UCGE Reports
Number 20191

Department of Geomatics Engineering

Simulation and Evaluation of the Performance of the Proposed Mars Network Constellation for Positioning, Orbit Improvement, and Establishment of a Spatial Reference Frame for Mars

by

Kyle O'Keefe

April 2004

UNIVERSITY OF
CALGARY

THE UNIVERSITY OF CALGARY

Simulation and Evaluation of the Performance of the Proposed Mars Network Constellation for Positioning, Orbit Improvement, and Establishment of a Spatial Reference Frame for Mars

by

Kyle O'Keefe

A THESIS
SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF GEOMATICS ENGINEERING

CALGARY, ALBERTA

April, 2004

Preface

This is an unaltered version of the author's Doctoral of Philosophy thesis. The thesis was accepted by the Faculty of Graduate Studies in April, 2004. The faculty supervisor for this work was Dr. Gérard Lachapelle. The co-supervisor was Dr. Susan Skone. Other members of the examining committee were Dr. M. Elizabeth Cannon, Dr. Patrick Wu, Dr. Wayne Cannon.

Abstract

In 1999, the NASA Jet Propulsion Lab presented a proposal for a six satellite navigation and communication network for Mars called the Mars Network. This thesis investigates the performance of the Mars Network both theoretically, using figures of merit commonly applied to satellite navigation systems on Earth, and in the position domain using simulated observations.

The Mars Network is evaluated in terms of availability, accuracy, and reliability as a function of position and time by simulating network geometry for users distributed across the planet. The Network is found to provide the best service to users in equatorial and polar regions. Instantaneous positioning is limited due to the small number of satellites in the constellation. The addition of a height constraint is shown to increase the availability of instantaneous positioning.

End-to-end simulation tools are developed for the simulation of Mars Network observations. The trajectories of the satellites are precisely modelled using numerical methods. Models are developed and implemented for the most significant error sources, including the effects on the navigation signals of the Mars ionosphere and troposphere. After the effect of orbital errors, the ionospheric effect is found to be the next largest error source.

A positioning and orbit determination algorithm is developed based on a decentralized processing strategy that only requires network elements to exchange state vectors and covariance matrices while making observations of each other. The algorithm is tested in several scenarios using a simulated range and Doppler observations between the six Mars Network satellites and eight simulated landers.

The ability of the Mars Network to position landers with sparse observations is demonstrated. The Mars Network is shown to be able to position landers on Mars to accuracies of 10 m after several hours of intermittent tracking. The limiting factor is found to be the growth of orbital errors. The ability of current and future lander missions to provide ground control for improving orbits and to improve lander positioning performance is demonstrated. The effect of inter-satellite observations is investigated.

Recommendations are given for ways to improve the navigation performance of the constellation and areas of future research are discussed.

Acknowledgements

I would like thank my supervisor, Dr. Gérard Lachapelle, and my co-supervisor Dr. Susan Skone, for their encouragement, insight, guidance, and patience. I would also like to thank several of my colleagues for their support and assistance. Mark Petovello, Glenn MacGougan, and Paul Alves in particular have helped with every area of my work over the last four years.

Throughout my studies I have been lucky to have had the unconditional support of four generations of my family. My grandparents, both teachers, have always inspired me to pursue my academic goals. I am also very grateful for their financial support over the years. My parents have always encouraged my curiosity and I am particularly lucky to have a mother who, beginning when I was about eight, decided I should learn proper grammar and syntax in English, Logo and BASIC. She taught me C when I was a bit older.

I thank my wife Kim for her support, encouragement, and analytical mind. I would never have made it through my first degree without her tutoring in vector calculus and differential equations. Finally, my daughter Amelia. At 21 months, her ability to play and learn is an inspiration. Her favorite shape is not a square or circle, but an ellipse. I look forward to discussing the algebra behind the picture book with her in 15 years or so.

Several agencies have generously supported my graduate studies. The Natural Science Research Council of Canada (NSERC), the Alberta Informatics Circle of Research Excellence (iCORE), the Killam Trusts, and the University of Calgary are gratefully acknowledged for their support.

Table of Contents

List of Tables

List of Figures

List of Symbols, Abbreviations, Nomenclature

Symbols

a	Semi-major axis
b	Semi-minor axis
$\boldsymbol{A}$	Design Matrix
c	Speed of Light
C_{nm}, $\bar{C}_{nm}$	Cosine Spherical Harmonic Coefficient
$\boldsymbol{C}_\bullet$	Covariance Matrix of $\bullet$
e	Eccentricity
f	Frequency
$f(x)$	Function of x
$\boldsymbol{F}$	Force Vector
$\boldsymbol{F}$	Dynamics Matrix
G	Gravitational Constant
h	Step size
H_0,H_a	Null Hypothesis, Alternate Hypothesis
i	Inclination
$\boldsymbol{I}$	Identity Matrix
$\boldsymbol{k}$	Perturbing Acceleration
$\boldsymbol{K}$	Gain Matrix
$\boldsymbol{\ell}$	Observation Vector
m	Order (of Spherical Harmonic), mass
M	Mass (of a planet or moon), Mean Anomaly
n	Degree (of Spherical Harmonic), mean motion
N	Integrated Doppler Count
$\boldsymbol{N}$	Nutation Matrix
P_{nm}	Legendre Polynomial
$\boldsymbol{P}$	Precession Matrix
$\boldsymbol{Q}$	Spectral Density Matrix
$\boldsymbol{Q}_{k,k-1}$	Process Noise Matrix
$\boldsymbol{Q}_\bullet$	Cofactor Matrix of $\bullet$
r	Radius, radial distance
$\boldsymbol{r}$	Position Vector
$\ddot{\boldsymbol{r}}$	Acceleration Vector
R	Reference Radius, Redundancy Number
$\boldsymbol{R}$	Redundancy Matrix
R_1, R_2, R_3	Rotation Matrices

S_{nm}, $\bar{S}_{nm}$	Sine Spherical Harmonic Coefficient
t	Time
$\boldsymbol{v}$	Residual Vector
V_n^2	Degree Variance
$\boldsymbol{w}$	Misclosure Vector
$\boldsymbol{x}$,$\hat{\boldsymbol{x}}$	State Vector, Estimated State Vector
α	Probability of Type I Error
β	Probability of Type II Error
δ_0	Non-centrality Parameter
$\boldsymbol{\delta x}$	Error State Vector
$\boldsymbol{\Delta x}$	Error in $\boldsymbol{x}$
$\Delta\epsilon$	Nutation in Obliquity
$\Delta\psi$	Nutation in Longitude
$\boldsymbol{\Theta}$	(Planetary) Rotation Matrix
λ	Longitude
$\boldsymbol{\nu}$	Innovation Sequence
$\boldsymbol{\Pi}$	Polar Motion Matrix
ρ,$\dot{\rho}$	Range, Range Rate
σ, σ^2	Standard Deviation, Variance
$\boldsymbol{\Phi}$	Transition Matrix
τ	Uplink or downlink time
ϕ	Latitude
ω	Argument of Perigee
Ω_0	Right Accession of Accending Node
$\dot{\Omega}$	Drift of Right Accession of Accending Node
∂	Partial Differential
∇	Marginally Detectable Blunder
$\bullet^{(-)}$	Predicted Value
$\bullet^{(+)}$	Updated Value

