Contributions to the foundations of a safety case for the use of GNSS in railway environments

Andrew Alexander Stolagiewicz

University College London

March 2009

I, Andrew Alexander Stolagiewicz, 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.

Andrew Stolagiewicz

Abstract

The use of GNSS in the railways for passenger information services and selective door opening is already commonplace but the advancement of this increasingly popular navigation technique into safety of life rail applications has been hindered by the unknown level of measurement error caused by the local rail environment, especially that due to multipath.

Current state of the art receiver technologies are discussed along with the additional advantages of signal differencing using local base stations. Limiting factors for hardware in a kinematic environment are also discussed and specific examples to the rail environment highlighted.

Safety critical analysis techniques such as FMEA, HAZOP and FTA are reviewed to illustrate the evaluation of safety integrity values and the possibility of system risk, leading to the formation of a structured safety case.

Three main data sets from electrified, rural and urban rail environments have been collected using dual frequency geodetic receivers in order to enable analysis of multipath effects in normal railway operations. The code and phase data have been combined to compute fluctuations in multipath errors and these have been used to characterise this effect in both space and time. Where phase positioning is possible comparisons with standard code-based positions have been made to assess the overall quality of the type of GNSS positioning expected to be operationally-viable on the railways. Experiments have also been undertaken to evaluate the possible effects of electromagnetic radiation from overhead cables used to power the trains.

Finally, the ways in which the results of these experiments can be used to help build a safety case for the use of GNSS on the railways are discussed. Overall it is concluded that it is unlikely that multipath errors or electromagnetic interference will be the limiting factors in utilising GNSS for safety-critical railway applications.

Table of Contents

List of Figures

List of Tables

Acronyms and Definitions

ALARP	As Low As Reasonably Practicable
ARP	Antenna Reference Point
AS	Antispoofing
ASK	Amplitude Shift Keying
Balise	Track located position transponder
BPSK	Binary Phase Shift Keying
BR	British Rail
BS	British Standard
C/A-code	Coarse/Acquisition code (1.023 MHz)
CDMA	Code Division Multiple Access
CMD	Cold Movement Detector
DGPS	Differential GPS
DoD	Department of Defence
DOP	Dilution of Precision
DR	Dead Reckoning
ECEF	Earth Centred Earth Fixed
ERTMS	European Railway Traffic Management System
ETA	Event Tree Analysis
ETCS	European Train Control System
FAA	Federal Aviation Authority
FBSR	Feedback Shift Register
FMEA	Failure Mode and Effects Analysis
FMECA	Failure Mode, Effects and Criticality Analysis
FSK	Frequency Shift Keying
FTA	Fault Tree Assessment
GALILEO	European Global Navigation Satellite System
GDOP	Geometric DOP

GLONASS	Global'naya Navigatsionnaya Sputnikkovaya Sistema
GMST	Greenwich Mean Sidereal Time
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
HAZOP	Hazard and Operability Analysis
HDOP	Height DOP
HOW	Hand Over Word
IGS	International GPS Service
INS	Inertial Sensor
ION	Institute Of Navigation
ITRF	International Terrestrial Reference Frame
JD	Julian Date
JPL	Jet Propulsion Laboratory
L1	L1 carrier (1575.42 MHz)
L2	L2 carrier (1227.6 MHz)
LAMBDA	Least-squares Ambiguity Decorrelation Adjustment
LHCP	Left Hand Circularly Polarised
M_{P1}	Code multipath on L1 signal
M_{P2}	Code multipath on L2 signal
P-Code	Precision Code (10.23 MHz)
PDOP	Positional DOP
PPP	Precise Point Positioning
PRN	Pseudorandom Noise
PSK	Phase-shift Keying
QPSK	Quadri-phase-shift Keying
RAIM	Receiver Autonomous Integrity Monitoring
RAMS	Reliability, Availability, Maintainability and Safety
RHCP	Right Hand Circularly Polarised

RINEX	Receiver Independent Exchange Format
RNP	Required Navigation Performance
RSSB	Rail Safety and Standards Board
RTK	Real Time Kinematic Positioning
SA	Selective Availability
SIL	Safety Integrity Level
SP3	Standard Product #3 for ECEF Orbital Files
SVN	Space Vehicle Number
SWD	Slant Wet Delay
TA	Train Awakening
TDOP	Time DOP
TEC	Total Electron Content
TECU	TEC Unit
TOW	Time of Week
TTA	Time to Arrival
UTC	Coordinate Universal Time
WAAS	Wide Area Augmentation System
WADGPS	Wide Area DGPS
WGS84	World Geodetic System 1984
Y-code	Encrypted P-code
ZHD	Zenith Hydrostatic Delay
ZWD	Zenith Wet Delay

Acknowledgements

I would like to thank Professor Paul Cross for his endless support throughout my work. I thank him for believing in my abilities and always providing a sound and positive outlook on my research even in times of crisis, without which I would have been unable to finish.

Thanks should also go to Professor Marek Ziebart for his personal and academic support during my research.
My thanks are extended to the other members of departmental staff who have all helped in their own way to make my time at UCL the most enjoyable and fulfilling of all my time in academia.

I would like to say a special thank you to my fellow researchers for all your help and putting up with my abuse (especially Tip and Joe), you have been the greatest colleagues a person could ask for!

A very big thank you to Martyn Thomas of the RSSB for his continual support and for making my industry linked experience so enjoyable. Without Martyn I would not have been able to obtain any data or understand the complexities of the UK rail system.

Personally, I would like to thank my friends and family for their unconditional understanding and support which I value more than anything in the world.

1 Introduction

1.1 Background

The current UK rail network is owned by several private companies following the breakup of British Rail in the 1993 railways act.
Because of this fragmented system, standardisation of the rail infrastructure has meant improvements and alterations to tried and tested methods are time consuming and costly. Despite this, there are constant changes to the rail network operational systems, the most major current change being the unification of all European rail systems through the implementation of the European Rail Traffic Management System (ERTMS) designed to allow trains from all EU countries to travel across country borders without issue.
The current methods used to locate trains relies upon track circuits which leads to a large redundancy in track occupation as safety 'buffer' zones have to be left between trains to avoid collisions. The ERTMS system has three levels, the most sophisticated of which (level 3) uses track mounted beacons (called Balises) to tell the train its location. These beacons together with the required under-carriage reader are costly and a very large number would be required in order to have continuous positioning accuracy along a line, so an alternate method of train location is required.

Global Navigation Satellite Systems (GNSS) offer the opportunity to position a user anywhere in the world to a very high accuracy. In this thesis the US Department of Defence (DoD) run Global Positioning System (GPS) will be concentrated on as it is the only viable option for reliable global positioning due to it currently being the only complete system. The methods used and results obtained in this work are however highly applicable to the other GNSS systems as they operate in the same frequency band with similar signal structure.
By using GPS as a location tool, physical additions to both train and track can be avoided, thus saving cost and inconvenience as line side equipment is difficult to maintain and is often subject to vandalism and harsh conditions. The current 'fixed block' train positioning system requiring large amounts of redundant track can be

switched to the ideal ‘moving block’ system where train separations are merely based on breaking distances, increasing track density and thus revenue.
There are however very stringent guidelines for any new technology in the railways, stating that it can only be applied if the same pre-existing level of safety is maintained, or improved upon.
There are levels of Required Navigation Performance (RNP) which define the accuracy (position compared to true value), integrity (ability of the system to inform the user of an error), continuity (the probability of maintaining navigational guidance) and availability (percentage of time the system is within limits of required performance) of a system so that when in design, the minimum system specification is known.

RNP’s were originally introduced when the International Civil Aviation Organisation (ICAO) Air Navigation Commission requested that the All Weather Operations Panel (AWOP) examine the possibility of extending the RNP concept, which was originally intended for en-route operations, to include approach, landing and departure operations (Filip, Polivka, & Suchanek, 2006).
The use of RNP’s have subsequently been adopted in the railways so that safety critical analysis of positioning systems can be carried out without confusion as to the levels being tested for.
Alcouffe and Barbu (Barbu & Alcouffe, 1999) defined values, but these were designed to be altered depending on the application and so finalised RNP’s for the UK railways are as yet undefined and only recommendations exist.
This is where GPS is an unknown as the accuracy and integrity of GPS positioning in the rail environment have not yet been fully tested and so has yet to be determined as a viable solution.
There have been numerous European projects, such as the Low Cost satellite based train location system for signalling and train Protection for Low-density traffic railway lines (LOCOLOC/LOCOPROL) (Lancien, 2005) that stated multipath (direct and reflected signals entering the receiver causing an error) as an unknown level of error, the Advanced Position Locator system (APOLO) (Barbu & Alcouffe, 1999) and the Galileo Demonstrator for Railway Operation System (GADEROS) (Ulrech, 2002). Both of which meant to define and test prototype equipment for a satellite based train positioning system, but did so without investigating the levels of multipath or signal

error caused by the lineside infrastructure and so multipath was an undefined errors source that needed to be quantified for the rail environment.

1.2 Objectives and Scope

The public applications of GPS in the rail environment to date have been for non safety critical or very low safety level systems allowing their operational integrity to be lower than that of a 'safety of life' system.
The objective of this research is to try and determine the level of GPS signal error in the rail environment so that a safety case can be produced for its use as a positioning system.
Through a range of experiments, the possibility of electro-magnetic (EM) interference from overhead power lines is investigated as well as degradation of GPS signals due to multipath in rural and densely urban rail environments from line side infrastructure and surrounding buildings.
Multipath is investigated using the L1/L2 code multipath observable and its effect on the final position solution is qualitatively assessed.
From this investigation, the results and conclusions are designed to help in the forming of a safety case for the use of GNSS as a positioning device in the UK rail network.

1.3 Research Limitations

This research considers the GPS constellation only, additional systems such as GLONASS and the future GALILEO will help to improve the accuracy and integrity of the final solution, but it is not within the scope of this research to evaluate these GNSS systems.
The UK rail safety culture and methods of analysis are used as a tool for evaluation of the systems discussed; it is not the aim of this research to alter pre-existing methods for rail system safety level determination.
The integrity of the GPS constellation is not part of this research, though this would of course be a contributing factor to any final system's safety qualification it was not possible to cover over the course of this research project.
It is not the purpose of this research to produce a full safety case, merely to investigate the possible error sources due to the rail environment and their possible effects.

1.4 Thesis Outline

Chapter 1: Describes the background and limitations of the research carried out as well as the contributions to the area of GNSS research.

Chapter 2: Background of key research sections, specifically GNSS signals and their interaction with given environments. The advantages of real time kinematic (RTK) networking and the underlying principles are discussed as a useful system for high accuracy real time positioning.
The safety culture in the UK rail network is described, leading to an explanation of current uses of GPS in the railways and the error sources encountered due specifically to the rail environment.

Chapter 3: Risk and Safety analysis in the railways is detailed, Hazard analysis and system safety failure modes are discussed together with other current safety integrity level determination techniques. The quantitative and qualitative aspects of the rail safety culture are also outlined.

Chapter 4: The departmental RTK software suite is presented and its modular format is discussed. The research focused library and alternate algorithm design which allows full user customisation is outlined and demonstrated with results validation.

Chapter 5: Three data collection exercises are designed and recorded so that different rail environments can be investigated for GPS signal interference. The first uses a private train line in the Severn Valley to provide a test data set, helping to prepare for larger data sets later in the project.
The second experiment is to detect any possible electromagnetic interference (EMI) due to overhead power lines on the electrified east cost main line.
The third exercise makes use of GPS data from a full location package consisting of a dual frequency GPS receiver, accelerometer, gyroscope and tachometer mounted on

an electric suburban train travelling through the urban Birmingham New street area, though only the GPS data will be used in this research.

Chapter 6: Data from the three experiments is processed and the results are presented. Geometric comparison is shown to be non ideal and code multipath observable techniques are used to determine the effects of the local environment on GPS signals.
Data snooping methods are devised to locate signal breaks so that possible EM effects can be isolated and recorded.

Chapter 7: Summation of the research carried out, leading to conclusions and possible future work. Limiting factors of the experimental results and possible improvements are also discussed.

1.5 Research Contributions

- Understanding of rail safety analysis techniques to help understand the limitations of GPS in a safety of life application so that further concept development can take place.
- Analysis of GPS data in the rail environment providing insight into the various error sources present and their expected levels.
- Electromagnetic interference from overhead cables is shown not to be an error source for GPs signals.
- A research toolkit has been developed by the department with contributions from this work for easy network based RTK evaluation.

2 Background of Research

2.1 Introduction

To understand the issues surrounding the application of GPS in the rail environment, the underlying principles and systems must be fully understood before any further research can be understood.
Section 2.2 explains the basics of the GPS system and goes on to detail the GPS signal in space (SIS) and the multiple signals that are modulated onto the main carrier frequencies. Section 2.3 provides the reasoning for and method used in RTK processing of GPS data. Section 2.4 discusses the Safety culture used in the UK railways and how this effects the changes and improvements made on a daily operational basis. Section 2.5 gives a brief overview of current applications of GNSS in the railways, specifically that of passenger information services and selective door opening. Section 2.6 outlines the possible error sources for GPS in the rail environment and focuses on that of possible EM interactions and the common error of multipath and its effects.

2.2 GPS signals, processing and atmospheric effects

2.2.1 Signal structures

GPS satellites all transmit on the same two carrier frequencies called L1 (1575.42 MHz) and L2 (1227.60 MHz), both of which are multiples (154 and 120 respectively) of an internally generated fundamental frequency of 10.23 MHz.
The signals are right hand circularly polarized (RHCP) in order to take advantage of the Faraday rotation effect and reduce ionospheric interference.
Modulated onto these carrier frequencies using binary phase shift keying (BPSK) are the individually assigned satellite pseudorandom noise (PRN) sequences. By using a spread spectrum of roughly 2MHz, code division multiple access (CDMA) allows the receiver to distinguish between multiple satellite signals and by using the PRN sequence it can identify which specific satellite it is receiving.

This PRN sequence makes up the coarse acquisition (C/A) code, which together with the precision (encrypted) (P(Y)) code and navigation message is modulated onto the L1 signal.
The C/A code has a wavelength of 300m and consists of a sequence of 1023 chips transmitted at a rate of 1.023Mcps. The P code consists of a very long (roughly 10^{14} chips) PRN code transmitted at a chip rate of 10.23Mcps and wavelength of 30m, allowing for comparatively high precision range measurements to be taken.
Since 1994 the P code has been called the Y code due to the encryption designed to limit its availability to authorised users only.
The Navigation message is 37500 chips long and contains the full constellation health, clock parameters and satellite ephemeris allowing the receiver to calculate precise satellite positions. It is transmitted at 50cps and it takes 12.5 minutes to receive the full message before it is repeated.

2.2.2 GPS receiver design

GPS receivers generally consist of an RF front end that feeds into a digital signal processing section which extracts the navigation data for use in the final data processing module (Ray, 2000)

Receivers use right hand circularly polarized antennae to match the GPS signal for maximum gain. These antennae can be actively amplified to reduce overall noise, non-target frequencies and the signal reduction due to coaxial cable attenuation.
The front end also has a frequency synthesizer driven by local oscillators which provide clock frequencies used to sample the GPS signal for analogue to digital conversion. These clock frequencies are then passed on to the digital signal processing section of the receiver.
Throughout the RF processing, automatic gain control keeps the signal level at the required level whilst minimising the amount of noise introduced to the system.

The digital signal processing uses multiple parallel channels to dynamically track individual satellites based on their PRN number.
Each channel has a code and carrier tracking loop that work together with internally generated versions of the received signals which are modulated in-phase and

quadrature-phase with the input signal. These combined signals are then correlated with early, late and prompt versions of internally generated code.
The early and late correlation values (traditionally 1 chip apart, though recently narrow spacing of 0.1chips have become popular for multipath mitigation) tend to be used for code tracking and the prompt correlation values for carrier tracking.
The prompt signal strips the C/A code from the input signal and outputs a continuous wave with phase transition caused only by the modulated navigation data. This feeds the carrier loop which outputs a continuous wave with the same frequency as the carrier input. This is then used to strip the carrier from the input signal in order to leave the C/A code, which then enters the code tracking loop (Wu & Zhang, 2005).
The discriminator algorithm gives a calculated number on the code phase error, with the discriminator output signal, ε, calculated by:

$$\varepsilon = \frac{y_E}{y_L} \qquad (2.1)$$

Where y_E and y_L are the early and late code correlation powers respectively.

If the value of ε is equal to 1, the prompt code is aligned with the input C/A code.
If $\varepsilon > 1$ the codes should be shifted to the right and if $\varepsilon < 1$, to the left.
Once the code lock has been achieved, the carrier phase lock loop attempts to match the locally generated carrier phase to the incoming signal. Generally, receivers use a Costas loop as the effect of a 180° shift due to data bit transitions is ignored.
The in-phase and quadrature-phase prompt correlation values are used to compute the carrier phase offset between the local and input signals.
In carrier tracking, the integer number of carrier cycles are determined so that range measurements can be calculated and high accuracy position calculations can occur.
If the signal is interrupted in some way, this can cause a ‘cycle slip’ where the tracking process must begin again as an unknown integer number of cycles may have been unaccounted for, causing an ‘integer ambiguity’.

2.2.3 Ionospheric effect

The ionosphere is located 50 to1000 km above the surface of the earth and is known to be a dispersive medium that can effect GPS range measurements by tens of meters.

The effects are due to electrons in the atmosphere interacting with the L1 and L2 signals, delaying the pseudoranges and advancing the phase. The severity of the effect is linked to the frequency of the signal passing through as well as the total electron content (TEC) of the signal path and so satellites observed at a low elevation angle to the horizon will have their signals affected to a greater extent due to their ionospheric path lengths being greater.

Unfortunately, the TEC can vary significantly as shown in figure 2.1 depending on the local time, magnetic latitude and how close to the solar maximum it is as sunspots and solar activity can affect the electron density in the atmospheric plasma considerably. The TEC can also fluctuate rapidly and with limited predictability so removing the ionospheric effect can be very difficult.

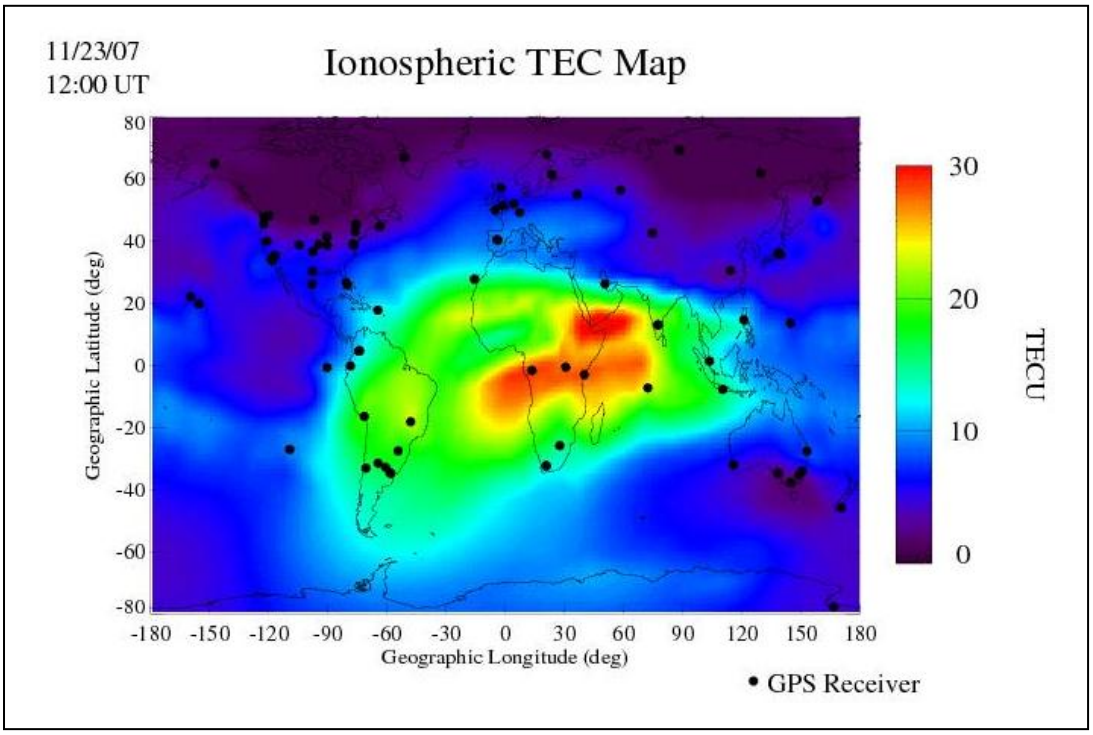


Figure 2.1 - Global TEC map showing GPS base station locations (JPL, 2007)

By utilising the frequency dependence of the ionospheric effect, the range delay error on L1 or L2 can be removed through a simple linear combination of the L1 and L2 frequencies, for L1:

$$d_{ion,L1} = \frac{f_2^2}{f_1^2 - f_2^2}(P_1 - P_2) \qquad (2.2)$$

Where $d_{ion,\,L1}$ is the range delay on L1 due to the ionosphere
f_1, f_2 are the frequencies of L1 and L2
P_1, P_2 are L1 and L2 range measurements

If only single frequency GPS data is available, 8 coefficients have been included in the navigation message so that at least a 50% reduction in ionospheric error can be achieved (Klobuchar, 1987).

2.2.4 Tropospheric effect

The troposphere ranges from the earths surface to an average of 11km with it thinning to 8km at the poles and thickening to 16km at the equator, the troposphere contains roughly 80% of the atmospheric mass and is made up of water vapour and gases. This vertical range of the troposphere is where the 'wet' part of the model is contained. The stratosphere stretches from the top of the troposphere, above the tropopause layer to about 50km in height. The 'dry' component of the tropospheric model stretches from ground level to roughly 40km and despite it occupying the stratosphere as well as the troposphere, most of its effect is from the lower region inside the tropospheric layer. Because the tropospheric effect on GPS signals is non dispersive, the dual frequency mitigation approach used for the ionosphere cannot be used as the effects are frequency independent.

The amount of path delay in the zenith direction due to the troposphere is in the range of 2m, increasing with zenith angle to several meters, so there needs to be a tropospheric model included when attempting precision position computations.

The wet component of the tropospheric delay accounts for up to 10% of the total, with the dry component causing the remaining 90% error.

The modified Saastamoinen tropospheric model (Saastamoinen, 1972) for calculating trop path delay requires variables such as the temperature, local atmospheric pressure, partial pressure of water vapour and height of the receiver location. The zenith angle to the satellite in question is also required and a mapping function is used to relate the combined wet and dry zenith delay to the received signal path delay.

$$Tro = \frac{0.002277}{\cos z}\left[P + \left(\frac{1255}{T} + 0.05\right)e - B\tan^2 z\right] + \partial R \qquad (2.3)$$

Where z is the zenith angle of the satellite

P is the atmospheric pressure in millibars

T is the temperature of the receiver in Kelvin

e is the partial pressure of water vapour in millibars

B and δR are correction terms dependent on receiver height and z and can be found on page 53 of Xu (Xu, 2003)

$$e = R_h \exp(-37.2465 + 0.213166T - 0.000256908T^2) \quad (2.4)$$

Where R_h is relative humidity in %

For simplicity standardised values for pressure, temperature and humidity can be calculated from the following equations:

$$P = P_0[1 - 0.000226(H - H_0)]^{5.225} \quad (2.5)$$

$$T = T_0 - 0.0065(H - H_0) \quad (2.6)$$

$$R_h = R_{h0} \exp[-0.0006396(H - H_0)] \quad (2.7)$$

Where P_0 is standard pressure ≈ 1013.25 millibars

T_0 is standard temperature ≈ 18° Celsius

R_{h0} is standard relative humidity ≈ 50%

H_0 is reference height ≈ 0m

There are alternative models such as the modified Hopfield model (Hopfield, 1963), Niell model (Niell, 1996) and Yionoulis model (Yionoulis, 1970), some of which separate the wet and dry components of the total zenith tropospheric delay, but all produce similar results.

2.3 Real Time Kinematic Relative positioning

2.3.1 Point positioning

For train signalling and control a positioning system needs to be able to produce real time positions within specified accuracy and integrity levels.
Using GPS, point positioning can be carried out using pseudorange C/A measurements to four or more satellites in order to provide a preliminary position solution using least squares (Kaplan, 2005).

$$\rho_j = \sqrt{(x_j - x_u)^2 + (y_j - y_u)^2 + (z_j - x_u)^2} + ct_u \qquad (2.8)$$

Where ρ_j is a pseudorange to satellite j

x_j, y_j and z_j are the x, y and z coordinates of satellite j

x_u, y_x and z_u are the x, y and z coordinates of the user

c is the speed of light

t_u is the user clock offset

From this equation, the user position is assumed to be a combination of an approximate user position and user clock offset x'_u, y'_u, z'_u, t'_u and incremental components $\Delta x_u, \Delta y_u, \Delta z_u, \Delta t_u$.

$$\begin{aligned} x_u &= x'_u + \Delta x_u \\ y_u &= y'_u + \Delta y_u \\ z_u &= z'_u + \Delta z_u \\ t_u &= t'_u + \Delta t_u \end{aligned} \qquad (2.9)$$

This allows us to write

$$f(x_u, y_u, z_u, t_u) = f(x'_u + \Delta x_u, y'_u + \Delta y_u, z'_u + \Delta z_u, t'_u + \Delta t_u) \qquad (2.10)$$

By performing an expansion about the approximate position and clock offset, a first order partial derivative Taylor series can be evaluated and substituted into equation 2.8 to produce 2.11:

$$\rho'_j - \rho_j = \Delta\rho_j = -\frac{x_j - x'_u}{r'_j}\Delta x_u - \frac{y_j - y'_u}{r'_j}\Delta y_u - \frac{z_j - z'_u}{r'_j}\Delta z_u - ct_u \qquad (2.11)$$

Where ρ'_j is an approximate pseudorange to satellite j

$\Delta\rho_j$ is the incremental change to the approximate pseudorange to satellite j

$$r'_j = \sqrt{(x_j - x'_u)^2 + (y_j - y'_u)^2 + (z_j - z'_u)^2}$$

By simplifying equation 2.10 with directional cosines a_{xj}, a_{yj} and a_{zj} ,it can be shown

$$\Delta\rho_j = a_{xj}\Delta x_u + a_{yj}\Delta y_u + a_{zj}\Delta z_u - c\Delta t_u \qquad (2.12)$$

Which can be put in matrix form due to the need for 4 satellites as there are 4 unknowns. This can then be rearranged to form

$$\mathbf{\Delta x} = \mathbf{H^{-1}}\Delta\rho \qquad (2.13)$$

Where $\Delta\mathbf{\rho} = \begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \Delta\rho_3 \\ \Delta\rho_4 \end{bmatrix}$ $\mathbf{H} = \begin{bmatrix} a_{x1} & a_{y1} & a_{z1} & 1 \\ a_{x2} & a_{y2} & a_{z2} & 1 \\ a_{x3} & a_{y3} & a_{z3} & 1 \\ a_{x4} & a_{y4} & a_{z4} & 1 \end{bmatrix}$ $\Delta\mathbf{x} = \begin{bmatrix} \Delta x_u \\ \Delta y_u \\ \Delta z_u \\ -c\Delta t_u \end{bmatrix}$

Equation 2.12 can be iterated until the position converges to a best estimate. This is ideal when there are 4 satellites, but if there are more than 4, equation 2.13 cannot work as Δρ and H are now:

$$\Delta\mathbf{\rho} = \begin{bmatrix} \Delta\rho_1 \\ \Delta\rho_2 \\ \vdots \\ \Delta\rho_n \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} a_{x1} & a_{y1} & a_{z1} & 1 \\ a_{x2} & a_{y2} & a_{z2} & 1 \\ \vdots & \vdots & \vdots & 1 \\ a_{xn} & a_{yn} & a_{zn} & 1 \end{bmatrix}$$

Where n is the number of satellites being observed (n≥4).

To overcome this, ordinary least squares is used to overcome a system with more equations than unknowns.

From 2.13 we require a value of $\mathbf{\Delta x}$ so that $\mathbf{H\Delta x}$ is close to $\mathbf{\Delta\rho}$. By forming a residual vector $\mathbf{r}$, so that $\mathbf{r} = \mathbf{H\Delta x} - \mathbf{\Delta\rho}$ and solving for the minimum value of its square ($R_{SE}(\mathbf{\Delta x})$), provided $\mathbf{H}^T\mathbf{H}$ is non-singular, $\mathbf{\Delta x}$ can be solved for as follows:
Given that:

$$R_{SE}(\mathbf{\Delta x}) = (\mathbf{H\Delta x} - \mathbf{\Delta\rho})^2 \qquad (2.14)$$

And:

$$(\mathbf{H\Delta x} - \mathbf{\Delta\rho})^2 = (\mathbf{H\Delta x} - \mathbf{\Delta\rho})^T(\mathbf{H\Delta x} - \mathbf{\Delta\rho}) \qquad (2.15)$$

Explaining 2.14:

$$R_{SE} = (\mathbf{\Delta x})^T\mathbf{H}^T\mathbf{H\Delta x} - 2(\mathbf{\Delta x})^T\mathbf{H}^T\mathbf{\Delta\rho} + (\mathbf{\Delta\rho})^2 \qquad (2.16)$$

By differentiating R_{SE}, its gradient is calculated as:

$$\nabla R_{SE} = 2\mathbf{H}^T\mathbf{H\Delta x} - 2\mathbf{H}^T\mathbf{\Delta\rho} \qquad (2.17)$$

As long as $\mathbf{H}^T\mathbf{H}$ is non-singular (all the columns of $\mathbf{H}$ are independent) we solve for $\mathbf{\Delta x}$ by equation this to 0 and obtain:

$$\mathbf{\Delta x} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{\Delta\rho} \qquad (2.18)$$

This is the least squares formulae for $\mathbf{\Delta x}$ as a function of $\mathbf{\Delta\rho}$ where $(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is sometimes known as the least squares solution matrix.
For now we will assume that $\Delta\rho$ and Δx contain errors in the linear forms:

$$\mathbf{\Delta\rho} = \mathbf{\rho}_T - \mathbf{\rho}_L + d\mathbf{\rho} \qquad (2.19)$$

$$\mathbf{\Delta x} = \mathbf{x}_T - \mathbf{x}_L + d\mathbf{x} \qquad (2.20)$$

Where $\mathbf{\rho}_T$ and $\mathbf{x}_T$ are the error free vector of pseudorange values and error free position and time values respectively

$\mathbf{\rho}_L$ and $\mathbf{x}_L$ are the vector of pseudorange values and position and time values respectively calculated at the linearization point.

$d\mathbf{\rho}$ and $d\mathbf{x}$ are the net errors in the pseudorange values and position and time estimates respectively.

Substituting 2.19 and 2.20 into 2.18, we get:

$$d\mathbf{x} = [(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T]d\boldsymbol{\rho} = \mathbf{K}d\boldsymbol{\rho} \qquad (2.21)$$

Where the matrix **K** is defined in [] brackets.

Equation 2.21 gives the relationship between the errors in pseudorange values and the errors in position and time. The covariance of $d\mathbf{x}$ is obtained by:

$$\text{cov}(d\mathbf{x}) = E(d\mathbf{x}\, d\mathbf{x}^T) \qquad (2.22)$$

Where E represents the expectation operator.

Substituting this into 2.21 and assuming a fixed geometry we get:

$$\begin{aligned}\text{cov}(d\mathbf{x}) &= \text{E}(\mathbf{K}d\boldsymbol{\rho}\, d\boldsymbol{\rho}^T\mathbf{K}^T) = \text{E}[(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\, d\boldsymbol{\rho}\, d\boldsymbol{\rho}^T\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}] \\ &= (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\,\text{cov}(d\boldsymbol{\rho})\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1} \qquad (2.23)\end{aligned}$$

The assumption that the components of $d\boldsymbol{\rho}$ are identically distributed, independent and have a variance equal to the square of the satellite user equivalent range error (UERE), $\boldsymbol{\sigma}^2_{\text{UERE}}$, allow the covariance of $d\boldsymbol{\rho}$ to be shown as:

$$\text{cov}(d\boldsymbol{\rho}) = \text{I}_{\text{nxn}}\boldsymbol{\sigma}^2_{\text{UERE}} \qquad (2.24)$$

Where I_{nxn} is the n×n identity matrix.

Combining 2.24 and 2.23:

$$\text{cov}(d\mathbf{x}) = (\mathbf{H}^T\mathbf{H})^{-1}\boldsymbol{\sigma}^2_{\text{UERE}} \qquad (2.25)$$

The vector $d\mathbf{x}$ has four components, representing the errors in the vector $\mathbf{x}_T = (x_u, y_u, z_u, ct_b)$.

The covariance of $d\mathbf{x}$ is a 4x4 matrix:

$$\text{cov}(d\mathbf{x}) = \begin{bmatrix} \sigma_{x_u}^2 & \sigma_{x_u y_u}^2 & \sigma_{x_u z_u}^2 & \sigma_{x_u ct_b}^2 \\ \sigma_{x_u y_u}^2 & \sigma_{y_u}^2 & \sigma_{y_u z_u}^2 & \sigma_{y_u ct_b}^2 \\ \sigma_{x_u z_u}^2 & \sigma_{y_u z_u}^2 & \sigma_{z_u}^2 & \sigma_{z_u ct_b}^2 \\ \sigma_{x_u ct_b}^2 & \sigma_{y_u ct_b}^2 & \sigma_{z_u ct_b}^2 & \sigma_{ct_b}^2 \end{bmatrix} \qquad (2.26)$$

From 2.26, dilution of precision (DOP) values can be calculated as a ratio of the components of cov($d\mathbf{x}$) and $\boldsymbol{\sigma}_{\text{UERE}}$.
Geometric DOP (GDOP) which represents the effect of the standard deviation of the measurement errors on the solution is defined by:

$$\text{GDOP} = \frac{\sqrt{\sigma_{x_u}^2 + \sigma_{y_u}^2 + \sigma_{z_u}^2 + \sigma_{ct_b}^2}}{\sigma_{UERE}} \qquad (2.27)$$

This can also be defined from components of $(\mathbf{H}^{\text{T}}\mathbf{H})^{-1}$ where:

$$(\mathbf{H}^{\text{T}}\mathbf{H})^{-1} = \begin{bmatrix} D_{11} & D_{12} & D_{13} & D_{14} \\ D_{21} & D_{22} & D_{23} & D_{24} \\ D_{31} & D_{32} & D_{33} & D_{34} \\ D_{41} & D_{42} & D_{43} & D_{44} \end{bmatrix} \qquad (2.28)$$

Therefore:

$$\text{GDOP} = \sqrt{D_{11} + D_{22} + D_{33} + D_{44}} \qquad (2.29)$$

Other values for position, time, horizontal and vertical dilution of precision can also be calculated using these components to give a measure of the final solutions accuracy.

If certain pseudoranges are known to be more degraded than others, or a system requires that certain signals have their effect on the final solution downgraded and others upgraded, a weight matrix $\mathbf{W}$ can be used in 2.18 as shown:

$$\boldsymbol{\Delta x} = (\mathbf{H}^{\text{T}}\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^{\text{T}}\mathbf{W}\boldsymbol{\Delta\rho} \qquad (2.30)$$

This can, as an example, be used in a position solution where there are limited satellites and so every available satellite must be used but if one of these satellites has a low elevation angle, its signal will be more prone to errors and so the effect of this pseudorange on the position solution can be reduced using a weight matrix.

By introducing tropospheric and ionospheric models, pseudorange measurements can be improved, leading to a better position fix, but there are also other corrections, such as the inclusion of a residual satellite clock error, rotation of the earth and movement of the satellites during signal propagation known as the light time equation (Blewitt, 1997) which are essential if the solution is to be within kilometres of the true position. When these error sources are included in the position solution, point positioning can occur with a general accuracy of 10m in the horizontal.
The signal received can be broken into the various error sources as shown in equation 2.31.

$$P = \rho + \delta\rho + c(dt_r - dt_s + d_r + d_s) + I + T + s \qquad (2.31)$$

Where P is the measured code range
ρ is the actual geometric range between satellite and receiver
$\delta\rho$ is the orbital error
c is the speed of light
dt_s is the satellite clock error
dt_r is the receiver clock error
d_r is the receiver hardware delay
d_s is the satellite hardware delay
I is the ionospheric delay error
T is the tropospheric delay error
s is the receiver noise and local signal errors

If the L1 and L2 signals are used for phase measurements and L1 for pseudorange, provided the integer number of wavelengths have been resolved, precise point positioning can be carried out with a single static receiver to achieve centimetre level position accuracy.

The problem with this setup is, for a receiver to be able to minimise the tropospheric and ionospheric errors it has to occupy a known point for a considerable amount of time which is not possible when trying to locate a kinematic receiver.

2.3.2 Relative positioning

By considering the fact that tropospheric and ionospheric errors are spatially correlated and that orbit and satellite clock errors cancel when dealing with multiple receivers, relative positioning can be used to facilitate high accuracy kinematic positioning.

Using a reference receiver that permanently occupies a known point, the local signal corrections can be calculated. This is possible as the receiver knows its position to a high accuracy and thus knows the true geometric distance to the satellite in question (usually the highest elevation satellite in view to both receivers.

Provided the second moving (rover) receiver stays within a localised area, the effect of the atmosphere should delay the received signals the same amount as for the reference receiver.

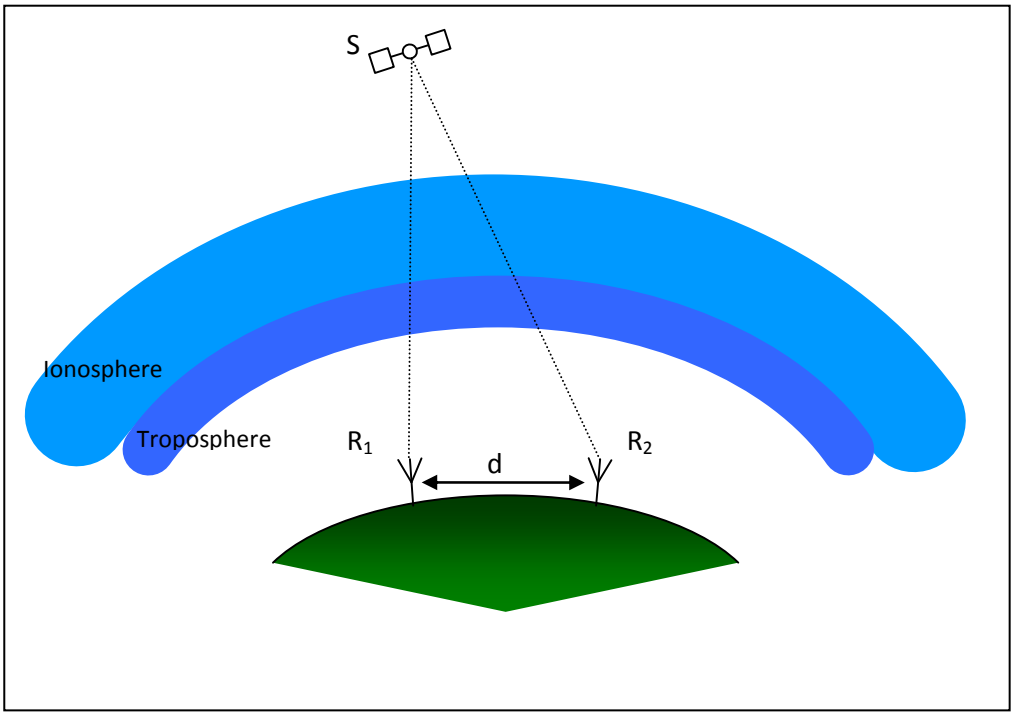


Figure 2.2 - Diagram of single differenced receivers

By single differencing, orbital errors are reduced as they are very similar for both receivers, depending on the baseline length and satellite clock errors are removed from the pseudorange equation as they are the same.

From equation 2.31 single differencing using receivers 1 and 2 gives equation 2.32 and 2.33 for phase and code measurements respectively.

$$\Phi_{12}^{s} = \rho_{12} + cdt_{r_{12}} + c\delta_{12} + \lambda\varphi_{12}(t_0) + \lambda N_{12}^{s} - I_{12} + T_{12} + s_{12} \qquad (2.32)$$

$$\mathrm{P}_{12}^{s} = \rho_{12} + cdt_{r_{12}} + cd_{12} + I_{12} + T_{12} + s_{12} \qquad (2.23)$$

Where Φ_{12}^{s} is the single differenced phase for receivers 1 and 2

P_{12}^{s} is the single differenced pseudorange for receivers 1 and 2

ρ_{12} is the single differenced true geometric range for receivers 1 and 2

$dt_{r_{12}}$ is the combined receiver clock offset

δ_{12} is the combined receiver carrier phase hardware delay

d_{12} is the combined receiver code hardware delay

$\lambda\varphi_{12}(t_0)$ is the between receiver non initial (t_0) zero phase in meters

N_{12}^{s} is the carrier phase ambiguity

I_{12} is the combined ionospheric error

T_{12} is the combined tropospheric error

S_{12} is the combined receiver noise

By observing two satellites simultaneously (double differencing – see figure 2.3) the receiver clock errors can be removed. Local environmental noise such as multipath cannot be eliminated as this is due to local reflectors and so is not linked to both receivers. By combining the observations as mentioned, noise is in fact increased.

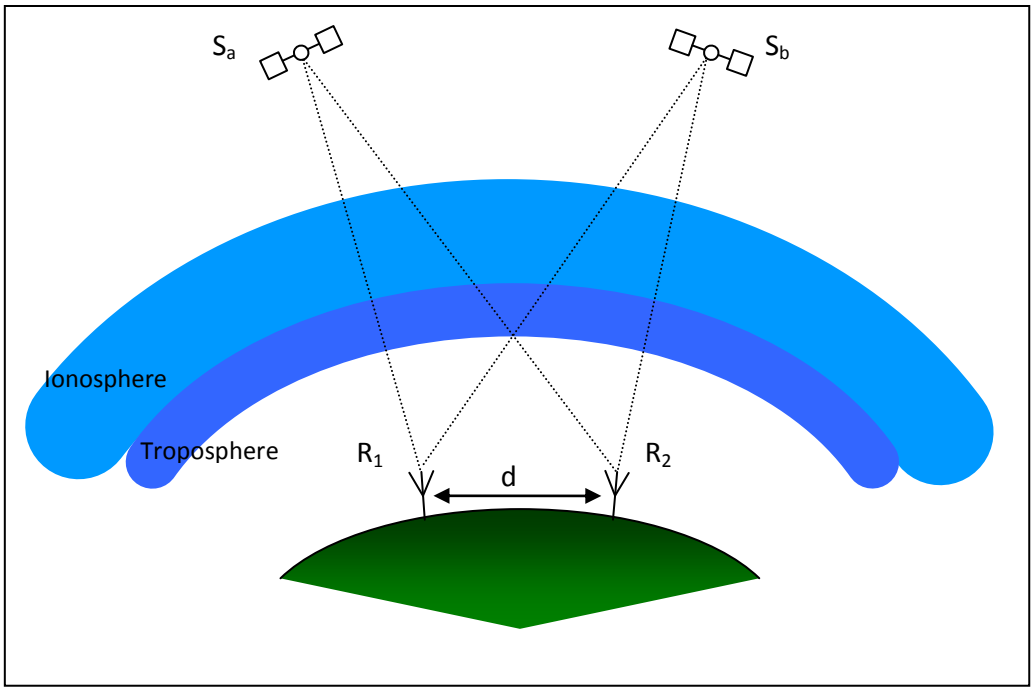


Figure 2.3 - Diagram of double differenced receivers

From equation 2.32 and 2.33, the addition of another satellite and receiver leads to equation 2.34 and 2.35 below.

$$\Phi_{12}^{ab} = \Phi_{1}^{ab} - \Phi_{2}^{ab} \qquad (2.34)$$

$$P_{12}^{ab} = P_{1}^{ab} - P_{2}^{ab} \qquad (2.35)$$

Where P_{12}^{ab} and Φ_{12}^{ab} are the double differenced pseudorange and phase for receivers 1 and 2 to satellites a and b

$P_{1}^{ab} = P_{1}^{a} - P_{1}^{b}$

$P_{2}^{ab} = P_{2}^{a} - P_{2}^{b}$

$\Phi_{1}^{ab} = \Phi_{1}^{a} - \Phi_{1}^{b}$

$\Phi_{2}^{ab} = \Phi_{2}^{a} - \Phi_{2}^{b}$

Even though the satellite clock offsets are removed as a possible error source in both single and double differencing, they must still be considered when calculating the positions of the satellites themselves.

The advantages of using these methods are such that if the two receivers are close enough, all the error sources can be eliminated provided there is no multipath or cycle slips (when considering phase). This is significant when trying to operate in a single

epoch mode in real time where all the eliminated error sources can degrade the final position solution.
These techniques also introduce their own problems, whereby the system noise is effectively doubled due to the measurement combinations and the constraint that the two or more receivers used in a baseline must be able to view the same two satellites. The addition of a real time system would also require a permanently operating radio link to be in place between rover and base station, something which in itself can fail, reducing the system to a single receiver point positioning solution with the accompanying reduced accuracy and integrity.

2.4 UK Rail safety culture

Due to privatisation of the railways, the UK government reviewed railway safety regulations and determined that the Health and Safety Executive (HSE) and the Health and Safety Commission (HSC) should regulate the entire rail network. In 2006 this responsibility was transferred to the Office of Rail Regulation (ORR) where it remains.
The laws that govern most rail safety practices are general safety laws that apply to all employers and employees in the UK, for instance; the health and safety at work act 1974 has produced the management of health and safety at work regulations 1992 and the construction (design and management) regulations 1994.
These are both non rail-specific regulations that place the responsibility of safety on the employer to keep their staff, passengers and anyone they interact with as safe as reasonable practicable.
The construction regulations also require those involved to keep records and share information with all those concerned in order to control risk.
The 1974 act also created the railways (safety critical work) regulations 1994 which ensures employers make sure their workers are properly trained in safe practice and the railways (safety case) regulations 1994 which states that train and station operators and railway infrastructure controllers must have a railway safety case before they begin work and that during the work it is followed.

From an engineering standpoint a defensive status has always been taken in respect to failures and possible risk. By designing all systems in a manner that facilitates a lower

level of risk in the event of failure, safety can be achieved even when the system does not operate as intended. The simplest example of this would be the automatic vacuum brake design in trains. Rather than the driver applying pressure to the braking surface with brake pads attached directly to brake pistons, the brake piston actually holds the brake pads away from the braking surface, then when the driver wishes to reduce speed, he simply allows the brake pads to apply pressure on the breaking surface by reducing the amount the pistons hold the pads away. In the event of the braking system failing, the brakes are automatically applied. This may cause a disruption to service, but is far safer than a train with no brakes at all. Unfortunately this system was only bought into service after 88 people were killed in the 1889 runaway carriage incident in Armagh. Previously a more conventional system of braking was used due to its simplicity and lower costing. Every alteration to the modern railways reflects this type of defensive engineering, including the way track junctions are set out and controlled, even though this can sometimes increase redundancy and reduce profits. After incidents such as the Hatfield rail crash on 17 October 2000 which exposed the shortcomings of the national railway infrastructure company Railtrack to ensure the sound condition of its assets, high risk industries have been trying to reduce the use of incident data for improving their safety principles and have moved towards the development of a 'positive safety culture' to manage risk before it occurs.
The HSE define a safety culture as (ACSNI, 1993):

"The safety culture of an organisation is the product of individual and group values, attitudes, perceptions, competencies and patterns of behaviour that determine the commitment to, and the style and proficiency of and organisation's health and safety management. Organisations with a positive safety culture are characterised by communications founded on mutual trust, by shared perceptions of the importance of safety and by confidence in the efficacy of preventative measures."

As well as promoting a positive safety culture, the ORR advises rail operators in good practice and has continued the 'blue book' of recommendations originally collated by the HSE. The blue book is designed to be a set of guidance notes and not specifically requirements for the rail industry as a whole.

The latest incarnation has been renamed ‘Railway safety principles and guidance’ and consists of two main parts, the first of which determines the top level principles (Figure 2.4) and the details required to make them work.

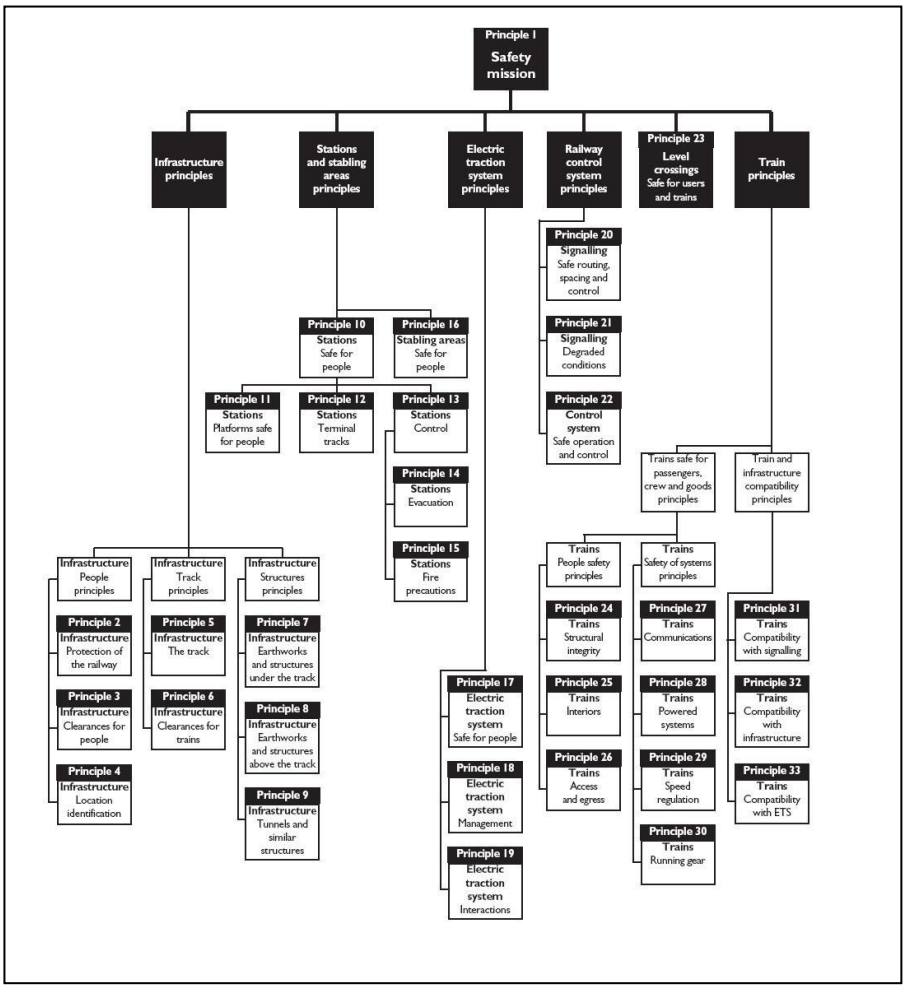


Figure 2.4 - Complete structure of safety principles **(Office of Rail Regulation, 2005)**

For some rail industry sections this provides enough guidance for them to then create their own standards but for others, part two of the guidance literature consists of seven separate publications focusing on specific aspects of railway construction and their associated good practices.

2.5 Applications of GNSS in the rail environment

There are many applications of GNSS in the rail industry with varying requirements on position accuracy and integrity.
Some systems are already making use of GPS signals for location purposes such as for passenger information services (PIS), freight management and automatic selective door opening.
PIS requires relatively low accuracy positioning ~100m as this simply informs passengers of times of arrival and general journey progress by using the train's current position to calculate an estimated travel time to the next station. This is a non 'safety of life' application that can fail without increasing risk and reducing safety, merely causing annoyance to passengers. If the PIS gives incorrect information however, this can cause issues with liability for costs incurred if a passenger were to alight at the incorrect station due to an incorrect PIS announcement. This is a clear distinction between total failure of a system and that of it operating incorrectly without knowingly doing so, as is the difference between reliability and integrity of a system.

Freight management systems use GPS locators combined with GSM communications to inform the user where a certain container or package is at any point in time. This can be especially important when dealing with valuable or hazardous goods that need to be closely monitored. Certain companies that offer this service also allow the user to be notified when the package in transportation comes within a pre-defined radius of its destination so the recipient can prepare accordingly, which is ideal if the package is large and requires a workforce to be prepared for its management. This type of system needs to have a higher level of integrity and reliability as its failure will cause the user to lose track of the package and render the entire system useless.
The failure would imply a financial loss to the company involved, but this is still considered a safety independent system as there is no danger to people's health in the event of the system failing.

Selective door opening (SDO) on Southeastern class 375 trains (Southeastern, 2005) was converted to use the on-board GPS units originally intended for PIS, for automatic opening of selected doors when entering a station without a full length

platform. Extending a platform is extremely costly and causes inconvenience for passengers as stations are closed.

The system works by encircling each station in a virtual zone that when entered, tells the main on-board computer which station the train is in and what the length of the platform is, thus which doors should be opened. The driver then stops the train in the correct position using traditional navigation techniques which then triggers the computer to open the correct doors as shown in figure 2.5 below.

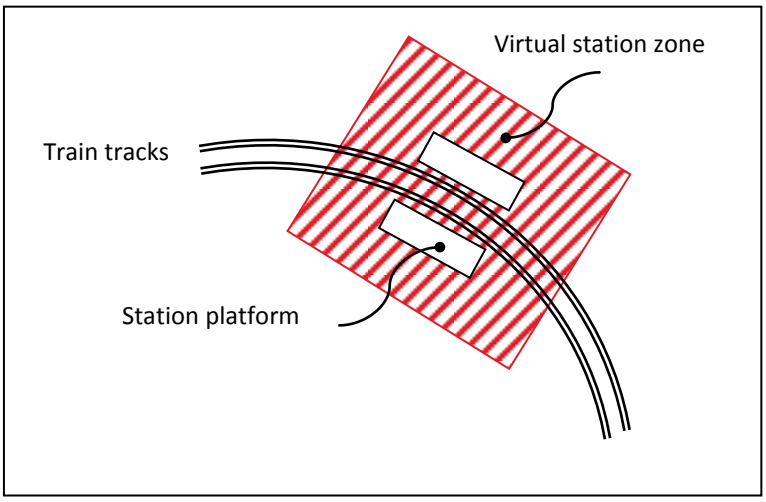


Figure 2.5– SDO virtual station zone

By using GPS to determine when the train is in the correct position, the safety of passengers is now in question as if the incorrect doors opened, the passengers could be exposed to extreme danger such as falling onto an electrified rail.

The initial problems involved with this system luckily only caused the doors to not open at all and after several updates to the software, the system performed as intended without incident. One of the main boundaries overcome was utilising hardware that was originally safety independent with no integrity values and adapting it for use in a system that could potentially cause risk to passengers. SDO by GPS is now used throughout the Southeastern fleet and has proved that if the hardware and software is robust enough, risk can be reduced to a level acceptable for daily use in the public domain.

The issues that arose during the development of the SDO system could have caused risk to passengers, but the possibility of a fatality was very low as, had the doors opened onto an electrified rail, the passengers would have a high chance of realising the error and would hopefully not fall out onto the track. SDO is currently the highest

safety critical application of GNSS in the UK railways and has provided automation using pre existing hardware. Despite this, multipath has still been flagged as being of an unknown error size that so far has simply been absorbed by the general error budget of the positioning system.

The next step in the use of GNSS in the UK railways is for train positioning and signalling which requires a much higher level of safety integrity and must be able to operate over the entire of the UK network. SDO only needs to operate in the vicinity of a train station, a controlled environment that tends to have a clearer view of the sky compared to when a train is travelling through deep cuttings and dense foliage between stations. The Alaskan Railroad Corporations Collision Avoidance System (ARRC-CAS) (Mokkapati & Pascoe, 2005) uses GNSS positioning to provide track authority limits and to enforce speed restrictions on the Fairbanks-Seward line in southern Alaska. They designed a system from the ground up, including the communications backbone and Automatic Train Protection (ATP) fail-safety architectures so that the existing Centralised Traffic Control (CTC) system could be optimised by automating many of the operator functions.

For GPS to replace the current track circuit based location system, the existing system must first be fully understood.
The UK system of train location does not use any on-board systems; the only train bourn part of navigation on a train is the driver who obeys the signals. The train is located using electronically isolated sections of track that carry a 10 volt DC signal. When the train is on a particular length of track (called a section) the axles act as an electrical contact between the rails, closing the circuit loop. This is then shown in the main control room as an ‘occupied section of track so the controller knows where each train is.

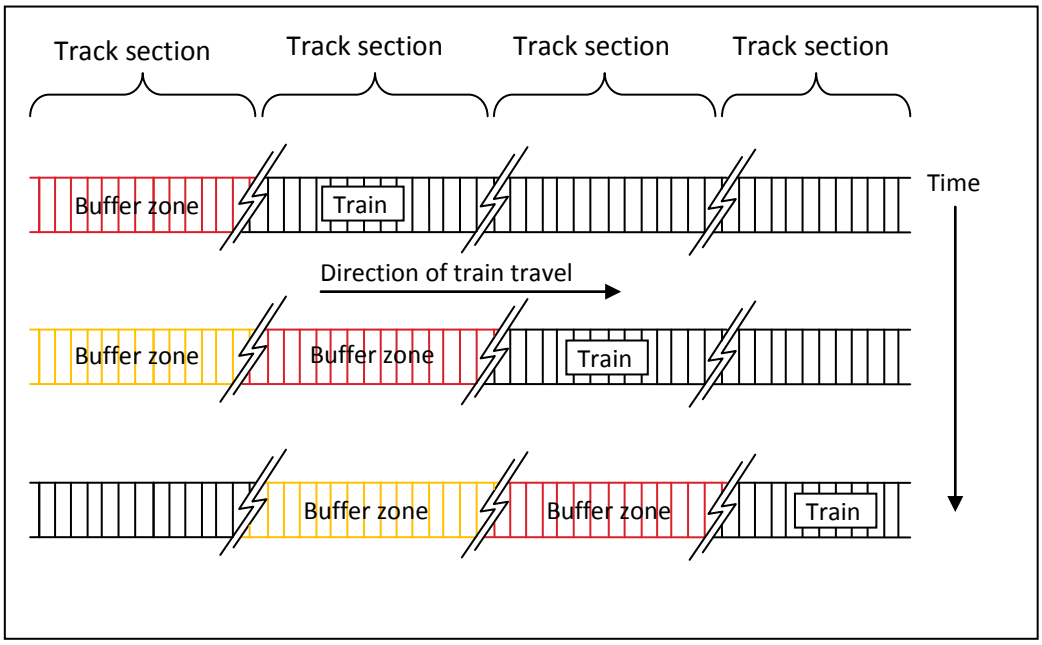


Figure 2.6 – Track sections and buffer zones for moving train

The problems occur when issues such as leaf fall cause an organic paste to be formed with the oil and water which causes conduction problems, leading to malfunction of the system.

Line side infrastructure like electrical cables and connections are also prone to vandalism and require constant maintenance and monitoring. These problems can become dangerous if not dealt with efficiently but it is more common for delays to occur than accidents.

Using track circuits is termed as being a ‘fixed block’ location technology as it uses fixed lengths of track place a train in a certain section. This reduces the possible train density of the line massively as there are usually two or more sections behind a train that are kept clear as a safety ‘buffer’ so that other following trains do not collide with it (see figure 2.6). This has to be done as the controller doesn’t know where inside the occupied section of track the train actually is, it could be at either end or in the middle. By leaving a buffer the controller can make sure there is no chance of two trains on the same line colliding. Two sections are used to allow the amber light system to be used. When entering a section of track, the train causes it to become occupied, the section of track following the train will have a red signal at its entrance and the section before that will have an amber signal at its entrance. This helps the train driver to know that a red signal is ahead and he can make speed adjustments accordingly. If he passes a red signal, this is known as a ‘signal passed at danger’ or

SPAD. This is seen as a safety violation and in each occurrence a full investigation is carried out as to why the driver could not stop before the signal.
Track sections can be tens of kilometres long and so knowing the actual position of a train can be very difficult which is why a location beacon system was brought about as part of the European rail traffic management system (ERTMS) and the European train control system (ETCS) which aims to standardise the methods of train location and signalling across Europe so that freight can pass through country borders without having to change trains. This would also remove the need for high speed trains to have multiple systems on-board which is costly and needlessly complex.
ETCS has 3 levels of operation that move from fixed block to moving block technology (see figure 2.8).

Level 1 is much the same as traditional track circuit location, but the use of beacons is employed. The beacons used are called 'Balises' and are placed in the centre of the track at known positions (as in Figure 2.7). When a train passes over, the electromagnetic induction powers the beacon enough to transmit a small signal with details on position and information on the track ahead (such as recommended speeds and possible hazards).

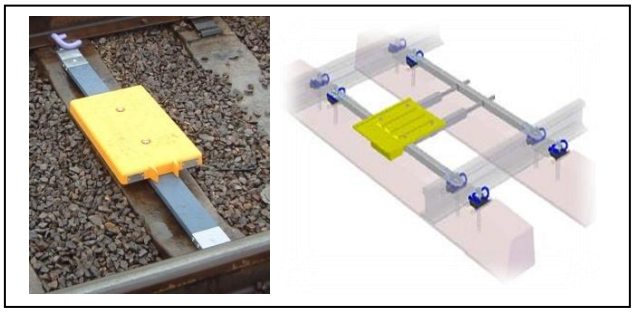


Figure 2.7– Photo and design image of a Balise

By combining this information with the track circuit positioning, the train is controlled by visual light signals at the line side as with traditional train control.
ETCS levels 2 and 3 rely heavily on the use of specific radio communications based on the mobile GSM telephony standard. GSM-R is used to:

- Exchange information between the main control centre and the train driver.
- Allow the drivers to speak with the traffic management centres.
- Transmit the maximum permitted speed from the control centre to the train driver.

By having a permanently operating communications channel between a train and the main control centre, the need for line side signalling is removed along with the possibility of its malfunction due to lack of maintenance or vandalism.
The issue of communication breakdown does however mean that GSM-R must have full coverage over the entire rail network as well as supporting full redundancy code and bit-checking in the signal protocol so that any distortion can be removed and possible jamming/spoofing can be avoided.

ETCS level 2 still uses track circuits for track occupancy detection, together with balise locators but the train operations are controlled through GSM-R, removing the lineside signals completely. This is already a reduction in the amount of lineside infrastructure required, but still uses a fixed block movement system with reduced freight capacity.
Level 3 removed the fixed block mode of track circuit detection, instead using train based odometry and balise locators. This information is then sent to the main control centre where the movement authority is determined and then transmitted back to the train via GSM-R.
This top level of operation provides moving block capability which can massively increase track occupancy levels without the increased risk.

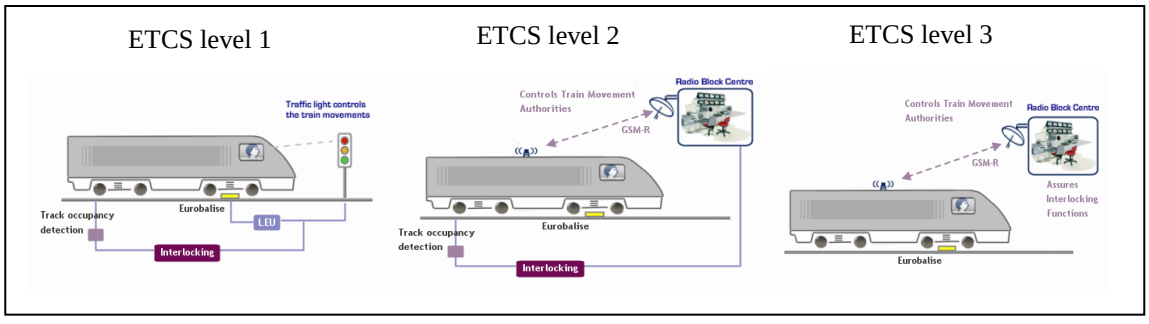


Figure 2.8– ETCS level 1, 2 and 3 **(UNIFE, 2007)**

The main problem with ETCS is the need for all trains to be fitted with balise readers which is very costly and takes the train being fitted out of service for an extended period of time. The cost of laying balises along the track is also considerable, coupled with the fact that the maximum accuracy achievable when using balise locators is limited by the distancing between each beacon as train odometry is notoriously poor at roughly 10% of the total distance travelled. For a reliable system to be in place, a level of redundancy also has to be included, which would require a fully working positioning system to be in operation even if one of the on track balises has malfunctioned, requiring double the number to be in place.

For these reasons, GNSS is being investigated as a viable alternative or addition to the use of balises, in a 'virtual balise' format (IERR, 2002). Because so much investment has already been made in the balise method of operation, rather than start again with an independent positioning system based on GNSS that would provide continuous location information, the use of a virtual balise has been chosen. The idea is that when the train passes a position on the track that would normally have had a physical balise mounted on the track; the on-board system would know this and produce a virtual readout using purely on-board location systems. Several projects have already used this approach when testing the possible use of GNSS in the railways, such as the European Space Agency (ESA) sponsored Railway User Navigation Equipment (RUNE) project where European Geostationary Navigation Overlay Service (EGNOS) signals were also used in an on board positioning system on the Torino-Chivasso line in 2005.
The Kayser-Threde GmbH sponsored INTEGRAIL (kayser-threde, 2004) project also focused on the use of EGNOS to produce a fully integrated ERTMS compliant railway traffic management and control system for Europe.

2.6 Error sources due to rail environment

The UK line side formations vary massively between urban and rural environments with deep cuttings and urban canyons, trees and foliage. GNSS signal obscuration and interference by physical surroundings as well at the train itself is of great concern. Rail specific interference from overhead cables and electrical engine noise is also possible.

These error sources are as yet unknown and so are of interest and are of great importance as large scale spatially correlated errors such as ionospheric and tropospheric errors can be largely removed by using network solutions. Other interference such as clock variations and receiver noise can be modelled to a certain extent, leaving the local non spatially correlated environment as the last significant source of error as differential processing is not effective.

Due to the physical surroundings in the rail environment, there are many objects that can effect GNSS signals. Without knowing the exact geometric makeup of the surroundings and satellites it is very hard to determine the possible effects of this interference (Bradbury, 2007).

As shown by A.Parkins (Parkins, 2006) the availability of satellites whilst in a rail cutting is dependent on the side wall elevation angle but independent of the cutting azimuth which amplifies the complexity and individuality of each train lines signal interference. This is a concern when analysing the feasibility of GNSS as a location tool for safety critical systems as possible position errors and the reduced integrity of any position calculated could allow for a catastrophic occurrence to happen.

This is why further investigation into the possible position errors and signal interference in the rail environment is needed, as without an initial understanding of the errors expected, further technological development to combat any issues is not possible.

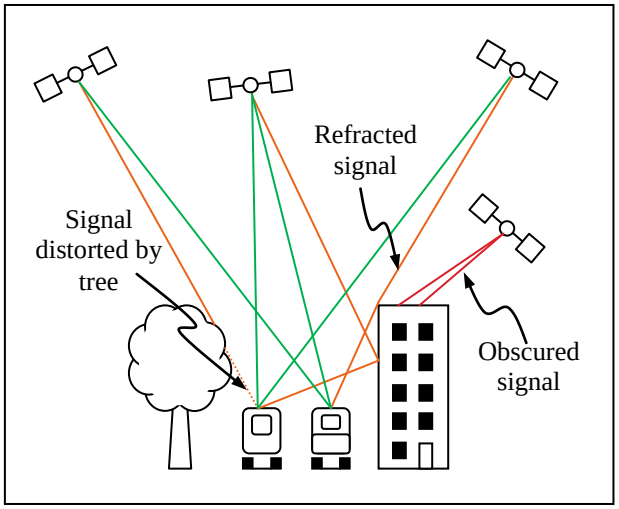


Figure 2.9 – Trains in urban/rural environment

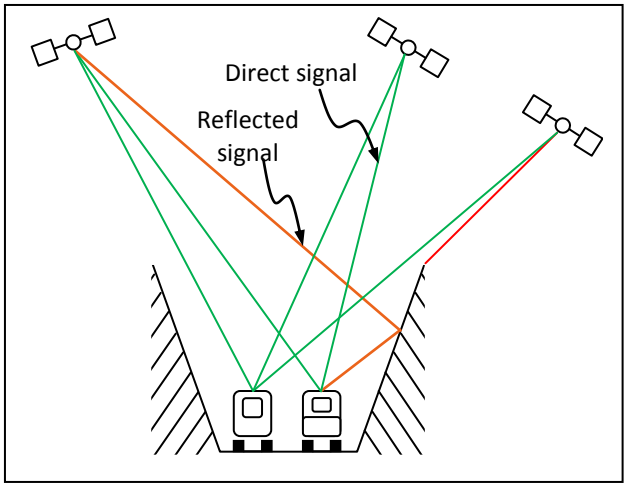


Figure 2.10 – Trains in deep cutting

As well as being completely blocked by physical barriers, signals can be received directly, diffracted, reflected only, attenuated as well as be subject to multipath just as shown for common rail environments in figures 2.9 and 2.10.

