho}^{(1)},\ \tilde{\rho}^{(2)} - \hat{\rho}^{(2)},\ \cdots,\ \tilde{\rho}^{(K)} - \hat{\rho}^{(K)}\right]_k^T \tag{7.14}$$

$$\delta\mathbf{z}_{\dot{\rho},k} = \left[\tilde{\dot{\rho}}^{(1)} - \hat{\dot{\rho}}^{(1)},\ \tilde{\dot{\rho}}^{(2)} - \hat{\dot{\rho}}^{(2)},\ \cdots,\ \tilde{\dot{\rho}}^{(K)} - \hat{\dot{\rho}}^{(K)}\right]_k^T \tag{7.15}$$

The predicted pseudorange, $\hat{\rho}$, and pseudorange rate, $\hat{\dot{\rho}}$, are calculated using (4.26) and (4.27).

Assuming the line-of-sight vector expressed in the local navigation frame is defined as $\hat{\mathbf{1}}^{(k),n} = \left[u_N^{(k),n},\ u_E^{(k),n},\ u_D^{(k),n}\right]^T$. The measurement matrix can be derived based on the same principles as described in Section 4.2.4.3 with sign differences, and is written here as

$$\mathbf{H}_k \approx \left[\begin{array}{ccccccc}
\mathbf{0}_{1\times3} & h_\rho^{(1)} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 1 & 0 \\
\mathbf{0}_{1\times3} & h_\rho^{(2)} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 1 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{1\times3} & h_\rho^{(K)} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 1 & 0 \\
\hline
\mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & (\hat{\mathbf{1}}^{(1),n})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & 1 \\
\mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & (\hat{\mathbf{1}}^{(2),n})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & 1 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & (\hat{\mathbf{1}}^{(K),n})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & 0 & 1
\end{array}\right]_{\mathbf{x}=\hat{\mathbf{x}}_k} \tag{7.16}$$

where $h_\rho^{(j)}$ for the $j$th satellite is defined as

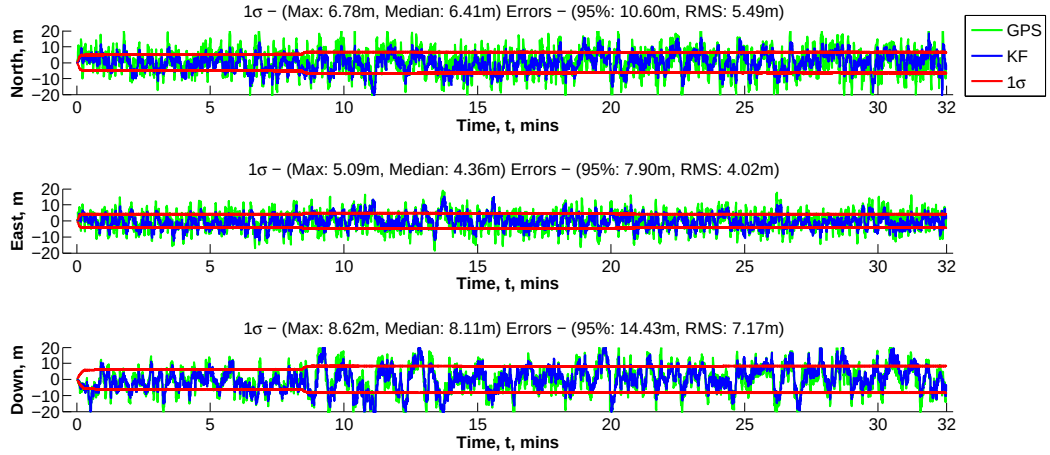
$$h_\rho^{(j)} = \left[(R_N(\hat{L}) + \hat{h})u_N^{(j),n},\ (R_N(\hat{L}) + \hat{h})\cos\hat{L}u_E^{(j),n},\ u_D^{(j),n}\right] \tag{7.17}$$

### 7.2.3. System Performance Analysis

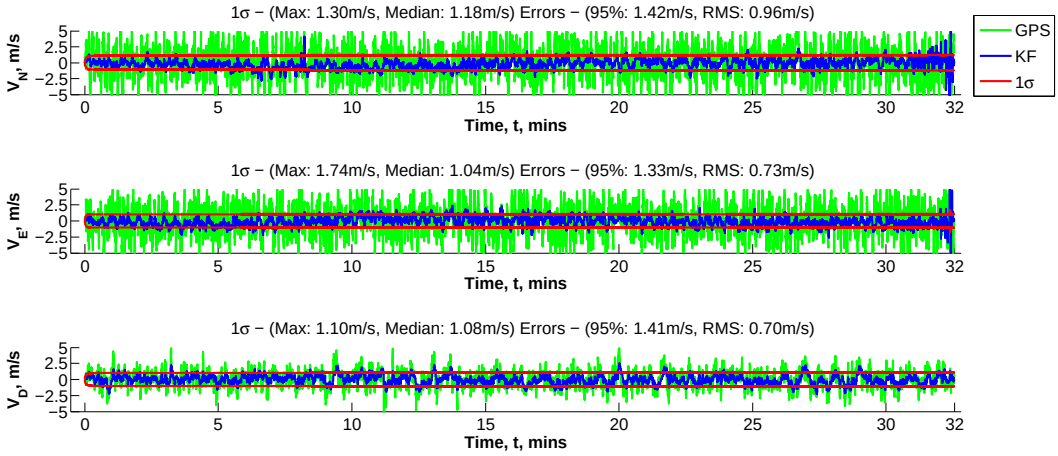
The performance of the tightly coupled integrated system is plotted in Figure 7.5, assuming access to the full simulated constellation and no physical obstructions simulated. Three graphs in the figure respectively show the position, velocity and attitude error in the integrated solution. All error information is plotted against the simulation time. The error in the GPS solution is plotted in each graph in green line providing a comparison. In addition, the predicted $1\sigma$ boundary of the system positioning solution error, provided by the state error covariance matrix, is plotted in each graph as the red lines. Statistics for the integrated positioning error and the predicted $1\sigma$ are also calculated and marked in each graph.

The position errors, shown in Figure 7.5a, are expressed in the direction of north, east and down in the local navigation frame in meters. The RMS errors for these three directions are respectively 5.49 m, 4.02 m and 7.17 m. In comparison with the RMS errors of GPS solutions, listed in Section 6.12 as 7.68 m, 5.53 m and 8.31 m, and the RMS errors of the loosely coupled solutions, listed in Section 7.1.3 as 5.94 m, 4.37 m and 8.24 m, the accuracy improvement of using tightly coupled approach can be observed. Because GPS still serves as the main contributor for the overall system performance in the integration process, the error pattern for the tightly coupled solution roughly follows the error pattern of the GPS only solution, similar to the loosely coupled solutions.

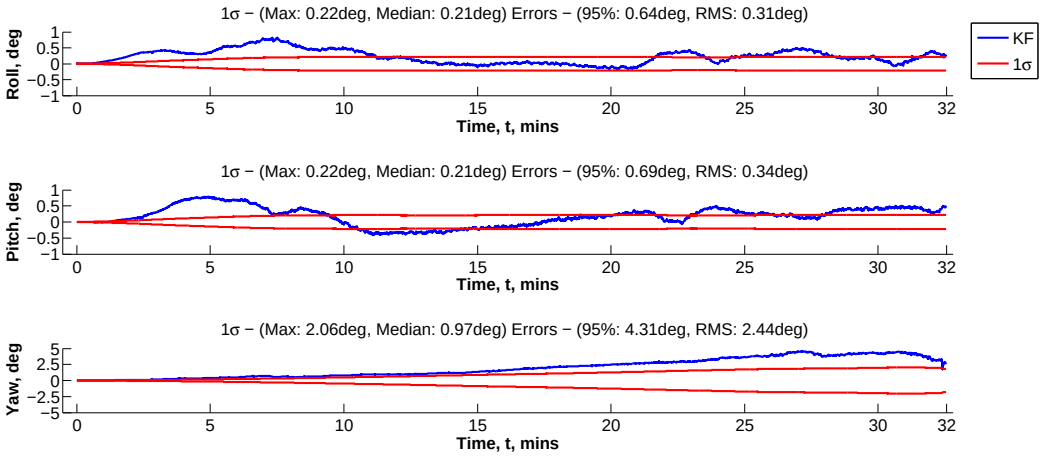
The velocity errors, shown in Figure 7.5b, are also expressed in the direction of north, east and down in the local navigation frame, in meters per second. The RMS error decreases from 0.98 m/s,




**(a)** Position errors




**(b)** Velocity errors

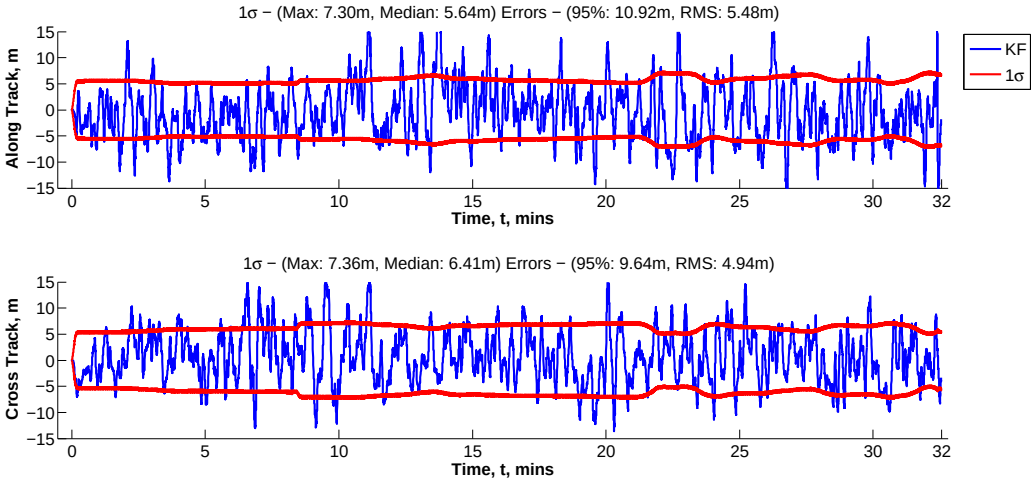



**(c)** Attitude errors

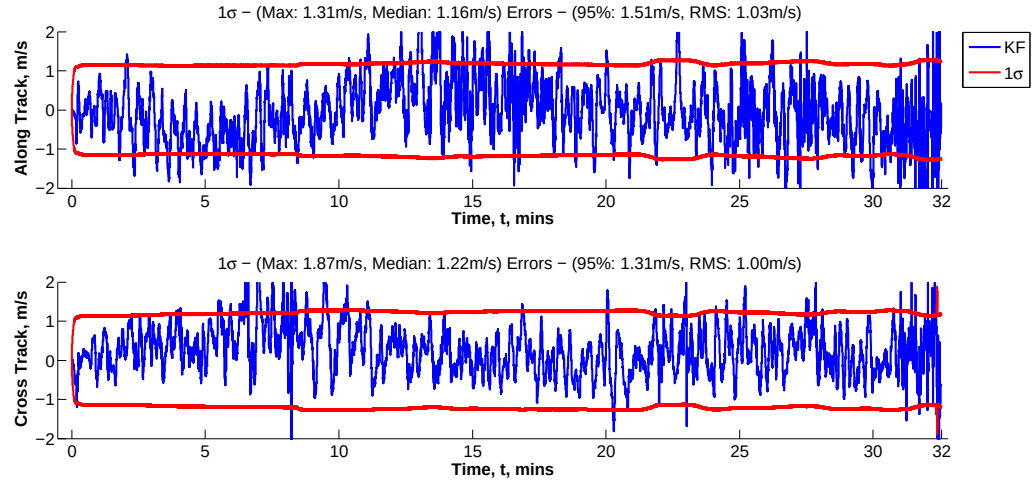
**Figure 7.5.:** Estimation errors of the tightly coupled system

0.73 m/s and 0.74 m/s for the loosely coupled solution, as listed in Section 7.1.3, to 0.96 m/s, 0.73 m/s and 0.70 m/s for the tightly coupled solution, a small accuracy improvement for the velocity estimation is achieved. As the same inertial sensors and same inertial calculation process is applied to both loosely coupled approach and tightly coupled approach, the improvement of accuracy is limited.

The attitude errors, shown in Figure 7.5c, are expressed in the form of Euler angles. The errors shown in the roll and pitch directions have approximately the same level of accuracy performance, and their RMS errors are respectively listed as 0.31° and 0.34°. The largest error is presented in the yaw direction, with a RMS error of 2.44°. The tightly coupled solution shows an almost similar performance to the loosely coupled system for the attitude errors, as no further measurement types are introduced to increase the observability of the attitude errors.




**(a)** Position errors




**(b)** Velocity errors

**Figure 7.6.:** Estimation errors of the tightly coupled system in along track and cross track

The estimation errors are also transformed into the along track and cross track directions, as shown in Figure 7.6. Accuracy improvements are shown through the RMS errors, 5.48 m and 4.49 m respectively for along track and cross track position errors, for example. Similar to the along track and cross track performance of the GPS only solution and loosely coupled solution, there is no strong indication of significant positioning performance difference between the along track and the

cross track direction.

**Table 7.1.:** An example of untuned loosely coupled system performance

| | | $1\sigma$ Errors | | | |
|---|---|---|---|---|---|
| | | Maximum | Median | 95% | RMS |
| Position (m) | N | 4.52 | 4.05 | 14.85 | 7.66 |
| | E | 3.58 | 3.09 | 10.85 | 5.59 |
| | D | 3.57 | 3.29 | 16.79 | 8.45 |
| Velocity (m/s) | N | 2.45 | 1.95 | 3.68 | 2.03 |
| | E | 3.68 | 1.80 | 3.28 | 1.70 |
| | D | 1.69 | 1.49 | 2.37 | 1.19 |

Unlike the tuning process for a loosely coupled system, a tightly coupled system only needs the minimum amount of tuning and the untuned system does not present performance differences as large as a loosely coupled system. Provided a loosely coupled system is finely tuned and there are no GNSS difficulty situations during the operation, the advantage of adopting a tightly coupled system is small in terms of accuracy. However, for practical usages, the information for sub-systems is not always available. Thus the tuning of an integrated system may be poor, as in the example shown in Table 7.1. Under such situations, using a tightly coupled approach shows advantages not only in accuracy but also in robustness of the system.

## 7.3. Test Scenarios

Three test scenarios are proposed in this section including typical situations on the railways where GNSS only positioning performance can be affected. Three types of systems are used for testing under proposed scenarios: a GPS only system with a Kalman filter based navigation processor, a loosely coupled integrated system using MEMS inertial units and the position and velocity solution provided by the GPS only system, and a tightly coupled integrated system using MEMS inertial units and satellite ranging measurements. The performance of all systems is then compared with each other and the performance of their nominal operations when all satellite signals are available and no DRM searching errors are reported within the same testing period.

It should be noted that GPS navigation processors using least-squares method to produce single-point solutions do not work under situations where the number of available satellites is less than four. Such GPS only systems do not work under proposed testing scenarios 1 and 2. Hence loosely coupled integrated systems using solutions from such GPS only systems rely purely on MEMS inertial calculation for testing scenario 1 and 2.

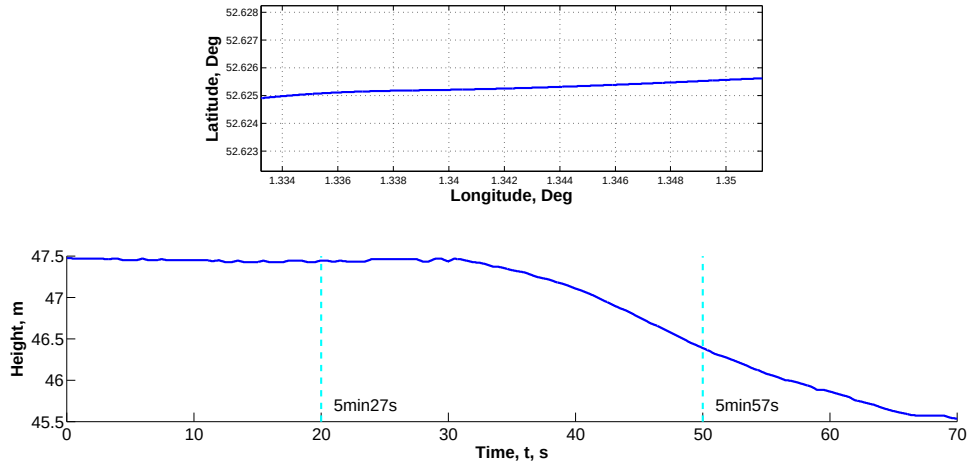
### 7.3.1. Scenario 1: Loss of GNSS Signals

This scenario simulates situations on the railways where the sky is completely blocked and GNSS signals are lost. This mostly happens when a train travels through a tunnel. According to the general GNSS availability study performed by LOCASYS (see Section 2.3.2), the GNSS unavailable situations on the railways usually only last for a short period, i.e. less than a minute. Therefore the length of the testing scenario is set as 30 s. Two sections during the overall operation period are chosen for testing scenario 1. The first section is chosen between $5'27''$ and $5'57''$ of the 32 min

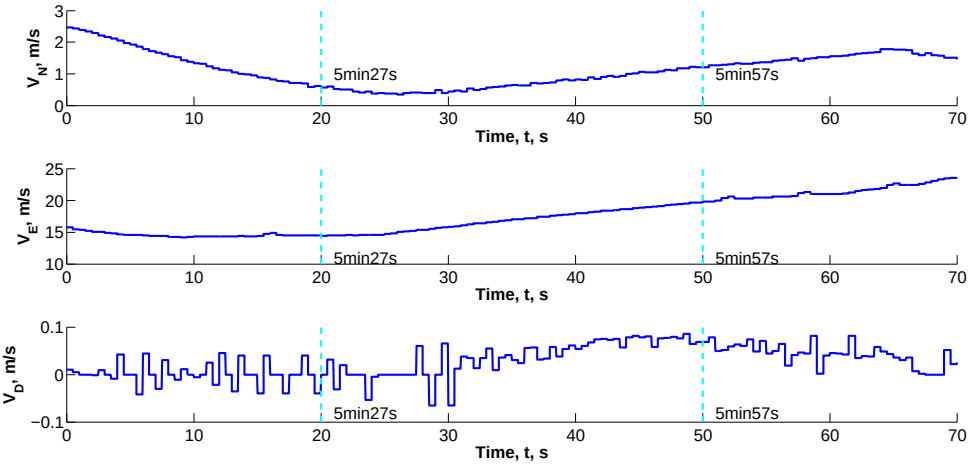
operation time, whereas the second section is chosen between 21′57″ and 22′27″. An overview of the 32 min operation time is illustrated in Figure 6.6 of Section 6.4.

The truth position and velocity profiles of the two testing sections are respectively shown by Figure 7.7 and Figure 7.8. As can be seen from the figures, both testing sections are chosen based on the direction of the rail route during the testing time, in order to clarify the effect of a hostile GNSS signal environment on the along track and cross track positioning performance of the systems. In addition, there is a speed difference between two testing sections, with one section travelling at a high speed and the other one travelling at a low speed. Thus the effect of travelling at different speed on the error growth of MEMS inertial calculation can be observed.

From Figure 7.7, it can be seen that the train is mainly moving along the east-west direction during the first testing section with a higher speed than the second testing section. The height decreases from 47.5 m to 46.5 m during the 30 s. A speed profile between 15 m/s and 20 m/s can be seen on the east direction in the time section marked by dashed lines, whereas the speed on the north direction is between 1 m/s and 2 m/s for the same time section.




**(a)** Position

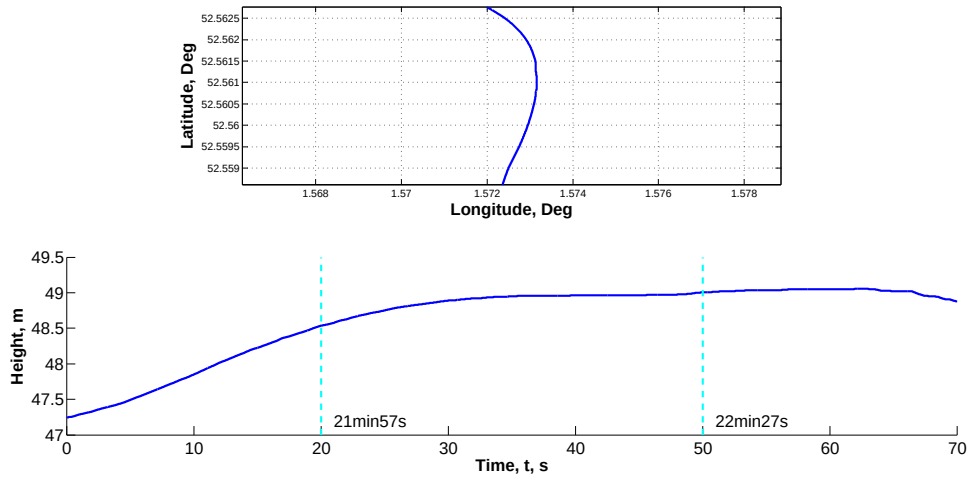



**(b)** Velocity

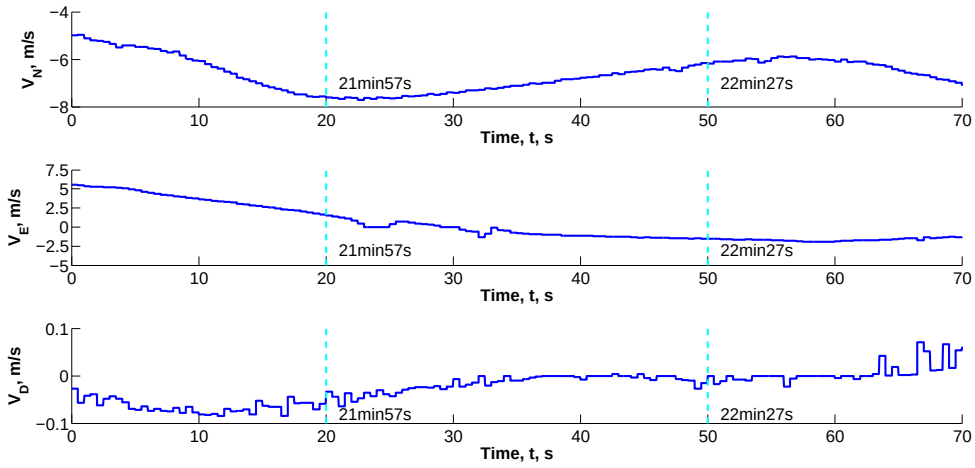
**Figure 7.7.:** Position and velocity truth between 5′27″ and 5′57″

From Figure 7.8, it can be seen that the train is mostly moving along the north-south direction

during the second testing section, and its speed is lower than the first section. The height increases from 48.5 m to 49 m during the 30 s. A speed profile between 6 m/s and 8 m/s can be seen on the south direction in the time section marked by dashed lines, whereas the speed on the east-west direction changes from 2.5 m/s in the east and 1 m/s in the west for the same time section.




**(a)** Position



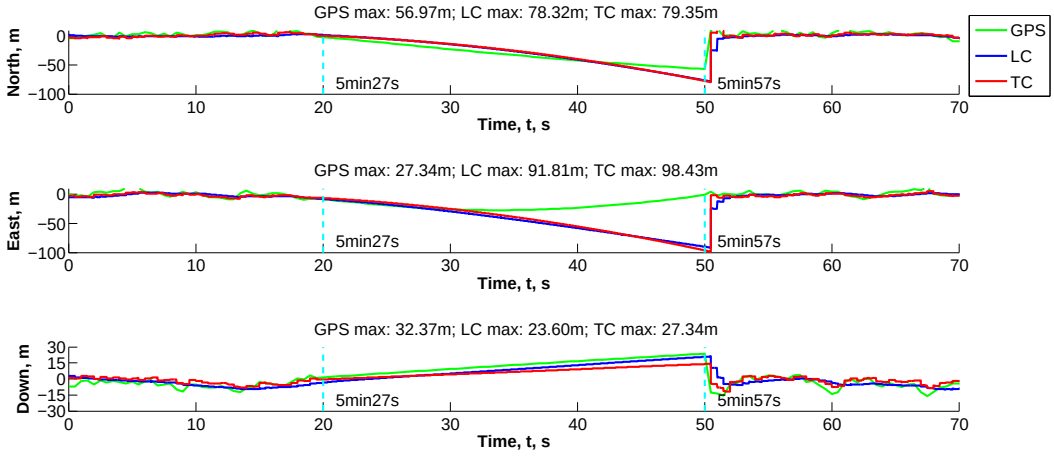

**(b)** Velocity

**Figure 7.8.:** Position and velocity truth between $21'57''$ and $22'27''$

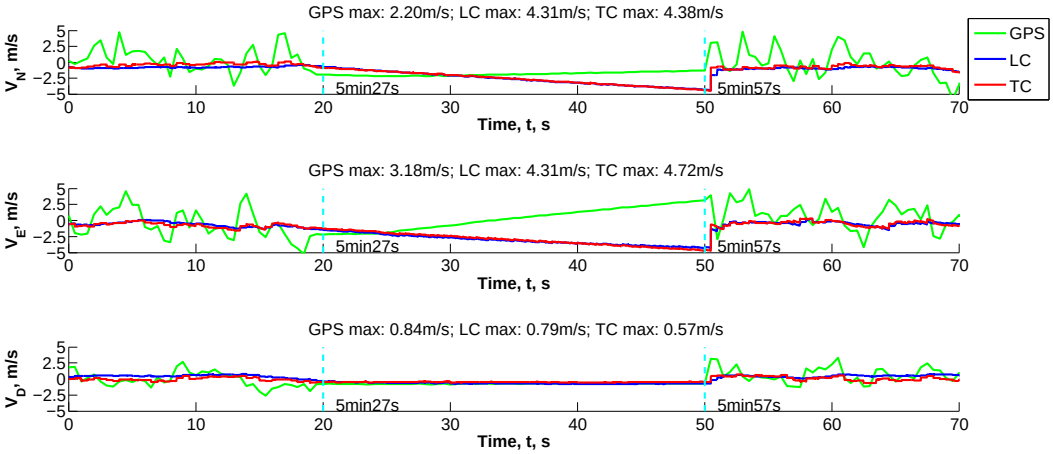
Figure 7.9 and Figure 7.10 respectively show the testing results for the first and the second testing sections. The nominal system performance under situations when all satellite signals are available and no DRM searching errors are reported of all systems for both testing sections are listed respectively in Table 7.2 and Table 7.3. Both position errors and velocity errors are shown in the direction of north, east and down, plotted against the simulation time, with position errors in meters and velocity errors in meters/second. The results for all tested systems are drawn respectively in green for the GPS only system, in blue for the loosely coupled system and in red for the tightly coupled system. The testing section is marked by dashed lines in all graphs. Clear error drifting patterns can be observed in all graphs on all systems in comparing with the normal system performances as listed in Table 7.2 and Table 7.3.

As noted at the beginning of Section 7.3, for a complete loss of satellite signal, a least-squares

based GPS navigation processor does not work because of insufficient satellites. On the other




**(a)** Position errors




**(b)** Velocity errors

**Figure 7.9.:** Position and velocity errors between $5'27''$ and $5'57''$ for scenario 1

hand, a filter-based navigation processor keeps on producing solutions by propagating the states through its system model from the last moment when satellite signals are still available. Therefore, the performance of a GPS only system using a filter based navigation processor is largely dependent on two factors: how the system dynamics model implemented in the filter system model fits with the actual system dynamics during the signal outage period, and the state errors status at the beginning of the signal outage period.

The first factor on system model assumptions can be shown by the performance difference of the GPS only system under two testing sections with different dynamic conditions. For higher speed situations such as the first testing section, the system dynamic model mostly fits the actual system dynamics, and an error drifting pattern that are smaller than the integrated systems can be observed. For instance, the maximum position errors for the GPS only system are respectively 56.97 m in the north, 27.34 m in the east and 32.37 m in the down direction, which are smaller than the errors of integrated systems. However, when the system dynamics do not fit the filter assumption as in the case of the second testing section, the maximum position errors for the GPS only system are larger

than the errors of integrated, respectively 15.48 m in the north, 68.03 m in the east and 48.42 m in the down direction.

The second factor affecting the initial error status of the signal outage can be shown by the error pattern in the east direction of the first testing section. As can be seen from the east velocity error plot in Figure 7.9b, the drifting pattern of the error follows the direction of the initial error state demonstrating an opposite drifting direction to the integrated systems. As the east velocity error grows across zero from negative to positive, a first increasing and then decreasing pattern can be observed on the correspondent east position error, as shown in Figure 7.9a.

Unlike the GPS only system, the integrated systems have access to MEMS sensors for updated information on the system dynamics. For the tightly coupled system, the only input available during the signal outage is from the MEMS inertial unit. The system positioning solution therefore performs similarly to the results of a pure inertial calculation process. For the loosely coupled system used in this testing scenario, the solution coming from the GPS only system is still used for the integration EKF during the signal outage period. The positioning results from the loosely coupled system consequently become a weighted combination of the MEMS measurements and the GPS solution.

As a result, for the first testing section, where the GPS only system presents the best performance, the loosely coupled system performs better than the tightly coupled system demonstrating smaller errors in the positioning solution as shown in Figure 7.9. But for the second testing section, where the GPS only system presents the worst performance, the loosely coupled system performs worse than the tightly coupled system demonstrating larger errors in the positioning solution as shown in Figure 7.10. It should be noted that when a loosely coupled system integrates positioning solutions from a GPS navigation processor using least-squares based algorithms, it does not have access to information other than MEMS sensor measurements just as the tightly coupled system. Under such circumstances, the performance difference between the two types of systems depends on the initial positioning errors when the signal outage starts.

It can also be noticed from the results that the size of positioning errors from integrated systems depends on the system travelling speed during the signal outage period. By comparing the maximum errors for both integrated systems in Figure 7.9 and Figure 7.10, the positioning errors drift slower under a slow travelling speed as in the case of the second testing section. This can also be proved through the INS error dynamics equations given in Appendix C, because inertial sensors are the main source for system states update during the signal outage.