Abbreviations and Acronyms

AAST	Airy Apparent Sidereal Time
AMST	Airy Mean Sidereal Time
AU	Astronomical Unit
CNES	Centre National d'Etudes Spatiales (French Space Agency)
CDMA	Code Division Multiple Access
CODE	Centre for Orbit Determination in Europe
DE	Adams-Bashford-Moulton Multistep Method

DOP (PDOP,GDOP,..)	Dilution of Precision (Position DOP, Geometrical DOP,...)
DSN	Deep Space Network
ECI	Earth-Centred Inertial
FDMA	Frequency Division Multiple Access
GMM2B	Goddard Mars Model 2B
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
iCORE	Alberta Informatics Circle of Research Excellence
INS	Inertial Navigation System
J2000	Standard Julian Epoch of 2000
JD	Julian Day Number
JPL	Jet Propulsion Laboratory
LMO	Low Mars Orbiting (satellite)
kbps	thousand bits per second
MDB	Marginally Detectable Blunder
MCI	Mars-Centred Inertial
MCMF	Mars-Centred Mars-Fixed
mas	milli-arcsecond
MER1, MER2	Mars Exploration Rovers 1 and 2
MJD	Modified Julian Day Number
MOLA	Mars Orbiting Laser Altimeter
MRSE	Mean Radial Spherical Error
MRT	Mean Response Time
MSD	Mars Sol Day Number
NASA	National Aeronautics and Space Administration
NEIGE	Netlanders Ionosphere and Geodesy Experiment
NSERC	Natural Sciences and Engineering Research Council
PE	Position Error
POD	Precise Orbit Determination
ppm	Parts per million
PRN	Pseudo-random Noise
RK4	Fourth Order Runge-Kutta
RMAN99	Roosbeek Martian Analytical Nutations 1999
TEC	Total Electron Content
UERE	User Equivalent Range Error
UHF	Ultra High Frequency

Nomenclature

AU	149 597 870.61 km
day	86400 SI seconds
year	365.25 days (1 julian year)
century	36525 days (1 julian century)
sol	Martian mean solar day: 24h 39m 35.244s
Mars Sol Day	Number of sols since noon December 29, 1873
Julian Day Number	Number of days since noon January 1, 4713 BCE
Modified Julian Day Number	JD - 2400000.5
J2000 Epoch	Julian Day Number 2451545.0

Language

In Mars related literature, there is a tendency among some authors to use Mars specific language. The most common examples of this involve the replacement of the prefix *geo* with its Martian analogue *areo*, for example Mars geology is called areology, the Mars geoid is named the areoid, and satellites in a Mars geostationary orbit are called areostationary. A more extreme example of this practice is found in the Mars Exploration Rover press releases currently being issued by NASA. For example, “(Spirit)...finished observations of yestersol’s rock abrasion tool depression” (JPL, 2004b). This type of usage seems to be more for fun than anything, as there is rarely any ambiguity when words like *today* or *geodesy* are used in the context of another planet. In this thesis, Mars specific usage is avoided where possible. The word geology means Mars geology and perigee means closest approach to any planet. Two exceptions are the word sol, which is used to specify a Mars mean solar day, and the word areostationary, which is used because it is mentioned in several important references. In the future, when the field of *Areomatics Engineering* is more mature, this thesis will probably be criticized for its non-inclusive geocentric language.

Chapter 1

Introduction

In the past ten years, satellite based positioning and navigation has developed from an experimental science to a mature technology that is rapidly becoming a global utility. The development of Global Navigation Satellite Systems (GNSS) and wireless communications systems have made instantaneous positioning, navigation and communication available at all times almost everywhere on Earth. However, beyond our planet, navigation and communication are both still slow and error prone procedures.

Since the mid 1990's there has been a renewed interest in the exploration of the solar system, particularly the most Earth-like planet, Mars. With this renewed interest in exploration, there is a need for more reliable communication and navigation to, and on, Mars. In the past, all navigation and communication has been accomplished through Earth-based radio transceivers limiting both the communication bandwidth and the navigation accuracy. To improve this situation, the NASA Jet Propulsion Laboratory (JPL) has proposed the development of a communication and navigation satellite system for Mars called the Mars Network.

The proposed Mars Network is still in the conceptual stage consisting only of a baseline satellite constellation design and performance specifications. The purpose of this thesis is to simulate the Mars Network and assess its performance for navigation and positioning on Mars, and to propose methods for it to become a self sufficient GNSS for Mars.

1.1 Background and Objectives

Though Mars is our second nearest neighbouring planet, traveling there is both expensive and dangerous. There have been many successful missions to the red planet, but many more have failed in launch, in transit, and upon arrival.

To address these and other problems, proponents of Mars exploration have begun to speak of establishing a permanent presence, or infrastructure for the exploration of the planet. Possibilities for this infrastructure include the deployment of Mars orbiting communication relay satellites and permanent unmanned ground stations. In 1999, JPL proposed the deployment of a navigation and communication satellite constellation for Mars to be called the Mars Network (Hastrup et al., 1999).

The primary mission of these satellites would be to provide a reliable communication link between future lander missions and Earth. The existence of such a service would make it unnecessary for future missions to be equipped with long range radios for Mars to Earth communication since all communications could be relayed through the Mars Network satellites. These satellites could also be used to provide location and navigation information by making range and Doppler observations of the communication signals.

In designing the Mars Network, JPL considered several constellations consisting of various combinations of Low Mars Orbiting satellites (LMOs) and areostationary (Mars geostationary) satellites. In 1999 JPL presented a series of papers describing their choice for the Mars Network constellation (Cesarone et al., 1999; Bell & Ely, 1999). The final proposed network consists of six microsatellites in low (800 km altitude) circular orbits around Mars. Four of the satellites are in highly inclined retrograde orbits while the remaining two are in almost equatorial orbits. Table

2.4 in Chapter 2 contains the initial orbital elements for the constellation. This constellation was designed to maximize its usefulness both for communication and navigation purposes primarily in equatorial regions of the planet. This was done because most of the planned and proposed missions to Mars intend to land at low latitudes.

While satellite navigation on and around Mars can be modelled on satellite navigation on Earth, there are major differences that need to be assessed and verified in detail. First, the Mars Network satellites will be transceivers, capable of broadcasting and receiving, unlike the Global Positioning System (GPS) where the satellites are beacons and the users are passive listeners. Secondly, GPS satellites are in high, nearly circular orbits that are only minorly affected by higher order terms in Earth's gravity field. In the planned Mars Network, the satellites are in low orbits which are intrinsically more affected by higher order gravity effects. This problem is compounded by the fact that Mars has relatively larger higher order terms in its gravity field. Thirdly, there is the issue of ground control. On Earth, GPS and other GNSS satellites are tracked from ground stations with precise coordinates determined by other means. On Mars, any ground observers, at least at first, will have to position themselves using the satellites, even if the satellite positions are not yet well known.

The research reported in this thesis has the following four objectives.