2.6.1 Multipath

Multipath is the phenomenon where a reflected signal is received along with a direct signal from the same satellite. This is then received by the receiver as a composite signal. Multipath is particularly difficult to detect as for short periods it is not zero mean and so a bias can still be left, even after time averaging of the signals, especially when considering a close reflecting surface.

Figure 2.11 shows a simplistic multipath diagram for a single reflection causing a second, delayed signal to reach the receiver.

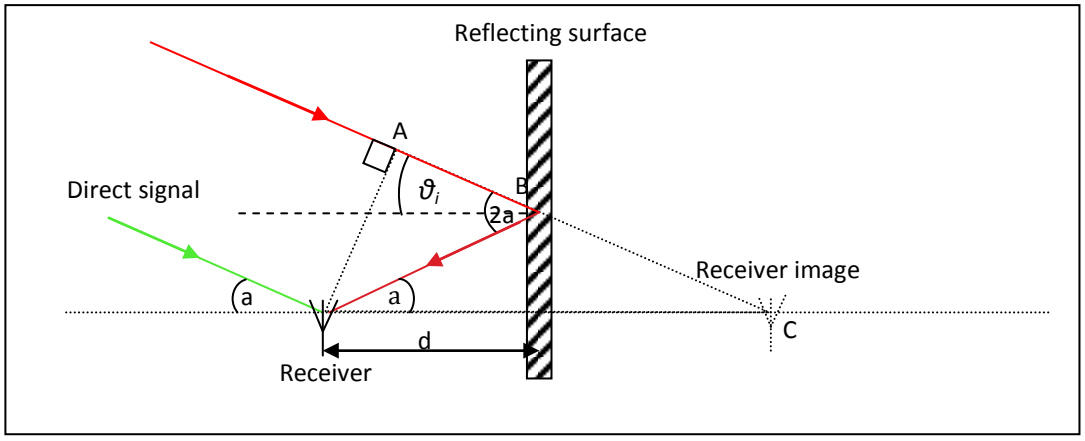


Figure 2.11 - – Geometry of a simple multipath occurrence

The extra distance δ_d travelled by the reflected signal is $\overrightarrow{AB}$+$\overrightarrow{BC}$, thus:

$$\delta_d = 2d \cos a \qquad (2.36)$$

From 2.36, if considering phase measurements converting to cycles and then radians gives the total phase shift θ:

$$\theta = \frac{2\pi\delta_d}{\lambda} \qquad (2.37)$$

Where λ is the signal wavelength

The receiver sees a combined composite of the direct and reflected signal which Braasch (Braasch, 1996) describes using phasor diagrams to give the total carrier phase multipath θ_c as:

$$\theta_c = \tan^{-1}\left(\frac{\alpha R(\tau_c - \delta)\sin\theta}{R(\tau_c) + \alpha R(\tau_c - \delta)\cos\theta}\right) \qquad (2.38)$$

Where α is the attenuation factor or multiple-to-direct signal ratio
$R(\tau_c)$ is the correlation function for the PRN at time lag τ_c
δ is the relative path delay

From 2.38 by assuming a worse case whereby α is equal to 1 and that τ_c and δ are considered small, it can be calculated that the maximum possible phase tracking error is π/2 radians. This is shown in figure 2.12. On the L1 frequency, a π/2 phase error equates to roughly 4.8cm.

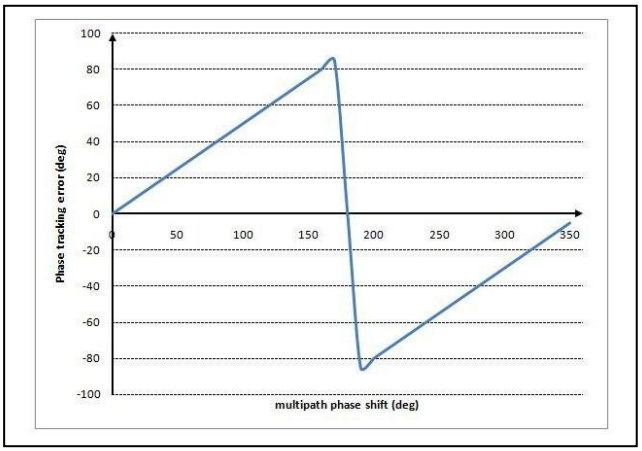


Figure 2.12 - – multipath phase tracking shift against phase tracking error

The actual value of α varies depending on the reflecting material, frequency of the incident signal and its angle of incidence, θ_i (Kavak, Xu, & Vogel, June 1996). With GPS signals being circularly polarized, the reflection coefficients for the perpendicular and parallel components are different.

For the perpendicular component, $\rho_{\perp,}$ provided the incident and reflecting mediums are both lossless nonmagnetic dielectrics the attenuation factor can described for phase and pseudorange measurements by:

$$\rho_{\perp} = \frac{\cos\theta_i - \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_i}}{\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_i}} \qquad (2.39)$$

Where ε_1 and ε_2 are the relative permittivity of the incident and reflecting media respectively.

For the parallel component, $\rho_{\parallel}$,for lossless nonmagnetic dielectrics:

$$\rho_{||} = \frac{-\left(\frac{\varepsilon_2}{\varepsilon_1}\right)\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_i}}{\left(\frac{\varepsilon_2}{\varepsilon_1}\right)\cos\theta_i + \sqrt{\frac{\varepsilon_2}{\varepsilon_1} - \sin^2\theta_i}} \qquad (2.40)$$

The assumption of dielectric properties is valid as most surfaces in an outdoor environment such as wood, concrete and glass are dielectric in nature and metal objects usually have a coating of silicon oxide or aluminium oxide for weatherproofing.
Research has been carried out to experimentally verify values for reflectance coefficients of various smooth surfaces such as concrete and leafy hills (Weiss & Axelrad, 2007) and relative permittivity for grass and asphalt (Kavak, Vogel, & Xu, 1998) in the GPS frequency band. Smooth surfaces produce specular reflections which are coherent and can to a certain extent be modelled, whereas rough surfaces produce incoherent diffusely scattered reflections which are difficult to model.

Composite multipath signals also cause the code tracking delay lock loop (DLL) to correlate in a constructive (reflected signal in phase with direct signal) or destructive (reflected and direct signal out of phase) manner as shown in figure 2.13.
The receiver uses an internally generated replica of the signal to sample the rising and falling edges of the peak of correlation so that the exact time of signal arrival can be deduced.
Multipath alters the form of the peak which affects the time of arrival calculation and causes the pseudorange measurements to appear shorter for destructive interference and longer for constructive interference.

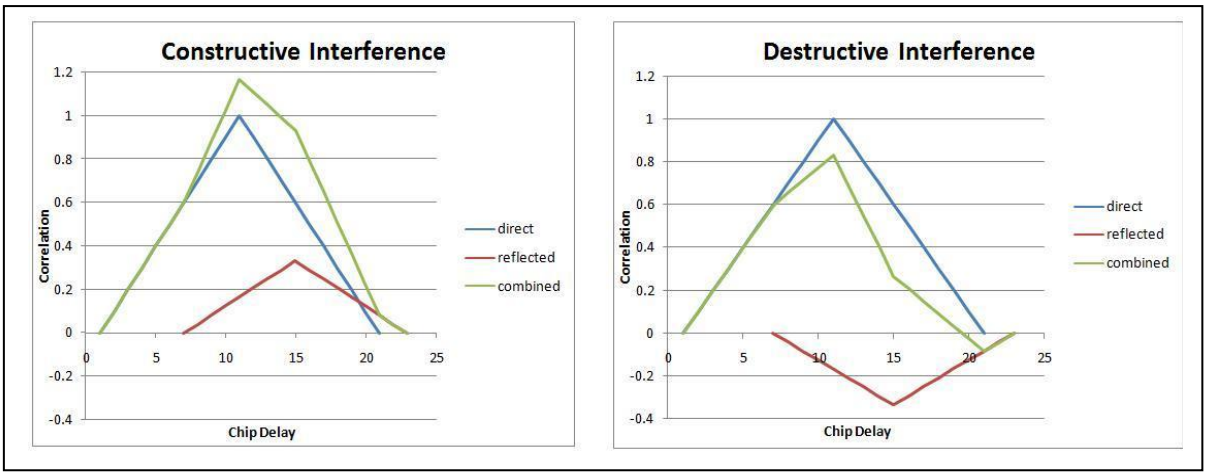


Figure 2.13 - Correlation effects of multipath

There are generally four main methods of reducing multipath in GNSS; antenna location, software, hardware a combination of the last two.
Receiver design can reduce the effect of interference by decreasing the correlator early-late spacing and increasing the precorrelator bandwidth a certain amount (Braasch, 1996). If the precorrelator bandwidth is too narrow, the correlation function peak becomes rounded and so any advantage of using a narrow correlator spacing is lost. Conversely, if the bandwidth is too great, the receiver is far more susceptible to electromagnetic interference and any satellite signal anomalies will produce larger pseudorange errors than that of a narrower bandwidth (Phelts & Enge, 2000).
The multipath delay and resulting tracking errors for the C/A code and P code for a basic receiver can be seen in figure 2.14.

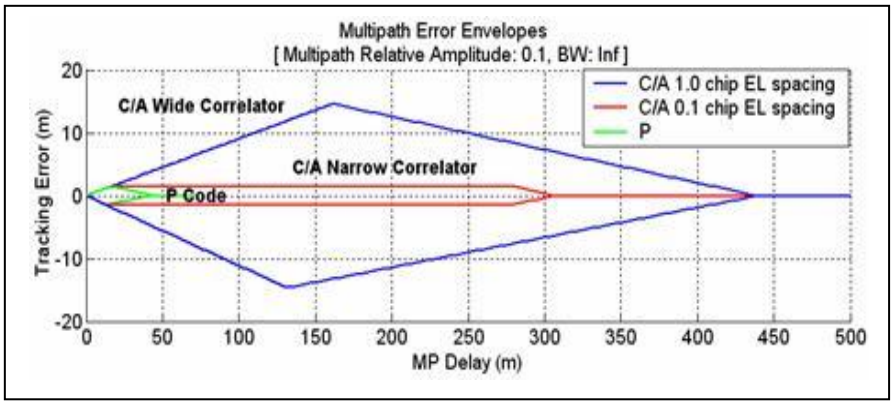


Figure 2.14– Diagram showing multipath delay and corresponding tracking error for C/A code and P code

The difference between a narrow and wide correlator spacing can be clearly seen, with the resulting tracking error for the narrow correlator reducing to a fraction of the wide correlator.

NovAtel communications (Fenton, Townsend, & Diedrendonck, 1995) designed a prototype correlator that uses multiple early and late samplings so that the effect of multipath on the correlation function gradients could be assessed. This system was reported to provide a reduction in multipath similar to that of a choke ring antenna. This was then improved upon in the NovAtel Pulse Aperture Correlator (PAC) by implementing the same technique in a receiver with higher bandwidth (Jones, Fenton, & Smith, 2004).

NovAtel have more recently reported on their ‘vision’ correlator (Fenton & Jones, 2005) that uses the PRN code chip transition shape as a multipath mitigation reference. By looking at the RF properties of a chip transition from 1 to 0 or the reverse, the effects of multipath can be easily distinguished compared to the traditional summation correlator design seen in figure 2.13.

By initialising the vision correlator with high elevation satellites with relatively low multipath, the basic transition curve can be stored and then compared against all incoming signals so that the effects of multipath can be determined.

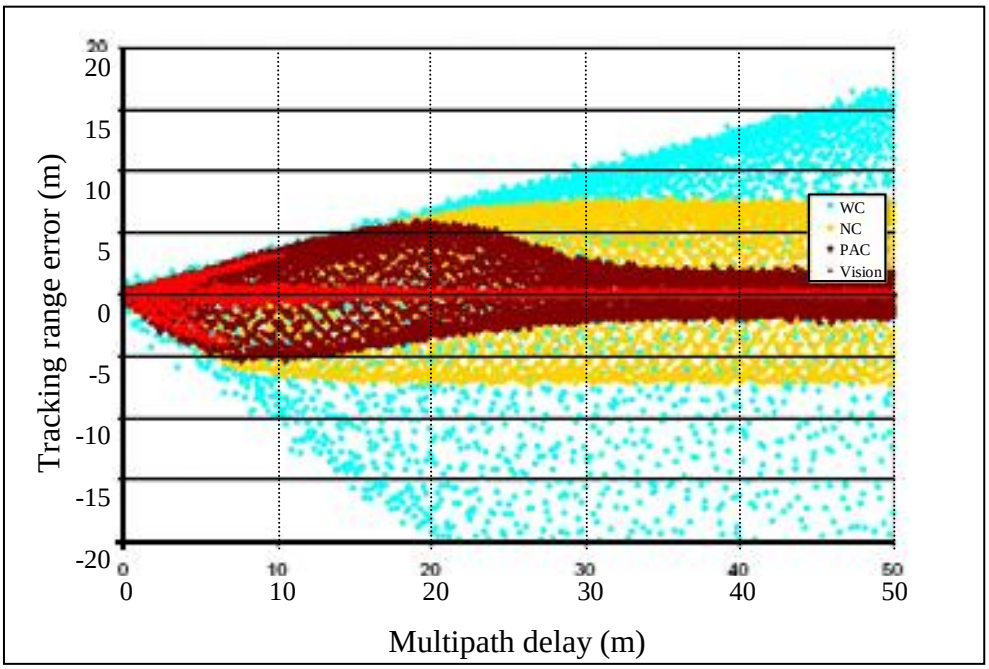


Figure 2.15– comparison of multipath processing techniques

The effects of different correlation techniques can be seen in figure 2.15. This figure shows the original wide correlator (WC) spacing results together with the now commonplace narrow correlator (NC) results. The NovAtel PAC and latest Vision correlator techniques are also shown, with the results proving that range errors can be limited to within 5m provided state of the art correlator designs are implemented in the receiver.

Some signal tracking loops can also remove multipath if it varies at a higher frequency than the loop bandwidth. High frequency multipath is more common for a kinematic receiver, especially if the reflecting surfaces are far away. In the rail environment, this would most likely come from the along-track direction and be due to overhead infrastructure or lineside infrastructure after a turn due to the across track visibility normally being limited to shorter distances.

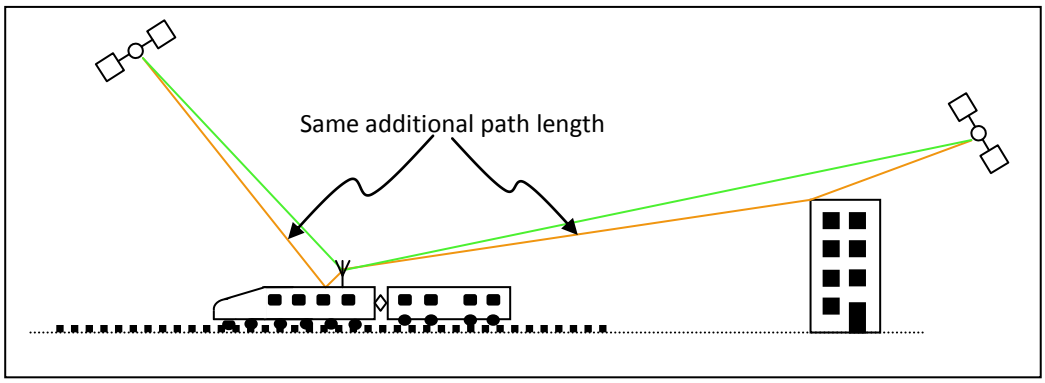


Figure 2.16– Near and far object multipath

From figure 2.16 we can see two theoretical situations that can produce the same level of multipath. Despite the building being a further distance away than that of the trains own roof, the additional path length equates to the same.
The only way to distinguish between the resultant multipath is by looking at the frequency with which they vary as local reflectors produce lower frequency multipath and far away reflectors produce higher frequency multipath variation. This is due to the changing satellite geometry's effect on the reflected signals additional path length. Per Enge (Enge, 1999) describes how the high frequency motion of an aircrafts skin produces 'white noise' multipath that can be removed by carrier smoothing but ground reflections produce low frequency multipath variation which leads to measurement errors.
Due to receiver design limiting the remaining code multipath to additional path lengths below 10m, the situations in figure 2.16 are of interest in the rail environment as they were the most likely sources of shorter additional path lengths.

Provided a reflecting surface is not a perfect conductor, single reflections of a signal can cause a switch from right hand circularly polarized (RHCP) to left hand elliptically polarized (LHEP) if the angle of incidence, θ_i, is less than the Brewster angle, θ_B, defined below:

$$\theta_B = \tan^{-1}\sqrt{\frac{\varepsilon_1}{\varepsilon_2}} \qquad (2.41)$$

If the angle of incidence is greater than the Brewster angle, the reflected signal is still elliptically polarised, but with the same polarisation as the incident signal. When the incident angle is equal to the Brewster angle, the reflected wave becomes linearly polarized perpendicular to the plane of the reflecting surface.

Most GPS antennas have been designed specifically to attenuate LHEP signals so single reflections (which also have the highest gain of any reflected signals) are removed. This feature does not completely remove the errors due to multipath as multiple reflections can have both RHEP and LHEP configurations but it does reduce the number of first order reflections that have the highest attenuation gain, thus reducing the level of correlation skew. The use of choke-ring antennas also reduces

the effects of multipath from below the horizon, but not from above, and due to their size, train mounting would prove problematic due to height and rooftop space restrictions.

Another method of reducing multipath is by limiting the minimum elevation angle for a satellite to be accepted and used. This reduces the use of signals with ground reflections as well as those from buildings and trees where interaction is more common, as well as more likely to produce short additional path lengths.
The problem with removing these satellites entirely is that availability is then reduced and so a weighting system is more appropriate in order to guarantee a maximum number of satellites used in the position calculation (Collins & Langley, 1999). This is on top of the reduced attenuation for low elevation satellites employed by most geodetic antennas.
Signal to noise ratios (SNR) can also be used as a weighting parameter as the attenuation of a reflected signal increases noise, allowing a user to determine if a signal is direct or attenuated in some way previous to reception. This is more common when dealing with phase observations.
One such approximation for the undifferenced phase variance σ^2 in mm^2 as a function of the carrier to noise power density ratio C/N0 was shown by Ward (Kaplan, 2005) as being:

$$\sigma_i^2 = C_i . 10^{-(C/N0_{measured}\ )/10} \qquad (2.42)$$

Where i indicates which L signal (L_1 or L_2)
C_i is the carrier loop noise bandwidth in mm^2

Equation 2.42 was built upon by Brunner (Brunner, Hartinger, & Troyer, 1999) by adding a scaled Δ value:

$$\Delta = C/N0_{template}\ - C/N0_{measured} \qquad (2.43)$$

Where $C/N0_{template}$ is derived from interference free elevation versus C/N0 readings at the GPS site in question.

Equation 2.42 then becomes 2.44.

$$\sigma_{\Delta}^{2} = C_i . 10^{-(C/N0_{measured} \ -\alpha.\Delta)/10} \qquad (2.44)$$

This SIGMA-Δ model proved to be successful at improving 2.42 in reducing multipath, but can only really be applied to stationary sites and in this instance, phase measurements only.

Other models, such as that proposed by Lau (Lau & Cross, 2005) have also been successful in improving and in this case, correcting phase multipath but have not been successfully applied in a highly dynamic environment due to its reliance on the sinusoidal characteristic of phase multipath. The applications of SNR weighting to code measurements is not as common.

Comparisons of elevation angle versus SNR multipath modelling have been made (Wang & Satirapod, 2000) with conclusions favouring the SNR approach as a more accurate system of multipath modelling.
Problems arose when trying to determine the SNR values used by each receiver manufacturer and the less than ideal format of SNR simply being given as a value between 0 and 9.

Code multipath error can be calculated as an observable if dual frequency phase data is available (Cross, Lecture Notes for Positioning 1 (GEOMG004) and Positioning 2 (GEOMGS05); Code Multipath Observable - UNAVCO, 2007), by breaking the signals into their elements (ignoring satellite and receiver clock offsets) as follows:

$$P_1 = D + I_1 + t + M_{P1} \qquad (2.45)$$

$$L_1 = D + B_1 - I_1 + t + M_{L1} \qquad (2.46)$$

$$L_2 = D + B_2 - I_2 + t + M_{L2} \qquad (2.47)$$

Where P_1 is the code measurement on L_1 frequency
L_1 is the phase measurement on the L_1 frequency
L_2 is the phase measurement on the L_2 frequency

D is the true range

$I_{1,2}$ are the ionospheric effects on the L_1 and L_2 signals respectively

$B_{1,2}$ are the integer ambiguities on the L_1 and L_2 signals respectively

t is the tropospheric effect on the signal

M_{P1} is the multipath effect on the P_1 code measurement

$M_{L1,L2}$ are the multipath effects on the L_1 and L_2 phase measurements respectively

By forming the L1 code/phase combination and assuming phase multipath is negligible, equations 2.45 and 2.46 become:

$$M_{P1} - B_1 = P_1 - L_1 - 2I_1 \qquad (2.48)$$

From 2.47 and 2.48:

$$L_1 - L_2 = I_2 - I_1 + B_1 - B_2 \qquad (2.49)$$

By using the ionospheric ratio $I_2 = I_1\left(\frac{f_1^2}{f_2^2}\right)$ 2.49 becomes:

$$L_1 - L_2 = I_1(\alpha - 1) + B_1 - B_2 \qquad (2.50)$$

Where $\alpha = \left(\frac{f_1^2}{f_2^2}\right)$

Therefore:

$$2I_1 = \frac{2}{(\alpha - 1)}(L_1 - L_2) + \frac{2}{(\alpha - 1)}(B_2 - B_1) \qquad (2.51)$$

From 2.48:

$$M_{P1} - B_1 = P_1 - L_1 - \frac{2}{(\alpha - 1)}(L_1 - L_2) - \frac{2}{(\alpha - 1)}(B_2 - B_1) \qquad (2.52)$$

Rearranging 2.52 we get:

$$M_{P1} - \left\{B_1 - \frac{2}{(\alpha - 1)}(B_2 - B_1)\right\} = P_1 - L_1 - \frac{2}{(\alpha - 1)}(L_1 - L_2) \qquad (2.53)$$

Which gives us the L_1 code multipath observable M_{P1}:

$$M_{P1} = P_1 - 4.0915L_1 + 3.0915L_2 + Constant \qquad (2.54)$$

The corresponding L_2 observable can also be calculated as:

$$M_{P2} = P_2 - 5.0915L_1 + 4.0915L_2 + Constant \qquad (2.55)$$

Where the *Constant* term is an offset due to the remaining integer ambiguity.
This offset can be reduced by time averaging, provided a long enough time period is observed to allow for low frequency multipath variation as the changes in M_{P1} and M_{P2} for static receivers are sinusoidal as the satellites move through the sky, thus altering the reflected signal path lengths. Moving antennas are more likely to have a white noise characteristic to multipath due to the reflecting surfaces being altered. This white noise multipath is also expected to have a normal distribution due to each reading being independent from another and there being a lower probability of large value multipath.
For an accurate multipath measurement to take place, averaging is used as it is assumed the level of multipath (both destructive and constructive interference) will have zero mean over time.
The $M_{P1,2}$ observables are normally used as a data quality check as any large changes in its value over time will indicate either a cycle slip or multipath. As an observable, cycle slips must first be removed if the results are to be a true indication of possible multipath.

2.6.2 EM interference

Interference with GPS signals can come from many sources and can be intentional or unintentional.
Intentional interference is not common unless dealing with a military situation or possible terrorism and is known as intentional electromagnetic interference (IEMI) (Radsky, Baum, & Wik, 2004).
The most basic form of intentional interference is by using high-power electromagnetics (HPEM) to introduce enough signal noise to make signal tracking and acquisition impossible. Due to GPS signals being relatively weak, high power is not necessarily needed for basic jamming of a receiver. HPEM can also disable the receiver itself by radiating the electronic components, causing them to malfunction.

Intentional spoofing is an act whereby a deliberately corrupt or incorrect signal mimics a correct GPS signal and is transmitted so that receivers in transmission range try to use the incorrect information. In 2002 an article describing how to build a short range L1 C/A jammer was published in Phrack magazine (issue 60). For US military applications, the anti-spoofing W code was modulated onto the P code for this specific reason, however spoofing has supposedly been attempted by forces in Afghanistan during the recent conflicts with little success according to the US military, despite the use of RF guided smart bombs being used to destroy the jamming beacons.
Because civil users don't have access to the W code, they are left vulnerable to spoofing from hackers and terrorists, something the Federal Aviation Aadministration (FAA) are aware of.

The L1 and L2 frequency bands have been earmarked for GPS and thus unauthorised for civilian users to transmit on, but unintentional jamming can still occur; one well known case was that of a malfunctioning TV preamplifier (Goodman, 2003) that caused the Moss landing incident in April of 2001 where sailors in Moss Landing harbour, California were left in a situation where GPS seemed to disappear.
Any unintended interference occurs either due to a radio system using a closely neighbouring frequency, where if enough power is used, saturation can occur and thus impede the functioning of a GPS receiver, or by radio systems operating outside the band but with in-band emissions from harmonics or frequency conversion products. Other spurious interference sources can be as diverse as engine ignition systems, TV and computer monitors, electric motors, fluorescent lights, ac-dc converters, alternators, generators and switching power supplies.
The increase in UWB devices such as medical equipment, communications, construction, ground radars and WiFi networking systems also may have an effect as despite their individual power being limited to -75.3 dBm/MHz in the protected GNSS band, the collective effect of multiple devices could significantly increase the noise floor and cause receiver tracking errors (ANASTASIA project, 2006).
Interference that is more specific to the rail environment would come from localised emitters such as GSMR radio beacons, diesel engine electrics, passengers personal electronic items and overhead electrified cables.

Figure 2.17– National grid pylons (© Andrew Dunn 2004)

There have been investigations into the effect of national grid electrified cables on GPS signals with conclusions showing no detectable interference during normal operation (Lee & Ge, 2006). The effect of the physical pylon supporting structure themselves (figure 2.17) and their possible scattering effect on GNSS signals has not been fully investigated due to the complexity of their structure.
Although the EM interference effects from the electric cables themselves seems to be below a detectable level, the wide band electronic noise caused by corona formed on conductor surfaces is of concern as the interference fields from conductor corona has been detected up to frequencies of 1GHz (Olsen, 1994).

M. Silva ((Silva J. M., 2001)) shows that for a line voltage of 500 kV at 50Hz, the effective incident noise power density ($N_{incident}$) on a receiver placed underneath a transmission line is -115.3 dB-Hz μW/m^2**.**

Given that the incident power density of a carrier signal from a directional transmitter is:

$$C_{incident} = [P_t(\text{dBw})+G_t(\text{dB}_i)-10\log_{10}(4\pi d2)+60] \text{ dB } \mu\text{W/m2} \qquad (2.56)$$

Where P_t is the transmitter power
G_t is the antenna gain relative to an isotropic antenna
d is the distance between satellite and receiver in meters

From 2.56, $C_{incident}$ is calculated as being -70 dB μW/m^2, giving a C/N0 value of 45.3 dB-Hz which is close to the minimum specified by some receiver manufacturers.
In practical tests however, the EM noise due to corona, even in worst condition foul weather where noise is 15-25 dB higher did not seem to degrade the receiver operation when using the NMEA amplitude measurement unit (AMU) output.
Hardware that is at different voltages and physically close can produce sparks known as gap discharge if the potential difference is high enough (see figure 2.18). These discharges become more frequent in dry weather to produce wideband EM noise similar to that of corona. Experimental results have found that the effects of gap discharge are similar to corona whereby the level of EM noise is undetectable.

Figure 2.18 - – Pylon detail common to cause gap discharge

Nationwide DGPS services broadcast in the UK by Trinity House in the 291 to 309.5 kHz frequency range are also liable to interference from corona noise, gap discharge and nearby electronic RF noise sources (Silva & Olsen, 2002). This reduces the availability of the DGPS network and can thus degrade a receiver's position and integrity performance if it is reliant upon these available signals.

The levels of safe operation as defined by RNP's, although not being a common specification for a discrete component of the system, need to be feasible given the number of possible error sources outlined in this chapter.
When a safety assessment is carried out on a system, Fault Tree Analysis (FTA - as mentioned in 3.6.2) would propagate any errors at a component level through to the

final system level, causing a failure of the system's ability to function within the defined levels. Further safety analysis techniques are discussed in the next chapter.

3 Risk and safety principles

3.1 Introduction

The dictionary defines safety as:

'A state in which or a place where you are safe and not in danger or at risk'

Safety of life systems are becoming increasingly more technical and complicated in their construction, requiring a safety evaluation method that can compile such intricacies to provide an overall statement for a system based upon its components. The safety case documents a system helping to prove it is of an acceptable safety standard based upon supporting evidence about its components, ranging from scientific outcomes from rigorous testing or engineering knowledge about similar systems and educated statements based upon this. The safety case is meant to tell a story for a system, detailing tests undertaken, the results and subsequent alterations applied to mitigate intolerable levels of error, culminating in a clear declaration of the system's ability to perform its task at the required level of safety.

3.2 The safety case

3.2.1 Overview

The main aim of a safety case is to tell a story about a certain system, culminating in an overall confidence statement on the systems safety level. This should include:

- Goals/Claims for a system.
- Evidence for claims made.
- Arguments linking evidence to the goals (claims).

The goals/claims are explicit requirements the final system has to satisfy in order for a successful safety claim to be made. They are based upon well defined criteria that are validated for the given environment and application, providing adequate safety level compliance.
Evidence is given by hazard logs, risk modelling and reduction methods, quality and safety reports, FMEA, etc and presented by methods such as ETA and FTA to give a complete picture of the system safety level.

Evidence is considered if it is not part of the original claims on the system, i.e. if additional evidence is to be used for a safety case, it cannot simply be as a result of the original system design parameters. Arguments must accompany all evidence, along with any assumptions and technical judgements that are made by the personnel involved.
Each section has different viewpoints within a project and each of these must be included, some of which are:

- Safety specialists
- Internal and external regulators
- Operators and managers
- Senior staff that decide equipment is safe to enter the public domain
- Lawyers

There is as such, no rigidly defined structure for a general safety case, making it difficult to know exactly what should be included, however the Railtrack yellow book has a section on the required content of a railway safety case (Railtrack, 2000) and it states that there should be:

- Executive summary
- Introduction
- Definition of system
- Quality management report
- Safety management report
- Technical safety report
- Related safety cases
- Conclusion

The safety case can be written for an entire system or a single component, it must prove that the safety level is achievable consistently based upon statistical evidence and expert opinion.

3.2.2 Project implementation of a safety case

Engineering safety management projects require an interaction between various parties such as project coordinators and engineers and an independent safety authority and safety assessor and ultimately, the customer.

For a safety case to be approved and a project to be released into the public domain, there are certain steps to endorse the safety case and review the risks involved, together with the running hazard log. Figure 3.1 shows how the safety case fits into the overall project safety assessment.

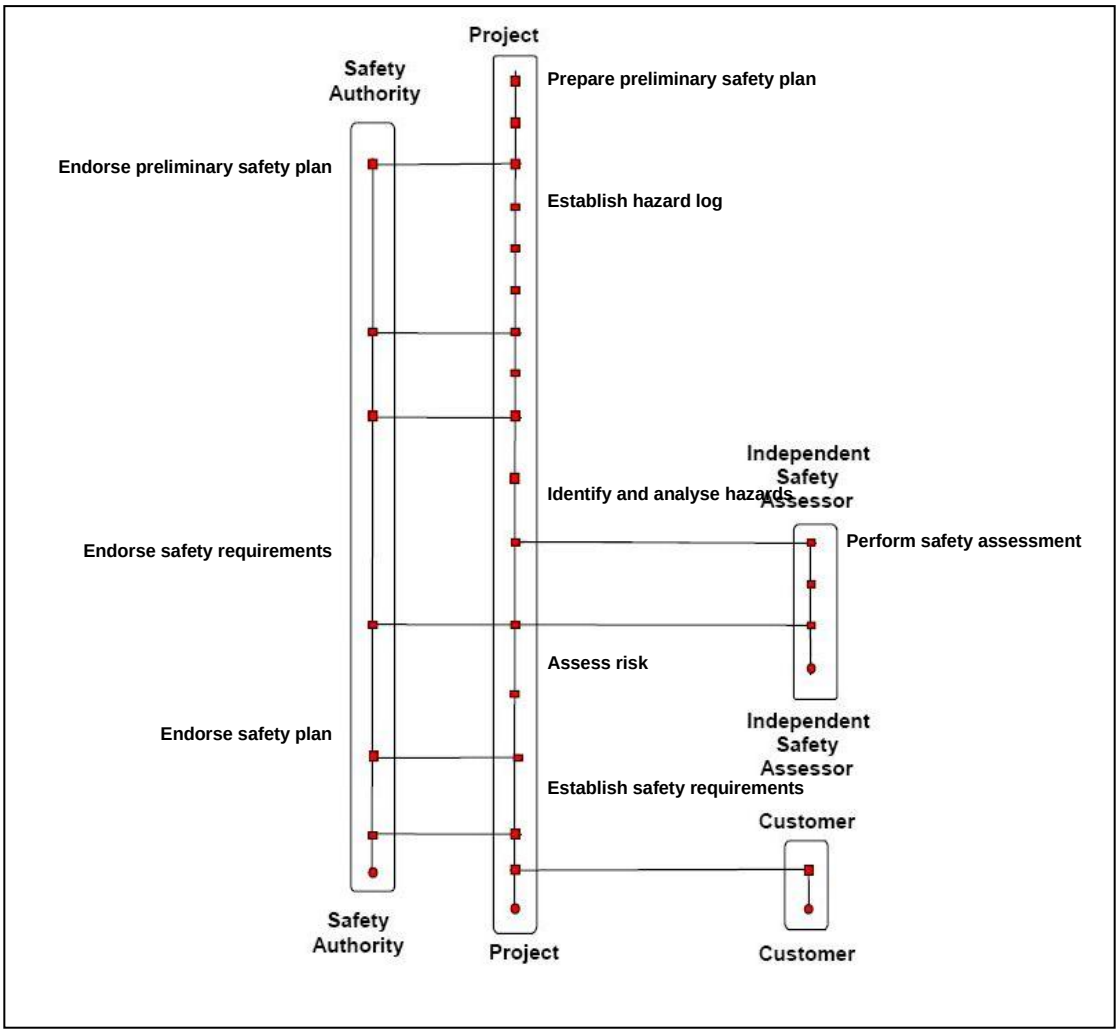


Figure 3.1 - Diagram showing project safety implementation links **(Railtrack, 2000)**

3.3 Risk definition and categories

Risk is the prime concern when constructing a safety case and it consists of two components, frequency and severity, in the relationship:

$$\textit{Risk} = \textit{Frequency} \times \textit{Severity}$$

This allows a more quantitative method of approach towards risk as the frequency of a systems failure can be measured statistically through analysis of data taken over a known timeframe. Severity can be quantified in various ways, from financial outlay for the correction of failures or the consequences of an accident where machinery or systems have become damaged or delayed or by the level of injury or number of fatalities that have occurred due to the error.

Safety cases have apportioned set levels of risk based upon the severity and frequency categories of an incident occurrence. There are six defined quantitative levels of frequency and four levels of hazard severity:

Severity level	**Consequences to persons or Environment**	**Consequence to service**
Catastrophic	Fatalities and/or multiple severe injuries and/or major damage to the environment	
Critical	Single fatality and/or severe injury and/or significant damage to the environment	Loss of a major system
Marginal	Minor injury and/or significant threat to the environment	Severe system damage
Insignificant	Possible minor injury	Minor system damage

Table 3.1 - Hazard severity breakdown

Category	Description
Frequent	Likely to occur frequently. Hazard will be continually experienced.
Probable	Will occur several times. Hazard can be expected to occur often.
Occasional	Likely to occur several times. Hazard can be expected to occur several times
Remote	Likely to occur sometime in the system life cycle. Hazard can be reasonably expected to occur.
Improbable	Unlikely to occur but possible. It can be assumed that the hazard may exceptionally occur
Incredible	Extremely unlikely to occur. It can be assumed the hazard will not occur.

Table 3.2 – Frequency category breakdown

Using Tables 3.1 and 3.2; risk evaluation takes place by formulating a frequency-consequence matrix to determine the acceptance level of the hazard defined by the qualitative risk categories:

Risk category	Actions to be applied against each category
Intolerable	Shall be eliminated.
Undesirable	Shall only be accepted when risk reduction is impracticable and with the agreement of the authority.
Tolerable	Acceptable with adequate control and with the agreement of the authority.
Negligible	Acceptable with/without the agreement of the authority.

Table 3.3 - Qualitative risk categories

Frequency of occurrence of a hazardous event	**Risk levels**			
Frequent	Undesirable	Intolerable	Intolerable	Intolerable
Probable	Tolerable	Undesirable	Intolerable	Intolerable
Occasional	Tolerable	Undesirable	Undesirable	Intolerable
Remote	Negligible	Tolerable	Undesirable	Undesirable
Improbable	Negligible	Negligible	Tolerable	Tolerable
Incredible	Negligible	Negligible	Negligible	Negligible
	Insignificant	Marginal	Critical	Catastrophic
	Severity Level of Hazard Consequence			

Table 3.4 - Frequency-Consequence matrix

Risk acceptance is based upon a number of principles defining when alterations to mitigate risk should be carried out.

3.4 Risk determination methods

In order to determine if a new system or paradigm should be implemented in the railways, several methods of analysis are used. These normally relate the current level of risk to the future level created after the change has been implemented.
Three of these analysis methods are detailed below, along with the countries in which they are normally practiced.

- **Globalement Au Moins Aussi Bon** (GAMAB as practiced in France)

  The main formulation being “All new systems must offer a level of risk globally at least as good as the one offered by any equivalent existing system”. This implies that progress must be made by the requirement “at least” and by using “globally” it does not single out a particular risk. The supplier is free to distribute allocation between the different risks within a system and applies the relevant approach, be it qualitative or quantitative.

- **Minimum Endogenous Mortality** (MEM as practiced in Germany)

  This accounts for death rates in the society caused by technological facts such as entertainment and sports, DIY, work machines and transport. It does not include deaths by illness and disease or congenital malformation.
  The risk is referred to as "Endogenous Mortality" is denoted by the letter R. In well developed countries R is lowest between the ages of 5 and 15 and is denoted by R_m (as it is the minimum value) and typically has the value $2x10^{-4}$ fatalities/(person × year). The impact of a new system upon the R_m value is the measure of the risk introduced by the system.

- **As Low As Reasonably Practicable** (ALARP as practiced in the UK)

  This is primarily analysis of a hazardous situations improvement being disproportionate to the financial outlay and effects on service operability.

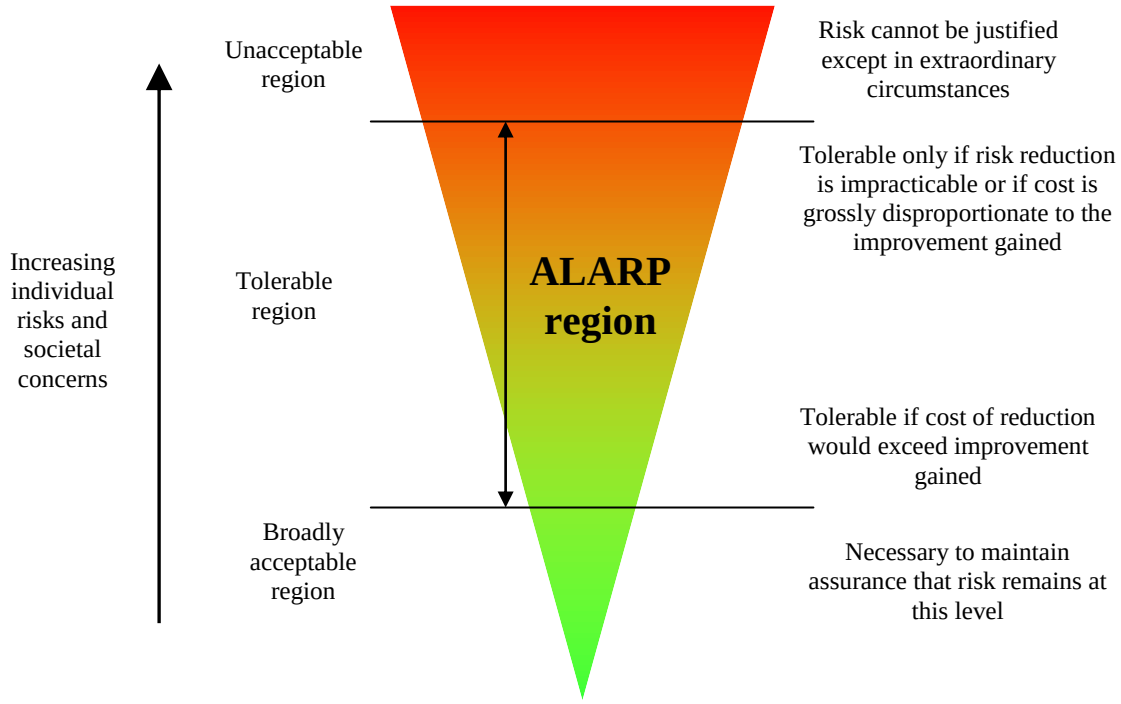


Figure 3.2 – Diagram of ALARP principal

3.5 Safety integrity levels

Safety integrity relates to the probability of a safety related system or component achieving its required safety function. The higher the safety integrity, the lower the probability of failure to perform the required safety functions like providing an accurate position.
Safety integrity is comprised of two components:

1. Systematic failure integrity.
2. Random failure integrity.

For the required safety level to be achieved, both types of failure mode must be considered when defining a SIL.
There are four discrete levels for specifying the safety integrity of a safety related system/component, SIL 4 the highest and SIL 1 the lowest, with SIL 0 being non safety dependent.
Tables 3.5 and 3.6 show the qualitative and quantitative levels defined for use in safety critical systems.

Safety Integrity Level		Descriptive words	
4	Very high	Vital	Fail safe
3	High		High integrity
2	Medium	Semi vital	Medium integrity
1	Low		Low integrity
0	Not specified	Non vital	Non safety

Table 3.5 - SIL qualitative level definition

SIL	PFD Prob Fail on Demand	THR/h Tolerable Hazard Rate/h	Availability required
4	10^{-5} to 10^{-4}	10^{-9} to 10^{-8}	>99.99%
3	10^{-4} to 10^{-3}	10^{-8} to 10^{-7}	99.9% to 99.99%
2	10^{-3} to 10^{-2}	10^{-7} to 10^{-6}	99.9% to 99.99%
1	10^{-2} to 10^{-1}	10^{-6} to 10^{-5}	99.9% to 99.99%

Table 3.6 - SIL quantitative level definition

The probability of a failure on demand is for a system that is not necessarily running all the time and so it determines how often the system will fail when you attempt to access it, thus it does not have any units.

If a SIL is defined for a system, it should address both the quantitative failures associated with random failures and the qualitative safety management systematic failures.

The required level of safety integrity for a given application is based on the results of hazard analysis and risk assessment carried out by qualified analysts.

In the UK railway safety case it is advised that RNP's are used as a guide for the assessment of a systems Accuracy, Integrity, Availability and Continuity as shown in table 3.7 from the APOLO project (Barbu & Alcouffe, 1999).

Class	Accuracy	Integrity	Availability	Continuity	Time to Alarm	Coverage
A	~ 1 m	high	> 0.99	high	< 1 s	Europe
B	~1m-10m	high	> 0.99	high	< 5 s	Europe
C	< 10 cm	high	> 0.99	high	< 5 s	Europe

Table 3.7 - Principal requirements for railway safety related application as defined in APOLO project

The Class system in 3.7 refers to the applications of the technology, defined as being:

Class A – Most demanding, automatic system including shunting and parallel track determination

Class B – Mainly refers to speed checking and line authority for drivers.

Class C – Shunting exercises and depot manoeuvres.

The high integrity and continuity generally refer to a value of 10^{-9} per operation, where the operation duration obviously plays a factor.

3.6 Hazard analysis and determination methods

Hazards and their possible outcomes have to be identified in many different ways as simply brainstorming for possible system insecurities and failure modes may not identify all the possibilities. Different methods have been designed to cope with a wide range of system architectures, from system level through to individual components so that the component level failures can be tracked through the system to assess their full impact.
When beginning a project, a hazard log is started so that the details of all hazards and potential accidents/failures that are identified during safety analysis can be recorded. All documentation linked to the safety of a system is logged and stored in the hazard log.

Some of the safety analysis methods used to identify hazards for hazard logs and safety cases are:

- Hazard and Operability studies (HAZOP) uses a set of guide words to identify possible hazards, this continues throughout the project lifecycle to ensure all risks are identified and rectified, with the results being stored in the hazard log.
- Fault tree analysis (FTA) uses a system level fault at the top of the tree to identify the combinations of events that can cause it.
- Event tree analysis (ETA) identifies the consequences of an initiating event in a complex system in a logical and simple way.
- Failure Mode and Effect Analysis (FMEA) reviews the full system at component level, identifying failures and their consequences. By identifying these faults early on, design modifications at a later stage can be avoided.
- Failure Mode, Effects and Criticality Analysis (FMECA) adds the criticality measurement to the consequences of the failures in FMEA.

Criticality analysis is a procedure where each potential failure is ranked according to the combined influence of severity and probability of occurrence. It is the American system for risk analysis.

- Safety Integrity Levels (SIL) are discrete quantitative values for defining a systems or a components integrity level. There are 5 levels from 0 to 4, with 0 being a safety independent level and 4 being the highest achievable level.

3.6.1 HAZOP process

Key features of HAZOP examination include the following.

- The examination is a creative process. The examination proceeds by systematically using a series of guide words to identify potential deviations from the design intent and employing these deviations as "triggering devices" to stimulate team members to envisage how the deviation might occur and what might be the consequences.
- The examination is carried out under the guidance of a trained and experienced study leader, who has to ensure comprehensive coverage of the system under study, using logical, analytical thinking. The study leader is preferably assisted by a recorder who records identified hazards and/or operational disturbances for further evaluation and resolution.
- The examination relies on specialists from various disciplines with appropriate skills and experience who display intuition and good judgement.
- The examination should be carried out in a climate of positive thinking and frank discussion. When a problem is identified, it is recorded for subsequent assessment and resolution.
- Solutions to identified problems are not a primary objective of the HAZOP examination, but if made they are recorded for consideration by those responsible for the design.

The HAZOP process consists of four main sections as detailed below:

1. **Definition**
   - Define scope and objectives
   - Define responsibility
   - Select team

2. **Preparation**
   - Plan the study
   - Collect data
   - Agree style of recording
   - Estimate the time
   - Arrange a schedule

3. **Examination**
   - Divide system into parts
   - Select a part and define design intent
   - Identify deviation by using guide words on each element
   - Identify consequences and causes
   - Identify whether a significant problem exists
   - Identify protection, detection, and indicating mechanisms
   - Identify possible remedial/mitigating measures (optional)
   - Agree actions
   - Repeat for each element and then each part of the system

4. **Documentation and follow-up**
   - Record the examination
   - Sign off the documentation
   - Produce the report of the study
   - Follow up that actions are implemented
   - Re-study any parts of system if necessary
   - Produce final output report

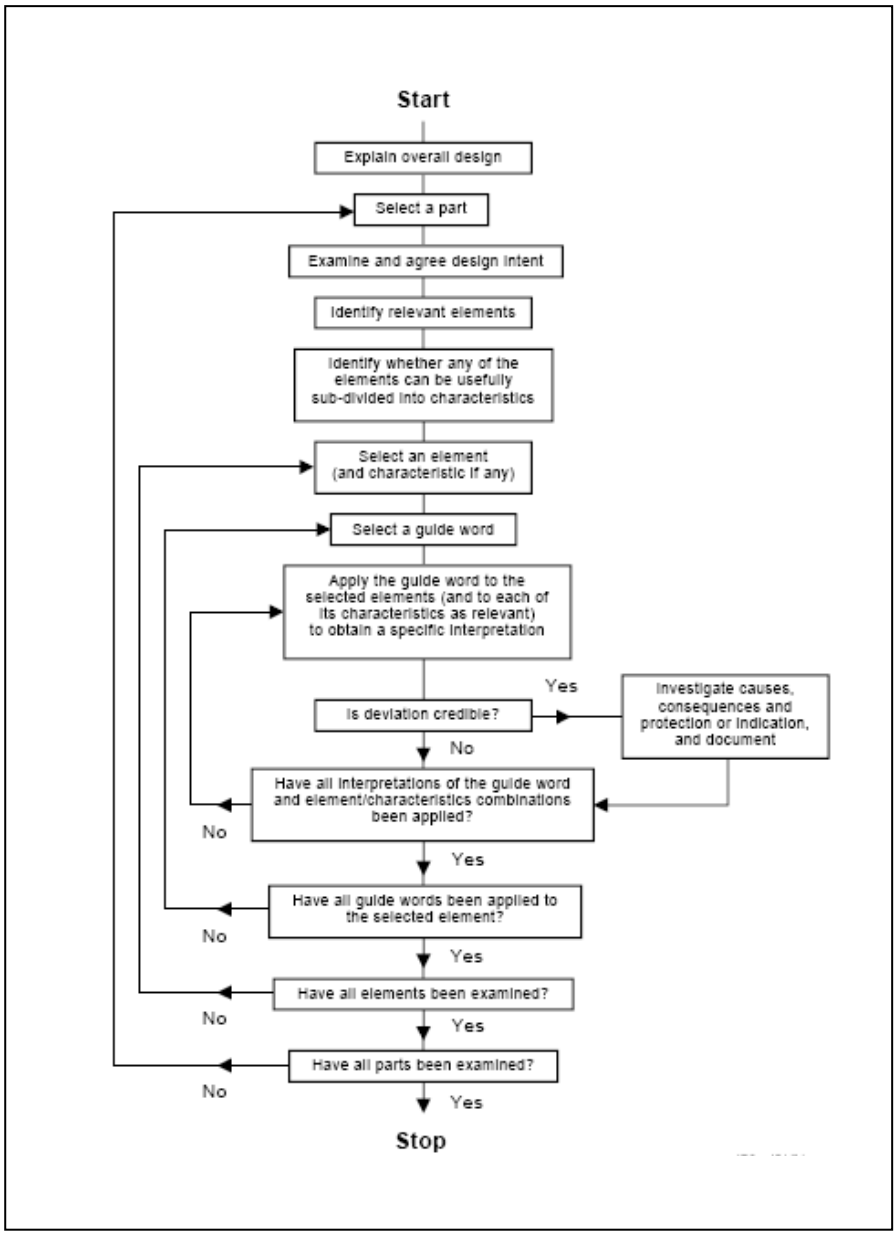


Figure 3.3 - The HAZOP process chain

Throughout the HAZOP process and the life of the system in question a hazard log must be kept for permanent reference purposes.
The Railtrack ‘yellow book’ outlines the hazard log structure as follows:

Introduction: This section describes the purpose of the hazard log and indicates the environment and safety requirements to which the system safety characteristics relate.
It will include:

- The aim, purpose and structure of the hazard log in sufficient detail to enable understanding by all project personnel.
- A unique identifier of the system to which the hazard log relates and a reference to a description and the scope of the system.
- A reference to the safety plan (in early stages of the project this will have to be omitted).
- A reference to the system safety requirements specification or, if this has yet to be written, the safety analysis documentation.
- The process for managing the hazard log, such as who may modify it and the approval process for each new entry.

Journal: The journal should describe all amendments to the hazard log in order to provide a historical record of its complication and provide traceability.

It should record, for each amendment:

- The date of the amendment (not necessary if a diary format used).
- A unique entry number.
- The person making the amendment.
- A description of the amendment and the rationale for it.
- The sections in the hazard log that were changed.

Directory: The directory, sometimes known as the safety record log, should give an up-to-date reference to every safety document produced and used by the project.

The documents referred to should include (but not be limited to) the following, where they exist:

- Safety plan.
- Safety requirements specification.
- Safety standards.
- Safety documents.

- Incident/accident reports.
- Analyses, assessment and audited reports.
- Safety case.
- Correspondence with the relevant safety authorities.

For each document the directory should include the following:

- A unique reference.
- The document title.
- The current version number and issue date.
- The physical location of the master.

Hazard data: This section should record every identified hazard.
For each hazard, the information listed below should be recorded, as soon as it becomes available. Data collected during hazard analysis and risk assessment should be transcribed to the hazard log when the reports have been endorsed.

- A unique reference.
- A brief description of the hazard which should include the system functions or components affected and their states that represent the hazard.
- The causes identified for the hazard.
- A reference to the full description and analysis of the hazard.
- Assumptions on which the analysis is based and limitations of the analysis.
- The severity for the related accident, the likelihood of the hazard occurring and the likelihood of an accident occurring with the hazard as a contributing factor.
- The predicted risk associated with the hazard.
- Target likelihood for its occurrence.
- The status of the hazard; Typically one of the following:

- Open (action to close the hazard has not been agreed).
- Cancelled (the event has been determined not to be a hazard or to be wholly contained within another hazard).
- Resolved (action to close the hazard has been agreed but not completed).
- Closed (action to close the hazard has been completed).

- If the hazard is not closed or cancelled then the name of a person or company who is responsible for progressing it towards closure.
- A description of, or a reference to, the action to be taken to remove the hazard or reduce the risk from the system to an acceptable level.
- This should include:

  - A statement as to whether the hazard has been avoided or requires further action (with a justification if no further action is to be taken).
  - Details of the risk reduction action to be taken.
  - A discussion of the alternative means of risk reduction and justification for actions considered but not taken.
  - A comment on the need for accident sequence re-evaluation following risk reduction actions.
  - A reference to any design documentation that would change as a result of the action.
  - A reference to all Safety Requirements associated with this hazard.

Incident data**:** This section should be used to record all incidents that have occurred during the life of the system or equipment. It should identify the sequence of events linking each accident and the hazards that caused it. For each incident the following should be provided:

- A unique reference.
- A brief description of the incident.
- A reference to a report describing an investigation of the incident.
- A description of any action taken to prevent recurrence or justification of the decision not to take any.

Accident data: This section should be used to record every identified possible accident. It should identify possible sequences of events linking identified accident with the hazards that may cause it.
For each accident the following should be provided:

- A unique reference.
- A brief description of the potential accident.
- A reference to a report giving a full description and analysis of the accident sequence.
- A categorisation of the accident severity and the highest tolerable probability of the accident (the accident probability target).
- A list of the hazards and associated accident sequences that could cause the accident.

3.6.2 Fault tree analysis

The International Standard (BSI, 1990) describes fault tree analysis, and gives guidance on its application, as follows:

- Defining basic principles.
- Providing the steps necessary to perform an analysis.
- Identifying appropriate assumptions, events and failure modes.
- Providing identification rules and symbols.

The fault tree itself is an organized graphical representation of the conditions or other factors causing or contributing to the occurrence of a defined undesirable event,

referred to as the "top event". The representation is in a form which can be understood, analyzed and, as necessary, rearranged to facilitate the identification of:

- Factors affecting the reliability and performance characteristics of the system, for example component fault modes, operator mistakes, environmental conditions, software faults;
- Conflicting requirements or specifications which may affect reliable performance;
- Common events affecting more than one functional component, which could cancel the benefits of specific redundancies.

Fault tree analysis is basically a deductive (top-down) method of analysis aimed at pinpointing the cause or combinations of causes that can lead to the defined top event. The analysis is mainly qualitative but, depending on certain conditions, it may also be quantitative.

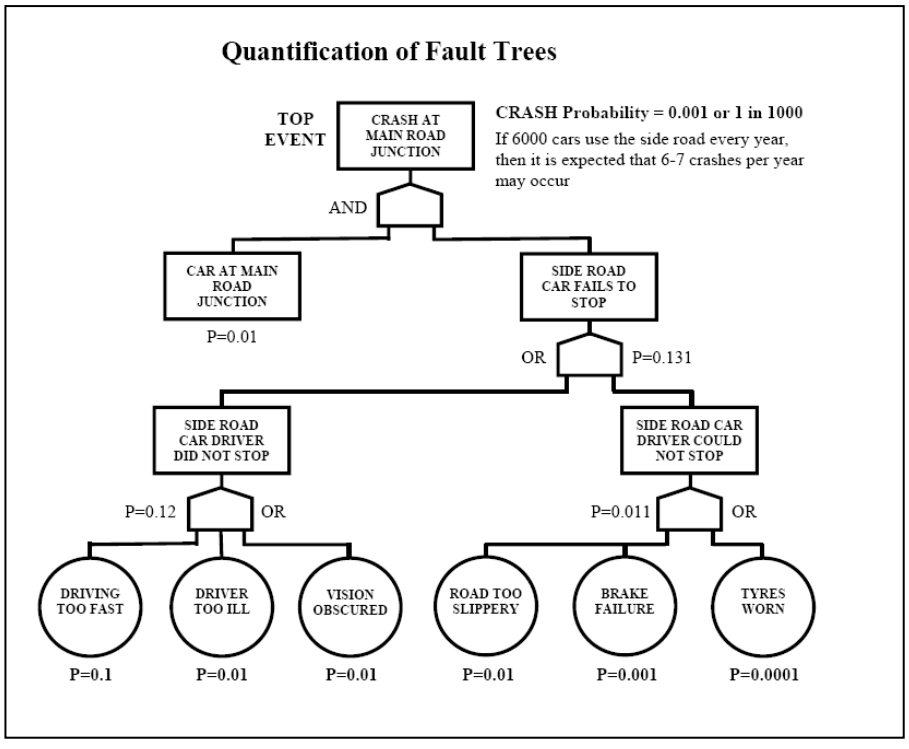


Figure 3.4 - Example of FTA for a car crash

From this evaluation technique, Figure 3.5 was created by A. Filip et al (Filip, Polivka, & Suchanek, 2006) for the Derivation of GNSS Signal In Space (SIS) integrity and continuity risk requirements for the Class C operations mentioned in 3.5.

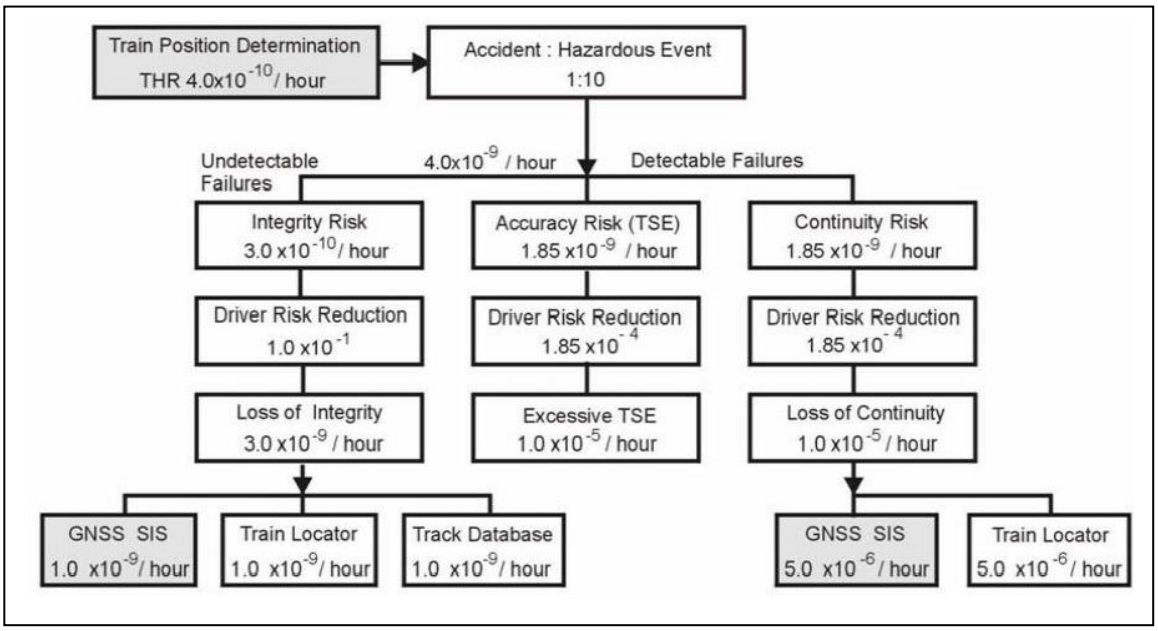


Figure 3.5 - Derivation of GNSS SIS integrity and continuity risk requirements for Class C operations.

3.6.3 Event tree analysis

By identifying an initiating event, all potential accident scenarios and outcomes in a complex system can be identified in a logical and numerical manner.

By tracing the effects of an incident through a system its weaknesses and design flaws can be identified and corrected and the resulting outcomes statistically documented.

Although there are no standards for the graphical layout of the final tree, the main steps for event tree analysis are logically as follows:

1. Identify the unwanted initiating event or accident.
2. Identify the integral system modules that are designed to prevent further consequences.
3. Design an event tree to graphically show these preventative measures.
4. Outline the resulting consequences.
5. Establish probabilities for the initiating event and the resulting event tree branches.
6. Calculate the frequency of the outcomes defined.
7. Record all calculated data on the tree diagram.

Event tree analysis can only be used on a single initiating event, requiring the possibility of multiple trees for one system, especially if there are variable conditions for the initiating event to occur.
Figure 3.6 is an example ETA for a fire alarm system:

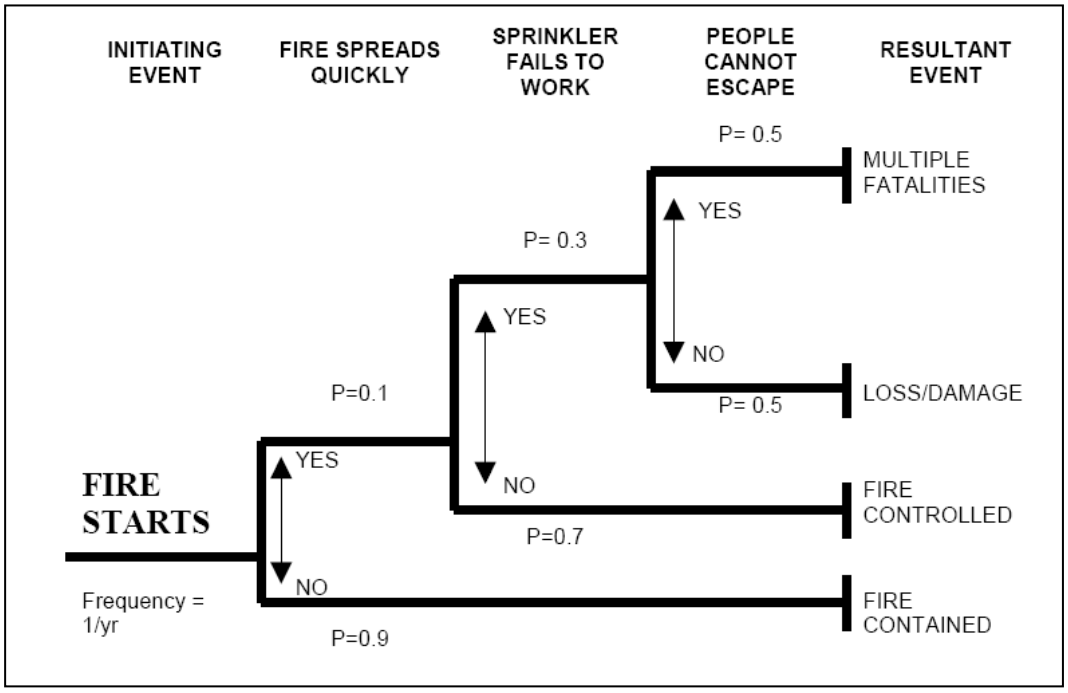


Figure 3.6 - Event tree analysis for fire alarm system

3.6.4 Failure mode and effects analysis and failure mode, effects and criticality analysis

FMEA is an inductive method of performing a qualitative system reliability or safety analysis from a low to a high level. A thorough understanding of the system under analysis is essential prior to undertaking FMEA.
Functional diagrams and other system drawings are normally necessary for this understanding. Reliability block diagrams, fault trees, event trees etc. are then usually derived from these in order to carry out the analysis. In many instances the block diagram descriptions and block diagram failure descriptions are included in the FMEA format.

The core details required for FMEA to be carried out are:

1. The name of the item in the system being analysed.
2. Function performed by the item.
3. Identification number of the item.
4. Failure modes of the item.
5. Failure causes.
6. Failure effects on the system.
7. Failure detection methods.
8. Compensating provisions.
9. Severity of effects.

Other information required for a particular system and project needs to be defined by the analyst assigned.
Some of the detailed applications and benefits of FMEA are listed below:

- To avoid costly modifications by the early identification of design deficiencies.
- Identify failures which, when they occur alone or in combination, have unacceptable or significant effects, and to determine the failure modes which may seriously affect the expected or required operation.
- To determine the need for the following:
    - — Redundancy.
    - — Design improvement.
    - — More generous stress allowances.
    - — Screening of items.
    - — Design of features that ensure that the system fails in a preferred failure mode, e.g. ‘fail-safe’ outcomes of failures.
    - — Selection of alternative materials, parts, devices, and components.

FMEA can be difficult and tedious for the case of complex systems that have multiple functions involving different sets of system components. This is because of the quantity of detailed system information that needs to be considered. This difficulty

can be increased by the existence of a number of possible operating modes, as well as by consideration of the repair and maintenance policies.

In FMEA, it's not sufficient to consider only random and independent failures. Some 'common-cause' (or 'common mode') failures (CCF) can occur, that cause system performance degradation or failure through simultaneous failures in several system components, due to a single source such as design or human error.
A CCF is the result of an event that, because of dependencies, a coincidence of failure states in two or more components (excluding secondary failures caused by the effects of a primary failure).
Table 3.8 is an example of an FMEA worksheet.

FMEA

Indenture level:					**Design by:**			**Prepared by:**				
Sheet no.:					**Item:**			**Approved by:**				
Mission phase:					**Issue:**			**Date:**				
Item ref	**Item description-function**	**Failure entry code**	**Failure mode**	**Possible failure causes**	**Symptom detected by**	**Local effect**	**Effect on unit output**	**Compensating provision against failure**	**Severity class**	**Failure rate (F/Mhr)**	**Data source**	**Recommendations and actions taken**
1.1.1	motor stator	1111	open circuit	winding fracture	low speed roughness	low power	trip	single phase protection temperature trip	4			
		1112	open circuit	connection fracture	low speed roughness	low power	trip	single phase protection temperature trip	3			
		1113	insulation breakdown	persistant high temp. manufacturing defect	protection system	overload	no output	annual inspection temperature trip	4			
		1114	thermistor open circuit	ageing connection fracture	protection system	none	no output	fitted spare	3			recommend consideration spare connected through to outside casing
		1115	thermistor short circuit	failure thermistor	protection system	reduced trip margin	no output if load high	fitted spare temperature trip	3			recommend consideration spare connected through to outside casing

Table 3.8 - FMEA worksheet example

FMECA is a logical extension of FMEA bringing together the failures and the criticality of the outcome. As with risk, criticality is evaluated by a subjective measure of the severity of the effect and an estimate of the probability or expected frequency of its occurrence.
Using the failure effects identified in the FMEA, each effect is allocated an appropriate severity class as defined by the risk matrix. The results of FMECA are recorded in a worksheet similar to the FMEA sheet, but with the additional columns, as per table 3.9:

								Exiting conditions						Resulting conditions			
Ref.	Process	Failure mode	Effect on	Potential effect	V	Potential cause	Existing controls	Occ	Sev	Det	APN	Recommended action	Action taken	Occ	Sev	Det	APN
01-01-01	Inserts	Incorrect size or shoulder bend angle	i)a	Inserts without load onto die. Reduced productivity		Poor manufacture or quality control	Producer and acceptance sampling plans	1	9	1	9	Review of sampling plans. Segregation of defective stock from good stock. Training assemblers					
02			i)b	Insert malaligned. Scrap													
03			i)a	Incorrect thickness of skirt surrounding insert. Scrap													
04			iv)b	Reduced performance													
05			iv)c	Reduced life													
01-02-01	Inserts	Poor flash nickel plating	ii)a	Corrosion. Rejected at finishing stage			Visual inspection during acceptance sampling plan	5	6	1	30	Include instructions in sampling inspection to carry out visual check for correct plating					

Table 3.9 - FMECA worksheet example

Modernisation of safety case presentation techniques has seen programs such as Adelards (Adelard, 2006) ASCE graphical hypertext safety case system offering new methods of presenting substantial amount of information (something that is common for the rail domain) in an easily navigated manner.

By using an overall structural network map to present a complete system breakdown, each module in the network can represent an individual component.

From these components, the program provides links for strategies, goals, assumptions and solutions to potential faults. All links have associated context and narrative embedded hypertexts so that the user or assessor has access to all the projects details.

The idea behind such graphical interface programs is that system producers can pass on a full goal structured notation to an assessing body in a style that is easy to navigate and understand rather than passing on all the raw data and expect the assessors to fully understand the system under scrutiny as a whole.

3.6.5 Use of Risk and Safety principals for GNSS in the Rail domain

Full risk analysis is not an easy thing to do for GPS as for the system itself, producing the SIS is massively complex and so it falls at organisations such as the CAA and the FAA to try and determine any failure modes, effects or hazards from failures in the network of ground stations and satellites that make up the GPS constellation operations. These failures would then need to be sent to users in a real time fashion so

that anyone relying on GPS as a means of navigation is fully aware of any errors. The satellites do transmit a health status along with the navigation message, but this is not updated in real time and simply states if a satellite should be used or not.
Future signals, like the addition of an integrity signal as proposed for Galileo should allow users to determine with confidence if a satellite should be used or not in real time, hopefully with a defined 'time to alarm' limit as defined by the operators of the system.