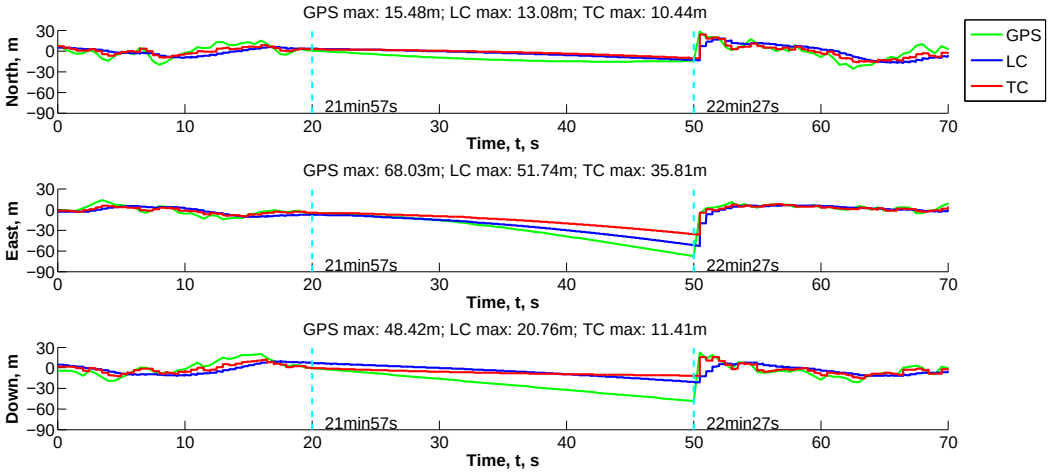
### 7.3.2. Scenario 2: Difficult GNSS Conditions

This scenario simulates situations on the railways where only a part of the sky is visible and the number of available satellites is less than four. This mostly happens when a train travels through a urban area, or stops at stations during the journey. The length of the testing scenario is set as 30 s. The first testing section used in scenario 1, which runs between $5'27''$ and $5'57''$, is chosen for this scenario. A detailed description of the testing section is given in Section 7.3.1.

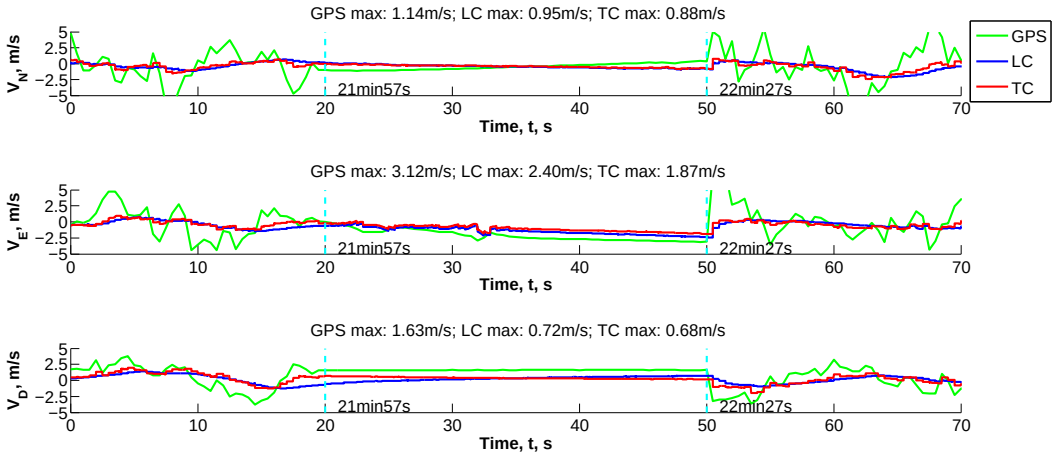
The GPS constellation status during the scenario is illustrated in Figure 7.11, where the visible satellites are marked in full colour and the blocked satellites are marked with dashed borders. A full description of the GPS constellation status over the 32 min operation is given in Section 6.12. Since the train is mainly moving along the east-west direction, two blockages are simulated along each side of the rail route blocking most of the constellation. As a result, only two satellites, G09 an G22, are visible during the testing scenario. G09 is ahead of the train with 120° azimuth angle and 80° elevation angle, whereas G22 is behind the train with 275° azimuth angle and 20° elevation

**Table 7.2.:** Normal System performances for the first testing section

| | | GPS | | LC | | TC | |
|---|---|---|---|---|---|---|---|
| | | Max | RMS | Max | RMS | Max | RMS |
| Position | N | 20.41 | 7.65 | 7.93 | 4.43 | 9.71 | 4.37 |
| (m) | E | 12.31 | 5.00 | 8.43 | 3.59 | 8.38 | 3.10 |
| | D | 14.18 | 6.61 | 10.63 | 5.97 | 11.67 | 5.87 |
| Velocity | N | 6.87 | 3.15 | 1.62 | 1.01 | 1.70 | 0.95 |
| (m/s) | E | 7.03 | 2.43 | 1.44 | 0.69 | 1.38 | 0.63 |
| | D | 3.05 | 1.19 | 1.17 | 0.61 | 1.37 | 0.62 |




**(a)** Position errors




**(b)** Velocity errors

**Figure 7.10.:** Position and velocity errors between $21'57''$ and $22'27''$ for scenario 1

angle.

Figure 7.12 shows the testing results for the testing scenario 2. The normal system performances of all systems for the testing section are listed in Table 7.2 on the preceding page. Both position errors and velocity errors are shown in the direction of north, east and down plotted against the simulation time, with position errors in meters and velocity errors in meters/second. The results for all tested systems are drawn respectively in green for the GPS only system, in blue for the loosely coupled system and in red for the tightly coupled system. The testing section is marked by dashed lines in all graphs. The benefits of having two extra satellites can be seen clearly from the improvements of all system performances.

No significant drifting pattern can be seen in the east direction of the velocity error, which leading to a much smaller drifting on the east position error comparing to other directions. This is because with two satellites along the direction of the route, the ranging measurements are able to constrain the error in the along track direction, in this case the east direction. The constraining effect of having two satellites along the route direction can be demonstrated by Figure 4.4b in Section 4.2.3.

The error drifting pattern is larger on both north and down directions in comparison to the results in scenario 1. For example, the maximum position errors for the north direction in scenario 1 are respectively 56.97 m for the GPS only system, 78.32 m for the loosely coupled system and 79.35 m for the tightly coupled system, whereas the maximum position errors for the north direction in scenario 2 are respectively 74.04 m for the GPS only system, 84.70 m for the loosely coupled system and 91.52 m for the tightly coupled system.

In addition, the contribution of the two satellites on positioning performance is also indicated by the variation of the position errors for both the GPS only system and the tightly coupled system during drifting in the east and the down direction. This is not clear in the loosely coupled system because it combines results from the GPS only system and the MEMS measurements, and the weighting of MEMS measurements in this case is heavier than the GPS only solution.

As already been pointed out in Section 7.3.1, a GPS only systems using least-squares based navigation processors still does not work under such a condition. Consequently, a loosely coupled system which integrates the solution from such GPS only systems can only access the inertial measurements during the signal difficult period. As a result, the performance of such a loosely coupled system resembles the performances shown by Figure 7.9.

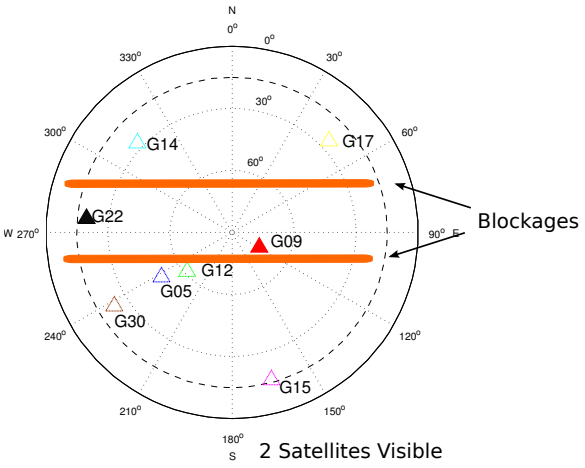
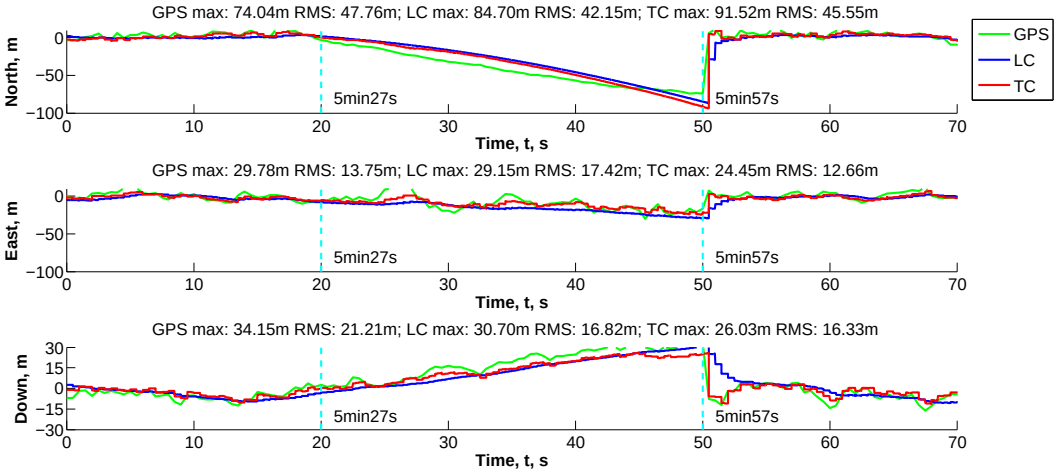


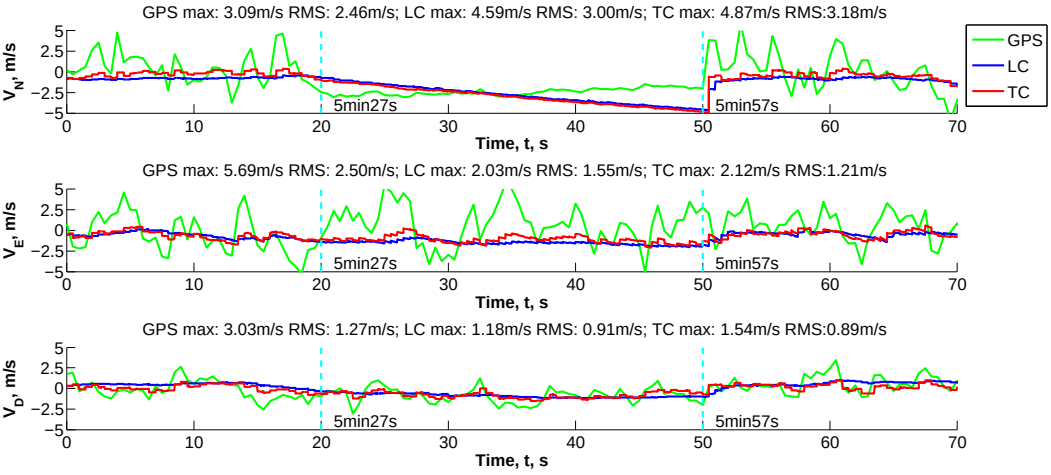

**Figure 7.11.:** Constellation status during scenario 2

**Table 7.3.:** Normal System performances for the second testing section

| | | GPS | | LC | | TC | |
|---|---|---|---|---|---|---|---|
| | | Max | RMS | Max | RMS | Max | RMS |
| Position | N | 18.38 | 7.65 | 12.44 | 4.42 | 14.07 | 4.37 |
| (m) | E | 11.69 | 5.00 | 7.80 | 3.59 | 9.19 | 3.10 |
| | D | 14.00 | 6.61 | 9.90 | 5.97 | 10.40 | 5.87 |
| Velocity | N | 5.51 | 3.14 | 1.20 | 1.01 | 1.45 | 0.96 |
| (m/s) | E | 4.89 | 2.43 | 1.41 | 0.69 | 1.69 | 0.63 |
| | D | 1.80 | 1.19 | 0.83 | 0.61 | 0.84 | 0.63 |




**(a)** Position errors




**(b)** Velocity errors

**Figure 7.12.:** Position and velocity error during scenario 2

### 7.3.3. Scenario 3: Bad Satellite Constellation Geometry

This scenario simulates situations on the railways where only half of the sky is visible but the number of available satellites is more than four. This mostly happens when a train travels through a urban area, or stops at stations during the journey. The length of the testing scenario is set as 30 s. The first testing section used in scenario 1, which runs between 5′27″ and 5′57″, is chosen for this scenario. A detailed description on the testing section is given in Section 7.3.1. When all visible satellites are in one half of the sky, this implies a bad satellite constellation geometry. The positioning performance for a GPS only system is therefore affected (see discussions in Section 4.2.3).

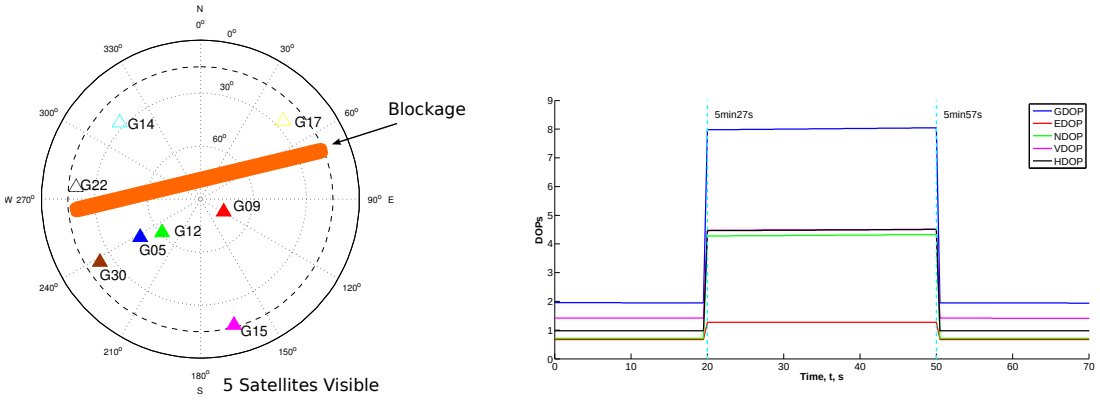
The GPS constellation status during the scenario is illustrated in Figure 7.13a, where the visible satellites are marked in full colour and the blocked satellite is marked with dashed borders. A full description of the GPS constellation status over the 32 min operation is given in Section 6.12. Since the train is mainly moving along the east-west direction, one blockage is simulated along the north side of the rail route blocking satellites on the northern part of the sky. As a result, five satellites in the southern part of the sky are visible during the testing scenario.

The DOP values are plotted in Figure 7.13b. All DOP values before and after the 30 s testing period indicate a reasonably good user-satellites geometry with all 8 satellites in view. During the testing period, increases on all DOP values demonstrates a decreased user-satellites geometry quality, since all visible satellites are in the southern half of the sky. It can be noticed that the increase on the EDOP value is not as large as other DOP values, which shows a still reasonably good satellites geometry in the east-west direction.

Figure 7.14 shows the test results for the testing scenario 3. The normal system performances of all systems for the testing section are listed in Table 7.2 on page 152. Both position errors and velocity errors are shown in the direction of north, east and down plotted against the simulation time, with position errors in meters and velocity errors in meters/second. The results for all tested systems are drawn respectively in green for the GPS only system, in blue for the loosely coupled system and in red for the tightly coupled system. The testing section is marked by dashed lines in all graphs. The effect of a decreased user-satellites geometry can be seen clearly from the differences among of performances in different directions.

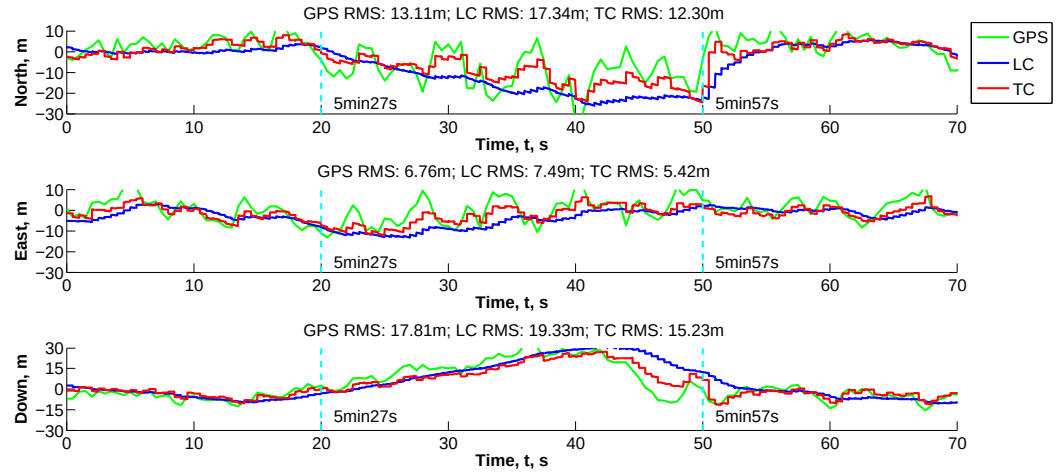
For position errors, an increase on the RMS errors can be seen on all three directions as a result of the performance decrease of the satellite ranging. However, the decrease of system performance in the east direction is not as large as other directions by comparing the RMS errors, 6.76 m for the GPS only system, 7.49 m for the loosely coupled system and 5.42 m for the tightly coupled system, with the normal operation performance, 5.00 m for the GPS only system, 3.59 m for the loosely coupled system and 3.10 m for the tightly coupled system. This is due to the reasonably good user-satellites geometry on the east-west direction as indicated by the DOP values in Figure 7.13b. On the other hand, the performance of the loosely coupled system is the worst among all three systems. This is caused by the existence of multiple Kalman filters in the system. When the performance of the GPS only system is affected, the tuned EKF parameters (see Section 7.1.3) are not longer optimal for the integration and the weighting between the GPS solution and the inertial measurements is thus affected. The tightly coupled system still performs as usual under such cases.

For velocity errors, slight increase on the RMS errors is seen on all three directions. Similar to performance under normal operation, the effect of having access to inertial measurements shows lower RMS errors for the integrated systems than the GPS only solution.

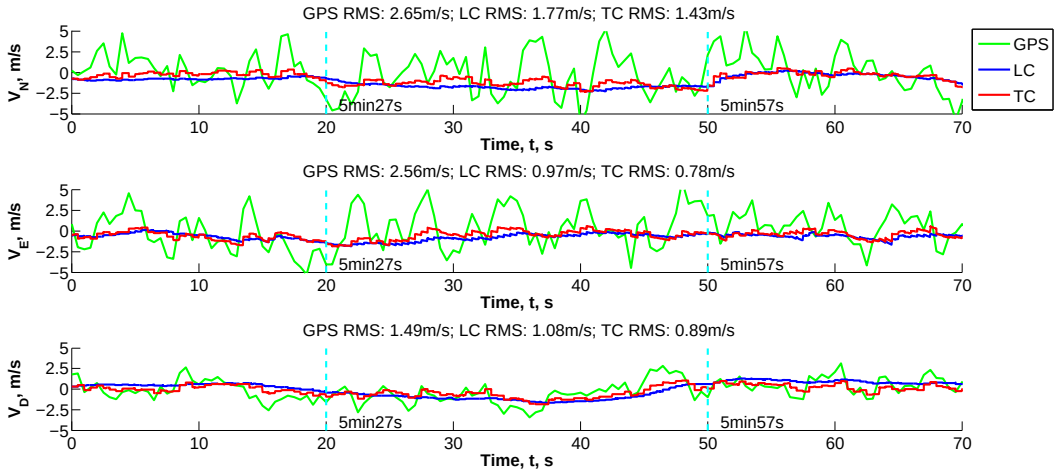



**(a)** Sky plot **(b)** DOPs

**Figure 7.13.:** Constellation status during scenario 3




**(a)** Position errors




**(b)** Velocity errors

**Figure 7.14.:** Position and velocity error during scenario 3

## 7.4. Conclusions

Investigations into different MEMS/GPS integration approaches for positioning on the railways have been reported in this chapter. The performances of different integration approaches are firstly compared and analysed under normal operation conditions. A filter based GPS only system is also operated along with the integrated systems, and its performance is analysed in comparison with other systems. Three testing scenarios covering common GNSS difficult situations on the railways are then proposed, and all systems are tested for all scenarios.

For the GNSS difficult situations when the number of available satellites is less than four, none of the three tested systems show satisfactory accuracy performances during the 30 s testing period comparing to the performance of the DRM-aided system under the same circumstances, as shown in the next chapter; whereas when there is a bad constellation geometry, the inclusion of MEMS sensor shows performance advantages in the velocity-domain. Because GPS serves as the main contributor to the system positioning accuracy, the performance of an integrated system in the position-domain is largely decided by the availability of GPS services. On the other hand, an integrated system shows better performance in the velocity estimation process, as the high update rate inertial measurements provide a better resolution than GPS.

For GPS only systems, although a filter based navigation processor enables a system to provide outputs under GNSS difficult situations, neither accuracy nor reliability of the solution can be guaranteed. Systems with a least-squares based navigation processor do not work under GNSS difficult situations with less than four satellites.

Under normal operating conditions, the inclusion of MEMS sensors in integrated systems decreases the variation on the position error indicating a smoother estimated trajectory, and greatly improves the quality of the velocity solution. However, when the GPS input is not available, the low performance of MEMS sensors cannot ensure a reliable system accuracy.

In terms of accuracy, assuming a loosely coupled system is tuned properly, a tightly coupled system performs only slightly better than a loosely coupled system. Nevertheless, the robustness of the tightly coupled integration approach shows advantages under situations where satellites are available but the number is less than four, as a result of using the satellite ranging measurement directly.

# 8. DRM-Aiding of GNSS/Inertial Navigation Systems

A Digital Route Model (DRM) is a database containing the position knowledge of surveyed points along the centre line of a rail route. The stored position knowledge of the surveyed points is normally expressed in the form of three dimensional coordinates. As discussed in Section 2.3.1.1, the railway is a guided transport system, which implies the movement of a train carriage is restricted to the rail tracks. For a positioning and navigation problem, a restricted movement of the object reduces the task of solving position in a three dimensional free space into a position seeking process along a predetermined trajectory, which is a one dimensional problem. The predetermined trajectory for the railways is described by a DRM. Under situations where GNSS is not available and low-cost inertial sensors cannot maintain a reliable positioning solution, the solution restriction provided a DRM can be helpful to maintain the reliability of the positioning solution.

Due to the nature of a guided transport system, the railway signalling and traffic control prefers the positioning solution provided by a train positioning system restrained to the correct rail track. This is particularly useful for complicated situations where multiple parallel or intersecting tracks exist. For a successful matching of the correct rail track, the cross track error of the position solution should be minimised.

The specification of a DRM is defined based on two aspects. The first aspect is the surveying quality of individual points in the database. The surveying quality decides the errors in the coordinates of each DRM point. In theory, a DRM with higher surveying quality is preferred. However, high surveying quality requires increased costs. The second aspect is the density of DRM points, in other words the distance between two successive DRM points. Longer distance between points decreases the restriction effects of a DRM, whereas short distance increases the work amount. Therefore, the building of a DRM is essentially a trade-off of both aspects.

This chapter investigates the performance of a DRM-aided GNSS/inertial navigation system. The integration of DRM with a tightly coupled MEMS/GPS system is described in Section 8.1. The general performance of a DRM-aided MEMS/GPS system is studied Section 8.2 by using DRMs with different specifications. Five test scenarios are proposed in Section 8.3 to study the performance of a DRM-aided system under various railways positioning situations.

## 8.1. Integrating DRM with a MEMS/GPS System

The centre line of a rail route defines the trajectory of a train journey. A DRM describes the centre line in a discrete way because of the limitation on its point density. In order to use DRM as a restriction of the position solution, it is assumed that the section of rail tracks between two successive DRM points can be represented by a straight line. Provided the distance between successive DRM points is small compared to the overall journey, this assumption is generally reasonable as a result of the fact that straight lines and long flat curves are preferred in rail track design conventions.

A tightly coupled MEMS/GPS system is chosen to integrate with DRM information in this chapter. Detailed description on the design of a tightly coupled MEMS/GPS system is given in

Section 7.2.

### 8.1.1. Using DRM Information

A rail route is divided into small sections by a DRM. When a positioning system is operating on the route, two problems must be firstly solved before a DRM can place restrictions on the positioning solutions. The first one is to determine the correct section of the DRM in which the train is running. The second one is how to use the two DRM point coordinates at each end of the section to form a restriction on the available position solution.

The first problem is essentially a database searching problem and needs a reliable searching algorithm. Because the initiation of the section searching requires knowledge of the train position in the first place, a recursive process of guessing and narrowing down the section is usually involved in the searching algorithm. A mis-matched section of the DRM introduces an integrity risk. An example of a reliable section searching algorithm is investigated in Zheng (2008). The focus of this research is the effect of having DRM information on an integrated positioning system. Hence it is assumed that the searching process is in general successful. Nevertheless, a mis-matched DRM scenario is proposed in the end to demonstrate the integrity risk of using mismatched DRM information.

The second problem can be solved in different ways based on different system designs. For a least-squares based GPS only system, researchers (Zheng, 2007; Simsky *et al.*, 2004) have shown that the minimum required number of satellites can be reduced by connecting the DRM coordinates with the estimated position through a geometric modelling process.

However, for a Kalman filter based algorithm, even when the number of available satellites is less than required, the system can still provide outputs by propagating its states through the system model, as shown by the results in Chapter 7. Thus the use of DRM information in a Kalman filter based algorithms should be focused on constraining the state estimates through its measurement model.

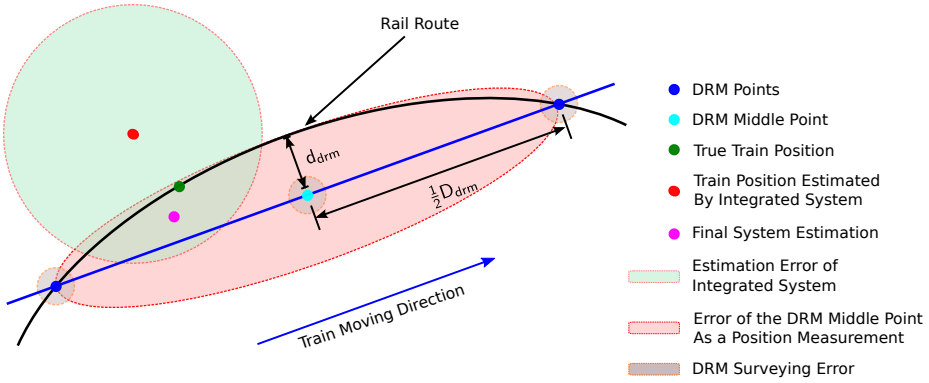



**Figure 8.1.:** The DRM-aiding process

Since most flat curves used in the rail track design are circular curves, the process of constraining the state estimates can be illustrated in Figure 8.1. When a train is travelling inside a DRM section, the middle point of the straight line between two DRM points can be used as an additional position measurement with constraining error models. Considering the middle point as a position measurement, its maximum position error ($\sigma_{max}$) occurs when the train enters or leaves the section in the direction of the DRM straight line, whereas its minimum position error ($\sigma_{min}$) occurs on the direction perpendicular to the DRM line when the train is at the middle of the circular curve. The maximum position error for the DRM middle point with respect to the true position of the train is

half of the distance between the two DRM points ($\frac{1}{2}D_{drm}$) as long as the train is still moving within the same section. The minimum error is the maximum discrepancy between the circular curve and the DRM line ($d_{drm}$) as marked in Figure 8.1.

The middle point of the straight line between two DRM points can be calculated using the DRM coordinates, and the quality of calculated coordinates can be obtained by propagating the standard deviations of the DRM points. When a Kalman filter combines positions of an integrated system estimation and the DRM middle points, the final estimated position is a weighted average between the two points and the weighted scheme takes considerations on the errors of both points. Hence a constraint on the final position solution is achieved.

### 8.1.2. Discussion on the Mid-point Approach for DRM-aiding

Among various methods of adopting the constraining information provided by DRM points, a mid-point approach is chosen in this research, i.e. the middle point of each DRM section is used with corresponding error covariance modelling to provide DRM-aiding for the integrated system. This process is illustrated in Figure 8.1. The mid-point approach, however, should not be considered as the only feasible way of applying DRM-aiding. Other methods, such as the direct use of the DRM points, could also be adopted if suitable searching algorithms for DRM sections are provided.

The main reason of the using mid-point approach in this research is the access to a reliable searching algorithm provided by previous research (Zheng, 2008). As part of a series of research projects carried out on DRM-aiding by the same research group, this research uses a searching algorithm previously developed in the group. Using an algorithm of which the reliability is proven reduces the chance of degraded system performance due to the effect of search algorithm failure and helps to identify the contribution to performance improvements that are introduced by DRM-aiding.

The approach that will actually be adopted when performing DRM-aiding in real time will depend on the form of the information provided and the searching algorithms available in the system, and may vary between different systems.

Because the usage of the DRM mid-points, the discrepancy between the rail track curve and the DRM mid-point, as marked by $d_{drm}$ in Figure 8.1, must be taken into account in the process of error covariance modeling. The error covariance modeling process used with the mid-point approach is described in Section 8.1.3. In that process, the discrepancy, $d_{drm}$, is modeled as the minimum position error of a DRM point as a positioning solution provided to the system when no random error is added, and has hence been taken into account of the error covariance modeling process.

In general, because the railway as a guided transport method is not tolerant to sharp turns of the vehicles during the journey, the rail track design tends to use straight lines and curves with large radii to ensure the smoothness of the train movement. As a result, discrepancies between the rail track curve and the DRM mid-points are usually small provided the intervals between DRM points are small.

An analysis of the scale of the discrepancies is performed in this section for the rail tracks used in this research to serve as the basis of the error covariance modeling carried out in the next section. The discrepancies, $d_{drm}$, are firstly generated for all DRM mid-points, statistics are then calculated for all $d_{drm}$. The calculated statistics include mean, standard deviation, minimum, maximum and median. Minimum, maximum and median are provided to show the full extent of the $d_{drm}$ changes throughout the journey. The process is repeated for all DRM intervals and the obtained statistics are listed in Table 8.1.

From the statistics listed in Table 8.1, it can be seen that the discrepancies increase along with the increasing of the DRM distance. The mean of the discrepancies for 10 m DRM is 0.021 m, whereas

**Table 8.1.:** Statistics of $d_{drm}$ for all DRM intervals

| DRM Distance (m) | Mean (m) | Standard Deviation (m) | Minimum (m) | Maximum (m) | Median (m) |
|---|---|---|---|---|---|
| 10 | 0.021 | 0.032 | $1.675 \times 10^{-6}$ | 0.553 | 0.011 |
| 50 | 0.155 | 0.220 | $0.637 \times 10^{-6}$ | 1.681 | 0.081 |
| 100 | 0.525 | 0.819 | $25.29 \times 10^{-6}$ | 6.603 | 0.249 |
| 500 | 11.286 | 17.813 | $113.7 \times 10^{-6}$ | 104.578 | 5.843 |

the mean for 500 m DRM is 11.286 m. For navigation applications, the scale of the discrepancies is in general around or smaller than the level of positioning errors of a single frequency GPS receiver, which is about 13 m at 95% confidence (RSSB, 2008).