1. To develop end-to-end simulation tools for the Mars Network to allow for study of the system before its deployment.
2. To evaluate the performance of the proposed Mars Network constellation for navigation in terms of availability, accuracy and reliability.
3. To develop and test algorithms for user navigation and network satellite orbit

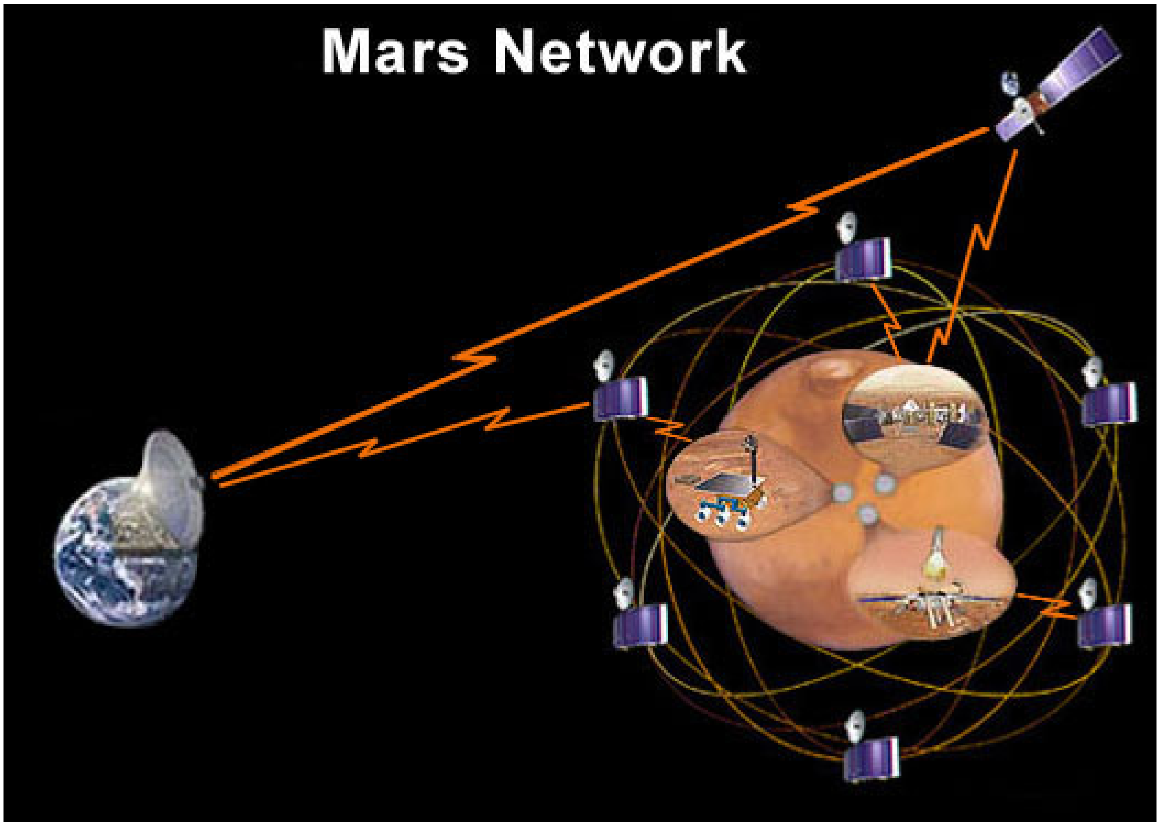

Figure 1.1: Artist's concept of the JPL Mars Network. Courtesy NASA/JPL-Caltech (from JPL (2000)).

determination on and around Mars that can operate with and independently of control from Earth.

4. To evaluate the need for ground control on Mars, and investigate the use of future lander missions to fulfill this role.

1.2 Contributions of this Thesis

The navigation performance of the Mars Network constellation proposal is assessed using measures commonly applied to Earth-based navigation systems such as availability, dilution of precision, and statistical reliability.

Previous work considering the Mars Network considered only statistical orbital errors and ignored all other errors. This thesis presents the first attempt to simulate navigation observations on Mars using a Mars-based GNSS, taking into account all of the significant error sources that affect GNSS signals. In particular, precisely determined orbital errors are generated by force modelling and realistic atmospheric error models are implemented using our limited knowledge of the Mars atmosphere.

The ability of static surface users to position themselves is determined in the position domain using simulated observations.

The usefulness of Netlanders, a proposed network of science landers, to be used as a ground control network for the Mars network is considered. Several algorithms for orbit improvement using observations from landers are presented and evaluated.

The effect of inter-satellite observations is also investigated through simulation.

1.3 Outline

In Chapter 2 of this thesis, there is a brief overview of Mars related science followed by a discussion of the Mars Network proposal. Relevant mathematical methods are discussed in Chapter 3.

The research components of this thesis are presented in three chapters. In Chapter 4, the navigation performance of the Mars Network constellation is evaluated from a user perspective based on the assumption of known satellite orbits.

In Chapter 5, a more detailed simulation of the Mars network is presented with the goal of generating simulated observations. The satellite and user trajectories are modelled numerically, observation error models are developed, and these models are applied to create a set of simulated observations.

The observations generated in Chapter 5 are used in Chapter 6 to evaluate positioning and orbit improvement algorithms in the position domain for several different scenarios. Conclusions and recommendations are discussed in Chapter 7.

Chapter 2

Mars and the Mars Network

This chapter contains an introduction to Mars and the Mars Network satellite constellation proposal. It begins with a brief summary of the history of Mars exploration followed by an overview of Mars planetary science, including Mars geology, geophysics, and atmospheric science. Some important elements of Mars geodesy are then presented followed by a detailed description of the Mars Network satellite constellation proposal, including a review of all relevant publications related to the project. The chapter concludes with a comparison of the Mars Network to other global satellite navigation systems.

2.1 Review of Mars Exploration

The planet Mars has been known and studied since antiquity. During the Renaissance, its pronounced retrograde motion was essential evidence for Copernicus' heliocentric model of the solar system and its relatively elliptical orbit lead Kepler to reject Ptolomy's concentric sphere model in favour of elliptical orbits for the planets.

It was not possible to observe the planet in detail until the development of large telescopes in the late 19th century. Even then, Earth-based observations consisted only of coloured regions on the surface and resulted in scientific interpretations that now resemble science fiction. Dust storms were interpreted as changes in surface vegetation, reports of channels were mistranslated from Italian as canals, and theories about advanced, but doomed, civilizations became popular. The public's fascination with the red planet, particularly advanced life on the red planet, continued into the mid 20th Century.

With the beginning of the space age, it finally became possible to travel to Mars and get a close up view. The first attempted Mars mission was Mars1960A, a Soviet fly by attempt that failed to launch in 1960. To date, there have been 39 attempted missions to Mars. Of these only twelve can be described as successful. There have been seven successful past missions, and five missions are currently in progress. A complete chronology of Mars missions has been compiled by Williams (2001). The successful past missions, and currently operating missions are summarized below. At the end of this section there is a discussion of some planned and proposed future Mars missions.