Receiver design is also very complex and at the time of writing this thesis, no company has undergone a full FMEA or FMECA of its receiver, making the uptake of GPS as a positioning technology by the UK railways difficult due to the unknown failures that could occur within the 'black box' receiver.
The receiver failures that could occur may cause a complete loss of function or an incorrect position. When considering the size of possible position errors, provided the navigation system uses additional sensors, such as INS, large instantaneous jumps in position or miss-matching with an additional GPS receiver should provide a level of risk management.
When considering the RNP's accuracy for normal line authority is in the region of 1m-10m with a time to alarm of 5s, if the train is travelling at roughly 55mph, it will cover 122m of track in this time and so the overall error budget allowed for a known position along the track comes to roughly 132m (122+10), a position error that a correctly functioning GPS receiver is able to function within provided a minimal amount of multipath mitigation is used. But as this amount includes the time to alarm, when a position error reaches this level, the system must be able to declare as such. Any internal integrity monitoring provided by the receiver may fail at this point and thus the possibility of a hazard is realised. In order to avoid this happening, system redundancy and error mitigation techniques such as additional sensors and backup navigation equipment would likely need to be employed, but as the current off the shelf systems do not provide any quantitative analysis of their failure, further investigation of any receiver used would be required.
This thesis attempts to investigate the error introduced to GPS signals by multipath and EM radiation from the rail environment so that further work (if required) can be carried out in a more focused manner and any failure mitigation steps can be defined as part of the HAZOP process.

4 Real Time Kinematic Network software suite

4.1 Introduction

The UCL department of civil, environmental and geomatic engineering has designed and written a fully functioning GNSS research library in the C++ programming language. The contributions to the library from this work were modules for RINEX file reading and Saastamoinen modelling, the remaining modules were created by other members of the development team. The library is primarily designed to function as a network based solution for high accuracy kinematic position solutions. Through integrating state of the art algorithms to solve for ionospheric and tropospheric disturbance as well as antenna offsets, geoid models and integer ambiguity validation techniques precision positions can be calculated. The addition of receiver autonomous integrity monitoring (RAIM) also allows the user to easily manipulate and test the latest integrity algorithms in a familiar experimental environment.

4.2 Modular construction breakdown

The RTK library has been designed from the outset to be modular in design, so that the user can apply whatever models are available by simply selecting them from a list, or they can create their own and add to the library for others to use. This methodology facilitates the ongoing evolution of the library as a departmental research tool.
The breakdown of the system at the highest level can be seen in figure 4.1.

Figure 4.1 – High level design of modular RTK library

(Parkins, 2007)

The main inputs required depend upon how the user specifies which library functions are to be used. This is specified in the settings file that is read at runtime every time the program in executed.

If precise SP3 (Spofford & Remondi, 1991) orbit interpolation is selected, providing the files are present in the projects folder, the program will provide satellite positions using the IGS SP3 product files.

The library can also use the broadcast ephemeris (BRDC) data transmitted in the navigation message to calculate the satellite positions if the higher accuracy SP3 files

are not available due to latency. This can allow the user to use the library in real time if rapid results are required, but this will be at slight detriment to the accuracy.
Depending on the network layout used, the library is designed to accept multiple roving receivers as well as multiple base station receivers. The accepted format for the receiver input is the industry standard Receiver Independent Exchange (RINEX) format derived by Werner Gurtner (Gurtner, 2006). The RINEX reader has been specifically designed to read multiple input files with the primary rover file acting as the seeker file, defining the cross-receiver epochs of interest.
Once the receiver data has been read, validation occurs in order to make sure common receiver errors and incorrect values are not passed on to the next processing stage.

The next stage is single epoch point positioning for all the receivers accounting for ionospheric, tropospheric, geoid, satellite clock and position offsets.
The reference station positions are assumed to be known and so the point positioning process is used to determine the receiver clock offsets.
After the reference station clock offsets have been computed, a minimum spanning tree is computed so that inter-station ambiguity resolution can take place with a higher success rate.

With the addition of antenna model corrections for phase offset and variance, the double differenced ambiguities to all common satellites are estimated, providing float solutions.

This is achieved in the same way as seen in section 2.3.2 by forming the double difference observation equation, shown in equation 4.1 for a baseline between two stations *A* and *B* using a reference (*R*) and commonly viewed (*i*) satellite.

$$\rho_{AB}^{Ri} = dT_{AB}^{Ri} - dI_{AB}^{Ri} + \varepsilon_{AB}^{Ri} + \frac{c}{f}\left(\varphi_{AB}^{Ri} - N_{AB}^{Ri}\right) \qquad (4.1)$$

Where N_{AB}^{Ri} is the double difference float ambiguity
φ_{AB}^{Ri} is the phase observable

ρ_{AB}^{Ri} is the double differenced geometric distance between receiver and satellite

dT_{AB}^{Ri} is the tropospheric effect

dI_{AB}^{Ri} is the ionospheric effect

ε_{AB}^{Ri} represents residual errors such as multipath and measurement noise

f is the GPS frequency (L1 or L2)

c is the speed of light

The covariance matrix for the least squares float solution is computed and fed into the LAMBDA (De Jonge & Tiberius, 1996) method along with the solution.

The least squares residuals provided by the LAMBDA method are then tested using the ratio method so that the chosen results are at least 2 to 3 times better than the next best solution (Zinas, 2007).

Provided this test is passed, the new, fixed ambiguities are used to interpolate for the rover so that its ambiguities can also be fixed.

If multiple epochs are to be processed and no cycle slips have occurred, the associated ambiguities can be compared as an additional form of solution validation.

The library has also been designed to provide as much information as possible to the user during operation so that if any faults arise, their origins are known. The program log file contains error reports and individual epoch statuses so that the user can easily identify any potential problems as well as track progress.

The processed results are written to multiple files based upon the output type and the individual receiver. Observations to all satellites are written to individual files for each receiver, stored according to epoch number. These observations have not been corrected for cycle slips and are a direct copy of the observations found in the corresponding station RINEX file. Among these observations are the M_{P1} and M_{P2} code multipath observables discussed in section 2.6.1.

Satellite files are also stored, giving the position, elevation angle and azimuth of each satellite from each receiver on an epoch by epoch basis.

For each rover, a file detailing the various position solutions is also produced. It contains the point position solution in cartesian earth centred, earth fixed coordinates as well as geodetic longitude, latitude and height values.

Along with the point position solution, the float ambiguity and, if available, fixed ambiguity solutions are stored on an epoch by epoch basis along with fixed ambiguity and ambiguity validation flags for the reference stations and rover.

4.3 Testing and validation

Using the RTK library to process single baseline ambiguities, the effects of the different available tropospheric and ionospheric models can be computed easily. Figure 4.2 shows the single baseline between the London (LOND) and Barking (BARK) Ordnance survey base stations used for initial testing.
The baseline is 15.3Km long with the Barking station using a Trimble 400ssi receiver and Trimble antenna and the London station using a Leica SR530 receiver and Ashtech antenna.

The 24 hour dataset selected was recorded on the 18th of October 2007 and features a 30 second epoch rate. Final version SP3 files were used for orbit determination.
For the single epoch ambiguity resolution testing LOND was used as a virtual rover and BARK as a reference station. Precise coordinates are known for both stations through long term averaging by the Ordnance survey, as shown in table 4.1.

Station	X coordinate	Y coordinate	Z coordinate
LOND	3979606.812	-8329.692	4967677.525
BARK	3977368.459	6726.346	4969508.937

Table 4.1 – Table showing OS station ETRS89 earth centred cartesian coordinates

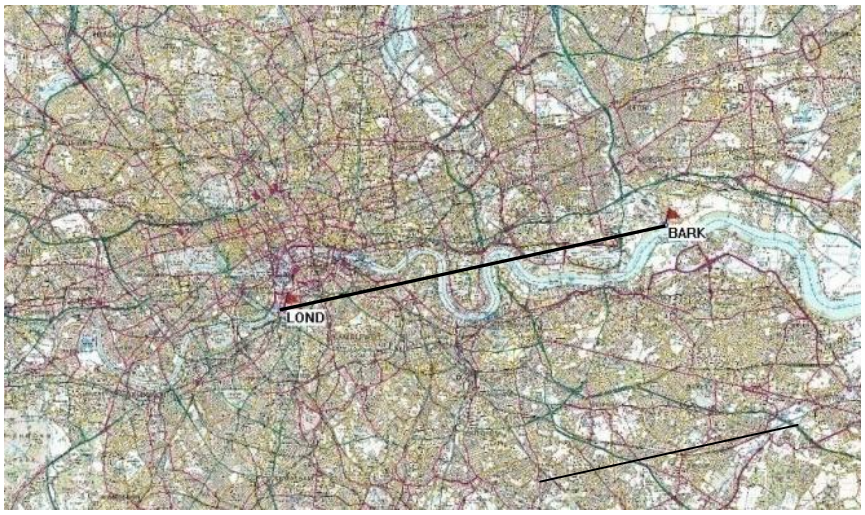


Figure 4.2 – Map showing LOND-BARK OS station baseline

By using base station data for the rover and reference and knowing the precise coordinates of both, we can check the accuracy of the final single epoch position solutions. We can also see the effects of the various options available to the user when selecting which tropospheric and ionospheric correction method to use.
X,Y and Z values were processed in 9 variations of atmospheric model as shown in table 4.2. The ionospheric models are either the Klobuchar model or the use of an IONEX file (ionosphere mapping exchange format) that is a 2D or 3D TEC map in a geographic grid.
The tropospheric models available are the Saastamoinen model or the ESA developed matlab model.

Plotting name	Ionospheric model used	Tropospheric model used
Saas, None	None	Saastamoinen
Saas, Klob	Klobuchar	Saastamoinen
Saas, Ionex	Ionex file	Saastamoinen
ESA, GMF, None	None	ESA Zenith delay with global mapping function
ESA, GMF, Klob	Klobuchar	ESA Zenith delay with global mapping function
ESA, GMF, Ionex	Ionex file	ESA Zenith delay with global mapping function
None, None	None	None
None, Klob	Klobuchar	None
None, Ionex	Ionex file	None

Table 4.2 – Atmospheric model variations used in initial test

The resulting plots using SP3 satellite positions are shown below in figures 4.3 to 4.5:

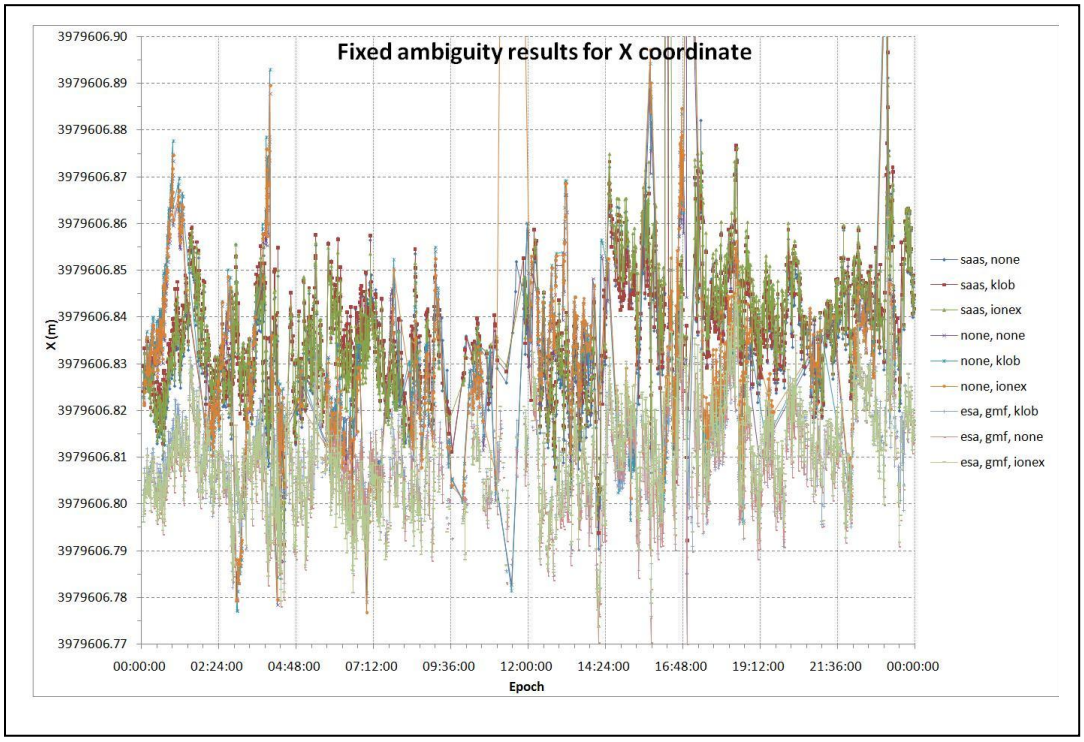


Figure 4.3 – Plot of X coordinates for LOND rover using different atmospheric model combinations

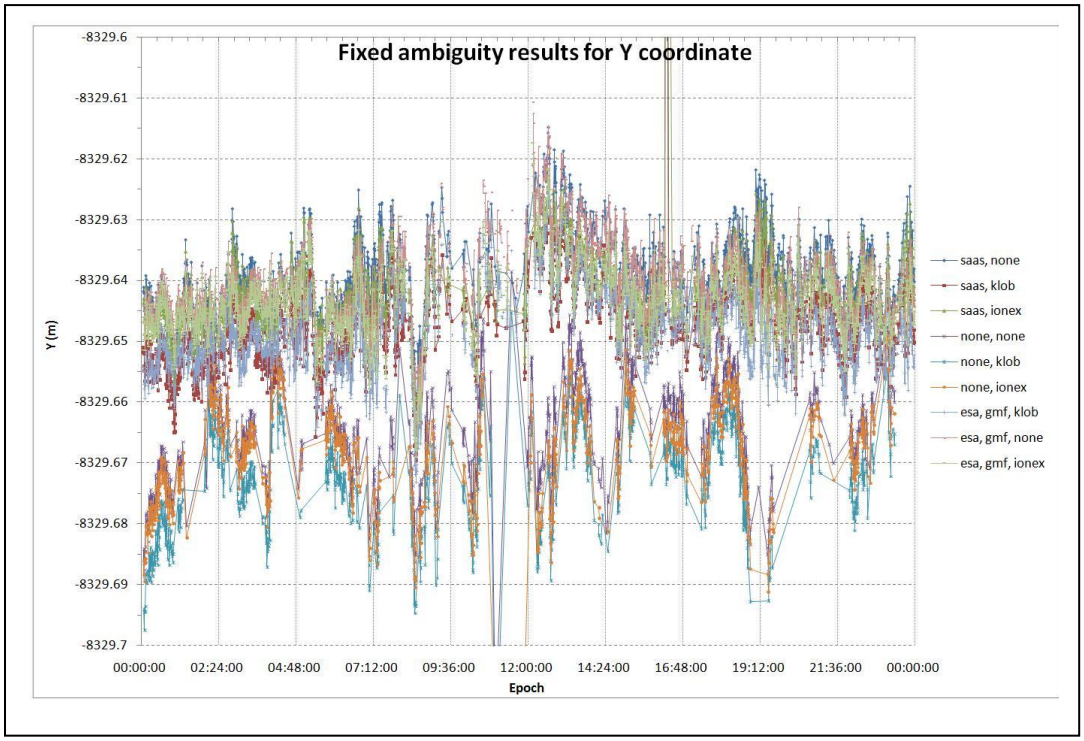


Figure 4.4 – Plot of Y coordinates for LOND rover using different atmospheric model combinations

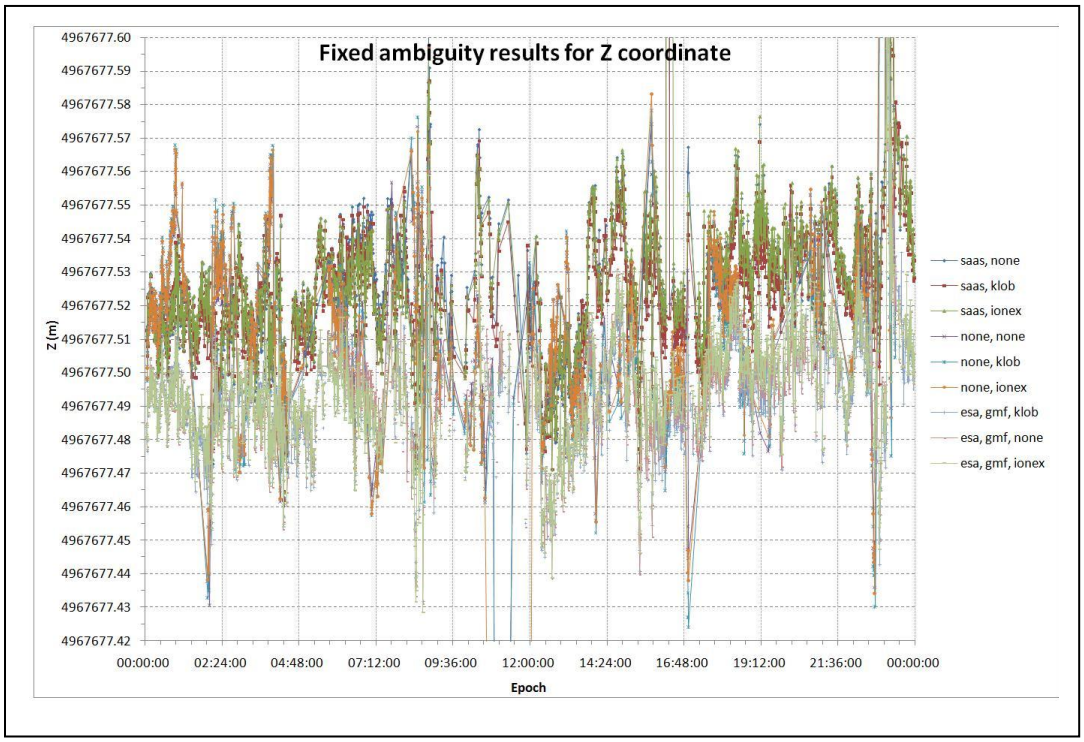


Figure 4.5 – Plot of Z coordinates for LOND rover using different atmospheric model combinations

The results are of interest when looking at the effects of the different atmospheric models, but to see the accuracy, the difference between the known values (supplied by the Ordnance Survey) and those computed. A graph for the X values is shown in figure 4.4, but for analytical purposes, table 4.3 shows the averages for the X, Y and Z coordinate errors for each of the atmospheric model combinations as well as the success rate for ambiguity resolution in each case.

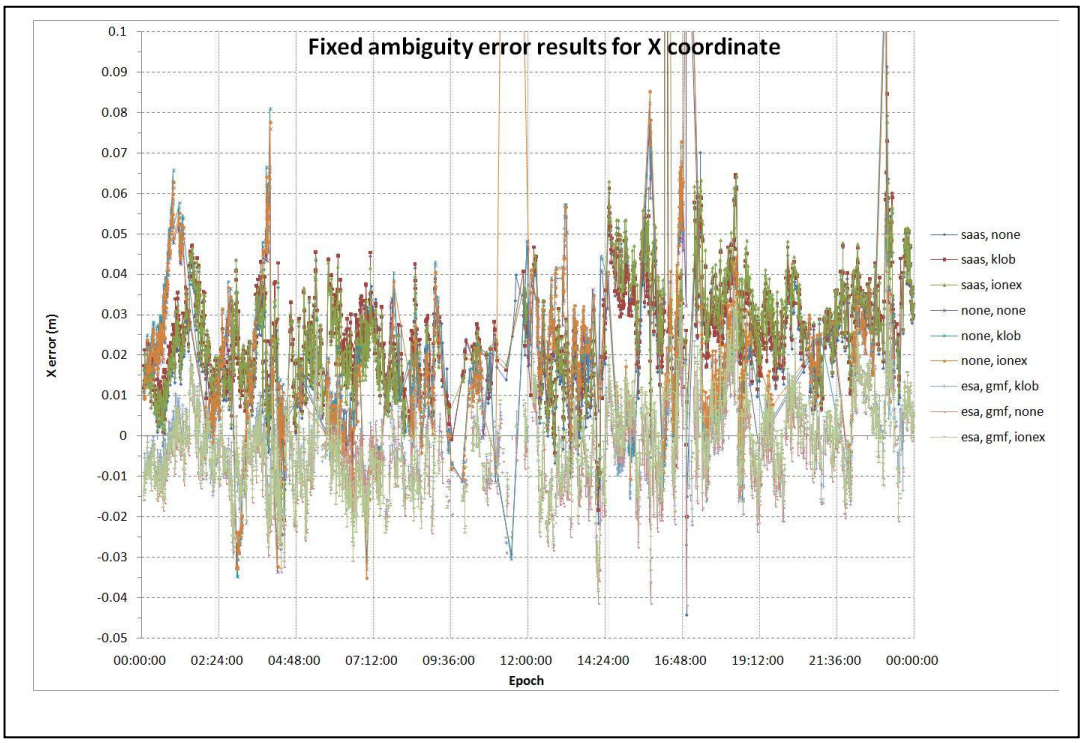


Figure 4.6 – Plot of X coordinate errors for various atmospheric model combinations

Atmospheric models used	Average X error (m)	Average Y error (m)	Average Z error (m)	Vector error (m)	Ambiguity resolution success rate (%)
Saas, None	0.0229	0.0534	0.0011	0.0581	70.7
Saas, Klob	0.0241	0.0448	-0.0003	0.0508	73.7
Saas, Ionex	0.0241	0.0500	0.0013	0.0555	68.0
ESA, GMF, None	-0.0036	0.0519	-0.0309	0.0605	81.0
ESA, GMF, Klob	-0.0026	0.0433	-0.0324	0.0541	82.7
ESA, GMF, Ionex	-0.0022	0.0483	0.0305	0.0571	77.8
None, None	0.0160	0.0273	-0.0143	0.0347	40.5
None, Klob	0.0170	0.0194	-0.0158	0.0302	36.1
None, Ionex	0.0171	0.0238	-0.0162	0.0334	35.7

Table 4.3 – Average coordinate errors and ambiguity success rates for different atmospheric model combinations

From table 4.3 it can be seen that the best method does not necessarily provide the best accuracy and the best success rate. Use of the ESA tropospheric model and

Klobuchar ionospheric model gave the best success rate, but only the 4th best accuracy. The highest accuracy combinations also have the lowest ambiguity success rate which is not ideal.
For our purposes, when seeking to create a high accuracy position solution using the RTK library, the ambiguity success rate is the most important factor and so the ESA/Klobuchar combination seems the best choice in this case.
The standard deviation of the vector error when using the ESA and Klobuchar models is 0.0240 which shows the results are not heavily spread and thus are more reliable than using only the Klobuchar ionospheric model which gives 0.0289.
The possible reasoning for the increase in accuracy by not using a tropospheric correction could be sue to its spatial variance having a longer wavelength and so trying to model it over such a short baseline as this could in fact cause an increase in the error. The low success rate does negate any possible advantage of not using a tropospheric correction and so it is recommended that one is always used for accurate positioning.
If the library was required to work in real time on board a train, the safety case would also have to include the operation times for each of the models used as this would be a factor in possible time to alert (TTA) in the event of a failure or unsafe position reporting. For the LOND-BARK baseline processing in table 4.3, the RTK runtimes ranged from 236 seconds when not using any modelling to 292 when using the EAS tropospheric and ionex ionospheric models. This can be explained by the use of input files by the ionex model adding significant extra processing a memory paging to the overall process. For the interests of this thesis, we will not be concerned with processing times, but it must be considered when applied to a safety critical process where processing times are an important factor.

When checking the data for ambiguity validation, it can be seen that on several occasions the ambiguities seem to have been incorrectly fixed, causing position errors for individual epochs. This is always possible as the LAMBDA method used is not guaranteed to have a 100% success rate due to the stochastic nature of the problem, even using the ratio test to check the results can allow incorrect ambiguity fixes to propagate to an incorrect final position solution.

Smoothing of the data by checking integer ambiguity values for single epoch changes could help solve this problem, but this would also require checking of other variables for alterations, such as a possible change in the reference satellite used.
Figure 4.2 also shows the Amersham (AMER) OS base station which is roughly 37Km from LOND. To show the effect of a longer baseline on ambiguity resolution, the LOND-AMER baseline was processed using the RTK library using the ESA tropospheric model and the Klobuchar ionospheric model. The results gave an integer ambiguity success rate of 53.64% as apposed to 82.7% for the 15Km LOND-BARK baseline. The x coordinate errors for the fixed ambiguity epochs are shown in figure 4.7 below:

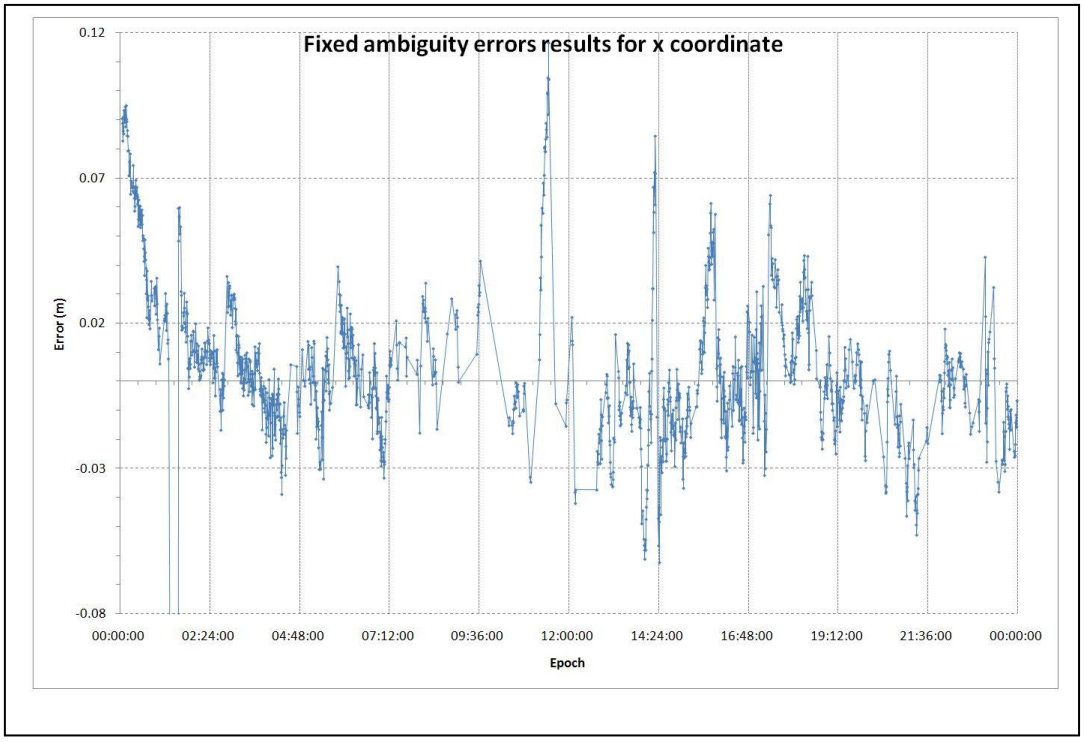


Figure 4.7 – Plot of X coordinate errors for AMER-LOND baseline using ESA and Klobuchar atmospheric models.

The results can be seen to be a little noisier than for the shorter baseline, with the average errors in X, Y and Z being 0.0067m, 0.0606m and 0.0321m respectively, giving a resultant average position error vector of 0.0689m, with standard deviation of 0.07668. The number of incorrectly fixed ambiguities is also higher.

The library is constantly being updated to improve accuracies over longer baselines so that wide area network operations are possible.

Despite the focus of this thesis being centred on the effects on code measurements in the rail environment, for multipath values to be calculated using geometric range and observed range a high precision position for the receiver must be calculated first and so the use of the RTK library is required

5 Data collection exercises

5.1 Introduction

For the effects of the UK rail environment on GPS signals to be understood, real world experimentation must take place as the highly complex nature of the surrounding infrastructure would prove very difficult to model in enough detail. We are interested in analysing data collected in three of the main types of environment found in the UK rail network. These are:

- Highly rural – trees and foliage being the main error sources
- Electrified – overhead cables and supporting infrastructure being the main error sources
- Highly urban – buildings and inner city infrastructure being the main error sources

For each of these situations, experiments have been designed and locations specifically chosen to give the best examples of each. The rural location is based in the Severn Valley where access has been granted to the recreational Kidderminster to Bridgenorth line. This is also the first, early dataset recorded so that an understanding of kinematic train data can be gained in preparation for the later exercise.
The electrified west coast main line was chosen due to the overhead electrification and supporting structures for the electrified environment experiment, this is not kinematic but is purely designed to check for possible interference. The urban environment chosen is located in the Birmingham new street central depot area and is part of a larger project called LOCASYS (Cross, 2007). Details of each site and the data collected are given in the following sections 5.2 to 5.4

5.2 Severn Valley railway

5.2.1 Track location and surrounding area

The Severn Valley railway is a 26 Km line that runs from Kidderminster to Bridgenorth, following the course of the river Severn.

The line was originally used for commercial transport for 101 years until 1963 when the profitability of the line was deemed impractical. The line is now run as a heritage railway by the Severn Valley Railway Company which has over 250,000 visitors a year travelling on the variety of steam and classic diesel locomotives owned by the company.

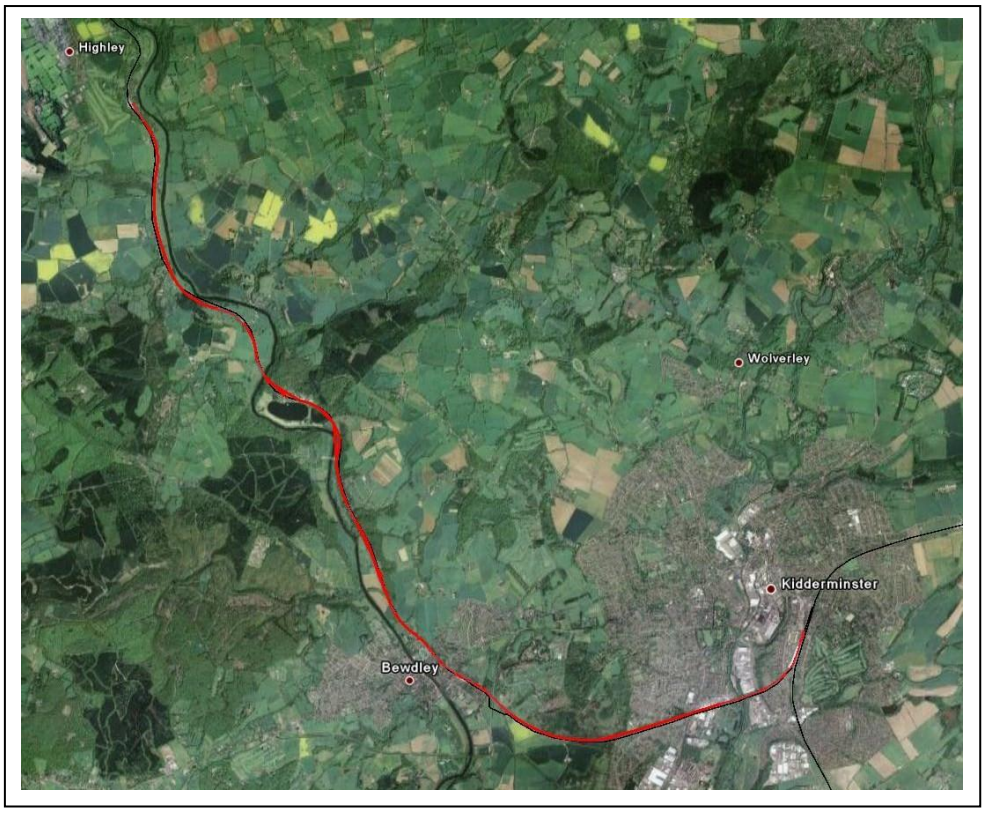


Figure 5.1 – Map showing Severn Valley Railway

The line has several features desirable when looking for examples of a rural rail environment. The line runs through both deep and shallow cuttings, through open valleys and also over hills and bridges. The line is mostly lined with trees; in some cases the trees form a complete canopy over the train line which is ideal for analysing their effects on GPS measurements. There is also a short tunnel in the Bewdley section of track.

Due to the operations of the line being purely for recreational use, access to the train was far more relaxed than that of a commuter train. The cooperation of staff proved to be invaluable when collecting data as access was granted not only to mount an antenna on the train, but also to set up a local base station in the nearby mechanics yard to facilitate the use of localised GPS corrections at the same epoch rate.
There are no overhead power cables and no other trains operating on the line which is 4ft 9½ inches wide and begins with two way sections and ends as a single track.
In the near vicinity to the track there were no high rise buildings with only the station depot buildings, water refilling tank and signal box being of significant height.

Figure 5.3 – Kidderminster platform and depot

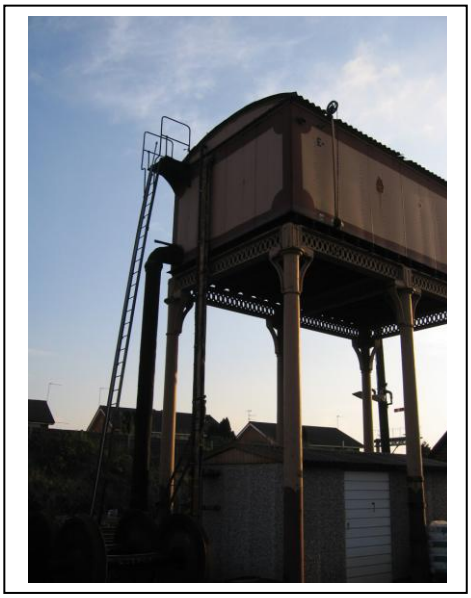


Figure 5.2– Water refilling tank

Figure 5.4– Main Kidderminster depot open area

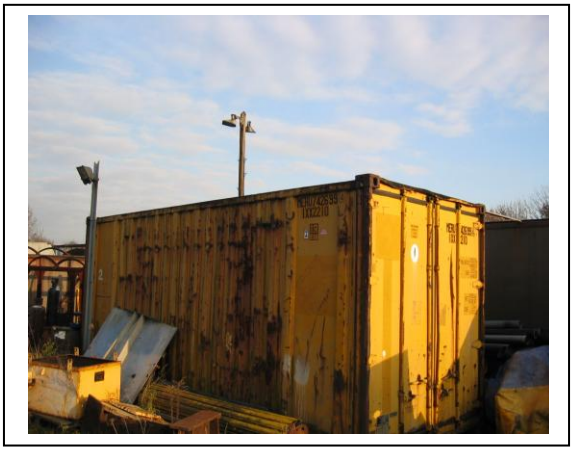


Figure 5.5– Container used to mount reference station receiver.

The data was collected on the 21st of December 2004 in dry, cold weather when the train was in full use transporting visitors from Kidderminster to Highley and back again. This was an ideal opportunity to collect data in a rail environment so that the 'quality' of the data could be assessed as well as any processing aspects specific to this as yet untested environment.

5.2.2 Equipment configuration

The receivers used were departmentally owned Leica system 500 units with manufacturer rated (Leica, 1999) 5cm rms code measurement precision using a narrow code correlator spacing.
The antennas used were Leica AT502 dual frequency units connected via coaxial cable.

The train mounting used consisted of a Leica standardised 2m pole attached to non heated service pipes with plastic zip-ties as shown in figure 5.6. The coaxial cable was then routed down the pole and along the outside of the train to the next carriage where it enters the window to the guards van as shown in figure 5.7.

Figure 5.6 – Antenna mounted to side of train

Figure 5.7 – Cable routing along side of train

The carriage skin was made of paint and lacquer covered plywood over a structural layer of pine slats. The roof is made of thin sheet steel covered in waterproofing membrane sealed with tar.

The guards van was chosen as the location for the receiver as it was not in use so caused the least amount of disturbance to the normal operation of the service as well as allowing easy manoeuvrability if required.

Figure 5.8 – Receiver and power pack located in guards van

The reference receiver was located in the nearby maintenance yard, also owned by the Severn Valley Railway company. The Leica receiver and antenna was mounted on top of a survey grade tripod (see figure 5.10) which was placed on top of an 8 foot high shipping container located in the yard so as to have the best possible, unobstructed view of the sky. The layout of the Kidderminster end of the line and depot is shown in figure 5.9 below with the reference receiver position marked. The train speed was lower than a conventional commuter train, peaking at around 30mph.

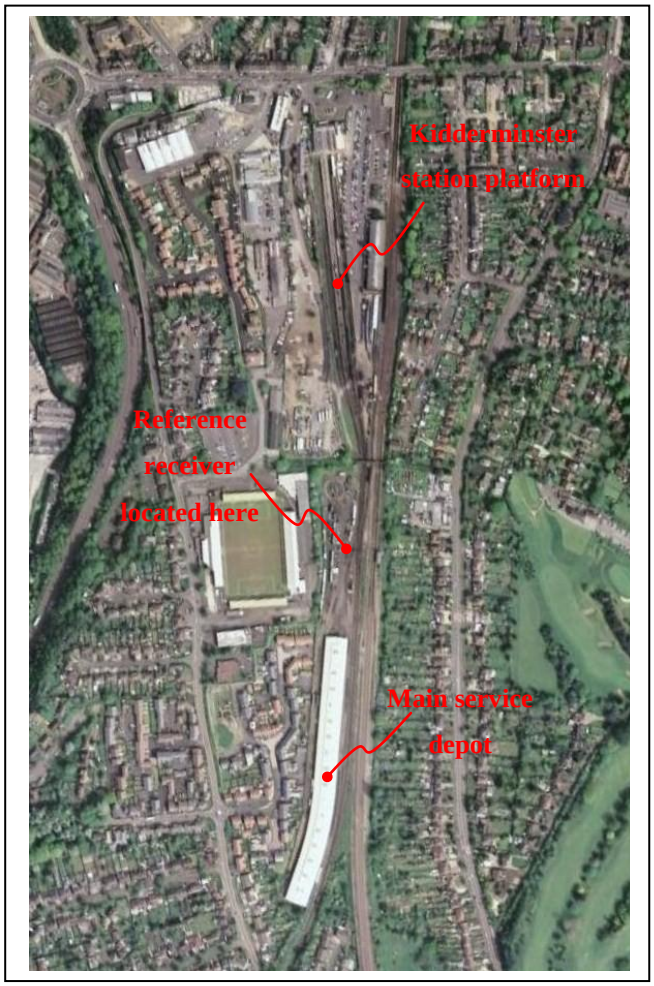


Figure 5.9 - Kidderminster station and depot layout

Figure 5.10 – Tripod mounted Leica reference receiver and antenna

The reference receiver was set to a 0.5 second epoch rate and left for the day to collect data so that any rover data collected would fall inside the operating 'window' of the reference receiver. The extra data collected outside the rover operating time can then be deleted to avoid unnecessary processing time.
The rover receiver on the train was also set to operate with a 0.5 second epoch rate but without the need for a static initiation so that kinematic initialisation could occur if the need arose.
Dual frequency code and phase data was collected in the receivers own ASCI binary format which was then post processed using Leica geo office (formally Ski-Pro) to produce RINEX 2.1 files for processing in the RTK library.

5.2.3 Data sets produced

The Rinex files produced are summarised in Table 5.1 below.

Detail	Rover	Reference	Reference (cut down)
File size	5.5Mb	19.1Mb	5.5Mb
Start time	12:27:6.5	09:01:50.0	12:27:6.5
End time	13:38:54.0	14:01:12.0	13:38:54.0
Total no. epochs	8614	359900	8614
No. epochs with 4 or more satellites	6607	359900	8614
% ambiguity fix when 4 or more satellites	26.5	N/A	N/A

Table 5.1 - Table showing Rinex file details for Rover and Reference

It can be seen that the beginning and end of the cut down reference station file has been removed so that skipping of the epochs is not required by the RTK library when trying to process double differences with the rover file. The reference receiver has enough satellites in view for at least 4 in every epoch so that a position can be calculated. This is not the same for the rover where 2007 epochs have below 4

satellites in view, some of which are during the time when the train is inside Bewdley tunnel.

Figure 5.11 shows a sky plot for the time when data was collected using Leica satellite availability software. A 10° cut off angle has been imposed as stochastically speaking satellites below 15° on the horizon tend to have a higher level of multipath and signal degradation but we wish to see as many available satellites as possible. By plotting the availability of satellites against time, we can see which satellites should be in view at any one time, provided they are above the minimum elevation of 10°, as shown in figure 5.12.

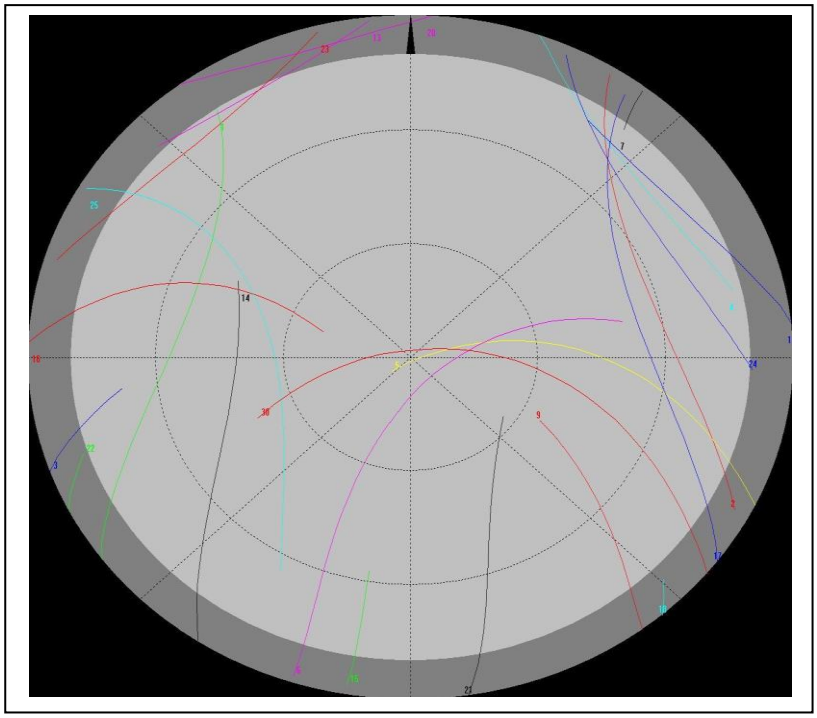


Figure 5.11– Sky plot for Kidderminster showing satellites visible from 9am to 2pm on the 21st December 2004. Dark grey band represents 10° cut off.

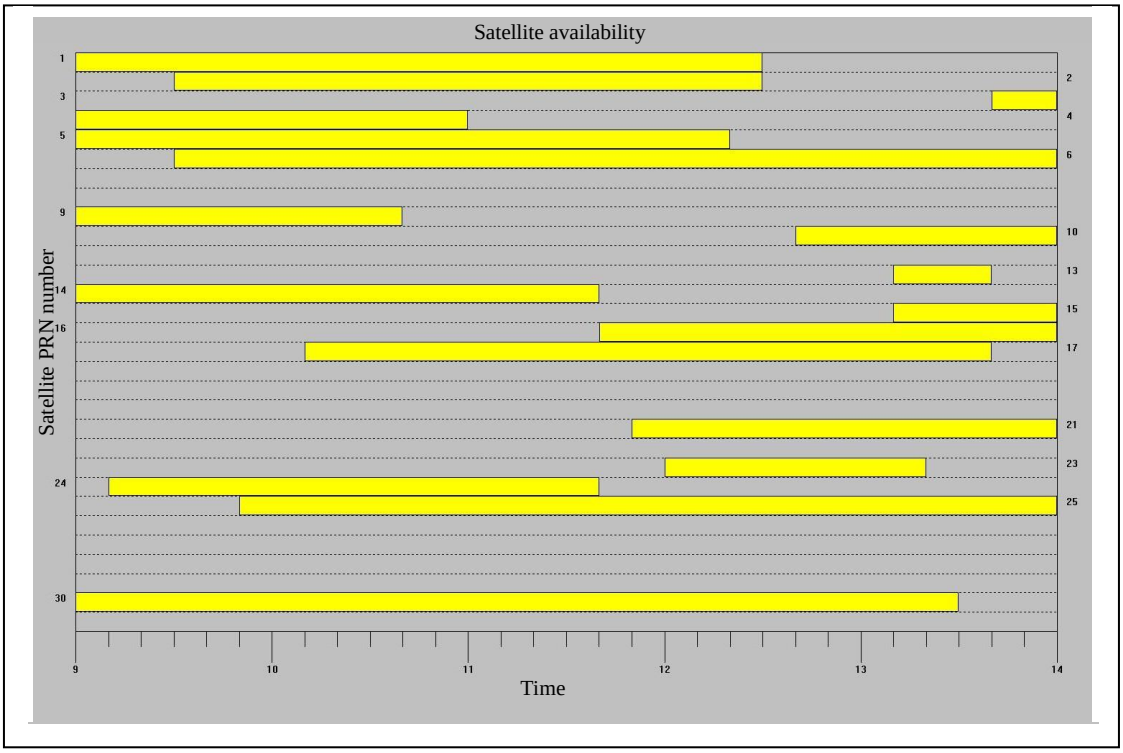


Figure 5.12 – Satellite availability plot for Kidderminster

As these plots can be calculated using forward prediction of orbits, they can be used to select an ideal time for data collection if particular satellites are of interest, or if simply the largest number of satellites is required.

For this location it can be expected that there are always at least 6 satellites in view, with 7 for the most part when using an orbit prediction program for the data collection date.

The epochs when rover data was collected can be seen to have at least 8 satellites, in some cases rising to 10 so this was considered an ideal time during the trains operation to collect data as a large number of satellites are available.

5.3 Watford Junction

5.3.1 Platform location and surrounding area

Watford Junction is located to the north west of central London roughly 10Km inside the M25 motorway.

The main train station has over 4.5 million commuters annually and looks to increase with the station being overhauled and upgraded by the year 2009 with the possibility of added links to the London Underground.

Figure 5.13 – Overview of Watford junction rail lines and local area

The station has 12 platforms and is in constant use by the public. Special permission was sought through Silverlink trains, Network Rail, C2C Rail and IVECO for access to both the platform and the roof of the station building (occupied by the HGV company IVECO).

For access to the platform to be granted, a brief safety case for the data collection process had to be compiled and formally submitted for scrutiny by the Silverlink property manager. Stipulation for the personnel involved to have full and valid Personal Track Safety (PTS) cards was also imposed requiring attendance of a two day lecture and testing course.

Once all safety procedures were completed access was granted the dates chosen for data collection were the 2nd and 3rd of November 2006. Two days were chosen so that any effects that were noted could be corroborated with the second days data in order to remove possible freak events.

By utilising the exact daily repetition (minus 4 minutes) of the satellite geometry, any effects from the stationary surrounding environment should be seen to repeat themselves.

Because the station was in constant use, the receiver used on the platform had to be constantly guarded as when trains pass through, the vibration of the ground and suction caused by the trains speed (some pendolino tilting trains pass through at near 200Km/h) could cause the tripod to move or even fall over.
The west coast mainline station platforms are on a rough bearing of 315° and are open to the sky with a clear field of view to roughly 15° elevation when close to the main entrance, though on the south side the main station building housing IVECO is a large block to the sky as it has several storeys. This will need to be accounted for when analysing the data produced as it could be a large source of signal reflections.

Figure 5.14 - View of main platform looking north

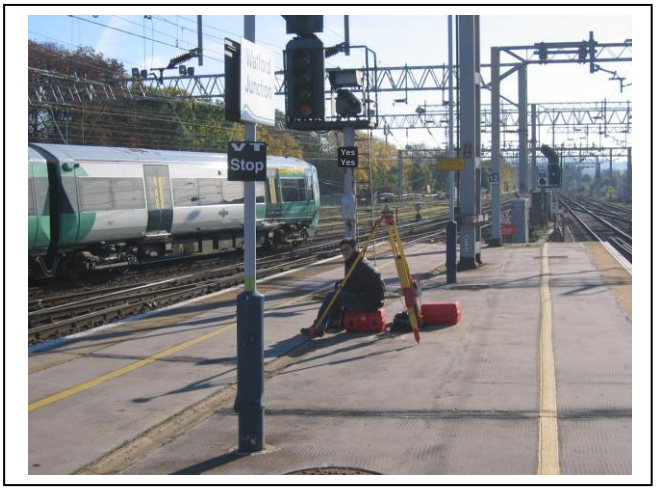


Figure 5.15– View of platform and receiver facing south

As can be seen in figure 5.15 to the south there are a large number of possible reflectors in the near vicinity to the receiver such as sign boards and a signal post. When looking to the north and north west of the receiver, the area is clear from obstructions, with only the main station building being a possible source of major reflections.

The layout of the surrounding cables and platforms was recorded in a simplified form and is shown in the following figure.

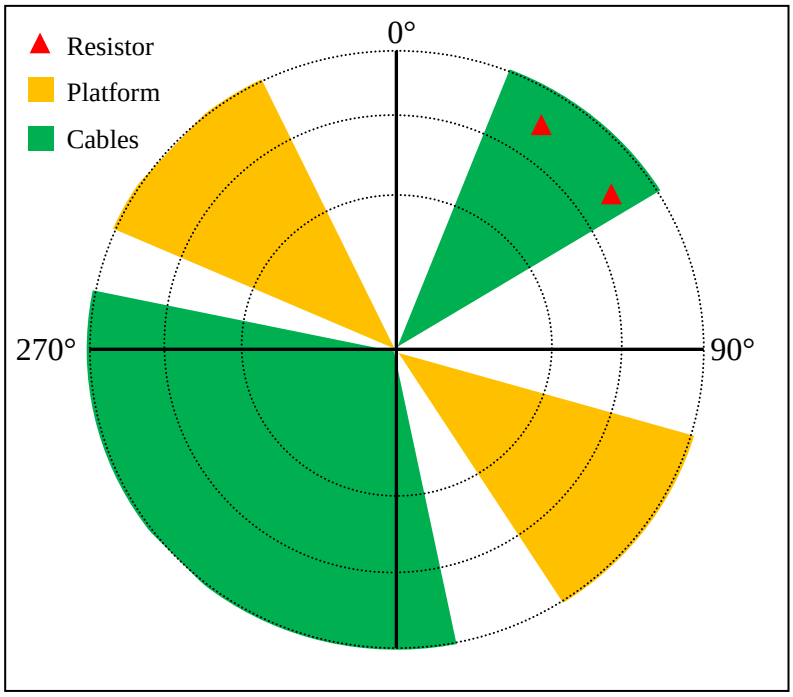


Figure 5.16 – Diagram showing receiver view of surroundings on platform

The northern cable section is also shown in the following figure; note the horizon of trees and buildings at roughly 7° elevation angle and supporting infrastructure at 60° azimuth.

Figure 5.17 – Picture showing north section of overhead catenary cables

When a train is either stopped beside the receiver or passing through the station, the rough elevation mask of the train does not exceed 14° within the cabled area of interest towards the north of the platform.

5.3.2 Equipment configuration

The receivers used were the same departmentally owned Leica system 500 units used for the Severn Valley data collection exercise. The antennas used were also the same Leica AT502 dual frequency units.

For the required observations to take place, a rover receiver was positioned on the end of the platform on a survey grade tripod as shown in figure 5.17 and a reference receiver was placed with the approval of IVECO on top of the main train station building mounted on a tripod with anti-perforation cups as shown in figure 5.18.

Figure 5.18 – Reference receiver on station building roof

The reference receiver has a full clear view of the sky without any obstructions as the building is the highest compared to all other visible buildings on the horizon.
The relative positions of the rover and reference receivers can be seen in the overhead picture in figure 5.19.

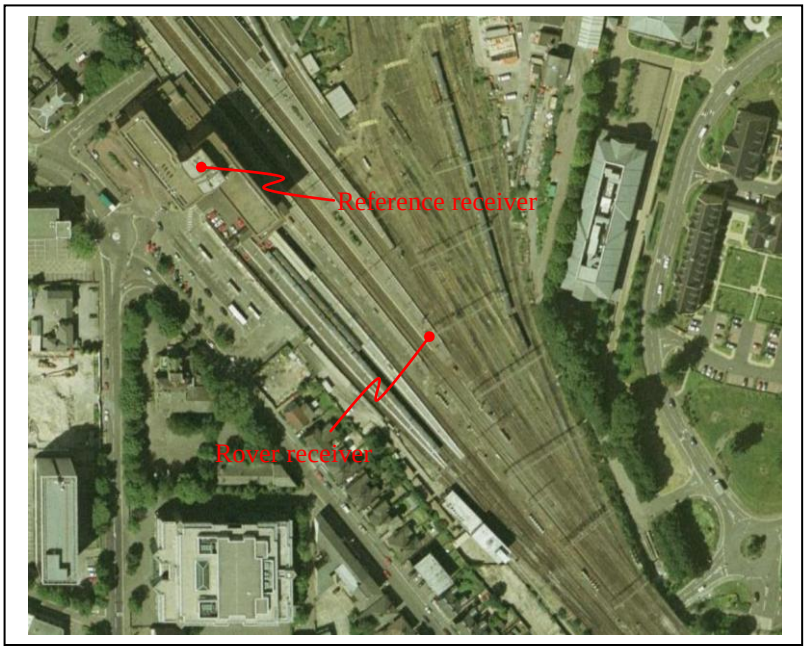


Figure 5.19 – Plan view of Watford Junction train station showing receiver locations

The baseline between the receivers is roughly 130 meters on the horizontal and so the spatially correlated atmospheric errors that can effect network RTK measurements are going to be minimal leaving only the effects of the highly localised environment.

Both the reference and rover receivers were set to log with one second epochs with the reference being set up first and shut down last so that the rover fits within the 'window' of the reference stations operating time. This was repeated on the two consecutive days at roughly the same time.
As before, dual frequency code and phase data was collected in the receivers own ASCI binary format which was then post processed using Leica geo office to produce RINEX 2.1 files for processing in the RTK library.

5.3.3 Data sets produced

Detail	Rover (platform) 2/11/06	Rover (platform) 3/11/06	Reference (roof) 2/11/06	Reference (roof) 3/11/06
File size	8.42Mb	11.0Mb	10.23Mb	11.3Mb
Start time	11:18:0.0	10:38:14.0	10:45:24.0	10:7:14.0
End time	15:31:37.0	15:37:23.0	15:51:20.0	16:1:57.0
Total no. epochs	15217	17949	18356	21283
No. epochs with 4 or more satellites	15217	17949	15217	21283
% ambiguity fix when 4 or more satellites	95.8%	93.5%	N/A	N/A

Table 5.2 – Table showing details of files produced by receivers on 2nd and 3rd of November 2006

The reference station datasets recorded on the roof were cut down to match the platform rover receiver files so as to reduce unnecessary processing time.
When looking at predicted availability, the sky plot (figure 5.20) shows that a large number of satellites are low on the horizon during the time when data collection takes place.

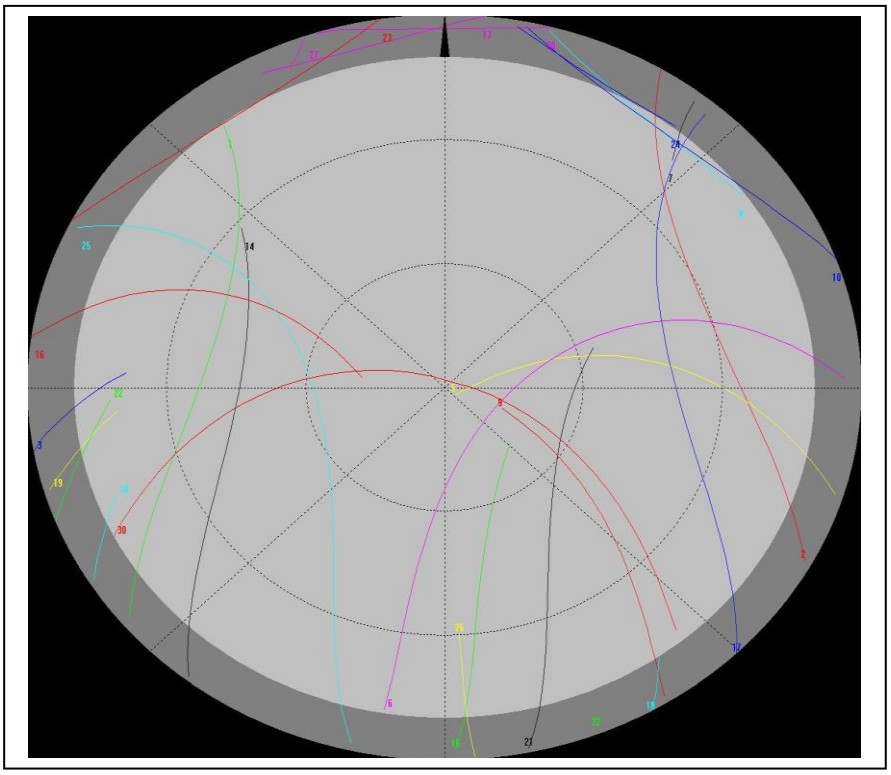


Figure 5.20 – Sky plot for Watford junction on 2nd/3rd November 2006, 10° elevation cut off shown by dark grey area.

The availability of specific satellites can be seen below, showing the fragmented visibility of some satellites with lower elevation. The number of satellites available during the data collection period does not drop below 6 and reaches a maximum of 11 for a short time during the first half hour.

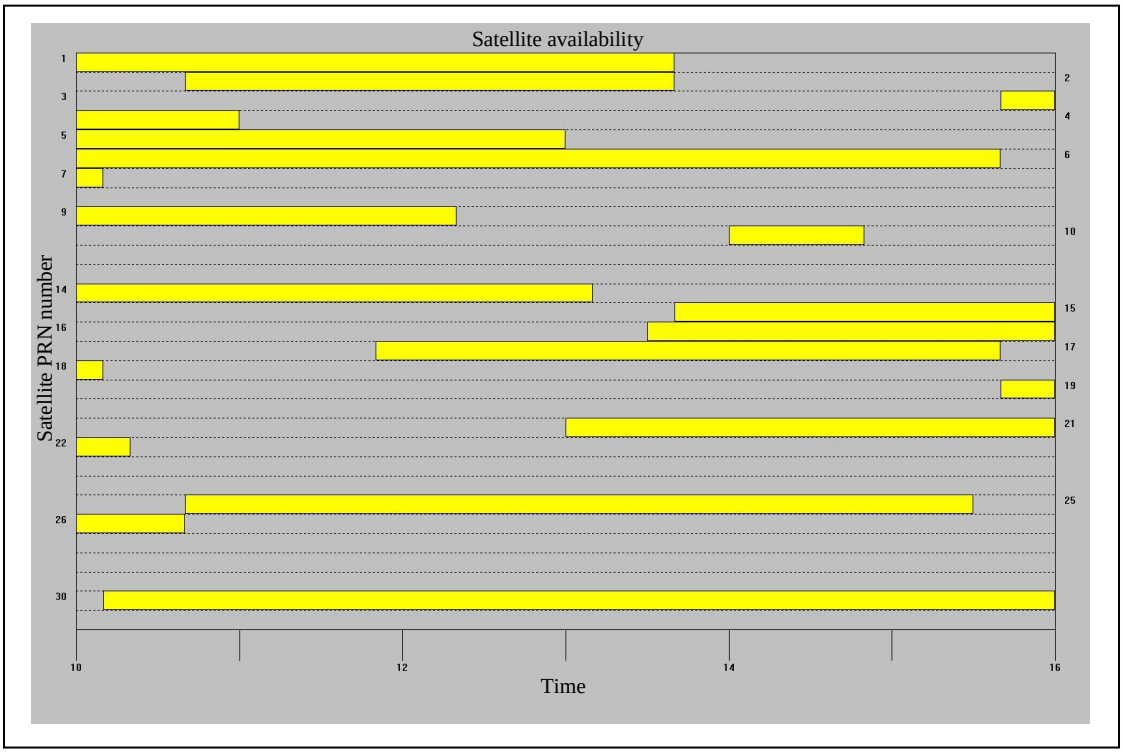


Figure 5.21 – Satellite availability plot for Watford Junction

When looking at the availability of satellites, it looks as though the first half would be the best to record data, but given that the purpose of the investigation is to ascertain the possible effects of EM radiation from overhead power lines, the period of investigation that is of most concern is when satellites are in a low enough elevation and within a certain azimuthal range that the signal propagation path coincides with the relative location of the overhead lines.

When comparing figure 5.20 with figure 5.16, the ideal candidate satellites are two subsets depending on whether the northern set of overhead lines (complete with resistors) are of interest, or the southern set of overhead lines are being investigated. For investigating the overhead lines and resistors to the north of the platform, satellites 2, 4, 10, 17 and 24 are of interest whereas for looking at the southern set of cables, satellites 1, 6, 14, 25 and 30 are of interest.

The sky plot for these specific satellites is shown below with the areas of interest marked on for reference.

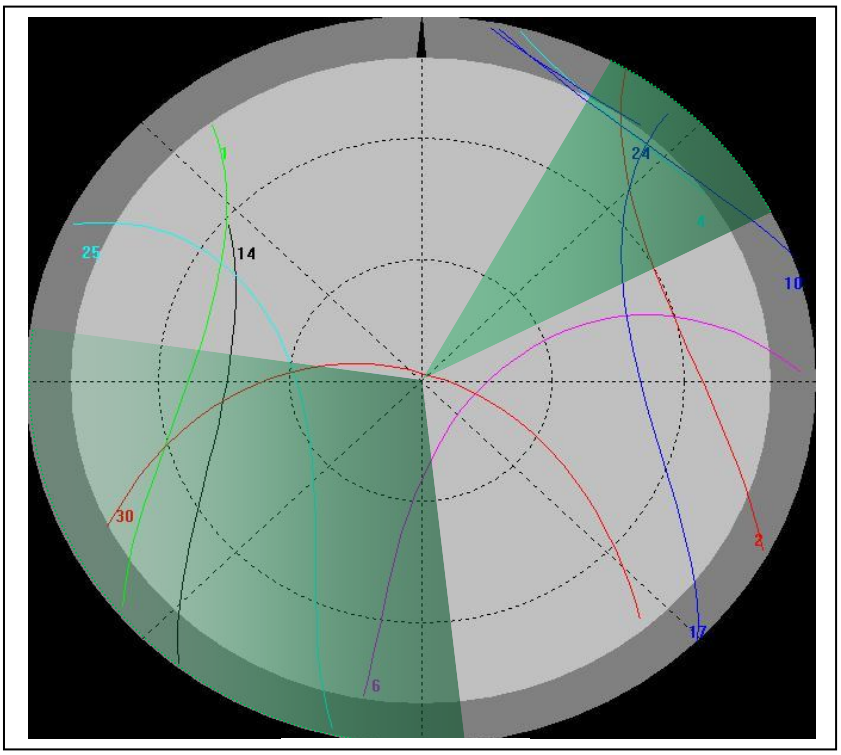


Figure 5.22 – Sky plot showing selected satellites and areas of interest highlighted in green

The northern section of interest contains only low elevation satellites, generally below a 30° elevation angle. The southern section contains satellites passing through a full

range of elevations with satellite 25 passing through the centre of the section in a descending elevation pattern.

This diagram will help in determining which satellites will be of interest and at what period during their visibility are of concern.

5.4 Birmingham New Street line

5.4.1 Track location and surrounding area

Birmingham New Street station is located in the centre of the city of Birmingham and lies on the Birmingham loop of the west coast main line.

The station is a major hub and has over 31 million passengers passing through every year. The current station was redeveloped in the 1960's when the electrification of the west coast main line took place and is due to be redeveloped again under the Gateway Plus scheme.

A map of the Birmingham rail network is shown below:

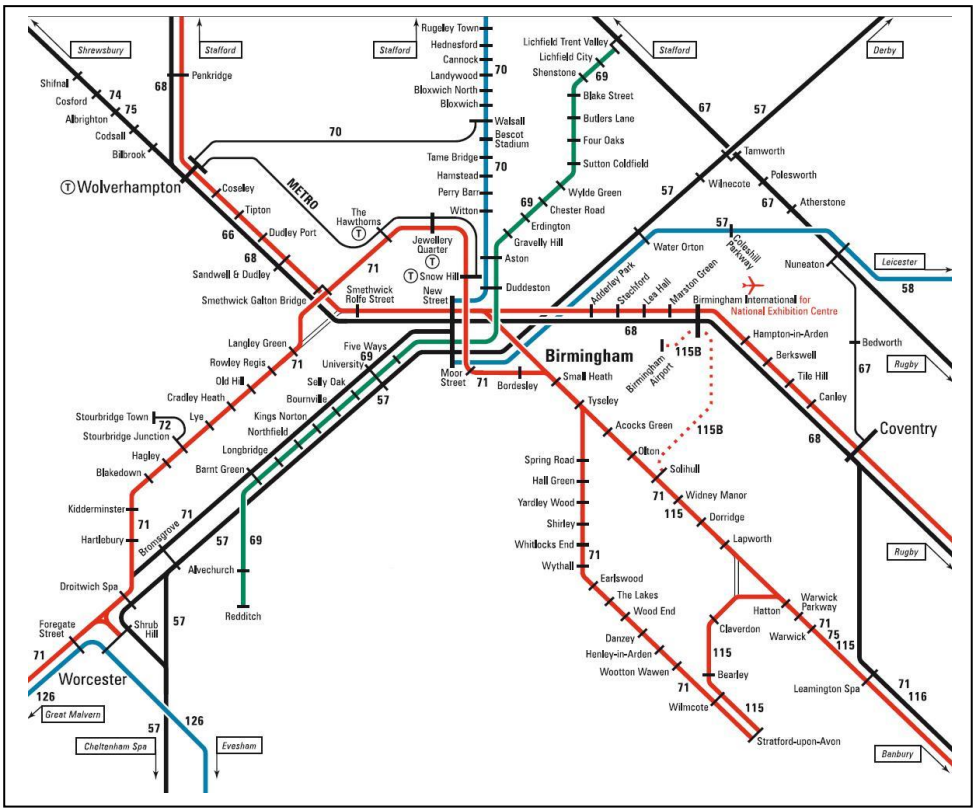


Figure 5.23 – Map of the local Birmingham rail network.

It can be seen that all the local lines pass through New Street station, causing the 12 platforms to be very busy.

The main lines that data will be collected on will be the green Lichfield to Redditch line and the Birmingham New Street to Wolverhampton line.

The lines used have a mix of rural and urban surroundings, with the latter going to the extremes of the New Street station area where high rise buildings, office blocks and overhead infrastructures such as roads and catenary line supports surround the lines, as can be seen in figure 5.24.

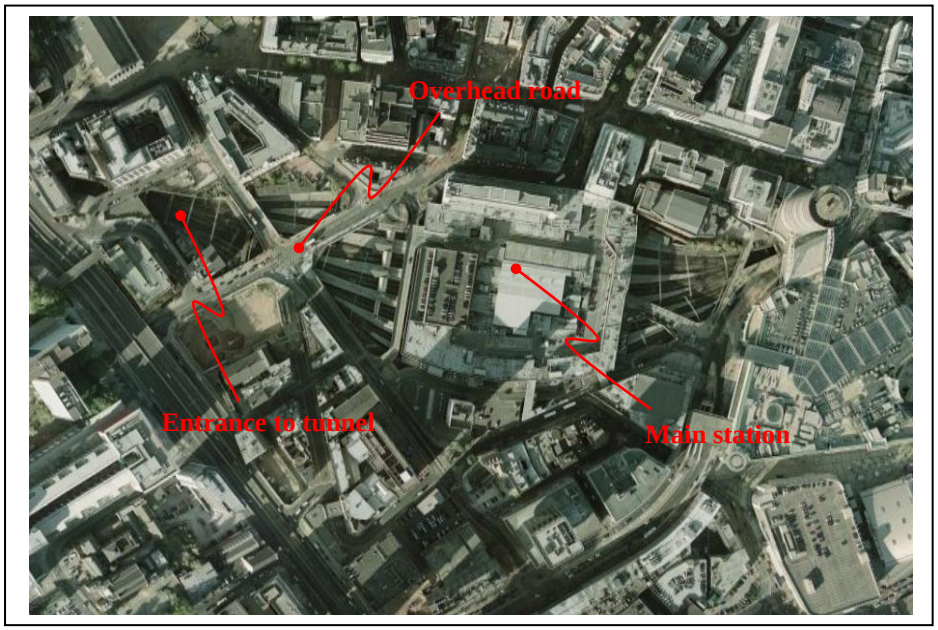


Figure 5.24 – New Street station overhead view

There are also a number of tunnels leading up to New Street to accommodate the overhead canals, roads and buildings, as shown in figure 5.25.