Nevertheless, the possible maximum discrepancies increase dramatically as the DRM distance increases. The maximum for 500 m DRM is as large as 104.578 m. This is though not the general case. As can be seen from the standard deviations and medians for all DRM intervals, very large discrepancies only happen occasionally when the radii of the rail curves are small and the DRM intervals are very large. For instance, whilst the maximum for 500 m DRM is 104.578 m, the median is only 5.483 m, which indicates that majority of the discrepancies for 500m DRM is still below the error level of single frequency GPS positioning.

From the medians of the $d_{drm}$ of all DRM intervals, it can be shown that for short intervals such as 10 m and 50 m, the scale of the discrepancies is at the centimeter level and therefore in the noise level. For longer intervals such as 100 m and 500 m, the discrepancies become an observable error source and must be calibrated during the integration of DRM-aiding to avoid the introduction of unnecessary errors.

In comparison with other DRM-aiding approaches, especially the direct use of the DRM points, the mid-point approach introduces the additional discrepancies between the track trajectory and the mid-points as part of the intrinsic errors when using mid-points as extra position constraint on the integrated positioning solution. Although the scale of the discrepancy is generally below the positioning error level of single frequency GPS positioning, the discrepancy must be accounted during the error covariance modeling process. The following sections will show that with the discrepancy taken into account properly, the performance improvement introduced by DRM-aiding can be achieved using the mid-point approach. On the other hand, the direct use of the DRM points would ensure that points provided to the integrated system are always on the track and hence avoid the introduction of discrepancies. But the geometric relationship between the DRM straight line and easting and northing variances and covariance is lost. This complicates the translation of the DRM error eclipses from the cross and along track direction to the easting and northing direction and requires additional calculations.

### 8.1.3. Reforming DRM for an Integrated System

Following the description given in Section 8.1.1, the conventional DRM in the form of three dimensional coordinates only needs to be reformed before integrating with the tightly coupled system. As a result, the middle point ($\tilde{\mathbf{p}}^n_{drm}$) and the constraining error model ($\mathbf{R}_{drm}$) at each DRM section is used as the output from a DRM.

The coordinates of middle point at each DRM section can be calculated using the coordinates of

DRM points at both end of the section, denoted as the *front* point and *behind* point,

$$\tilde{\mathbf{p}}^n_{drm} = \frac{1}{2}(\tilde{\mathbf{p}}^n_{front} + \tilde{\mathbf{p}}^n_{behind}) \tag{8.1}$$

where all coordinates are expressed in the local navigation frame. The surveying quality of the DRM points are known respectively as $(\sigma^2_{front,N},\ \sigma^2_{front,E},\ \sigma^2_{front,D})$ and $(\sigma^2_{behind,N},\ \sigma^2_{behind,E},\ \sigma^2_{behind,D})$. Thus the quality of the middle point, $(\sigma^2_{drm,N},\ \sigma^2_{drm,E},\ \sigma^2_{drm,D})$, can be derived from (8.1),

$$\sigma^2_{drm,N} = \left(\frac{1}{2}\right)^2 \sigma^2_{front,N} + \left(\frac{1}{2}\right)^2 \sigma^2_{behind,N} \tag{8.2}$$

$$\sigma^2_{drm,E} = \left(\frac{1}{2}\right)^2 \sigma^2_{front,E} + \left(\frac{1}{2}\right)^2 \sigma^2_{behind,E} \tag{8.3}$$

$$\sigma^2_{drm,D} = \left(\frac{1}{2}\right)^2 \sigma^2_{front,D} + \left(\frac{1}{2}\right)^2 \sigma^2_{behind,D} \tag{8.4}$$

As described in Section 8.1.1, the maximum position error of the DRM middle points with respect to the true position of the train is half the length of the DRM straight line, hence

$$\sigma_{max} = \frac{1}{2} D_{drm} \tag{8.5}$$

where $D_{drm}$ is specified in Section 6.5. The minimum position error is the maximum discrepancy between the circular curve and the DRM line ($d_{drm}$), and is calculated using geometrical relationships

$$\sigma_{min} = d_{drm} = \bar{R}_c - \sqrt{\bar{R}_c^2 - (\frac{1}{2} D_{drm})^2} \tag{8.6}$$

where $\bar{R}_c$ is the average radius of all circular curves along fitted the simulated rail route, as shown in Section 6.6.2. The direction of the DRM straight line is expressed by its heading towards the north, $\psi_{drm}$, and is calculated with

$$\psi_{drm} = \arctan 2(\Delta E_{drm},\ \Delta N_{drm}) \tag{8.7}$$

where $\Delta E_{drm}$ and $\Delta N_{drm}$ are the easting and northing difference of the DRM points at both ends of the section.

The maximum and minimum standard position errors of the DRM middle points with respect to the true position of the train is expressed in the direction of the DRM straight line. They need to be transformed into the local navigation frame before been used in the constraining error model. The transformation uses the concept of error eclipses. Following Cross (1983), the maximum and minimum position error can be expressed as

$$\sigma^2_{max} = \cos^2 \psi_{drm} \sigma^2_n + \sin^2 \psi_{drm} \sigma^2_e + 2 \cos \psi_{drm} \sin \psi_{drm} \sigma_{en} \tag{8.8}$$

$$\sigma^2_{min} = \sin^2 \psi_{drm} \sigma^2_n + \cos^2 \psi_{drm} \sigma^2_e - 2 \cos \psi_{drm} \sin \psi_{drm} \sigma_{en} \tag{8.9}$$

where $\sigma^2_n$, $\sigma^2_e$ and $\sigma_{en}$ are respectively the standard error in north, east and the covariance of east and north. The direction of the DRM straight line can be expressed in terms of $\sigma^2_n$, $\sigma^2_e$ and $\sigma_{en}$ as well

$$\tan(2\psi_{drm}) = \frac{2\sigma_{en}}{\sigma^2_n - \sigma^2_e} \tag{8.10}$$

By solving (8.8), (8.9) and (8.10), the values for $\sigma^2_n$, $\sigma^2_e$ and $\sigma_{en}$ can be obtained. Finally, the

constraining error model of the DRM middle point can be given as

$$\mathbf{R}_{drm} = \begin{bmatrix} (\sigma_{drm,N} + \sigma_n)^2 & \sigma_{en} & 0 \\ \sigma_{en} & (\sigma_{drm,E} + \sigma_e)^2 & 0 \\ 0 & 0 & \sigma^2_{drm,D} \end{bmatrix} \tag{8.11}$$

which is used as the measurement error covariance matrix for the integration Kalman filter. The elements of $\mathbf{R}_{drm}$ of a DRM with 10 m point intervals and 1 m point accuracy is plotted in Figure 8.2 as an example. As can be seen from the figure, the variation of each element accompanies the changing of the heading as the train goes through its journey. For instance, the heading of the train is mainly along the east-west direction between 5′ and 10′ as shown in Figure 6.6, and the easting element $\sigma_{drm,E} + \sigma_e$ is also larger than the northing element $\sigma_{drm,N} + \sigma_n$ for the same period indicating a constraint on the cross track direction.

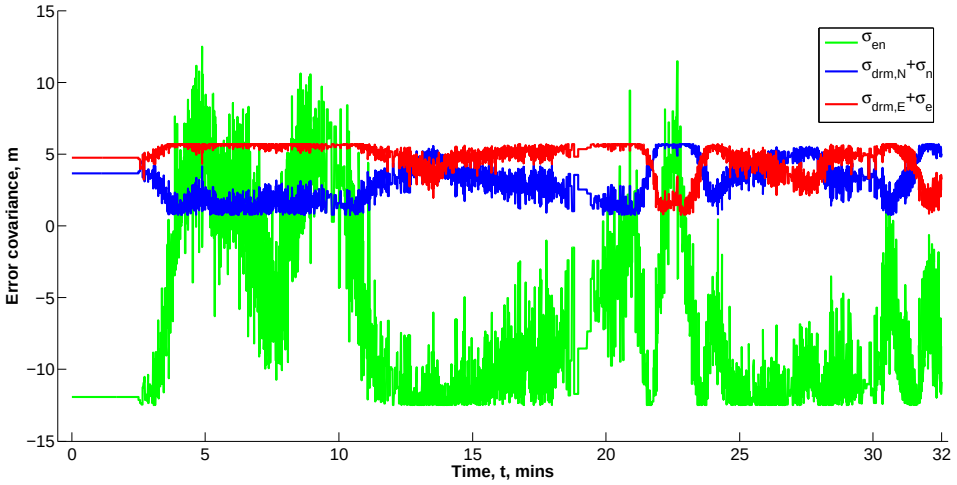



**Figure 8.2.:** The constraining error model of a DRM with 10 m point intervals and 1 m point errors

### 8.1.4. System Description for DRM-Aided MEMS/GPS Integration

The system architecture of DRM-aided MEMS/GPS system is shown in Figure 8.3. The system is a centralised system based on the tightly coupled MEMS/GPS system as described in Section 7.2. DRM middle point measurements and the constraining error model is fed into the integration EKF at the same update rate as GPS ranging measurements.

As mentioned in Section 8.1.1, the focus of using DRM information in the integration process is to constrain the state estimates through the measurement model of integration EKF. Hence the system model of the integration EKF for a DRM-aided MEMS/GPS system can be described by (7.10), (7.11) and (7.12).

The measurement model of a DRM-aided MEMS/GPS system needs to include the constraining information from the DRM. Hence the measurement vector for epoch $k$, assuming there are $K$

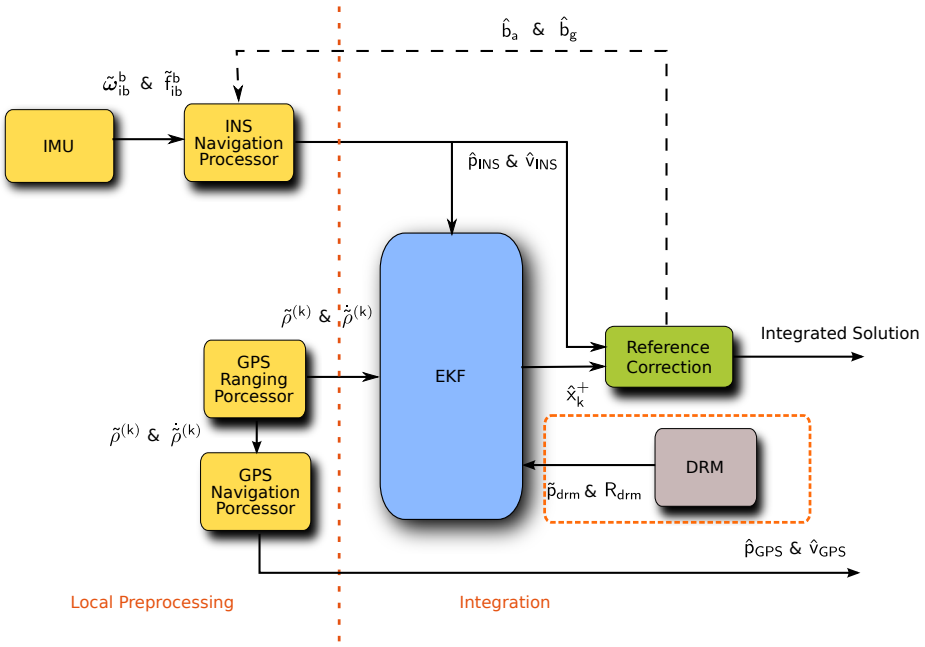



**Figure 8.3.:** DRM-aided tightly coupled MEMS/GPS integration

satellite in view, is given as

$$\delta\mathbf{z}_k = \begin{bmatrix} \delta\mathbf{z}_{\rho,k} \\ \delta\mathbf{z}_{\dot{\rho},k} \\ \tilde{\mathbf{p}}_{drm} - \hat{\mathbf{p}}_{INS} \end{bmatrix} \tag{8.12}$$

where $\delta\mathbf{z}_{\rho,k}$ and $\delta\mathbf{z}_{\dot{\rho},k}$ are respectively defined in (7.14) and (7.15). The measurement matrix for the DRM information is given as

$$\mathbf{H}_{drm} = \begin{bmatrix} \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \tag{8.13}$$

Thus the system measurement matrix can be written as

$$\mathbf{H}_k = \begin{bmatrix} \mathbf{H}_{GPS} \\ \mathbf{H}_{drm} \end{bmatrix} \tag{8.14}$$

where $\mathbf{H}_{GPS}$ is defined in (7.16).

## 8.2. General Performance Analysis

Based on the integration architecture described in Section 8.1.4, the general performance of a DRM-aided MEMS/GPS system is investigated in this section. The study of the system performance is done by varying two aspects of the DRM specifications: the point intervals and point accuracy levels.

### 8.2.1. DRMs with Different Point Intervals

Integrated systems aided by DRMs with different point intervals are firstly tested. The DRM point interval used for this research includes 10 m, 50 m, 100 m and 500 m. The point accuracy for all DRMs used in this section is assumed to be the same and is set as 1 m. All systems operate through the 32 min train journey with the access to full visible satellite constellation and no physical obstructions simulated. Positioning performances of all proposed systems are compared at the end.

A summary of system performances aided by DRMs with different point intervals is given in Table 8.2. The performance of the system aided by the DRM with 10 m point interval is plotted in Figure 8.4 and Figure 8.5 in this section as an example, whereas detailed performance figures on other systems are given in figures in Appendix D.1.

Two graphs in Figure 8.4 from the top to the bottom respectively show the position and velocity errors in the direction of east, north and down, and graphs in Figure 8.5 show the position and velocity errors in along track and cross track direction. All error information is plotted against the simulation time. The errors in the GPS solution are plotted in Figure 8.4 in green line providing a comparison. In addition, the predicted $1\sigma$ boundary of the system positioning solution error, provided by the state error covariance matrix, is plotted in each graph as the red lines. Statistics for the integrated positioning error and the predicted $1\sigma$ are also calculated and marked in each graph.

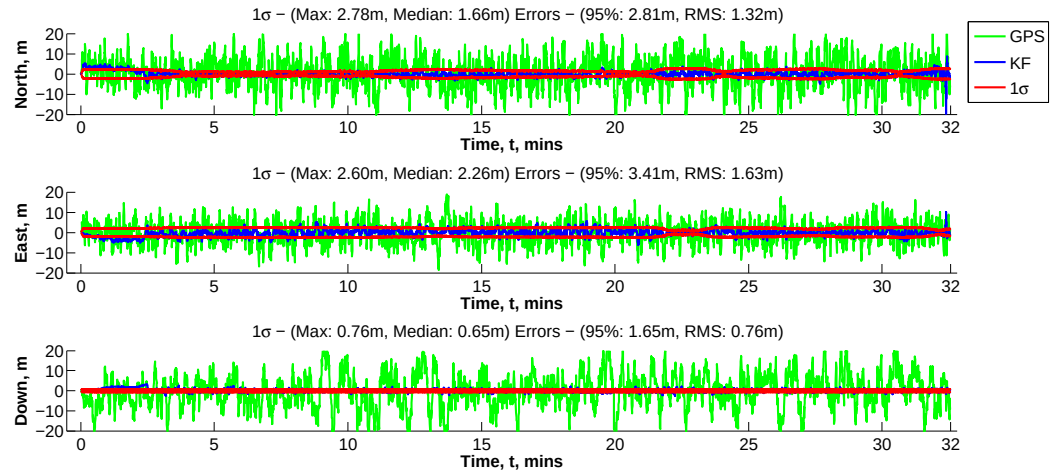
For position errors, the improvements can be seen clearly on all three directions of the local navigation frame in Figure 8.4a. For example, the RMS position error is reduced from 5.49 m using a tightly coupled system (see Figure 7.5a on page 145) to 1.32 m using a DRM-aided system. The improvement of positioning accuracy on the height estimation is particularly effective by achieving a sub-meter level accuracy with DRM-aiding. This is because the height change over the journey is less than 10 m (see Figure 6.6). Using DRM with 1 m point accuracy and 10m point interval under such conditions can provide a very strong constraint on the growth of position error on the height direction in comparison to other two directions.

More importantly, the position error on cross track direction is well constrained through DRM-aiding. This can be seen from Figure 8.5a, where a sub-meter level accuracy on the cross track direction (RMS error 0.89 m) is achieved in comparison with a 1.19 m RMS error on the along track direction. In addition, the cross track error is less noisy than the along track error, and is well bounded by the predicted $1\sigma$ limits.

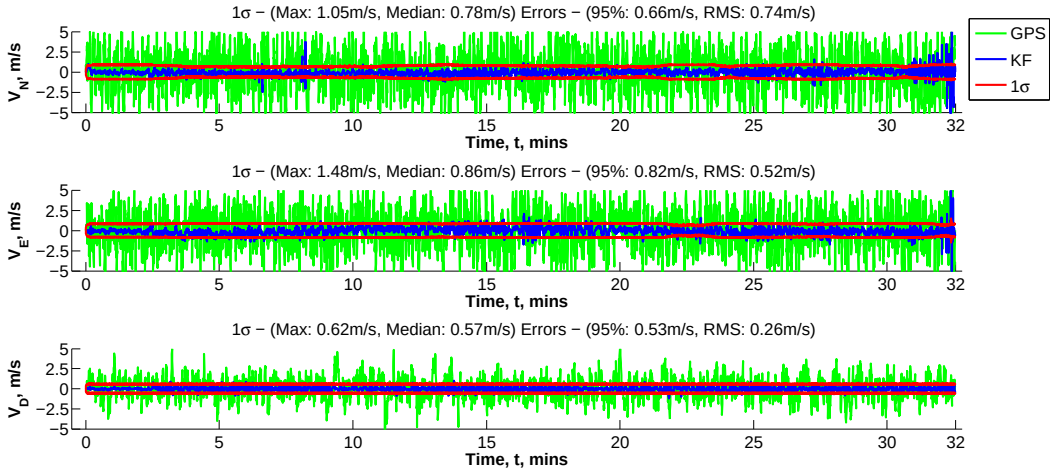
For velocity errors, small improvements can be seen on all three directions in Figure 8.4b. The RMS estimation error is reduced from 0.96 m/s in north, 0.73 m/s in east and 0.70 m/s in down using a tightly coupled system (see Figure 7.5b on page 145) to 0.74 m/s in north, 0.52 m/s in east and 0.26 m/s in down using a DRM-aided system. On the other hand, the estimation errors in cross track direction is less noisy than along track direction by showing a smaller 95% error.

Four DRMs with increasing point intervals are used to aid the integrated system in order to find a suitable point interval. Results listed in Table 8.2 indicates although in general an increased point interval decreases the overall performance of a DRM-aided system, an improved positioning results can still be achieved with reasonably long point intervals. For instance, an integrated system aided by a DRM with 100 m point interval could still achieve 4.56 m along track error and 3.46 m cross track error. However, when the point interval is too large, the assumption of using a straight between successive DRM points to represent the rail route is not necessarily true. Hence the constraining error model is no longer optimal. As shown in the table, a system using a DRM with 500 m point interval produces 7.56 m along track error and 9.59 m cross track error. The positioning performance of the system is actually decreased and the constraining effect on cross track direction is lost.

Moreover, the positioning performance decrease in the down direction is lower than other two




**(a)** Position errors



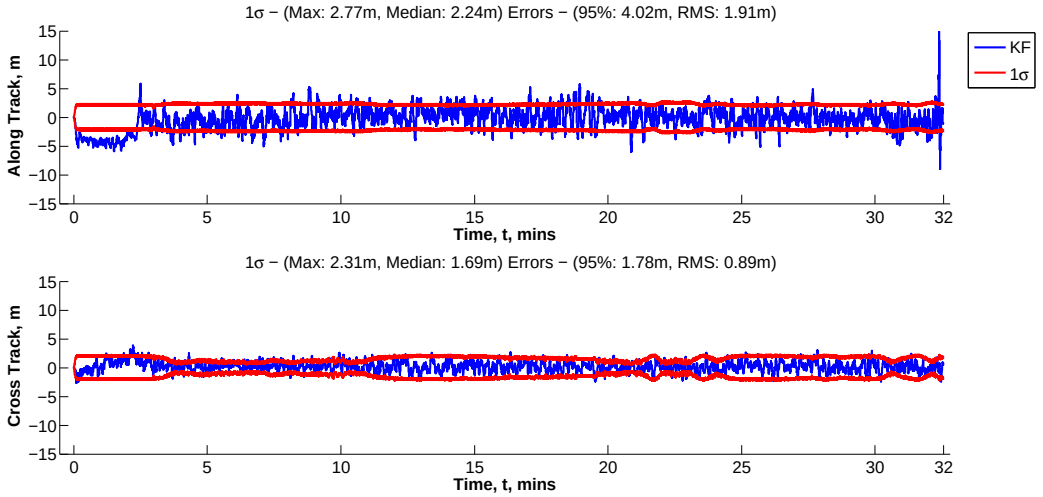

**(b)** Velocity errors

**Figure 8.4.:** General performance of a system using a DRM with 10 m point intervals and 1 m point errors

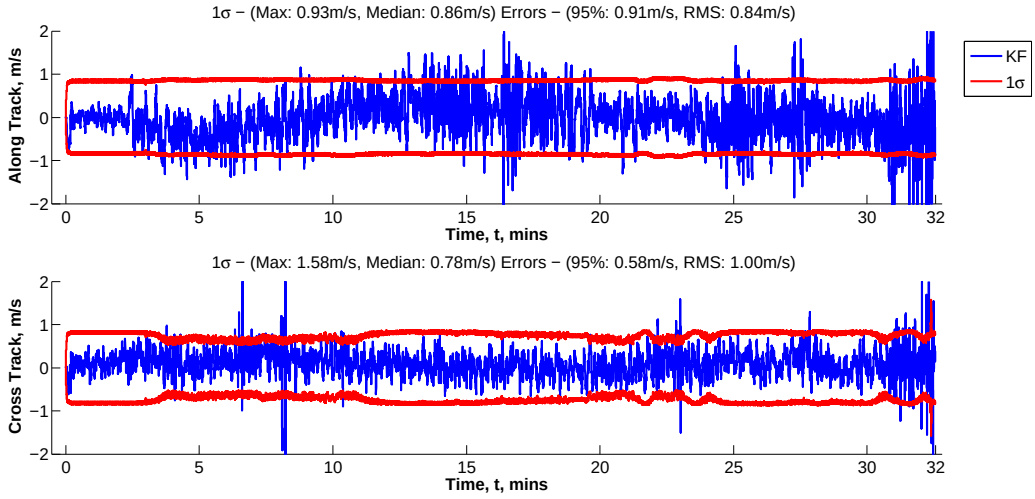
**Table 8.2.:** RMS errors for integrated system aided by DRMs with different point intervals

| | | Tightly coupled | 10 m | 50 m | 100 m | 500 m |
|---|---|---|---|---|---|---|
| Position | N | 5.49 | 1.32 | 3.60 | 4.27 | 10.38 |
| (m) | E | 4.02 | 1.63 | 3.63 | 3.81 | 6.44 |
| | D | 7.17 | 0.76 | 0.77 | 0.81 | 1.01 |
| | Along Track | 5.48 | 1.91 | 4.48 | 4.56 | 7.56 |
| | Cross Track | 4.94 | 0.89 | 2.47 | 3.46 | 9.59 |
| Velocity | N | 0.96 | 0.74 | 0.82 | 0.89 | 1.36 |
| (m/s) | E | 0.73 | 0.52 | 0.68 | 0.72 | 0.98 |
| | D | 0.70 | 0.26 | 0.23 | 0.19 | 0.11 |
| | Along Track | 1.03 | 0.84 | 0.95 | 0.98 | 1.16 |
| | Cross Track | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

directions. Even for a system using a DRM with 500 m point interval, a 1.01 m position error and a 0.11 m/s velocity error could still be achieved in the down direction. This can be explained by the small height variation (less than 10 m over a 33 km journey) during the operation, in which cases the constraint of position error growth provided by a DRM with 1m accuracy and 10m interval is stronger on the height direction to other two directions. For a DRM point with 1 m accuracy over a less than 100 m section, the constraining effect applied by the DRM provides a better resolution on the height changes than other sensors in the system.




**(a)** Position errors




**(b)** Velocity errors

**Figure 8.5.:** Along track and cross track performance of a system using a DRM with 10 m point intervals and 1 m point errors

From the performance statistics, a 10 m DRM provides the best performance (3284 used over a 33 km journey), whereas a 50 m DRM provides a good balance between performance and construction effort (658 points used over a 33 km journey). A DRM with smaller point interval presents a higher performance requirement for the searching algorithm (searching speed, reliability) and also produces a large amount of unused DRM sections when the train is moving at a high speed. Therefore, a DRM using different point intervals at different part of the journey, i.e. only using small intervals at critical parts of the journey such as intersections and stations, can be an ideal solution.

### 8.2.2. DRMs with Different Accuracy Levels

The performances of integrated systems aided by DRMs with different point accuracy levels are investigated in this section. The DRM point accuracy level used for this research includes 0.1 m, 1 m and 10 m. The point interval for all DRMs used in this section is assumed to be the same and is set as 10 m. All systems operate through the 32 min train journey with the access to full visible satellite constellation and no physical obstructions simulated. Positioning performances of all proposed systems are compared at the end.

A summary of system performances aided by DRMs with different point accuracy levels is given in Table 8.3. The performance of the system aided by the DRM with 10 m point accuracy level is plotted in Figure 8.6 and Figure 8.7 in this section as an example, whereas detailed performance figures on other systems are given in figures in Section D.2.

Two graphs in Figure 8.6 from the top to the bottom respectively show the position and velocity errors in the direction of east, north and down, and graphs in Figure 8.7 show the position and velocity errors in along track and cross track direction. All error information is plotted against the simulation time. The errors in the GPS solution is plotted in Figure 8.6 in green line providing a comparison. In addition, the predicted $1\sigma$ boundary of the system positioning solution error, provided by the state error covariance matrix, is plotted in each graph as the red lines. Statistics for the integrated positioning error and the predicted $1\sigma$ are also calculated and marked in each graph.

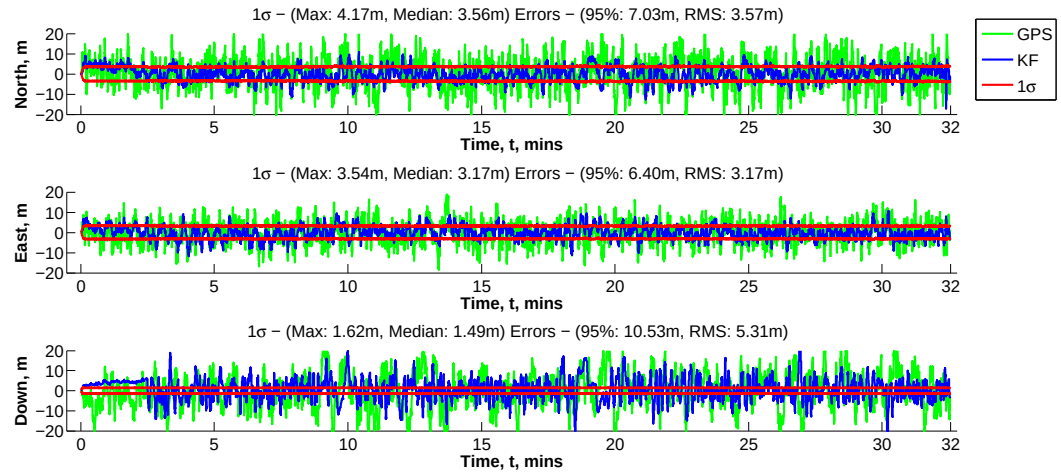
No significant performance improvement is shown by using a DRM with 0.1 m point accuracy, as shown in Table 8.3, compared to the performance of using a DRM with 1 m point accuracy. Although a slightly smaller sub-meter level position accuracy (RMS error 0.57 m) is achieved in the cross track direction, the along track position error (RMS error 2.07 m) is slightly larger than the error produced by using a 1 m accuracy DRM (RMS error 1.91 m). This can be explained by the accuracy limitation of the DRM constraining error model, which is caused by simplifications used during the calculation process for the constraining error model such as using averaged circular curve radius (see Section 8.1.3).

On the other hand, using a less accurate 10 m DRM causes the system performance to deteriorate rapidly, as can be seen from Figure 8.7 and Table 8.3. The system accuracy on all directions of the local navigation frame decreases as the DRM accuracy decreases, especially on the down direction, where the position error increases from sub-meter level to 5.31 m. The performance change on the down direction is mainly due to the small height change during the journey. A 10 m error on the height of DRM middle points measurements is too large for a journey with less than 10 m height changes, and hence the height information provided by the DRM is less useful during the integration. In addition, positioning solution in the cross track direction is less constrained when using a less accurate DRM. As a result, the DRM-aiding is not as effective as expected.

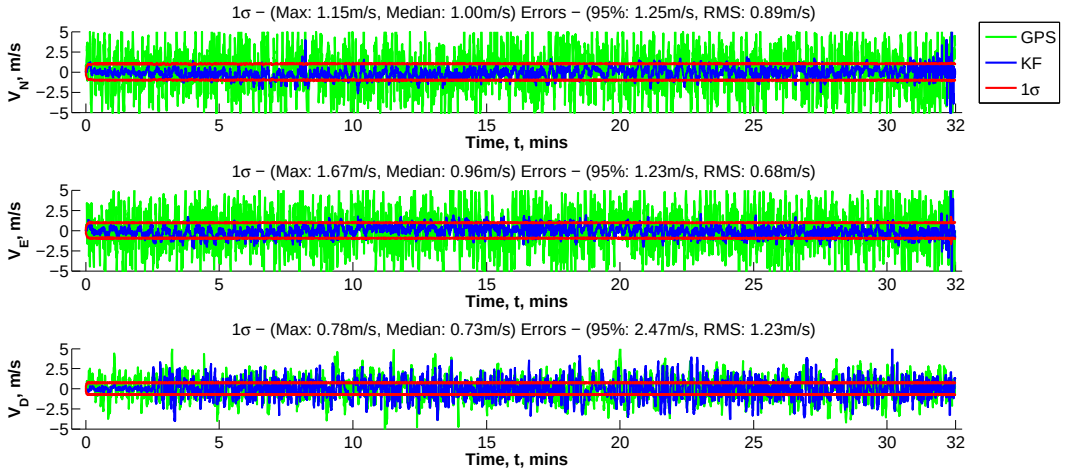
Based on the performance statistics, 1 m accuracy level is the most suitable for DRM-aiding among all three proposed candidates. Since a more accurate DRM does not provide better positioning performance and a less accurate DRM deteriorates the performance rapidly, the 1 m accuracy level provides a good balance between the system performance and construction effort.