2.1.1 Successful Past Missions

Mariner

In the 1960s, the United States carried out an ambitious campaign of inner solar system exploration using Mariner probes. Four Mariner probes succeeded in returning data from Mars. On July 14, 1965, Mariner 4 flew by Mars returning television images of the surface. Mariner 6 and 7 performed fly bys on July 31 and August 5, 1969. Mariner 9 was the first artificial satellite of Mars, orbiting and photographing the planet between 1971 and 1972. It successfully photographed 100% of the surface of the planet revealing its volcanos and the Valles Marineris canyon system.

Viking

The two Viking missions were launched in 1975 and arrived at Mars in 1976. Viking 1 and 2 each consisted of an orbiter and a lander and were the first successful safe landings on another planet. The orbiters imaged the surface until 1980 and 1978 respectively while the landers conducted imaging, biology, and meteorology experiments and continued to operate until 1982 and 1980 respectively. The main scientific results of the Viking Missions are reported in the September 30, 1977 issue of the Journal of Geophysical Research (Vol. 82, No. 28).

Mars Pathfinder

Twenty years after the Viking missions, NASA successfully landed the Mars Pathfinder Probe on July 4, 1997. It consisted of a landing platform and the Sojourner rover. Pathfinder was meant mainly to demonstrate landing and rover technology that would be used on future missions, but it also returned a large amount of useful scientific data, including meteorological and geological data about the landing site. Results of Earth-based tracking of the lander allowed for improved Mars rotation and precession models which are discussed in Section 2.3.

2.1.2 Present Missions

Mars Global Surveyor

Mars Global Surveyor (MGS) is perhaps the most successful Mars exploration mission to date. The orbiter arrived on September 12, 1997. It carries a high resolution camera, a laser altimeter, a thermal emission spectrometer, a magnetometer, and a communications relay system. MGS conducted its primary mapping mission between 1999 and 2001, mapping the entire surface of the planet from a polar orbit. Scientific results from MGS include detailed topography derived from both laser altimetry and photogrammetry, the discovery of local crustal magnetic fields, atmospheric profiles from radio occultation observations, and gravity field determinations from orbital tracking. MGS is still operating and continues to capture images of Mars and has been used as a communications relay with the Mars Exploration Rovers.

2001 Mars Odyssey

2001 Mars Odyssey is NASA's most recent science orbiter. It is currently in a 400 km polar orbit making observations with several different imaging systems. Its primary missions are to search for water and ice beneath the surface of Mars and to evaluate the radiation environment of the planet. It is also being used as a communications relay with the Mars Exploration Rovers. It is expected to remain in operation until 2005.

Mars Express

Mars Express is a European science orbiter that arrived at Mars on December 25, 2003. Like Global Surveyor and Odyssey, its primary mission is to map the surface and atmosphere of Mars and continue the search for water and evidence of life. Mars Express also carried the British Beagle II lander that has not yet made contact with Mars Express and was probably lost during the landing.

Mars Exploration Rovers

NASA's most recent Mars missions, the twin Mars Exploration Rovers, arrived at Mars in January 2004. The first rover, Spirit, successfully landed on January 4, 2004, and Opportunity arrived on January 25, 2004. Though neither rover has travelled very far from its landing site, the MER rovers' ability to travel and make geological observations is a huge technological advance over previous lander missions. The major scientific results of the MER project will likely be published in late 2004.

2.1.3 Future Missions

Mars Reconnaissance Orbiter

Mars Reconnaissance Orbiter is the next NASA Mars science orbiter mission, scheduled for launch in 2005. It will continue the search for water using remote sensing techniques and will have a considerably higher data return capability than any previous Mars science orbiter.

Phoenix

The Mars Phoenix polar lander mission will be launched in 2007. It will consist of a small lander that will dig into the northern polar cap to search for evidence of life. The spacecraft is the identical twin of the Mars Polar Lander that failed to safely land in 1999.

Mars Science Laboratory

Mars Science Laboratory is the next NASA lander/rover mission, planned for launch in 2009. It will likely consist of a larger and longer lasting rover that will carry out experiments similar to the Mars Exploration Rovers.

Netlanders

Netlanders is a European mission concept that plans to deploy a network of four identical landers on the surface of Mars. Each lander will be equipped with seven different science payloads including a weather station, an electric field sensor, a magnetometer, a ground penetrating radar system, a stereoscopic multispectral camera, a seismometer, and a UHF transceiver capable of making range and Doppler measurements (Dehant et al., 1994; Barriot et al., 2001). The purpose of this mission is to accomplish network science, meaning scientific observations that benefit from observations being made simultaneously at multiple locations. This mission will be discussed in Chapter 6 as a possible network of ground control stations for the Mars Network.

Premier and Marconi

Two European orbiter missions are currently in development. CNES, the French space agency, is developing the Premier program, which proposes to send a series of orbiters that would become the return vehicles for sample return missions. It was also proposed that the first Premier orbiter, originally scheduled for launch in 2007, would carry the Netlanders mission. The Italian Space Agency also had plans to send a dedicated Mars communications satellite, named Marconi, during the 2007 launch opportunity. Unfortunately, both of these missions have recently been subject to budget cuts and may be delayed or cancelled entirely (Moomaw, 2002).

Sample Return

Though no specific sample return missions are currently being developed, several have been proposed over the last decade and JPL states that the first sample return mission will be launched no earlier than 2014 (JPL, 2004a). The goal of these mission concepts is to return a small piece of Mars for study on Earth. This type of mission presents several challenges, including rendez-vous of the sample return canister with a vessel that would return it to Earth. This type of mission will not likely be attempted until the Mars Network, or some other positioning infrastructure is deployed.

Figure 2.1 is a map of the locations of past, present and future lander missions.

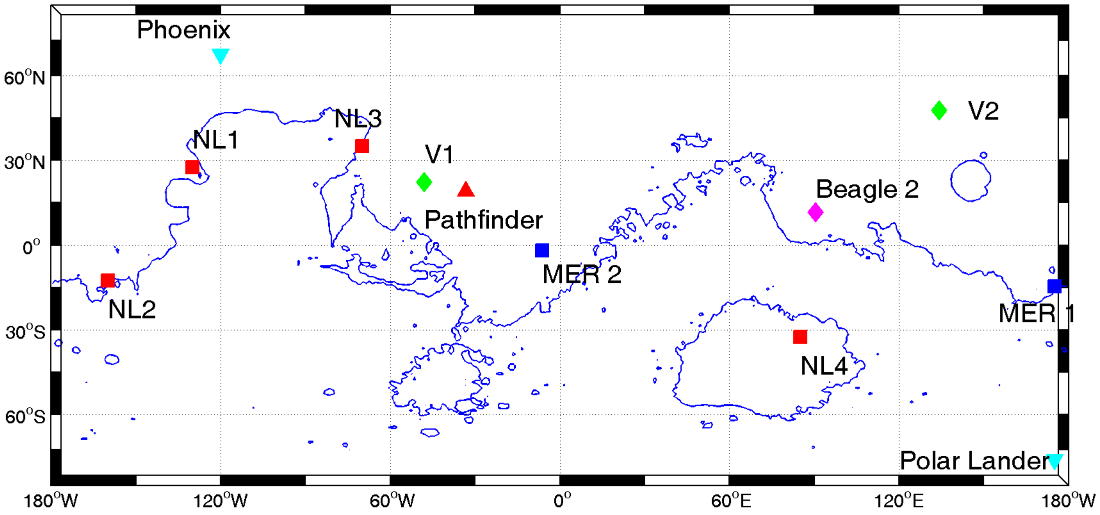

Figure 2.1: Map of Mars showing the locations of past, present, and future lander missions. The contour lines indicate zero elevation, with the northern hemisphere being generally below zero elevation and the southern hemisphere above. Viking 1 and 2 are labelled V1 and V2, the Mars Exploration Rovers are labelled MER1 and MER2, and the four Netlanders are labelled NL1-NL4. The target landing sites of two failed lander missions are also indicated: Beagle 2 (2004) and Polar Lander (1999). Polar Lander will be used in a simulation scenario in Chapter 6.