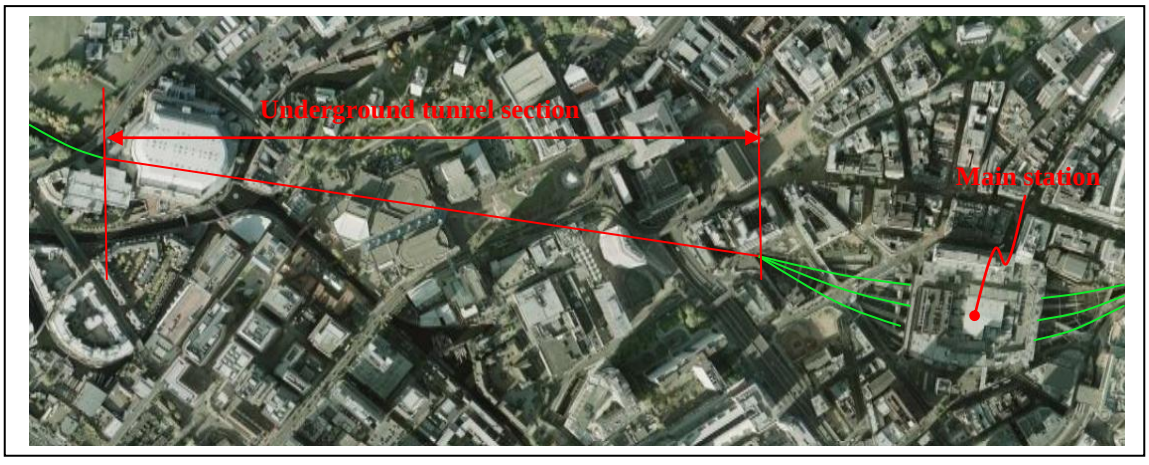


Figure 5.25 – Overhead view of main New Street tunnel section

The routes used run for roughly 16 hours a day with the 1 hour trip being repeated 7 to 8 times a day. The train used is a normal revenue generating train so as to maximise

the amount of data collected during the 36 month operation period for the larger NSL run LOCASYS project (Cross, 2007) of which this work is an introduction. The length of the major project is such that changing weather seasons can be documented, such as using GPS to check on-board odometers, allowing wheel slip due to leaf fall to be quantified.

5.4.2 Equipment configuration

The data collected is from multiple sensors and receivers mounted on the train, as well as corresponding base station data from the local Ordnance Survey active base stations.

The train mounted sensor system was built by Bangor University and designed to occupy a minimal footprint for installation in the passenger cabin under a seat.

Figure 5.26 – Sensor system installed on train under seating in main passenger compartment

Figure 5.27 – Sensor system built by Bangor University

The sensor unit contains a dual frequency geodetic quality NovAtel OEM-V3 receiver producing raw measurements in a binary format that can be converted to RINEX.

There is also a combined INS/GPS (single frequency) Microbotics MIDG unit that provides combined INS/GPS position solutions as well as raw INS measurements at a maximum of 50Hz.
Due to the Baud rate being maxed out by the raw INS data stream from the combined unit, a third single frequency NovAtel GPS receiver is also installed.
An analogue to digital converter card that allows the unit to receive information from the trains onboard systems is interfaced with the ignition system so that the moment when the driver turns on the train, the system can begin logging. The trains on-board speedometer is also fed into the system so that the unit can be used to check the calibration of the trains tachometer system.

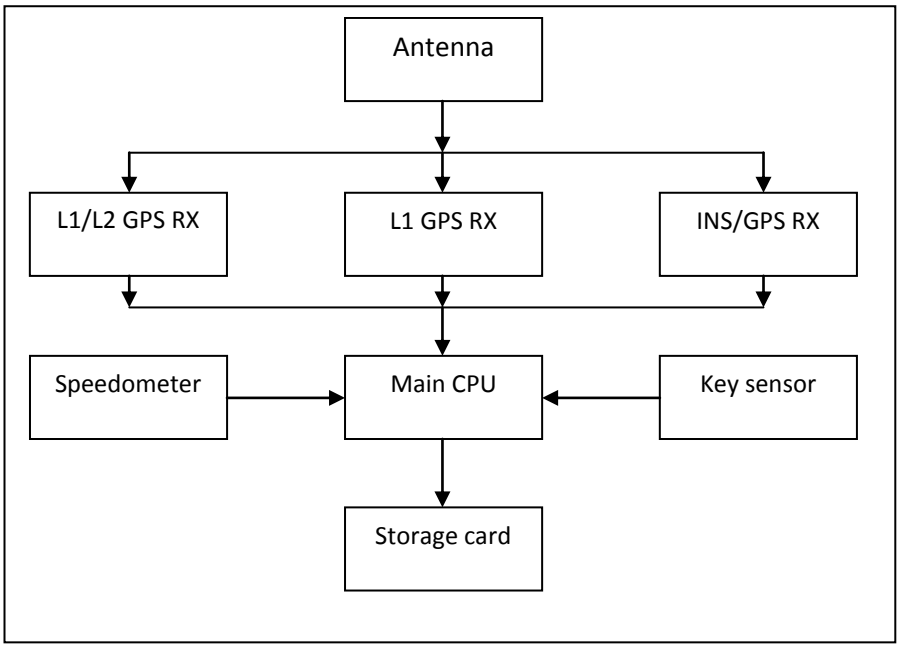


Figure 5.28 – On board sensor and logging system configuration

The unit also contains an integrated GSM modem so that status updates can be automatically transmitted for easy monitoring of the system as a whole as well as the memory status of the integrated memory card adaptor used to store all the recorded data.
The compact flash card used to store the data can run for several weeks before the data needs to be downloaded and the card wiped clean. This is ideal as access to the unit is only possible when the train is in the depot due to the security and safety concerns for the travelling public.

The Antenna used for the GPS signals is a low profile Aeroantenna AT2775-41 aviation grade dual frequency antenna connected by a split feed coaxial cable enabling the three receivers to use the same antenna.

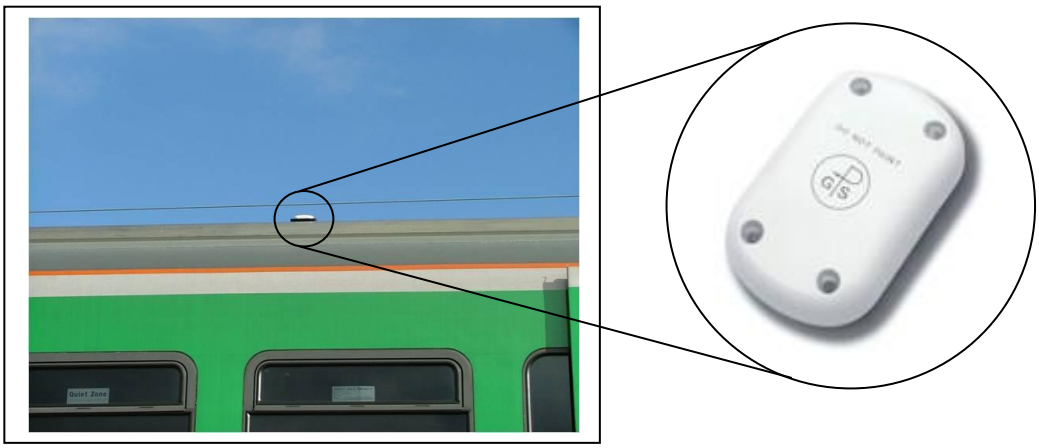


Figure 5.29 – Picture showing roof mounted low profile GPS antenna

The trains used in the Birmingham area are Hunslet TPL 3 vehicle electric suburban trains as shown in figure 5.31 below with a maximum line speed of 75mph. The trains are in constant public use and so for safety all items installed must be checked for possible EMC emissions.

Figure 5.30 – Sensor system undergoing EMC testing

Figure 5.31 – Hunslet TPL train as used in this experiment

The epoch rate used for the GPS data logging was 1 second with the INS raw data collection rate being set at 50Hz and with a maximum line speed of 75mph, this allows a position at roughly every 33m when at full speed, but given the more common speed of around 20-40mph in urban areas, an average of 13m between positions is more common, close to the accuracy levels in the RNP's listed in section 3.5 and so given the positions plotted are so close to the accuracy requirements, along track errors nearing 10m will be clearly seen.

In order to make full use of this 1 second data, the base stations used must also have 1 second data available if double differenced position solutions are going to be calculated.

The Ordnance Survey OS net GPS network not only provides 30 second epoch rates, but on request 1 second data rates are available. There is no base station (surprisingly) in Birmingham itself, but there are in the neighbouring towns of Church Lawford (CHUH), Droitwich (DROI), Lichfield (LICH) and Shrewsbury (SHRE), all of which produce 1 second epoch rate RINEX files.

The baseline lengths and geographical layout of these stations relative to Birmingham New Street can be seen in figure 5.32.

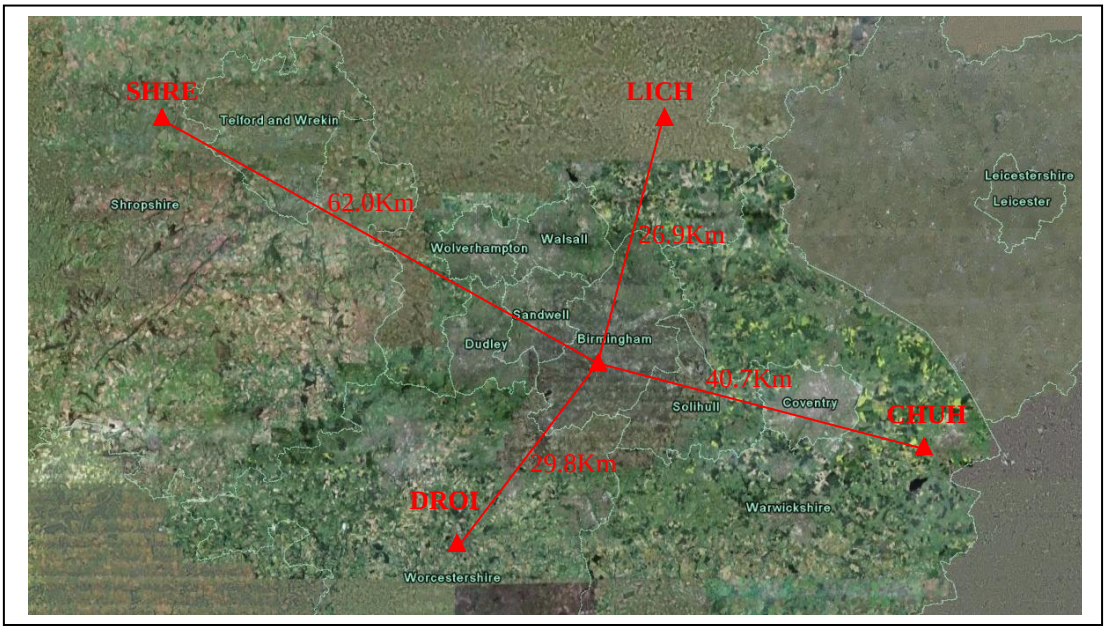


Figure 5.32 – Map showing OS net stations and their distances to Birmingham New Street

The coordinates of each of the stations can be seen in figure 5.35 below.

Station	X coordinate	Y coordinate	Z coordinate
CHUH	3902468.454	-90706.003	5027408.638
DROI	3909833.011	-147097.137	5020322.461
LICH	3870544.458	-122565.380	5051157.820
SHRE	3867417.468	-183495.554	5051679.744

Table 5.3 – Table showing OS net base station ETRS89 cartesian coordinates

All the stations listed use a Leica SR530 receiver and a Leica LEIAT504 antenna apart from DROI which uses an Ashtech ASH700936E antenna.
All antenna offsets are contained in the ANTEC file used in the RTK library and so the phase offsets are accounted for when calculating differential positions.

5.4.3 Data sets produced

The data sets produced by the combined system vary in size depending on the active period during each run of the train. The RINEX data produced by the dual frequency NovAtel receiver is of prime interest to this thesis and so this will be what is discussed. The combined INS/GPS data is also available for analysis as is the single frequency receiver data, but the use of this is limited as the need for dual frequency GPS observables is required for the formation of the code multipath observables.

The data received from the on board speedometer cannot be used due to the interface between the sensor system and the train incorrectly functioning. This situation is to be resolved, but due to the time constraints of this thesis the results will not be available in time. The data collection exercise will carry on as the LOCASYS project continues for over 12 months (to allow for a full range of seasonal variations to be experienced by the equipment)

The initial data sets produced were for the dates and journeys listed below:

02/11/07: Return trips between Birmingham and Wolverhampton plus a trip from Birmingham to the depot.

03/11/07: Return trips between Birmingham and Redditch/Lichfield (or other intermediate stations) plus 1 return trip between the depot and Birmingham

04/11/07: 1 trip from Birmingham to the depot

05/11/07: Mostly between Birmingham and Redditch/Lichfield (or other intermediate stations) with 1 return trip between Birmingham and Wolverhampton and a single journey to Wolverhampton

06/11/07: Mostly between Birmingham and Wolverhampton with a single return journey between Birmingham and Longbridge and a single return journey between Birmingham and Coventry

08/11/07: Return trips between Birmingham and Redditch/Lichfield (or other intermediate stations) plus 2 return trips between the depot and Birmingham and a single trip from Birmingham and the depot.

09/11/07: Return trips between Birmingham and Redditch/Lichfield (or other intermediate stations) plus 1 return trip between the depot and Birmingham

10/11/07: Return trips between Birmingham and Wolverhampton and between Birmingham and Walsall with a single trip from the depot to Birmingham

13/11/07: Return trips between Birmingham and Redditch/Lichfield (or other intermediate stations)

14/11/07: Return trips between Birmingham and Redditch/Lichfield (or other intermediate stations) plus 1 return trip between the depot and Birmingham

15/11/07: Return trips between Birmingham and Redditch/Lichfield (or other intermediate stations)

These RINEX files and the corresponding OS base station files had a total size of roughly 3.2Gb with a daily file size being roughly 264Mb. The Train RINEX files

were on average 60Mb for a full day of operation which lasted roughly from 5.30am to midnight the same day.
This gives an average number of daily epochs as 66,600.
Because of the continuing nature of the data collection exercise, the data sets will not be discussed in detail, with the main discussion being based on the analysis and results obtained.

When looking at the number of satellites available to the train during some of the above listed runs, the visibility is seen to be very good with observed satellites reaching as many as 13 in some cases.
Obvious breaks can also be seen in the data due to various obstructions and tunnels, but these will be further discussed in section 6.4.

6 Data analysis

6.1 Introduction

Dual frequency code and phase data has been recorded for all three rail environments so that investigations into their effect on GPS signals can be as thorough as possible. Multiple analysis techniques are used to identify individual attributes in the rail environment and their specific effects.

By utilising the current abundance of satellite imagery, positioning data can be plotted straight onto overhead images of the local area being investigated. This can facilitate visual mapping of individual signal errors as the resultant position degradation can be seen in deviations from the expected results.

The RTK library allows easy manipulation of complex datasets so that individually tailored outputs can be designed for analysis of specific variables. Phase solutions are also provided for high accuracy geometric comparisons with observed pseudoranges. M_{P1} and M_{P2} observables are also readily available so that mapping of abrupt changes in multipath values can be analysed in the position domain, linking the abstract concept to its physical location.

Commercially available processing tools such as Leica's Geo office and Geoquest solutions' Grid inquest also provide alternative sources of position information and independent data analysis.

Multiple datasets were also used as not only an additional solution check, but also as an additional level of data analysis, utilising the daily repetition of satellite orbital positions to investigate independent yet geometrically identical position solutions.

6.2 Severn Valley data analysis

6.2.1 Introduction

As shown in figure 5.2.3 there are two main datasets produced from the collection exercise at Kidderminster, a rover file and a reference file. When dealing with these results, the period when both the rover and reference station are actively collecting data will be focused on.

Initially, the satellite availability will be compared to the predicted availability seen in figure 5.13 to assess the reference station and rover environments for obscuration. The data will be plotted onto overhead satellite imagery so that position errors and outages can be seen in the context of the train's surroundings. From this, areas of particular interest can be analysed in more detail in order to determine multipath levels and sources as well as signal quality and the local environmental effects.

6.2.2 Initial data analysis

Satellite availability was computed for both the reference station and rover receivers on an epoch by epoch basis. The results are shown below in figures 6.1 and 6.2 respectively.

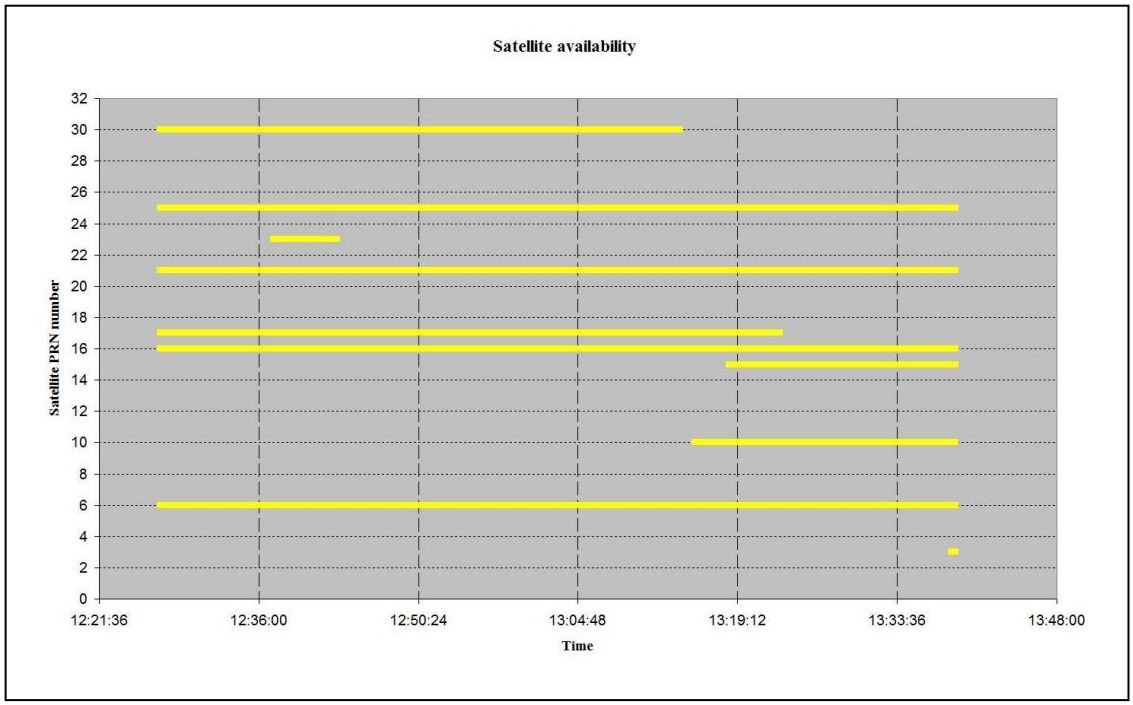


Figure 6.1 – Satellite availability for Kidderminster reference station receiver

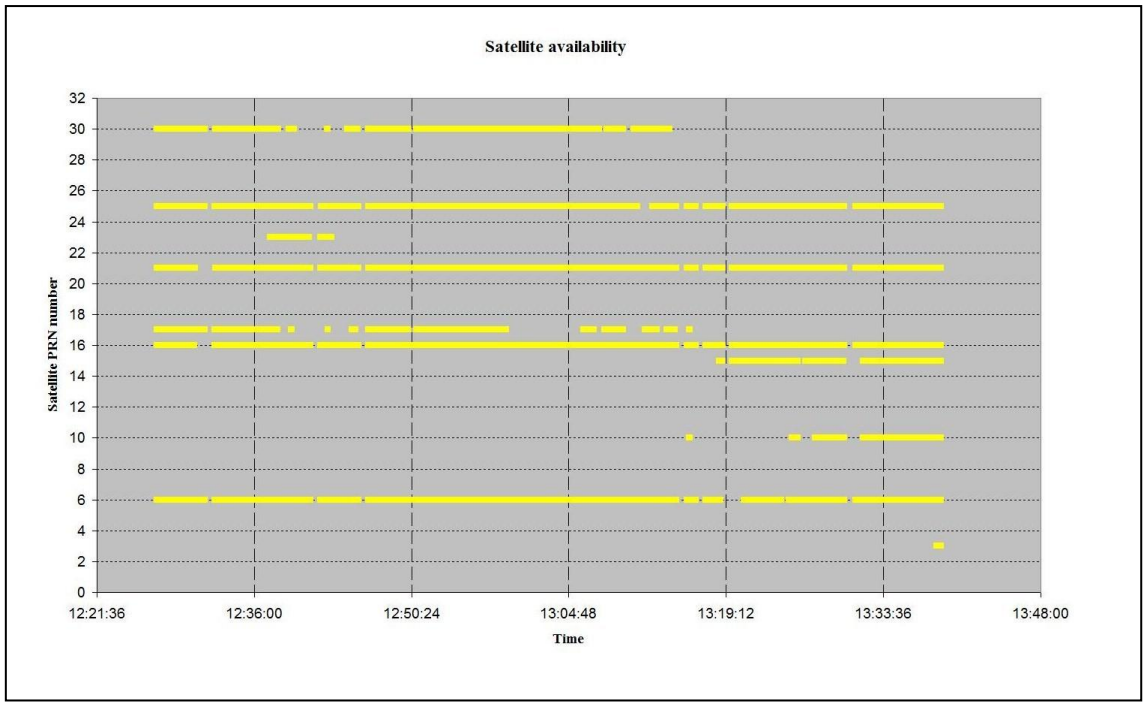


Figure 6.2 – Satellite availability for train mounted rover receiver

It can be seen that the reference station has a good and constant view of 6 satellites, with 7 satellites visible for a subset of the data. This differs to the predicted availability of 8 to 10 satellites during this time period. This can be explained when looking at the predicted sky plot during the rover data collection period, as shown in figure 6.3.

The satellites above a 10° mask angle are numbers, 30, 25, 21, 17, 16 and 6. These are the same as the satellites seen by the reference and rover satellite for the duration of the collection exercise. Satellites 10 and 15 are seen for a short period due to their ascending in the sky towards the end of the time period being considered. Satellite 23 is also seen for a short section of time as despite it being above the cut off angle, it remains very low on the horizon and so is far more susceptible to obscuration.

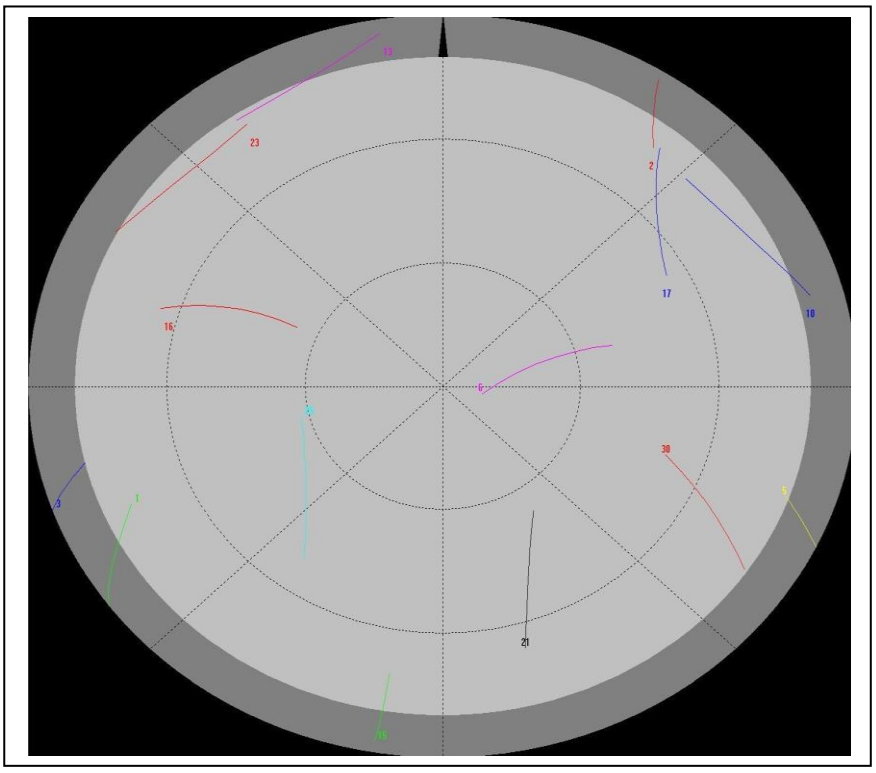


Figure 6.3 – Sky plot for satellites from 12:20 to 13:30 on 21st December 2004. Dark grey band represents 10° cut off.

The breaks in rover data can be clearly seen, unlike the continuous availability bars seen for the reference station. The reasons for these breaks are what will be investigated first.

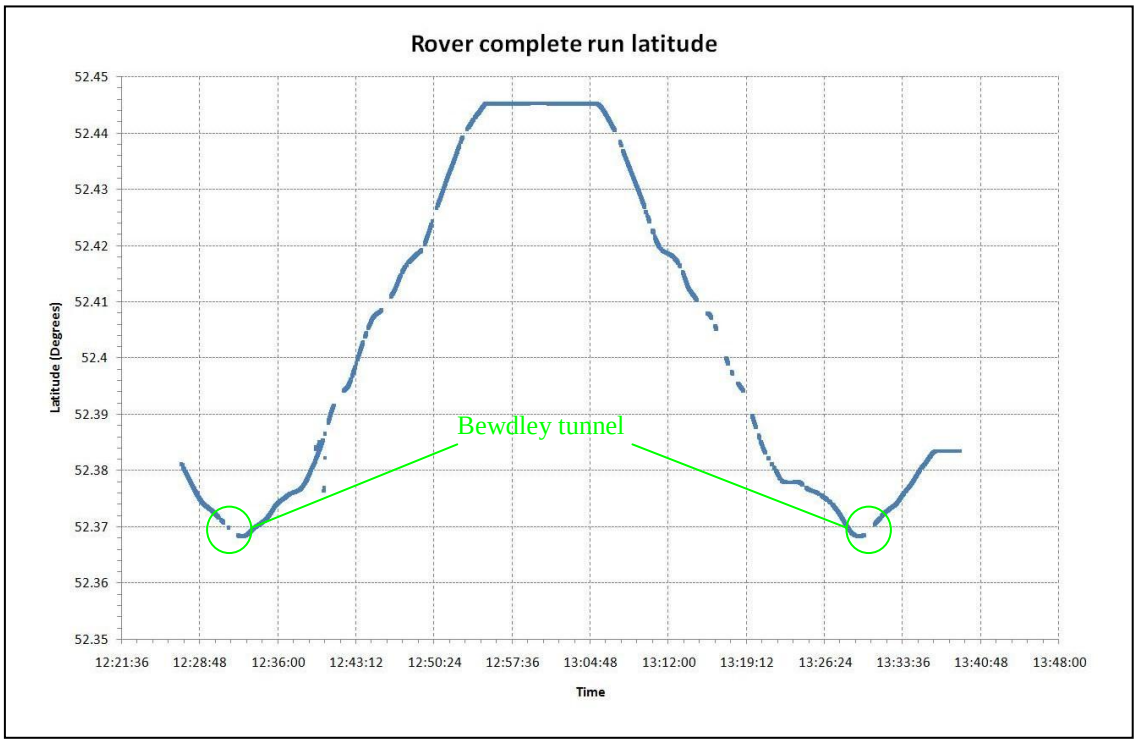


Figure 6.4 – Latitude plot for complete rover dataset using pseudorange observations

Using the RTK library to process the rover pseudorange data, longitude and latitude positions were plotted for the full train run from Kidderminster to Highley and back again. The results are shown in figures 6.4 and 6.5.

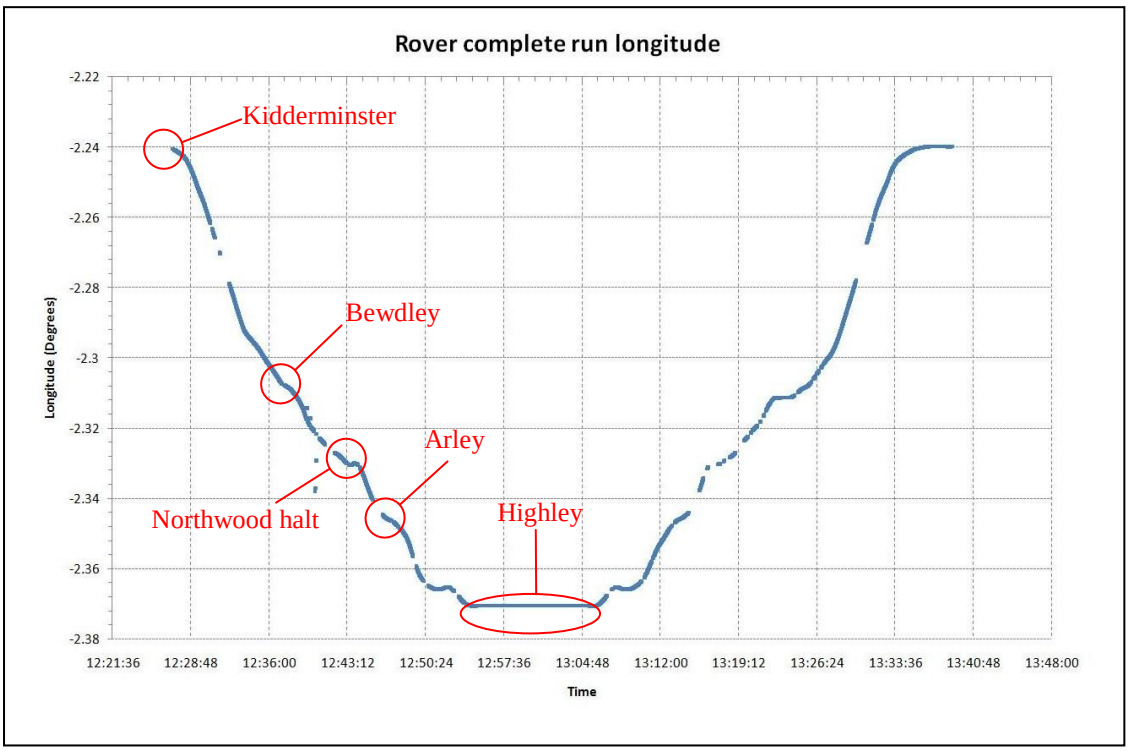


Figure 6.5 – Longitude plot for complete rover dataset using pseudorange observations

The two way journey can be seen to repeat many of the same breaks on the outbound and return path, one of which is Bewdley tunnel, as shown on figure 6.4.
There are also level sections where the train was stationary whilst waiting at stations along the line. By cross-referencing the coordinates, the individual stations could be identified and marked on figure 6.5.
For a complete view of the two dimensional track geometry, a plot of longitude against latitude is in figure 6.6 was produced.

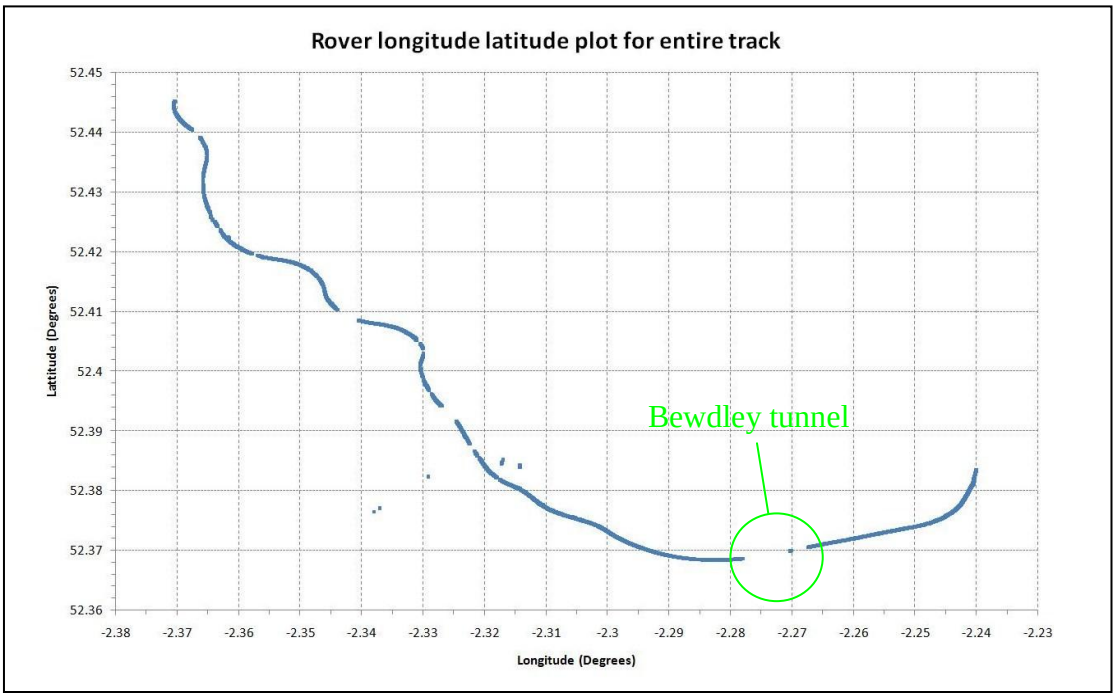


Figure 6.6 – Longitude latitude plot for rover dataset using pseudorange observations

Other breaks in the track data suggest areas of high surrounding trees and buildings leading to complete obscuration of satellites, or at least less than 4, removing the opportunity for a position solution to be computed.

By utilising the open source Google earth KML script language, positions from pseudorange observations can be plotted as a path onto satellite imagery. An example of a KML file used is shown in figure 6.7. This specific file plots a three dimensional line onto the surface of a map within Google earth, we have used the longitude and latitude as horizontal coordinates and initially the height value for position as a vertical component.

We can also plot a simple two dimensional line onto a map, which is shown in figure 5.1 at the start of the previous section. By plotting the code solutions, deviations from the expected path can be visually seen as the physical train route can be seen on the satellite imagery.

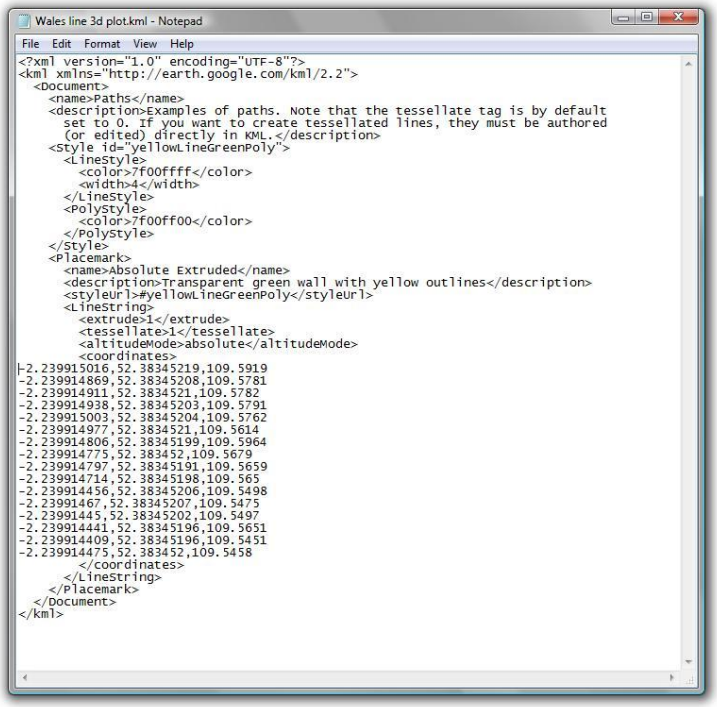


Figure 6.7 – Example KML file format for plotting 3D path

The first area of interest is the main break due to the Bewdley tunnel. When looking at the data in figure 6.6 the break is seen to not be entirely clear, with position solutions seeming to appear in the middle of the tunnel. On further investigation, the reasoning behind this is clear. Below are three images (figures 6.8, 6.9 and 6.10) of the Bewdley tunnel area aerial photo as standard and with 2D and 3D plots produced by code position solutions for the full rover dataset.

Figure 6.8 – Aerial image of Bewdley tunnel area

Figure 6.9 – Aerial image of Bewdley tunnel area with 2D line plot

Figure 6.10 – Aerial image of Bewdley tunnel area with 3D line plot

When looking at these images, the tunnel section can be clearly seen and is marked on figure 6.8.

When looking at a simple two dimensional plot of the rover code based position solutions, due to line fitting, seeing the individual points is very difficult and so distinguishing between what may be a very straight line made up of lots of points and a long line simply fitted between two points is very difficult. This uncertainty was removed by using a three dimensional line plot and including the height value from the position solution (long, lat, height). Figure 6.10 clearly shows (in the form of vertical lines) each individually plotted position. In further figures, the height component from the 3D plot has been replaced by levels to show the number of satellites used for the position at each epoch, i.e. (long, lat, sats used).

The break in data can be seen to cover not only the tunnel section, but also sections before and after as reacquisition of the GPS signals occurs. The data plotted is for both the outbound and return journeys. If the data is broken into two sets, the directional differences can be shown, as in figures 6.11 and 6.12.

Figure 6.11 – Aerial image of Bewdley tunnel area with 3D plot of train out (travelling left) journey

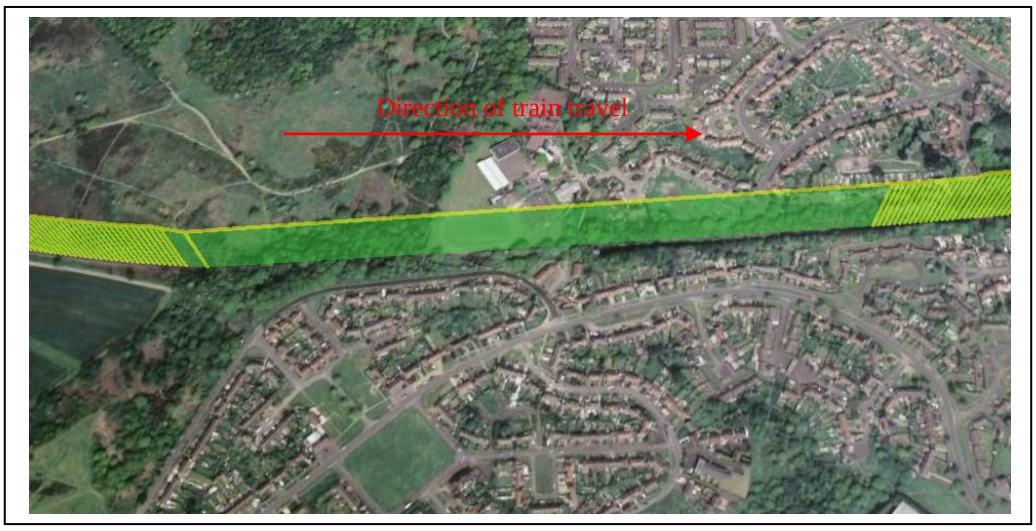


Figure 6.12 – Aerial image of Bewdley tunnel area with 3D plot of train return (travelling right) journey

Figure 6.11 shows that on the outbound journey, the data gave position solutions in the middle of the major blackout region; this was not the case when the train was travelling in the opposite direction an hour later on the return journey. This could be due to the reacquisition time required by the receiver to lock onto the GPS satellite signals which would delay the first position solution, or due to the satellites seen on the outward journey being in a different position on the return journey due to the

time difference being 1 hour. Any alterations in speed would also affect the distance travelled without a position fix, but at this point on the track, the train was travelling at roughly 20mph.
When looking at the data for these two times, the following table was produced (table 6.1).

Satellite number	Outward journey		Return journey	
	Elevation	Azimuth	Elevation	Azimuth
6	72	79	46	73
16	N/A	N/A	58	292
17	30	52	N/A	N/A
21	N/A	N/A	54	142
25	56	239	35	210
30	31	109	N/A	N/A

Table 6.1 – Table showing available satellites and corresponding elevation and azimuth angles for Bewdley tunnel area during outward and return journeys

The table shows that there are only two satellites in common with the two journeys, both of which are descending in elevation. When looking at the layout of the track along this section, the direction of travel is either at a bearing of 78.7° or 258.7°.

From the sky plot in figure 6.3 it can be seen that satellites 6 and 25 are also travelling close to and parallel with this azimuth.
Satellite 17 is seen to begin in an ideal position, but it's azimuth and elevation decrease and so it is not seen on the return journey. Conversely, satellite 16 is too low on the horizon and possibly obscured by the trees due to its large azimuth. This azimuth then decreases and the elevation increases, bringing the satellite into view on the return journey.

If the raw Rinex data is used to determine the number of satellites seen by the receiver, rather than the number of satellites with data used for the position solution during the outbound journey whilst entering Bewdley tunnel we produce figure 6.13.

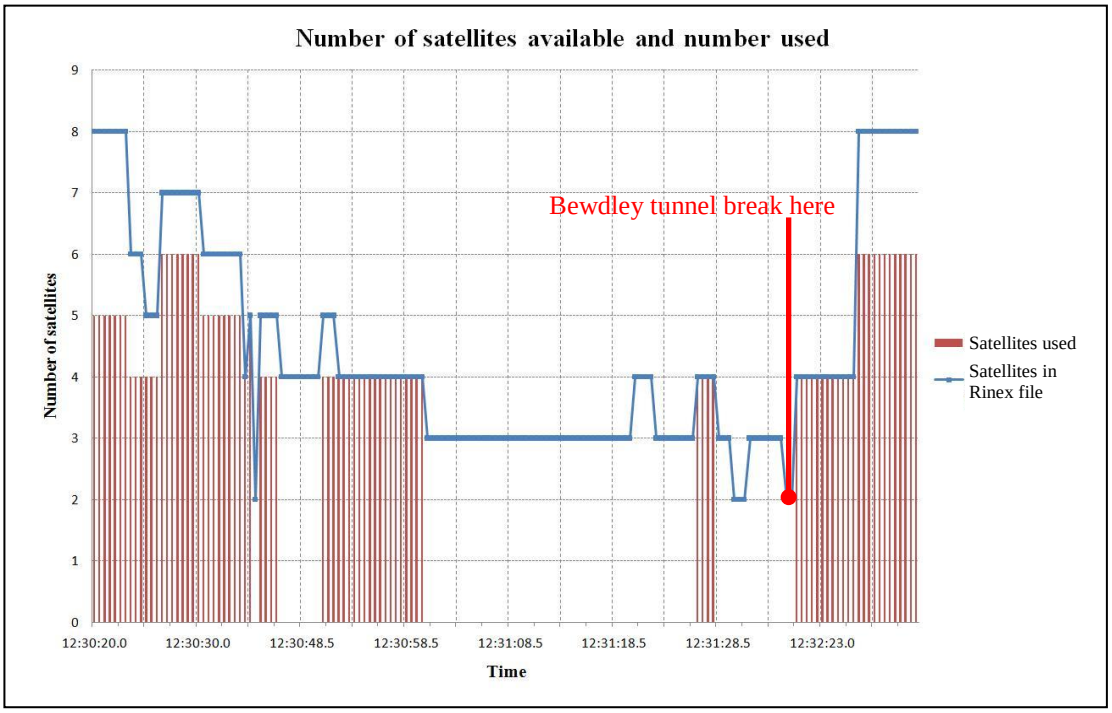


Figure 6.13 – Graph to show number of satellites recorded in Rinex file and number of satellites used for position solution in Bewdley tunnel area on outward journey

The number of satellites available during the section of tree covered track can be seen to drop to three and so a position solution is not attempted by the RTK library. When looking at the Rinex data directly, the L2 signals are lost for some of the satellites while the L1 signal is retained. This is assumed to be due to the weaker signal strength of the L2 band as well as the additional processing in the receiver on the L2 signal. Loss of the L2 frequency removes the possibility for the dual frequency ionospheric correction to take place, increasing pseudorange and eventual position error.

Given that the focus of this thesis is code multipath, unless otherwise stated, L1 and L2 will be used to refer to the code observation on each frequency as phase multipath is beyond the scope of this investigation.

Further breaks in the train's data can be linked to tree canopies, bridges and buildings, as shown in figures 6.14 to 6.16.

Figure 6.14 – Data breaks due to tree canopy signal obstruction

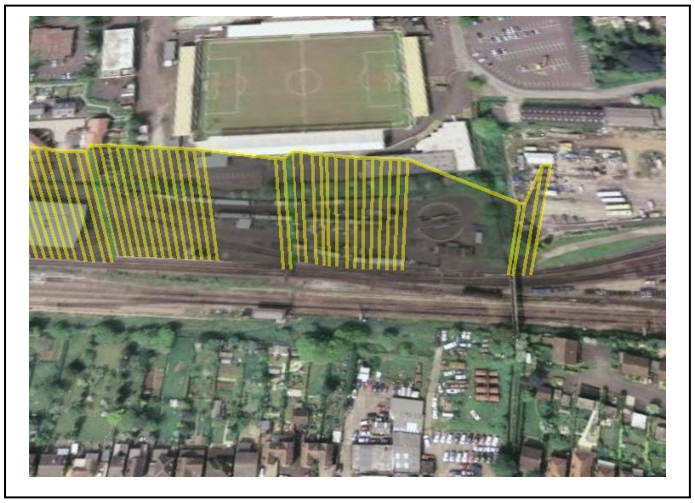


Figure 6.15 – Data breaks due to buildings obscuring signals

Figure 6.16 – Data breaks due to overhead bridge

When considering the entire dataset, 83.2% of the epochs have four or more satellites available in the Rinex file, however only 76.7% of the total data set provides position solutions using four or more satellites. The exclusion of 6.5% of these epochs can be due to several factors.

The most obvious when using the RTK library is the need for a common subset of available satellites between all base stations and rovers. With the data used in this instance, given that the reference station is in an ideal position for maximum satellite availability, this is not such an issue and so other reasons are more likely, such as the previously mentioned loss of the L2 band due to trees. Errors on the signals received due to the interaction with the surroundings can cause them to be considered as unusable.

Multipath mitigation has not been implemented in the processing used so far and so this would not cause discrimination in signal use. Satellite health status flags were all set to healthy during the data collection process and so that was also not a contributing factor to the reduced number of satellites used.

Figures 6.15 to 6.17 also show additional information in the height of the three dimensional line. The KML input file has been altered to accommodate the number of satellites used in the pseudorange position solution as a variable.
This allows the user to verify if a sudden change in an otherwise smooth path is caused by the loss of a satellite or by other errors such as multipath or signal interference.
An obvious example of the former is shown below where the plan view shows an abrupt change in direction for the rover (figure 6.18) and when the view includes the height parameter (the number of satellites used) the error source can be seen as a change in the number of satellites, causing a shift in position (figure 6.17).

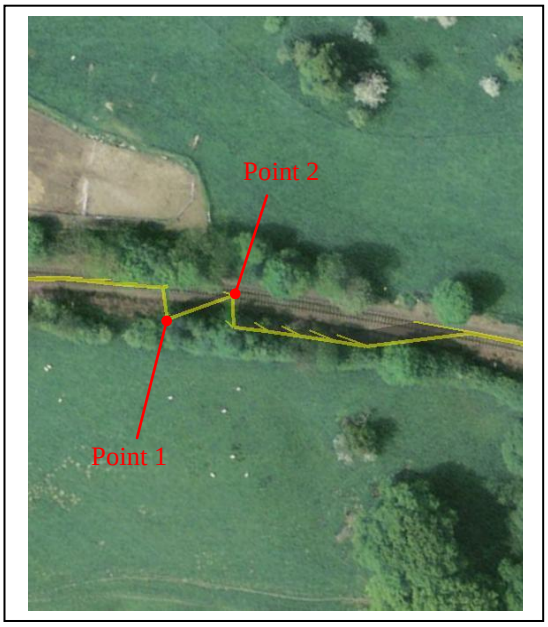


Figure 6.17 – Plan view of abrupt position change for rover path (epoch 12:49:53.0)

Figure 6.18 – Side view of abrupt position change for rover path showing change in number of satellites used

The abrupt change in direction is due to the number of satellites falling from 5 to 4 and then back to 5 for a single epoch, then to 4 for 5 epochs before returning to 5 satellites and the correct path. The satellite that is lost and then re-acquired is number 25 which has a 51.6° elevation and 227.4° azimuth which is perpendicular to the track direction in figure 6.18. This shows how the loss of a satellite can drastically shift a train's position, especially if the satellite is located perpendicular to the train's line of travel. This is a possible issue as the along-track direction is considered to have the best chance of continuous individual satellite acquisition, i.e. satellites seen in the along-track direction will be in view without regular obstruction provided the train path is relatively straight. Due to the nature of the rail environment, satellites seen to the side (perpendicular to the line) will be more prone to obscuration due to the line side infrastructure not necessarily being constant in its blocking effect. Inclusion of accelerometers and gyroscope data together with doppler measurements to provide a bearing for the direction of train travel could be used to down weight the effect of satellites perpendicular to the line of travel so as to reduce this effect. Commercial systems use down weighting of newly acquired satellites for this reason.

Jumps also occur when the number of satellites is constant. This is less common and is assumed to be due to signal errors due either to multipath, diffraction or interference.

From the dataset recorded, some of these shifts in position will be investigated to try and determine the reason. Figures 6.19 and 6.20 show the plan and side view of such an event.

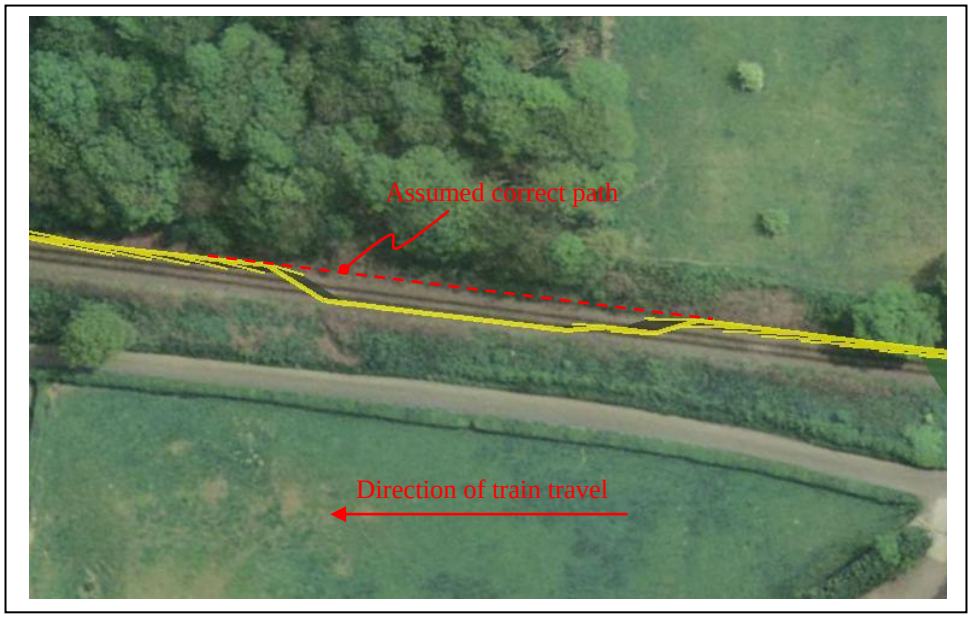


Figure 6.19 – Position jumps with constant number of satellites plan view

Figure 6.20 – Position jumps with constant number of satellites seen from side

The jump seen is attributed to the loss of the L2 signal on satellite 6 whilst retaining the L1 signal. This means the ionosphere free combination cannot be carried out and

causes a range error for this satellite. The minor shifts seen within the main section are due to the L2 signal for satellites 23 and 25 being regained, thus causing a slight shift to the correct position (if averaging was to occur between both ends of the shifted section) Then the path is pulled back onto the expected line by regaining the L2 signal on satellite 16. This is shown more clearly on the annotated graph of Longitude in figure 6.21.

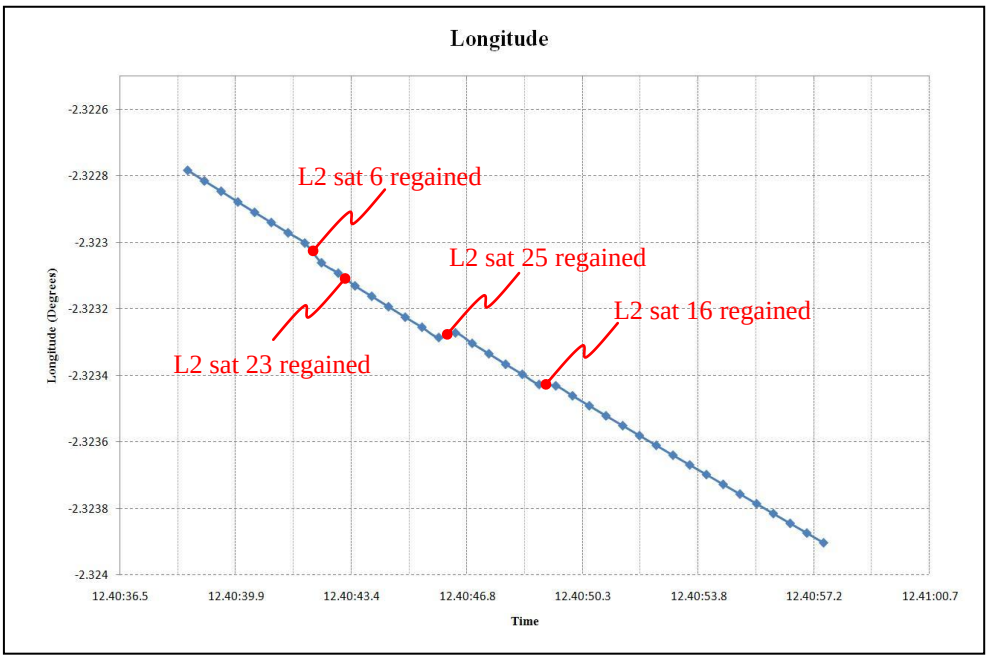


Figure 6.21 – Plot to show position shifts due to regaining the L2 code signals of specific satellites

From this it is clear that not only does losing/regaining a satellite affect the position but losing/regaining a single frequency from a satellite can affects the position in a similar manner.

Another example of additional satellite data not necessarily improving the position solution is a section of track located north of Bewdley and is the next section along from that shown in figure 6.19. Figure 6.22 shows an overhead view of the section being investigated.

Figure 6.22 – Overhead view of section of track surrounded by trees (epoch 12:40:58.0)

The small deviation seen in the plotted line is due to a single epoch where the signals of satellite 23 was momentarily observed.
The section of interest consists of 24 epochs from 12:40:58.0 to 12:41:09.5.
During this period satellites 6, 16, 21 and 25 were also observed but without any cycle slips or data gaps.
This raises the issue that the position of a train calculated when only L1 signals are available is possibly better than when only some have L2 signals available, or if the availability of L2 signals is fluctuating.
When looking through the entire route, most position shifts that were visible in Google earth were deduced as being due to either the loss or acquisition of the L2 signal of one or more satellites.

6.2.3 Dual frequency investigation

When looking at the signal availability for the entire run for each of the satellites mentioned in relation to figure 6.22, the ratio of epochs where L2 is observed as well as L1 is shown in figure 6.24 along with the availability of L1 when considering epoch breaks.

Satellite PRN number	L1 availability over whole run	L2 availability over whole run	L2 availability when L1 present
6	81.7%	73.5%	90.0%
16	81.5%	75.3%	92.2%
21	81.7%	80.7%	98.7%
25	81.7%	75.0%	91.8%

Table 6.2– Table showing availability of L1 and L2 signals during

These results do not follow the expected availability when considering only the elevation angles as when averaging for the duration of the journey, satellite 6 has the highest elevation angle yet has the lowest L2 visibility (though joint highest L1 visibility with satellites 21 and 25). The lowest elevation satellite, PRN 21 also has the highest L2 availability. None of the elevation angles are however below 40° and the differences between the satellites discussed are small (~10°) so the data does not cater for the extreme cases very well.

The breaks in L2 availability consequently cause breaks in the calculation of the M_{P1} and M_{P2} code observables. These breaks mean that when trying to average a long run of data in order to extract the actual multipath values, only short runs can be averaged and depending on the variance of the multipath, it may not be possible to identify actual values.

One section of track that allows a significant stream of data to be logged without breaks is located just north of the Bewdley tunnel in a valley passing the local safari park just before the Bewdley stop is reached as can be seen in figure 6.23.

Figure 6.23 – Valley section of line with few signal breaks

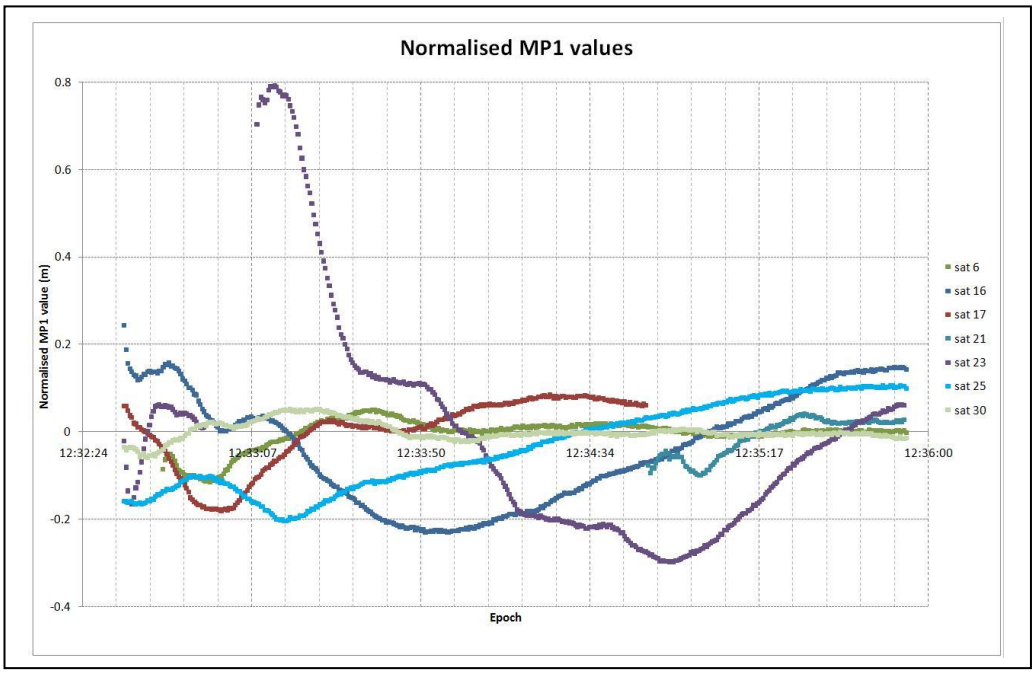


Figure 6.24 – Averaged MP1 values for valley section of track on outgoing journey

Despite this section of track having a comparatively clear view of the sky, there are still breaks in the data as well as complete obscuration of some of the satellites at different times.

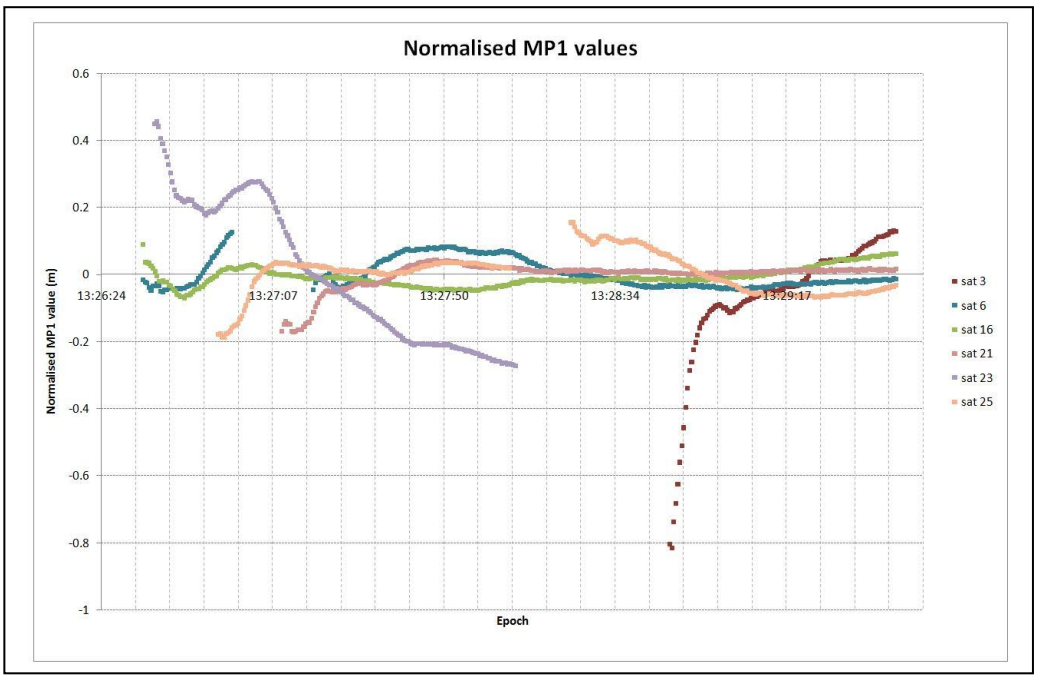


Figure 6.25 – Averaged MP1 values for valley section of track on return journey

By looking at the elevation angles of the visible satellites during the outbound and return journeys, some of the breaks in data and high M_{P1} values can be seen to be elevation angle dependant.

Satellite PRN	Outward journey		Return journey	
	Elevation	Azimuth	Elevation	Azimuth
3	N/A	N/A	10.2°	256.0°
6	70.9°	77.8°	47.8°	73.4°
16	34.2°	288.8°	57.2°	292.9°
17	29.54°	51.3°	N/A	N/A
21	30.9°	160.1°	53.2°	144.7°
23	12.5°	308.4°	14.6°	320.5°
25	56.0°	238.0°	37.2°	212.3°
30	30.6°	110.3°	N/A	N/A

Table 6.3 – Table showing elevation and azimuth angles for all visible satellites on the outgoing and return journeys in the valley section of track seen in figure 6.23

Satellite 23 has the greatest variance in M_{P1}, reaching 1.1m in total during the earlier run. This is most probably due to the very low elevation angle of 12° to 13°, something that may not normally be possible, but due to the azimuth of the satellite relative to the train being along the line of the track, the visibility is improved dramatically, especially as this particular section of the valley is relatively straight and feeds onto a level area where Bewdley town is situated.

This level of multipath follows on from what was hypothesised in section 2.6 regarding the along track, low elevation satellite geometry providing the highest risk of multipath.

Satellite 3 comes into view at 13:28:48.0 with a high level of rapidly changing multipath. The direction of the satellite relative to the train receiver is perpendicular to the track direction over the shallow side of the valley. The low elevation angle of the satellite and the fact that it is ascending into the sky means the initial signals received will be more susceptible to multipath and interference from the ground

plane of the valley. Table 6.3 illustrates this in the late acquisition and rapid change in M_{P1} observable for satellite 3. The level of change is then reduced dramatically and begins to stabilise to a variance similar to that of the other satellites present.

Analysing the extremes of M_{P1} and M_{P2} values seen on all visible satellites for the out and return journeys gives a rough estimation of the values expected. The results are shown in table 6.4 by calculating the difference between the maximum and minimum values for the entire run.

Satellite PRN	**Maximum normalised difference**	
	M_{P1} (m)	**M_{P2} (m)**
3	1.162	0.821
6	0.303	0.960
10	0.420	0.326
13	0.781	1.140
15	0.638	1.538
16	0.617	1.061
17	0.464	1.735
21	0.500	0.704
23	1.220	1.471
25	0.685	0.821
30	0.401	1.438

Table 6.4 – Table showing maximum variations in M_{P1} values for satellites in view

From the results, it can be seen that M_{P2}, the code multipath seen on the L2 signal is generally larger, but not in a proportionate sense as can be seen in figure 6.26.

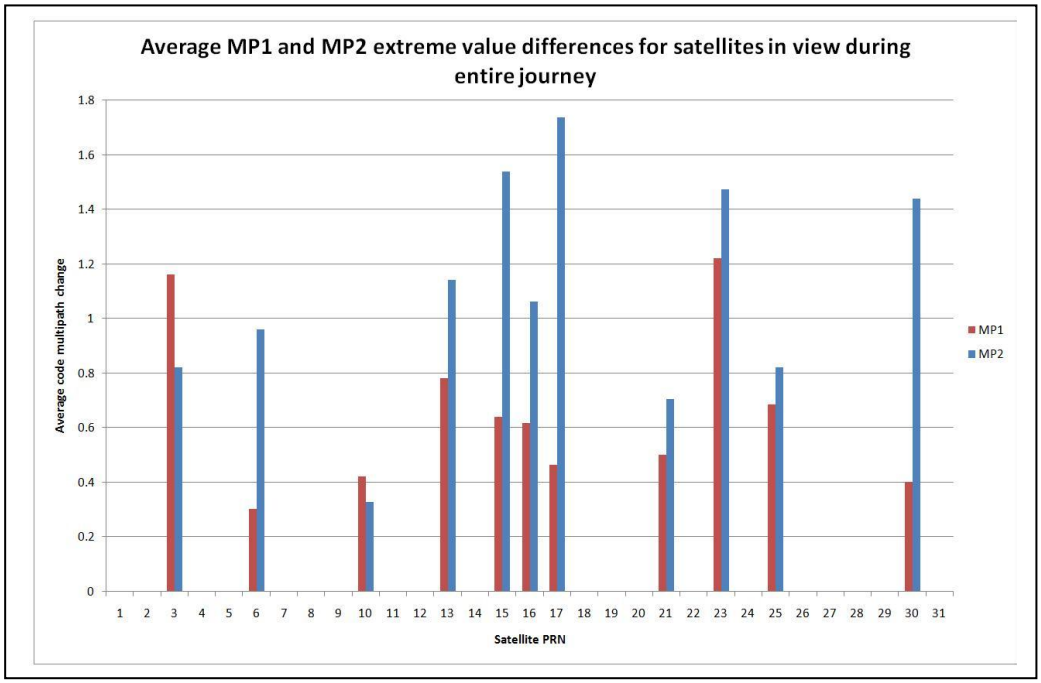


Figure 6.26 – Graph showing comparison of average MP1 and MP2 differences for the out and return journey.

The reasons for the overall increased, yet not proportionate M_{P2} values are due to the nature of the L2 signal and its reflection and interaction properties being different to that of the L1 signal. The different surfaces that interact with the signals also affect the regularity of the received signals, increasing the seemingly random ratio of M_{P1}/M_{P2} values calculated. When the ratio of L2 availability was compared to L1 in table 6.2 it was also seen that L2 was generally a weaker signal with a consistently lower availability.

When analysing static data recorded from some of the satellites in view whilst the train was in Kidderminster station, as discussed in chapter 2, the changes in code multipath observables are seen to have a far smoother, sinusoidal variation, as in figure 6.27 and 6.28.

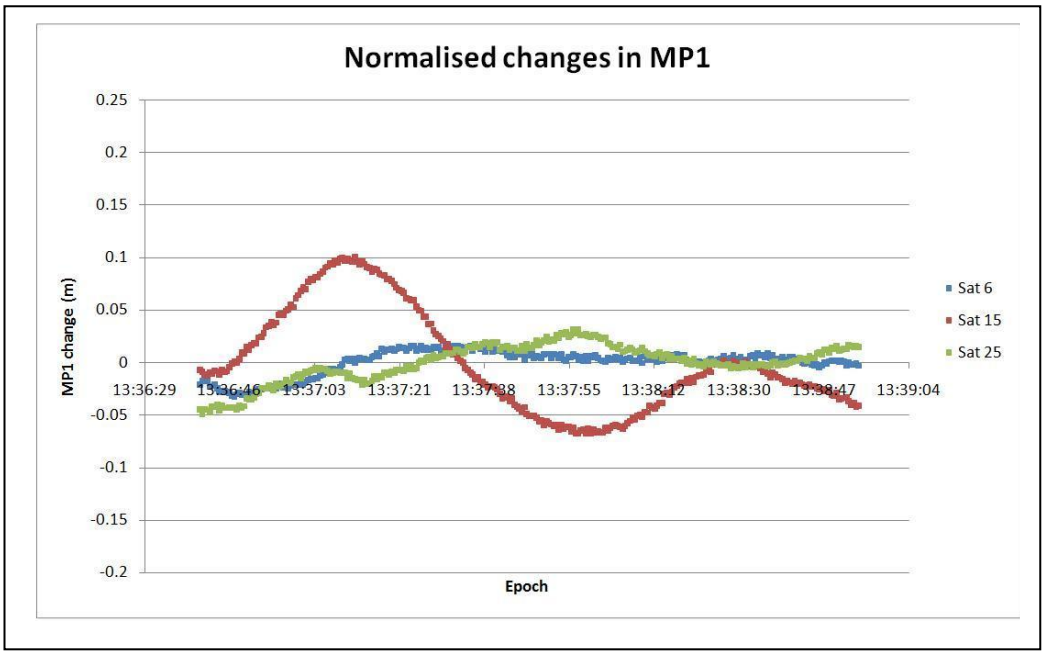


Figure 6.27 – Graph showing change in MP1 for static train data

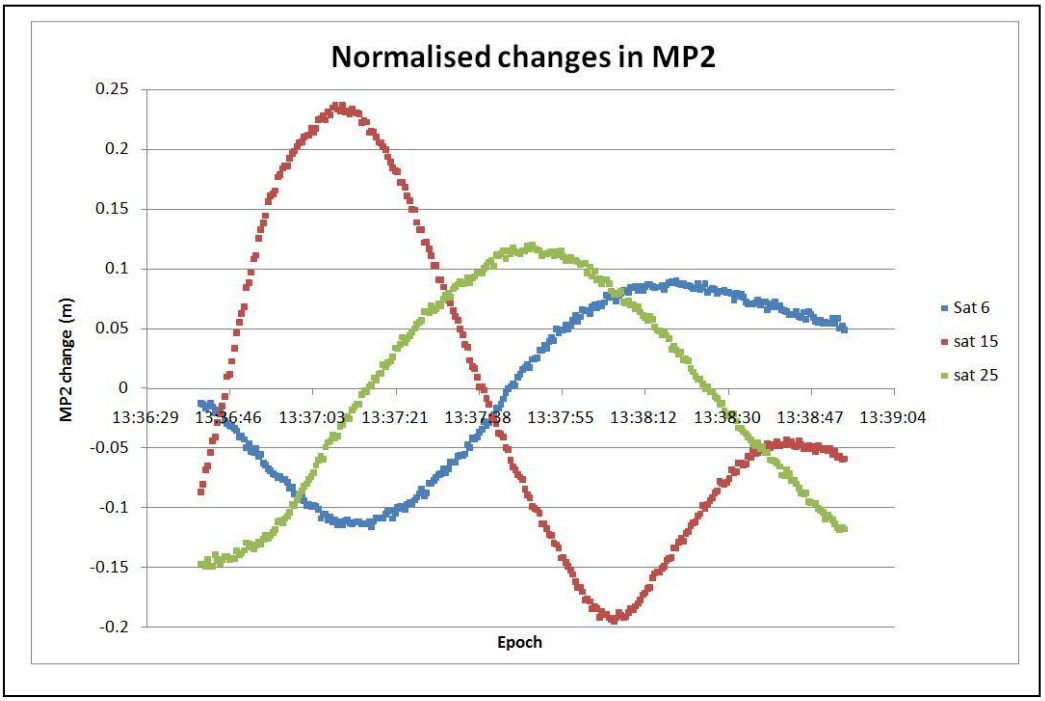


Figure 6.28 – Graph showing change in MP2 for static train data

The scale of change on M_{P2} is over double M_{P1}but with similar frequency, though the M_{P1} values for satellites 6 and 25 are not as clearly sinusoidal as that of satellite 15. When comparing the frequency of the multipath, all 3 satellites appear to have a very similar frequency of roughly 1 minute. This is due to the multipath on all the satellite signals either being from the same surface or from surfaces the same distance from the receiving antenna.