## 8.3. Test Scenarios

Five test scenarios are proposed in this section including typical situations on the railways where GNSS only positioning performance can be affected. Two types of systems are used for testing under proposed scenarios: a tightly coupled integrated system using MEMS inertial units and the satellite ranging measurements and a DRM-aided MEMS/GPS system. Detailed description on




**(a)** Position errors

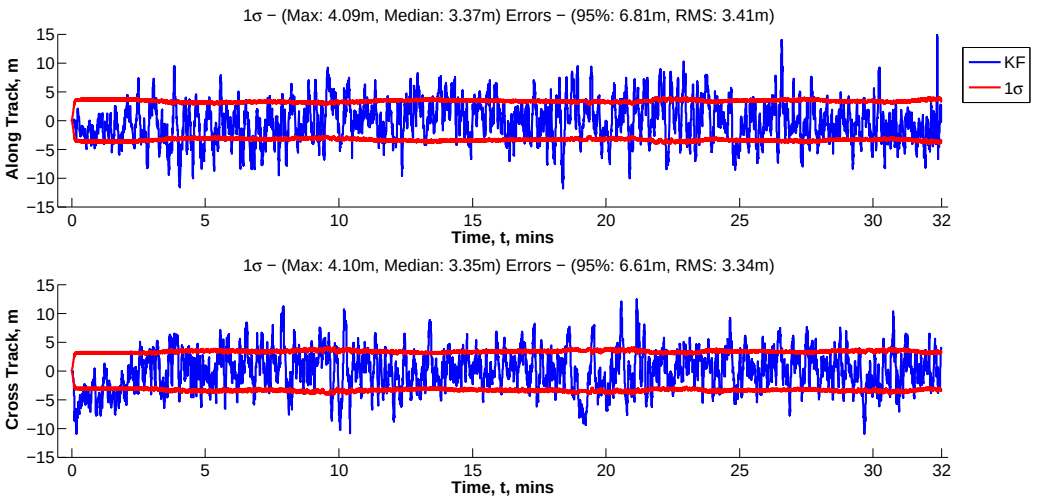



**(b)** Velocity errors

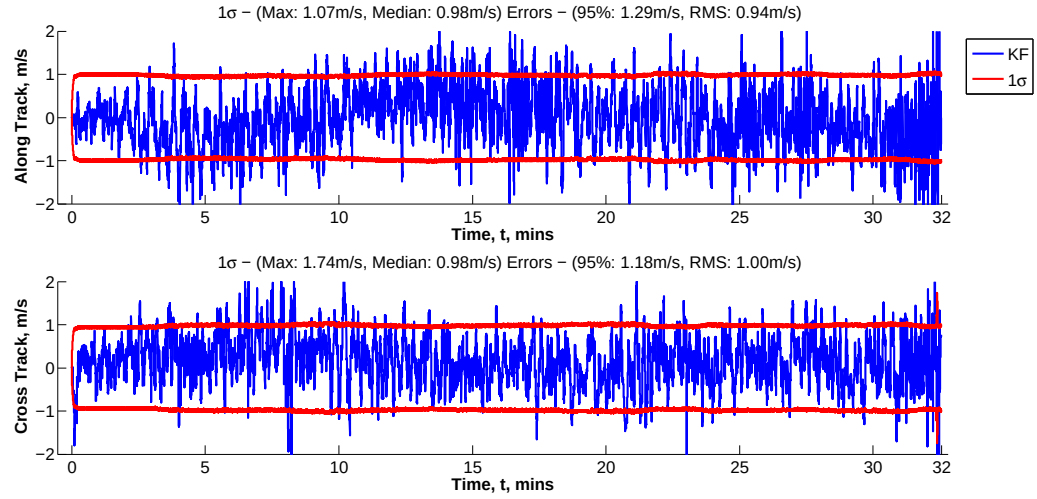
**Figure 8.6.:** General performance of a system using a DRM with 10 m point intervals and 10 m point errors

**Table 8.3.:** RMS errors for integrated system aided by DRMs with different point accuracy levels

| | | Tightly coupled | 0.1m | 1 m | 10 m |
|---|---|---|---|---|---|
| Position | N | 5.49 | 1.42 | 1.32 | 3.57 |
| (m) | E | 4.02 | 1.61 | 1.63 | 3.17 |
| | D | 7.17 | 0.25 | 0.76 | 5.31 |
| | Along Track | 5.48 | 2.07 | 1.91 | 3.41 |
| | Cross Track | 4.94 | 0.57 | 0.89 | 3.34 |
| Velocity | N | 0.96 | 0.72 | 0.74 | 0.89 |
| (m/s) | E | 0.73 | 0.50 | 0.52 | 0.68 |
| | D | 0.70 | 0.08 | 0.26 | 1.23 |
| | Along Track | 1.03 | 0.84 | 0.84 | 0.94 |
| | Cross Track | 1.00 | 1.00 | 1.00 | 1.00 |




**(a)** Position errors



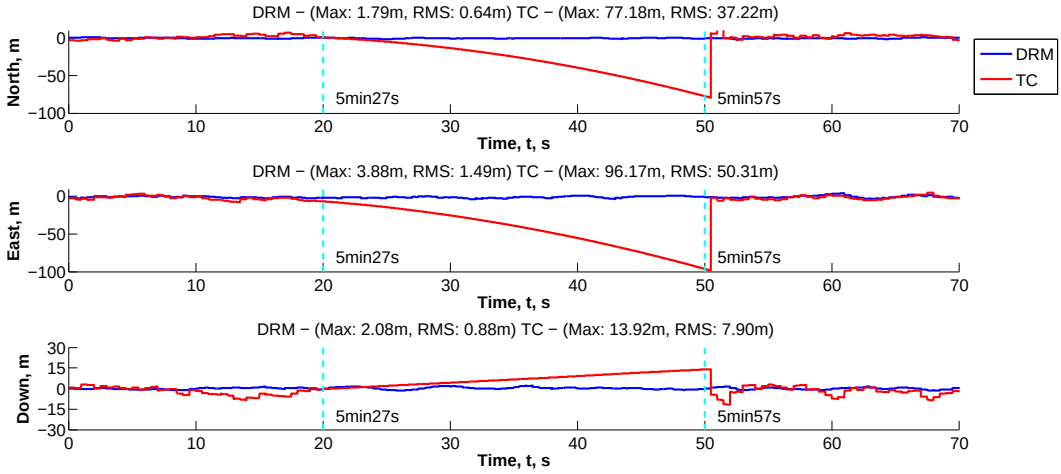

**(b)** Velocity errors

**Figure 8.7.:** Along track and cross track performance of a system using a DRM with 10 m point intervals and 10 m point errors

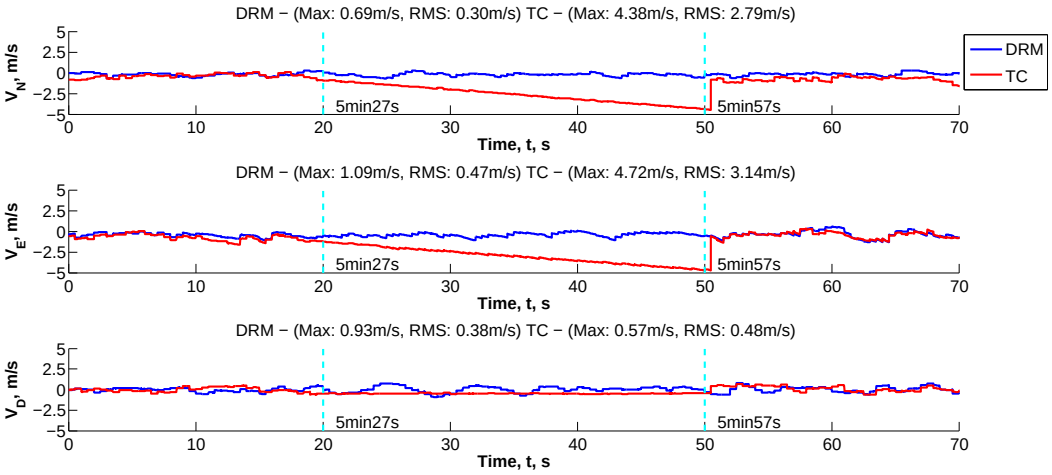
both systems can be respectively found in Section 7.2 and Section 8.1.4. A DRM with 10 m point interval and 1 m accuracy level is used for aiding in all proposed scenarios. The performances of both systems are then compared with each other and the performance of their normal operations within the same test period when no interruptions such as signal outages or lost and DRM failure would affect the system.

### 8.3.1. Scenario 1: Loss of GNSS Signals

This scenario simulates situations on the railways where the sky is completely blocked and GNSS signals are lost. This mostly happens when a train travels through a tunnel. According to the general GNSS availability study performed by LOCASYS (see Section 2.3.2), complete GNSS outage on the railways usually only last for a short period, i.e. less than a minute. Therefore the length of the test scenario is set as 30 s. Two sections of during the overall operation period are chosen for test scenario 1. The first section is chosen between 5′27″ and 5′57″ of the 32 min operation time, whereas the second section is chosen between 21′57″ and 22′27″. An overview of the 32 min




**(a)** Position errors



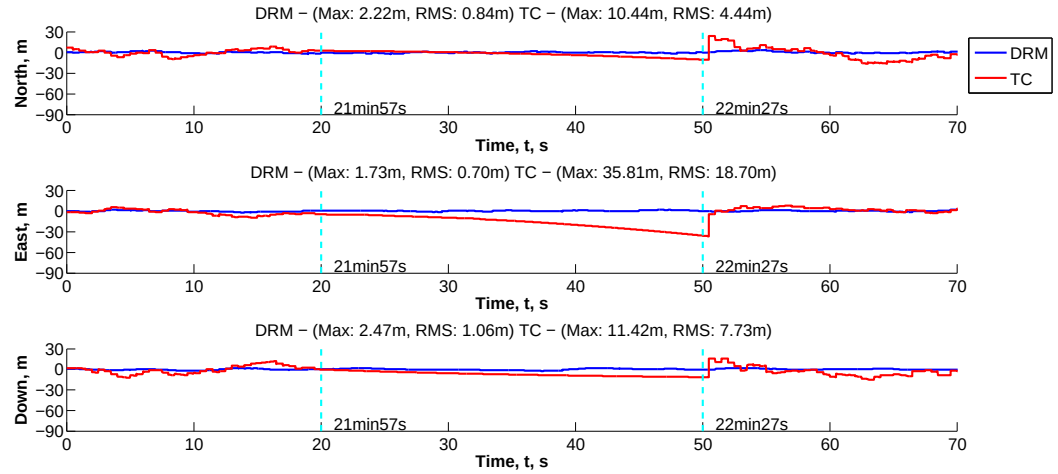

**(b)** Velocity errors

**Figure 8.8.:** Position and velocity errors between $5'27''$ and $5'57''$ for scenario 1

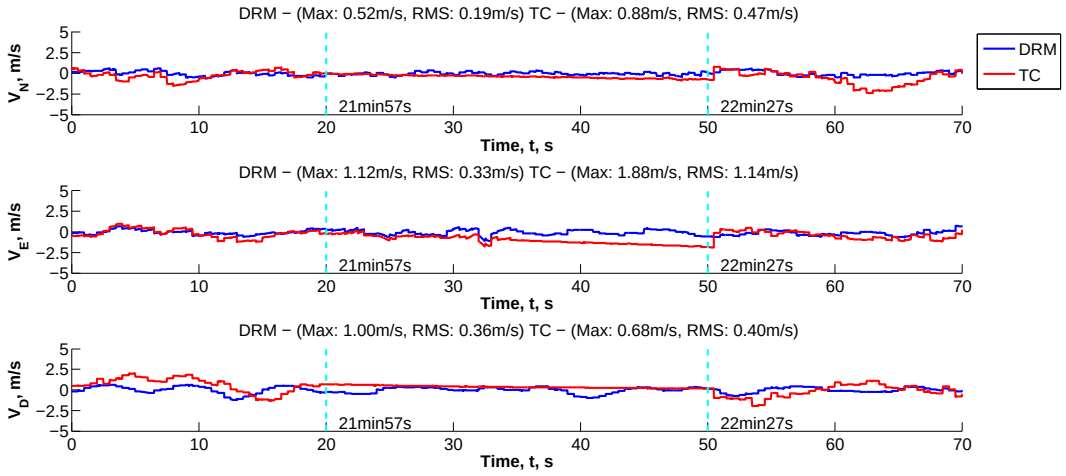
**Table 8.4.:** Normal System performances for the first test section

| | | DRM | | TC | |
|---|---|---|---|---|---|
| | | Max | RMS | Max | RMS |
| Position | N | 1.76 | 0.61 | 9.71 | 4.37 |
| (m) | E | 3.51 | 1.49 | 8.38 | 3.10 |
| | D | 2.23 | 0.88 | 11.67 | 5.87 |
| Velocity | N | 0.67 | 0.29 | 1.70 | 0.95 |
| (m/s) | E | 1.11 | 0.46 | 1.38 | 0.63 |
| | D | 0.86 | 0.38 | 1.37 | 0.62 |

operation time is illustrated in Figure 6.6 of Section 6.4, whereas descriptions on both test sections are given in Section 7.3.1.




**(a)** Position errors




**(b)** Velocity errors

**Figure 8.9.:** Position and velocity errors between $21'57''$ and $22'27''$ for scenario 1

Figure 8.8 on the preceding page and Figure 8.9 on the current page respectively show the test results for the first and the second test sections. The normal system performances of all systems for both test sections are listed respectively in Table 8.4 on the preceding page and Table 8.5 on the next page. Both position errors and velocity errors are shown in the direction of north, east and down plotted against the simulation time, with position errors in meters and velocity errors in meters/second. The results for all tested systems are drawn in blue for the DRM-aided system and in red for the tightly coupled system. The test section is marked by dashed lines in all graphs.

While the positioning error of a tightly coupled system increases rapidly, a DRM-aided system shows the capability of maintaining a reliable positioning solution during the satellite signal outage. For the first test section, RMS position errors of 0.64 m in the north, 1.49 m in the east and 0.88 m in the down can be achieved by a DRM-aided system. The RMS position errors for a DRM-aided system under the normal operating conditions is listed in Table 8.4 as 0.61 m in the north, 1.49 m in the east and 0.88 m in the down. No performance degradation can be seen from the RMS error

statistics. Since the train is mainly moving along the east-west direction during the first test section, a smaller error in the north direction indicates a constrained cross track error. This shows that the DRM constraining effect is also maintained under satellite signal outage situations.

The error growth of a tightly coupled system is related to the dynamic status of the train during satellite signal outages, as discussed in Section 7.3.1, because of MEMS sensor error characteristics. The DRM position errors for the second test section, where the train is moving at a low speed, is shown by Figure 8.9 as 0.84 m in the north, 0.70 m in the east and 1.06 m in the down. By comparing these results with the DRM-aided system performance in the first test section, no strong dependency on the train dynamics can be seen. The drifting of MEMS sensors without GPS calibration can be well controlled by the DRM aiding information.

**Table 8.5.:** Normal System performances for the second test section

| | | DRM | | TC | |
|---|---|---|---|---|---|
| | | Max | RMS | Max | RMS |
| Position | N | 2.22 | 0.84 | 9.71 | 4.37 |
| (m) | E | 1.73 | 0.70 | 8.38 | 3.10 |
| | D | 2.47 | 1.06 | 11.67 | 5.87 |
| Velocity | N | 0.52 | 0.19 | 1.70 | 0.95 |
| (m/s) | E | 1.12 | 0.33 | 1.38 | 0.63 |
| | D | 1.00 | 0.36 | 1.37 | 0.62 |

### 8.3.2. Further Result Analysis for Scenario 1

Section 8.3.1 presented the system performance of a DRM-aided integrated system under the situations of complete loss of GNSS signals based on two 30s long outage periods. The two periods is chosen because of the orientation of the tracks during the outages. The effect of DRM-aiding is can therefore be shown distinctively with respect to the orientation of the tracks. A further analysis is carried out in this section to further investigate the performance of a DRM-aided system under the situations of complete loss of GNSS signals, and the results are presented in this section to further enhance previous conclusions on the performance improvement introduced by DRM-aiding

**Table 8.6.:** Statistics for the position RMS errors of the DRM-aided system for all disjoint 30s periods

| | Mean (m) | Standard Deviation (m) | Minimum (m) | Maximum (m) | Median (m) |
|---|---|---|---|---|---|
| Easting | 1.062 | 0.386 | 0.570 | 2.597 | 0.977 |
| Northing | 1.017 | 0.314 | 0.548 | 2.000 | 0.989 |
| Down | 0.625 | 0.150 | 0.350 | 1.109 | 0.596 |
| Along Track | 1.259 | 0.415 | 0.690 | 2.599 | 1.137 |
| Cross Track | 0.791 | 0.157 | 0.504 | 1.128 | 0.819 |

The GNSS signal outages are simulated for all possible 30s disjoint sections during the train journey. For each individual outage simulation the root mean square errors of the positioning solution is calculated for all directions including easting, northing, down, along- and cross-track

direction. Statistics, such as mean, standard deviation, median, minimum and maximum, are calculated for all simulated 30s periods. The same simulation is performed for both the tightly integrated system and the DRM-aided system. The statistics are listed respectively in Table 8.6, 8.7, 8.8 and 8.9.

As can be seen from the statistics listed in all tables in this section, while the RMS positioning error of a tightly coupled system increases significantly comparing to its nominal performance with no signal outages, the performance degradation of a DRM-aided system is small. For instance, the RMS positioning error on the easting direction for the tightly coupled system simulated in this research increases from its nominal value 4.02 m (indicated in Figure 7.5a) to an averaged 35.406 m (indicated in Table 8.8) with a 680% increase. The scale of increase in system positioning error also shows the risk of using low-cost inertial sensor alone for safety-of-line navigation application. Whereas the DRM-aided system developed does not show significant performance degradation on the same direction with a nominal performance of 1.63 m (indicated in Figure 8.5a) and an averaged error of 1.062 m for all possible 30s outages (indicated in Table 8.6).

**Table 8.7.:** Statistics for the velocity RMS errors of the DRM-aided system for all disjoint 30s periods

| | Mean (m/s) | Standard Deviation (m/s) | Minimum (m/s) | Maximum (m/s) | Median (m/s) |
|---|---|---|---|---|---|
| Easting | 0.318 | 0.091 | 0.169 | 0.534 | 0.319 |
| Northing | 0.379 | 0.100 | 0.157 | 0.638 | 0.370 |
| Down | 0.268 | 0.062 | 0.064 | 0.426 | 0.266 |
| Along Track | 0.361 | 0.111 | 0.169 | 0.639 | 0.339 |
| Cross Track | 0.340 | 0.070 | 0.227 | 0.639 | 0.332 |

**Table 8.8.:** Statistics for the position RMS errors of the tightly integrated system for all disjoint 30s periods

| | Mean (m) | Standard Deviation (m) | Minimum (m) | Maximum (m) | Median (m) |
|---|---|---|---|---|---|
| Easting | 35.406 | 23.410 | 3.902 | 112.336 | 32.027 |
| Northing | 86.233 | 40.198 | 3.185 | 146.075 | 91.727 |
| Down | 141.655 | 89.688 | 5.943 | 312.598 | 140.238 |
| Along Track | 69.911 | 38.806 | 6.380 | 141.320 | 69.625 |
| Cross Track | 58.986 | 31.414 | 2.847 | 152.423 | 57.423 |

The cross track constraining provided by DRM-aiding can also be confirmed even under signal outage situations from the statistics shown in this section. All statistics for the cross track position error of the DRM-aided system as shown in Table 8.6 are consistently smaller than the along track error, which indicates a consistent better performance on the cross track direction. On the other hand, statistics for the tightly coupled system as listed in Table 8.8 do not suggest a superior performance for either direction.

### 8.3.3. Scenario 2: Difficult GNSS Conditions

This scenario simulates situations on the railways where only a part of the sky is visible and the number of available satellites is less than four. This mostly happens when a train travels through a urban area, or stops at stations during the journey. The length of the test scenario is set as 30 s. The first test section used in scenario 1, which runs between 5′27″ and 5′57″, is chosen for this scenario. A detailed description on the test section is given in Section 7.3.1.

The GPS constellation status during the scenario is illustrated in Figure 7.11, where the visible satellites are marked in full colour and the blocked satellite is marked with dashed borders. A full description of the GPS constellation status over the 32 min operation is given in Section 6.12. Since the train is mainly moving along the east-west direction, two blockages are simulated along each side of the rail route blocking most of the constellation. As a result, only two satellites, G09 an G22, are visible during the test scenario. The G09 satellite is in ahead of the train with 120° azimuth angle and 80° elevation angle, whereas the G22 satellite is behind the train with 275° azimuth angle and 20° elevation angle.

Figure 8.10 shows the test results for the test scenario 2. The normal system performances of both systems for the test section are listed in Table 8.4 on page 173. Both position errors and velocity errors are shown in the direction of north, east and down plotted against the simulation time, with position errors in meters and velocity errors in meters/second. The results for all tested systems are drawn respectively in green for the GPS only system, in blue for the loosely coupled system and in red for the tightly coupled system. The test section is marked by dashed lines in all graphs.

As the two satellites available are along the east-west direction, the positioning performance of a tightly coupled system is improved on the east direction. However, the overall system performance is still unreliable because of the error drifting on the north direction. For example, a maximum position error of 91.52 m on the north direction is presented by the end of the signal outage. On the other hand, a reliable performance is still maintained by the DRM-aided system. The maximum and RMS position errors on the east direction are respectively 3.98 m and 1.44 m, as shown in Figure 8.10. A small performance improvement is shown as a result of having extra satellites information when comparing with the east direction performance during scenario 1.

### 8.3.4. Scenario 3: Bad Satellite Constellation Geometry

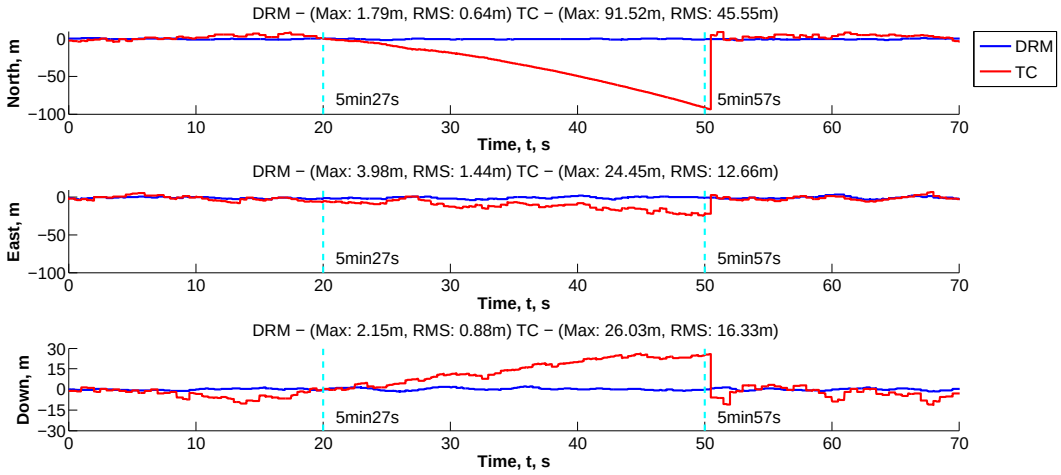
This scenario simulates situations on the railways where only half of the sky is visible but the number of available satellites is more than four. This mostly happens when a train travels through a urban area, or stops at stations during the journey. The length of the test scenario is set as 30 s. The first test section used in scenario 1, which runs between 5′27″ and 5′57″, is chosen for this scenario. A detailed description on the test section is given in Section 7.3.1. When all visible satellites are in one half of the sky, this implies a bad satellite constellation geometry.

The GPS constellation status during the scenario is illustrated in Figure 7.13a, where the visible satellites are marked in full colour and the blocked satellite is marked with dashed borders. One blockage is simulated along the north side of the rail route blocking satellites on the northern part of the sky. As a result, five satellites in the southern part of the sky are visible during the test scenario. The DOP values are plotted in Figure 7.13b. During the test period, all DOP values increases because of a decreased user-satellites geometry quality. The increase on the EDOP value is not as large as other DOP values indicating a still reasonably good satellite geometry in the east-west direction.

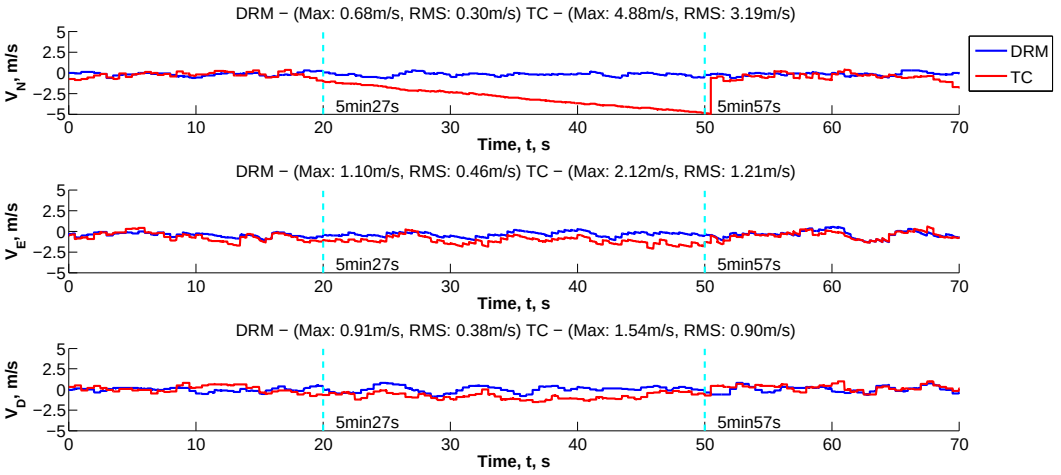
Figure 8.11 shows the test results for the test scenario 3. The normal system performances of both systems for the test section are listed in Table 8.4 on page 173. Both position errors and velocity

**Table 8.9.:** Statistics for the velocity RMS errors of the tightly integrated system for all disjoint 30s periods

| | Mean (m/s) | Standard Deviation (m/s) | Minimum (m/s) | Maximum (m/s) | Median (m/s) |
|---|---|---|---|---|---|
| Easting | 1.127 | 0.730 | 0.156 | 3.929 | 0.965 |
| Northing | 3.089 | 1.240 | 0.522 | 5.447 | 3.378 |
| Down | 3.247 | 1.929 | 0.109 | 8.075 | 3.008 |
| Along Track | 2.112 | 1.191 | 0.130 | 4.659 | 2.140 |
| Cross Track | 2.445 | 1.018 | 0.462 | 5.001 | 2.522 |

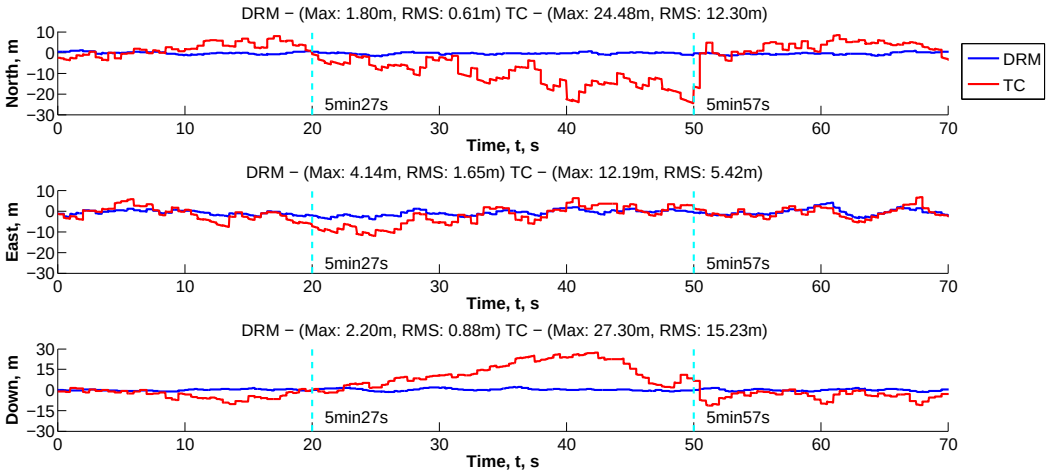



**(a)** Position errors

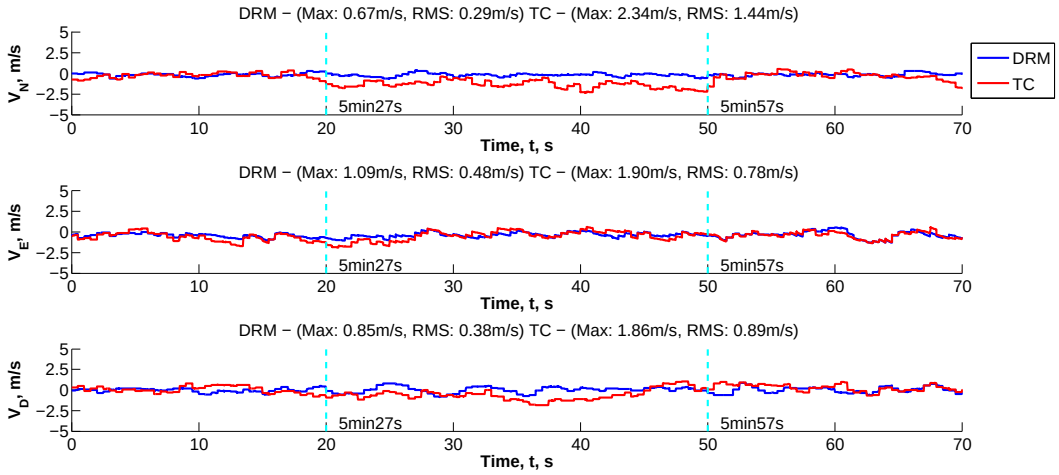



**(b)** Velocity errors

**Figure 8.10.:** Position and velocity error during scenario 2




**(a)** Position errors




**(b)** Velocity errors

**Figure 8.11.:** Position and velocity error during scenario 3

errors are shown in the direction of north, east and down plotted against the simulation time, with position errors in meters and velocity errors in meters/second. The results of the DRM-aided system are drawn in blue, and red is used for the results of the tightly coupled system. The test section is marked by dashed lines in all graphs.

The positioning performance of a tightly coupled system is affected by the decrease of GPS positioning performance. Increased positioning errors can be observed on all directions. On the other hand, maximum position errors for a DRM-aided system are shown in Figure 8.11 as 1.80 m in the north, 4.14 m in the east and 2.20 m in the down, indicating a noisier results when compared with the normal operating performance. It can be noticed that the only direction showing an increase on the RMS position error for an DRM-aided system is the east direction. Considering the main degradation of constellation geometry is in the north-south direction, the DRM aiding information provides the system a GPS independent error constraining mechanism.