2.2 Overview of Mars Planetary Science

Mars is the fourth planet in the solar system. It is a small rocky planet, with a thin atmosphere, and a cold surface. Despite these attributes, Mars is the most Earth-like planet in the solar system. Its most Earth like features are its surface, which consists of mountains, plains and valleys, its near 24 hour rotation period, and its axial tilt and resulting seasons. Because of these features, and its relative proximity to Earth, Mars has been the subject of more planetary science investigations than any other planet other than Earth. This section will review the planetary science of Mars and present some comparisons to Earth. More detailed reviews of Mars planetary science are given in Lodders & Fegley Jr. (1998) and Jones (1999). Some physical properties of Mars, as reported by Lodders & Fegley Jr. (1998), King (2001) and Duxbury et al. (2002), are listed in Table 2.1, along with corresponding values for Earth.

Table 2.1: Physical properties of Mars and Earth (from Lodders & Fegley Jr. (1998),King (2001), and Duxbury et al. (2002)).

Property	Mars Value	Earth Value
Mean Radius	3389500 m	6371010 m
Equatorial Radius	3396190 m	6378136 m
Polar Radius	3376200 m	6356753 m
Mean Density	3933.5 kg m^{-2}	5.515 kg m^{-2}
Surface Gravity	3.69 ms^{-2}	9.78 ms^{-2}
Tropical Orbit Period	686.973 days	365.242 days
Mean Orbital Radius	1.52 AU	1.0 AU
Inclination of Equator to Orbit	25.189°	23.45°
Length of Day	24.6597 hours	24.0000 hours
Mean Surface Pressure	6.363 HPa	1013 HPa
Mean Surface Temperature	214 K	288 K

2.2.1 Geology and Geophysics

Mars, like Earth, is thought to have formed approximately 4.5 billion years ago. The surface of Mars can be divided into two regions, the lightly cratered northern plains and the heavily cratered southern highlands. Figure 2.2 is a low resolution topographic map of the planet showing the main surface features. The southern highlands appear to be much older than the northern planes. Local crustal magnetization was recently discovered in the south, suggesting that Mars had a planetary magnetic field when the southern highlands were formed that was no longer present when the northern hemisphere was resurfaced. Prominent surface features include two large impact basins in the southern hemisphere, Hellas and Argyre, as well as the Tharsis bulge, which includes several large shield volcanos. Olympus Mons, just North of Tharsis, is the largest volcano in the solar system, rising 26 km above the surrounding planes. East of Tharsis is the Valles Marinaris canyon system.

The composition of Mars has been inferred from the study of meteorites believed to originate from Mars and from models of the formation of the solar system. Similar to Earth, it is believed that Mars has a iron core surrounded by a silicate mantle and crust. Since Mars lacks a planetary dipole magnetic field, it has been suggested that either Mars has a completely sold core, or it has a liquid core without a solid inner core.

2.2.2 Phobos and Deimos

Mars has two moons that were discovered in 1877, named Phobos and Deimos. Both are small, highly irregularly shaped objects. Phobos has a mean radius of 11.1 km, while Deimos has a radius of 6.1 km. Both moons are in nearly circular prograde equatorial orbits. The orbital period of Phobos is only 0.32 days long, meaning that

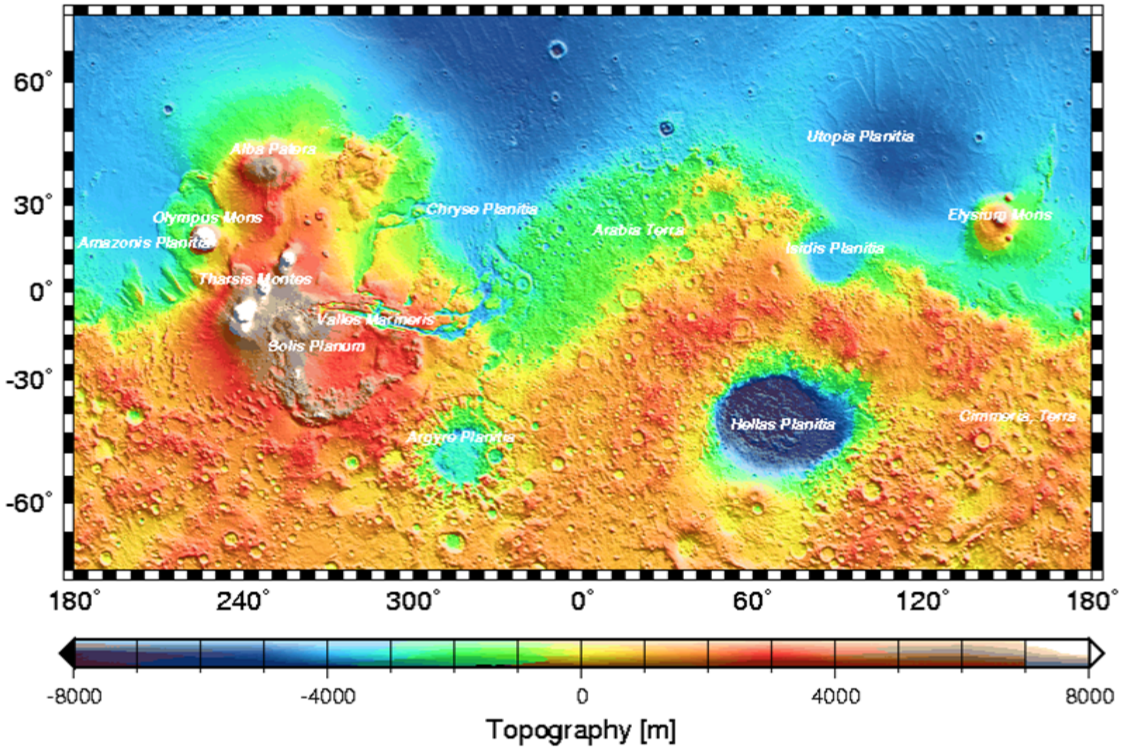

Figure 2.2: Labeled topographic map of Mars generated by the Mars Orbiting Laser Altimeter (MOLA) (from MOLA Science Team (2004)).

it rises in the west and sets east, while Deimos is in a higher 1.26 day orbit. Due to tidal friction, Phobos' orbit is decaying and it will likely impact Mars in approximately 100 million years (Lodders & Fegley Jr., 1998).