It is possible that the multipath seen in these figures is in fact due to the train roof reflecting the incoming signals at a low angle. The difference in values for the L1 and L2 code multipath can be due to many reasons, some with larger effects than others.

One contributing factor is the weaker signal strength on L2 and the tendency for its absorption coefficients to be higher than L1 for some surfaces, causing a greater circular signal retardation and a greater perceived reflected signal path length.

Due to the internals of the hardware receiver being unknown, it is difficult to say with certainty that the L2 signal undergoes more processing and so introduces more possible errors (mainly as system noise) than for L1, but it is generally considered as another source of degradation on the L2 signal.

When investigating the dual frequency data collected, the major limiting factor for obtaining useful and valid results has been the intermittent characteristics of the two frequencies in the presence of foliage and buildings. Due to the frequency of breaks in the signals, trying to produce M_{P1} and M_{P2} results over a sufficient period of time in order to reduce them and determine the actual physical multipath values, rather than simply the amount they vary by is very difficult.
By using larger data sets it may be possible to reduce this issue as over the course of a day, satellite positions change and so the way in which their signals interact with the local environment will also change, possibly providing longer sections of break-free data.
If this is not possible, in order for the true value of multipath to be determined, the exact position of the receiver and satellites must be calculated and then compared to the observed ranges. The initial results from implementing this process are described in the following section.

6.2.4 Geometric comparisons

Through the use of the RTK library mentioned in section 4, dual frequency data is used to provide highly precise double differenced positions for the rover receiver by utilising the local base station set up in the nearby depot yard.
If a highly accurate position is known for the rover receiver as well as the satellites in use, the exact geometric (true) range can be calculated, provided the satellite positions used are for the time of transmit and not the time of reception by the receiver. Clock offsets for both the receiver and satellites are removed during double differencing, but these must be calculated and included when calculating satellite positions.

Provided the best estimates of the geometric ranges between the satellites and receiver have been calculated, the observed pseudoranges can be used to estimate the multipath on the received signals. The raw measurements cannot be used without alteration though as the ionospheric and tropospheric errors need to be accounted for as otherwise, they will show up as an additional unidentified path length that could be misunderstood as multipath.
Once these errors have been removed, some initial results were produced and those for satellite 6 for the beginning of the data set are shown in figure 6.29.

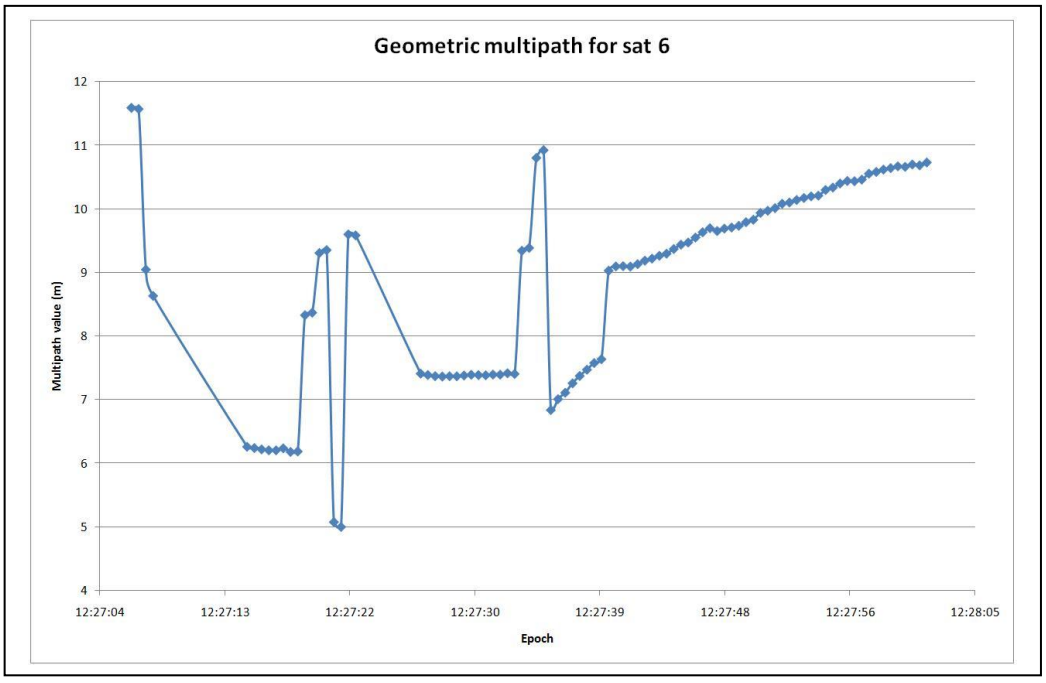


Figure 6.29 – Graph showing geometric multipath for satellite 6

From this figure, it seems as though the multipath calculated from differencing between the geometric range to the satellite and the observed pseudorange is fluctuating by almost 7 meters.
This is not the case as the errors seen are mainly due to the receiver and satellite clock offsets not being modelled or calculated to a high enough accuracy. The inaccuracies in clock offsets, although small (nanosecond level), affect the pseudorange value by levels similar to those seen in 6.29 due to any timing error being multiplied by the speed of light.
When comparing some of the data with M_{P1} values for satellite 6, an interesting correlation appeared. This can be seen in figure 6.30.

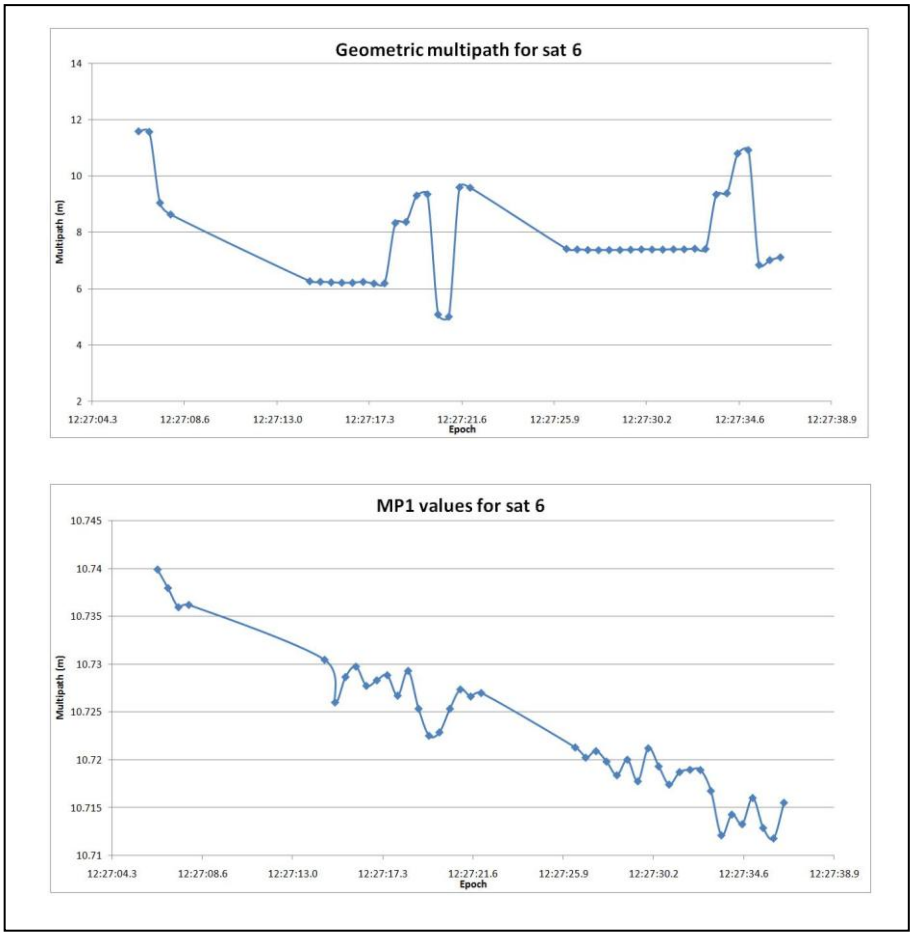


Figure 6.30 – Graphs showing correlation between geometric multipath with clock error and MP1 values for satellite 6

When investigating this correlation, it is hypothesized that the reason M_{P1} seems to change by the same sign at the same epochs, yet by a different scale could be due to the L1/L2 correlator frequencies or offset being affected by the clock offset jumps seen in the upper graph. The changes seen in M_{P1} are in the region of 5mm and so are

not of concern in this thesis, but further investigation of the receiver clock jump effects on the L1-L2 correlator offset could be of interest to very high precision applications.

In order to remove these unwanted clock effects, double differencing can be used as was seen in section 2. The only remaining error factor in the signal measurements will be due to multipath from the local environment and the internal receiver noise.

The problem with using the double differencing method is that the multipath values calculated will be a combination of those seen by both receivers for both satellites and so the combination of M_{P1} and M_{P2} values must also be calculated if a physical level of the combined multipath is to be derived.

When using the double differencing equation 2.34, the corresponding combination of multipath code observable values (MP) can be expressed as in equation 6.0 shown below.

$$MP_{12}^{ab} = MP_1^a - MP_2^a + MP_2^b - MP_1^b \qquad (6.0)$$

Where MP_{12}^{ab} is the double difference combination of code multipath observables for satellite a and b observed using receivers 1 and 2
MP_1^a is the code multipath observable on satellite a from receiver 1
MP_2^a is the code multipath observable on satellite a from receiver 2
MP_1^b is the code multipath observable on satellite b from receiver 1
MP_2^b is the code multipath observable on satellite b from receiver 2

The double differenced geometric multipath results are shown in figure 6.31 for a small initial section of the Severn Valley Rail data.

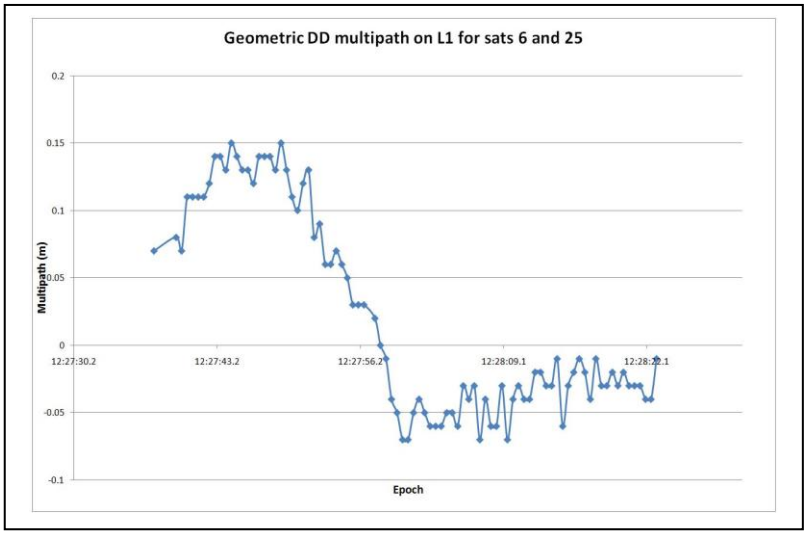


Figure 6.31 – Graph showing double differenced geometric range multipath for satellite 25, using satellite 6 as reference.

From figure 6.31, the multipath calculated using the observed minus computed residuals from the double differencing process is seen to vary by roughly 20cm over the 1 minute period. This is a combination of the multipath seen at the base station and rover sites for signals received from satellites 6 and 25.

The variation in multipath shown in 6.32 can be compared to the M_{P1} observable for the same time period in order to show both the validity of the code multipath observables to accurately show the multipath on a signal and to correct for the double differenced combined integer ambiguity offset seen as a result of the M_{P1} combination in equation 6.0.

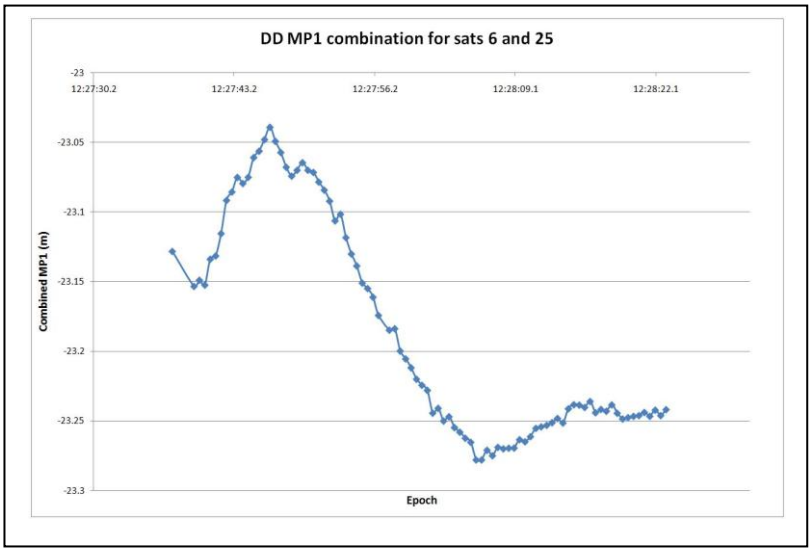


Figure 6.32 – Graph showing double differenced combination of MP1 observables for satellite 25 using satellite 6 as reference

Figure 6.32 shows the same time period as figure 6.31 and the combined integer ambiguity offset can be seen to be roughly -23m. When comparing the amount by which the geometric multipath varies by, the two figures show the same 20cm change as well as a high sinusoidal correlation. Figure 6.33 clearly shows this by normalising the M_{P1} output by adding the average offset of 23.186m and plotting it along with the geometric values on the same graph.

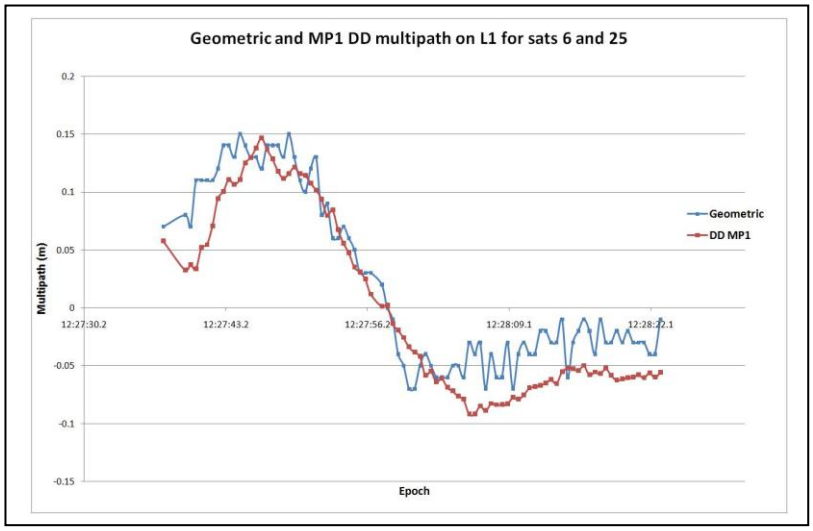


Figure 6.33 – Graph showing geometric and normalised double differenced MP1 observables for satellite 25 using satellite 6 as reference

From this correlation, the absolute values for the combined M_{P1} observables can be computed, but the integer offsets at the rover and reference sites are still not known. Due to the reference receiver being static and ideally with a clear view of the sky, large periods of continuous data are more commonly available and so averaging of the data can occur with more certainty. From this averaging, the M_{P1} integer offsets for both the observed and reference satellite can be approximated, along with the maximum levels of change in the site specific multipath.

Despite this extra information allowing the integer offsets for the reference station to be calculated, the integer offsets for the rover cannot be calculated as the sign of the offset are not known and so the combination of the code multipath observables cannot be broken up to solve for the rover observable offsets .

What this process of double differencing does allow, as shown in figure 6.33 is the checking of the M_{P1} observable and whether or not the values observed (although combined) represent the true physical variation of multipath on the satellite signals.

By using the RTK library to calculate the double differenced position for the trains receiver, a truth is obtained so that positions calculated purely with code pseudoranges can be compared for accuracy and precision as the additional benefits of phase solutions may not be cost effective if the pure code solutions are within the required limits of positional accuracy required by the rail network. The possibility for further research in this area has been provided by the construction of the RTK library, but is beyond the scope of this research project.

6.2.5 Conclusions

The initial purpose of this experiment was to collect data from a rail environment in order to gain an understanding of how it may differ from that collected from more common sources such as base stations and rover receivers in cities and open spaces. The UK rail environment has not been investigated using GPS very much and multipath analysis has never been undertaken.

Because of the relative lack of available information on the area and the complete lack of RINEX data from a train based receiver, this initial experiment was an ideal chance to record something that would provide data similar to that expected from the Birmingham data collection exercise.

The data characteristics in terms of breaks, kinematic variance and signal errors were new and gave an idea of how a GPS receiver would function in kinematic mode along a smooth track where repeatability of position is possible within very low tolerance due to the outgoing and return journeys of the train.

The rural nature of the Severn Valley also helped to quantify the possible effects of trees and foliage on a kinematic receiver.

Being the first data set for evaluation, many processing tools were developed using the data as the erratic availability of dual frequency data and whole satellites forced a single epoch approach to calculating positions. The number of cycle slips was also larger than normally expected, adding to the need for single epoch positioning as multiple epoch data processing would rarely function.

Due to the main Birmingham data set availability being delayed by 2 years, the processing system needed to be automated so that when the large data sets were available, errors in data did not cause problems.
Simple errors such as RINEX breaks and receiver glitches were prevalent in the Kidderminster data set and this allowed a number of software improvements to be undertaken to improve the resilience of the processing library used.

Given the predicted satellite availability from the Leica availability program, the actual availability generally agreed favourably with the rover and reference visibility however the rover visibility altered significantly depending on the surroundings.
The major errors seen when trying to produce a track line plot in Google earth were produced either by whole satellite availability changes or L2 signal dropouts. It was clear when analysing specific instances of these jumps that satellites in the along-track direction cause less of a position error than when satellites perpendicular to the track direction fluctuate in availability. The results are similar when assessing L2 signal availability, when the L2 drops out, the position jumps can be similar to losing a satellite entirely.
In order to counteract this, a form of smoothing is usually used by commercial location devices in order to reduce the instantaneous effects of satellite acquisition. This does not however alleviate the sudden jumps seen when a satellite is lost from one epoch to the next. For certain satellite geometries, the effects of a satellite being removed may have more or less effect, by looking at Geometric Dilution of Precision (GDOP), the even spreading of satellites may reduce any shift in position along a certain line but as calculation of GDOP simply produces a unit-less number, a full 3D geometric analysis of each epoch would need to be carried out in order to determine possible position effects.
Given the obvious fact that a train is on a fixed line and that the maximum curvature of the line is limited relative to the line speed, this could be used when attempting to smooth continuous position data. The use of extra sensors such as tachometers and accelerometers could also aid in the reduction of large position jumps due to satellite and signal availability, provided an appropriate kalman filter is used.

Another problem due to L2 dropping out is when trying to assess the fluctuations in multipath through the computation of the M_{P1} and M_{P2} code observables. Due to the

unknown offset in the values calculated, averaging must take place. This averaging requires a significant period of break-free data in order to confidently determine the average value. Due to the constantly varying availability of some L2 signals, the process of determining an average and thus trying to determine the physical level of multipath as opposed to simply the amount by which it changes is a very difficult process given the amount of data that has been recorded is limited.

It is for this reason that future data collection exercises of this kind would ideally produce a far larger data set so that averaging is possible due to the increased likelihood of longer periods of break-free data.

The tools developed during this initial phase have also paved the way for future analysis to be carried out easily by providing a tool set that can be manipulated and added to in order to analyse specific scenarios and data sets.
The use of Google earth also helped by providing visual results that can be quickly analysed without the need for complex coordinate mapping and the production of multiple graphs for specific areas of interest.

As results, although this initial data set was designed as a test execution methodology for the later Birmingham data set, they would be included in a safety case for GPS in the railways as it is supporting evidence due to the results obtained but also for the development of the analysis techniques used. Given that the integrity level requirements rely on analysis of very large data sets (or reliable extrapolation from smaller sets) any additional data collected is of use when trying to define any possible errors that can occur, along with their probability.

6.3 Watford junction data analysis

6.3.1 Introduction

The four data sets produced over the two days of observation by a rover and reference receiver were processed using the RTK library and the outputs analysed. The 4 minute offset between successive day satellite positions provides a truth to check for any repeated breaks in the data or anomalies seen.
General satellite availability is not the focus of this experiment due to the area being relatively open to air and the receivers used not being mounted on the train itself but on the platform and nearby building roof.
By focusing on the elevation and azimuth angles where possible sources of EM radiation are located, changes in the signals received can be analysed without having to perform a detailed analysis of all the data.

6.3.2 Initial data analysis

The satellite availability for the two days for the rover and reference receivers are shown in figures 6.38 and 6.39. The graphs show the link between the roof and platform receivers in the satellites available being very similar. The platform results for both days follow the roof results very closely, confirming initial thoughts that the platform receiver has a very good view of the sky.

The broken availability of satellite 30 is also seen to repeat itself on the second day of observation for both the roof and the platform receivers and so the possibility of the breaks being due to overhead cables or infrastructure unique to the platform area is unlikely.

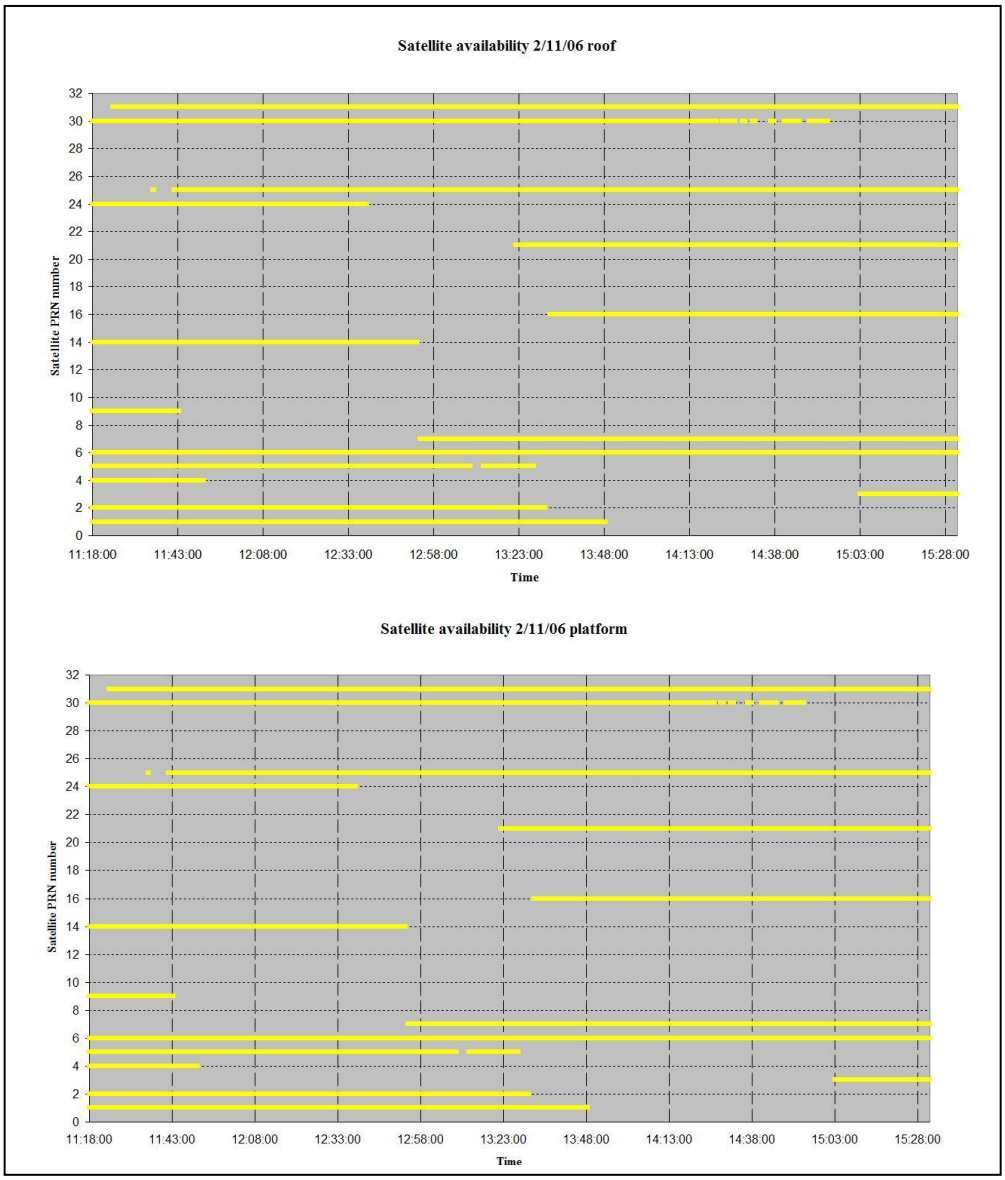


Figure 6.34 – Graph showing satellite availability for roof and platform Watford Junction receivers on 2nd November 2006

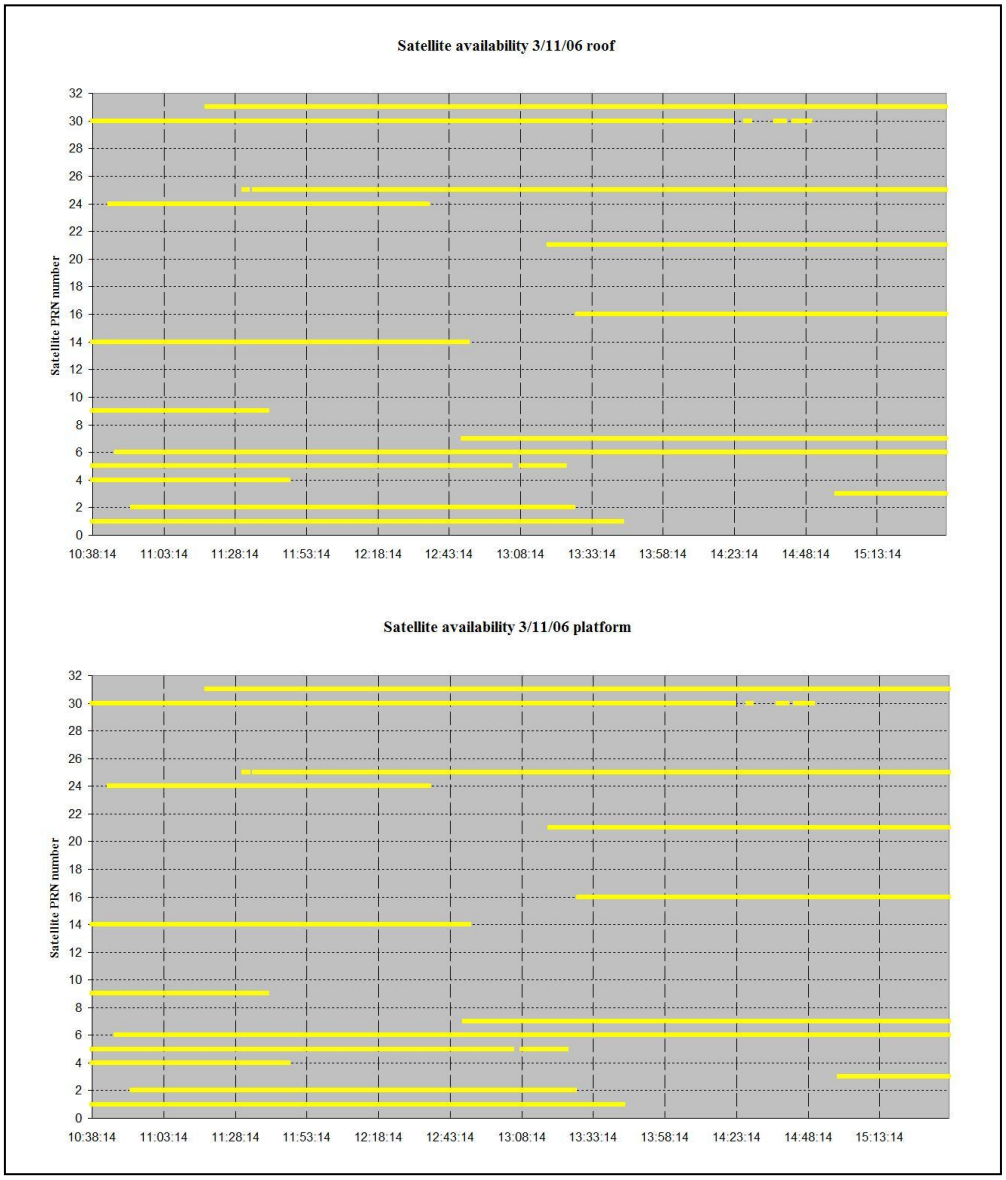


Figure 6.35 – Graph showing satellite availability for roof and platform Watford Junction receivers on 3rd November 2006

There is generally a good availability of satellites, with there always being at least 4 and up to as many as 11 on numerous occasions.

Due to the number of epochs represented in figures 6.34 and 6.35 and the possible size of any breaks being very short, it is unlikely that any breaks due to overhead cables would be seen in either of these figures.

The rough satellite positions seen in the skyplot (figure 5.20, section 5.3.2) are valid for the times when data was collected as the full days worth of data was collected without breaks.

As mentioned in section 5.3.2, the 10 satellites of specific interest are 2,4,10,17 and 24 for the northern section of sky and 1,6,14,25 and 30 for the southern.
When looking at these satellites and comparing them to the availability shown by figures 6.34 and 6.35, most of the northern satellites of interest are not present, in fact only satellite 2 has a significant amount of data at around 2 hours, with satellite 24 providing under an hour on both days. Given that the satellites are not available to either the platform or roof mounted receivers, it is assumed that either the orbital prediction program was incorrect, or other factors were involved. For satellite 4 it is possible that its availability was premature and so during the time of data collection it was missed, satellite 10 seems to be too low on the horizon but the absence of satellite 17 is unexplained.

When looking at the south western satellites, the situation is better with none missing for both days and both receivers.
This is ideal for the southern mounted overhead cables, but the northern set of cables also have the additional resistors mounted within the region of interest and so any possible effects from these may not be observed due to the poor availability of satellites with signals passing through that area.

6.3.3 Interference investigation

The first investigation into possible interference on satellite signals due to EM interference from overhead cables and the supporting infrastructure is an overview of the M_{P1} and M_{P2} multipath observables for the satellites listed in section 6.2.2.
Due to the platform receiver being static and the sky view being almost ideal, the number of breaks in data are minimal, thus the code multipath observables can be normalised with confidence. The following set of graphs (figures 6.36 – 6.39) show the M_{P1} and M_{P2} values for the platform receiver during the two days of observation for the northern and southern satellites respectively.

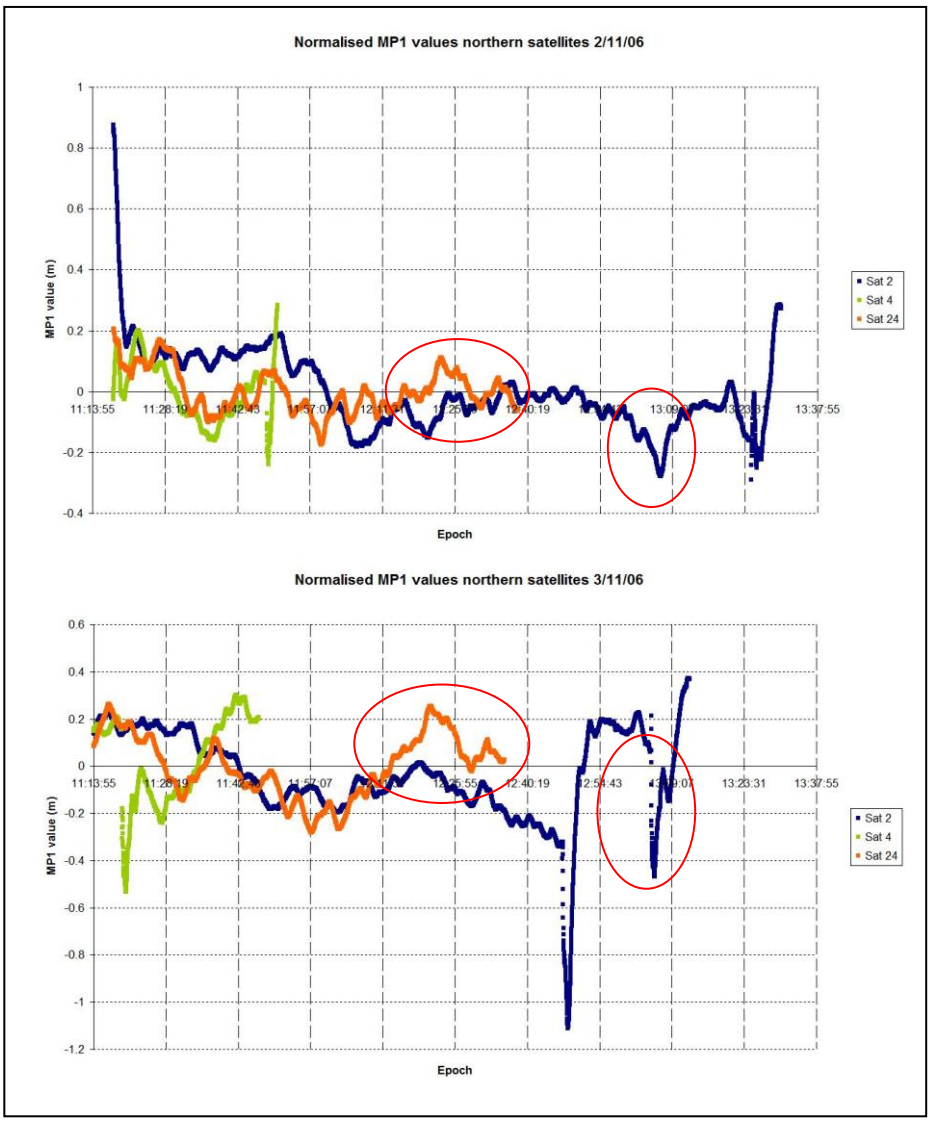


Figure 6.36 - Graphs showing MP1 variations for northern satellites on the 2nd and 3rd of November 2006

Figure 6.36 and 6.37 show the M_{P1} and M_{P2} values for the northern satellites for the two days of data collection. The obvious similarities between the two days have been shown in red but given the scale difference the magnitude of the M_{P1} and M_{P2} changes are different.

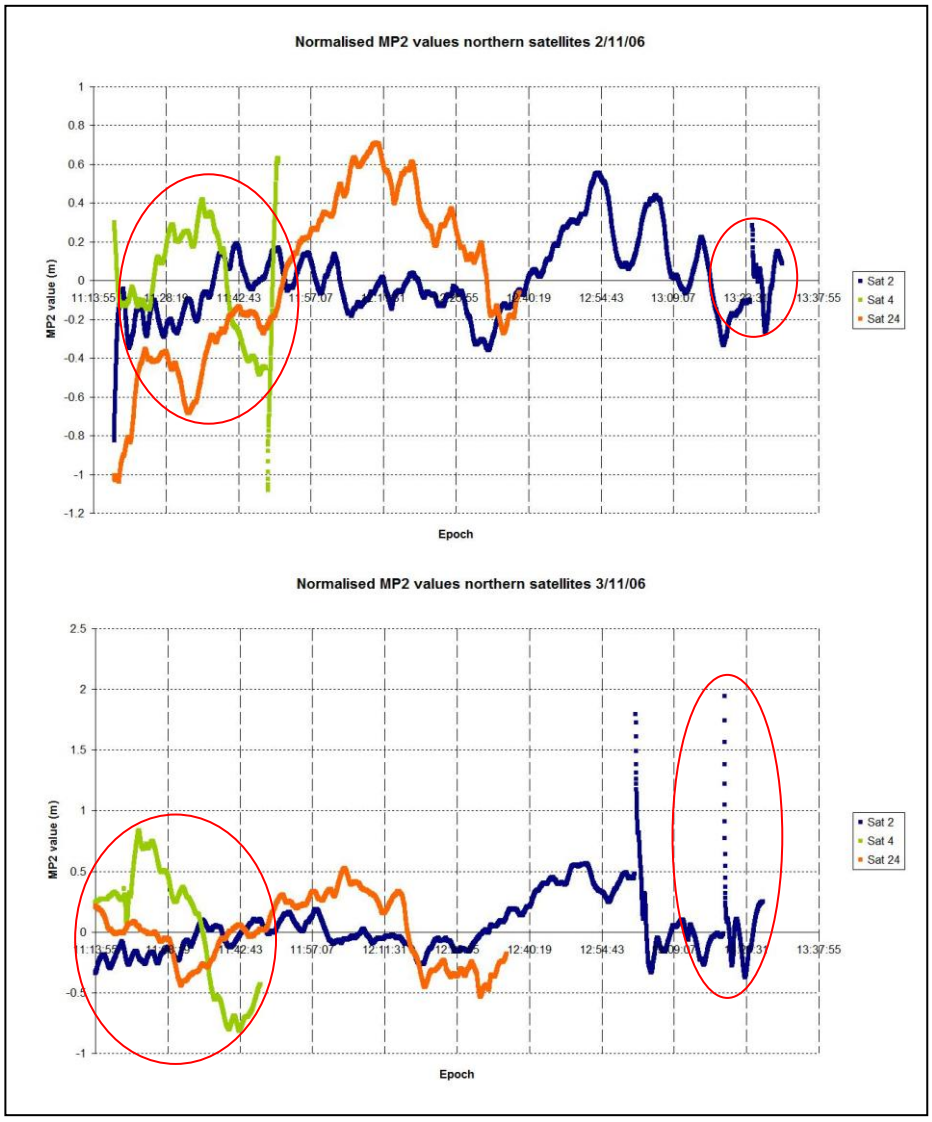


Figure 6.37 - Graphs showing MP2 variations for northern satellites on the 2nd and 3rd of November 2006

When viewing these figures, if there is a difference from one day to the next, it is assumed to be due to either a person passing the receiver causing an error, or due to freight trains that pass through as they are not timetabled and so do not repeat from day to day (Though timetabled trains do not tend to run on time anyway).
This can be seen on the 3rd of November for satellite 2 in both the M_{P1} and M_{P2} graphs. When looking at satellite 4 on the first day, towards the end, there is a significant increase in both M_{P1} and M_{P2}. This can be attributed to the low elevation of the satellite causing reflections from objects.

Satellite 24 is seen to be stable in its multipath fluctuation and follows a similar pattern for both days with M_{P1} and M_{P2}.

From this brief initial analysis of the northern satellite signals, there do not seem to be any sections of data that arouse suspicion of signal errors. The data will however be analysed further as the possible EM interaction with the overhead cables could cause a block of the signal instead of errors in its value. Due to the density of the data plotted in figures 6.36 and 6.37 any breaks in the data with duration shorter than about 2 minutes are very hard to spot and so further automatic processing is required.

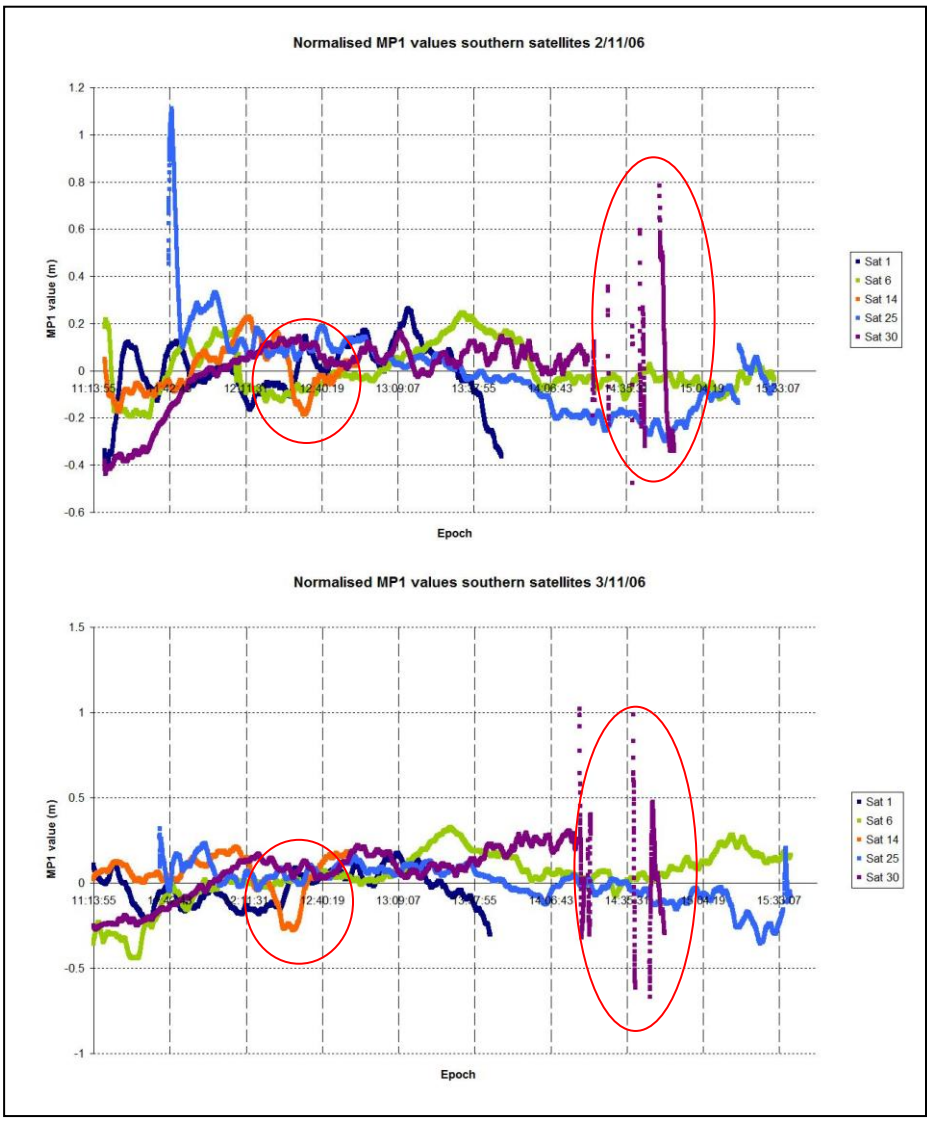


Figure 6.38 – Graphs showing MP1 variations for southern satellites on the 2nd and 3rd of November 2006

Figure 6.38 and 6.39 show the M_{P1} and M_{P2} values for the southern satellites for the two days of data collection again with the similar sections ringed in red.

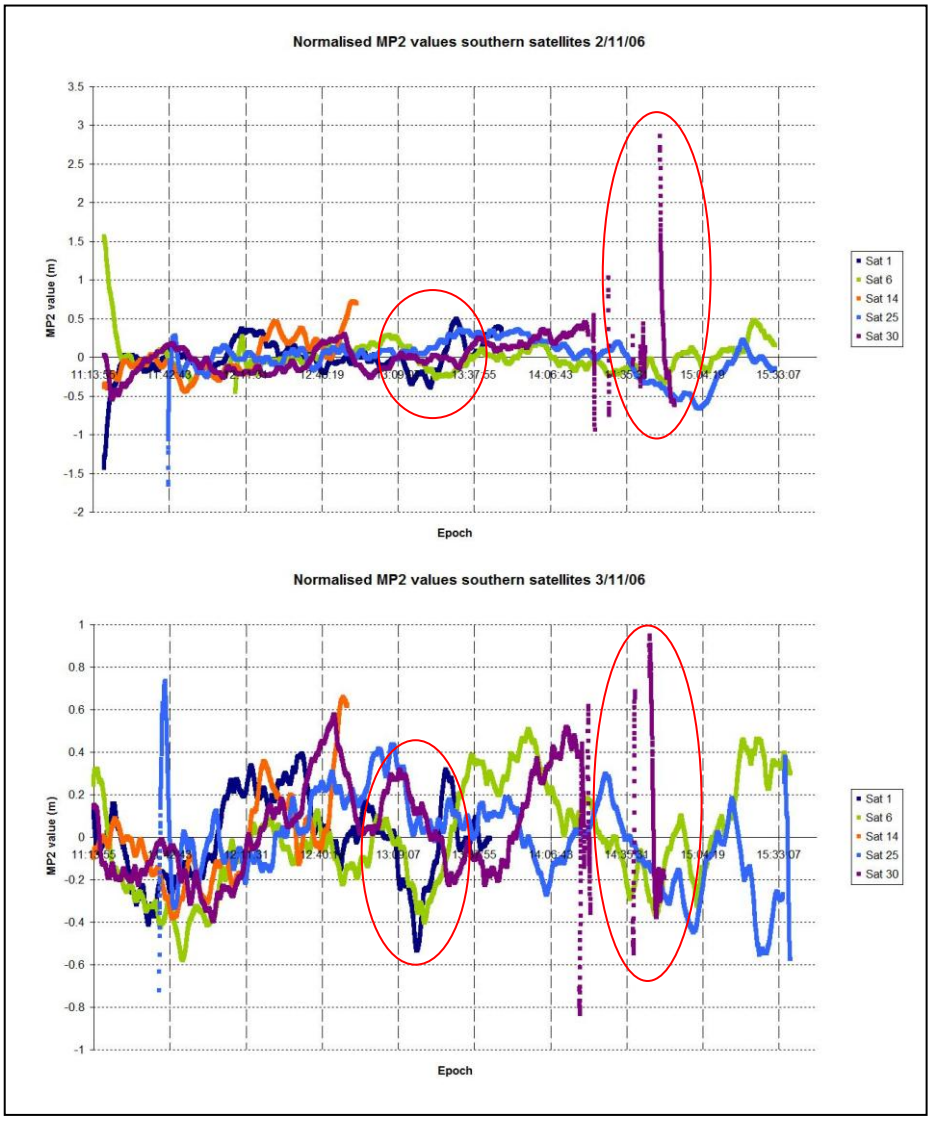


Figure 6.39 – Graphs showing MP2 variations for southern satellites on the 2nd and 3rd of November 2006

Satellites 1,6,14 and 25 do not seem to have any epochs of particular interest with regard to abnormal values, despite satellite 25 beginning with a large variation in M_{P1} and M_{P2}, this is most likely due to the low elevation of the satellite as it comes into view of the receiver.

Satellite 30 shows two distinctive peaks in M_{P1} and M_{P2} around 14.30. Despite this being near the end its period of visibility and thus possibly due to multipath from other objects low on the horizon, the initial spike in the code multipath observable is less likely to be due to this. On further investigation it was seen that the azimuth angle of the satellite towards the end of the data collection exercise was around 120° and so the cause of the multipath values and breaks in the signal were most likely the platform itself and the infrastructure supporting the overhead cables and signals.

Due to the apparent lack of EM interference from overhead cables altering the information on the incoming signals or distorting the code or phase measurements, the possible blocking of signals (or assumed rejection by the receiver hardware if scrambled) due to EM interference was investigated.

When analysing the full data set for both days manually, there were very few breaks inside the specific zones designated as containing EM radiation sources of interest. For both the northern and southern sets of satellites, only satellites 2 and 25 had breaks on the first day of data and satellites 2, 4, 6 and 25 had breaks on the second. Given that the break in signal from satellites 6 and 4 are not repetitions, they have been ruled out as being due to EM interference, leaving only satellites 2 and 25 for further investigation.

The results for these two satellites are shown in table 6.5.

Day	Satellite PRN	Epoch	Duration	Elevation	Azimuth
2/11/06	2	13:24:24	14s	16°	39°
	25	15:17:58	10s	29°	214°
3/11/06	2	13:01:18	3s	22°	45°
	2	13:18:57	6s	17°	39°
	25	15:34:41	9s	21°	210°

Table 6.5 – Table showing signal breaks, elevation and azimuth angles for satellites 2 and 25

The first break that can be removed is for satellite 2 on the 3rd. This is because it is very short and probably due to a passing passenger or freight train. It also has no correlated break on the previous day, ruling out a repeating error source.
When comparing the other break in this satellite, the durations are not the same, and the starting epoch is not inside the 4 minute window, but the elevation and azimuth angles are almost identical.

The breaks shown in table 6.5 were also cross checked with the reference receiver to make sure that there were no chances of satellite malfunction in order to qualify the breaks as being entirely due to the local platform environment.

Given that the receivers had to be set up both days in exactly the same position for the results to follow theory precisely, it is possible that the positioning of the receivers on both days were in fact not exact, and due to the constant rail traffic causing large vortices and ground movement, some additional movement of the receivers is possible. From this additional assumption and the fact that a minute change in the receiver position will only alter the elevation and azimuth a very small amount it is assumed to be the same object causing the obscuration.

The change in initial epoch is most likely due to the near-field effect of an obscuring object causing a larger effect on the visibility of far objects, thus a small change in the receivers position causes a significant change in the timing of satellite visibility. The effects seen on the signal from satellite 25 are not as closely correlated, with the difference in elevation and azimuth angles between the two days being 8° and 4° respectively. The time offset is also far larger than for satellite 2 and is not seen earlier the second day, but later so the likelihood of the two days observations being for the same obscuration are far lower.

By looking at the elevation and azimuth angles at which the satellites have been blocked, the local platform site was then checked to see if either the overhead cable EM radiation or local objects were to blame for the breaks.

6.3.4 Mapping errors to local infrastructure

From the final results of section 6.3.3, only 2 satellites are seen to be blocked or affected by the local platform infrastructure. The lines in figure 6.45 show the azimuth direction of the satellites relative to the location of the platform receiver.

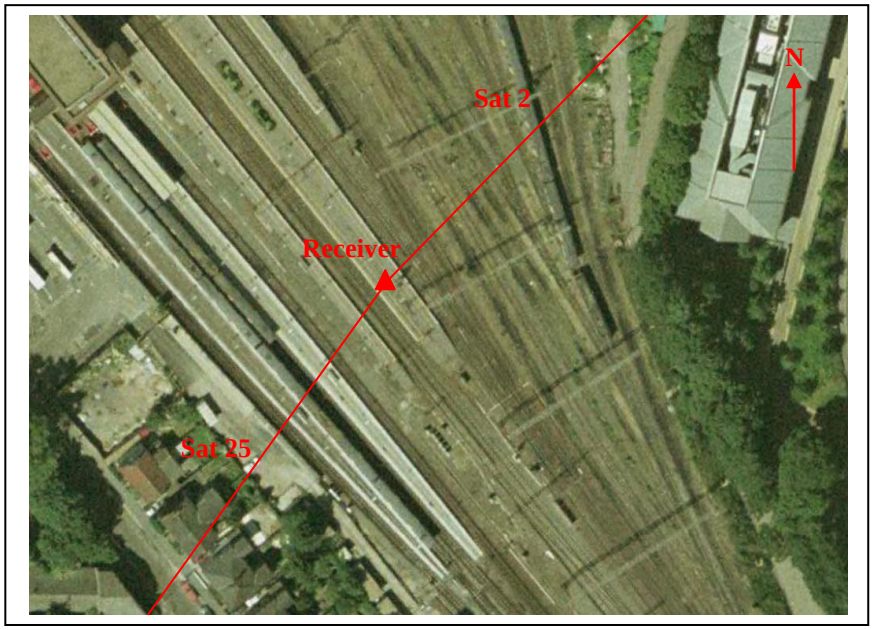


Figure 6.40 – Overhead view of Watford Junction platform receiver location, showing azimuth of satellites 2 and 25.

When checking the receiver locations for the northern side of the platform as seen in figure 5.19, the breaks in signal for satellite 2 coincide with the location of the northern pair of resistors in both azimuth and elevation. Figure 5.18 also shows the resistor pair to the left of the photo. Unfortunately, due to the timeframe of observation, there are no other satellites observed passing this point as the prediction program shows satellites 4, 10 and 24 passing in very close proximity.

The obscuration of satellite 25 on both days due to a fixed object has already been shown to be doubtful and so the next obvious step was investigated, the train times. Due to the movement of apparatus when trains enter and leave the station, exact timings of train arrivals and departures were not taken as the equipment had to be held to the ground. But on checking the timetable, the two times when satellite 25 was seen to have breaks in signal, at 17 and 35 past the hour, trains were due (± 4 minutes) and so this is believed to be the reason. The trains using the southern section of track were Virgin pendolino trains and so despite the train body not obscuring the sky to such a high elevation angle of 29°, the pantograph does in fact reach this high as it makes contact with the overhead cables.

We have already seen from analysing all the recorded data that the overhead cables do not seem to cause any interference or breaks in GPS signals as several satellites

have been observed through the area containing cables without issue and so the two breaks observed are assumed to be due to other factors.

The change in azimuth and elevation for the break points in the signal from satellite 25 can also be explained as being due to the slightly varying position of the train as it stops by the platform. Whereas with satellite 2 the differences were small, the 8° elevation and 4° azimuth change in break position for satellite 25 are significant. The second day's elevation and azimuth increasing by these amounts can be seen to fit with the physical situation of the train stopping fractionally further along the platform on the second day, thus increasing azimuth, and due to the platform being roughly perpendicular to 225° azimuth, the apparent elevation angle to the pantograph would also increase.
There is however the duration of the breaks that make the possibility of the pantograph being the cause less possible. Due to the breaks only lasting in the order of 10 seconds, this seems too short a timeframe for a train to stop at a platform and so the only additional reasoning for this may have something to do with gap discharge or the nature of the trains coupling with the overhead cables once the train has reached a standstill. Further investigation would be needed to confirm this theory.

6.3.5 Conclusions

Due to an increasingly large proportion of the UK rail network adopting overhead electrification as a form of rolling stock propulsion, any possible effects from the overhead lines need to be investigated if GNSS is to be used for signalling and control purposes.
The experiment involved using receivers mounted near to the overhead cables on Watford Junction platform and on a nearby roof free from obstructions to the sky. By calculating the code multipath observable, any possible errors or abnormal values in either the code or phase for the L1 and L2 frequencies would be shown when plotting a time series.
Any values of interest were first checked with the reference receiver to see if they were in fact due to the localised platform environment or the constellation itself.

Further to this, individual investigations were carried out in order to determine the possible reasons for the errors.
This is important when considering the evaluation of continuity and availability RNPs due to the breaks in signals possibly causing a loss of performance within the required limits or, ultimately a total loss of position integrity and therefore a break in system availability.

The results show that the only effects seen by the receiver due to overhead infrastructure were obstruction by larger objects, such as the in-line resistors and train mounted pantograph. These obstructions may be due to the physical size of the obstruction or due to the possibility of corona or gap discharge as the pantograph maintains a dynamic coupling with the cable and the resistors have been noted as a possible source of electrostatic build-up leading to small discharges, but the limited time noted for the breaks require further investigation as the source of signal obscuration may not be the physical object of the pantograph itself.

The plots created for the code multipath observables show that there are no apparent signal errors or range anomalies due to the cables themselves or the surrounding area.
No cycle slips were observed whilst the GPS signals were passing through the cabled area, increasing confidence in the correct operation of GPS in the presence of overhead cables.
When mapping the individual signal breaks to physical objects located in the near vicinity, the inaccuracies in receiver position caused a certain level of uncertainty and the advantage of repeating observations over the two days was diminished slightly.
Ideally, the local area would have undergone a full survey using either a total station and pogo or a laser scanner mounted in exactly the same position as the receiver. This would have allowed precision mapping of any signal breaks to individual infrastructure details in an automated manner.

The overall levels of multipath experienced by the rover receiver were also a good indication of what a train would see in an open air station like Watford Junction.

The level of multipath variation on L2 is again seen as being larger than for L1, with the average over the two days being ~0.65m for L2 compared with ~0.45m for L1. When looking at extreme values for multipath fluctuation, the L2 signal was more prone to large differences, reaching a change of 3.5m for satellite 30 on the 3rd of November compared with the maximum L1 variance being 1.5m for satellite 2 on the 2nd and satellite 25 on the 3rd or November.

These results should be integrated into any HAZOP process evaluating GPS interference due to overhead lines as the analysis shows that there are no detectable errors and that they should therefore not cause an issue in the operation of GPS in the vicinity of quaternary overhead power lines.

Further work towards a more stringent and quantitative value for the possible emission of interfering EM frequencies from overhead lines would require the use of specialist equipment to monitor EM fields in an anechoic chamber. The results from such an experiment would hopefully determine exact values for a standardised overhead power network that could be included in a final safety case for the use of GNSS under power lines.

Due to the effects not being detected in this experiment, the FMEA worksheet would not contain EM radiation from power lines as a possible cause of system failure and thus would not require any additional features to mitigate its risk.

Given the levels of confidence required for SIL qualification, the data collected and analysed in this report would not be considered sufficient for a level to be determined, but through extrapolation and further testing, the possibility of undetected errors may be determined, though it is expected to be of a very low likelihood given the results already shown.

6.4 Birmingham data analysis

6.4.1 Introduction

Following on from the results obtained from the Severn Valley railways data collection and analysis, the methods of analysis to be used for the Birmingham datasets have been refined and developed.

The use of Google Earth is again ideal for visually representing the position data and the local infrastructure effects on satellite visibility and interference.

Due to the high volume of data, statistical testing can be carried out in order to determine quantitative statements about the variability of multipath and possibly its dependence on the length of break-free observations for each satellite.
Although inertial sensor data is available to accompany the GPS data recorded, for the purposes of this thesis, this data is not included and will not be considered.

6.4.2 Initial data analysis

Given the large volume of data collected, individual files will not be referenced, but the day of collection will be referred to for particular data sets. The two main lines, Birmingham – Redditch and Birmingham – Wolverhampton will be dealt with separately and any differences in results noted.

The first data sets recorded in November will be broken into the two routes according to the day's schedule. The data collected varies in length depending on how many journeys are carried out and which line the train is operating on.
The first two days data cover the two lines and will be part of the initial analysis, then the days when these routes are repeated will be integrated into the investigation in order to statistically improve the results obtained.

The first day of data collected (2nd November 2007) covers the Birmingham to Wolverhampton line several times, as can be seen from the longitude and latitude plots in figures 6.41 and 6.42 from code pseudorange measurements processed using the RTK library in point positioning mode.

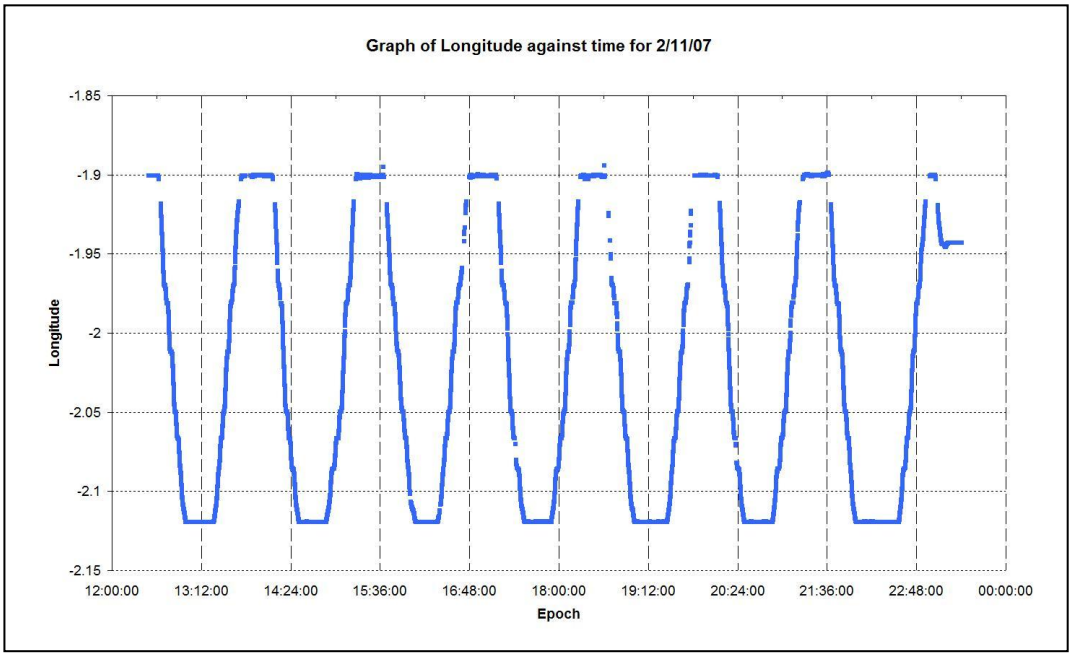


Figure 6.41 – Graph of trains longitude against time for 2/11/07

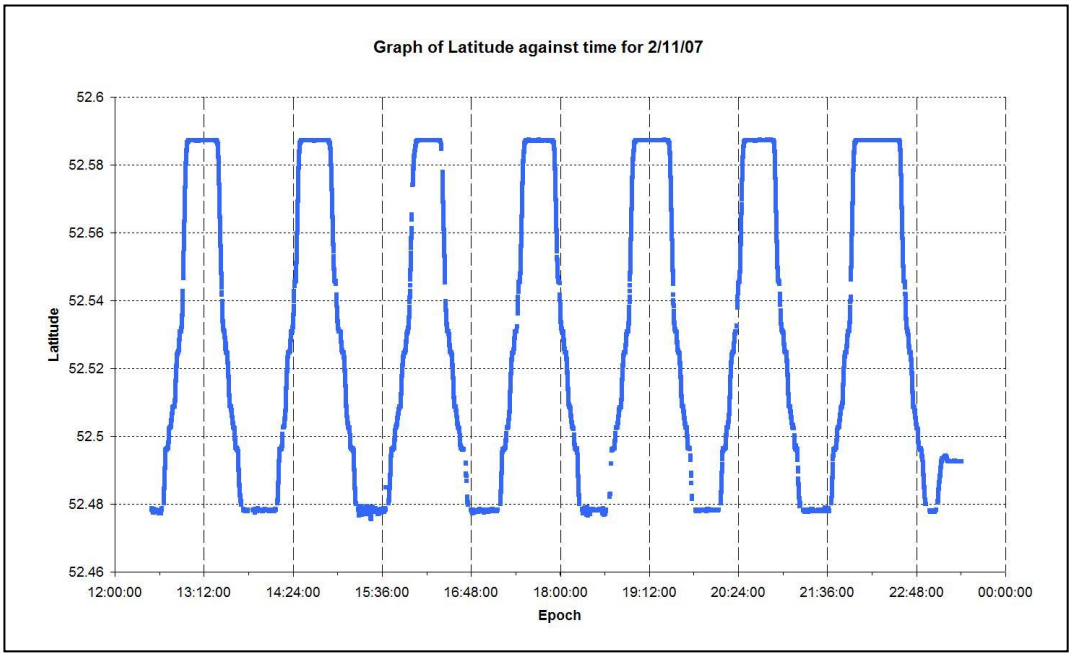


Figure 6.42 – Graph of trains latitude against time for 2/11/07

The train is seen to repeat the return journey 7 times during the running time of 12:30 to 22:50 with the final leg showing the train travelling to the depot.

The large repeated breaks seen in the longitude plot are due to New Street station and tunnel and can also be seen as smaller breaks in latitude due to the latitudinal change when entering the station being minimal compared to the longitude.

This can be seen in figure 6.43 when plotting the Birmingham – Wolverhampton track coordinates.

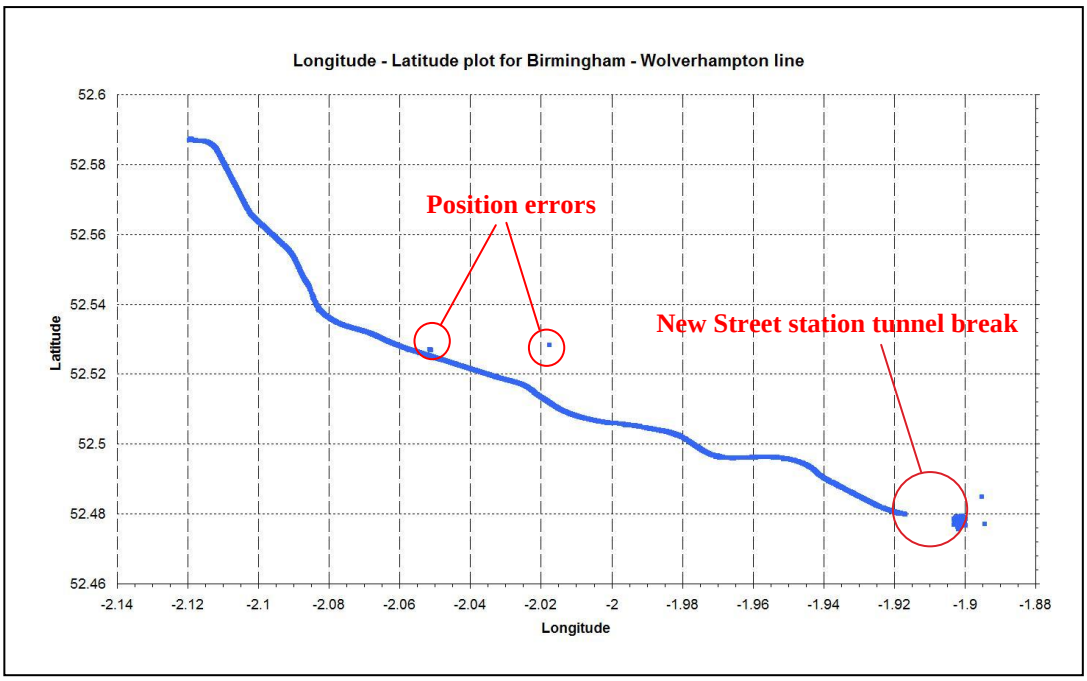


Figure 6.43 – Plot of Longitude against Latitude for Birmingham – Wolverhampton line

The large break in data is mainly due to the tunnel at the western entrance to New Street station and is made longer by the lack of re-acquisition of satellites signals after the tunnel due to the roads and bridges located above the tracks in and around the main station area. Once the train has entered the platform area for a while and is able to re-acquire enough satellites for a position, the errors are large due to the surrounding buildings and infrastructure.

The second day of data collected (3rd November 2007) covers the Redditch to Lichfield line which also passes through Birmingham New Street station. The number of repeated runs can be seen from the longitude and latitude plots in figures 6.44 and 6.45.