### 8.3.5. Scenario 4: Initialisation with DRM-Aiding

This scenario simulates situations on the railways where the train stops at a train station under a roof and its positioning system needs to be initialised. Under a normal operation condition, the initialisation process for a inertial unit can be aided by the GPS solution. When the GPS solution is not available, the initialisation process for a conventional integrated system has difficulties in completing.

This test scenario is set up at the first 3 min of the overall operation, during which the train gradually starts to move away from the station. Only 2 satellites, G09 and G22, are available during the first 1 min of the test time. The GPS constellation status during the scenario is illustrated in Figure 7.11, where the visible satellites are marked in full colour and the blocked satellite is marked with dashed borders. Both a tightly coupled system and a DRM-aided system are tested in this scenario. Both systems have a 10 m position error and a 5 m/s velocity error included in their initial status.

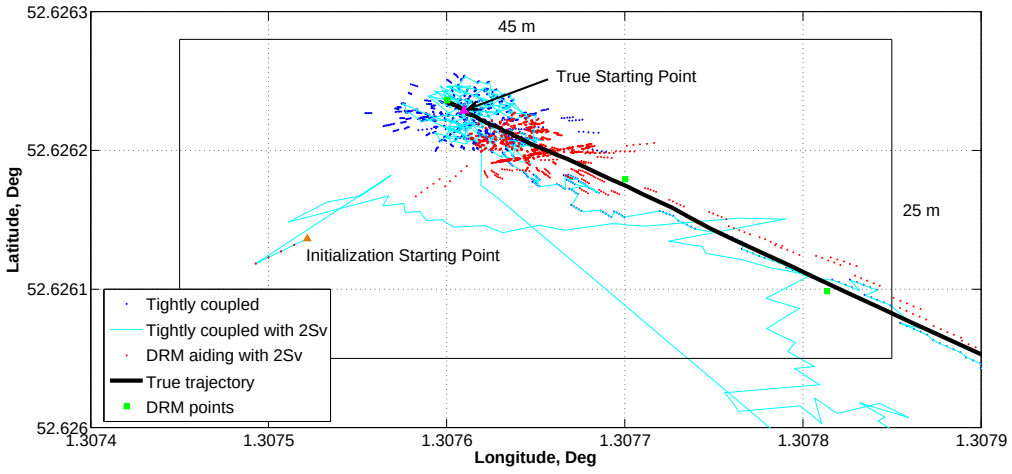
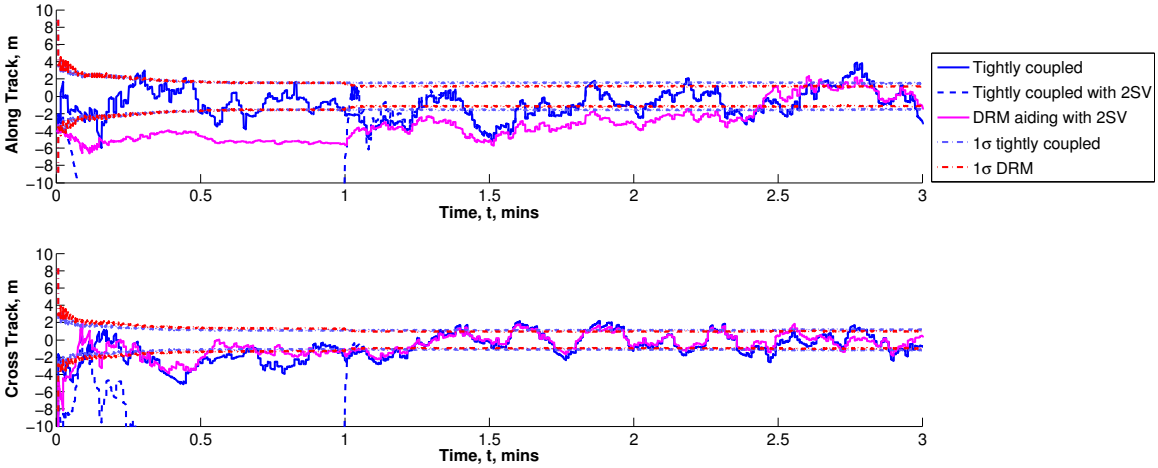



**Figure 8.12.:** Initialisation with 2 satellites in view

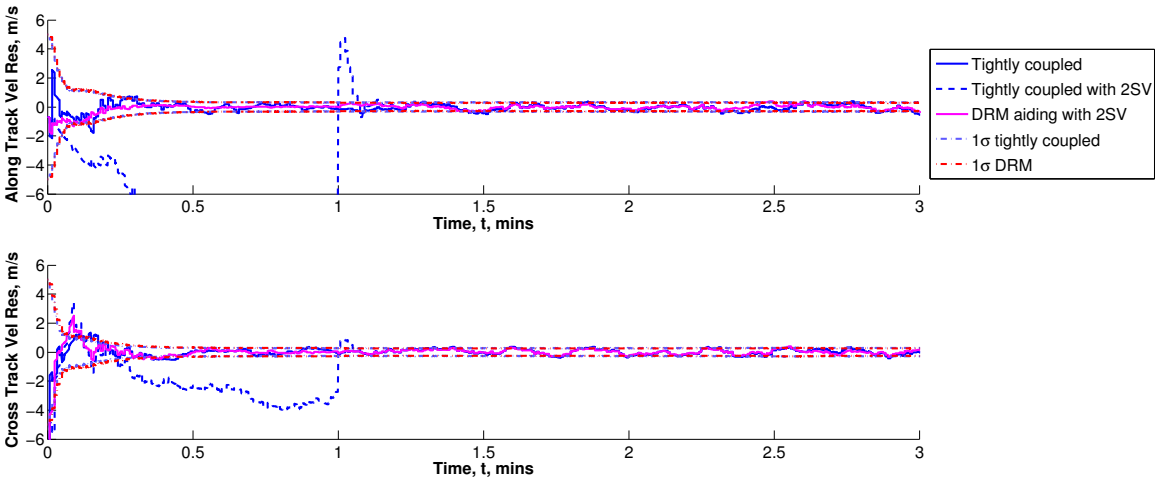
Figure 8.12 shows a position plot of the initialisation process tested in this section. The truth trajectory is marked by a black line, whereas the DRM is marked by green square dots along the truth trajectory. The initial position given to both systems are marked by a brown triangle and the actual initial position is marked by a purple triangle. The cyan line shows the position solution of the tightly coupled system during the initialisation process. The red dots show the position solution

of the DRM-aided system. Note that the position solution of a tightly coupled system with full access to the constellation in view is also given out in blue dots as a reference.

As can be seen from Figure 8.12, the tightly coupled system failed to initialise the system during the first minute with access to only two satellites. The position solution of a tightly coupled system with full access to the constellation in view scattered around the truth solution. On the other hand, the solution of a DRM-aided system gradually converges along the middle point of the DRM section indicating a stronger constraint on the position solution.




**(a)** Position errors




**(b)** Velocity errors

**Figure 8.13.:** Position and velocity error during scenario 4

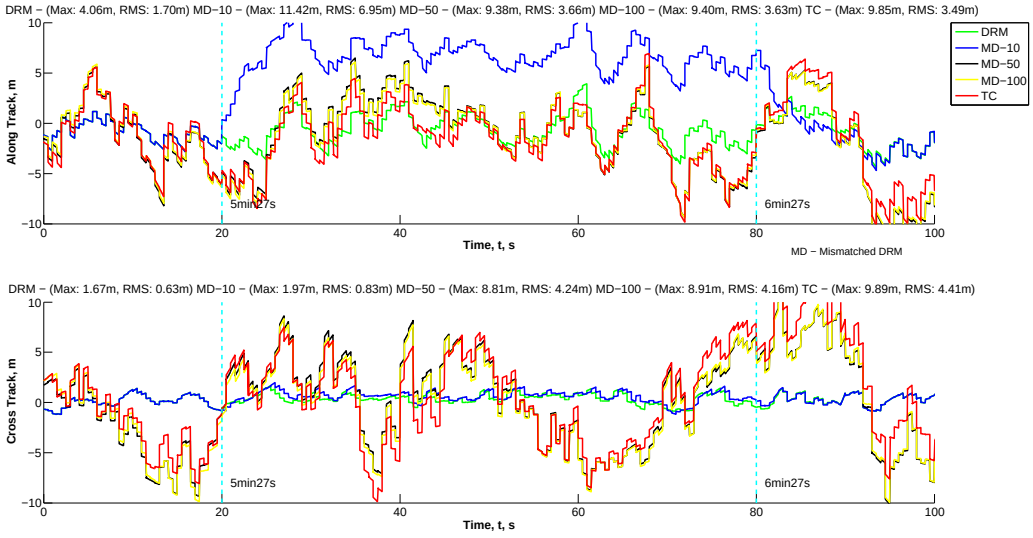
The position and velocity error of all systems during the test time is plotted in Figure 8.13 in along track and cross track directions. Positioning errors for tightly coupled systems with and without access to full constellation are respectively plotted in blue solid and dashed lines, whereas errors for the DRM-aided system are plotted in purple lines. The $1\sigma$ boundaries from both the tightly coupled system with access to full constellation and the DRM-aided system are also plotted in the graphs.

As can be seen from Figure 8.13, the velocity initialisation for a DRM-aided system is quite successful and quick. This is due to the fact that the same position measurement throughout the first minute implies its stationary status, and thus produces a satisfactory velocity initialisation result. The position initialisation process for a DRM-aided system, on the other hand, is only as quick and successful as the tightly coupled system with access to full constellation in the cross track direction. The convergence process of the along track position error of a DRM-aided system is slower

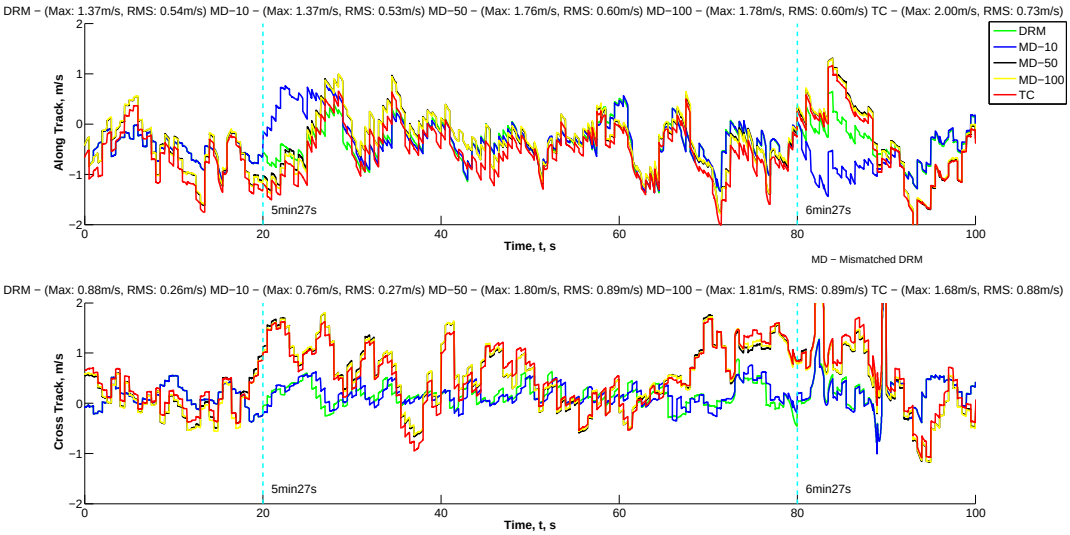
than the normal tightly coupled system. However, for a situation without the GPS solution, the DRM-aided system shows its advantage over the tightly coupled system by performing a slightly slower but still successful initialisation.

### 8.3.6. Scenario 5: Searching Algorithm Failure

This scenario simulates an integrity risk situation where the searching algorithm fails to find the correct DRM section. When a searching algorithm reports a wrong DRM section to the integrated system, the integration EKF processes the DRM aiding information with an incorrect error constraining model and a hazardous positioning solution may be produced by the system. Hence it is important to study the system behaviour under false DRM aiding information.




**(a)** Position errors




**(b)** Velocity errors

**Figure 8.14.:** Position and velocity error during scenario 4

A 60 s period is simulated in this scenario assuming that all reported DRM information is incorrect,

i.e. the reported DRM section is one section ahead of the section where the train actually is. The test section is set between $5'27''$ and $6'27''$. A detailed description on the first half of the test section is given in Section 7.3.1, and the second half follows the same pattern as the first half. The point accuracy for all tested DRMs in this section is 1m, and the tested DRM includes point intervals of 10m, 50m and 100m.

Figure 8.14 shows the test results for this scenario. Both position and velocity errors are shown in the direction of along track and cross track plotted against the simulation time, with position errors in meters and velocity errors in meters/second. Positioning results for the tightly coupled system and the DRM-aided system using a correct DRM with 10m point interval and 1m point accuracy are also presented for comparison. The results are drawn in green for the DRM-aided system using incorrect information and in red for the tightly coupled system. Positioning results of the DRM-aided system using wrong-section DRM information with different point intervals are presented in different colour as shown by the legends.

The test results show that when using a DRM with short point interval, the incorrect DRM-aiding information mainly causes system performance degradation on the along track direction of the position error. For instance, when using a 10m point interval DRM, the RMS position error on the along track direction is 6.95 m because of using the incorrect DRM information, and a maximum error of 11.42 m is reported during the test section. No significant performance degradation is seen on velocity estimation error and the cross track position error. This can be explained by the fact that straight lines and long flat curves are preferred in rail track design conventions. As a result, an incorrect DRM aiding information is mainly misleading in the along track direction.

On the other hand, the use of incorrect section information from DRMs with longer point intervals (such as 50m and 100m) significantly weakens the DRM constraining effect on integrated solutions, and the overall system performance is reduce to similar level of a tightly integrated system. As can be seen from the statistics, the RMS position error of using a mismatched DRM with 50m and 100m point intervals are respectively 3.66 m and 3.63 m on the along track direction. This indicates a similar performance level with a tightly coupled system of which the RMS error for the same time period is 3.49m. As a result of the way the error covariance is modelled for a mid-point approach DRM-aiding, the DRM constraining effect is intrinsically weaken as the point interval increases (as shown in Section 8.2.1). A DRM point measurement provided to the integrated system using long point intervals usually comes with a large standard deviation and covariance between easting and northing direction. Because the satellite ranging measurements are still available during the simulation period, a wrong section information with long point intervals presents little effect during the error-covariance-based weighting process of an integrating algorithm. This could be a dangerous integrity risk in situations where satellite measurements are not available and only wrong DRM information is provided to the system. Nevertheless considering the size of error presented with a tightly coupled system with a 30s signal outage (as shown in Table 8.8, as large as 69.91 m in the along track direction), a wrong section information can still be useful to constrain the drift of low-cost inertial sensors provided the distance of mismatched sections are kept at a small level.

## 8.4. Conclusions

Studies have been reported in this chapter that investigate the DRM-aiding of GNSS/inertial navigation systems. A DRM-aided integrated navigation system is developed based on the tightly coupled MEMS/GPS system as described in Section 7.2. The general performance of a DRM-aided system is studied by varying the point intervals and point accuracy levels of DRMs used in the integration. Five test scenarios are then proposed to study the performance of DRM-aided system under

typical railway GNSS hostile situations. The first three proposed scenarios are also used in Section 7.3 to investigate the performance of various conventional systems. Hence the performance of the DRM-aided system is studied by comparing with possible performances of conventional systems. The last two scenarios respectively studied the initialisation process of an integrated system with DRM-aiding under GNSS difficult conditions and a potential integrity risk caused by the failure of DRM searching algorithms.

The conventional DRM information is reformed in this chapter to integrate with other sensors in the system. Instead of referencing each DRM point in the database, a DRM middle point is provided to the integrated system at each section as measurements. A constraining error model is used along with the reformed DRM measurement and is incorporated in the measurement error covariance during the Kalman filtering process.

It has been discovered that the main advantages of a DRM-aided system over conventional integrated navigation systems are shown on two aspects including improved positioning accuracy and well constrained cross track errors. By using DRMs with proper specification, a reliable sub-meter level positioning performance can be achieved by a DRM-aided system. The accuracy improvement is especially prominent for estimations in the height direction. In addition, the positioning results of a DRM-aided system show well constrained errors in the cross track direction, whereas solutions of conventional systems show no dependence on directions. A positioning solution with minimised cross track errors suits the nature of a guided transport system as the railways, and is crucial for signalling under busy and complicated traffic conditions.

An increased point interval in the DRM decreases the general performance of a DRM-aided system and weakens the cross track constraining effect of the DRM-aiding. However, an improved positioning result, comparing with conventional systems, can still be achieved with reasonably long point intervals, as shown by investigations in this chapter. Among the proposed DRM specifications, a 50 m point interval DRM provides a good balance between performance and construction work amount. Moreover, a DRM with variable point intervals based on locations of the tracks is also advisable.

On the other hand, a sub-meter level accuracy for DRM points does not show significant performance improvement on the DRM-aided system, but a decreased point accuracy weakens the DRM constraining effect significantly even under short point interval conditions. As a result, a 1 m accuracy level is chosen among proposed specifications as an ideal standard for DRM points.

As discussed in Section 2.3.1.2, the major problem presented by the railway environment is various GNSS difficult situations characterised by physical obstructions blocking the path of satellite signals. A DRM-aided system deals with these situations by providing a satellite-independent constraint on the MEMS sensors used in the system. A consistent and reliable positioning performance can be achieved with DRM-aiding throughout the journey including GNSS difficult situations.

The initialisation process is important for a system to deliver a consistent high performance solution. Conventional MEMS/GPS systems fails to initialise under GNSS difficult situations, whereas a DRM-aided system shows the ability to initialise the MEMS unit and providing a reliable positioning solution with a small sacrifice on the along track positioning accuracy and converging speed.

One potential failing mode for a DRM-aided system is the failure of its searching algorithm to identify the correct DRM section where the train is. Investigations in this chapter show that a mismatched DRM measurement from a short point interval (such as 10m) caused performance degradation mainly on the position error of along track direction. Whereas mismatched DRM measurements from longer point intervals (such as 50m and 100m), a system performance similar to the tightly coupled system as simulated in this research is reported, and the cross track constraint provided by DRM is lost.

The performance shown by a DRM-aided MEMS/GPS system as described in this chapter under various situations indicates the benefits of having extra DRM information in an integrated system using low-performance inertial sensors. With properly configured aiding information, a reliable positioning performance can be achieved by low-cost low-performance inertial sensors without having to compromise the system accuracy.

# 9. Summary and Conclusions

## 9.1. Summary

### 9.1.1. Railway Signalling and Integrated Navigation

The development of signalling and control on the railways is moving towards more efficient and reliable systems that do not rely on traditional track side train detection sensors. By shifting the function of determining the train position from track side infrastructure to onboard positioning equipment, a moving box signalling system allows a more flexible way of calculating train separations in the network. In order to determine the train separations based on real time information about position and velocity of trains, it is necessary to provide the railways with a reliable high-performing positioning system that is capable of maintaining the specified working standard under all railway situations. In comparison with track side sensors, satellite positioning using GNSS features low maintenance costs and high positioning accuracy. GNSS has already been used in many non-safety-of-life applications on the railways, such as providing train arrival and departure information. However, for a safety-of-life application in the signalling and train controlling process, GNSS only systems suffer from reliability issues due to the different railway operating environment. This thesis investigates the potential performance of MEMS/GPS integrated systems with conventional integration architectures and proposes an integrated GNSS/inertial navigation system using low-cost MEMS sensors and aiding information from a digital route model.

The current railway signalling system is known as the fixed block system, in which the rail route is separated into rigid small sections and the train separation is defined in the unit of small sections. Traditional train detection sensors, such as track circuits and treadles, are widely used across the railways. These train detection sensors are installed along the rail tracks to detect the presence of a train in a section or whether the train has passed a particular point. Train locations are indicated by an occupied track section instead of actual coordinates, and velocities are usually not available. The fixed signalling system limits the transport potential of a railway network because of the inefficiency in determining the vehicle status in the network. In addition, the maintenance cost for train detection systems are increasing along with the rapid growing of the railway network, but the general reliability level of train detection systems are decreasing as a result of the complex operating environment of modern railways.

Unlike the fixed block system, a moving block system is able to maintain a dynamic train separation based on the current position and velocity of trains. Instead of indicating an occupied block using track side sensors, the train is able to use onboard positioning system to determine its own position and velocity, and reports to the central traffic control. GNSS has been proved capable of been used for safety-of-life applications in other transport systems, and presents satisfactory accuracy performance for most applications such as aviation. However, the reliability for a GNSS only positioning system has been proved insufficient for safety-of-life applications on the railways (see discussions in Section 2.3.1.2). Traditional integrated systems are commonly used for situations where GNSS is not always available. Sensors with an independent ability of performing navigation tasks are usually used in an integrated system, such as high-performance inertial sensors. For a

large scale deployment of positioning systems across the railways network, the cost factors must also be considered. Therefore the potential usage of low-cost low-performance inertial sensors aided by DRM for a safety-of-life application on the railways is explored in this thesis.

The main advantage of an integrated navigation system is an improved performance in both positioning accuracy and system reliability, with the help of multiple sensors and data sources. Critical aspects for the design of an integrated system include sensor selection, system integration architecture and the adaption of the integration algorithm. Systems based on position fixing and systems based on dead reckoning are usually chosen to integrate due to their complementary characteristics. Low-cost low-performance inertial sensors are not commonly used independently for navigation purposes, but their applications in an integrated system are still beneficial. Since the railway is a guided transport system, a known trajectory, e.g. a digital route model, can be provided to the positioning system for aiding purposes. Different system integration architecture affects the performance of integrated systems.

### 9.1.2. Integrated Navigation on the Railways with MEMS and DRM-aiding

A GPS navigation processor and an inertial navigation processor were realised in the software. A train journey was simulated based on GPS data collected along a suburban and countryside rail route. The simulated journey database includes the status of the train carriage throughout the journey, information on the tracks along the route, and sensor outputs. Two MEMS/GPS integrated systems using different integration architectures, respectively based on the principle of loosely coupled and tightly coupled INS/GPS integration, were developed. Finally a DRM-aided MEMS/GPS system was proposed and realised based on the tightly coupled INS/GPS architecture. All of the developed integrated systems use sensor measurements from the simulated database to study their general performances. Five testing scenarios were set up covering typical situations on the railways where GNSS only positioning performance can be affected. All developed systems were tested under proposed testing scenarios to study their performance characteristics.

The simulated train journey lasted for 32 min and the length of the journey was approximately 33 km. The train movement was shown by a truth file of its position, velocity and acceleration built from collected GPS data. Attitude changes of the carriage along the journey were simulated with a connection to the actual physical and dynamic status of the train, and were expressed in the form of Euler angles. The position of the tracks and superelevation variations along the route were simulated based on the real railways track design convention in order to reflect their effects on the train attitude changes. DRMs were built based on the obtained position truth file using different specifications including point intervals and point accuracy levels. The simulated DRMs were used by the developed DRM-aided MEMS/GPS system to study the relationship between DRM specifications and the DRM-aided system performance. The simulated sensor outputs are simulated measurements from a MEMS inertial unit, satellite ranging measurements, and GPS position and velocity solutions.

The first three proposed testing scenarios were set up to study possible system positioning performance under various railway operating situations including GNSS outage period, GNSS difficult situations, and poor constellation geometry. All developed systems were tested for the first three testing scenarios, and system performances were compared with each other. Scenario 4 was set up to study situations on the railways where the train stops at train station under a roof and its positioning system needs to be initialised. Both the tightly coupled system and the DRM-aided system were tested for scenario 4. Scenario 5 simulated an integrity risk situation where the searching algorithm fails to find the correct DRM section. The potential behaviour of the DRM-aided system under a

fault condition was studied during the test.

Investigations into different MEMS/GPS integration approaches for positioning on the railways have been done through both general performance study and testing under different railway operation scenarios. Positioning solutions from the filter based GPS navigation processor were also stored along with the tested integrated systems, the performance of which is analysed to compare with other systems. Three testing scenarios covering common GNSS difficult situations on the railways were then proposed, and all systems were tested for all scenarios.

The conventional DRM information was reformed before integrating with other sensors. Instead of referencing each DRM points in the database, a DRM middle point was provided to the integrated system at each section as measurements. A constraining error model was used along with the reformed DRM measurement and was incorporated in the measurement error covariances during the Kalman filtering process.

The general performance of a DRM-aided system was studied by varying the point intervals and point accuracy levels of DRMs used in the integration. Five testing scenarios were then proposed to study the performance of DRM-aided system under typical railway GNSS hostile situations. For the first three testing scenarios, the performance of the DRM-aided system was studied by comparing with possible performances of the GPS navigation processor and conventional integrated systems. The last two scenarios respectively studied the initialisation process of an integrated system with DRM-aiding under GNSS difficult conditions and a potential integrity risk caused by the failure of DRM searching algorithms.

## 9.2. Conclusions

In accordance with the research objectives set up at the beginning of this thesis (Section 1.3), the major conclusions of this research are summarised as following:

- *What performance levels can low-cost MEMS/GNSS positioning systems achieve with various integration architectures in different GNSS signal situations in a railway environment?*

  For the situations where the number of available satellites is less than four, none of the tested MEMS/GNSS positioning systems showed satisfactory accuracy performances during the 30 s testing period comparing to the performance of the DRM-aided system under the same circumstances. The largest maximum position errors shown during test scenario 1 is as large as 98.43 m. On the other hand, when there is a bad constellation geometry, the inclusion of MEMS sensor shows performance advantages in the velocity-domain. For example, the RMS velocity error of a tightly coupled system in the east direction is 1.21 m/s, in comparison the 2.50 m/s error produced by a filter based GPS only system. Because GPS serves as the main contributor to the system positioning accuracy, the position-domain performance of an integrated system using low-cost low-performance sensors is largely decided by the availability of GPS services.

  The velocity-domain performance of MEMS/GNSS integrated systems presents improved performance under normal operation conditions. For example, the RMS velocity error of a tightly coupled system in the east direction is 0.73 m/s, compared with the 2.48 m/s error produced by a filter based GPS only system. Although low-cost low-performance inertial sensors are not commonly used independently for navigation purposes, integrating with GNSS provides a good real time calibration for the sensor error characteristics. The inclusion of GNSS-calibrated high

update rate inertial measurements provides a better resolution in train velocity estimation than GPS over the short term.

For GPS only systems, although a filter based navigation processor enables a system to provide outputs under GNSS difficult situations, neither accuracy nor reliability of the solution can be guaranteed. Systems with least-squares based navigation processor do not work under GNSS difficult situations.

Under normal operating conditions, the benefits of using MEMS sensors in integrated systems includes a reduced level of variations on the position error indicating a smoother estimated trajectory, and a more accurate velocity solution. For example, the 95% and RMS position error achieved by a tightly coupled system in the east direction are respectively 7.90 m and 4.02 m, compared with the 10.91 m and 5.53 m position error of a filter based GPS only system. However, when the GPS input is not available, the low performance of MEMS sensors cannot ensure a reliable system accuracy.

In terms of accuracy, assuming a loosely coupled system is tuned properly, the performance improvement of using a tightly coupled system is limited. Nevertheless, the robustness of the tightly integration approach shows advantages under situations where satellites are available but the number is less than four, as a result of using the satellite ranging measurement directly.

- *What improvement can the introduction of a DRM bring to an integrated low-cost MEMS/GNSS positioning system in an open railway environment?*

  The benefits of introducing a DRM to an integrated low-cost MEMS/GNSS positioning system are shown on two aspects including a largely improved system accuracy and well constrained cross track errors.

  A consistent metre level positioning accuracy can be achieved by a DRM-aided system using DRMs with proper specification. For example, the RMS position errors in the direction of north, east and height are respectively reduced from 5.49 m, 4.02 m and 7.17 m to 1.32 m, 1.63 m and 0.76 m using a DRM with 10 m point intervals and 1 m point accuracy. The accuracy improvement is especially prominent for estimations in the height direction.

  The positioning results of a DRM-aided system shows minimised errors in the cross track direction in both position- and velocity-domains because of the constraining error model used with the DRM measurements, whereas solutions of conventional systems show no dependent on track directions. For example, the along track and cross track position errors of a DRM-aided system using a 10 m point intervals and 1 m point accuracy DRM are respectively 1.91 m and 0.89 m, in comparison with errors of 5.48 m and 4.94 m from a tightly coupled system. A positioning solution with minimised cross track errors suits the nature of signalling control on the railways as a guided transport system, especially for busy and complicated traffic conditions.

- *How can a tightly integrated low-cost MEMS/GNSS positioning system with aiding from a DRM perform under difficult GNSS conditions in a railway environment? What happens to this performance if the GNSS signal is completely lost?*

Both difficult GNSS conditions and GNSS signal outages in a railways environment show minimal interferences on the positioning accuracy level of a tightly integrated low-cost MEMS/GNSS positioning system with aiding from a DRM. For example, when there is a GNSS outage (as in test scenario 1), position errors of a DRM-aided system in the direction of north, east and height are respectively 0.64 m, 1.49 m and 0.88 m, in comparison with its normal position errors of 0.61 m, 1.49 m and 0.88 m.

The major problem presented by the railway environment is various GNSS service interruptions characterised by physical obstructions blocking the path of satellite signals. A DRM provides a GNSS-independent error constraining mechanism on the MEMS inertial sensors in the system. DRM measurements performance level can be maintained through a journey without regarding to the GNSS signal status. Therefore even for situations of a complete GNSS outage, the low-cost MEMS measurements constrained by a DRM can still maintain the system accuracy level for a short period.

- *How can a DRM aid the initialisation process of a tightly integrated low-cost MEMS/GNSS positioning system when no, or only limited, GNSS signals are available?*

  The initialisation process of a tightly integrated low-cost MEMS/GNSS positioning system aided by a DRM can be achieved with limited position-domain performance decrease on the along track direction when GNSS signals are insufficient. During the 1 min test period of scenario 4, final RMS position errors achieved by a DRM-aided system after system initialisation are respectively 5.28 m in the along track direction and 1.28 m in the cross track direction, in comparison with position errors of 1.91 m in the along track direction and 0.89 m in the cross track direction under normal operation. The conventional tightly coupled integrated system fails to initialise under the same conditions.

- *How should the information in a DRM be used so that it can be most beneficial to an integrated low-cost MEMS/GNSS positioning system?*

  Instead of referencing each DRM points in the database, a new method of integrating DRM information is proposed in this thesis. The middle point of each DRM section is provided to the integrated MEMS/GNSS system at each section as measurements. A constraining error model is used along with the reformed DRM measurements and is incorporated in the measurement error covariances during the Kalman filtering process.