2.2.3 Atmospheric Science

Our knowledge of Mars' atmosphere is derived from three sources: Earth-based and Mars orbiter observations, refractivity profiles obtained from radio-occultation measurements made of Mars orbiters, and observations made by the Viking and Pathfinder landers.

Unlike Earth's relatively dense nitrogen and oxygen atmosphere, the atmosphere of Mars is very thin and composed almost entirely of carbon dioxide. The composition by volume of the atmospheres of Mars and Earth are shown in Table 2.2.

The surface pressure at mean radius (the Mars analogue to sea level) is 6.36 mb though it can vary seasonally between 4.0 and 8.7 mb (King, 2001). The surface density is 0.020 kgm^{-3} and the mean molecular weight is 43.34 g/mole.

Table 2.2: Atmospheric composition by volume of the atmospheres of Mars and Earth (from King (2001)).

Gas	Mars	Earth
Carbon Dioxide (CO_2)	95.32%	0.035%
Nitrogen (N_2)	2.7%	78.084%
Argon (Ar)	1.6%	0.9340%
Oxygen (O_2)	0.13%	20.946%
Water Vapour (H_2O)	0.021%	$\approx 1\%$

Jones (1999) divides the Mars atmosphere into three layers: the lower atmosphere, below 45 km; the mesosphere, up to 110 km; and the exosphere above. The lower atmosphere is characterized to the presence of thin CO_2 ice, H_20 ice, and dust clouds.

Several standard temperature and pressure models for Mars' atmosphere exist and are discussed in more detail in Sections 5.1.4 and 5.4.2 where the effects of the atmosphere on satellite orbital motion and navigation signals are discussed.

Ionosphere

Unlike Earth's four layer ionosphere, the Mars ionosphere consists of a single layer extending from about 100 km altitude to several hundred km (Ho et al., 2002). The dayside Mars ionosphere is generally modelled as an ideal Chapman layer with a peak electron density around an altitude of 130 km. There have been very few observations of the nightside Mars ionosphere, as the electron density on the nightside has generally been too low to detect using radio occultation observations. An excellent review of Mars' ionosphere prior to Mars Global Surveyor results is presented in Shinagawa (2000) while Ho et al. (2002) discuss MGS results as well.

One unique feature of the Mars ionosphere is its interaction with local crustal magnetic fields in the southern hemisphere. While Mars has no planetary magnetic field, local crustal magnetization was detected by Mars Global Surveyor (Acuna et al., 1999). Crider et al. (2001) report increased ionopause heights above regions with local crustal magnetic fields.

Study of the effect of the Mars ionosphere on radio signals has generally been limited to estimating ionospheric effects on communication signals. Ho et al. (2002) present a detailed ionospheric model for the planet, but then only discuss its application to ground-to-ground communications. Mendillo et al. (2003) are the only authors to address the effect of the Mars ionosphere on potential satellite navigation signals. They conclude that the Mars ionosphere would have very little effect on GPS frequency signals, but that the impact of the ionosphere would be much greater at lower frequencies. They further propose that dual-frequency measurements would be an excellent method to observe the structure of Mars' ionosphere. A similar scheme, using dual frequency UHF and S-band range and Doppler observations has been proposed as a component of the Netlanders mission (Barriot et al., 2001).

The ionospheric models presented in Mendillo et al. (2003) and Ho et al. (2002) are discussed in further detail in Chapter 5 where they are used to simulate the ionospheric effect on navigation signals.

2.3 Martian Geodesy

Mars geodesy is concerned primarily with two tasks: the establishment of a spatial reference system for the planet and the determination of the Mars gravity field. In this section, the current state of Mars geodesy will be discussed and conventions will be established that will be used throughout the rest of this thesis.

2.3.1 Mars Reference Systems

Analogous to Earth, the Mars-Centred Mars-Fixed (MCMF) and Mars-Centred Inertial (MCI) reference systems can be defined by the position of the Mars centre of mass (origin), the orientation of the Mars rotation axis (z-axis), and an arbitrary conventional direction (x-axis). For the MCMF system, the x-axis is in the direction of the prime meridian of Mars. This was initially defined by a dark spot on the planet known as Meridiani Sinus that could be easily observed by Earth-based telescopes. It is now defined as the meridian running through the centre of a small crater called Airy-0 (Duxbury et al., 2002). The orientation of the MCI system can be defined analogously to the Earth Centred Inertial (ECI) system by making the x-axis point in the direction of the Mars Vernal equinox. Some authors do not define an MCI system, and prefer to use the ECI system for Mars by defining the orientation of the of the Mars pole vector in ECI coordinates (Folkner et al., 1997; Duxbury et al., 2002). In this thesis, the MCI and MCMF systems are used because they are analogous to their Earth counterparts. The transformation between MCMF and MCI system is shown in equation 2.1 and depends on four Mars rotation matrices, $\mathbf{\Pi}(t)$, $\mathbf{\Theta}(t)$, $\boldsymbol{N}(t)$, and $\boldsymbol{P}(t)$ that account for polar motion, rotation, nutation, and precession respectively.

$$\boldsymbol{r}_{\mathrm{MCMF}} = \mathbf{\Pi}(t)\mathbf{\Theta}(t)\boldsymbol{N}(t)\boldsymbol{P}(t)\boldsymbol{r}_{\mathrm{MCI}} \tag{2.1}$$

This transformation is analogous to the transformation between ECEF and ECI on Earth. The effects represented by each of these rotation matrices are described below.

Rotation

The sidereal rotation period of Mars was first precisely determined by astronomers in the late 18th century to be 24 hours, 37 minutes, 9.9 seconds or 350.9425°/day (Sheehan, 1996). The most recent estimates of Mars rotation parameters were derived from Earth-based tracking of the Pathfinder lander. Folkner et al. (1997) report a rotation rate for Mars of 350.89198226°/day ±0.00000008°/day with an initial value of 133.61259° at the J2000 standard epoch.

Precession

Folkner et al. (1997) were also able to estimate the precession rate of Mars using observations of Pathfinder in 1997 in conjunction with observations of the Viking landers in between 1976 and 1982. They estimate the precession rate of the Mars pole vector to be −7576 milli-arcseconds/year (mas/year).

Nutation

While Mars nutation has yet to be observed, Roosbeek (1999) and Bouguillon & Souchay (1999) have recently published analytical nutation theories for Mars. Both theories agree with each other to within the precision of each. Roosbeek's theory, called the Roosbeek Martian Analytical Nutations 1999 (RMAN99), has been used in this thesis. Similarly to the IAU 1980 Earth Nutation Theory, RMAN99 consists of two Fourier series, for the nutation in longitude and obliquity respectively, that are functions of linear combinations of the positions of various solar system bodies. The theory consists of a 23 term series for nutation in longitude and a 10 term series for nutation in obliquity. The two series are shown in Table B.1 in Appendix B. Evaluations of the series for two years beginning at the J2000 epoch are shown in Figure 2.3.