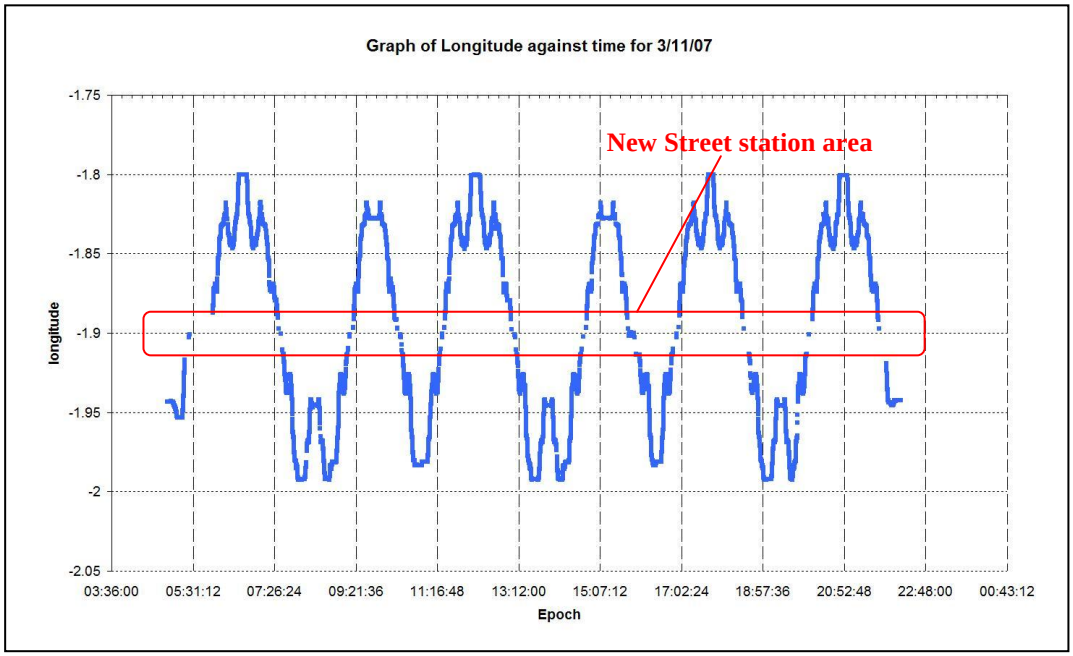


Figure 6.44 - Graph of trains longitude against time for 3/11/07

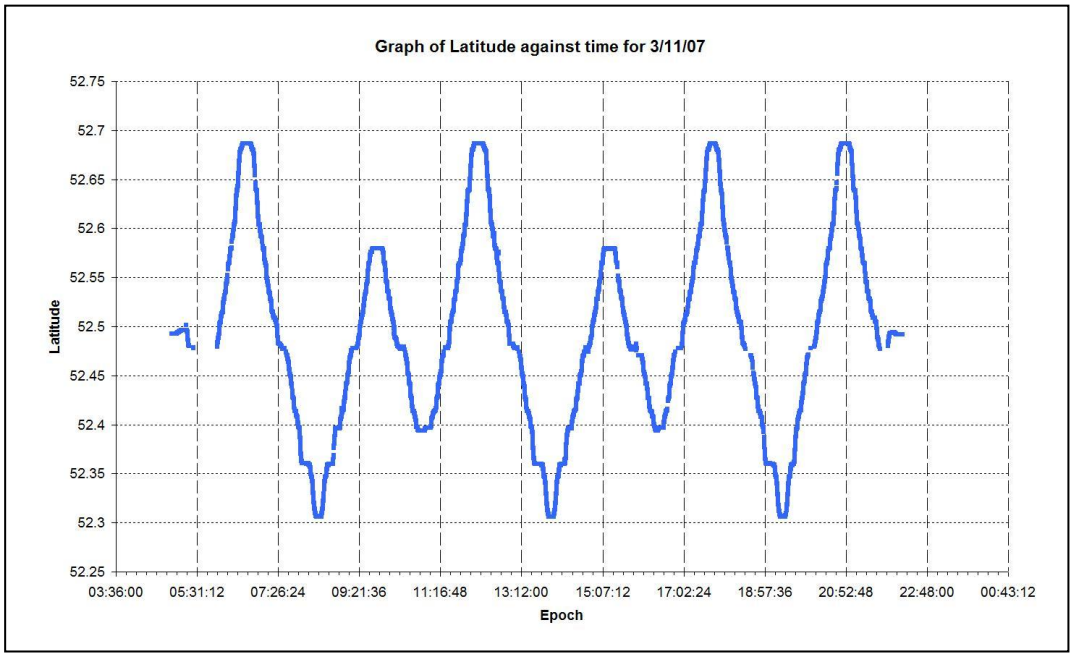


Figure 6.45 - Graph of trains latitude against time for 3/11/07

The main route that is split into two sections by New Street station can be seen to repeat 4 times during the day. The longitude plot also shows Birmingham New Street station as the small breaks around -1.90°.

Figure 5.45 shows the route in its entirety, with New Street station in the centre where a return journey to the depot near Smethwick is also plotted.

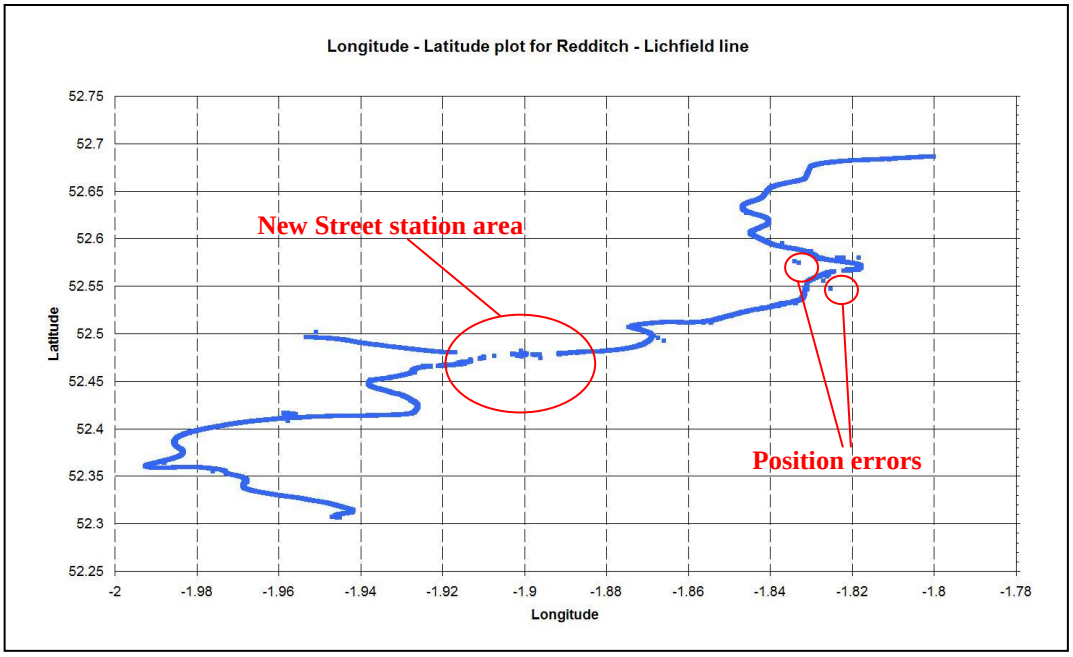


Figure 6.46 – Plot of Longitude against Latitude for Redditch – Lichfield line.

As with figure 6.43, several position errors are visible, even at this scale. These are mostly due to loss of satellites and poor satellite geometry and will be briefly investigated in the next section as the reasons are similar to those previously seen in the analysis of the Severn Valley data in section 6.2.

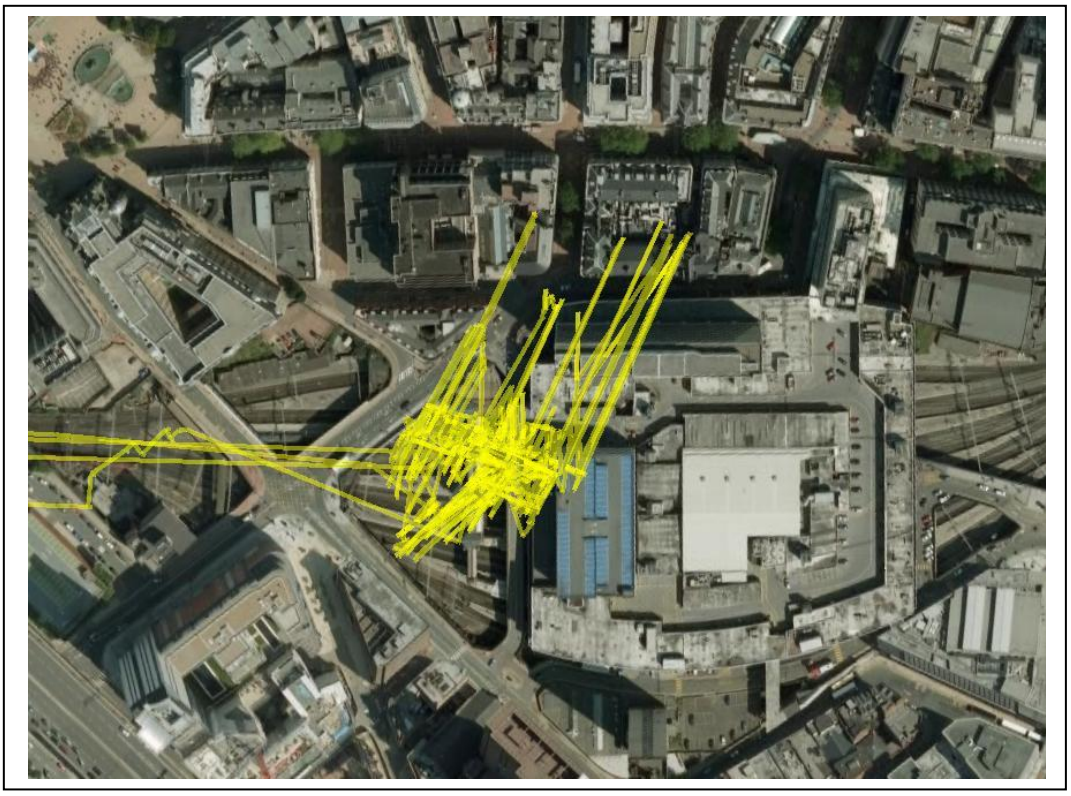


Figure 6.47 – Picture showing Birmingham New Street station train position errors for data on 2/11/07 caused by surrounding infrastructure

By using Google earth, the positions calculated using code pseudoranges can again be used to visually show the effects of the local infrastructure. The effects of the New Street station infrastructure are shown in figure 6.47.
The position errors due to multipath will be discussed in section 6.4.3 but the errors due to satellite availability will be discussed next.

When looking at satellite availability for the entire data set for 2/11/07, figure 6.48 was produced when plotting against time and longitude.

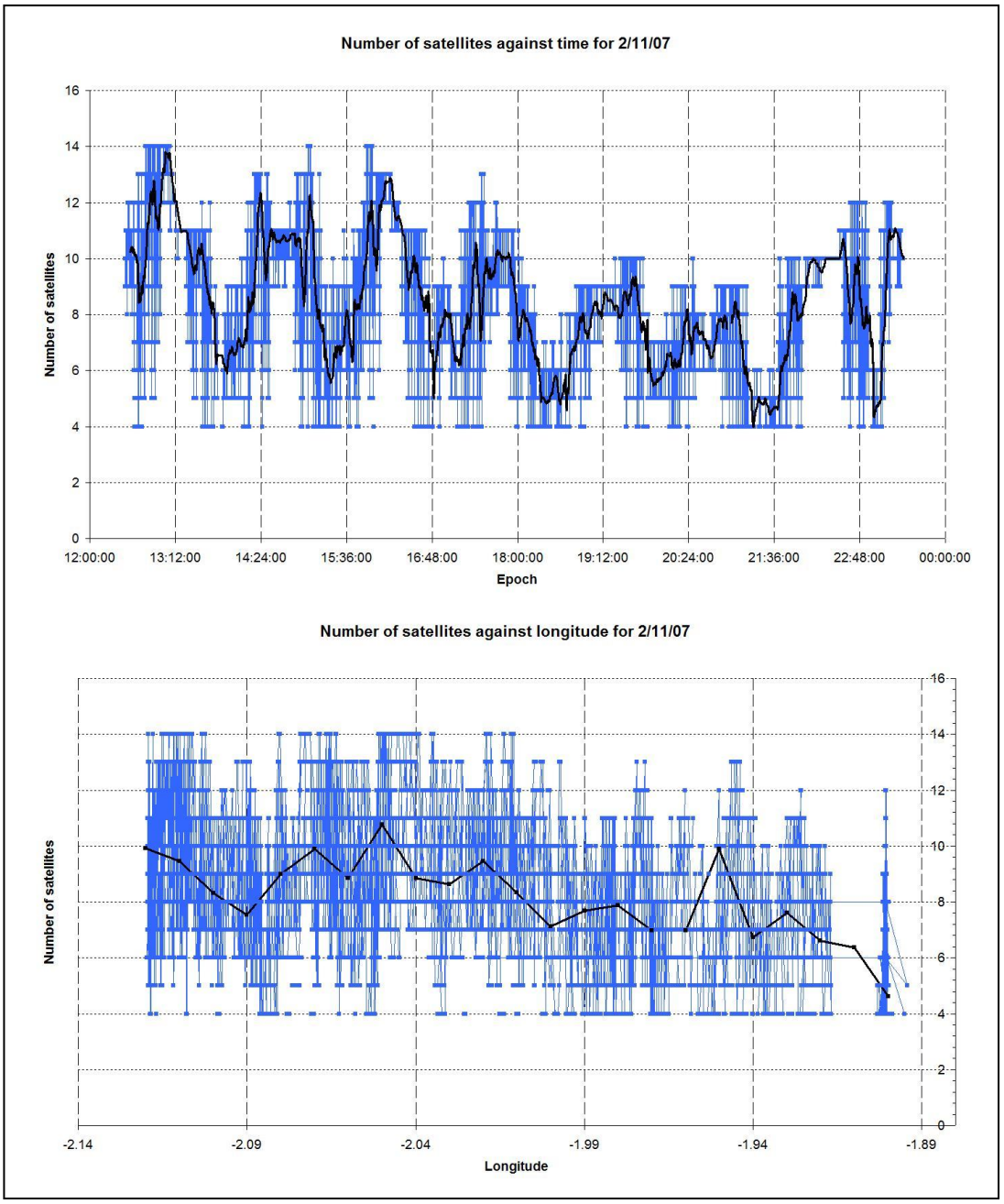


Figure 6.48 – Plots of satellite availability against time and longitude for all data recorded on 2/11/07

The time based plot shows a visible cyclic pattern that repeats 7 times during the day. The minimum number of satellites shown is 4 and if fewer are visible then a break in data is produced.

This output has been chosen as the focus of this research is to assess the errors on signal and position when one is obtained and not on implementing alternative positioning algorithms to obtain a solution with fewer than 4 satellites, and so epochs where there are fewer than 4 satellites will be seen as blank.

The graphs in figure 6.48 show that throughout the entire journey, then number of satellites in view change rapidly and does not stay constant for very long. It can be seen that towards the end of the day, the maximum number of visible satellites reduces from 14 to 12, with the period of 18:00 to 22:00 only having a maximum of 10. The average number of satellites for each 1 minute period has also be plotted as a black line on the time varying graph in order to remove some of the difficulty encountered when dealing with such a high volume of data.
The black line on the longitude graph has been broken into 0.01° averaged sections and shows that the average number of available satellites near the station reduces to below 5 per epoch.

When looking at the repeated pattern during the day, the number of satellites visible when the train is in New Street station is also reduced, with this being more obvious later in the day. This can again be attributed to the large number of high rise buildings in and around the station blocking incoming signals.
The small section of data from 22:50 onwards shows availability of satellites when the train is in transit from New Street station to the local depot at the end of its daily service.

As the aim for this research is to aid in the construction of a safety case, the worst possible operating conditions must be considered when analysing data. Despite the average number of satellites not falling below the minimum 4, there are a large number of breaks in the data and certain sections of track contain obstructions that hinder the performance of GPS.
One such area is along from the Smethwick Rolfe Street section of track where several overhead bridges in close succession cause the position errors seen in figure 6.49.

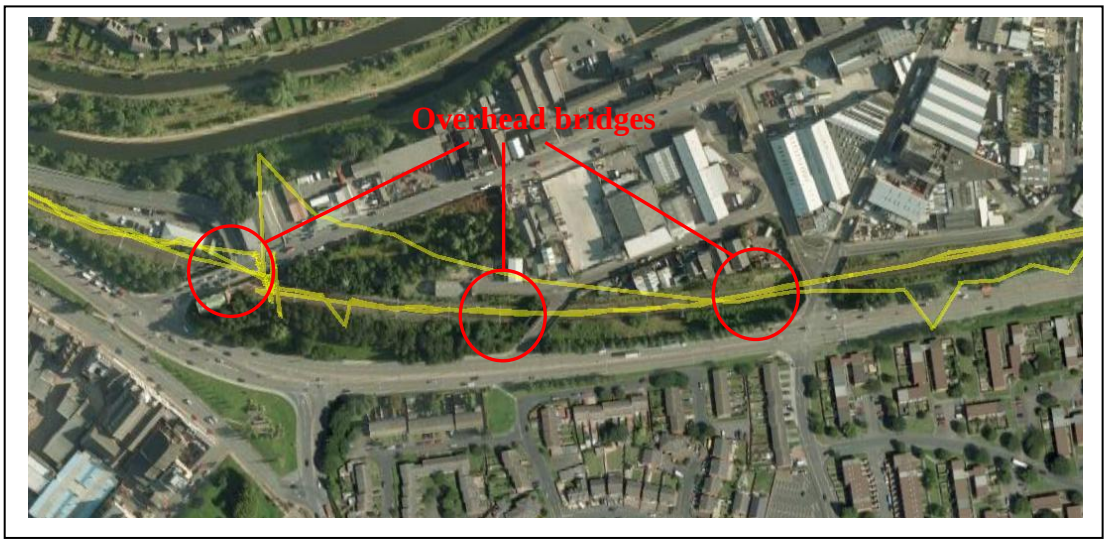


Figure 6.49 – Plot of train positions in Google Earth near the Smethwick Rolfe Street area.

The figure shows all the runs from 2/11/07. The times when the train passes through the area pictured are shown in table 6.6, the timetable can be seen to repeat itself at roughly 10 and 40 minutes past the hour.

Train enters section	Train leaves section
12:41:31	12:43:23
13:37:31	13:39:07
14:13:37	14:15:15
15:10:32	15:12:37
15:44:01	15:46:03
16:39:20	16:42:05
17:13:49	17:16:01
18:11:42	18:13:25
18:42:25	18:44:12
19:43:07	19:44:13
20:11:55	20:13:56
21:09:37	21:11:18
21:41:26	21:43:26
22:49:49	22:51:30

Table 6.6 – Table showing times when train enters and leaves the overhead bridge section near the Smethwick Rolfe Street section of track

Due to figure 6.48 representing a full day of data, the resolution for the section of track being investigated is poor, because of this figure 6.50 is plotted in order to show the satellite availability for the area in greater detail.

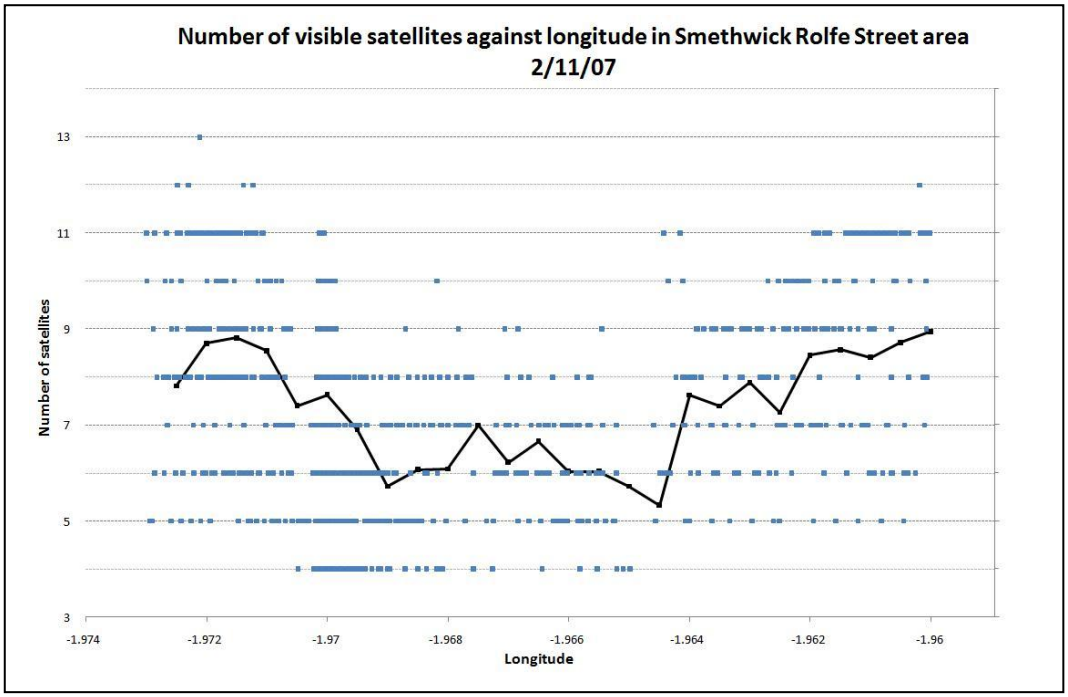


Figure 6.50 – Graph showing number of satellites and average number against longitude in Smethwick Rolfe Street area for 2/11/07

The average number of satellites for sections of 0.0005° (roughly 46m long) is also plotted on the graph as a black line due to the rapidly changing values making determining the most common number of satellites at each point difficult.
The gaps in data are also clear as when the train passes under the bridges, sometimes complete satellite obscuration occurs. Smethwick Rolfe Street station is located at 1.970° longitude and can be seen on the graph as having a greater density of data collected and with an average number of around 7 visible satellites.
The run that produced the highest position deviation from the track was investigated further so as to ascertain whether or not the position jumps were due to satellite visibility.

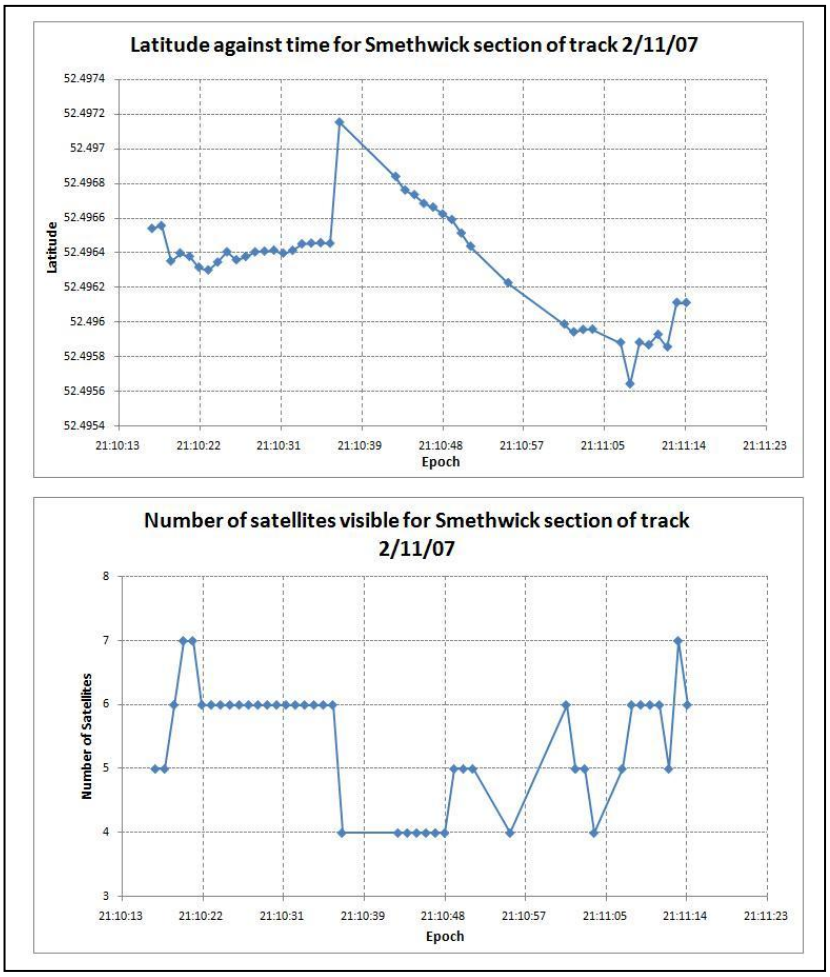


Figure 6.51 – Graphs showing latitude and number of visible satellites for Smethwick section of track for highest position error run.

The graphs show that due to a loss of 2 satellites a large position jump occurs, there are also a large number of gaps in the data where the number of satellites has fallen below 4.

When investigating the specific satellites that come into and go out of view, two satellites have very low elevation angles as well as azimuth angles that are perpendicular to the direction of travel. The general fluctuation in the number of visible satellites can be attributed to these factors, with the loss of more satellites most likely being due to the overhead infrastructure present in the area.

This pattern is also seen in figure 6.48 where the average number of visible satellites is greater earlier in the day when more satellites had a higher elevation angle than later in the day. The number of satellites visible still fluctuates, but towards the end of the day, the lower elevation angles of some start to remove them from view altogether.

These results mirror those seen for the Severn Valley Railway data analysed in section 6.2 whereby when there are a limited number of satellites available and one satellite is lost, the computed position can change significantly. Because of the earlier conclusions, the data collected from Birmingham will not be fully analysed for satellite availability issues causing position shifts.
For this reason, the next section will involve the dual frequency investigation of the data utilising the code multipath observables M_{P1} and M_{P2}.

6.4.3 Dual frequency investigation

An initial assessment of the level of change in M_{P1} and M_{P2} for all the satellites in view was carried out for the first two days of operation as both of the major lines under investigation are covered during this time.
The results can be seen in figures 6.52 to 6.55 for seconds of day.

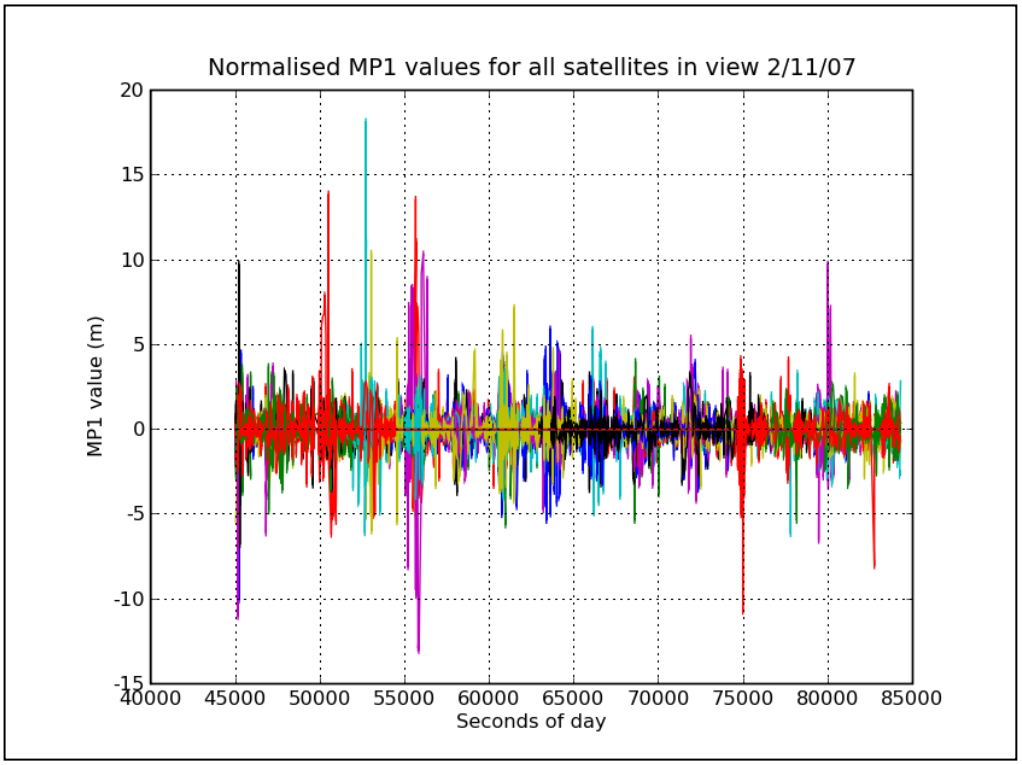


Figure 6.52 – Graph showing MP1 values for all visible satellites (colours represent different satellites) on 2/11/07 (Birmingham – Wolverhampton line)

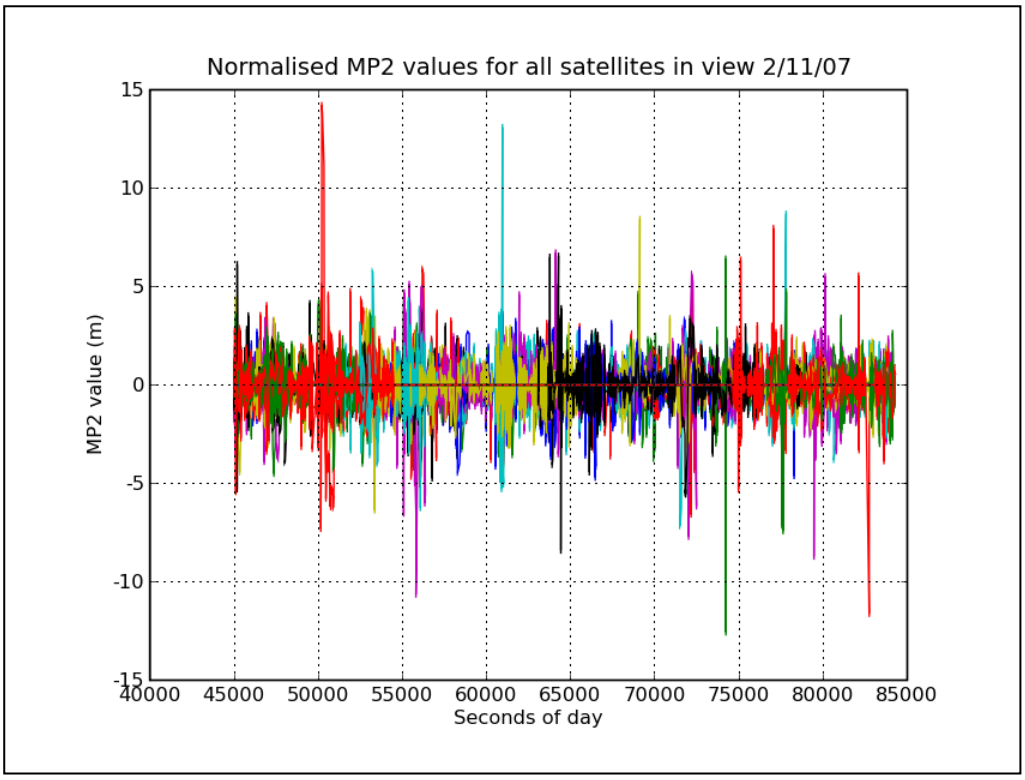


Figure 6.53 - Graph showing MP2 values for all visible satellites (colours represent different satellites) on 2/11/07 (Birmingham – Wolverhampton line)

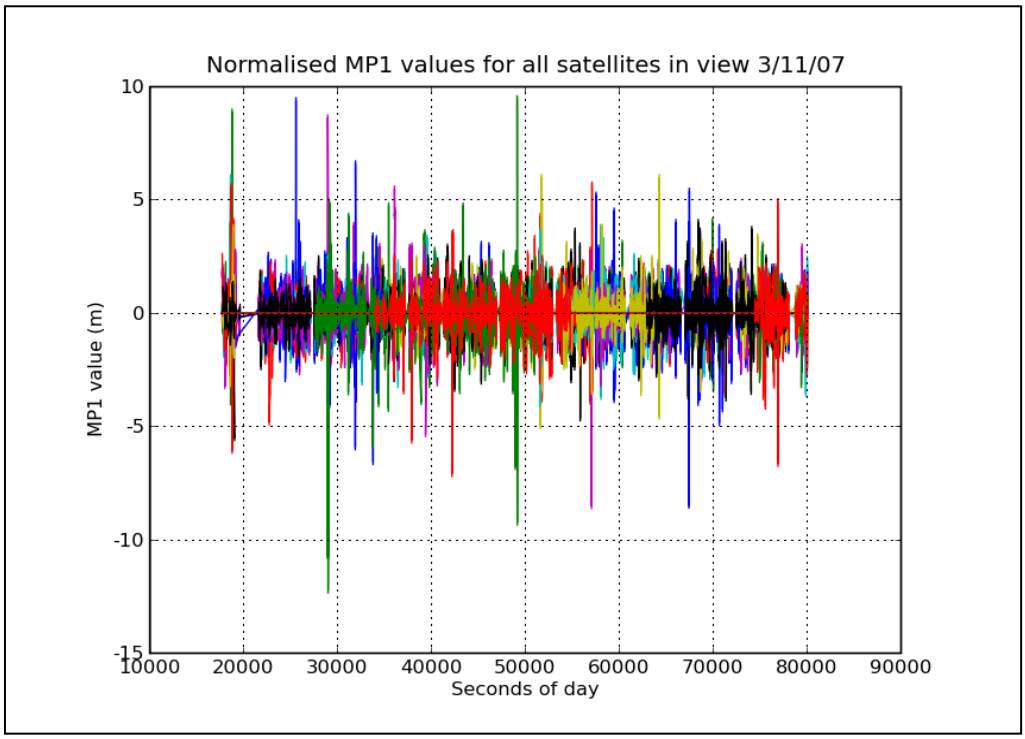


Figure 6.54 – Graph showing MP1 values for all visible satellites (colours represent different satellites) during first half of 3/11/07 (Redditch - Lichfield line)

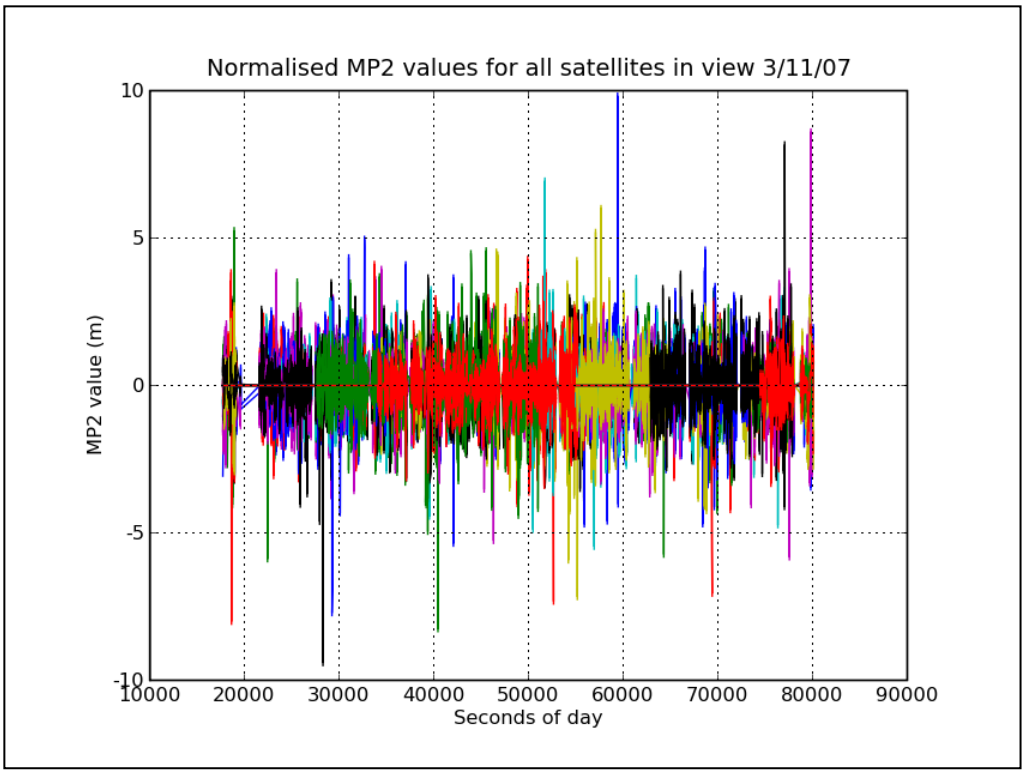


Figure 6.55 - Graph showing MP2 values for all visible satellites (colours represent different satellites) during first half of 3/11/07 (Redditch - Lichfield line)

As can be seen from the figures, the average change in M_{P1} and M_{P2} for both lines rarely exceeds 4m. The extreme cases are assumed to be due to specific local infrastructure such as overhead bridges and high buildings or due to long periods of break free data at stationary points providing significant variation.

The M_{P1} values seen for the 2nd of November generally vary by around the same amount as the M_{P2} values for the same day but extreme cases where high levels of multipath are observed for shorter periods of time are more common and larger on M_{P1} than on M_{P2}. This is contrary to the usual expectation that due to the L2 signal being slightly weaker and more prone to interference, a higher multipath value than that for L1 is normally seen.

This specific section of data will be investigated, but when comparing it to figure 6.44 and 6.45 it can be seen that the train is inside New Street station where high levels of reflecting surfaces from the surrounding high rise structures are present.

The values of the multipath recorded on both L1 and L2 frequencies were also checked for a possible bias in terms of the distribution.
Given that the graphs in figures 6.52 to 6.55 contain so much information, each value has been placed in integer sets to show the distribution of multipath values. A Gaussian distribution is formed when graphing the results, as shown in the following figures:

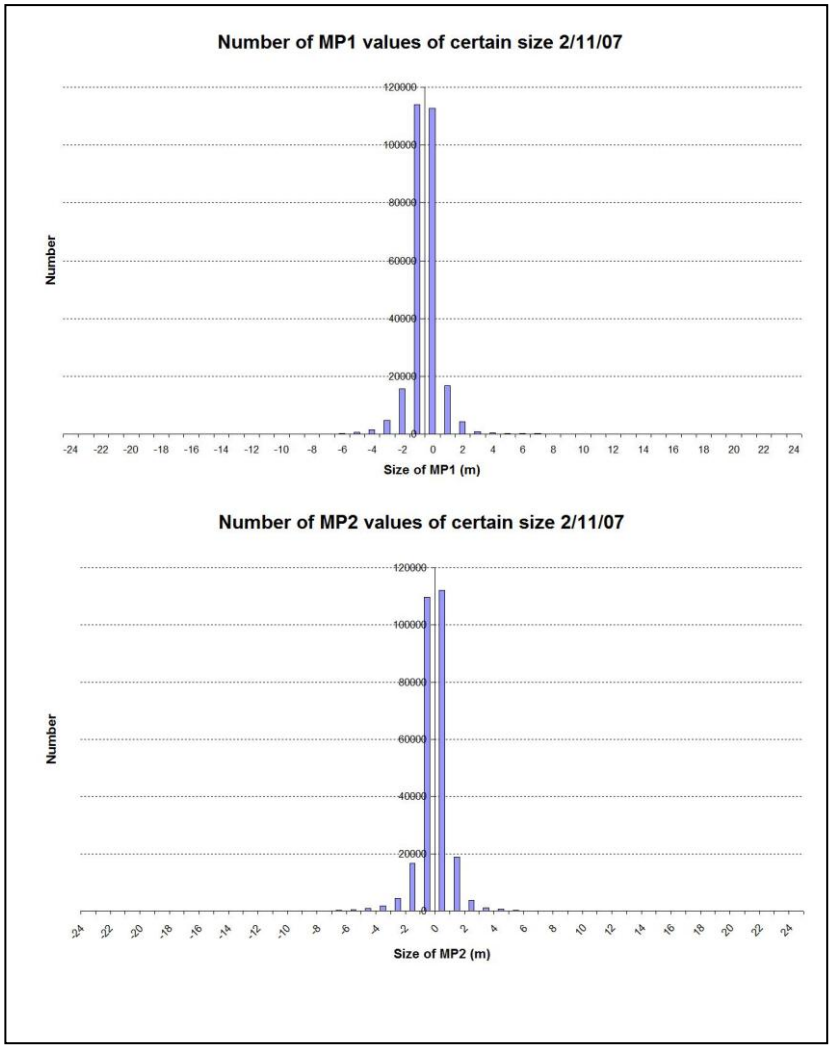


Figure 6.56 – Graphs showing distribution of normalised MP1 and MP2 values for 2nd November data

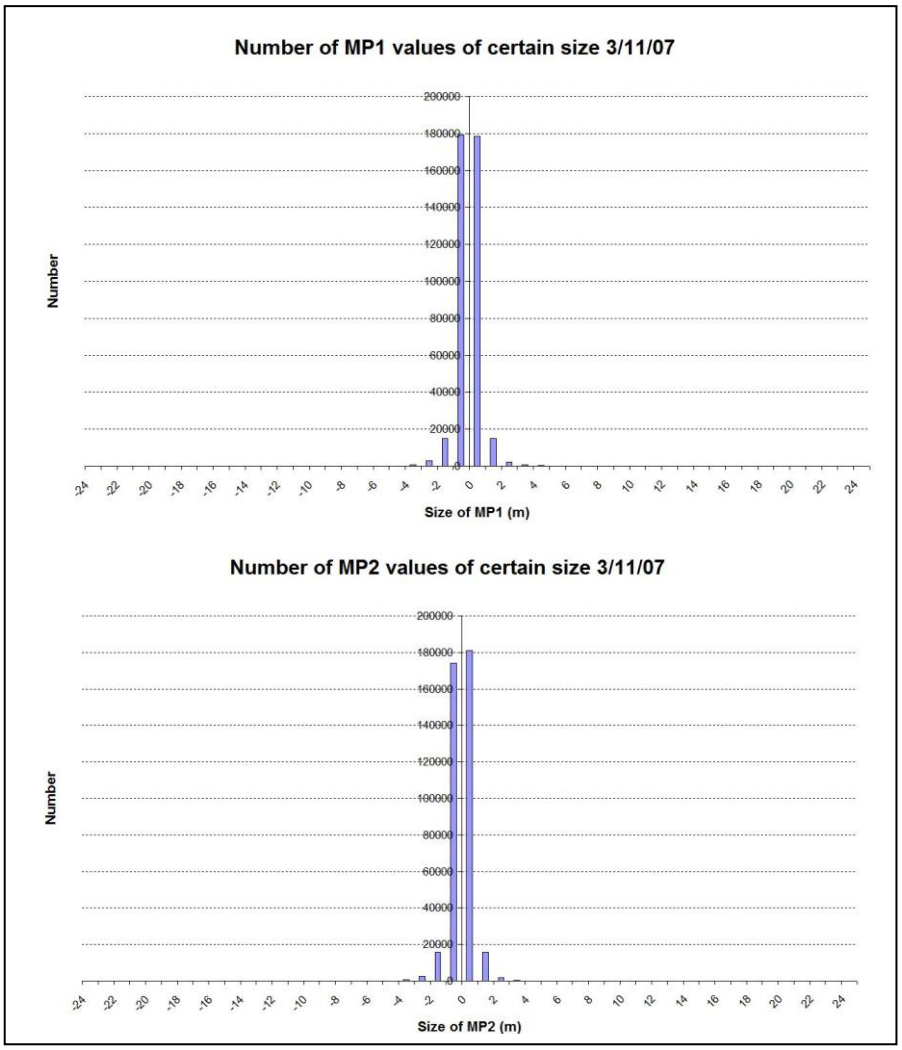


Figure 6.57 - Graphs showing distribution of normalised MP1 and MP2 values for 3rd November data

The results in figures 6.56 and 6.57 show the distribution properties for the multipath values centred around 0 with a very narrow distribution.

The distributions do not look 'normal' due to the very narrow distribution of values so although there are values distributed over a large range of M_{P1} and M_{P2} values, the centrally located values are so highly populated that the multipath values over 4m are relatively invisible.

This gives a clear indication that the expected level of change in multipath for the two main tracks is within 4 meters or so.

When discussing safety critical systems, σ values are normally used to describe the distribution of possible outcomes (1σ contains 84% of the data, 2σ contains 97% etc.) if the data fits a normal distribution.

If σ values are to be used to quantify the expected values of multipath in the rail environment, the results seen in figures 6.56 and 6.57 have to be normally distributed.

The results obtained were checked using the χ^2 test. This was done by first calculating the expected normal distribution using the standard deviation and mean of the collected data and then compared against the actual results as seen in 6.56 and 6.57.

The χ^2 test determines if the results obtained follow the expected normal distribution by using equation 6.1.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad (6.1)$$

Where n is the frequency of each category

O_i is the observed value

E_i is the expected value

Once the χ^2 value is calculated it is compared with tabulated values corresponding to the same number of degrees of freedom (categories present) and depending on the value being below or above the table value, the distribution is considered as being normal or not.

The results shown in 6.56 for the Wolverhampton line were analysed and found to be vastly differing from the expected normal distribution. Initially the full set of data where categories have 5 or more values present were graphed and compared, as seen in figure 6.58 for M_{P1} and M_{P2} values. Given that such a wide distribution was used to calculate the mean and σ values, a smaller, centrally located subset was then used (from -6m to +6m) to see if that changed the resulting distribution function. The resulting graphs are shown in figure 6.59.

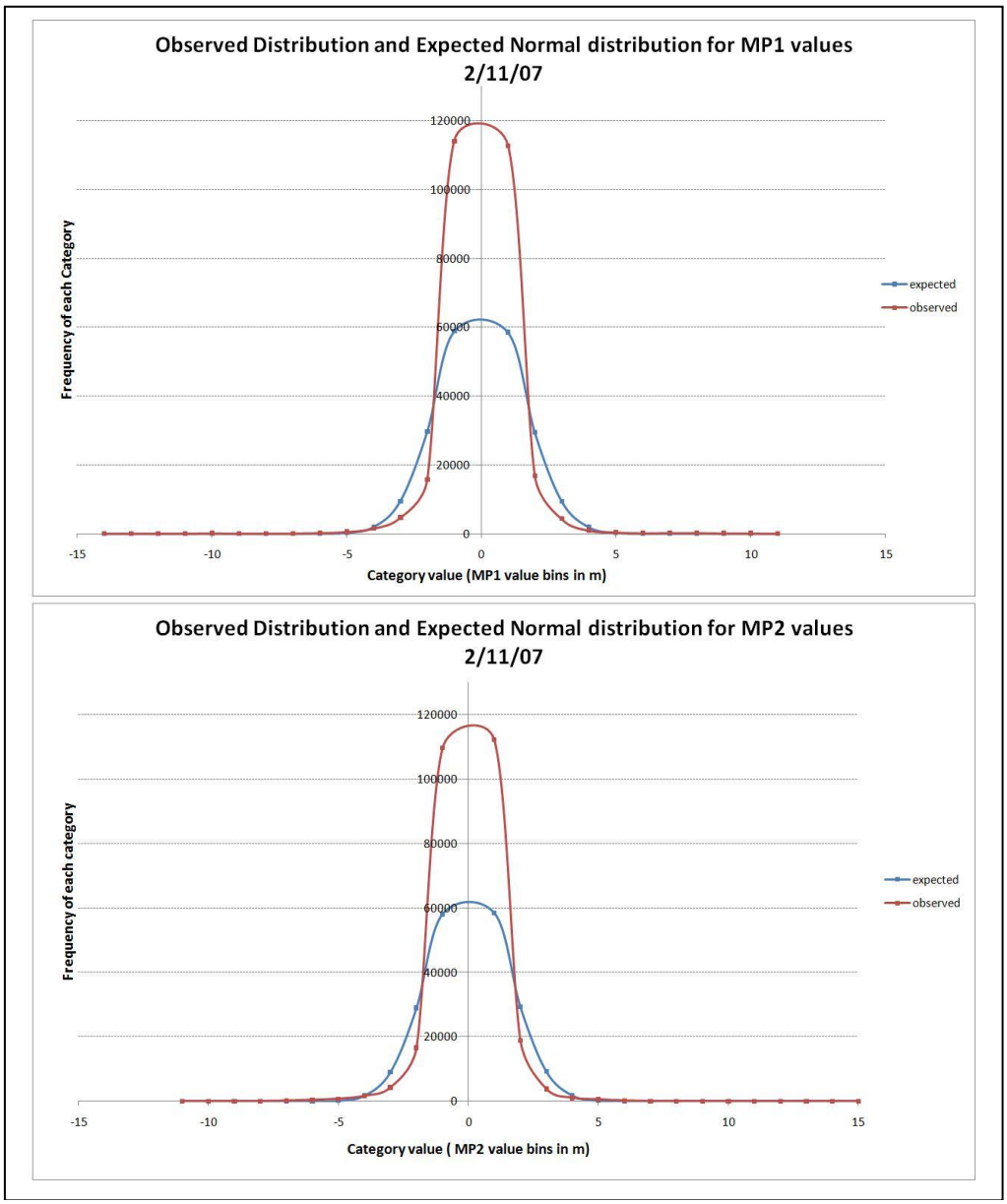


Figure 6.58 – Graphs to show observed distribution and expected normal distribution for MP1 and MP2 values 2/11/07

From 6.59 it can be seen that despite reducing the spread of data used, the observed values still do not follow a normal distribution.

The next check to see if a normal distribution was possible involved reducing the size of the bins in which the multipath values were stored. By reducing the size from 1m to 0.5m, the same calculations were carried out and figure 6.59 was produced.

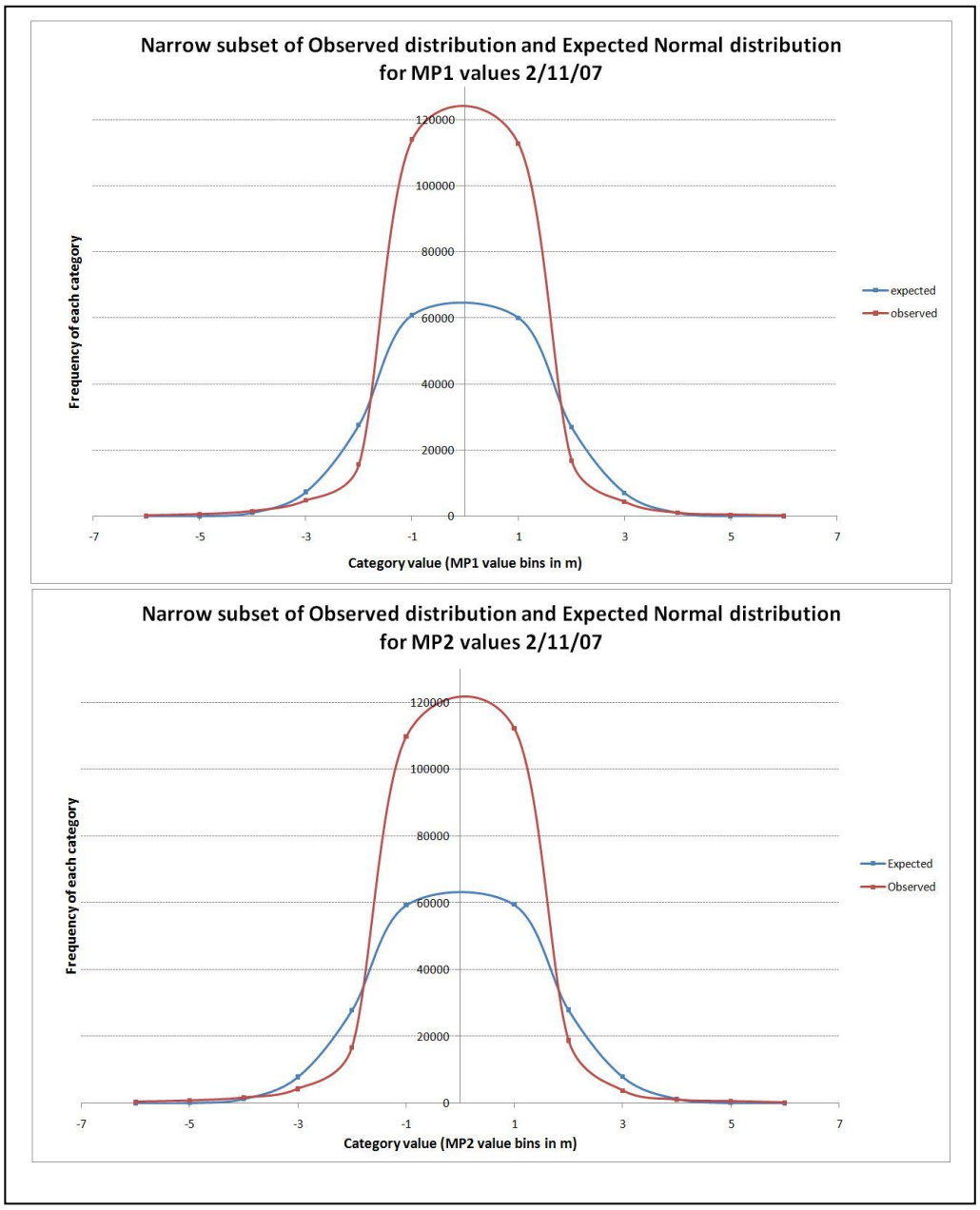


Figure 6.59 – Graphs showing narrow subset of observed distribution and expected normal distribution for MP1 and MP2 values 2/11/07

Once again the distribution was seen not to be normal with a much narrower distribution than is normally described using the datasets mean and standard deviation.

The same process used to produce figures 6.58 to 6.60 was also carried out on the data collected on the 3rd of November and the results were similar. The graphed results can be seen in figure 6.61.

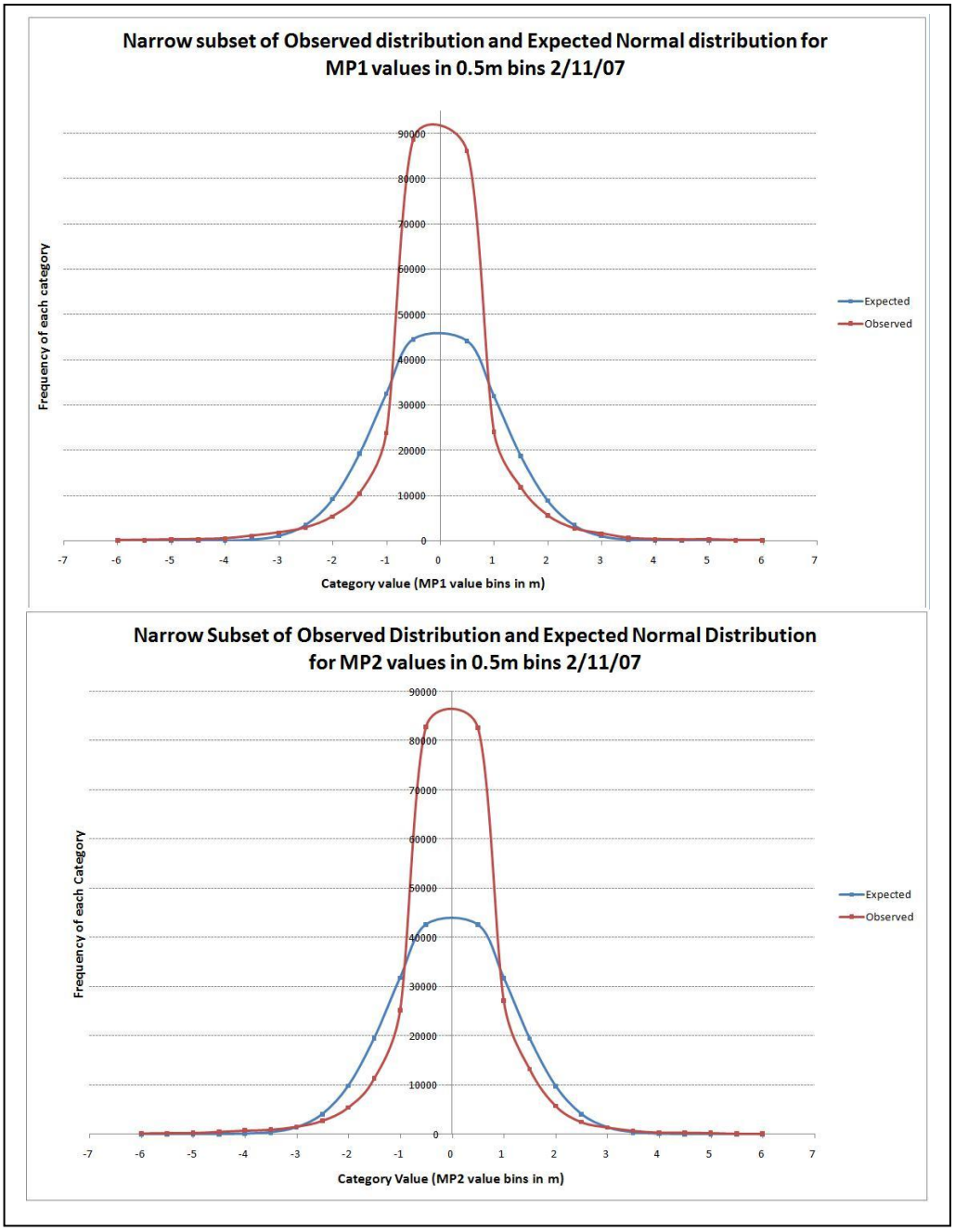


Figure 6.60 - Graphs showing narrow subset of observed distribution and expected normal distribution for MP1 and MP2 values 2/11/07 using 0.5m bins

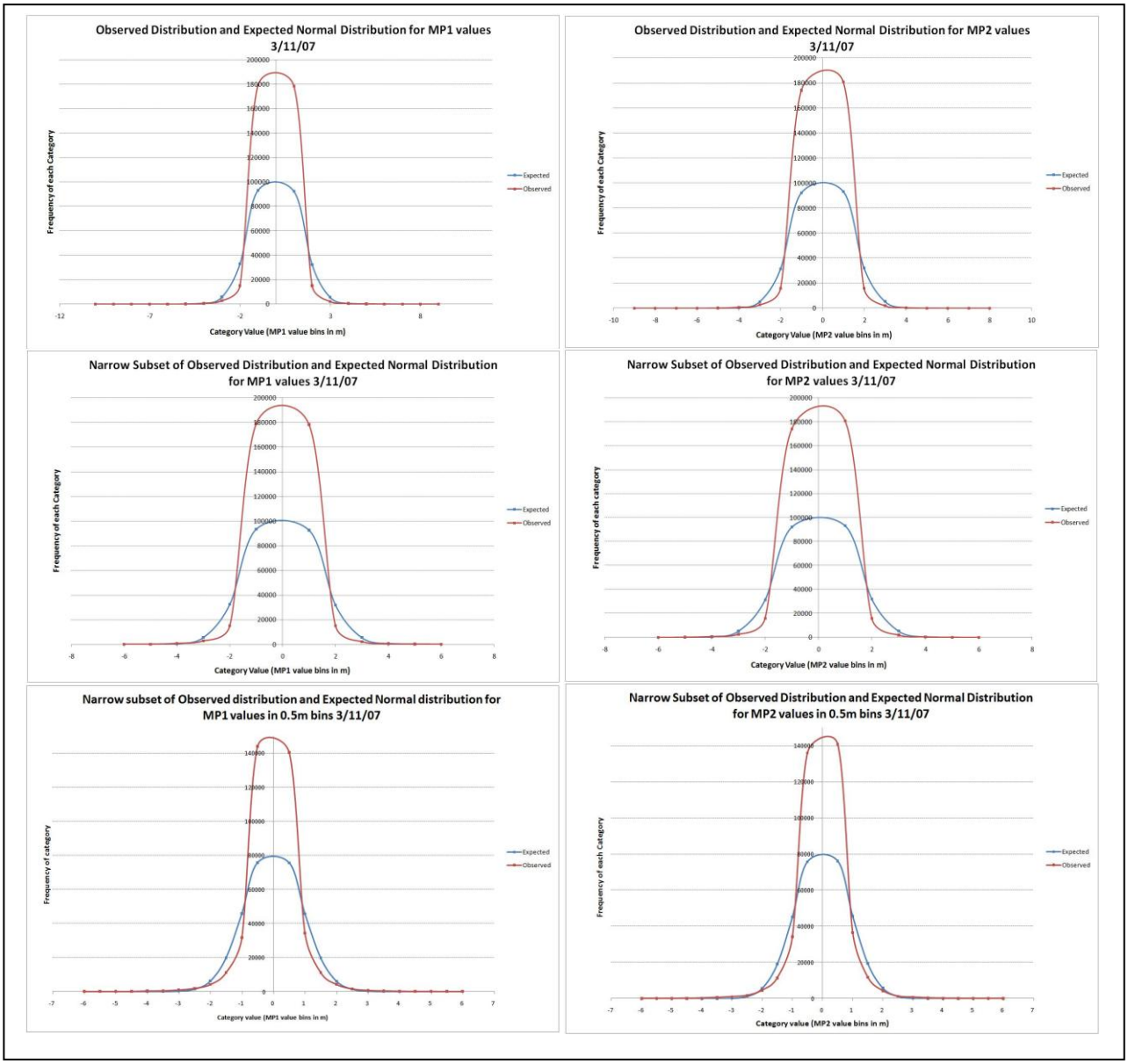


Figure 6.61 – Graphs showing observed distributions and expected normal distributions for MP1 and MP2 values for the full data set, narrow data set and 0.5m bins data set on 3/11/07

Given these results, it is concluded that statistical terms used to describe a normal distribution cannot be applied to multipath levels in the rail environment as the expected function has a larger spread of data and a lower central frequency when using the mean and σ calculated from the observed data.

This is significant as the rail community normally use σ values to describe safety critical systems and the expected hazard rates and error margins. Due to the results not adhering to a normal distribution, these statistical descriptions cannot be used with confidence and so a qualitative approach and tailored statistical terms must be used.

In order to devise a statistical term for the data recorded and analysed, the nature of the values recorded must be fully understood.

Due to the length of break-free data being a major factor in the level of M_{P1} and M_{P2} change calculated, the length of data without breaks collected was also analysed. As expected, the number of short duration break-free data sets is far greater than the longer sets.
This can be seen for both the 2nd and 3rd of November data sets in figures 6.62 and 6.63 respectively.

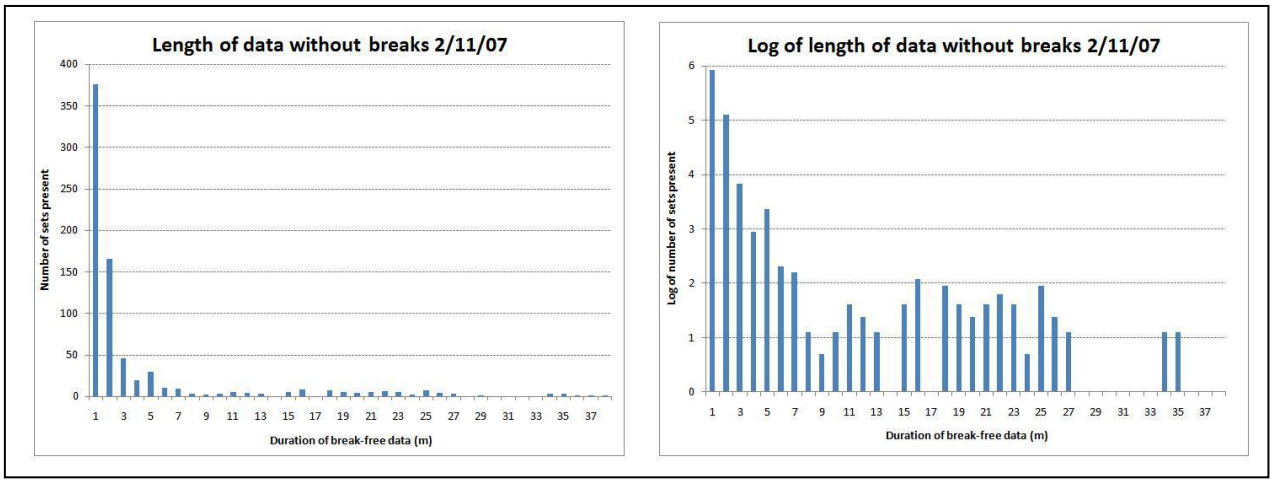


Figure 6.62 – Charts showing the number of break-free data sets for discretely grouped durations for 2/11/07 and the equivalent natural log values

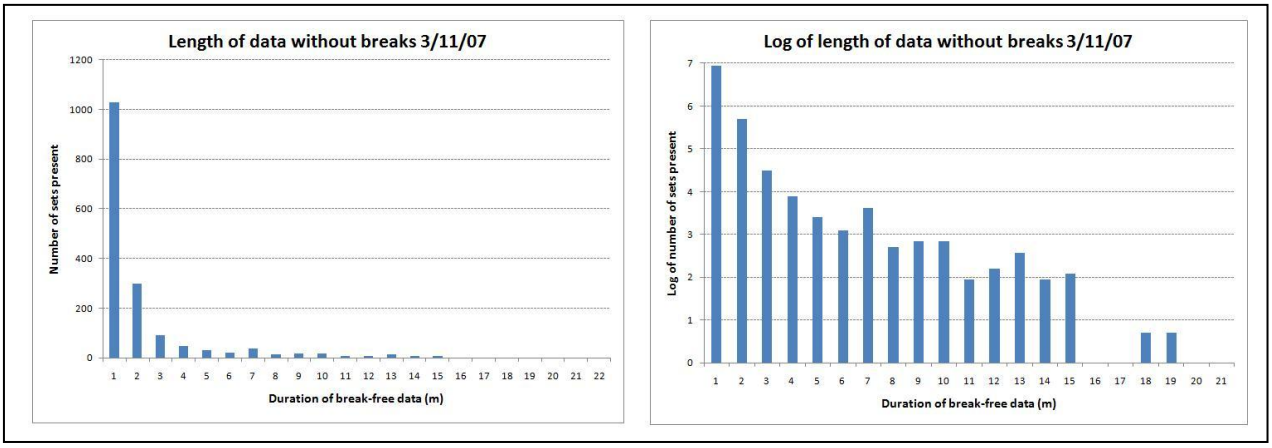


Figure 6.63 – Charts showing the number of break-free data sets for discretely grouped durations for 2/11/07 and the equivalent natural log values

As there is a large difference between the number of shorter and longer periods of data, the equivalent natural log values have also been plotted in the figures.

The values for break-free data periods of 1 minute or less have not been plotted, but for reference there were 8,905 and 12,396 for the 2nd and 3rd of November respectively, far higher than any other duration, indicating the extent to which the data is broken up.

The reasons for data breaks vary; the most common being the local infrastructure causing a momentary loss of signal. Other reasons include the loss of satellite due to low elevation or an azimuth change relative to the direction of train travel such that the satellite is no longer parallel to the line of travel.

Breaks due to receiver failure are not considered here as none were known to have occurred and are assumed to be infrequent enough not to have an impact on the results of this experiment. Receiver failure should however be a major part of the final safety case for GNSS in the railways and the various reasons for receiver failure must be addressed if system integrity is to be determined.

The average calculated change in M_{P1} and M_{P2} over the duration of each break-free section of data was also added as an output option for the RTK library so that any correlation between the duration of break-free sections of data and the corresponding maximum change in M_{P1} and M_{P2} can be analysed. This resulted in data that was of interest when trying to describe the expected value of M_{P1} and M_{P2} for a given length of break free data, but the main interest when discussing safety critical systems is the maximum possible (worst case) values that can appear. Because of this, the maximum recorded M_{P1} and M_{P2} value was also added as an output, the results of which for the 2nd and 3rd of November can be seen in figures 6.64 and 6.69.

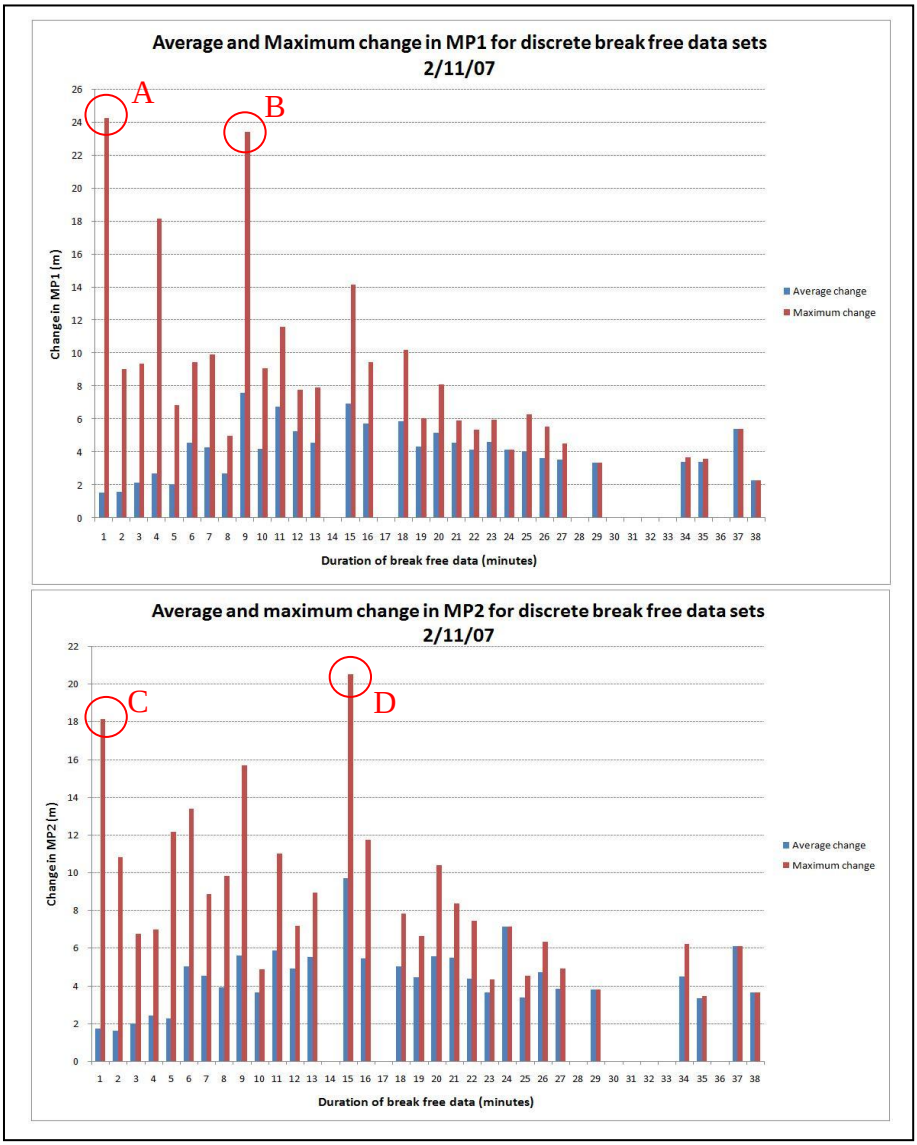


Figure 6.64 – Charts showing the average and maximum change in MP1and MP2 for discrete durations of break-free data for 2/11/07

Given the extreme values of M_{P1} and M_{P2} seen in figure 6.64, the two largest values for each graph (labelled A-D) were individually investigated.