- *How does the accuracy of individual DRM points affect the performance of a DRM in aiding a low-cost MEMS/GNSS integrated positioning system?*

  A sub-meter level accuracy for DRM points does not show significant performance improvement on the DRM-aided system with the proposed straight line approximation. For fixed 10 m point intervals DRMs, RMS position errors achieved by the system using a DRM with 0.1 m point accuracy in the direction of north, east and height are respectively 1.42 m, 1.61 m and 0.25 m. Only improvement in the height direction can be seen using a higher accuracy DRM. However, potential benefits of using a higher accuracy DRM may be achieved using a polynomial line assumption for each DRM section.

  A decreased point accuracy level (10 m) weakens the DRM constraining effect significantly

even under short point interval conditions. For fixed 10 m point intervals DRMs, RMS position errors achieved by the system using a DRM with 10 m point accuracy in the direction of north, east and height are respectively 3.57 m, 3.17 m and 5.31 m.

As a result, the 1 m accuracy level is recommended among proposed specifications as an ideal standard for DRM points.

- *What is the optimal distance between adjacent DRM points in a DRM-aided integrated low-cost MEMS/GNSS positioning system?*

  An increased point interval in the DRM decreases the general performance of a DRM-aided system and weakens the cross track constraining effect of the DRM-aiding. However, an improved positioning result, comparing with tightly integrated low-cost MEMS/GNSS positioning system, can still be achieved with reasonably long point intervals. For example, RMS position errors of the system using a DRM with 100 m point interval and 1 m point accuracy in the direction of along track and cross track are respectively 4.56 m and 3.46 m. The accuracy improvement can still been seen in comparison with a tightly coupled system.

  Among the proposed DRM specifications, a 50 m point interval DRM provides a good balance between performance and surveying work load.

  Moreover, a DRM with variable point intervals based on locations of the tracks may also be beneficial during the DRM surveying. Stations and junctions are the busiest areas in the railway network and hence require the highest positioning accuracy requirement. Because the train moves slower when approaching stations and junctions, and hence a small distance DRM can be used to improve the positioning accuracy without increasing the load of the searching algorithm and risks of integrity failures. On the other hand, a larger distance DRM could suffice the positioning requirement for countryside areas.

- *How would the use of a mismatched DRM (e.g. for the wrong section) affect the results of a DRM-aided integrated low-cost MEMS/GNSS positioning system?*

  A mismatched DRM measurement, e.g. for the wrong section, from short DRM intervals such as 10m presents an integrity risk of using the DRM-aided system by showing no performance degradation alert from the predicted system estimation information under such a system fault condition. While the usage of a mismatched DRM measurement from short DRM intervals causes a performance degradation mainly on the position error of along track direction, no indications can be reflected on the prediction of $1\sigma$ system estimation error its own positioning error. The actual amount of performance decrease on the along track direction depends on the DRM specification. For example, during test scenario 5, the along track position error of a system using a DRM with 10 m point interval and 1 m point accuracy increases from 1.70 m under normal operation to 6.95 m without alert from the $1\sigma$ prediction of system positioning error.

  On the other hand, the DRM-aided system using mismatched DRM measurements from longer DRM intervals such as 50m and 100m shows a similar performance level compared with the simulated tightly coupled system. The DRM-aiding effect is lost when using such DRM measurements.

Through the conclusions listed in this section, all questions as asked by the initial research objectives have been answered. The use of an integrated positioning system using low-cost MEMS sensors at the measurement level with aiding of a DRM to maintain high accuracy positioning performance on the railways under GNSS difficult situations is therefore proved as feasible.

## 9.3. Future work

In order to explore the full potential of high-performance positioning on the railways using DRM-aiding, some more investigations are suggested in this section which could be beneficial for further positioning performance improvement of a guided transport system such as the railways.

#### Real-time System Testing

Simulated sensor measurements are used for the convenience of sensor error characteristics control and scenario testing in this thesis. Due to the time limitation of doctoral research, real-time system testing is not performed in this thesis. As only limited DRM information is available on the railways at the moment, an operational DRM is a prerequisite for real-time system testing. The building of DRM involves large amount of surveying and data collection along rail routes. However, the testing of the developed systems can provide positioning results which are useful for required navigation performance studies on the railways.

#### Using GPS Augmentations on the Railways

EGNOS is the European satellite based augmentation system. EGNOS is still only certificated by the European Commission for Non-Safety-of-Life use, but it is expected to be certified for Safety-of-Life applications in 2010. According to RSSB (2008), 1 to 2 m GPS positioning accuracy at 95% confidence can be expected with currently available augmentation services such as EGNOS. However it is difficult to receive information directly from geostationary satellites on the railways because of their low elevation angles. ESA released an internet service for continuous delivery of EGNOS signals, called SISNet, to address this visibility issue. Nevertheless, communication arrangements within the railways signalling system and the compatibility with future ERTMS is yet to be standardised. The ability of reliably using augmentations increases the general accuracy of GNSS on the railways.

#### Multipath Detection and Mitigation for the Railways

The usage of GNSS receivers in the railways environment currently relies mostly on the code-phase based positioning techniques. Unlike the carrier multipath error that is usually on centimetre levels, the code multipath error varies greatly for different environment and can exceed hundreds of meters in extreme cases. For an environment such as the railways, large numbers of physical obstructions and reflector sources along the rail tracks in the urban area increases the magnitude of short-delay multipath errors on the positioning solution. Further difficulties for the receiver can also be introduced when the train comes out of tunnels and signal reacquiring is still in the progress.

The knowledge of DRM can help to detect and mitigate the multipath error, especially when only single-frequency receiver is available. Techniques such as using DRM to aid satellite residuals monitoring, EKF innovation filtering etc. would be interesting areas for further research.

### GNSS Multi-frequency and Multi-constellation Positioning on the Railways

The GPS L2C signal was introduced since the first launch of Block IIR-M on September 2005. The new GPS signal frequency L5 is available along with the launch of Block IIF satellites this year. The new L5 frequency is located in ARNS band and is intended for civil safety-of-life application. Further introduction of civil signal L1C will be realised by the launch of GPS Block III satellites. New civil signals in ARNS band will allow safety-of-life applications, such as positioning on the railways, to take the advantage of positioning with multiple frequencies. An integrated positioning system combines GNSS multi-frequency positioning with developed DRM-aided system could further increase the performance of current systems.

Currently, the American GPS and the Russian GLONASS are the only two GNSSs in full operational capacity. However, more GNSSs are under development, such as the European Galileo system and the Chinese Compass, and are expected to research full operational capacity in the near future. Therefore GNSS multi-constellation positioning can be expected on the railways. The presence of multiple constellations will potentially increase the overall availability along a rail line. But the GNSS interoperability issue as discussed in Section 4.1.4 will need to be tackled first.

### Integration with Reduced-level IMU

Due to the nature of the railways as a guided transport system, instead of knowing the full 3D coordinates of a train, traffic control on the railways is more interested in knowing how far the train has travelled along the rail track. It does not require dead reckoning measurements in all three axes to achieve the travel distance along the track. Therefore the full set of an IMU might not be needed for an integrated positioning system. An integrated positioning using a reduced-level IMU can also potentially reduce the complexity of a DRM searching algorithm, decreasing the chance of an integrity risk because of searching algorithm failure.

### Integrated Positioning with Multiple Receivers/Antennas

A train usually consists of several carriages and is mostly moving along straight line or a long flat circular curve. In addition, the dimensions for train carriages follow specific design conventions. Current GNSS availability studies on the railways indicate that most GNSS outage situations on the railways are for a short period. It is a common situation for one part of the train is blocked of satellite signals whereas other parts can still receive normal GNSS services. An array of GNSS receivers/antennas could therefore be integrated into a DRM-aided MEMS/GNSS system by level arm modelling. Using multiple receivers/antennas not only increases the system ability to overcome GNSS outage period but also can be used for the train integrity monitoring.

# Bibliography


Albanese, A. & Marradi, L. (2005). The RUNE project: the integrity performances of GNSS-based railway user navigation equipment. *Proc. the 2005 ASME/IEEE Joint Rail Conference*, 211–218. 2.3.3.2

Allan, D., Ashby, N. & Hodge, C. (1997). The Science of Timekeeping. *Hewlett Packard Application Note*, **1289**. 4.2.5

Alston, B. (2000). Guidance Note: Train Detection. Tech. rep., Railtrack PLC (UK), Railway Group Guidance Note (GK/GN0611). 2.1.2

Ashton, P. (2006). Exit from New Street. `http://www.flickr.com/photos/peteashton/108898654/`, Image cited from Flickr under Creative Commons, Attribution-Non-Commercial License. 2.11b

Atlantic Inertial Systems (2008). SiIMU02. *www.atlanticinertial.com*, Product Datasheet. 5.2.3, 6.9

Axelrad, P. & Brown, R.G. (1996). *Global Positioning System: Theory and Applications*, vol. I, chap. GPS Navigation Algorithms, 409–434. AIAA. 4.2.4.2

Bailey, C. (1995). *European railway signalling*. A & C Black. 1.2.1, 2.1.4

Barbour, N. (2004). Inertial Navigation Sensors. *NATO RTO Lecture Series*, **232**. 5, 5.2.1.1

Barbu, G., Miguel, S. & García, M. (2004). GADEROS: Galileo Demonstrator for Railway Operation System. Tech. rep., EC Competitive and Sustainable Growth Programme.

Bedrich, S. & Müncheberg, S. (2004). GNSS-Based Sensor Fusion for Telematic Applications in Railway Traffic. *Proc. ION GNSS 2004, Long Beach, CA*. 2.3.3.2

Bertran, E. & Delgado-Penin, J. (2004). On the use of GPS receivers in railway environments. *IEEE Trans. on Vehicular Technology*, **53**, 1452–1460. 2.3.2

Bloomfield, R. (2006). Fundamentals of European rail traffic management system (ERTMS). *Proc. the 11th IET Professional Development Course on Railway Signalling and Control Systems*, 165–184. 1.2.1, 2.2

Böehringer, F. (2003). Train location based on fusion satellite and train-borne sensor data. *Proc. SPIE*, **5084**, 76. 2.3.3.2

Braasch, M.S. (1996). *Global Positioning System: Theory and Applications*, vol. 1, chap. Multipath Effects, 547–568. AIAA. 4.3.5

Brown, R. & Hwang, P. (1997). *Introduction to Random Signals and Applied Kalman Filtering: With MATLAB Exercises and Solutions*. John Willey & Sons, Toronto, 3rd edn. 3.2.4.1, 3.2.4.2, 3.2.7.2

Brown Jr, K. (1991). The theory of the GPS composite clock. *ION GPS-91: Proceedings of the 4th International Technical Meeting of the Satellite Division of the Institute of Navigation, Albuquerque, NM*, **1**, 223–241. 3.2.1

Carlson, N.A. (2002). Federated Filter for Distributed Navigation and Tracking Applications. *Proc. ION 58th AM, Albuquerque, NM*, 340–353. 3.1.3

Carlson, N.A. & Berarducci, M.P. (1994). Federated Kalman filter simulation results. *Navigation: JION*, **Vol. 41, No. 3**, 297–321. 3.1.3

Chatfield, A. (1997). *Fundamentals of high accuracy inertial navigation*. AIAA Progress in Astronautics and Aeronautics. 1.2.4, 5, 5.4.4.3

Conley, R., Cosetino, R. *et al.* (2006). *Understanding GPS: Principles and Applications*, chap. Performance of Stand-Alone GPS, 301–378. Artech House, 2nd edn. (document), 1.2.3, 4.2.1.2, 4.3.1, 4.3.2, 4.3.3, 4.3.3, 4.3.4, 4.3.5, 4.5

Counselman, C. & Gourevitch, S. (1981). Miniature Interferometer terminals for Earth Surveying: Ambiguity and Multipath with Global Positioning System. *IEEE Transaction on Geodscience and Remote Sensing*, **GE-19**, 244–252. 4.3.5

Crassidis, J. (2006). Sigma-Point Kalman Filtering for Integrated GPS and Inertial Navigation. *IEEE Trans. on aerospace and electronic systems*, **42**, 750–756. 5.4.3.3, 6.9, 6.9

Crassidis, J. & Junkins, J. (2004). *Optimal estimation of dynamic systems*. Chapman & Hall. 1.2.2, 3.2, 4.2.4, 5.3.2

Crews, M. (2008). Long-Term Future of GPS. *Proc. ION NTM 2008, San Diego, CA*. 4.1.1.1

Cross, P., Kelsey, D. & Pierce, R. (2006). Safety cases for electromagnetic compatibility of T121 track circuits and eurostar trains. *Proc. the 1st IET International Conference on System Safety*, 121–129. 2.1.3.1

Cross, P.A. (1983). *Working Paper No. 6: Advanced least squares applied to position-fixing*. Departmetn of Land Surveying, North East London Polytechnic, ISSN 0260-9142. 8.1.3

Davies, S. (2009). The future is seamless - [control transport]. *Engineering Technology*, **4, Issue:10**, 42–44. 2.2.3, 2.2.3.2

Department for Transport (2007). ERTMS National Impelementation Plan. Tech. rep., Formally notified to the European Commission in September 2007, and its target dates are thus now binding on the UK. 1.2.1, 2.2

Dorsey, A.J., Marguis, W.A., Fyfe, P.M. *et al.* (2006). *Understanding GPS: Principles and Applications*, chap. GPS system segments, 67–112. Artech House, 2nd edn. 4.1.1.1, 4.1.1.1

El-Rabbany, A. & El-Diasty, M. (2004). An Efficient Neural Network Model for De-noising of MEMS-Based Inertial Data. *The Journal of Navigation*, **57**, 407–415. 5.3.2

El-Sheimy, N. & Niu, X. (2007). The Promise of MEMS to the Navigation Community. *Inside GNSS*, **March-April**, 46–56. 5.3.2, 5.3.3, 6.11

Ernest, P., Mázl, R. & Přeučil, L. (2004). Train locator using inertial sensors and odometer. *Proc. IEEE Intelligent Vehicles Symposium*, 860–865. 1.1, 2.3.3.2

European Space Agency (2010). Galileo OS SIS ICD (Open Service Signal-In-Space Interface Control). Tech. rep., ESA. 4.1.3

Falcone, M., Erhard, P. & Hein, G.W. (2006). *Understanding GPS: Principles and Applications*, chap. GALILEO, 559–594. Artech House, 2nd edn. 4.1.3

Farrell, J. (2008). *Aided navigation: GPS with high rate sensors*. McGraw-Hill Professional. 5.4.1, 6.7, 6.10.4, C

Feairheller, S. & Clark, R. (2006). *Understanding GPS: Principles and Applications*, chap. Other Satellite Navigation Systems, 595–634. Artech House, 2nd edn. 4.1.2

Fenner, D. (2007). Railway Signalling. *Proc. the 3rd IET Professional Development Course on Railway Electrification Infrastructure and Systems*, 153–160. 2.1.1, 2.2.3.2, 2.3.1.1

Filip, A., Beugin, J., Marais, J. & Mocek, H. (2008). A relation among GNSS quality measures and railway RAMS attributes. *Manuscript of paper for CERGAL 2008 symposium, Braunschweig, Germany*.

First Great Western (2010). Regional Fleet Class 150. *http://www.firstgreatwestern.co.uk/Content.aspx?id=57*, Train operator fleet specification, visited on 29 July 2010. 6.6.3

GCOR Committee (2005). General Code of Operating Rules (Fifth Edition). Tech. rep., General Code of Operating Rules Committee (US).

Gemaakt, Z. (2004). Eurobalise in Lutherstadt Wittenberg. `http://en.wikipedia.org/wiki/File:Siemens_Eurobalise.jpg`, Image cited from Wikipedia under GNU Free Documentation License. 2.7

Genghi, A., Marradi, L., Martinelli, L., Campa, L., Labbiento, G., Cianci, J. *et al.* (2003). The RUNE Project: Design and Demonstration of a GPS/EGNOS-Based Railway User Navigation Equipment. *Proc. ION GPS/GNSS 2003, Portland, OR*. 1.1, 2.3.3.2

Goddard, E. (2006). Overview of signalling and train control systems. *Proc. the 9th IET Professional Development Course on Electric Traction Systems*, 336–350. 2.1.4

Goverde, R. (2005). *Punctuality of railway operations and timetable stability analysis*. Ph.D. thesis, TU Delft, TRAIL Thesis Series. 2.1.3.1

Grewal, M. & Andrews, A. (2008). *Kalman filtering: theory and practice using MATLAB*. John Wiley & Sons, Inc., 3rd edn. 3.2

Grewal, M., Weill, L., Weill, L. & Andrews, A. (2007). *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons, Inc., 2nd edn. 3.1.1

Groves, P. (2008). *Principles of GNSS, inertial, and multisensor integrated navigation systems*. Artech House. 1.2.2, 3.1, 3.1.1, 3.1, 3.2, 3.3, 3.1.3, 3.3, 4.1.1.2, 4.1.3, 4.2.4.2, 4.2.4.2, 4.2.4.3, 4.2.5, 4.3.1, 5.4, 5.3.2, 5.3.3, 5.4.3.4, 5.4.4.1, 5.4.4.3, 6.10.4, A.1.1, C

Hanse, J. (2004). Honeywell MEMS inertial technology & product status. *Proc. IEEE PLANS*, 43–48.

Health and Safety Executive (2005). Railway Safety Principles and Guidance (Part 2, Section D, Guidance on signalling). Tech. rep., Office of Rail Regulation (UK). 1.2.1, 2.1.1, 2.1.2, 2.1.4

Hein, G.W. *et al.* (2007). Envisioning a Future GNSS System of Systems, Part 1. *Inside GNSS*, **January/February**, 58–67. 4.1.1.3

Hide, C., Moore, T. & Smith, M. (2003). Adaptive Kalman filtering for low-cost INS/GPS. *The Journal of Navigation*, **56**, 143–152. 3.2.7.2

Honeywell Aerospace (2007). Honeywell Tactical Inertial Measurement Units (IMU). *www.honeywell.com*, Product Datasheet. 5.2.3

ITRF (2007). ITRS and WGS84. ITRF Website `ftp://itrf.ensg.ign.fr/pub/itrf/WGS84.TXT`, visited on 3 Aug 2010. 6.10.1

Jenkins, R. (1960). *Curve surveying*. Cleaver-Hume Press. 6.6.1, 6.6.2

Julier, S. & Uhlmann, J. (1997). A new extension of the Kalman filter to nonlinear systems. *Int. Symp. Aerospace/Defense Sensing, Simul. and Controls*, **3**, 26. 3.2.7.2

Kalman, R. (1960). A new approach to linear filtering and prediction problems [CITATION]. *Transactions of the ASMEJournal of Basic Engineering*, **82 (Series D)**, 35–45. 1.2.2

Kreyszig, E. (2007). *Advanced engineering mathematics*. John Wiley & Sons, Inc., 9th edn. 5.4.4.1

Kuipers, J. (1999). *Quaternions and rotation sequences*. Princeton Univ. Press. 5.4.1, B

Lisi, M. (2009). GALILEO Program Status Update to ION GNSS 2009. *Proc. ION GNSS 2009, Savannah, GA*. 4.1.3

Lockyear, M. (1996). Changing track: moving-block railway signalling. *IEE Review*, **42**, 21–25. 2.1.4

Lucic, I. (2005). Holistic safety performance forecasting for train detection system. *Proc. IEEE International Conference on Systems, Man and Cybernetics*, **3**, 2973–2978. 2.1.3.3, 2.1.3.3

Madden, D. (2008). GPS Program Update to ION GNSS 2008. *Proc. ION GNSS 2008, Savannah, GA*. 4.5

Madden, D. (2009). GPS Program Update to ION GNSS 2009. *Proc. ION GNSS 2009, Savannah, GA*. 4.1.1.1

Mázl, R. & Přeučil, L. (2003). Sensor data fusion for inertial navigation of trains in GPS-dark areas. *Proc. IEEE Intelligent Vehicles Symposium*, 345–350. 2.3.3.2

McGuirk, J., Foster, G., Fixler, J., Snadden, M. & Kasevich, M. (2002). Sensitive absolute-gravity gradiometry using atom interferometry. *Physical Review A*, **65**, 33608. 5.1.3

Mertens, P., Franckart, J. & Starck, A. (2003). LOCOPROL: A low cost Train Location and Signalling system for "Low Density" Lines. *Proc. WCRR 2003, Cherryfield, Maine*. 1.1, 1.2.3, 2.3.3.1

Mirabadi, A., Mort, N. & Schmid, F. (1998). Fault detection and isolation in multisensor train navigation systems. *Proc. UKACC International Conference on Control '98 (Conf. Publ. No. 455)*, **2**. 2.3.3.2

Mirabadi, A., Mort, N. & Schmid, F. (1999). Design of fault tolerant train navigation systems. *Proc. the 1999 American Control Conference*, **1**. 2.3.3.2

Mirabadi, A., Schmid, F. & Mort, N. (2003). Multisensor Integration Methods in the Development of a Fault-Tolerant Train Navigation System. *The Journal of Navigation*, **56**, 385–398, Cambridge Univ Press. 2.3.3.2

Mirobotics Inc. (2009). MIDG IIC Specifications. *www.microboticsinc.com*, Product Datasheet. 5.2.3, 5.4.1, 6.9

Misra, P. & Enge, P. (2006). *Global Positioning System: Signals, Measurements and Performance*. Lincoln, MA: Ganga-Jamuna Press, 2nd edn. 4.1.1.2, 4.1.4, 4.2.1.1, 4.2.1.2, 4.2.5, 4.3.3, 4.3.4, 4.3.5, 6.10.4

Mueller, K.T., Dow, D., Brawner, J. & Bortins, R. (2003). Field Test Results with a High Speed Rail Prototype GPS Locomotive Location System. *Proc. ION GPS/GNSS 2003, Portland, OR*. 2.3.3.2

NGA (1997). Department of Defense World Geodetic System 1984. Tech. rep., National Imagery and Mapping Agency (now NGA). (document), 5.4.3.1, 6.10.1, A.1, C.2.3

Ober, P. (2003). *Integrity Prediction and Monitoring of Navigation Systems*. Integricom Publishers. 1.4

Palmer, J. (2006). The need for train detection. *Proc. the 11th IET Professional Development Course on Railway Signalling and Control Systems*, 47–53. 1.2.1, 2.1.3.1, 2.1.3.3, 2.1.3.3

Plaza, J., Llobera, A., Dominguez, C., Esteve, J., Salinas, I., Garcia, J. & Berganzo, J. (2004). BESOI-based integrated optical silicon accelerometer. *Microelectromechanical Systems, Journal of*, **13**, 355–364. 5.1.3

Revnivykh, S. (2008). GLONASS Status and Progress. *Proc. ION GNSS 2008, Savannah, GA*. 4.1.2

Rockwell Collins (2008). Micro INS. *www.rockwellcollins.com*, Product Datasheet. 5.2.3, 6.4

Rome, H.J. (2003). One Dimensional Integrated GPS–Odometer Navigation With Applications To Rail Track Analysis. *Proc. ION GPS/GNSS 2003, Portland, OR*. 2.3.3.2

RSSB (2008). Guidance on the Use of Satellite Navigation. Tech. rep., Rail Safety and Standards Board Limited (UK), Railway Group Guidance Note (GE/GN8578). 1.1, 2.3.2, 8.1.2, 9.3

RSSB (2009). ERTMS Operational Concept. Tech. rep., Rail Safety and Standards Board Limited (UK). 2.2

RSSB (2010). ETCS System Description. Tech. rep., Rail Safety and Standards Board Limited (UK), Railway Group Guidance Note (GE/GN8605). 1.2.1, 2.2, 2.2.1, 2.2.2, 2.2.3.2

Russian Space Agency (2010). GLONASS constellation status. `http://www.glonass-ianc.rsa.ru/pls/htmldb/f?p=202:20:1288933936855007::NO`, visited on 22nd April, 2010. 4.1.2

Siemens AG (2008a). Eurobalise S21 for Trainguard. *http://www.mobility.siemens.com*, Product Datasheet. 2.2.2, 2.8

Siemens AG (2008b). Trainguard - Full interoperability for European railways. *http://www.mobility.siemens.com*, Product Datasheet. 2.2.2

Signalhead (2007). Axle counter detection unit. `http://en.wikipedia.org/wiki/File:Axlecounter.jpg`, Image cited from Wikipedia under GNU Free Documentation License. 2.3

Signalhead (2009). A treadle fitted to a rail in the UK. `http://en.wikipedia.org/wiki/File:Rail_treadle.jpg`, Image cited from Wikipedia under Creative Commons Attribution-Share Alike 3.0 Unported license. 2.2

Simsky, A., Wilms, F. & Franckart, J. (2004). GNSS-based failsafe train positioning system for low-density traffic lines based on one-dimensional positioning algorithm. *Proc. 2nd ESA Workshop on Satellite Navigation User Equipment Technologics NAVITEC 2004, Noordwijk, Netherlands*. 1.2.3, 2.3.3.1, 8.1.1

Spilker, J. (1996). *Global Positioning System: Theory and Applications*, vol. I, chap. Satellite constellation and geometric dilution of precision, 177–208. AIAA. 4.1.1.1

Taylor, J., Pace, J., Mendicki, P.J. & Dorsey, A.J. (2008). GPS Control Segment Upgrade Goes Operational - Enhanced Phased Operations Transition Details. *Proc. ION NTM 2008, San Diego, CA*.

Tennent, R. (1971). *Science data book*. Oliver and Boyd, Edinburgh. 5.3.1

Thomas, M., Ejimogu, L., Lowe, D., Dumville, M., Roberts, W., Cross, P. *et al.* (2007). Dependability of GNSS on the UK Railways. *Proc. Nav 07, London*. 2.3.2

Thryduulf (2006). New Street station approach. `http://en.wikipedia.org/wiki/File:New_Street_station_approach.jpg`, Image cited from Wikipedia under GNU Free Documentation License. 2.11a

Titterton, D. & Weston, J. (2004). *Strapdown inertial navigation technology*. Peter Peregrinus Ltd, 2nd edn. 5, 5.1, 5.2, 5.3.2, 5.1, 5.4.3, 5.4.4.2, 6.7

Vallado, D. & McClain, W. (2001). *Fundamentals of astrodynamics and applications*. Kluwer Academic Publishers. 4.3.2

Van Dierendonck, A., Fenton, P. & Ford, T. (1992). Theory and performance of narrow correlator spacing in a GPS receiver. *Navigation: JION*, **39**, 265–283. 4.3.5

Walter, T., Blanch, J. & Enge, P. (2010). Future Augmented: Coverage Improvement for Dual-Frequency SBAS. *GPS World*, **March**, 36–41. 4.1.1.3

Wie, B. (2008). *Space vehicle dynamics and control*. AIAA Education Series, 2nd edn. 5.4.1

Zheng, Y. (2007). Improving Positioning Accuracy and Integrity in Rail Safety-Critical Applications through the Integration of GNSS with a Track Data Base. *Proc. ION GNSS 2007, Fort Worth, TX*. 1.2.3, 2.3.3.1, 8.1.1

Zheng, Y. (2008). *Integration of Satellite Positioning and a Track Database for Safety-Critical Railway Control Systems*. Ph.D. thesis, University College London. 1.2.2, 1.2.3, 2.3.1.1, 2.3.3.1, 6.2, 8.1.1, 8.1.2

# A. Coordinate Reference Frames and the Earth

Some basic mathematical functions used by navigation systems are explained in this appendix. The concept of a coordinate reference frame is introduced in Section A.1, along with the definitions of references frames used in this research. The transformations of coordinates and vectors between different reference frames are described in Section A.2. The modelling of the Earth surface and common constants of Earth models are given out in Section A.3.

## A.1. Coordinate Reference Frames

A coordinate reference frame defines the position and orientation of an object for navigation applications. The basic elements of a reference frame include an origin and a set of axes, in terms of which the motions of the navigated object can be described. Based on different origins and sets of axes, various reference frames are commonly used in navigation. Most navigation reference frames are orthogonal reference frames, i.e. axes of the frame are perpendicular to each other. Reference frames defined here are only the frames involved in this research.

### A.1.1. Inertial Frame

A inertial frame is a reference frame that does not accelerate or rotate with respect to the rest of the Universe (Groves, 2008). For navigation applications, the origin of the inertial frame is defined at the mass centre of the Earth, and hence been known as Earth-centred inertial frame. The inertial frame is usually denoted by a lower case character $i$.

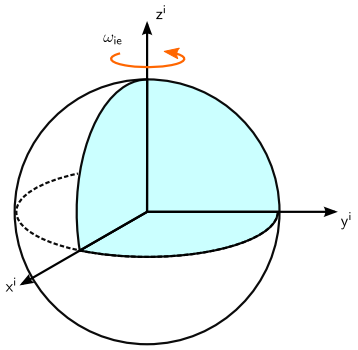



**Figure A.1.:** The inertial frame

An illustration of the inertial frame is shown in Figure A.1. The $z$-axis of the inertial frame is set along the Earth's rotation axis pointing the north pole. The $x$ and $y$-axis of the frame are set in the equatorial plane, where the $y$-axis is set 90° ahead of the $x$-axis along the Earth's rotation direction. The $x$ and $y$-axis stay still in the inertial space and do not rotate with the Earth. In order to uniquely define a inertial frame, the knowledge of the time when axes of the inertial frame coincide with the Earth frame is also needed.

### A.1.2. Earth-Centred Earth-Fixed Frame

The Earth-centred Earth-fixed (ECEF) frame, referred as the Earth frame in this thesis, rotates with the Earth, and its axes are fixed with respect to the Earth. The Earth frame is very important for navigation as the user requires positioning solution relative to the Earth. The Earth frame is usually denoted by a lower case character $e$.

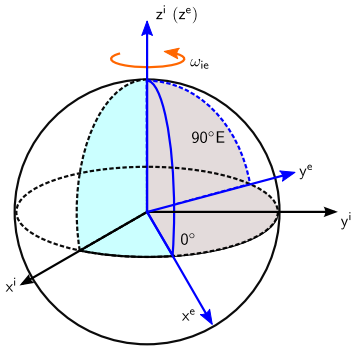



**Figure A.2.:** The Earth-centred Earth-fixed frame

An illustration of the Earth frame is shown in Figure A.2. The origin of the Earth frame lies at the centre of the ellipsoid Earth model. Similar to the inertial, the $z$-axis of the Earth frame is set along the Earth's rotation axis pointing the north pole, whereas the $x$ and $y$-axis of the frame are set in the equatorial plane. The $x$-axis points through the intersection of the equator and the zeros meridian. The$y$-axis points through the intersection of the equator and the 90° east meridian.

### A.1.3. Local Navigation Frame

The local navigation frame is defined by fitting a tangent plane to a fixed point on the surface of the Earth, usually the point a navigation solution is sought for. This point of interest is also used as the origin of the local navigation. For the expressing of the position and velocity vector, the local navigation frame therefore provides a convenient set of resolving axes. The local navigation frame is usually denoted by a lower case character $n$.