Polar Motion

Mars' polar motion has not yet been observed (Gauchez & Souchay, 2000), however, the Netlanders Ionosphere and Geodesy Experiment (NEIGE) intends to detect it (Hoolst et al., 2000; Barriot et al., 2001). If present, it would be caused by free core nutation and by Chandler wobble. The detection and measurement of polar motion would also help settle the debate about whether Mars has a liquid core. For the purposes of this thesis, the effects of polar motion have been neglected.

2.3.2 Reference Ellipsoid and Curvilinear Coordinates

Though a reference ellipsoid is not required to define a reference system, it is useful for producing maps and for the definition of geodetic (ellipsoidal) coordinates. The process of selecting a suitable reference spheroid for Mars-based on laser altimetry observations is described in detail in Duxbury et al. (2002). In their paper, six

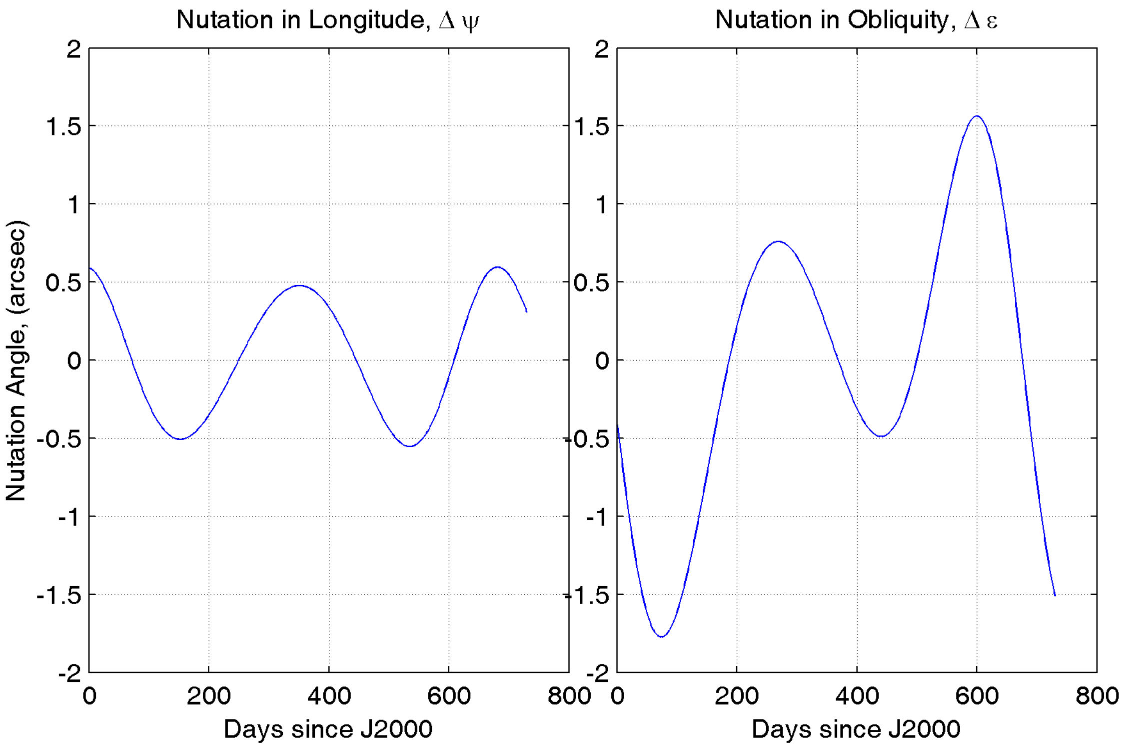

Figure 2.3: RMAN99 Mars nutation series evaluated for two years beginning at the J2000 epoch.

different reference surfaces were considered including the best fitting sphere, bi-axial ellipsoid and tri-axial ellipsoid, both centred and not centred at the centre of mass of Mars. Duxbury et al. (2002) recommend the use of the best fitting bi-axial ellipsoid, centred at the centre of mass of Mars. This is the ellipsoid that will be used in this thesis. The Mars reference ellipsoid has a equatorial radius, a, of 3396190 m ± 100 m and a polar radius, b, of 3376200 m ± 100 m. From these values, the flattening of the planet can be computed as

$$f = \frac{a - b}{a} = \frac{1}{170} \tag{2.2}$$

which is considerably greater than the 1/298 flattening of Earth (Schwarz, 1998).

Care must be used when interpreting curvilinear coordinates (latitude and longitude) on Mars. Astronomers and cartographers have traditionally used what is known as a planetographic coordinate system for mapping Mars. In this system, longitude is measured west of the prime meridian so that the longitude facing an Earth-based observer increases with time. Historically, this convention has been applied to all prograde rotating bodies in the solar system. In a planetographic system, latitude is defined analogously to geodetic latitude on Earth, i.e. the angle between the equator and the normal to the reference ellipsoid. More recently there has been a shift to using planetocentric coordinates, where longitude is measured east of the prime meridian and latitude is analogous to geocentric latitude on Earth. According to Duxbury et al. (2002), all pre-2002 maps of Mars employ planetographic coordinates, though they recommend that planetocentric coordinates be used for all future work. They also recognize that some software in use uses a mix of the two systems, for example, East longitude and geodetic latitude. In order to maintain compatibility with existing Earth-based satellite navigation software, and to maintain the analogy with Earth, this mixed system (east longitude, geodetic latitude, which Duxbury et al. (2002) refer to as East-ographic coordinates) is used throughout this thesis except when dealing with gravity field spherical harmonics, where geocentric latitude is required.

2.3.3 Gravity Model

Gravity models of Mars have been developed first by observation of Phobos and Deimos and later by radio tracking of successive Mars orbiters starting with Mariner 9 in 1971. The best current model for the gravity field of Mars is the Goddard Mars Model 2B (GMM2B) (Lemoine et al., 2001). This model was developed by the Goddard Space Flight Center using orbital data from Mars Global Surveyor. JPL simultaneously published their own model, MGS75D, based on a subset of the same observations (Yuan et al., 2001). GMM2B was used in this thesis as it was available for download online. The model consists of a reference mass and radius for the planet and a series of spherical harmonic coefficients to degree and order 80. The Mars gravitational constant (GM), reference radius, and the first few spherical harmonic coefficients of the model are shown in Table 2.3.

The gravitational potential can be evaluated at any point (r, ϕ, λ) outside the reference radius as

$$V(r, \phi, \lambda) = \frac{GM}{r} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{R^n}{r^n} P_{nm}(\sin\phi)(C_{nm}\cos(m\lambda) + S_{nm}\sin(m\lambda)) \tag{2.3}$$

where R is the reference radius, and r, ϕ, and λ are the Mars-centred Mars-fixed spherical polar coordinates of the point. C_{nm} and S_{nm} are the unnormalized spherical harmonic coefficients and are related to the normalized coefficients by

$$\begin{Bmatrix} \bar{C}_{nm} \\ \bar{S}_{nm} \end{Bmatrix} = \sqrt{\frac{(n+m)!}{(2-\delta_{0m})(2n+1)(n-m)!}} \begin{Bmatrix} C_{nm} \\ S_{nm} \end{Bmatrix} \tag{2.4}$$

where δ_{0m} is 1 when $m = 0$ and zero otherwise. One method of representing the coefficients of a gravity model and of showing their relative importance is by computing the degree variances, V_n^2, which are the quadratic means of all of the normalized spherical harmonic coefficients for a particular degree.

$$V_n^2 = \frac{1}{2n+1} \sum_{m=0}^{n} (\bar{C}_{nm}^2 + \bar{S}_{nm}^2) \tag{2.5}$$

On Earth, the square roots of the degree variances follow the relation

$$V_n \approx \frac{10^{-5}}{n^2} \tag{2.6}$$

which is known as Kaula's Rule (Montenbruck & Gill, 2000). For Mars, this relationship has been determined to be

$$V_n \approx \frac{1.3 \times 10^{-5}}{n^2} \tag{2.7}$$

Table 2.3: Parameters of Goddard Mars Model 2B (from Lemoine et al. (2001)).