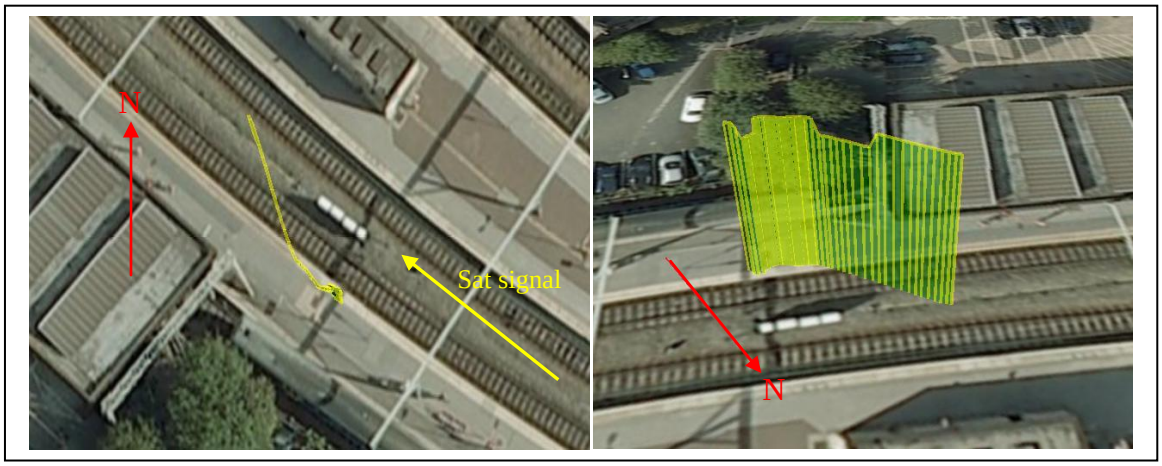


Figure 6.65 – Overhead and side view of position error due to signals seen at Wolverhampton station 2/11/07 (mark A)

From figure 6.65, an almost linear position error is visible, possibly due to the station housing or platform. Satellite 18 has an azimuth of 129º and is shown in the picture, indicating that the satellite should be visible directly, but reflection due to the station house is still possible. The M_{P2} value on satellite 18 at this time was 3.2128m.

Mark B was due to satellite 19 when the train was stationary in New Street station between 15:28:56 and 15:38 43 reaching am M_{P1} value of 23.423m. The position errors associated with this period of time can be seen in figure 6.66.

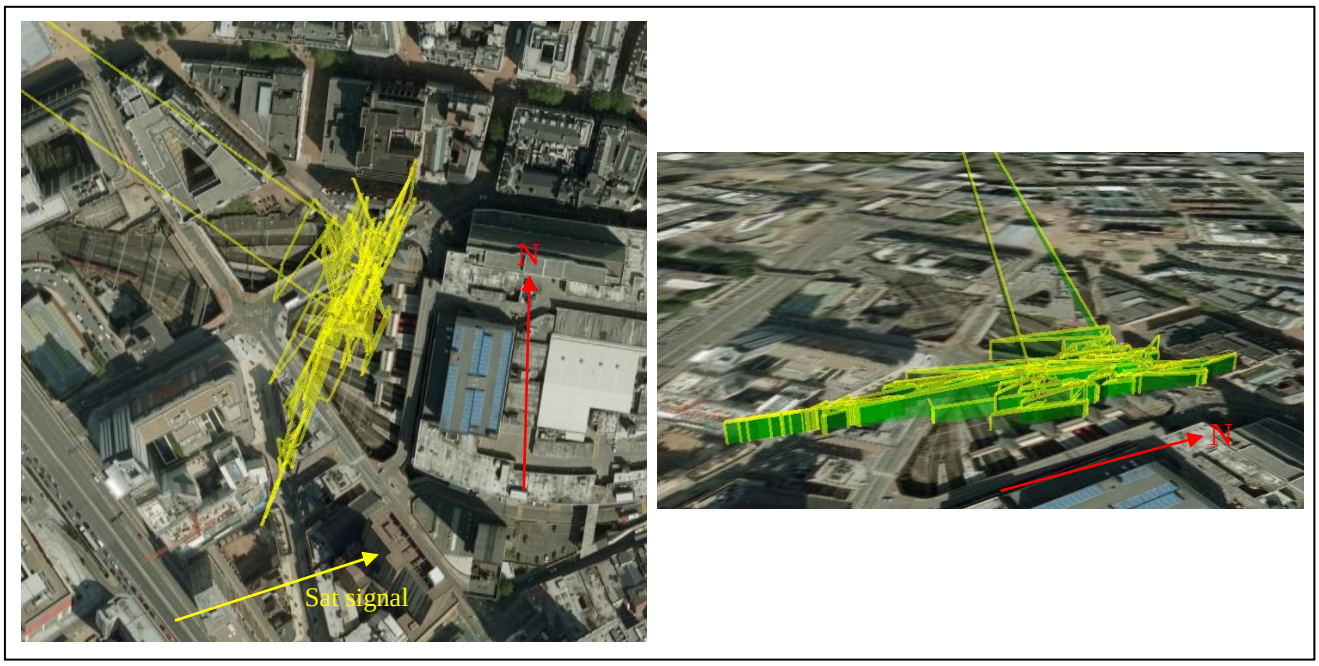


Figure 6.66 – Overhead and side view of position error due to signals seen at New St station 2/11/07 (mark B)

Given the large amount of surrounding infrastructure, as seen before, the level of multipath seen in Birmingham New Street station is very high, with signal obscuration and reflection being the more likely causes of the extreme position errors seen in 6.66 as several are above the error likely to be caused by an M_{P1} value of 23m. The corresponding M_{P2} value for this satellite during these epochs was also too low at 15.679m which again is lower than M_{P1} for the same satellite, negating the assumption that M_{P2} is normally recorded as being higher than M_{P1}.

The position jumps produced during point C are shown in figure 6.67 and relate to the epochs between 16:55:20 and 16:57:10 for satellite 25.

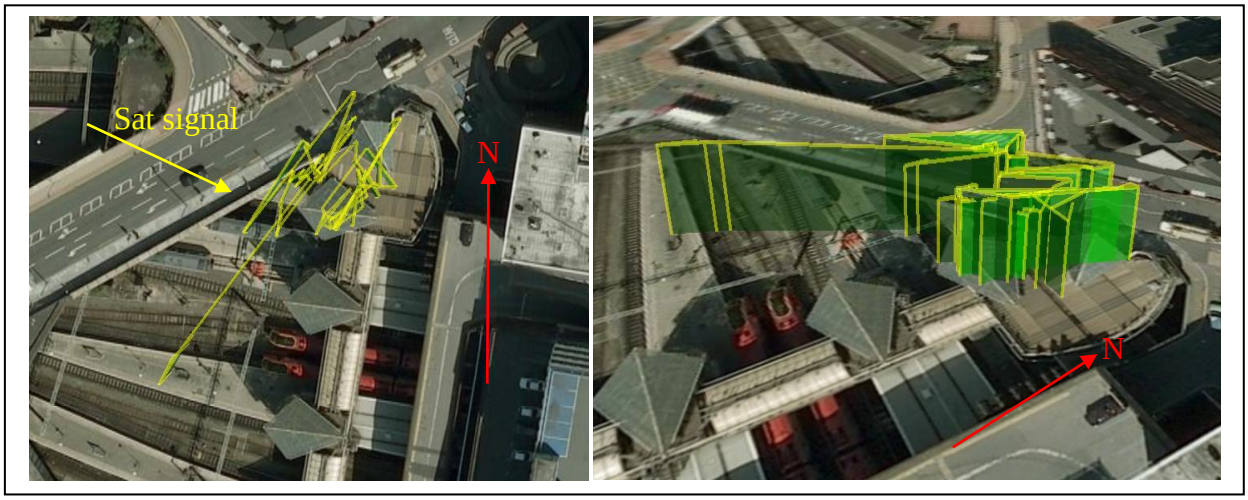


Figure 6.67 - Overhead and side view of position error due to signals seen at New St station 2/11/07 (mark C)

The train is again located in New Street station, but the duration of data is shorter and the position errors are smaller. The direction of satellite 25 relative to the receiver is directly in line with the overhead roadway, which may be the major source of multipath, especially as the satellite is at a low elevation of just 15º.

Mark D relates to satellite 3 from 13:52:45 to 14:08:43 where an M_{P2} value of 20.515m was observed, again in New Street station. The corresponding M_{P1} value was 14.185m which can be seen on the M_{P1} chart in figure 6.64.

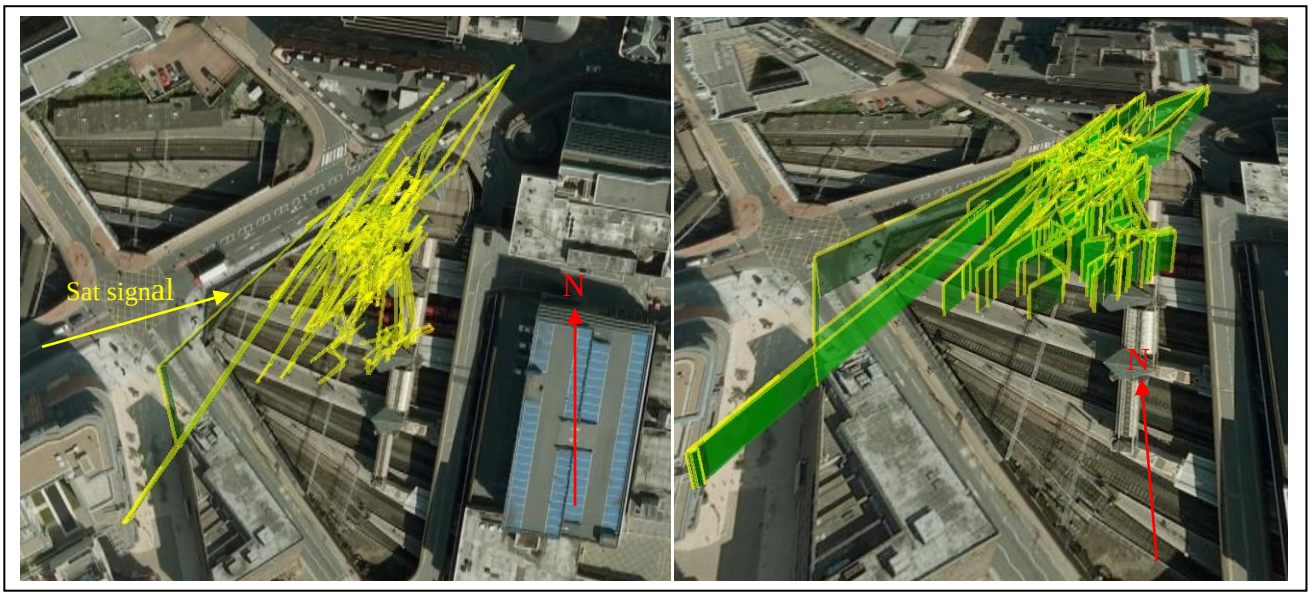


Figure 6.68 - Overhead and side view of position error due to signals seen at New St station 2/11/07 (mark D)

The signal from satellite 3 comes in at a very low elevation of 9º over the overhead roadway and so the level of interference and multipath from the surrounding infrastructure is significant. The position errors are again very large and show that areas of main concern when considering the integrity of a position solution are in highly built up areas, especially where there are overhead structures and high rise buildings in the near vicinity.

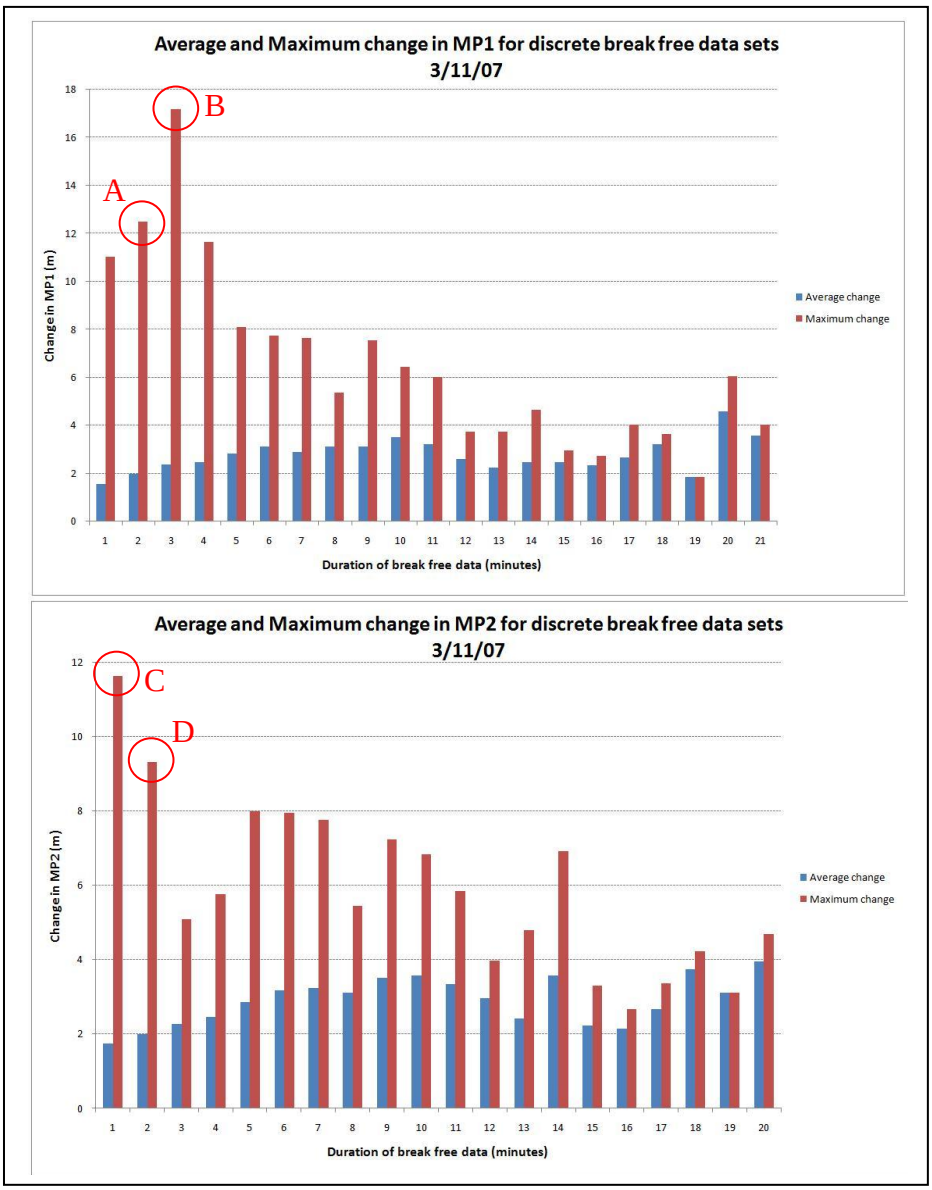


Figure 6.69 – Charts showing the average and maximum change in MP1and MP2 for discrete durations of break-free data for 3/11/07

The two extreme cases labelled A and B in 6.69 are both related to the same epochs. Point A relates to satellite 5 and B relates to satellite 30. The longer time period is

from 8:02:40 until 8:06:31 and can be seen in figure 6.70. The azimuth angles of the two satellites are 227º and 228 º respectively with elevations of 29 º and 12 º.

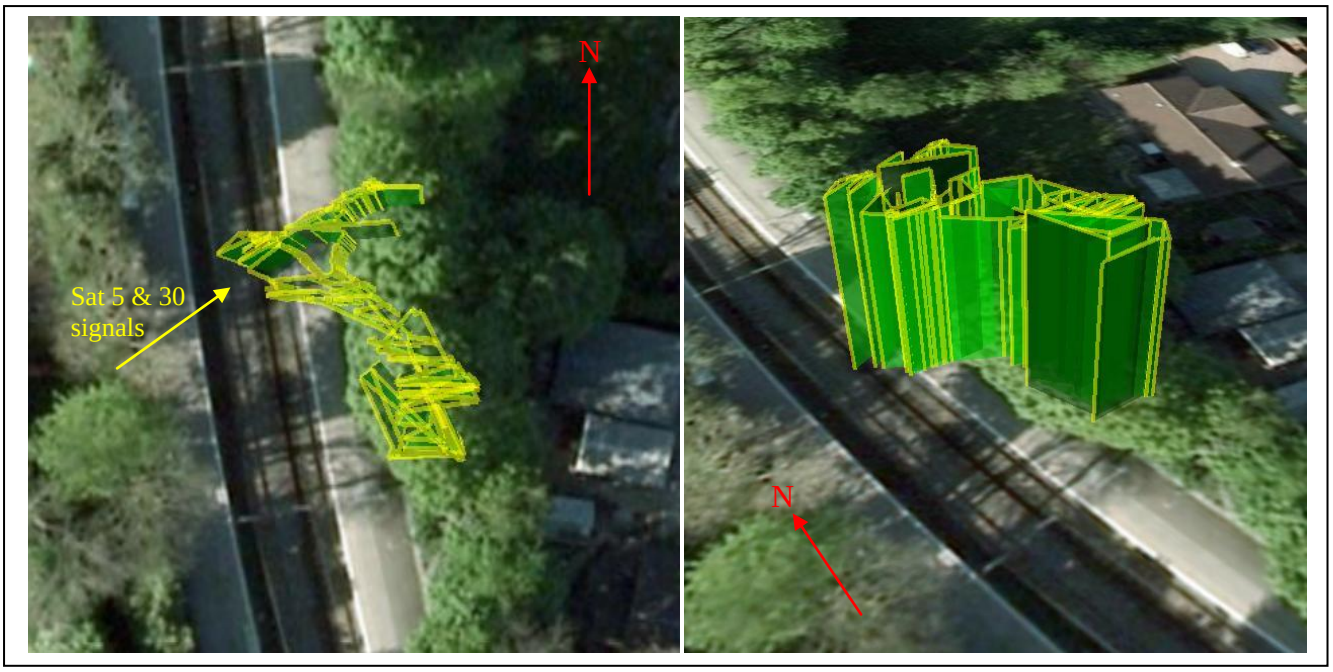


Figure 6.70 - Overhead and side view of position error due to signals seen at Barnt Green station 3/11/07 (marks A and B)

The position errors seen are when the train was stationary in the station surrounded by trees and foliage. The equivalent M_{P2} error for satellite 5 was 3.13m and 2.39m for satellite 30, both considerably less than the corresponding M_{P1} values of 12.49m and 17.16m.

This may be due to the weaker L2 signal having its power reduced even further by the surrounding foliage not having enough power to reflect on the surroundings and still register with the receiver whilst the stronger L1 signal has enough power to form a combined signal at the receiver.

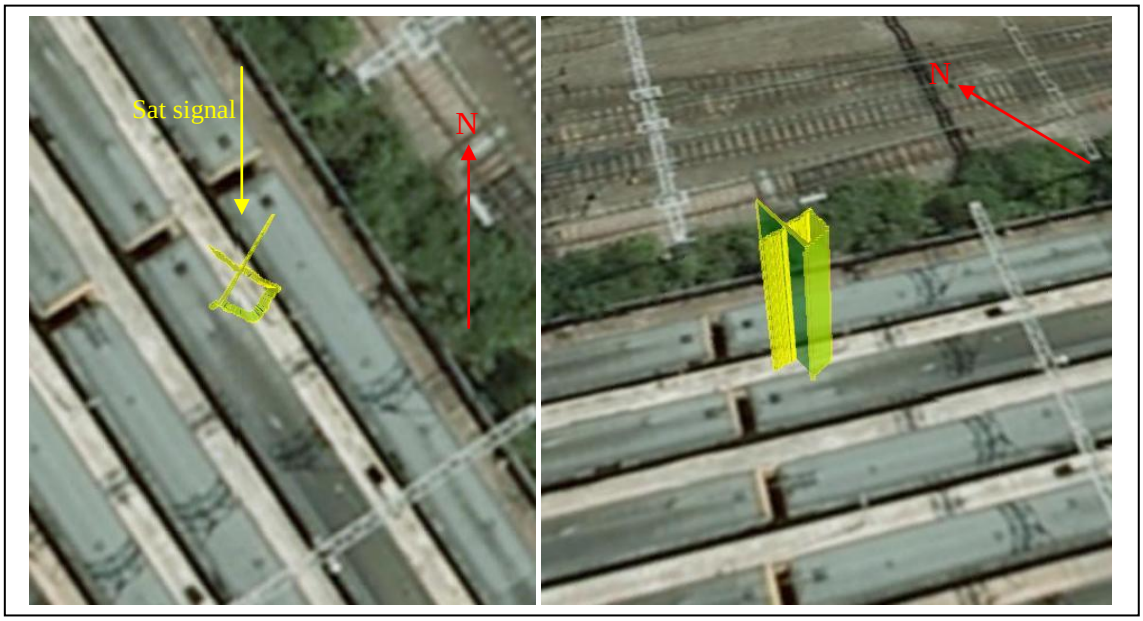


Figure 6.71 - Overhead and side view of position error due to signals seen at Winson Green sidings 3/11/07 (mark C)

Figure 6.71 shows the position error when the train was stationary in Winson Green sidings (mark C). It is not known if there was another train along side during the data collection period and so only the overhead infrastructure and nearby bushes are known to be possible causes of multipath (as well as the trains own roof). The M_{P2} value for satellite 5 changed by 11.61m with corresponding M_{P1} change of 1.48m.

The satellite was positioned very low on the horizon and so the level of multipath, despite there being few surrounding high objects, is likely to be big. The change in multipath may be due to the satellite geometry changing or due to a train moving nearby; it is unknown without a detailed geometric evaluation of the situation which is beyond the scope of this thesis.

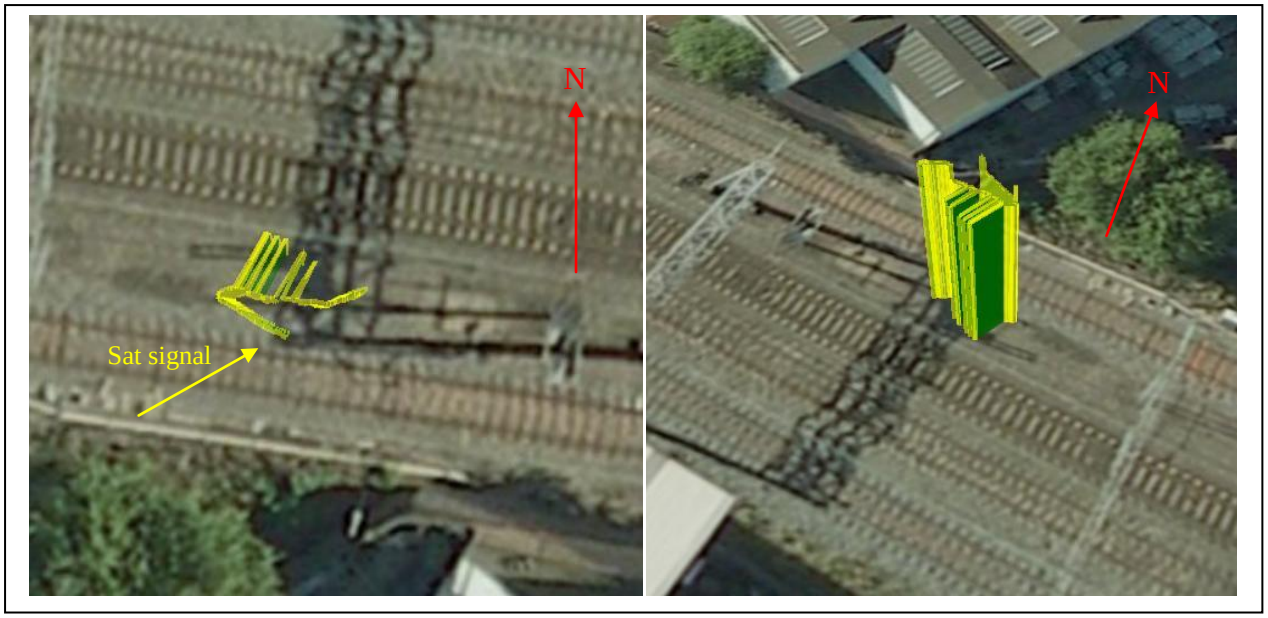


Figure 6.72 - Overhead and side view of position error due to signals seen at Smethwick 3/11/07 (mark D)

Figure 6.72 shows the epochs for mark D where M_{P2} on satellite 9 reached 9.31m between 5:12:30 and 5:14:44. The M_{P1} value changed by 4.36m over the same timescale with most likely source of multipath being the two large warehouses located on either side of the tracks. Overhead infrastructure is also visible in the area and so this could also be a contributing factor.
The satellites low elevation of 11.8º would have also caused a higher level of multipath from the surroundings with the loss of satellite 9 most likely due to its decreasing elevation causing obscuration by one of the industrial buildings.

The results shown in figures 6.64 and 6.69 follow two patterns for the average results.
The initial expectation is that the longer a data set is without breaks, the greater the observed change in M_{P1} will be, purely as there is the possibility for a greater data spread to be normalised. This is shown in the figures for both days of data as the longer streams of break-free data have an average M_{P1} variation of around 6 to 10 meters.

The other noticeable result for the average values is that some of the shorter sections of break-free data can also contain larger variations in M_{P1}.
This is best explained by considering that the most hostile signal environment will usually cause the highest level of multipath. This hostile environment will also be

most likely to cause signal breaks due to obscuration of satellites, so despite the comparatively short duration of these data sets, the changes seen in M_{P1} can in fact be much larger than for the longer break-free data sets, as seen in 6.64 for the 2nd of November, though this is mainly due to there only being two 9 minute break free data sections, one of which providing 23.4m the other 18.2m when the train was at New Street station between 15:17:25 and 15:38:43, showing the high levels of multipath experienced when inside such a heavily built environment. The average value for M_{P1} for the 2nd was 4.18m for the averaged values and 8.40m for the maximum values. For M_{P2} it was 4.43m and 8.59m for average and maximum averages respectively. This result agrees with the earlier conclusion that the weaker L2 signal is more prone to multipath than L1.

The results for the 3rd do not have the same level of difference for the shorter periods recorded, but as with the first day of data, there is no clear linear relationship between the duration of break-free data and the level of change in M_{P1}. The averaged values for M_{P1} are 2.75m and 6.66m for the average and maximum changes and 2.90m and 6.11m for M_{P2} which again shows the average being larger for M_{P2}, but the average of the maximum values if smaller which is more due to the random nature of the possible maximum values than the general trend of the data.

The values for maximum M_{P1} and M_{P2} change follow a slightly more random pattern (as expected), but with a much higher level of multipath for shorter data sets. This fits with the theory that the high multipath areas (such as New Street station) will also cause more breaks in the data due to obstruction of the signals.
The results also show the difference between an averaged multipath value and the maximum for the same set which seems far less for the longer data sets, but this is in fact mainly due to there being fewer long data sets and hence the average is sometimes for only 2 results. The differences seen for the shorter data sets clearly show how the average is not indicative of the possible changes in high multipath environments, which is a fundamental characteristic of the seemingly random, white noise nature of multipath.

Given the results show general characteristics of the collected data and that any conclusions should be based on the widest available set of data, the processing of results was extended to include all the other data from the set provided from 2/11/07 to 15/11/07.
The data was again split into the days when the train was either travelling on the Redditch/Lichfield line or the Wolverhampton line. The results were collated and analysed so as to give a more 'statistically sound' evaluation of the multipath values expected in the rail environment.

The first collated data set will be for the Birmingham – Wolverhampton line and will be made up of data collected on the 2nd, 6th and 10th of November 2007.

The second collated data set will be for the Redditch – Lichfield line and will be made up of data collected on the 3rd, 5th, 8th, 9th, 13th, 14th and 15th of November 2007.

There is an obvious bias in the amount of data for the second data set, but this is unavoidable as the train timetable does not provide an even split of train time on both tracks and unfortunately further data was not available at the time of writing.

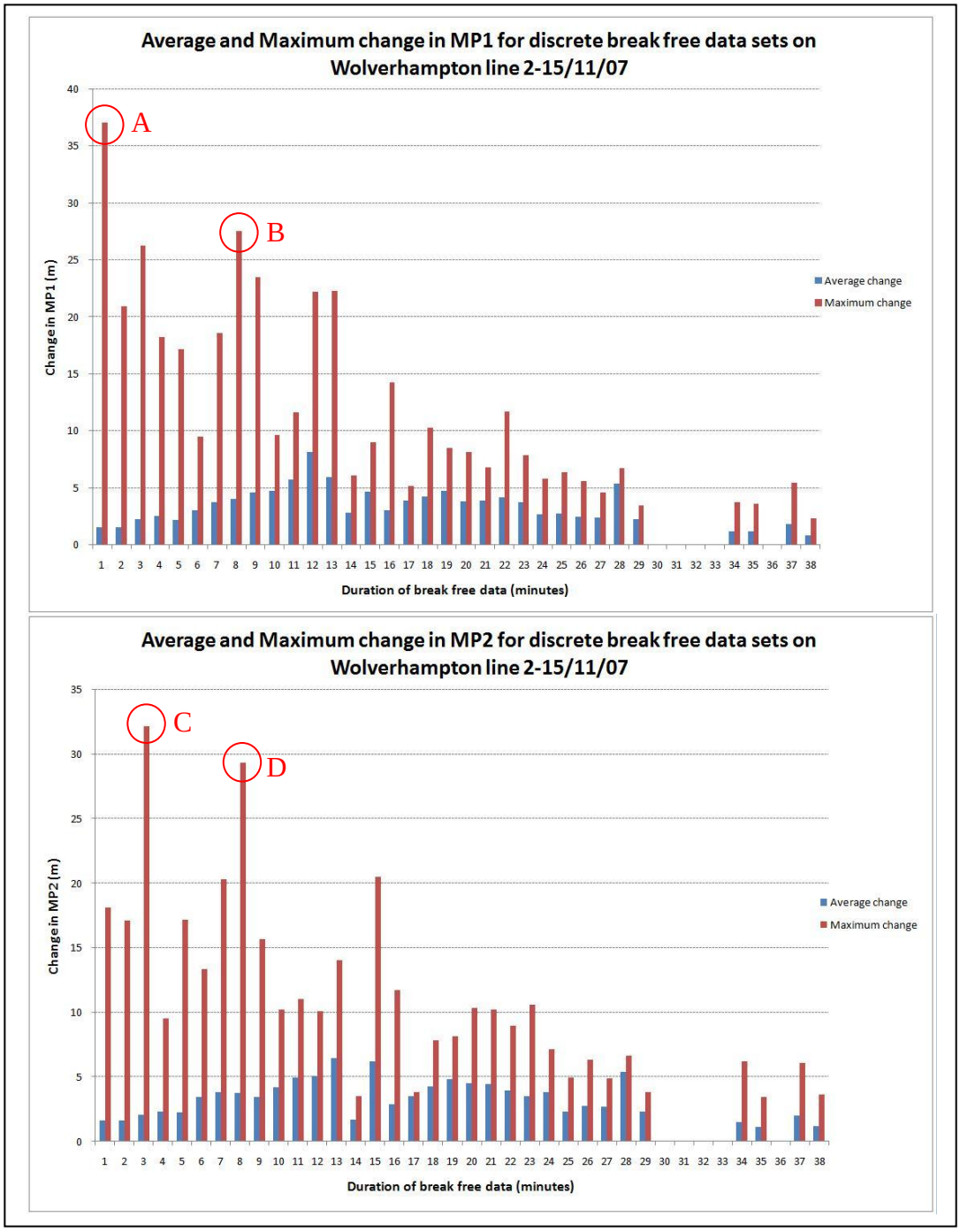


Figure 6.73 – Graphs showing the average and maximum change in MP1 and MP2 for discrete break free data sets for all available Wolverhampton line data 2-15/11/07

Mark A references a very high level of multipath on satellite 19 between 15:21:04 and 15:22:25 on the 6th of November that should be filtered out by modern receivers as the second (reflected) signal received would have a very large delay that a receivers' internal correlator should reject.

Given the azimuth of the satellite (shown in figure 6.74, ~270º) and the elevation of 22º, high levels of multipath are expected, especially as the train is once again located in New Street station. The M_{P2} value for satellite 19 during the same epochs changes by 6.28m, again showing that the levels of multipath are high.

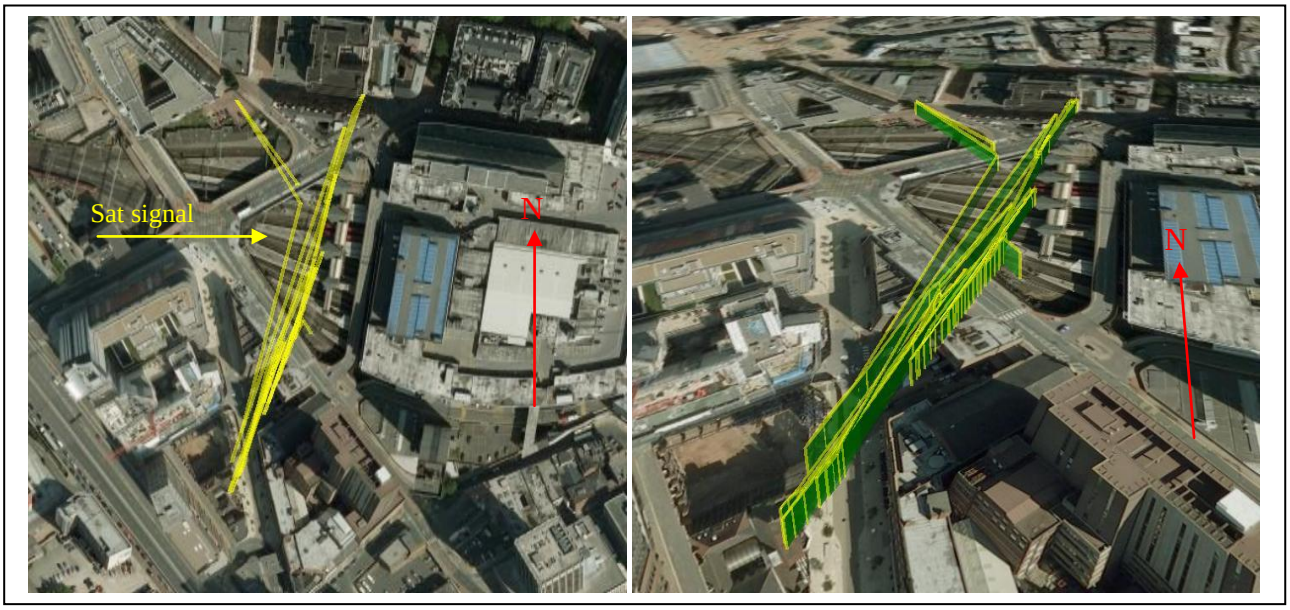


Figure 6.74 – Overhead and Side view of position errors due to signal errors in New Street Station 6/11/07 (Mark A)

Figures 6.76 and 6.77 represent marks C and D and are both also from satellite 19 on the 6th but from 15:23:36 to 15:27:04 and 15:27:25 to 15:36:19 respectively. The azimuth of the satellite only changes by 1º with the elevation decreasing to 20º.

The M_{P2} changes for satellite 19 are 32.20m and 29.34m with the corresponding M_{P1} changes being 19.24m and 25.89m for C and D. These values are very high and suggest that given the satellites position at that moment, there is a very strong source of multipath or there is an error on the satellite signal.

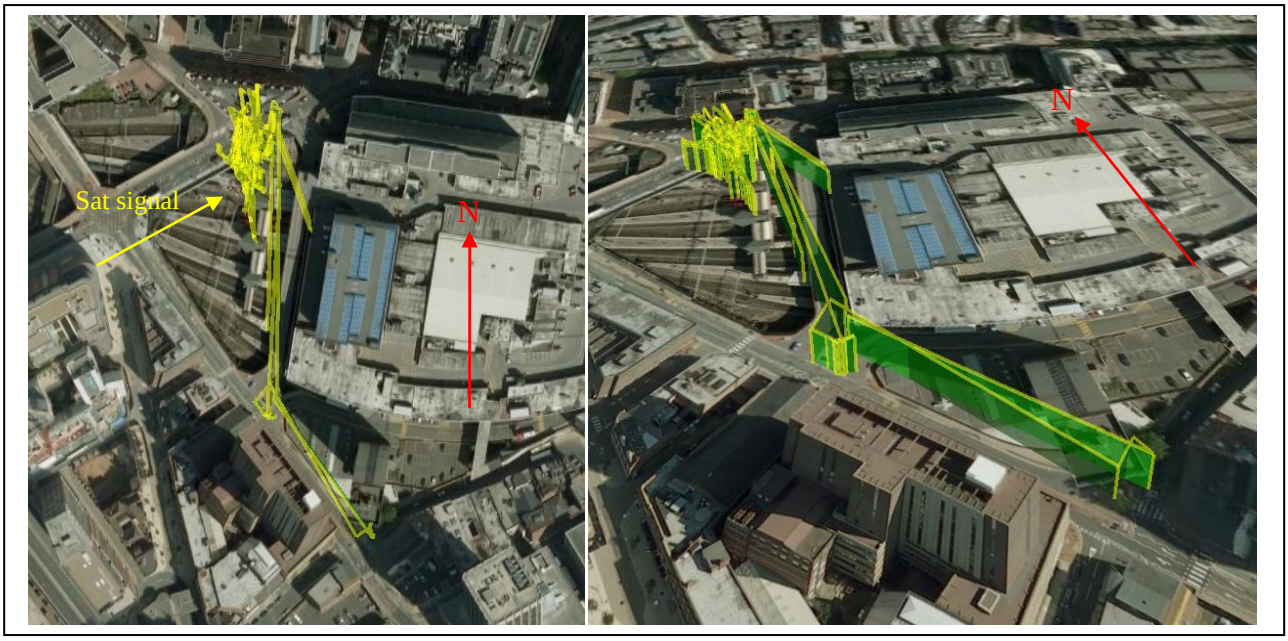


Figure 6.75 - Overhead and Side view of position errors due to signal errors in New Street Station 10/11/07 (Mark B)

Figure 6.75 relates to satellite 1 on the 10th of November between 11:28:04 and 11:36:22. The M_{P1} value changed by 27.49m and M_{P2} by 6.32m with very large position errors being visible, again in the New Street station area.

Due to the size of the main station building, satellites from the eastern section of sky are obscured and so this may be a major contributing factor to the greatly decreased positional accuracy observed whilst the train is in the station.

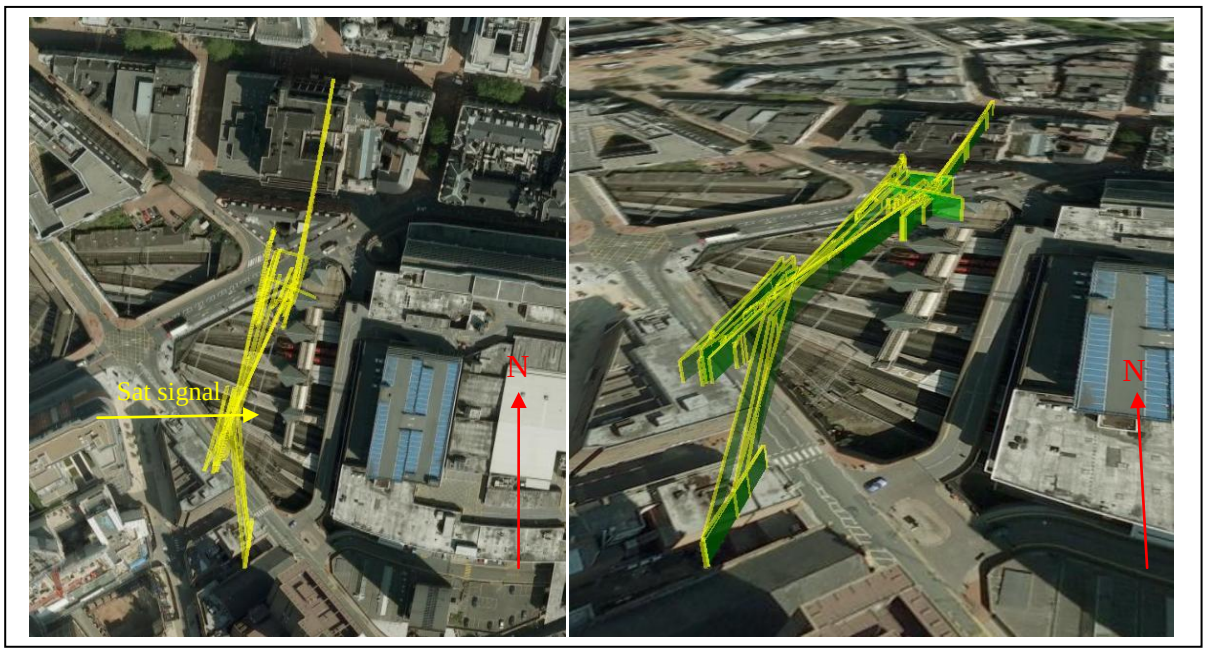


Figure 6.76 - Overhead and Side view of position errors due to signal errors in New Street Station 6/11/07 (Mark C)

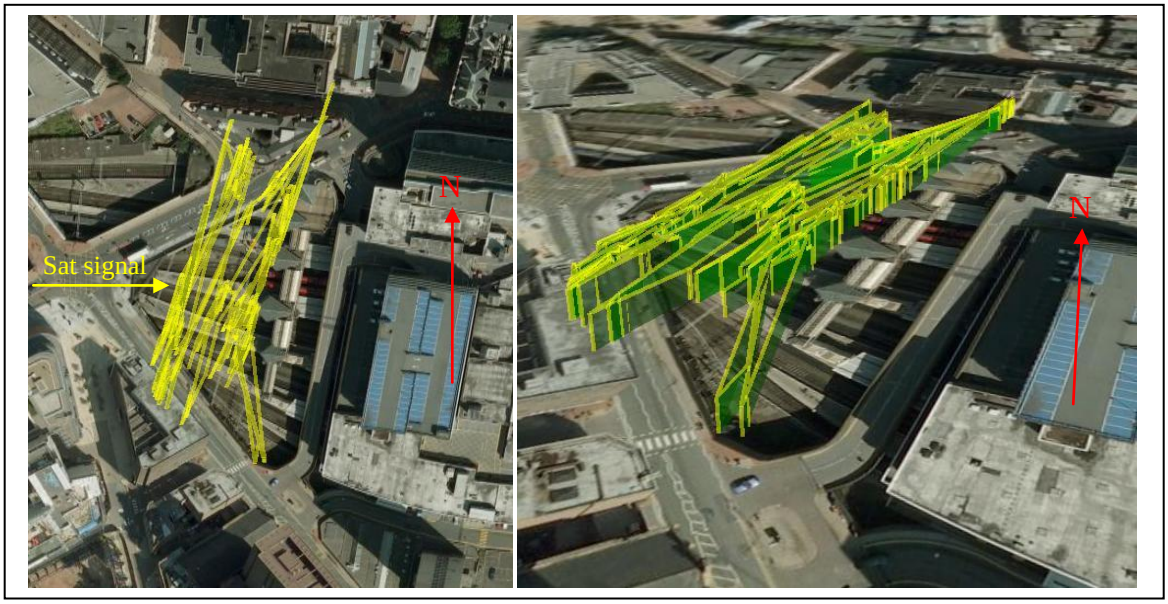


Figure 6.77 - Overhead and Side view of position errors due to signal errors in New Street Station 6/11/07 (Mark D)

The graphs in figure 6.78 follow a similar pattern to the M_{P1} and M_{P2} graphs in figure 6.73 with the average value tending to be below 6m with a rough peak in the 12 to 15 minute length sections of data.
There is also a lack of data sets that are 30-33 minutes long; the reason for this is unknown, but when looking at figure 6.44 breaks are seen in the position around every 30 minutes, this is likely due to overhead infrastructure along the line and when entering New Street station, which would match the return journey time being roughly 30 minutes.

The maximum values are as expected, larger in most cases as there is a larger volume of data that is more likely to have the possibility of extreme values, despite this, the maximum errors do not exceed 38m on L1 and 33m on L2 with them rarely exceeding 20 to 25 meters, which when translated to a position error, in the worst case scenario would be unlikely to produce and along-track error above 38m which is assumed to be within the operating limits of the railway if compared to the traditional track circuit location systems mentioned in section 2 of this thesis.

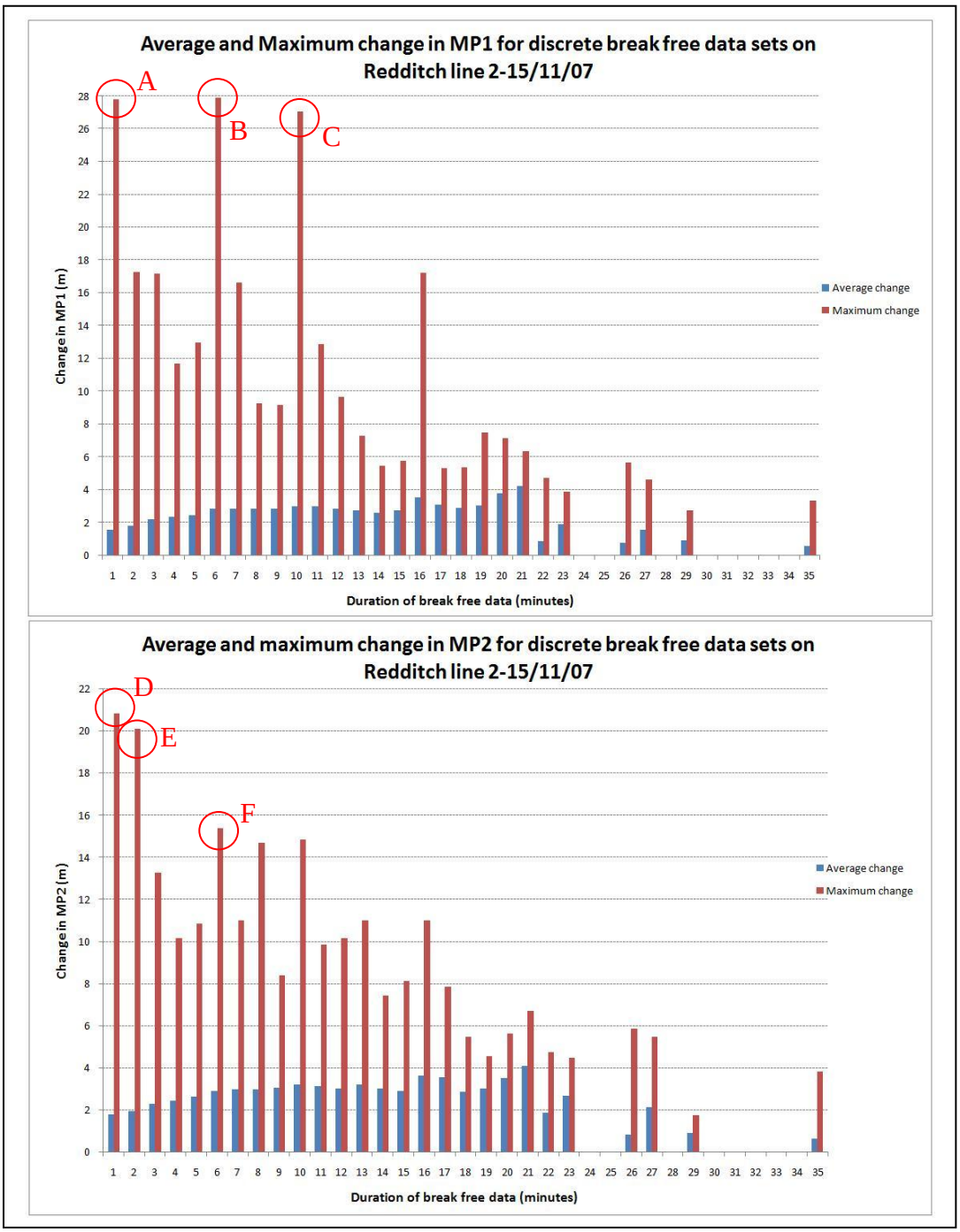


Figure 6.78 - Graphs showing the average and maximum change in MP1 and MP2 for discrete break free data sets for all available Redditch line data 2-15/11/07

Marks A, B and C were investigated due to their high values compared with the other data lengths collected. The resulting position errors and incoming signal azimuth angles can be seen in figures 6.79 to 6.81.

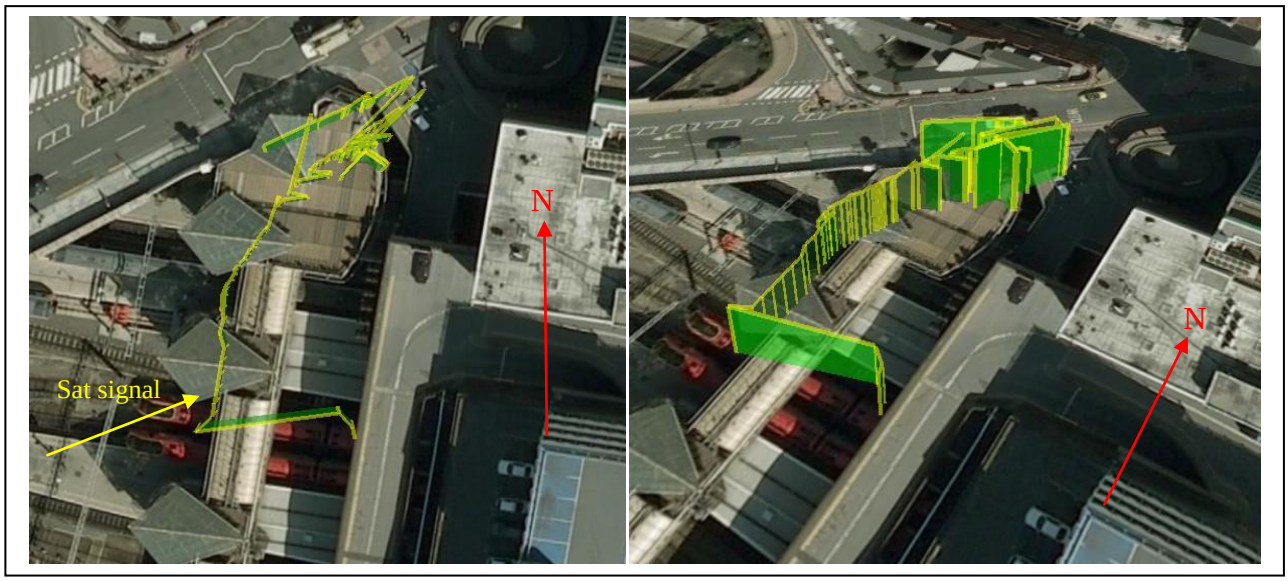


Figure 6.79 - Overhead and Side view of position errors due to signal errors in New Street Station 5/11/07 (Mark A)

Mark A was produced by satellite 17 between 22:32:28 and 22:34:22 on the 5th of November with an incoming azimuth of 237º and an elevation of 39º. The resulting M_{P1} change was recorded as being 27.80m with the corresponding M_{P2} value being 2.23m. The position errors seen in 6.84 show a steady walk from the southernmost point, closes to the true position towards the north of the station.

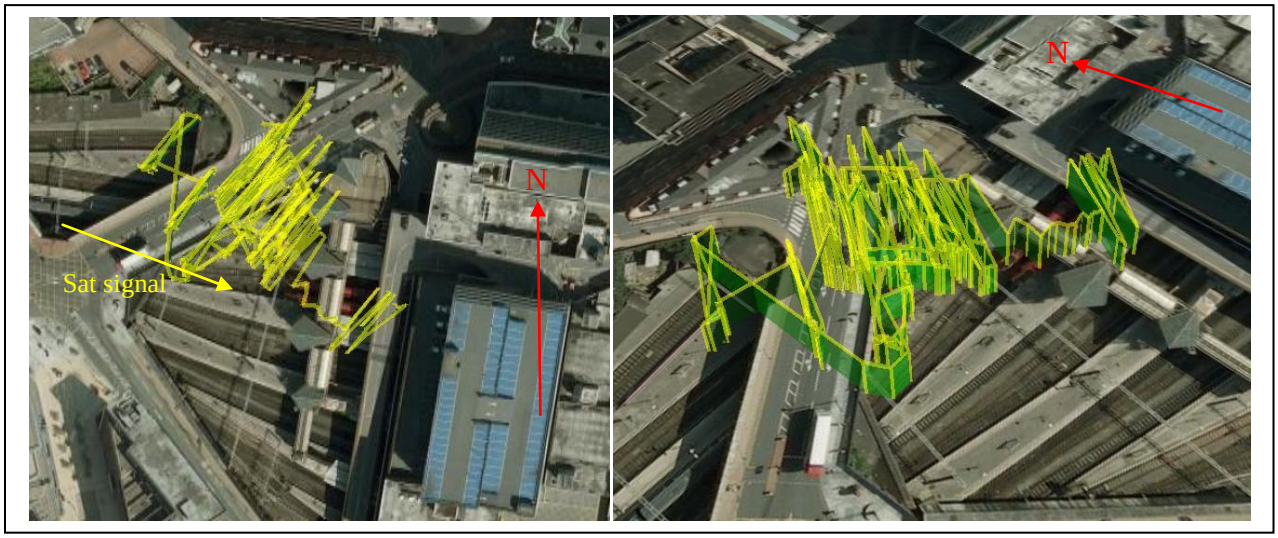


Figure 6.80 - Overhead and Side view of position errors due to signal errors in New Street Station 13/11/07 (Marks B & F)

Mark B also coincides with mark F as the same satellite (no. 2) during the same epochs (23:47:37 to 23:54:29) had both an M_{P1} change of 27.92m and an M_{P2} change of 15.36m. The satellite was at an azimuth of 295º with elevation 44º causing the signal to enter the station over the overhead road structure.

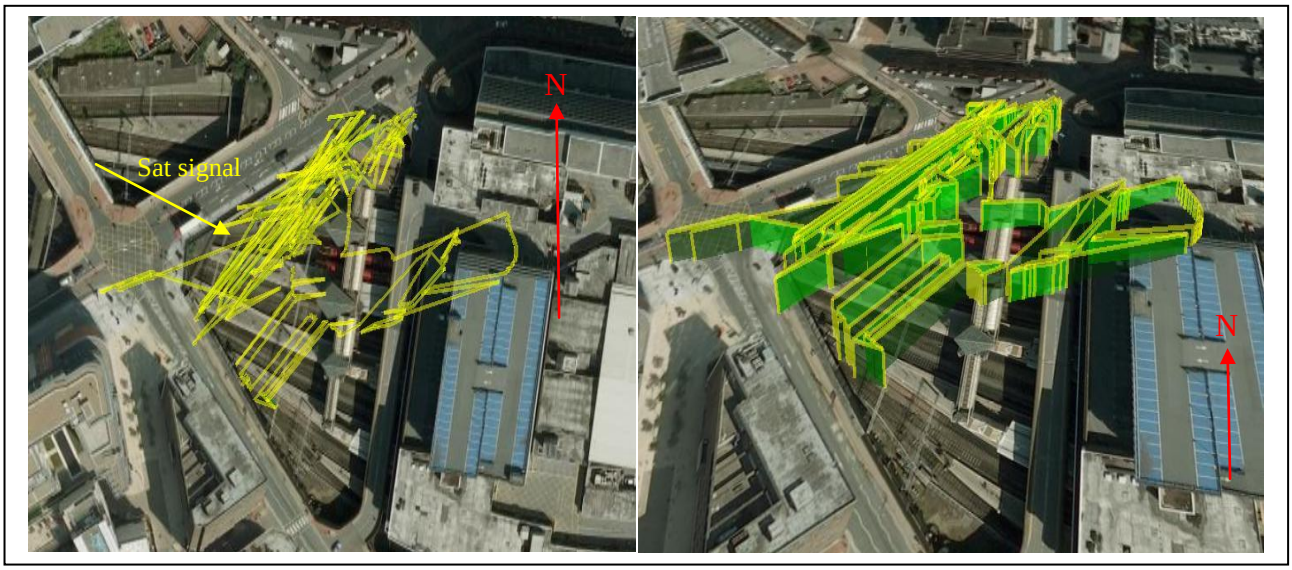


Figure 6.81 - Overhead and Side view of position errors due to signal errors in New Street Station 5/11/07 (Mark C)

Figure 8.81 refers to mark C, again when the train is located in New Street station, causing an M_{P1} change of 27.04m and a corresponding M_{P2} change of 14.82m on satellite 4 with azimuth 298º and elevation 36º.

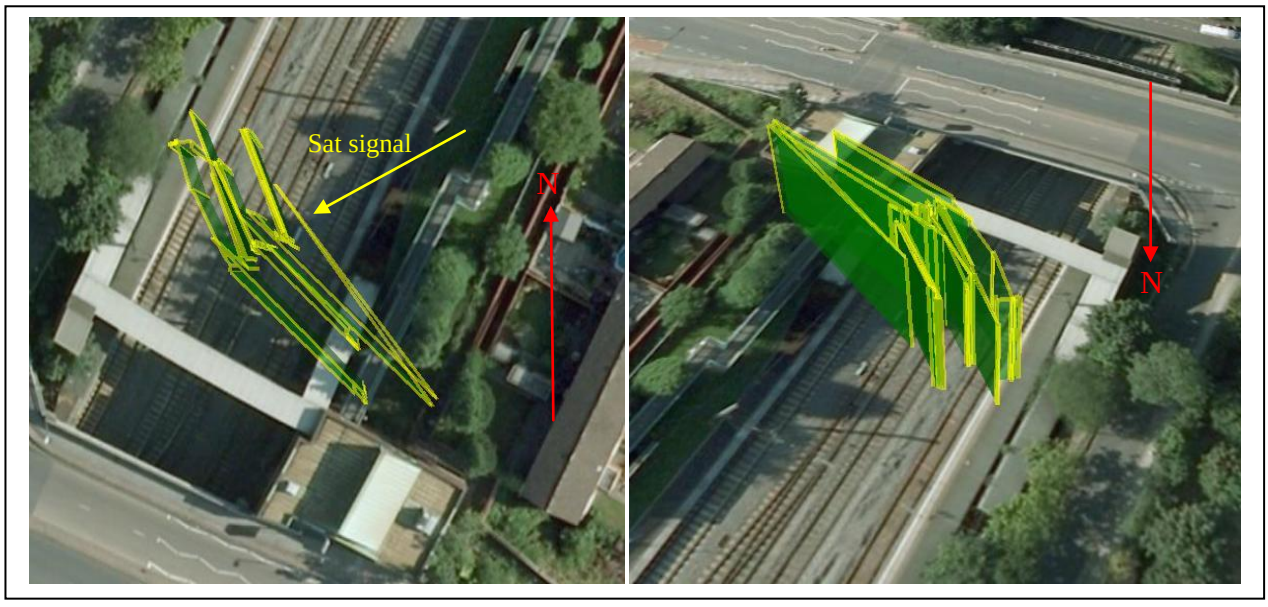


Figure 6.82 - Overhead and Side view of position errors due to signal errors in Longbridge station 14/11/07 (Mark D)

Figures 6.85 and 6.86 show position errors at Longbridge station caused by satellites 10 and 4 respectively. Mark D was cause by satellite 10 on the 14th of November when at 58º azimuth and 11º elevation. This low elevation is likely to be the major contributing factor to the high level of multipath observed between 12:25:56 and 12:27:27 giving rise to an M_{P2} change of 20.81m and M_{P1} change of 9.00m.

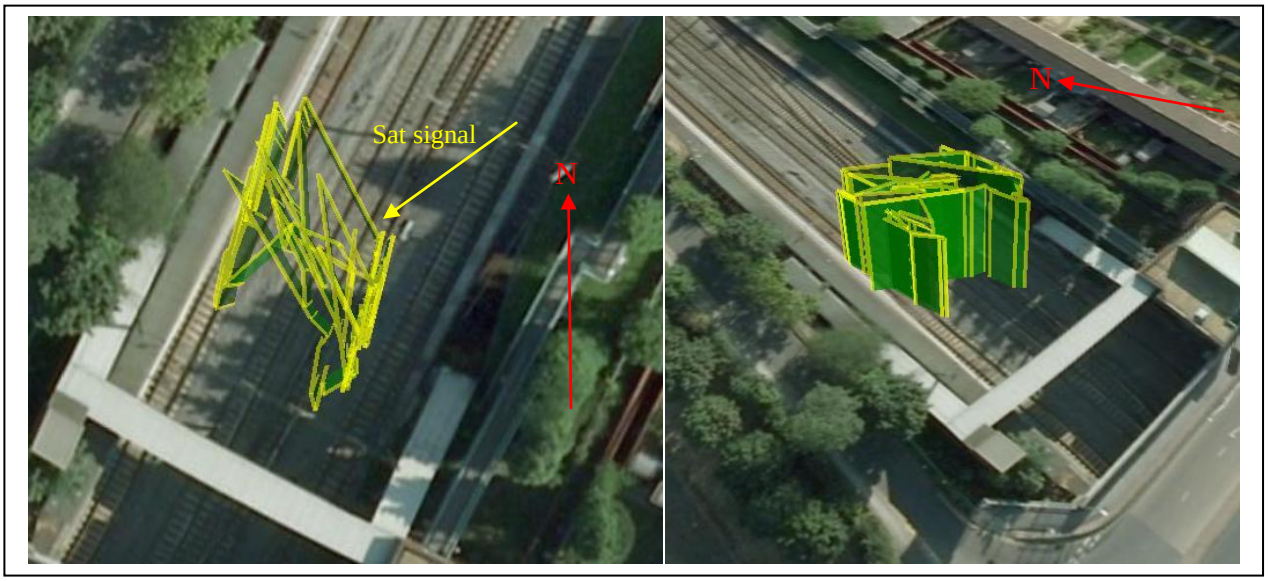


Figure 6.83 - Overhead and Side view of position errors due to signal errors in Longbridge station 5/11/07 (Mark E)

Mark E was caused by satellite 4 on the 5th of November between 10:00:49 and 10:03:18 when at an azimuth of 55º and elevation of 17º. The M_{P2} change observed was 20.09m and the M_{P1} change was 8.54m.

The overhead infrastructure surrounding Longbridge station combined with the train being stationary for an extended period of time allow for high levels of multipath change to be observed. Given the low elevations of satellites 10 and 4, the possibility of higher multipath is further increased.

The trends for the average change in M_{P1} and M_{P2} for Redditch are the same as for the data from the 3rd of November with peaks at around 10 and 20 minute lengths of data. There is more data available for the Redditch line and so longer break free data sets have increased the spread of results, again there are no data sets between 30 and 34 minutes in length for M_{P1} M_{P2}.

The Averaged values for M_{P1} are 2.46m for the average values and 11.22m for the maximum values, for M_{P2} the averages are 2.44m and 9.00m which is the opposite to the expected result and what has been shown previously.

The reason the average maximum value for the two frequencies exhibit this non typical result due to the maximum multipath values being random and so a ‘typical’ results is not really known.

The results for the averaged mean would normally be greater for L2 than L1 as seen previously but this is not the case.
This may be due to the fact that the Redditch line is more rural compared to Wolverhampton and so the L2 signal could be blocked by trees (which allow the L1 signal to pass through), causing no L2 measurements but possibly high multipath L1 signals but this would produce a break in both M_{P1} and M_{P2} data due to the linear combinations shown in equations 2.50 and 2.51 requiring both frequencies to be present for the calculation. Because of this the reason for greater M_{P1} values than M_{P2} would most likely be due to the reflecting material interacting in a different way for both of the signals.

The Wolverhampton line is considerably more urbanised and so signals from both L1 and L2 will mainly effected by building reflections and L2 would normally receive a greater multipath error due to the signal strength and correlation function discrepancies in the receiver.

In order to identify the specific areas of high multipath, the values of M_{P1} and M_{P2} were plotted against Longitude and Latitude and the specific areas of interest investigated.
The data for the 2nd of November is shown in figures 6.84 and 6.85.

The multipath levels at both Wolverhampton and New Street stations can be clearly seen as much larger than elsewhere along the line.
Wolverhampton station is not heavily urban and is not surrounded by high buildings, but due to the train being stationary for an extended period of time, the observed change in M_{P1} and M_{P2} can become very large. There is also a larger volume of data for the trains stationary position in the station which also increases the probability of a high multipath observation being recorded.

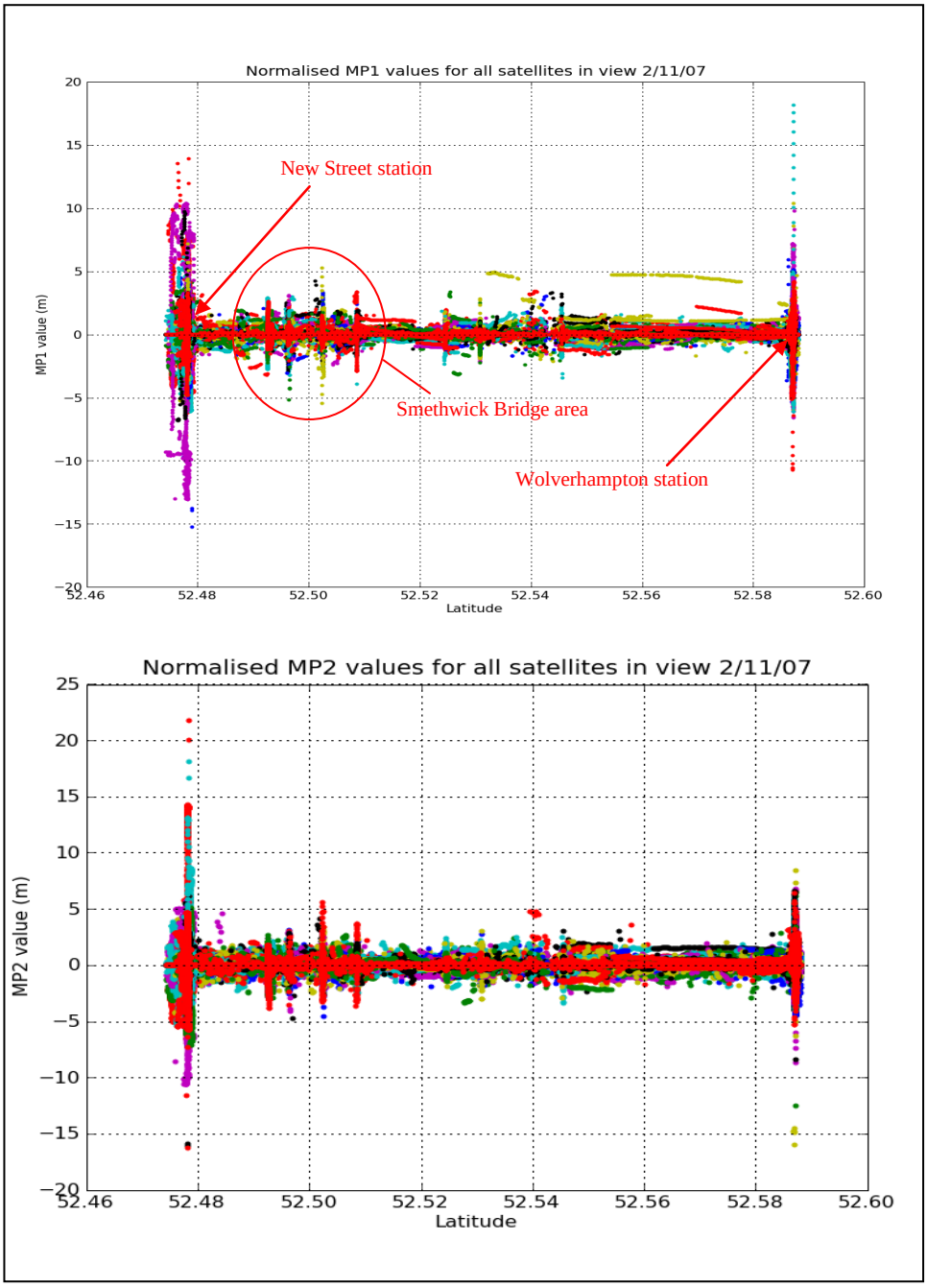


Figure 6.84 - Graphs of MP1 and MP2 against Latitude for Wolverhampton line 2/11/07 (colours represent different satellites)

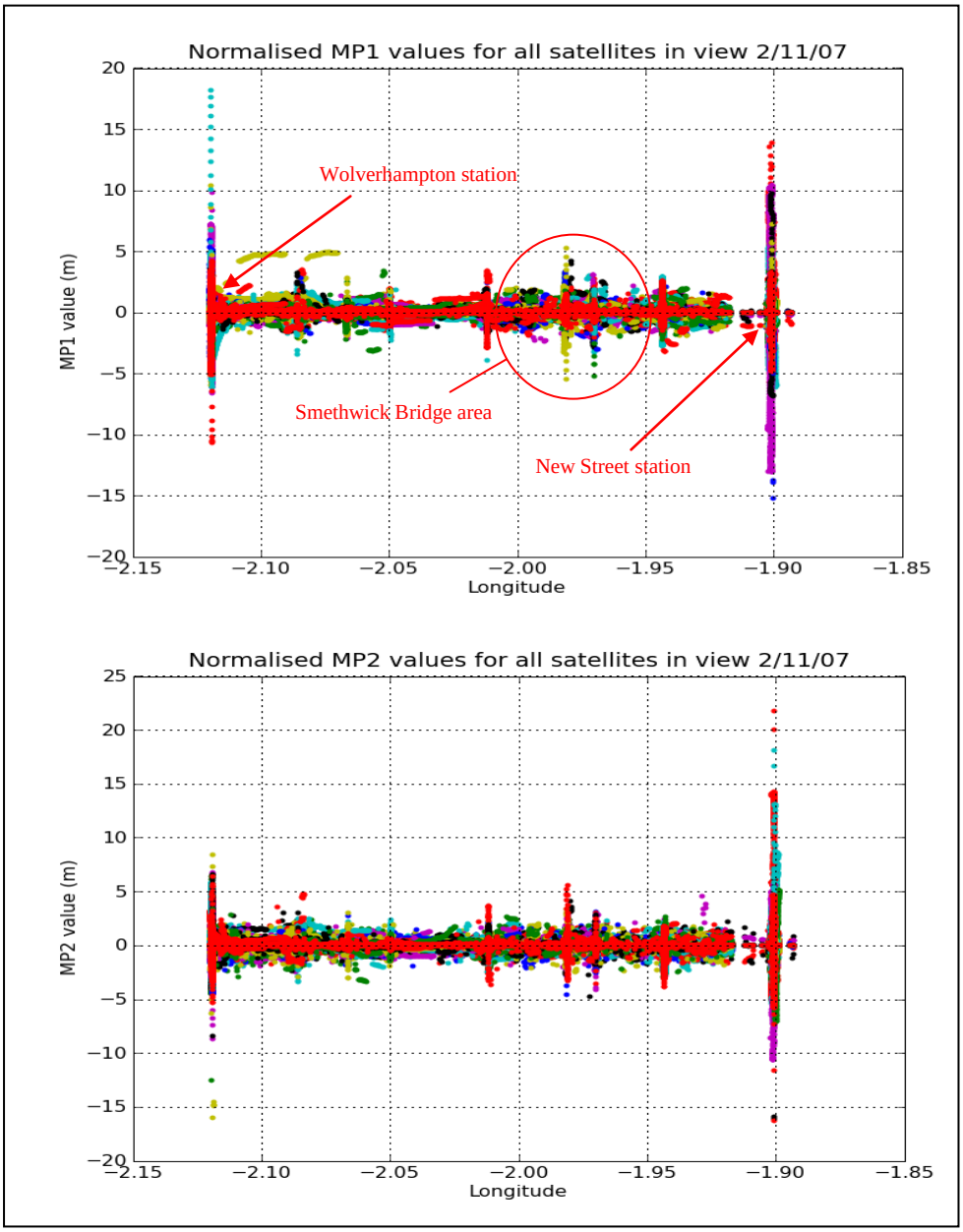


Figure 6.85 - Graphs of MP1 and MP2 against Longitude for Wolverhampton line 2/11/07 (colours represent different satellites)

When analysing the M_{P1} and M_{P2} values for Wolverhampton station, despite them being high, the average is still below 5m whereas for New Street, a comparatively more urban environment, the average multipath level is expectedly higher at 9m.