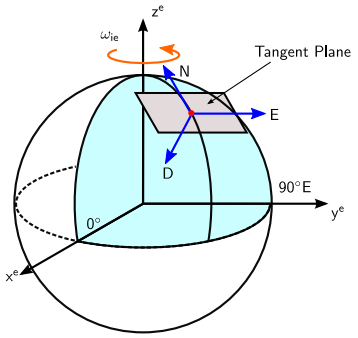



**Figure A.3.:** The local navigation frame

The definition of axes for a local navigation frame can be different for various applications. Figure A.3 demonstrates the definition of the local navigation frame used for this research. The $z$-axis of the local navigation frame is pointing from the origin to the centre of the Earth, and is usually denotes as $D$-axis. The $x$-axis, or $N$-axis, is set along the projection of the line from the origin to the north pole in the tangent plane, pointing to the north. The $y$-axis completes a right-handed orthogonal axes set with other axes, and is referred as $E$-axis.

### A.1.4. Body Frame

The body frame defines the orientation of the object with respect to other frames. All inertial measurements are initially expressed in the body frame, and is then transformed into other frames for navigation calculation. The body frame is usually denoted by a lower case character $b$.

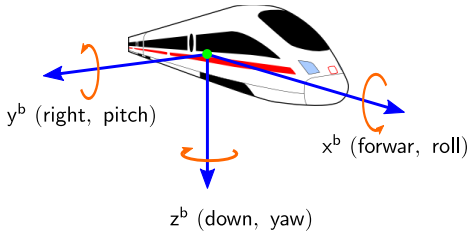



**Figure A.4.:** The body frame

The definition of the body frame in this thesis is illustrated by Figure A.4 with respect to a moving train. The body frame shares the same origin with the local navigation frame defined in Section A.1.3. The $x$-axis is pointing at the forward moving direction of the train. The $z$-axis is pointing at the down direction, along the direction of the gravity. The $y$-axis points at the right side of the train, completing a right-handed orthogonal coordinate frame. As the angular motion is also measured in the body frame, $x$-axis is also known as roll axis, $y$-axis is the pitch axis, and $z$-axis is the yaw axis.

## A.2. Frame Transformations

As can be seen from definitions of various reference frames in previous sections, the navigation process usually involves more than one coordinate reference frame. As a result, transforming quantities between different reference frames becomes necessary. This section introduces the transformation between the Earth frame and the local navigation frame, and the transformation between Cartesian ECEF position and curvilinear position. Note the transformation between the local navigation frame and the body frame is introduced in Section B.3.

### A.2.1. Earth and Local Navigation Frames

In order to transform vectors and matrices between the Earth and the local navigation frame. Direction cosine matrices, $\mathbf{C}_e^n$ and $\mathbf{C}_n^e$, are used, and they are defined as following

$$\mathbf{C}_e^n = \begin{bmatrix} -\sin L\cos\lambda & -\sin L\sin\lambda & \cos L \\ -\sin\lambda & \cos\lambda & 0 \\ -\cos L\cos\lambda & -\cos L\sin\lambda & -\sin L \end{bmatrix} \tag{A.1}$$

$$\mathbf{C}_n^e = \begin{bmatrix} -\sin L\cos\lambda & -\sin\lambda & -\cos L\cos\lambda \\ -\sin L\sin\lambda & \cos\lambda & -\cos L\sin\lambda \\ \cos L & 0 & -\sin L \end{bmatrix} \tag{A.2}$$

### A.2.2. Cartesian ECEF Position and Curvilinear Position

For navigation equations implemented in the local navigation frame, it is more natural to express the position solution in curvilinear position, i.e. latitude ($L$), longitude ($\lambda$) and height ($h$). The Cartesian ECEF position, $\mathbf{x}^e = [x,\ y,\ z]^T$, can be expressed using the curvilinear position as

$$x = (R_E(L) + h)\cos L \cos\lambda \tag{A.3}$$

$$y = (R_E(L) + h)\cos L \sin\lambda \tag{A.4}$$

$$z = \left[(1 - e^2)R_E(L) + h\right]\sin L \tag{A.5}$$

where $R_E(L)$ is defined by (A.10) and $e$ is the eccentricity of the ellipsoid. The inverse transformation is given by

$$\sin L = \frac{z}{[(1 - e^2)R_E(L) + h]} \tag{A.6}$$

$$\tan\lambda = \frac{y}{x} \tag{A.7}$$

$$h = \frac{\sqrt{x^2 + y^2}}{\cos L} - R_E(L) \tag{A.8}$$

## A.3. Earth Surface

The surface of the Earth is modelled as an ellipsoid. Various datum are built with different realisations of the global ellipsoid model. Each datum consists some common parameters including the length of the semi-major axis ($R$), the length of the semi-minor axis ($r$), the flattening of the ellipsoid ($f$), and the major eccentricity ($e$). WGS84 is the datum used by GPS. Its model is listed by Table A.1.

**Table A.1.:** WGS84 models (NGA, 1997)

| | |
|---|---|
| Semi-major axis, R | 6378137.0 m |
| Semi-minor axis, r | 6356752.3142 m |
| Flattening, f | 1/298.257223563 |
| Major eccentricity, e | 0.0818191908426 |
| Earth's rate, $\omega_{ie}$ | $7.292115 \times 10^{-5}$ rad/s |

Another two common terms used for navigation computation are the meridian radius of curvature ($R_N$) and a transverse radius of curvature ($R_E$). They defined as following

$$R_N(L) = \frac{R(1 - e^2)}{(1 - e^2 \sin^2 L)^{3/2}} \tag{A.9}$$

$$R_E(L) = \frac{R}{(1 - e^2 \sin^2 L)^{1/2}} \tag{A.10}$$

# B. Navigation Kinematics

The basic navigation kinematics related the attitude representation of a vehicle is briefly introduced here. It should be noted that only Euler angles and direction cosine matrix is introduced. Other common attitude representation methods include using quaternion and rotation vector. More information on attitude and frame rotation representation can be found in Kuipers (1999).

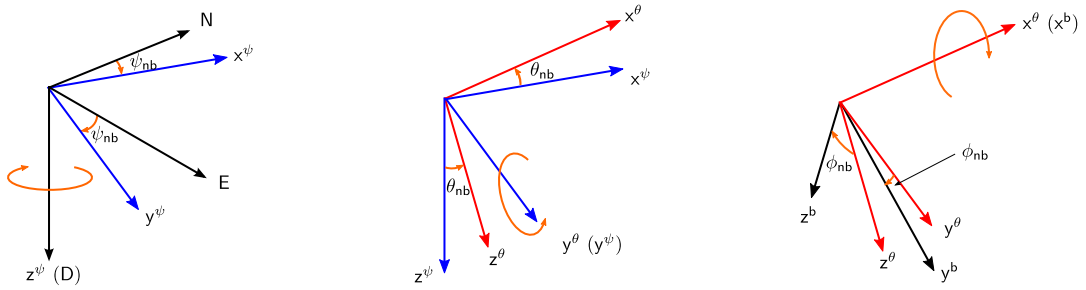
## B.1. Euler Angles

Euler angles are one of the common and intuitive ways to describe the attitude of a vehicle. For the case of inertial navigation calculation, they are mostly used for the transformation between the vehicle body frame and the local navigation frame ($\boldsymbol{\psi}_{nb}$). Euler angles are usually expressed in the form of a vector with three components including yaw ($\psi_{nb}$), pitch ($\theta_{nb}$) and roll ($\phi_{nb}$), and is written as

$$\boldsymbol{\psi}_{nb} = \begin{bmatrix} \phi_{nb} \\ \theta_{nb} \\ \psi_{nb} \end{bmatrix} \tag{B.1}$$

Note that although three angles are expressed together in the vector, they actually represent three successive rotations. Therefore, successive transformations cannot be simply expressed by add the Euler angles vectors together.

The sequence of performing the three rotations follows the order of roll, pitch and yaw, which is opposite to the order written in Equation (B.1). The rotations are illustrated as their implementing sequence in Figure B.1. The definitions of coordinate frames are given in Section A.1. There are two intermediate reference frames during the frame transformation process, caused by rotations of yaw and pitch, and they are respectively defined here as ($x^{\psi}$, $y^{\psi}$, $z^{\psi}$) and ($x^{\theta}$, $y^{\theta}$, $z^{\theta}$).




**(a)** Yaw **(b)** Pitch **(c)** Roll

**Figure B.1.:** Successive rotations represented by Euler angles

The first rotation, represented by yaw angle ($\psi_{nb}$), is performed by rotating the axis $D$ of the local navigation frame with the specified yaw angle. The resulting first intermediate reference frame, ($x^{\psi}$,

$y^{\psi}$, $z^{\psi}$), therefore shares a common axis $z^{\psi}$ ($D$) with the original body frame. The rotation can be written as

$$\begin{bmatrix} x^{\psi} \\ y^{\psi} \\ z^{\psi} \end{bmatrix} = \begin{bmatrix} \cos\psi_{nb} & \sin\psi_{nb} & 0 \\ -\sin\psi_{nb} & \cos\psi_{nb} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} N \\ E \\ D \end{bmatrix} \tag{B.2}$$

where ($N$, $E$, $D$) are coordinates in the local navigation frame.

The second rotation, represented by pitch angle ($\theta_{nb}$), is performed by rotating the axis $y^{\psi}$ of the first intermediate frame with the specified pitch angle. The resulted second intermediate reference frame, ($x^{\theta}$, $y^{\theta}$, $z^{\theta}$), therefore shares a common axis $y^{\theta}$ ($y^{\psi}$) with the first intermediate frame. The rotation can be written as

$$\begin{bmatrix} x^{\theta} \\ y^{\theta} \\ z^{\theta} \end{bmatrix} = \begin{bmatrix} \cos\theta_{nb} & 0 & -\sin\theta_{nb} \\ 0 & 1 & 0 \\ \sin\theta_{nb} & 0 & \cos\theta_{nb} \end{bmatrix} \begin{bmatrix} x^{\psi} \\ y^{\psi} \\ z^{\psi} \end{bmatrix} \tag{B.3}$$

Note that the pitch angle is usually limited between $-90°$ and $90°$ to avoid duplication.

The third rotation, represented by roll angle ($\phi_{nb}$), is performed by rotating the axis $x^{\theta}$ of the second intermediate frame with the specified roll angle. The destination of the rotation is the desired body frame, ($x^{b}$, $y^{b}$, $z^{b}$). A common axis $x^{b}$ ($x^{\theta}$) is shared by the body and the second intermediate frame. The rotation can be described as

$$\begin{bmatrix} x^{b} \\ y^{b} \\ z^{b} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_{nb} & \sin\phi_{nb} \\ 0 & -\sin\phi_{nb} & \cos\phi_{nb} \end{bmatrix} \begin{bmatrix} x^{\theta} \\ y^{\theta} \\ z^{\theta} \end{bmatrix} \tag{B.4}$$

## B.2. Direction Cosine Matrix

The direction cosine matrix is a $3 \times 3$ orthonormal matrix which is used to transform a vector or a matrix from one resolving frame to another. A direction cosine matrix is usually denoted by a bold upper case letter $\mathbf{C}_{\alpha}^{\beta}$, with a superscript $\beta$ represents the original frame before the transformation and a subscript $\alpha$ represents the frame after the transformation.

Each element of the direction cosine matrix is the product of the unit vectors describing frame axes of two frames. Thus $\mathbf{C}_{\alpha}^{\beta}$ can be defined as

$$\mathbf{C}_{\alpha}^{\beta} = \begin{bmatrix} \mathbf{u}_{\beta x} \cdot \mathbf{u}_{\alpha x} & \mathbf{u}_{\beta x} \cdot \mathbf{u}_{\alpha y} & \mathbf{u}_{\beta x} \cdot \mathbf{u}_{\alpha z} \\ \mathbf{u}_{\beta y} \cdot \mathbf{u}_{\alpha x} & \mathbf{u}_{\beta y} \cdot \mathbf{u}_{\alpha y} & \mathbf{u}_{\beta y} \cdot \mathbf{u}_{\alpha z} \\ \mathbf{u}_{\beta z} \cdot \mathbf{u}_{\alpha x} & \mathbf{u}_{\beta z} \cdot \mathbf{u}_{\alpha y} & \mathbf{u}_{\beta z} \cdot \mathbf{u}_{\alpha z} \end{bmatrix} \tag{B.5}$$

where $\mathbf{u}$ are unit vectors. The manipulation of direction cosine matrix is relative easy comparing to the Euler angles. The transpose of a direction cosine matrix represents a reverse transformation,

$$\mathbf{C}_{\alpha}^{\beta} = (\mathbf{C}_{\beta}^{\alpha})^{T} \tag{B.6}$$

In addition, successive transformations can be represented by simply multiplying individual direction cosine matrix according to their performed order,

$$\mathbf{C}_{\alpha}^{\beta} = \mathbf{C}_{\gamma}^{\beta}\mathbf{C}_{\alpha}^{\gamma} \tag{B.7}$$

In order to transform a vector into another reference frame, the direction cosine matrix can be applied as following

$$\mathbf{x}^{\beta} = \mathbf{C}_{\alpha}^{\beta}\mathbf{x}^{\alpha} \tag{B.8}$$

where $\mathbf{x}$ is the vector to be transformed. In case the transformation of a matrix is required, the direction cosine matrix needs to be applied using

$$\mathbf{\Omega}^{\beta} = \mathbf{C}_{\alpha}^{\beta}\mathbf{\Omega}^{\alpha}\mathbf{C}_{\beta}^{\alpha} \tag{B.9}$$

## B.3. Transformations

The transformation between different attitude representations is needed for this research to build the simulated attitude information and process navigation calculations. The direction cosine matrix denoting the transformation from the local navigation frame to the body frame, $\mathbf{C}_n^b$, can be expressed in terms of Euler angles by multiplying the three successive rotations Equation (B.2), Equation (B.3) and Equation (B.4), and is written here as

$$\begin{aligned}\mathbf{C}_n^b &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_{nb} & \sin\phi_{nb} \\ 0 & -\sin\phi_{nb} & \cos\phi_{nb} \end{bmatrix} \begin{bmatrix} \cos\theta_{nb} & 0 & -\sin\theta_{nb} \\ 0 & 1 & 0 \\ \sin\theta_{nb} & 0 & \cos\theta_{nb} \end{bmatrix} \begin{bmatrix} \cos\psi_{nb} & \sin\psi_{nb} & 0 \\ -\sin\psi_{nb} & \cos\psi_{nb} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \cos\psi_{nb}\cos\theta_{nb} & \sin\psi_{nb}\cos\theta_{nb} & -\sin\theta_{nb} \\ \begin{pmatrix} -\sin\psi_{nb}\cos\phi_{nb}+ \\ \cos\psi_{nb}\sin\theta_{nb}\sin\phi_{nb} \end{pmatrix} & \begin{pmatrix} \cos\psi_{nb}\sin\phi_{nb}+ \\ \sin\psi_{nb}\sin\theta_{nb}\sin\phi_{nb} \end{pmatrix} & \cos\theta_{nb}\sin\phi_{nb} \\ \begin{pmatrix} \sin\psi_{nb}\sin\phi_{nb}+ \\ \cos\psi_{nb}\sin\theta_{nb}\cos\phi_{nb} \end{pmatrix} & \begin{pmatrix} -\cos\psi_{nb}\sin\phi_{nb}+ \\ \sin\psi_{nb}\sin\theta_{nb}\cos\phi_{nb} \end{pmatrix} & \cos\theta_{nb}\cos\phi_{nb} \end{bmatrix} \end{aligned} \tag{B.10}$$

and the Euler angles can be obtained using elements of $\mathbf{C}_n^b$ by

$$\phi_{nb} = \arctan 2(C_{2,3},\ C_{3,3}) = \arctan 2(C_{3,2},\ C_{3,3}) \tag{B.11}$$

$$\theta_{nb} = -\arcsin C_{1,3} = -\arcsin C_{3,1} \tag{B.12}$$

$$\psi_{nb} = \arctan 2(C_{1,2},\ C_{1,1}) = \arctan 2(C_{2,1},\ C_{1,1}) \tag{B.13}$$

where $C_{i,j}$ denotes the $j$th elements of the $i$th row of the direction cosine matrix $\mathbf{C}_n^b$.

# C. INS State Propagation in the Local Navigation Frame

Both loosely coupled system and tightly coupled system share the same propagation process for their inertial states. As the error-state approach is used for this research, the dynamic equations for the inertial error states should be derived first from the inertial equations given in Section 5.4.3. All obtained equations are expressed in the local navigation frame. The inertial matrices used for the Kalman filtering process can then be designed based on the dynamic equations.

This section presents a concise description on the design process of inertial matrices. The derivation shown in this section mainly follows Groves (2008) and Farrell (2008), where more detailed derivations are presented.

## C.1. INS State Vector

The inertial states selected for this research consist of 13 error states and 2 biases terms, and are given by the following $15 \times 1$ vector,

$$\mathbf{x}_{INS} = [\delta\boldsymbol{\psi}_{nb},\ \delta\mathbf{p}^n,\ \delta\mathbf{v}^n_{eb},\ \mathbf{b}_g,\ \mathbf{b}_a]^T \tag{C.1}$$

The attitude error, $\delta\boldsymbol{\psi}_{nb}$, contains the small-angle rotations defined with respect to the local navigation frame, which rotate the truth local navigation frame to align with the computed local navigation frame. $\delta\boldsymbol{\psi}_{nb}$ can therefore be expressed as

$$\delta\boldsymbol{\psi}_{nb} = [\epsilon_N,\ \epsilon_E,\ \epsilon_D]^T \tag{C.2}$$

where $\epsilon_N$ and $\epsilon_E$ are referred as tilt errors, and $\epsilon_D$ is the yaw error.

The position error, $\delta\mathbf{p}^n$, is expressed as the errors in the latitude, longitude and height estimations, and is written as

$$\delta\mathbf{p}^n = [\delta L,\ \delta\lambda,\ \delta h]^T \tag{C.3}$$

The velocity error, $\delta\mathbf{v}^n_{eb}$, is expressed with respect to the local navigation frame in the direction of north, east, and down. The vector is defined as

$$\delta\mathbf{v}^n_{eb} = [\delta v_N,\ \delta v_E,\ \delta v_D]^T \tag{C.4}$$

The biases terms, $\mathbf{b}_g$ and $\mathbf{b}_a$, are introduced in Section 5.3.1 and are defined by (5.2) and (5.1).

## C.2. INS Error State Dynamic Equations

The state propagation model for the inertial states are first written as

$$\dot{\mathbf{x}}_{INS} = \mathbf{F}_{INS}\mathbf{x}_{INS} + \mathbf{G}\mathbf{w}_{INS} \tag{C.5}$$

where the inertial system matrix, $\mathbf{F}_{INS}$, is defined by the following $15 \times 15$ matrix

$$\mathbf{F}_{INS} = \begin{bmatrix} \mathbf{F}_{\psi\psi} & \mathbf{F}_{\psi p} & \mathbf{F}_{\psi v} & \hat{\mathbf{C}}_b^n & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{F}_{pp} & \mathbf{F}_{pv} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{F}_{v\psi} & \mathbf{F}_{vp} & \mathbf{F}_{vv} & \mathbf{0}_{3\times3} & \hat{\mathbf{C}}_b^n \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \end{bmatrix} \tag{C.6}$$

Note that sub-matrices in (C.6) contain partial derivatives obtained from the inertial error dynamic equations, and are developed gradually throughout this section.

The system noise vector, $\mathbf{w}_{INS}$, models the random noises in the state propagation process, and is written as

$$\mathbf{w}_{INS} = [\mathbf{w}_g,\ \mathbf{w}_a,\ \mathbf{w}_{gd},\ \mathbf{w}_{ad}]^T \tag{C.7}$$

where $\mathbf{w}_g$ and $\mathbf{w}_a$ are respectively the random noise presented on the measurements from gyros and accelerometers, $\mathbf{w}_{gd}$ and $\mathbf{w}_{ad}$ are zero-mean Gaussian white-noise processes caused by in-run biases presented respectively on $\mathbf{b}_g$ and $\mathbf{b}_a$. $\mathbf{w}_g$ and $\mathbf{w}_a$ are introduced in Section 5.3.2 and are defined by (5.4) and (5.3). The system noise distribution matrix, $\mathbf{G}$, is given by

$$\mathbf{G} = \begin{bmatrix} \hat{\mathbf{C}}_b^n & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \hat{\mathbf{C}}_b^n & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \end{bmatrix} \tag{C.8}$$

The transition matrix, $\mathbf{\Phi}_{INS}$, for the discrete propagation model can be computed with $\mathbf{F}_{INS}$ using a first order power-series expansions of (3.4), and is written here as

$$\mathbf{\Phi}_{INS} = \mathbf{I}_{15\times15} + \mathbf{F}_{INS}\tau_s \tag{C.9}$$

where $\tau_s$ is the system update interval. The system states propagates at the same rate with the inertial calculation, therefore $\tau_s = \tau_i = 0.01\,\text{s}$.

## C.2.1. Attitude

The attitude error in the local navigation frame mainly arises from the measurement error of the rotation between the body frame the local navigation frame. The rotation between these two frames, as shown in Section 5.4.3.1, comprises gyro measurement ($\delta\boldsymbol{\omega}_{ib}^b$), Earth rate ($\delta\boldsymbol{\omega}_{ie}^b$), and transport rate terms ($\delta\boldsymbol{\omega}_{en}^b$). Thus, $\delta\boldsymbol{\psi}_{nb}$ can be written and expanded as

$$\begin{aligned}\delta\dot{\boldsymbol{\psi}}_{nb} &\approx \hat{\mathbf{C}}_b^n(\tilde{\boldsymbol{\omega}}_{nb}^b - \boldsymbol{\omega}_{nb}^b) \\ &\approx \hat{\mathbf{C}}_b^n\delta\boldsymbol{\omega}_{ib}^b - \hat{\mathbf{C}}_b^n\delta\boldsymbol{\omega}_{ie}^b - \hat{\mathbf{C}}_b^n\delta\boldsymbol{\omega}_{en}^b\end{aligned} \tag{C.10}$$

where the Earth rate and transport rate terms can be expanded as

$$\hat{\mathbf{C}}_b^n\delta\boldsymbol{\omega}_{ie}^b + \hat{\mathbf{C}}_b^n\delta\boldsymbol{\omega}_{en}^b \approx \hat{\boldsymbol{\Omega}}_{in}^n\delta\boldsymbol{\psi}_{nb} + \delta\boldsymbol{\omega}_{ie}^n + \delta\boldsymbol{\omega}_{en}^n \tag{C.11}$$

Note that the products of error states are neglected.

The rotation of the local navigation frame term $\hat{\boldsymbol{\Omega}}_{in}^n$ can be obtained with

$$\hat{\boldsymbol{\Omega}}_{in}^n = \hat{\boldsymbol{\Omega}}_{ie}^n + \hat{\boldsymbol{\Omega}}_{en}^n \tag{C.12}$$

where the Earth rate$\hat{\boldsymbol{\Omega}}_{ie}^n$ and the transport rate $\hat{\boldsymbol{\Omega}}_{en}^n$ are calculated using (5.12) and (5.14).

The Earth rate term resolved in the local navigation frame, $\delta\boldsymbol{\omega}_{ie}^n$, can be derived from (5.11) as

$$\delta\boldsymbol{\omega}_{ie}^n = \tilde{\boldsymbol{\omega}}_{ie}^n - \boldsymbol{\omega}_{ie}^n = -\omega_{ie}\begin{bmatrix}\sin L \\ 0 \\ \cos L\end{bmatrix}\delta L \tag{C.13}$$

whereas the transport rate term in the local navigation frame, $\delta\boldsymbol{\omega}_{en}^n$, can be derived from (5.13), neglecting products of error states and the variation of curvature transverse radius,

$$\delta\boldsymbol{\omega}_{en}^n = \tilde{\boldsymbol{\omega}}_{en}^n - \boldsymbol{\omega}_{en}^n = \begin{bmatrix}\frac{\delta v_{eb,E}^n}{R_E(\hat{L})+\hat{h}} \\ -\frac{\delta v_{eb,N}^n}{R_N(\hat{L})+\hat{h}} \\ \frac{\delta v_{eb,E}^n \tan\hat{L}}{R_E(\hat{L})+\hat{h}}\end{bmatrix} - \begin{bmatrix}0 \\ 0 \\ 1\end{bmatrix}\frac{\hat{v}_{eb,E}^n}{(R_E(\hat{L})+\hat{h})\cos^2\hat{L}}\delta L + \begin{bmatrix}-\frac{\hat{v}_{eb,E}^n}{(R_E(\hat{L})+\hat{h})^2} \\ \frac{\hat{v}_{eb,N}^n}{(R_N(\hat{L})+\hat{h})^2} \\ \frac{\hat{v}_{eb,E}^n \tan\hat{L}}{(R_E(\hat{L})+\hat{h})^2}\end{bmatrix}\delta h \tag{C.14}$$

The partial derivative sub-matrices for the attitude can therefore be defined as

$$\mathbf{F}_{\psi\psi} = -\hat{\boldsymbol{\Omega}}_{in}^n \tag{C.15}$$

$$\mathbf{F}_{\psi p} = \begin{bmatrix}\omega_{ie}\sin\hat{L} & 0 & \frac{\hat{v}_{eb,E}^n}{(R_E(\hat{L})+\hat{h})^2} \\ 0 & 0 & -\frac{\hat{v}_{eb,N}^n}{(R_N(\hat{L})+\hat{h})^2} \\ \omega_{ie}\cos\hat{L} + \frac{\hat{v}_{eb,E}^n}{(R_E(\hat{L})+\hat{h})\cos^2\hat{L}} & 0 & \frac{-\hat{v}_{eb,E}^n\tan\hat{L}}{(R_E(\hat{L})+\hat{h})^2}\end{bmatrix} \tag{C.16}$$

$$\mathbf{F}_{\psi v} = \begin{bmatrix}0 & \frac{-1}{R_E(\hat{L})+\hat{h}} & 0 \\ \frac{1}{R_N(\hat{L})+\hat{h}} & 0 & 0 \\ 0 & \frac{\tan\hat{L}}{R_E(\hat{L})+\hat{h}} & 0\end{bmatrix} \tag{C.17}$$

### C.2.2. Position

The time derivative of latitude, longitude and height is given in (5.22) to (5.24). The time derivative of the position error is derived based on those equations, neglecting products of error states and the

variation of curvature radii,

$$\delta\dot{L} = \frac{\delta v^n_{eb,N}}{R_N(\hat{L}) + \hat{h}} - \frac{\hat{v}^n_{eb,N}\delta h}{(R_N(\hat{L}) + \hat{h})^2} \tag{C.18}$$

$$\delta\dot{\lambda} = \frac{\delta v^n_{eb,E}}{(R_E(\hat{L}) + \hat{h})\cos\hat{L}} + \frac{\hat{v}^n_{eb,E}\sin\hat{L}\delta L}{(R_E(\hat{L}) + \hat{h})\cos^2\hat{L}} + \frac{\hat{v}^n_{eb,E}\delta h}{(R_E(\hat{L}) + \hat{h})^2\cos\hat{L}} \tag{C.19}$$

$$\delta\dot{h} = -\delta v^n_{eb,D} \tag{C.20}$$

The partial derivative sub-matrices for the position can therefore be defined as

$$\mathbf{F}_{pp} = \begin{bmatrix} 0 & 0 & -\frac{\hat{v}^n_{eb,N}}{(R_N(\hat{L})+\hat{h})^2} \\ \frac{\hat{v}^n_{eb,E}\sin\hat{L}}{(R_E(\hat{L})+\hat{h})\cos^2\hat{L}} & 0 & \frac{\hat{v}^n_{eb,E}}{(R_E(\hat{L})+\hat{h})^2\cos\hat{L}} \\ 0 & 0 & 0 \end{bmatrix} \tag{C.21}$$

$$\mathbf{F}_{pv} = \begin{bmatrix} \frac{1}{R_N(\hat{L})+\hat{h}} & 0 & 0 \\ 0 & \frac{1}{(R_E(\hat{L})+\hat{h})\cos\hat{L}} & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{C.22}$$

## C.2.3. Velocity

The time derivative of the velocity can be expressed in terms of the specific force, gravity, and centrifugal acceleration, as given in (5.18), and the velocity error dynamic equation is given as following

$$\begin{aligned} \delta\dot{\mathbf{v}}^n_{eb} \approx & -(\hat{\mathbf{C}}^n_b\hat{\mathbf{f}}^b_{ib}) \wedge \delta\boldsymbol{\psi}_{nb} - (\boldsymbol{\Omega}^n_{en} + 2\boldsymbol{\Omega}^n_{ie})\delta\mathbf{v}^n_{eb} + \mathbf{v}^n_{eb} \wedge (\tilde{\boldsymbol{\omega}}^n_{en} - \boldsymbol{\omega}^n_{en}) \\ & + 2\mathbf{v}^n_{eb} \wedge (\tilde{\boldsymbol{\omega}}^n_{ie} - \boldsymbol{\omega}^n_{ie}) - \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \frac{2g_0(\hat{L})}{r^e_{eS}(\hat{L})}\delta h + \hat{\mathbf{C}}^n_b\mathbf{b}_a \end{aligned} \tag{C.23}$$

where $\wedge$ denotes the cross product of vectors, $g_0$ is the gravity, $r^e_{eS}$ is the geocentric radius at the surface.