The position error for the train when stationary inside New Street station (after the tunnel section and overhead roads) was calculated using an averaged values for the first stationary period of the day and plotted against time, this is shown in figure 6.86 along with the M_{P1} and M_{P2} values for all visible satellites in figures 6.87 and 6.88.

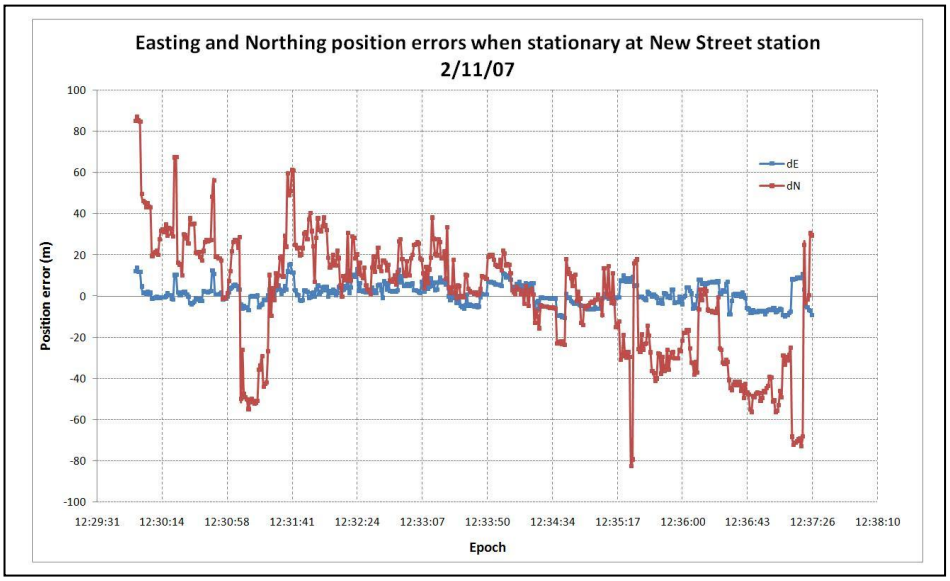


Figure 6.86 - Graph showing Easting and Northing position error for train stationary in New Street station 2/11/07

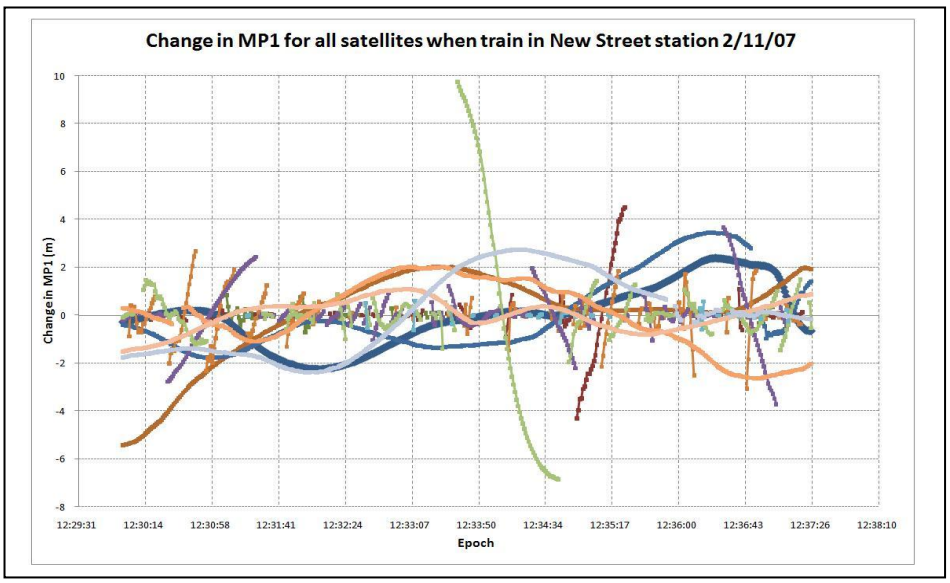


Figure 6.87 - Graph showing change in MP1 for all satellites when train located in New Street station 2/11/07

Given the high levels of multipath shown in figure 6.84 for New Street when plotted against position, the values against time shown in figures 6.87 and 6.88 are similar to the average levels seen along the rest of the track with only a few satellites reaching values of ±10m for M_{P1} and ±6m for M_{P2}.

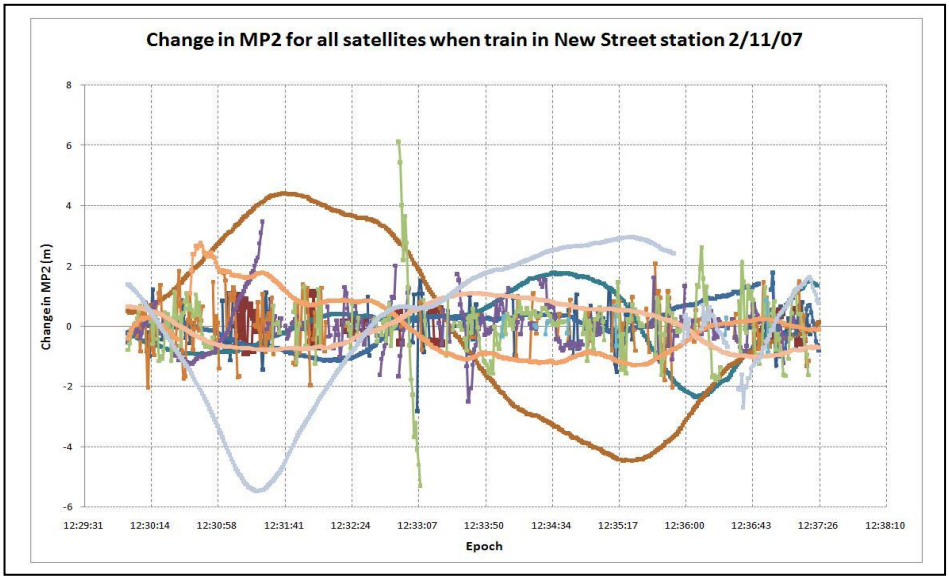


Figure 6.88 - Graph showing change in MP2 for all satellites when train located in New Street station 2/11/07

This is due to the continuously visible satellites having to be of a high elevation angle due to the obstructions and so they produce a lower multipath compared to the lower elevation satellites, but these lower satellites have very broken data sets and so the normalisation for each stream of data produces smaller values for M_{P1} and M_{P2}, skewing the data.

The elevation angles for the satellites are shown in figure 6.89, the breaks in data for the lower elevation satellites can be clearly seen, compared to the continuous data from higher elevation satellites.

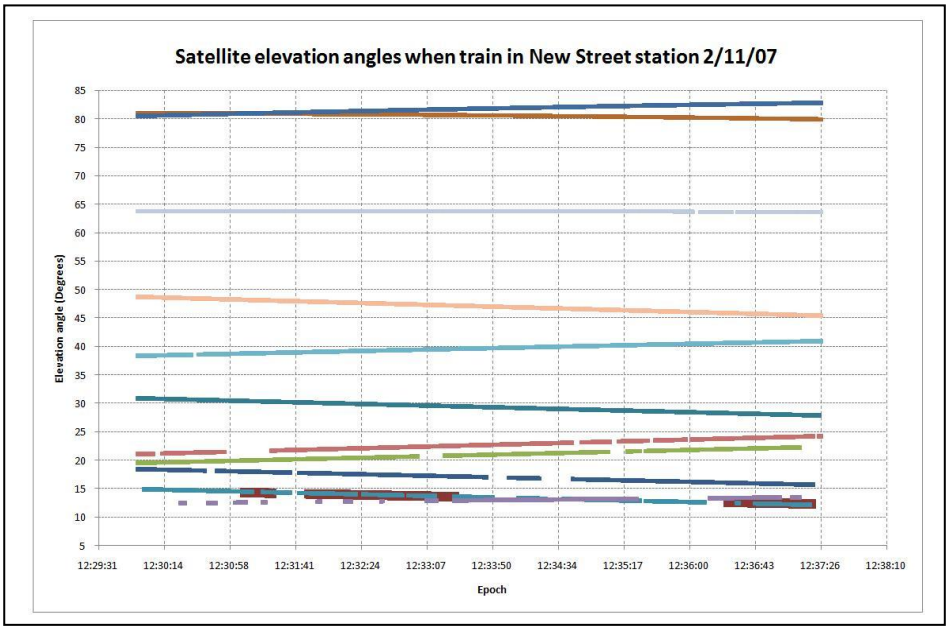


Figure 6.89 - Graph showing Satellite elevation angles for New Street station 2/11/07

The position errors from figure 6.86 seem to be more closely linked to the M_{P2} values seen in figure 6.88 than the corresponding M_{P1} values in 6.87. This is due to the extreme M_{P2} values being much larger (~6m) than those for M_{P1} (~2m) causing the effect on position error to be greater.

The error in Northing is also much larger than the error in Easting due to the nature of the station layout having exposure to the sky being east-west in nature and having large buildings and obstructions perpendicular to the track direction causing an urban canyon effect.

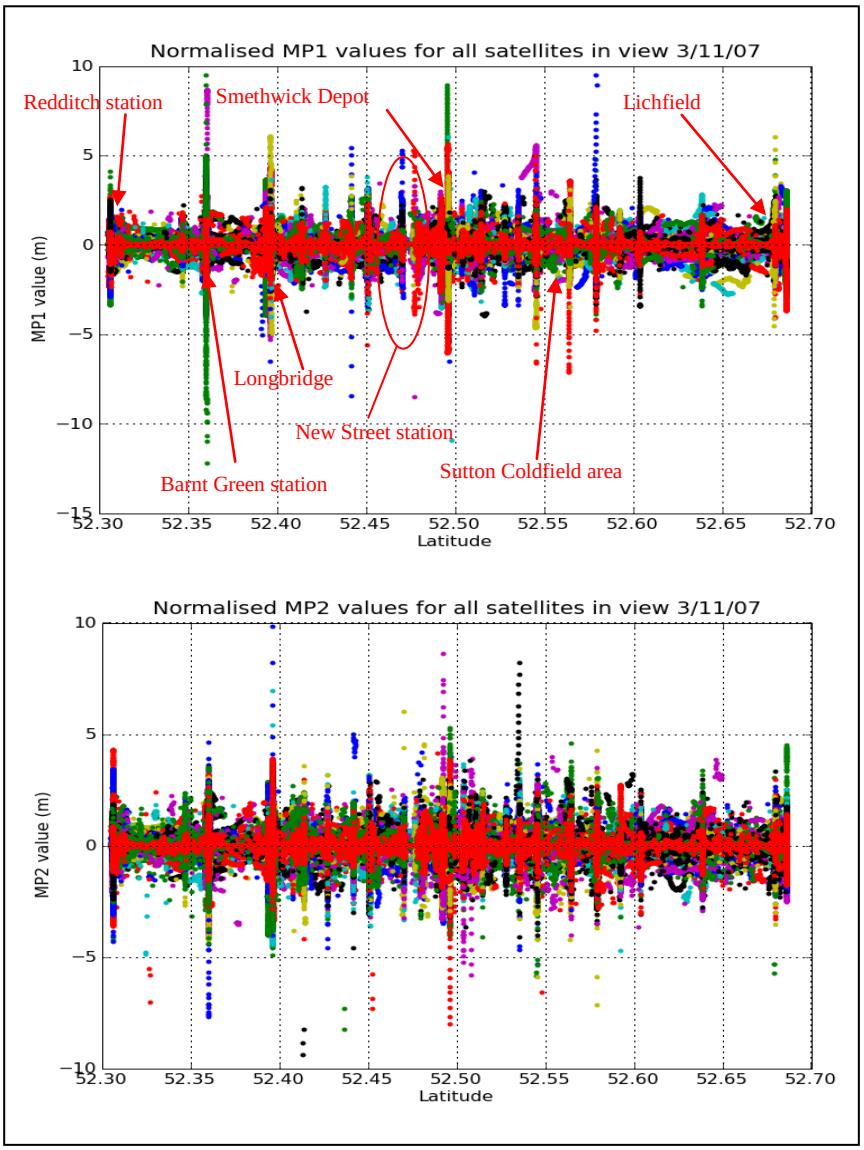


Figure 6.90 - Graphs of MP1 and MP2 against Latitude for Redditch line 3/11/07 (colours represent different satellites)

Figures 6.90 and 6.91 show the M_{P1} and M_{P2} values for the Redditch section of track for the 3rd of November to have a similar level of multipath to the Wolverhampton line when not at either end of the track.

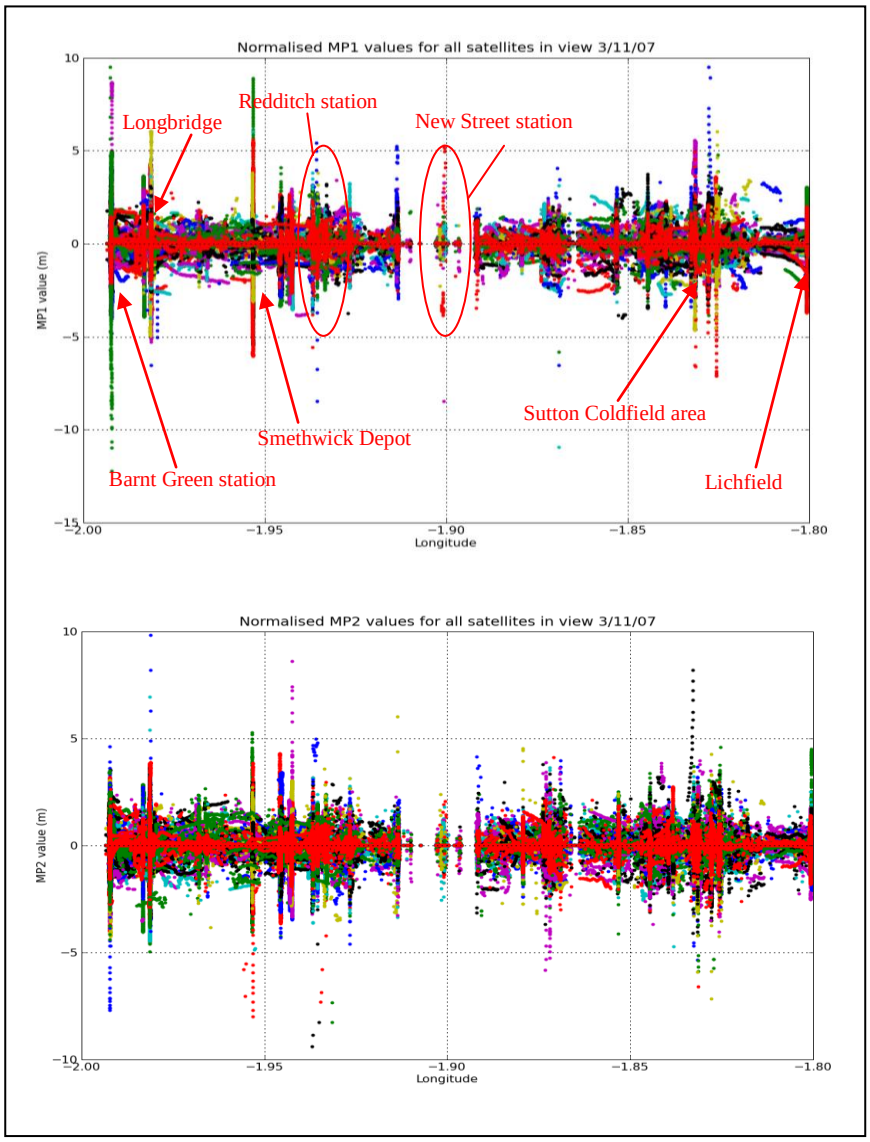


Figure 6.91 - Graphs of MP1 and MP2 against Longitude for Redditch line 3/11/07 (colours represent different satellites)

The individual spikes in multipath can be attributed to the various stations along the line where the train was stationary for an extended period of time and so highlight the importance of a long break-free section of data in order to record enough of a change in either M_{P1} or M_{P2} so that an accurate value can be determined by normalising the data.

In order to compare the M_{P1} and M_{P2} values more closely, the values for satellite 8 have been plotted against latitude in figure 6.92, showing the way in which the M_{P2}

values vary compared to M_{P1}, especially how the larger M_{P2} values for Longbridge are spread out whereas the M_{P1} values are in a more densely populated, smaller spread.

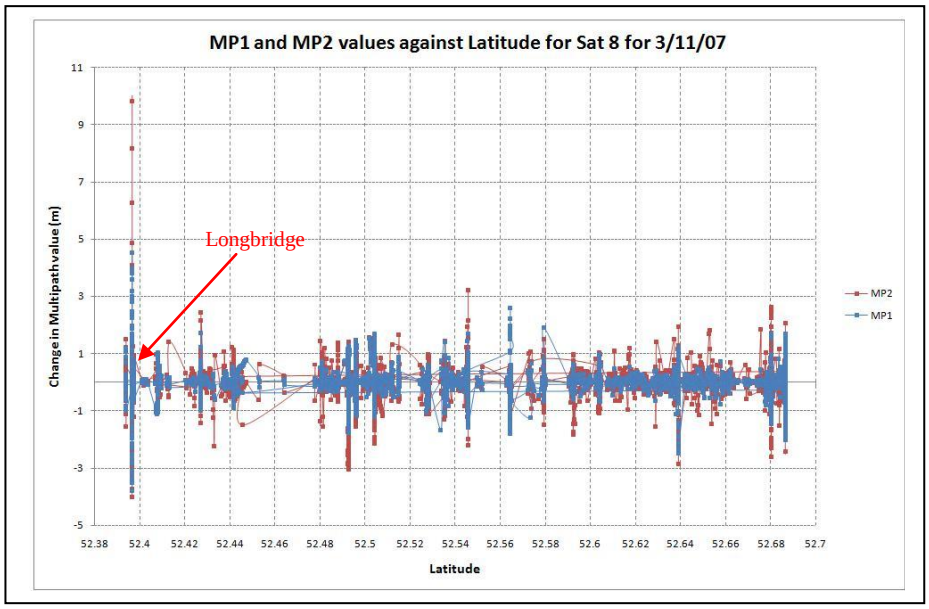


Figure 6.92 - MP1 and MP2 values for satellite 8 for 3/11/07

Figure 6.93 shows the overhead Google earth plot of the positions for Longbridge station. The station has an overhead footbridge as well as a main road overhead that obscures the train's view of the sky and causes significant multipath. On one side of the tracks there are houses and trees and on the other side, behind the main station building is an access road to the elevated section of carriageway.

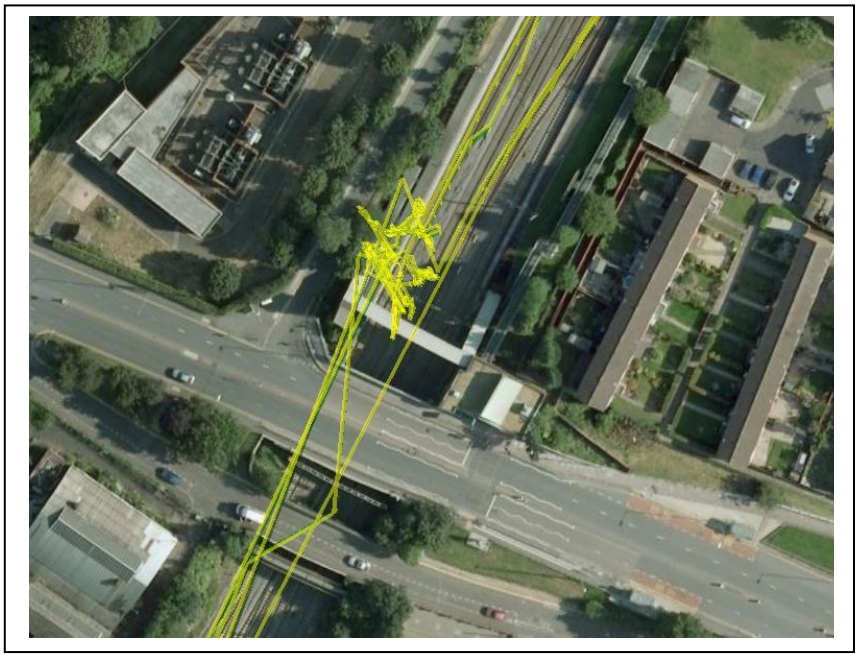


Figure 6.93 - Overhead view of position plots for Longbridge station 3/11/07

Both sides therefore also contribute to the urban nature of the surroundings at Longbridge station and thus to the high levels of multipath experienced by the receiver when stationary in the station.

Another station that has high levels of multipath is at Barnt Green, though the area is mainly rural and without high rise buildings, the trees surrounding the platform area encroach on the view of the sky and appear to be quite large as shown in figure 6.100.

Figure 6.94 - Overhead view of position plots for Barnt Green station 3/11/07

The resulting multipath levels seen are not only due to the train being stationary for a long period of time but also due to the surrounding trees causing interference, as was previously shown in the Severn Valley data.

The average values of multipath change seem to be below 4m until the train enters a built up urban area where they average to around 9m which could cause a problem if a train was trying to distinguish between two parallel tracks (due to the distances between them being roughly 2m) but if the train knows which track it is on, the along track position tends to have lower accuracy requirement.

6.4.4 Conclusions

The volume of data created by the Birmingham exercise provides a better statistical grounding for analysis of multipath in the rail environment than the limited amount of data collected from the Severn Valley.
The data also gives a better idea of the levels of multipath expected from an urban area as seen when comparing the results from New Street against other stations along the route that were in more rural areas, though the effects from foliage and trees is still substantial.

The data analysis shows that when trying to model an expected level of multipath, the assumption of normally distributed data is not sufficient and qualitative generalisations must be included if the expectation of a large error is to be extrapolated.
The data shows that despite maximum levels of multipath change reaching 35m, the average value expected, even for areas surrounded by large buildings is around 4m and the position errors seen for New Street station are primarily in the across track direction due to the along track urban canyon effect. Using GDOP as a measure of possible position shift is not ideal and a full geometric analysis would need to be carried out if trying to predict a position shift.
These worse case scenarios would need to be included in the safety case log as examples of possible hazards and situations where additional sensors and techniques may be required to produce results within the required safety margins.
The possible faults due to these extreme cases would also need to be documented, such as the reduced position accuracy and integrity as well as any possible safety monitoring system alarms that may occur.
Mitigation of the risks associated with these faults would also need to be detailed so that a solution complying with the ALARP principal can be designed.

When collating all available data the results were very similar to those from individual days and so assumptions and models of data from limited sets could produce results comparable to those extracted from far larger sets, provided they are for the same line, as the difference in results for the Wolverhampton and Redditch

line can be seen for individual stations, though the average level of multipath change is around 4m for both.

One of the main issues arising from the data and the analysis undertaken is the lack of availability of long break free data streams for both low and high elevation satellites along the length of the track. Because of this, the levels of multipath change calculated may not be entirely correct.
Despite this, when multipath is observed and position errors result, it is possible that track database integration into the position solution could reduce the level of position error as well as improve the level of receiver autonomous integrity monitoring by reducing the number of satellites required for a position fix, thus increasing the possibilities of detecting measurement errors on redundant satellites, something that would also be possible with additional satellites provided by Galileo.

7 Summary and Conclusions

7.1 Summary of results obtained

This thesis provides a qualitative and quantitative view of multipath experienced in various environments in the UK network, along with an explanation of the different safety analysis techniques that can be applied in order to assess the danger multipath poses to the safety of life aspect.

The multipath investigations were carried out on the Severn Valley and Birmingham New Street area lines.
The Severn Valley railway was used for an initial experimentation exercise utilising the rurally located 26 Km line to collect moving data from a train in order to gain an understanding of how a geodetic grade receiver would work in kinematic mode in such an environment. One of the main issues encountered was that for the 6607 epochs of data collected (over a 1 hour journey); only 26.5% allowed a phase ambiguity fix when 4 or more satellites were in view and so high precision phase based position solutions were difficult to compute.

Despite dual frequency code and phase data being collected, code based positions were calculated using the RTK Library and Google maps allowed the trains position to be plotted. Breaks and jumps were seen due to tunnels and the loss of the L2 signal due to tree canopies and other line side infrastructure. The change in the calculated code multipath observable on both the L1 and L2 frequency was within 1.8m for all the observed satellites, but due to the limited data set, this was not considered a full investigation, merely a useful exercise to prepare for the later Birmingham data where a higher volume of data can provide better statistical certainties for possible signal errors that may contribute to a system error that would need to be added to a HAZOP. The speed of the train was roughly 30mph but when entering stations, this was obviously reduced, as can be seen by the closer spacing of each epoch position. This provided a higher resolution to possible position errors and caused no detrimental effects to the overall data.

The Birmingham New Street data was collected from two main lines, one from Lichfield to Redditch and the other between New Street and Wolverhampton, both of which have dense urban (around New Street station) and rural surrounding infrastructure.
A geodetic quality receiver collecting dual frequency code and phase data was used and the code multipath observable calculated. As this method is normally used for the assessment of a static base stations local environment, the breaks in both L1 and L2 signals again caused issues as the resulting jumps in phase counter measurements meant the normalisation applied to the changing multipath had to be restarted.
The resulting data was limited in its use as signal breaks were common, especially in highly urbanised or overgrown areas of track. Geometric analysis of the data was found to be non ideal due to receiver clock jumps causing errors in the results and so the novel application of code multipath observable formation was used for larger sets of data in order to compensate.
The data used was collected over 11 days and is a subset from a larger exercise lasting 12 months. This larger data set allowed for averaging to take place and the quantitative and qualitative analysis of a much greater range of break free data sets. In doing this, the results obtained are expected to be closer to the true values, but are still believed to be a lower estimate of the average value as the breaks still present in much of the data are likely to reduce the normalised change in M_{P1} and M_{P2}.

Using Google earth, position errors were also plotted so that jumps due to specific buildings, such as rural stations and the main New Street station could also be seen. Signal reflections and obscuration were also seen due to overhead walkways and roads for limited periods, highlighting the need for additional positioning inputs for short periods of GNSS outage.
This outage issue would be reduced when future systems such as Galileo and the updated Russian GLONASS system are used as the number of satellites available would be nearer 90. This would increase availability for more enclosed areas, but tunnels and bridges would still need to be accounted for.

The results show that the average level of multipath did not tend to exceed 4m with roughly 60% of the break free data sets lasting between 1 and 2 minutes. When analysing the numbers of each change in multipath, the underlying distribution cannot be described as normal (shown by χ^2 tests), with a very high probability of smaller multipath values expected.

The extreme levels of multipath can exceed 35m, but the occurrence of these extremes are limited and generally located in stations either allowing long break free data sets in comparatively high multipath environments, or short data sets in very high multipath environments like New Street station, but due to the nature of the receivers internal correlator, the likelihood of multipath of this level is unlikely and thus it is more likely that these extreme values are due to reflected only signals not being rejected.
These extremes, although uncommon, would need to be included in any evaluation of multipath in a safety case with the effects being categorised for FMECA analysis. This would be accounted for in an FMECA table as the failure being a position error value higher than a level considered safe with the effect being logged as possible position error outside the desired RNP and cause being a high level of multipath.

For a better evaluation of these areas, traditional static survey techniques could be carried out in order to collect sufficient exposure to individual satellites for a full multipath examination to be carried out. Together with computer modelling of the surrounding infrastructure, expected direct, reflected and multipath signals could be calculated.

Recommended actions could either be the mitigation of multipath through advanced receiver design or the addition of external sensors such as three axis inertial units to reduce the overall positional error provided it conforms to ALARP principals and that system redundancy is not an issue.

The investigation into possible EM interference from overhead cables was done using data recorded from a dual frequency geodetic grade receiver placed on the station platform near the electrified overhead lines at Watford Junction station. Another dual frequency receiver was placed on top of a nearby building as a reference.
By looking for abrupt changes in the code multipath observable as the satellite line of sight passes through the overhead cables, any apparent EM effects were investigated. The results show that any anomalies were not detected and thus it is assumed the possibility of errors caused by electric fields from quaternary lines is unlikely.
Due to large sections of the UK railways being converted to overhead electric power, these findings should provide significant support for the use of GNSS in these areas without worrying about signal interference.

It is hoped that the results and conclusions of this research will provide a solid scientific grounding when attempting to create a safety case for the use of GNSS in the railways. As exact multipath values cannot be determined due to the variation and volume of available data, the safety analysis techniques used would have to cater for the qualitative aspects of the conclusions as well as the results defined from the statistical analysis of the collected data in order to come to an informed conclusion.

The results obtained outline the main issue for the use of GNSS in the UK railways as being satellite availability rather than multipath. As discussed, the code multipath mitigation techniques used by modern receivers are designed to produce results well within the expected requirements for along track position accuracy.
If across-track position is not required for track selection purposes during service, or for train awakening, GNSS is a viable option as a location technology, provided it is augmented with systems to facilitate position determination during periods of satellite outage.

Given the volume and variance of data collected and analysed in this thesis, the characterisation of multipath in the rail environment provided here is unlikely to be sufficient evidence for a full safety case to be produced.
It is difficult to precisely predict the level of multipath experienced by a train mounted receiver in the UK railways and so further analysis would be required in order to increase the level of confidence in the findings, especially when trying to quantitavely justify compliance with SILs and RNPs.

7.2 Contributions to a safety case

Past European projects to determine the viability of GNSS as a positioning system for railways have been based upon new ideas and possible proof of concept and so there are no 'Good Practice' references and no historical data for implementing GNSS as a safety of life application in the rail domain. Because of this, any attempt to introduce GNSS in this capacity requires a 'Value Based' approach using qualitative and quantitative analysis with conservative approximates of the possible hazards that may occur as the system would be required to meet SIL4 if used for passenger trains.

Through this thesis, the possibilities of these hazards due to EM interference from overhead cables and multipath have been investigated.
The results obtained have removed EM radiation as a hazard (though further conclusive testing is advisable) and have defined the levels of multipath to be expected in various types of surroundings found in the UK rail network.
For a safety case to be complete, all possible outcomes from multipath (within reasonable limits) need to be recorded, no matter how unlikely they are to occur so that not only the random events (such as the 35m levels recorded in New Street station) are detailed so that stringent mathematical analysis can take place, but the systematic errors (such as the average 4m of multipath seen) are also recorded so that SIL categorisation can take place.

Availability of satellites was a known issue from the beginning and has been detailed in previous UK reports (RSSB, 2007) concluding that roughly 10.5% of the British rail network at route level receive no or poor GPS coverage for distances of 150m,

this equates to 1885km of track where GPS availability is below standards required for even the most basic RNPs. This could be improved by including track databases in the positioning algorithm, reducing the number of satellites required, increasing the overall coverage but the issue of tunnels and overhead infrastructure is still unavoidable, thus the inclusion of additional sensors is mandatory for any positioning system required to operate continuously.

Causal analysis of multipath would attribute the surrounding infrastructure as the major reason with satellite position as an enhancement or reduction of its value. By implementing advanced receiver architectures, the reduction of multipath and increase in integrity can be obtained with current Commercial off the Shelf (COTS) equipment and so when undertaking consequence analysis the potential for hazardous levels of position error and low integrity are reduced.

The potential for false trust in the systems output due to multipath is possible but given the levels of multipath observed on a general level, it is believed that the errors in position would be within the required levels and that any additional sensors or track database integration should improve errors and facilitate higher RAIM availability so that the system could be trusted.

If the seven stage process detailed in the Engineering Safety Management book (Railtrack, 2000) were to be carried out, the analysis would take the form similar to figure 7.1 using the data and conclusions from this thesis as well as from other projects and further work.

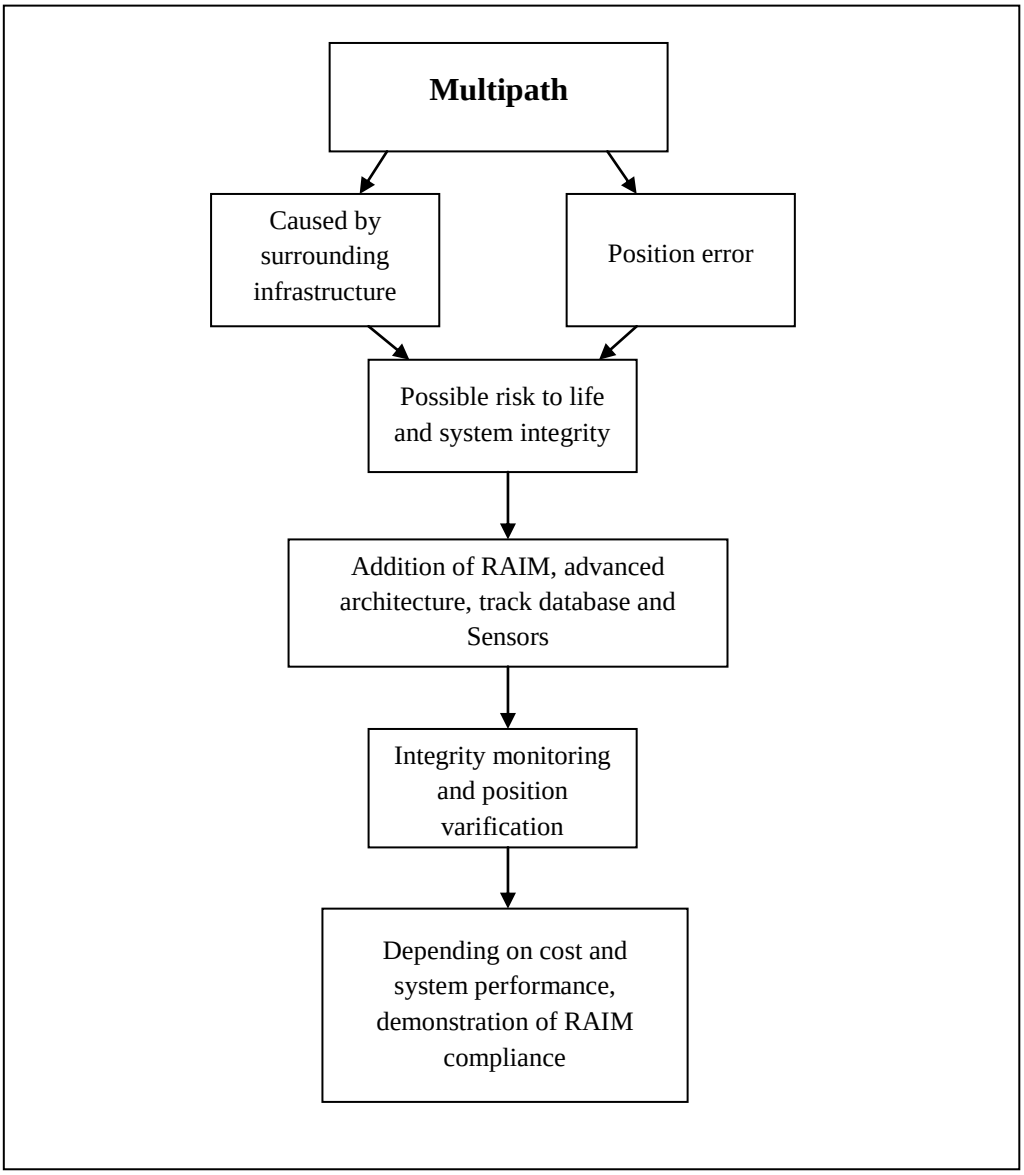


Figure 7.1 - Example of safety case structure for Multipath risk assessment stages

7.3 Further work

The work carried out in this thesis has given a solid grounding into the effects of multipath in the UK rail environment. The results are from three specific lines in the UK and so will not necessarily represent the entire of the UK or Europe in terms of the possible variation in multipath.

Larger data sets from other tracks throughout the UK would help to strengthen the claims made in the conclusions of this work by increasing the statistical reliability and

including the possibility of any geographical variance in satellite observations either due to the track heading or any orbital anomaly (though this is highly unlikely).
The tracks under investigation have had little if any sections located in deep cuttings and so the possible effects of these is not well represented, this would also be another factor to consider when locating any possible further data collection sites.

Using complex clock modelling techniques would provide an accuracy increase to allow highly precise point positioning to occur in order for the geometric multipath value to be calculated on a single epoch basis, rather than having to rely on a continuous stream of data for normalisation.

Further investigation into possible EM effects from overhead cables may need to be carried out in order to provide a definite conclusion, as the experiment carried out could have been improved on in order to be more rigorous, with a full site survey of the train station being carried out and precise cable positions mapped. Given the location of the receiver used in Watford Junction, any direct interference from the overhead cables with the antenna itself will be limited due to the distances being roughly over 3m. There may be possible effects when placing the antenna on top of a train as any EM field created by the overhead cables will be far stronger in close proximity due to the inverse square law. This should also be investigated further but the results from the Birmingham data did not show any errors when passing under electrified sections of track.
In addition to this, testing of a receiver in an anechoic chamber using high sensitivity detectors and clearly defined signal transmission paths would be ideal so that all possible background and infrastructure related interference is removed and thus any interference experienced can be definitively apportioned to any test EM source present.

Given the use of GNSS in the railways as a future possibility for control and signalling, augmentation would almost certainly be required due to the obvious satellite availability issues encountered throughout the data. The various types of augmentation with other devices (dead reckoning) have different positive and negative selection criteria and depending on the RNP limits set and the cost of the

alternatives an ALARP combined system could be produced, though if this would be available as a commercial, off the shelf solution is unknown.

One of the major concerns with devices such as accelerometers and gyroscopes is the amount and speed by which they drift from the true value. Cost – benefit analysis must be carried out when selecting as normally low cost inertial sensor almost always mean a high drift rate and therefore a short time for inaccuracies to build up in the position solution. This could be directly linked to conditions such as tunnel length and overhead bridge frequency combined with the receivers time to re-acquisition of satellites in order to determine the maximum likely period without satellite coverage.

The Figure below shows the possible change in position error for a train travelling at 30 km/h if a GPS outage was to occur, i.e. if a train enters a tunnel, how the predicted position error increases over time.

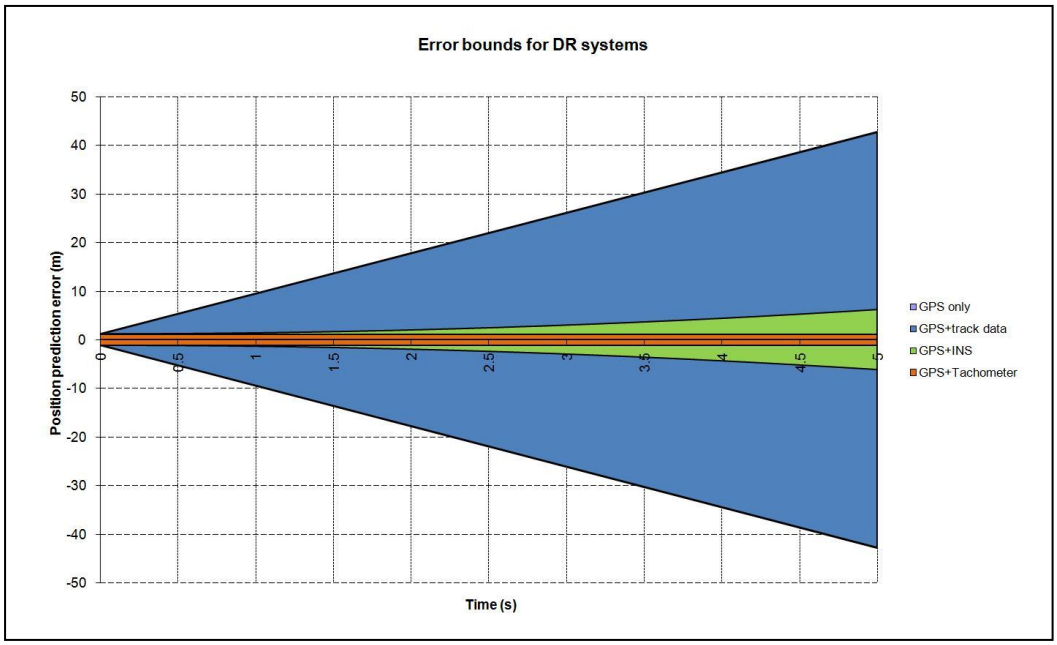


Figure 7.2 - Graph showing along track position error drift rates after a GPS outage

The graph shows that despite a track database being integrated with the GPS in order to obtain a better initial position, as there is no GPS for the duration of the plot, the position error on the train is purely down to the speed of the train being an unknown (assuming the train is breaking/accelerating within a 30 km/h bound) and the assumption that it is still travelling at 30 km/h.

The INS drift rate used is that of the system installed on board the Birmingham train and is the manufacturer stated 0.2m/s^2. The use of a tachometer is assumed to be without any noticeable drift within the 5 second time frame.

This graph should be considered when designing an integrated system as despite tachometry having a significant drift rate over long distances, this can be re-set when GPS signals are available so that when a tunnel or overhead obstruction causes a GPS outage, the tachometer can take over the along-track position determination for a limited time until GPS is re-acquired. This solution, although making use of pre-existing sensors on the train may not be the cheapest option as tapping into the tachometer output may be more complicated and thus more expensive (and potentially less reliable) than to simply add INS to the installed GPS package.

Additional augmentation systems can be considered to remove the possibility of position error, specifically, the errors encountered due to the loss of satellites or signals. When the number of satellites drops to 4, the position error can increase significantly (as seen throughout the results in this thesis). This is due to the possibility of a satellite signal error being detected being very low as there are no 'spare' satellites to compare with.
This is the basis of RAIM, mentioned in section 4.1 and can be improved using a known track database to reduce the three dimensional position problem normally used to a one dimensional problem (along track) (Zeng & Cross, 2008). Rather than requiring 4 satellites to get a position, only 2 are needed, thus giving the user 2 'spare' satellites to provide a higher level of RAIM availability.

The level of modern hardware sophistication available for multipath mitigation should reduce the errors seen in this thesis even further, strengthening the conclusion that for the UK rail environment, multipath should not be a major concern when considering along track position determination provided adequate additional sensors are used to cater for satellite availability issues.

References

ACSNI. (1993). *Study Group on Human Factors : 3rd Report: Organising for safety*. Health and Safety Executive (HSE).

Adelard. (2006). *Safety Course Manual* (Vol. Vol. 1&2). London: Adelard.

ANASTASIA project. (2006). Characterisation of interference and multipath signal environments. In *Interference and Multipath Mitigation Study Report* (pp. 24-30). ANASTASIA PMC.

Barbu, G., & Alcouffe, F. (1999). *Telematics Application Program: User needs and technology assessment.* APOLO Project TR 4003.

Blewitt, G. (1997). Basics of GPS Technique: Observation Equations. In *Geodetic Applications of GPS.* Swedish Land Survey.

Braasch, M. S. (1996). Multipath Effects. In J. S. B. W. Parkinson, *Global Positioning: Theory and Applications* (pp. 547-568). AIAA.

Bradbury, J. (2007). Prediction of Urban GNSS Availability and Signal Degradation Using Virtual Reality City Models. *ION GNSS 2007* (pp. 2696-2706).

Brunner, F. K., Hartinger, H., & Troyer, L. (1999). GPS Signal Diffraction Modelling: The Stochastic Sigma Model. *Journal of Geodesy , Vol. 73(5),* (pp. 259-267).

BSI. (1990). *Reliability of Systems, Equipment and Components - Guide to Fault Tree Analysis.* Britich Standards International.

Collins, J., & Langley, R. (1999). *Possible Weighting Schemes for GPS Carrier Phase Observations in the Presence of Multipath.* New Brunswick: U.S. Army Research Office.

Cross, P. (2007). High Integrity Safety-Critical GNSS Applications on The Railways. *2007 IET Seminar on Global Navigation Satellite Systems,* (pp. 123-142). Portsmouth: IEEE.

Cross, P. (2007). Lecture Notes for Positioning 1 (GEOMG004) and Positioning 2 (GEOMGS05); Code Multipath Observable - UNAVCO. London.

De Jonge, P., & Tiberius, C. (1996). *The LAMBDA method for integer ambiguity estimation: Implementation aspects.* Delft: Delft Geodetic Computing Centre.

Enge, P. (1999). Local Area Augmentation of GPS for the Precision Approach of Aircraft. *Vol. 87, No.1* (pp. 111-132). IEEE.

Fenton, P., & Jones, J. (2005). The Theory and Performance of NovAtel Inc.'s Vision Correlator. *ION GNSS 18th International Technical Meeting* (pp. 2178-2186). Long Beach, California: ION.

Fenton, P., Townsend, B., & Diedrendonck, K. (1995). L1 Carrier Phase Multipath Error Reduction Using MEDLL Technology. *ION GPS-95* (pp. 1539-1544). Palm Spring: ION.

Filip, A., Polivka, A., & Suchanek, J. (2006). Practical Analysis of Satellite Navigation Performance for Train Control. *7th World congress on railway research.* Montreal.

Goodman, J. L. (2003). A Software Perspective on GNSS Receiver Integration and Operation. *Satellite Navigation Systems: Policy, Commercial and Technical Interaction, International Space University, Strasbourg ,* (pp. 119-126).

Gurtner, W. (2006). *The Receiver Independent Exchange Format 3.0.*

Hopfield, H. S. (1963). The Effect of Tropospheric Refraction on the Doppler Shift of Satellite Data. *Journal of Geophysical Research , Vol. 68,* (pp. 5157-5168).

IERR. (2002). *Virtual BALISE Functional Requirement Specification - project 2544.* Union Internationale des Chemins de Fer (UIC).

Jones, J., Fenton, P., & Smith, B. (2004). Theory and Performance of the Pulse Aperture Correlator. *Pulse Aperture Correlator* . NovAtel Inc, http://www.novatel.com/Documents/Papers/PAC.pdf.

JPL. (2007, November 23rd). *JPL TEC map page*. Retrieved November 23rd, 2007, from JPL: http://iono.jpl.nasa.gov/latest_rti_global.html

Kaplan, E. (2005). *GPS: Principals and Applications, Second Edition.* Artech House.

Kavak, A., Vogel, W., & Xu, G. (1998). Using GPS to Measure Ground Complex Permittivity. *Electronics letters 5th February , Vol. 34,* (pp. 254-255).

Kavak, A., Xu, G., & Vogel, W. (June 1996). GPS Multipath Fade Measurements to Determine L-Band Ground Reflectivity Properties. *Proceedings of the 20th NASA propagation experimenters meeting. Fairbanks* (pp. 257-263).

Kayser-Threde. (2004). *INTEGRAIL*. Retrieved 07 09, 2008, from INTEGRAIL: http://integrail.kayser-threde.com/ig-proj.html

Klobuchar, J. (1987). Ionospheric Time-Delay Algorithm for Single-Frequency GPS Users. *IEEE Transactions on Aerospace and Electronic Systems, Vol. 23,* (pp.325-331).

Lancien, D. (2005). *Compiled evaluation results, system safety report.* LOCOPROL IST 2001-28103.

Lau, L., & Cross, P. (2005). Use of Signal-to-Noise Ratios for Real-Time GNSS Phase Multipath Mitigation. *Nav 05.* London: Royal Institute of Navigation.

Lee, I.-S., & Ge, L. (2006). The Performance of RTK-GPS for surveying under challenging environmental conditions. *Earth, Planets, Space , Vol 58,* (pp. 515-522).

Leica. (1999). System 500 Technical Specification. Leica Geosystems.

Mokkapati, C., & Pascoe, R. (2005). ARRC Collision Avoidance System. *2005 ASME/IEEE Joint Rail Conference* (p. 34). Union Switch & Signal Inc.

Niell, A. E. (1996). Global Mapping Functions for the Atmospheric Delay at Radio Wavelengths. *Journal of Geophysical Research , Vol 101(B2),* (pp. 3227-3246).

Office of Rail Regulation. (2005). *Railway Safety Principles and Guidance.* Health and Safety Executive.

Olsen, R. G. (1994). Power-Transmission Electromagnetics. *IEEE Antennas and propogation magazine, Issue 36* , 7-16.

Parkins, A. (2006). GPS on the railway: A study of the value of incorperating knowledge of the track in a train positioning solution. *MSc Thesis.* UCL.

Parkins, A. (2007). RTK Library modular design from group RTK work.

Phelts, E., & Enge, P. (2000). The Case for Narrowband Receivers. *National Technical Meeting.* San Diego, CA: Institute of Navigation.

Radsky, W., Baum, C., & Wik, M. (2004). Introduction to the Special Issue on High-Power Electromagnetics (HPEM) and Intentional Electromagnetic Interference (IEMI). *IEEE Transactions on Electromagnetic Compatibility , Vol. 46(3),* (pp. 314-321).

Railtrack. (2000). *Engineering Safety Management Issue 3.* Railtrack PLC.

Ray, J. K. (2000). *PhD Thesis, Mitigation of GPS Code and Carrier Phase Multipath Effects Using a Multi-Antenna System.* Calgary Dept. Geomatic Engineering.

RSSB. (2007). *T740 Project Report - GNSS Data Coverage.* London: RSSB.

Saastamoinen, J. (1972). Atmospheric Correction for the Troposphere and Stratosphere in Radio Ranging Satellites. *The Use of Artificial Satellites for Geodesy, Geophysical Monograph Vol. 15* , (pp. 247-251).

Silva, J. M. (2001, April). Evaluation of the Potential for Power Line Noise to Degrade Real Time Differential GPS Messages Broadcast at 283.5-325 kHz. *Power Engineering Review, Vol. 21* , (pp. 69-69).

Silva, M., & Olsen, R. (2002, October). Use of Global Positioning System (GPS) Receivers Under Power-Line Conductors. *IEEE Trans. Power Delivery Vol. 10(4)* , (pp. 938-944).

Southeastern. (2005). *Internal Presentation: Selective Door Operating on Class 375.* London: Southeastern.

Spofford, P., & Remondi, B. (1991). *The National Geodetic Survey Standard GPS Format SP3.* Maryland: NGS.

Ulrech, A. (2002). *Report: GADEROS Project Overview.* Contract: GRD1-40230-2001-SI2.324586.

UNIFE. (2007, November 14th). *ERTMS.* Retrieved November 14th, 2007, from ERTMS: http://www.ertms.com/

Wang, J., & Satirapod, C. (2000). Comparing the Quality Indicators of GPS Carrier Phase Observations. *Geomatics Research Australasia , Vol. 73,* (pp. 75-92).

Weiss, J., & Axelrad, P. (2007). Estimation of Simplified Reflection Coefficients for Improved Modeling of Urban Multipath. *Institute of Navigation 63rd Annual Meeting* (pp. 635-643). Cambridge, MA: ION.

Wu, F., & Zhang, K. (2005). GPS Signal Acquisition and Tracking Using Software GPS Receiver. *2005. Proceedings of the Eighth International Symposium on Signal Processing and Its Applications,* (pp. 843-846).

Xu, G. (2003). *GPS Theory, Algorithms and Applications.* Springer.

Yionoulis, S. M. (1970). Algorithm to compute Tropospheric Refraction Effects on Range Measurements. *Journal of Geophysical Research Vol. 75* , (pp. 7636-7637).

Zeng, Y., & Cross, P. (2008). *Report: Improvements of GNSS Integrated with a Digital Route Map for Safety Critical Railway Applications.* London: University College London, Department of Civil, Environmental and Geomatic Engineering.

Zinas, N. (2007, June). Sensor Network Approach to GPS RTK. *Mphil/PhD Transfer Report* . University College London, Department of Civil, Environmental and Geomatic Engineering.

A. Appendix A - The Safety Case

7.4 Safety case example – Watford Junction experiment

Below is the safety case used for the Watford Junction data collection exercise. It was submitted to the station manager and approved.

Watford junction train station data collection safety case

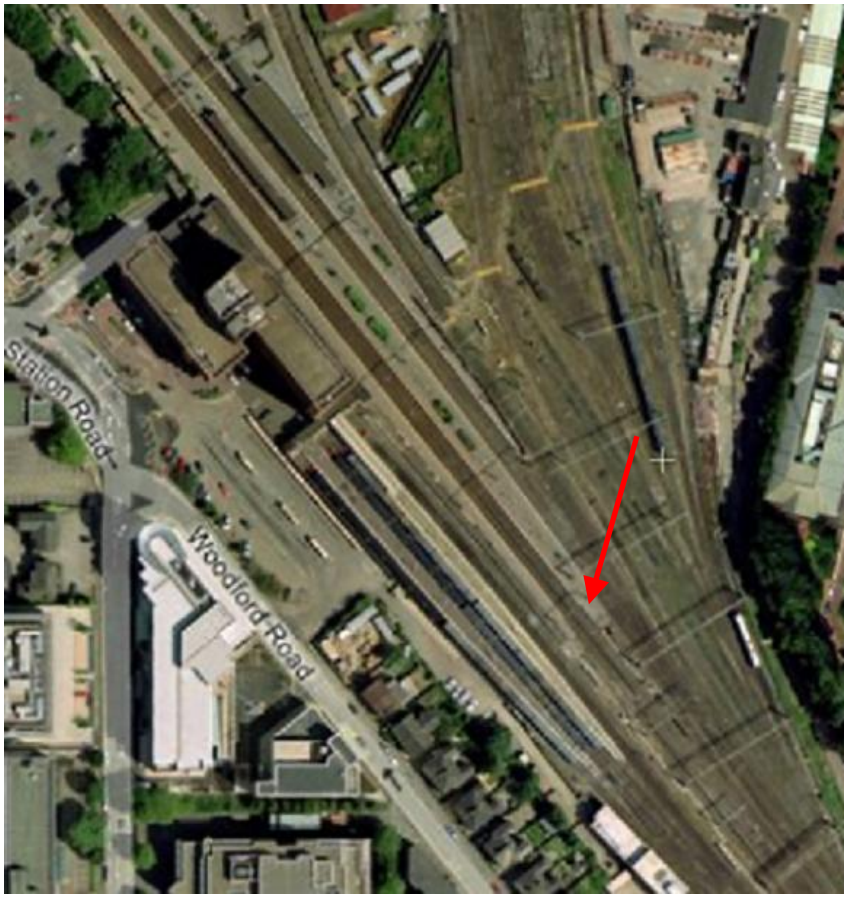
Proposed experiment:

We wish to record GPS signals on the platform at Watford Junction station using a Leica GPS500 receiver system which consists of a tripod mounted antenna, receiver pack and self contained battery. The tripod is a standard issue heavyweight surveying tripod (that is commonly seen on building sites) and the antenna increases its height by roughly 30 centimetres. The receiver unit and battery attach to the side of the tripod and are boxes the size of an A4 file and a standard video camera battery.
We also wish to use a tripod mounted theodolite to measure the vertical angle of the over head lines, this involves looking through an eyepiece of a telescope until it is aligned with the item being surveyed and is purely a remote sensing device.
A picture of the equipment is shown below;

GPS equipment on tripod

Theodolite on tripod

The positioning of the GPS receiver and the total station would be above the same ground mark and so only one tripod will be used, with the different measurement devices being interchanged accordingly. The position of the ground mark would be in the south east end of the main central platform, next to one of the main supporting pylons as this is away from the main flow of passengers and gives the best position for data collection., as shown on the map below;



Map of Watford junction station showing preferred location of equipment

The equipment will be positioned in the centre of the platform at least 2500mm from the platform edge and over 3000mm from the overhead lines, and will be accompanied by someone at all times who will also adhere to these distances and make sure that members of the public are kept clear of the equipment.

Procedures:

The persons carrying out the experiment will register themselves before starting work, by 'signing-in' in the station visitors book, and receiving a Silverlink safety briefing, or as directed by the Silverlink Group station manager or their nominated representative. A record of the Silverlink safety briefing is to be retained by Silverlink for audit purposes. The persons will also 'sign-out' of the Station visitor book, at the end of any working period, or when obliged to quit site, or as directed by the Silverlink Group station manager or their nominated representative.

All persons associated with the experiments will have in their possession at all times, current and valid personal photo ID.
Silverlink is to be fully indemnified against any claims which may arise as consequence of, or attributable to, the experiments.
In the event of an incident, the procedures detailed in the Silverlink safety briefing will be complied with.
The persons will wear at all times, Hi Visibility vest / jacket, to an approved colour and specification, issued by Silverlink.
All vehicles will be parked in designated areas as agreed with the Silverlink Group station manager or their nominated representative, and Silverlink recompensed accordingly.
Access to station facilities (e.g. toilets) will not be required and local public services will be used if required.
Access to any secure station areas is not required.
Access to station electrical supplies is not required.
Access to electrical intake areas is not required.
All cabinets, compounds, storage areas etc. are to remain secure at all times.
Access to station services (e.g. water) is not required.
Sufficient protection will be provided to mitigate injury or loss to any persons from the equipment during unloading / loading, transfer to / from site, or positioning on site.
Means of protecting surfaces including but not limited to, car park, platforms, stairs, and subway, will be provided to mitigate damage from the equipment during unloading / loading, transfer to / from site, or positioning on site.
All damage caused to the station fabric and its facilities, is to be made good to the written satisfaction of the Silverlink Group station manager or their nominated representative, and route property manager.
The experiment site and access routes within the station are to be kept clear and free from litter, obstructions or other hazards. These areas are to be examined with the Silverlink Group station manager or their nominated representative, to ensure the area is safe for use.
No means of gaining height is to be used in association with these experiments.

Cleaning regimes are to be invoked to ensure that all waste and residue caused by the experiment (including litter), is cleared from the station fabric and its facilities. These areas are to be examined at the end of each period with the Silverlink Group station manager or their nominated representative, to ensure the area is acceptable to Silverlink.

Date of work:

The date and time when the experiment will be carried out is to be agreed between the station manager and the Department of Geomatic Engineering but is expected to be during normal operating hours.

Although this was approved as a valid safety case, the layout normally used in the UK railways contain more detail and supporting evidence for the risks.

B. Appendix B – UCL RTK Library

7.5 Settings File

Example RTK library settings file:

```
Settings.cfg - Notepad
File  Edit  Format  View  Help
//List of receiver file names, with rovers at the top of the list. Fixed co-ordinates are specified as e.g.: , X = 3976502.0422,

//Possible options after receiver name:
//RECEIVER:  NAME, filename=auto, x=3977368.470, y=6726.262, z=4969508.959
//Note that these are not required.
//Specifying co-ordinates overides the values in the RINEX header
//filename=auto is used to allow the automatic generation of RINEX file names based upon the receiver name specified and the date and time of the first observation in the rover file. Th
//If filename=auto is not present then the rover RINEX obs file will be assumed to be called [NAME].txt


RECEIVER:  all_2_11_07
//RECEIVER:  all_3_11_07
//RECEIVER:  all_4_11_07
//RECEIVER:  all_5_11_07
//RECEIVER:  all_6_11_07
//RECEIVER:  all_7_11_07
//RECEIVER:  all_8_11_07
//RECEIVER:  all_9_11_07
//RECEIVER:  all_10_11_07
//RECEIVER:  all_13_11_07
//RECEIVER:  all_14_11_07
//RECEIVER:  all_15_11_07
//RECEIVER:  all_2_11_07_short
//RECEIVER:  all_13_11_07_short

//RECEIVER:  all
//RECEIVER:  drot311A
//RECEIVER:  lich311A
//RECEIVER:  drot306A


NUMBER OF ROVERS:                       1                                       //The rest of the receivers are assumed to be reference stations

//Output
OUTPUT TYPE:                    Overwrite                               //"None", "Overwrite", "Unique"

//Orbits
ORBIT TYPE:                     SP3                                     //"SP3", "BRDC"

//GPS observations to use
GPS CODE:                               C1C,C2P                                 //Seperate with commas and use the RINEX 3 naming convention.
GPS CODE PRECISION:             0.3,0.3  //0.06,0.15                            //metres, standard deviation. Seperate with commas. Order is same as GPS_CODE:

GPS PHASE:                              L1C  ,L2P                               //Seperate with commas and use the RINEX 3 naming convention.  "L1","L2"
GPS PHASE PRECISION:            0.003, 0.003 //0.002, 0.005                     //metres, standard deviation. Seperate with commas. Order is same as GPS_PHASE:

//Galileo observations to use
//GALILEO CODE:                         C1X,C5X                                 //Seperate with commas and use the RINEX 3 naming convention.
//GALILEO CODE PRECISION:                       0.3,0.3                         //metres, standard deviation. Seperate with commas. Order is same as GALILEO_CODE:

//GALILEO PHASE:                        L1X,L5X                                 //Seperate with commas and use the RINEX 3 naming convention.  "L1","L2"
//GALILEO PHASE PRECISION:      0.003,0.003                                     //metres, standard deviation. Seperate with commas. Order is same as GALILEO_PHASE:

AMBIGUITY VALIDATION:           ratioTest                               //"ratioTest", "marginTest", "fixedFailRateRatioTest", "AFM", "None"
AMBIGUITY RESOLUTION:           LAMBDA                          //"LAMBDA", "SubsetAR"
SUBSET AR TYPE:                         analyseAll                              //"analyseAll", "progressivelyDeeper"
SUBSET AR CRITERIA:                     ADOP                                    //"ratio", "ADOP", "SNR", "SNR/ADOP"
SUBSET MIN SUCCESS RATE:        0.7

//Reference Station Ambiguity Resolution & Atmospheric Estimation
REFERENCE STATION AMBIGUITY RESOLUTION:    singleEpoch      //"singleEpoch", "multipleEpoch"
NUMBER OF MULTIPLE EPOCHS:                  4                           //An integer number of epochs to attempt Ref Station Ambiguity Resolution
REFERENCE STATION IONOSPHERE:           true                    //"true","false"
REFERENCE STATION TROPOSPHERE:          false                   //"true","false"

//GPS observations to use for Multiple Epoch Reference Stations Ambiguity Resolution
REFERENCE STATION GPS CODE:                             C1C,C2P                         //Seperate with commas and use the RINEX 3 naming convention.
REFERENCE STATION GPS PHASE:                    L1C,L2P         //Seperate with commas and use the RINEX 3 naming convention.  "L1","L2"
//REFERENCE STATION GALILEO PHASE:              L1X,L5X                 //Seperate with commas and use the RINEX 3 naming convention.  "L1","L2"
//REFERENCE STATION GALILEO CODE:               C1X,C5X                         //Seperate with commas and use the RINEX 3 naming convention.

//Linear combination to use for the phase solution.
//Make sure that the required observations are specified above (and only these)
LINEAR COMBINATION:                     None            //"None", "ionoFree", "widelane"
PP LINEAR COMBINATION:          ionoFree        //"None", "ionoFree"

//Atmospheric Modelling
TROPOSPHERIC ZD:                        ESAZD                                   //"None", "ESAZD", "Saas"
TROPOSPHERIC MF:                        GMF                                     //"None", "GMF", "Saas", "None" assumes a value of 1.

IONOSPHERIC ZD:                         Klobuchar                               //"None", "Klobuchar", "IONEX"
IONOSPHERIC MF:                         Klobuchar                               //"None", "Klobuchar", "CODE", "JPL", "None" assumes a value of 1.

//Antenna model
ANTENNA MODEL:                  Elev                                    //"None", "Elev"

//Minimum satellite elevation (degrees)
SATELLITE ELEVATION MASK ANGLE:         10
SNR MASK VALUE L1:                                      0
SNR MASK VALUE L2:                                      0

//Stochastic Models
PP CODE VARIANCE MULTIPLIER:            1/sinSq(elev)                   //"1/sinSq(elev)", "1", "Multipath", 1/cosSq(elev), b+a/cosSq(elev)
//used in geometryBasedFloat and geometryBasedFixed
ROV CODE VARIANCE MULTIPLIER:           1                                       //"1/sinSq(elev)", "1", "Multipath", 1/cosSq(elev), b+a/cosSq(elev)
ROV PHASE VARIANCE MULTIPLIER:          1                                       //"1/sinSq(elev)", "1", "Multipath", 1/cosSq(elev), b+a/cosSq(elev)
//used in twoRefStatAmbiguityResolution
REF CODE VARIANCE MULTIPLIER:           1                                       //"1/sinSq(elev)", "1", "Multipath", 1/cosSq(elev), b+a/cosSq(elev)
REF PHASE VARIANCE MULTIPLIER:          1                                       //"1/sinSq(elev)", "1", "Multipath", 1/cosSq(elev), b+a/cosSq(el

//Integrity
HPL:                                    2.5                                                     //Horizontal protection limit, metres
ALPHA:                          1.0e-6                                                  //Probability of false alarm
BETA:                                   1.5e-4                                          //Probability of missed detection

//Projection parameters
PROJECTION TYPE:                                transverseMercator                      //"transverseMercator", (add other projections here)
PROJECTION PARAMETERS:                  nationalGrid                            //"nationalGrid", "Custom"
PROJECTION PARAMETER VALUES:            -                                       //If "Custom" above then specify the values here, if not then put "-"
//For transverse Mercator these parameters are (in this order and seperated by commas): E0,N0,F0,phi0,lambda0

//Misc
NUMBER OF EPOCHS TO PROCESS:            -1                                      //An integer number of epochs to process, or -1 to process all in file
FIRST EPOCH TO PROCESS:                 0                               //The location of the first epoch to process (0 to start at the beginning of the rover file)
MAX CONSECUTIVE ERRORS:                 100                                     //The maximum number of consecutive errors allowed before program terminates
SENSOR NETWORK:                         false                                   //"true","false"

//Sensor network
SENSOR NETWORK:                         false                                   //"true","false"

//Kalman filtering
KALMAN FILTER:                          false                                   //"true","false"
```

Figure B.1 - Example Settings file for UCL RTK library software suite