The gravity model here used is provided by the WGS 84 datum, known as the Somigliana model (NGA, 1997), and is given as following

$$g_0(\hat{L}) \approx 9.7803253359\frac{(1 + 0.001931853\sin^2\hat{L})}{\sqrt{1 - e^2\sin^2\hat{L}}} \tag{C.24}$$

The geocentric radius at the surface is written as

$$r^e_{eS}(\hat{L}) = R_E(L)\sqrt{\cos^2 L + (1 - e^2)^2\sin^2 L} \tag{C.25}$$

The partial derivative sub-matrices for the Velocity can therefore be defined as

$$\mathbf{F}_{v\psi} = -\left[(\hat{\mathbf{C}}_b^n \hat{\mathbf{f}}_{ib}^b)\wedge\right] \tag{C.26}$$

$$\mathbf{F}_{vp} = \begin{bmatrix} \begin{pmatrix} -\frac{(\hat{v}_{eb,E}^n)^2 \sec^2 \hat{L}}{R_E(\hat{L})+\hat{h}} \\ -2\hat{v}_{eb,E}^n \omega_{ie} \cos \hat{L} \end{pmatrix} & 0 & \begin{pmatrix} \frac{(\hat{v}_{eb,E}^n)^2 \tan \hat{L}}{(R_E(\hat{L})+\hat{h})^2} \\ -\frac{\hat{v}_{eb,N}^n \hat{v}_{eb,D}^n}{(R_N(\hat{L})+\hat{h})^2} \end{pmatrix} \\ \begin{pmatrix} \frac{\hat{v}_{eb,N}^n \hat{v}_{eb,E}^n \sec^2 \hat{L}}{R_E(\hat{L})+\hat{h}} + 2\hat{v}_{eb,N}^n \omega_{ie} \cos \hat{L} \\ -2\hat{v}_{eb,D}^n \omega_{ie} \sin \hat{L} \end{pmatrix} & 0 & \begin{pmatrix} -\frac{\hat{v}_{eb,N}^n \hat{v}_{eb,E}^n \tan \hat{L}}{(R_E(\hat{L})+\hat{h})^2} \\ -\frac{\hat{v}_{eb,E}^n \hat{v}_{eb,D}^n}{(R_E(\hat{L})+\hat{h})^2} \end{pmatrix} \\ 2\hat{v}_{eb,E}^n \omega_{ie} \sin \hat{L} & 0 & \begin{pmatrix} \frac{(\hat{v}_{eb,E}^n)^2}{(R_E(\hat{L})+\hat{h})^2} + \frac{(\hat{v}_{eb,N}^n)^2}{(R_N(\hat{L})+\hat{h})^2} \\ -\frac{2g_0(\hat{L})}{r_{eS}^e(\hat{L})} \end{pmatrix} \end{bmatrix} \tag{C.27}$$

$$\mathbf{F}_{vv} = \begin{bmatrix} \frac{\hat{v}_{eb,D}^n}{R_N(\hat{L})+\hat{h}} & \left(-\frac{2\hat{v}_{eb,E}^n \tan \hat{L}}{R_E(\hat{L})+\hat{h}} - 2\omega_{ie} \sin \hat{L}\right) & \frac{\hat{v}_{eb,N}^n}{R_N(\hat{L})+\hat{h}} \\ \left(\frac{\hat{v}_{eb,E}^n \tan \hat{L}}{R_E(\hat{L})+\hat{h}} + 2\omega_{ie} \sin \hat{L}\right) & \frac{\hat{v}_{eb,N}^n \tan \hat{L} + \hat{v}_{eb,D}^n}{R_E(\hat{L})+\hat{h}} & \left(\frac{\hat{v}_{eb,E}^n}{R_E(\hat{L})+\hat{h}} + 2\omega_{ie} \cos \hat{L}\right) \\ \frac{2\hat{v}_{eb,N}^n}{R_N(\hat{L})+\hat{h}} & \left(-\frac{2\hat{v}_{eb,E}^n}{R_E(\hat{L})+\hat{h}} - 2\omega_{ie} \cos \hat{L}\right) & 0 \end{bmatrix} \tag{C.28}$$

## C.3. System Noise

As described in Section 3.2.2, the system noise accounts for the increase of the system estimation uncertainty during the system state propagation. For the inertial error propagation, the main source of the system noise comes from the random noises on the inertial measurement and the variation of sensor in-run biases. The modelling of these random noises is discussed in Section 6.9. Based on the inertial error propagation matrices described in previous sections, the system noise covariance matrix is

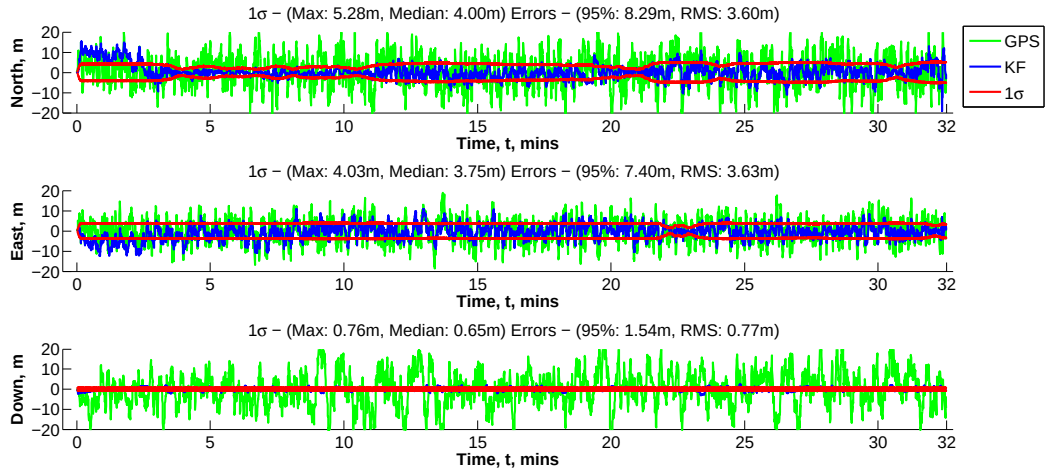
$$\mathbf{Q}_{INS} = \begin{bmatrix} n_g^2 \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & n_a^2 \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & n_{gd}^2 \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & n_{ad}^2 \mathbf{I}_{3\times3} \end{bmatrix} \tau_s \tag{C.29}$$

where $n_g^2$ and $n_a^2$ are power spectral densities of the random noises on angular rate and specific force measurement, $n_{gd}^2$ and $n_{ad}^2$ are power spectral densities of the in-run biases of gyros and accelerometers.

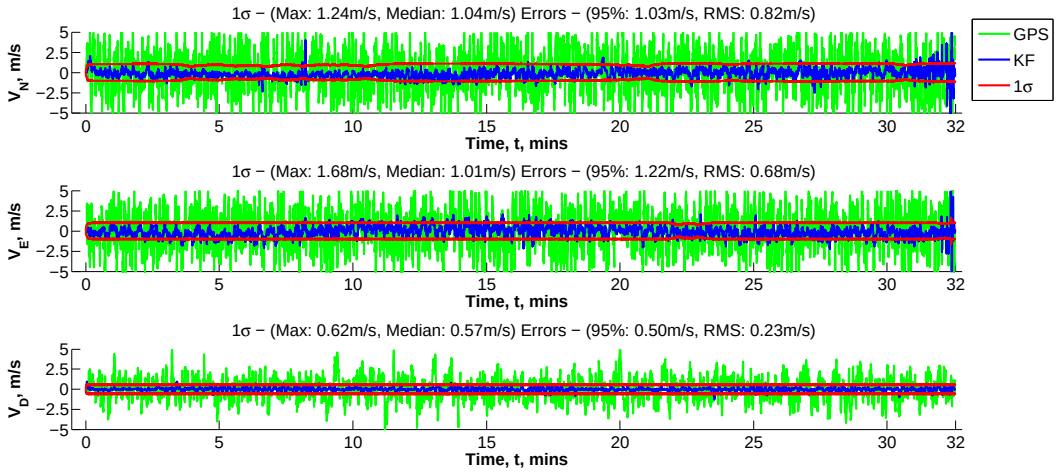
# D. More Results

## D.1. More Results on DRMs with Different Point Intervals

The following graph shows more results of experiments on using DRMs with various point intervals.

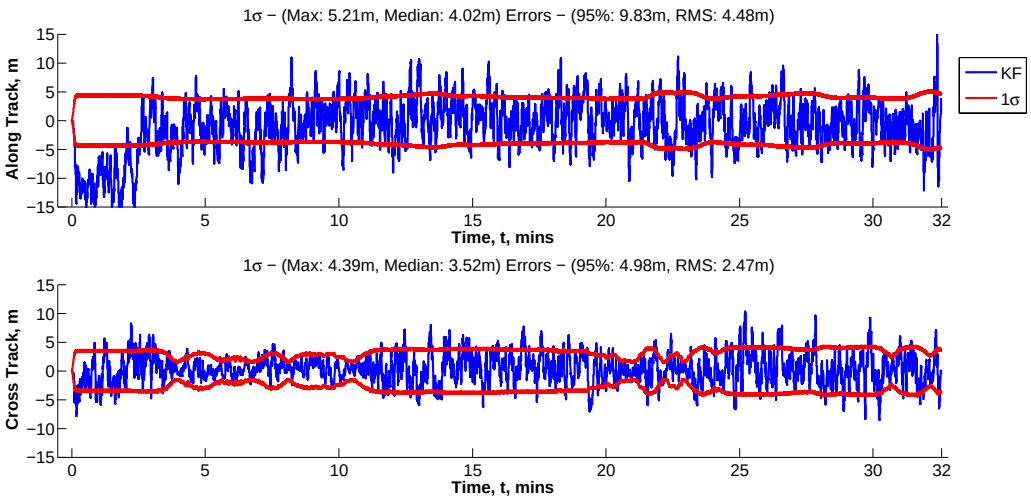



**(a)** Position errors

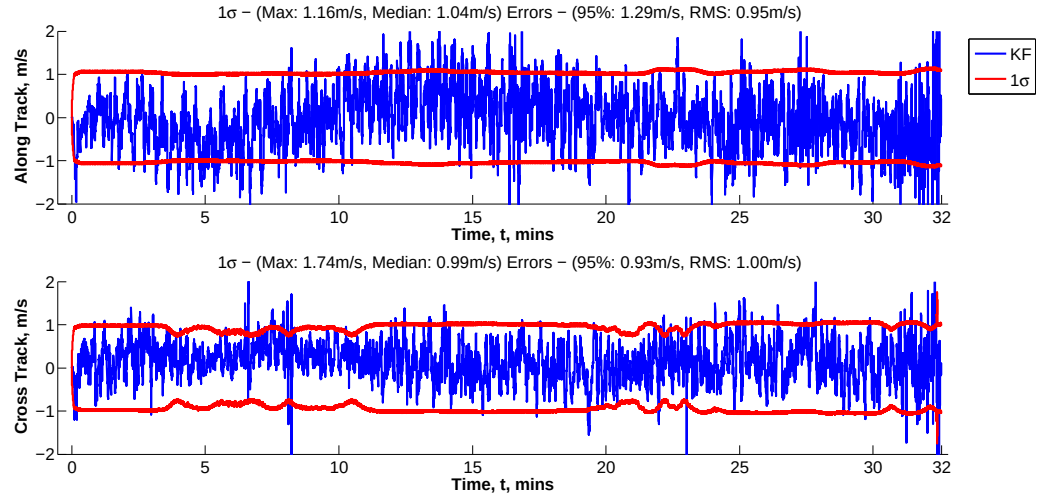



**(b)** Velocity errors

**Figure D.1.:** General performance of a system using a DRM with 50 m point intervals and 1 m point errors

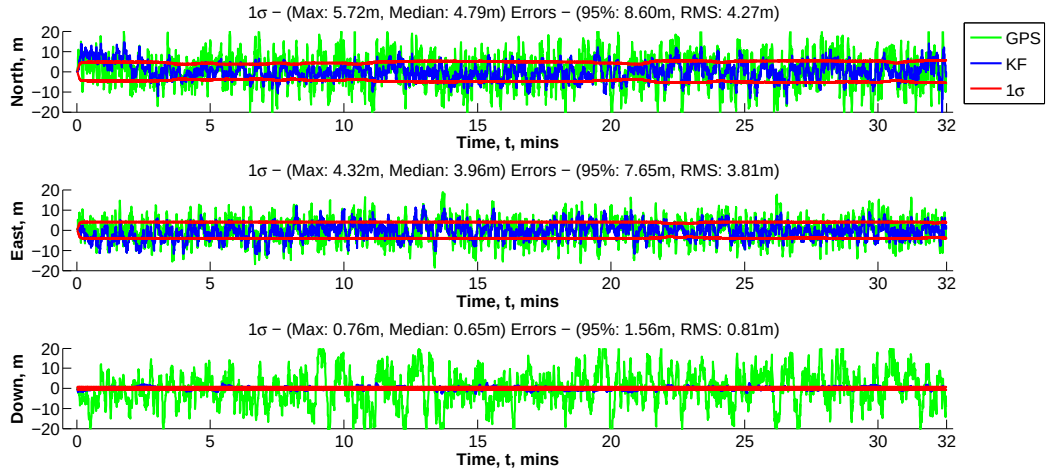



**(a)** Position errors

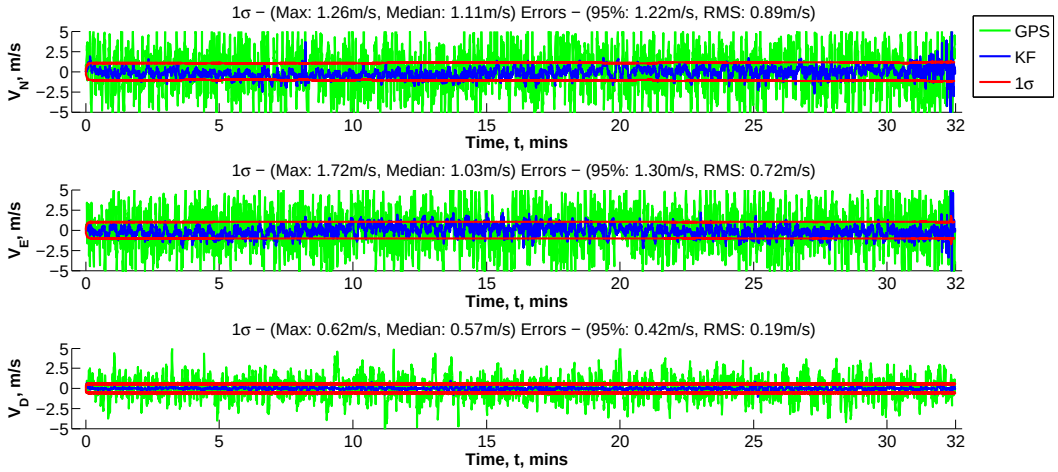



**(b)** Velocity errors

**Figure D.2.:** Along track and cross track performance of a system using a DRM with 50 m point intervals and 1 m point errors

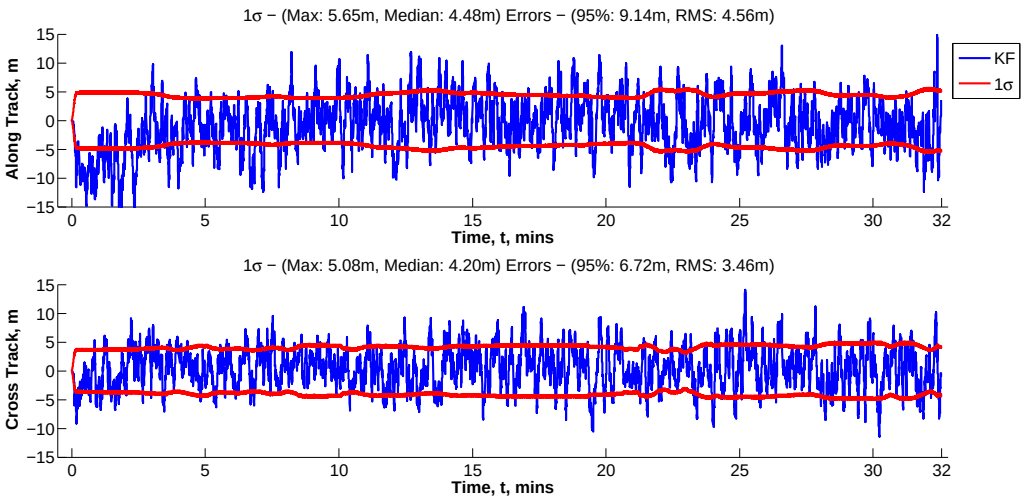



**(a)** Position errors

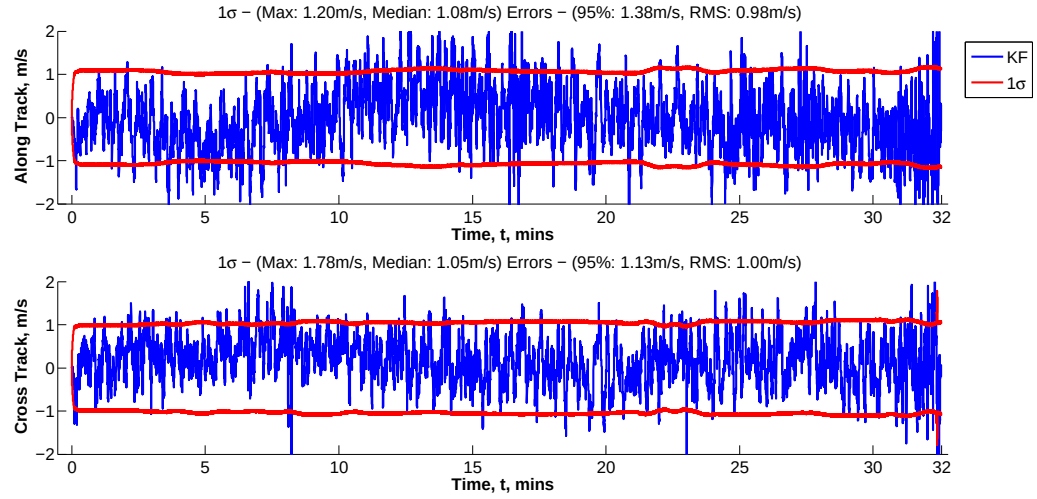



**(b)** Velocity errors

**Figure D.3.:** General performance of a system using a DRM with 100 m point intervals and 1 m point errors

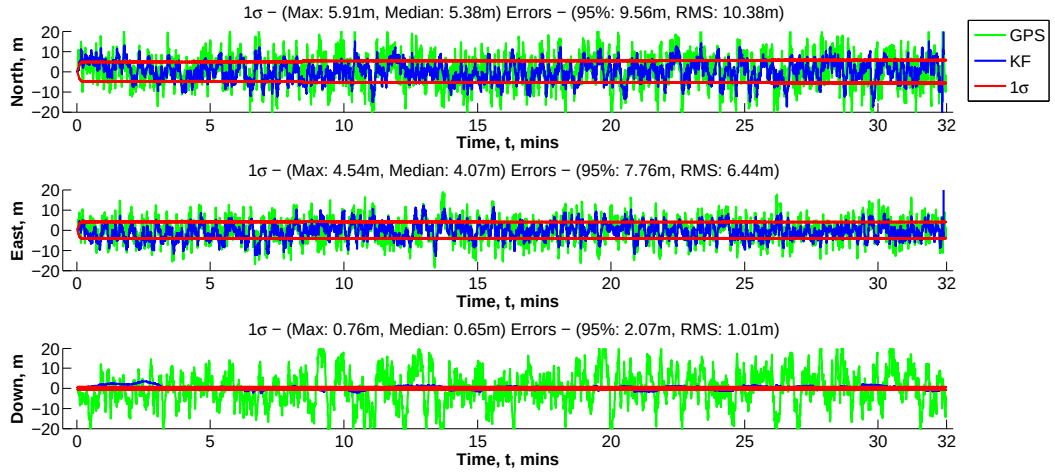



**(a)** Position errors

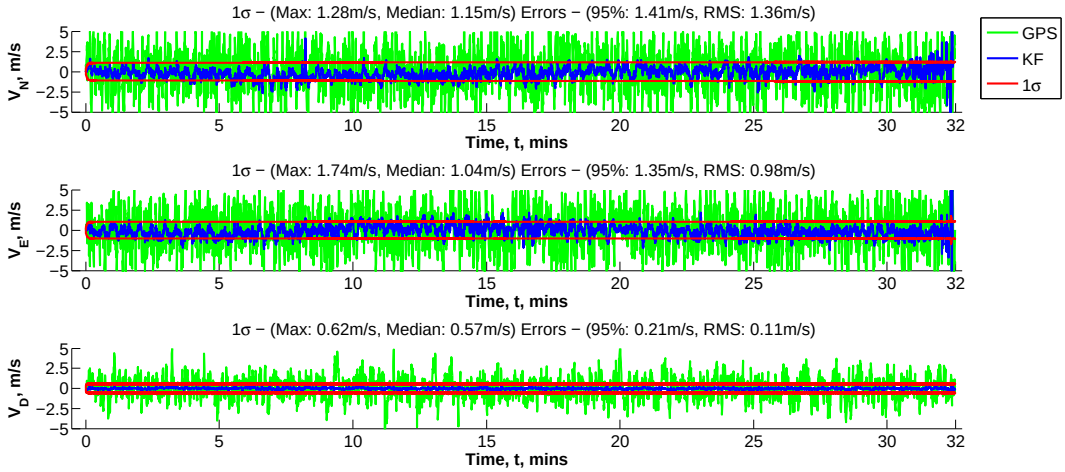



**(b)** Velocity errors

**Figure D.4.:** Along track and cross track performance of a system using a DRM with 100 m point intervals and 1 m point errors

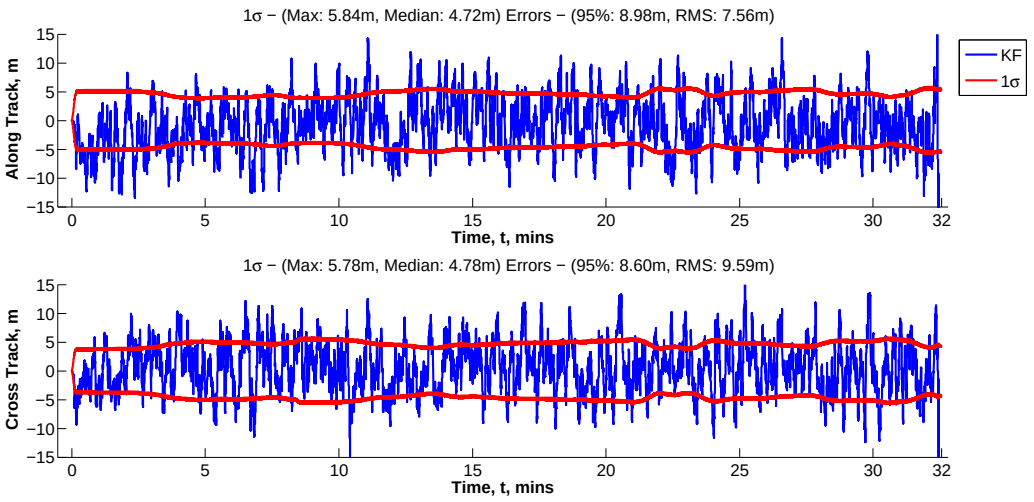



**(a)** Position errors

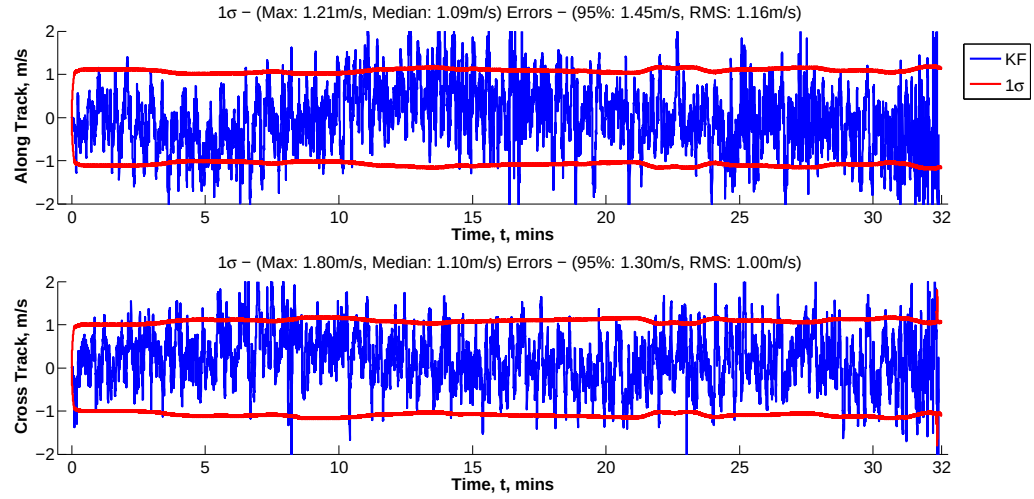



**(b)** Velocity errors

**Figure D.5.:** General performance of a system using a DRM with 500 m point intervals and 1 m point errors




**(a)** Position errors

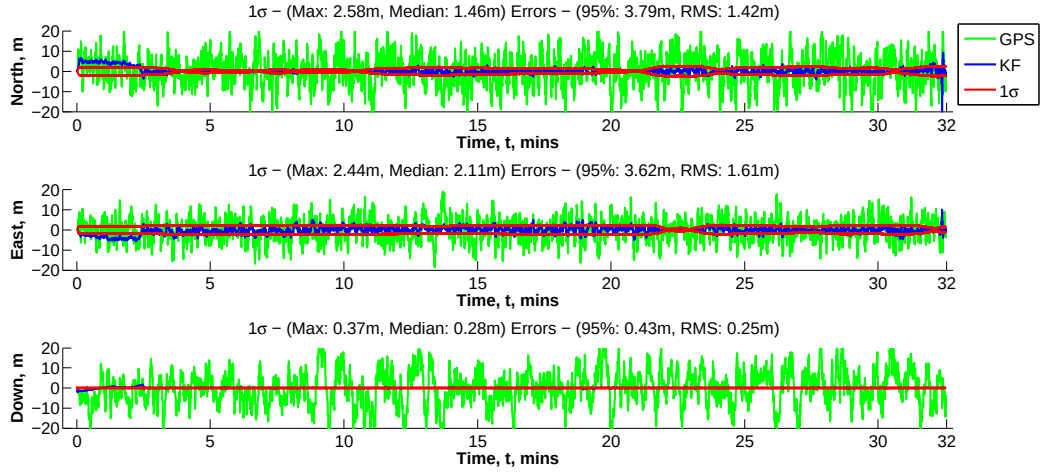



**(b)** Velocity errors

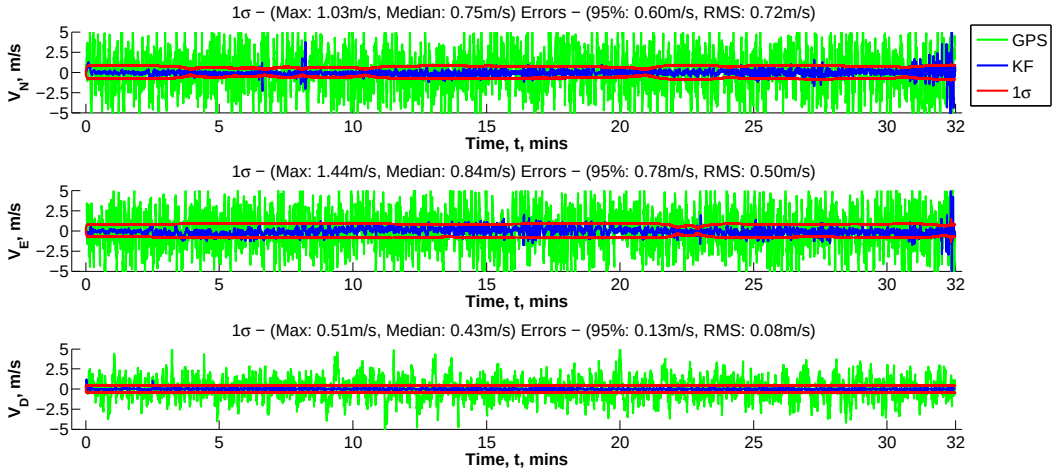
**Figure D.6.:** Along track and cross track performance of a system using a DRM with 500 m point intervals and 1 m point errors

## D.2. More Results on DRMs with Different Accuracy Levels

The following graph shows more results of experiments on using DRMs with various surveying accuracy levels.

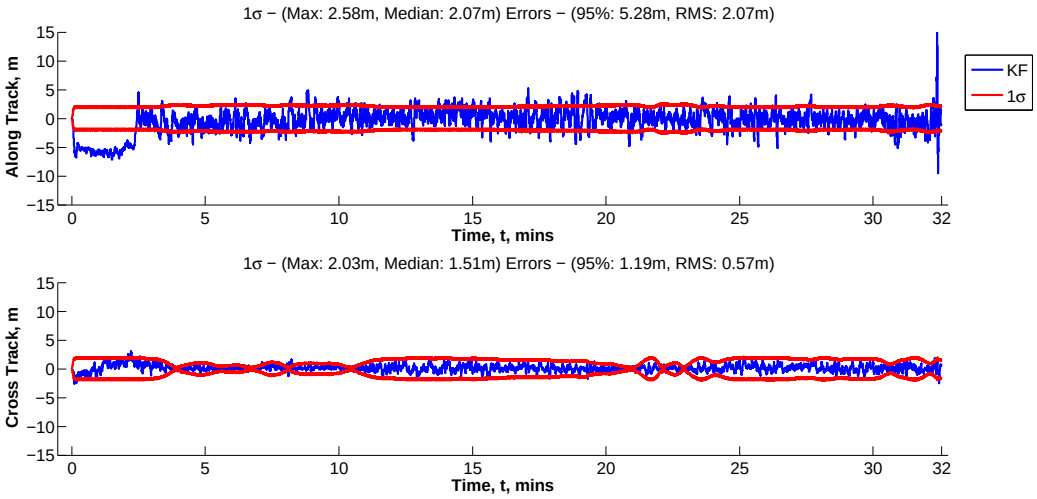



**(a)** Position errors

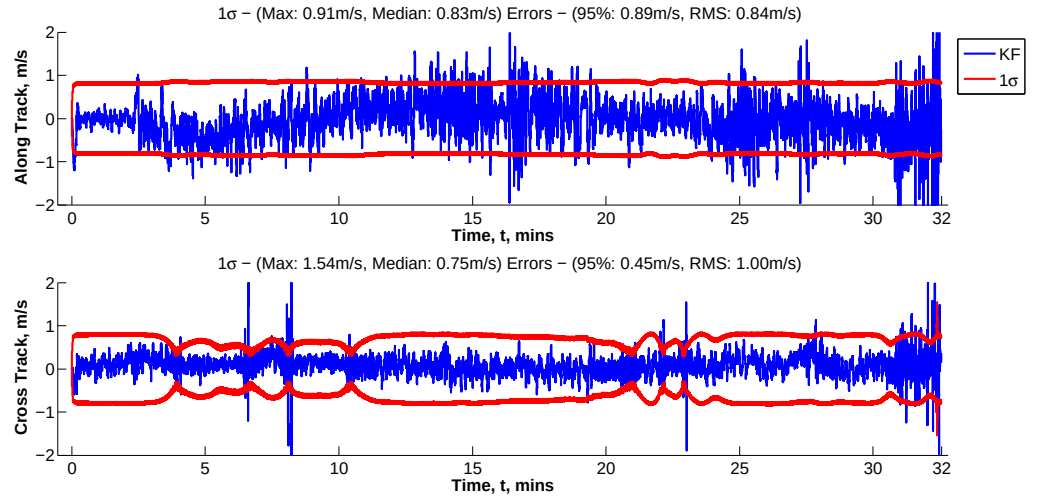



**(b)** Velocity errors

**Figure D.7.:** General performance of a system using a DRM with 10 m point intervals and 0.1 m point errors




**(a)** Position errors




**(b)** Velocity errors

**Figure D.8.:** Along track and cross track performance of a system using a DRM with 10 m point intervals and 0.1 m point errors