Reference Radius, R	3397000 m
Gravitational Constant, GM	42828.371901 km^3s^{-2}

Low Degree Normalized Spherical Harmonic Coefficients			
Degree n	Order m	$\bar{C}_{nm}$	$\bar{S}_{nm}$
2	0	$-8.7450547081842009 \cdot 10^{-4}$	
2	1	$1.3938449166781359 \cdot 10^{-10}$	$1.7044280642328221 \cdot 10^{-10}$
2	2	$-8.4177519807822603 \cdot 10^{-5}$	$4.9605348841412452 \cdot 10^{-5}$
3	0	$-1.1886910646015641 \cdot 10^{-5}$	
3	1	$3.9053442315700724 \cdot 10^{-6}$	$2.5139324037413419 \cdot 10^{-5}$
3	2	$-1.5863411026265399 \cdot 10^{-5}$	$8.4857987158792132 \cdot 10^{-6}$
3	3	$3.5338541142774030 \cdot 10^{-5}$	$2.5113984262622799 \cdot 10^{-5}$

The degree variances of the GMM2B model as a function of degree are shown in Figure 2.4.

2.3.4 Mars Time Systems

Analogous to Earth, it is possible to define both solar and sidereal time on Mars. Solar time on Mars, as on Earth, is derived from the hour angle of the Sun. The Mars equivalent of a mean solar day is called a sol, and is 24 hours, 39 minutes, and 34.9 seconds long. Sidereal time on Mars is defined differently by different authors. In analogy to Earth, Roosbeek (2003) defines Martian Sidereal Time (or Airy Mean Sidereal Time) as the angle between the Martian vernal equinox and the Martian prime meridian. This differs from the definition given in Duxbury et al. (2002), where sidereal time is replaced by an orientation angle of the Mars prime meridian with respect to the Earth vernal equinox. Regardless of the reference orientation, the Mars sidereal rotation period is 24 hours, 37 minutes, and 22.4 seconds.

In analogy to the Julian day number (JD), Allison & McEwan (2000) have pro-

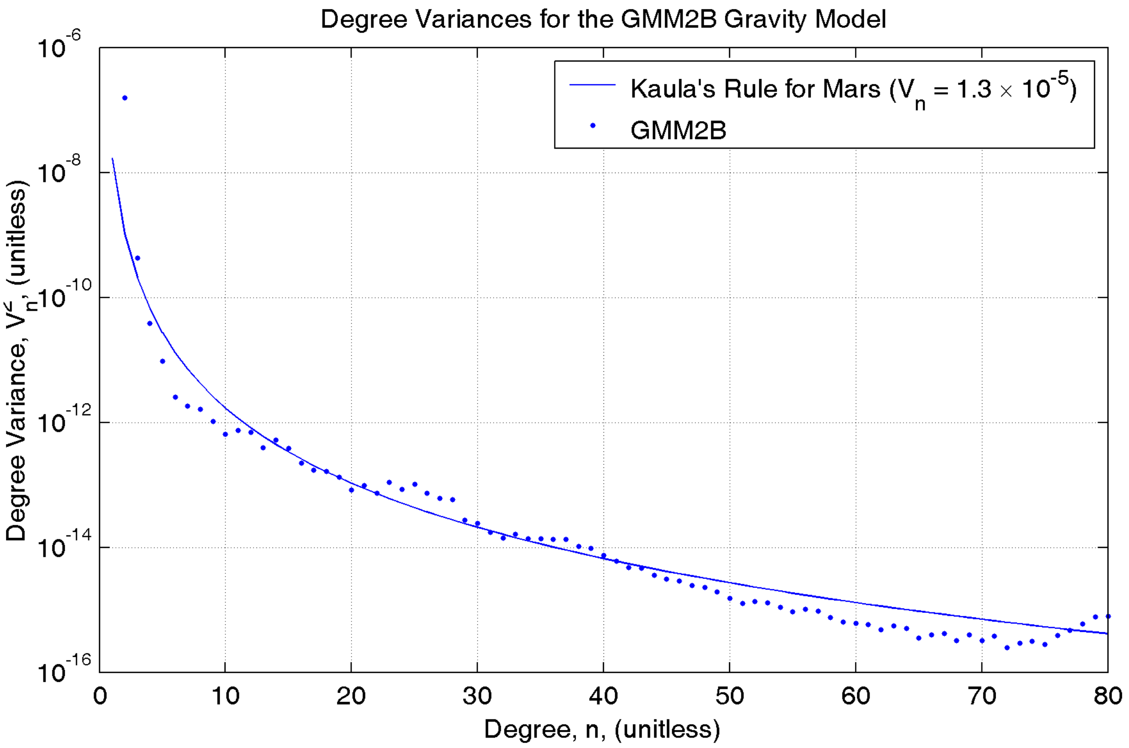

Figure 2.4: Degree variances of the Goddard Mars Model 2B compared to the Kaula rule for Mars (after Lemoine et al. (2001)).

posed the Mars Sol Day (MSD) as a method for continuous counting of sols. However, all of the models for calculating precession, nutation, and rotation above have time arguments of days or years, typically the Julian day number. MSD, and other solar timing formulas presented by Allison & McEwan (2000) are very useful for determining ellipse conditions and modelling solar radiation pressure. This is discussed in more detail in Chapter 5.

2.4 Introduction to the Mars Network Proposal

2.4.1 Literature Review

Work conducted by JPL in the design and evaluation of the Mars Network has been reported in ten conference papers and two refereed papers published between 1999 and 2003.

The fundamental design parameters of the Mars Network proposal were presented in three conference papers in August, 1999. A high level overview of the Mars communications and navigation infrastructure proposal is presented in Cesarone et al. (1999). It describes the possible constellation elements, including low orbiting communication and navigation microsatellites as well as the possibility of larger areostationary communications relay satellites. The need for this kind of orbital infrastructure is then reviewed in the context of several planned and proposed Mars missions. Cesarone et al. (1999) then discuss the philosophy behind the constellation design noting five communication and navigation performance goals:

1. High communications capacity in equatorial regions.

2. Global coverage, even in the event of the loss of one spacecraft.

3. Maximized performance at all latitudes and longitudes.
4. Minimized performance variations across all latitudes and longitudes.
5. Minimized orbital maintenance and coverage variability due to precessional effects.

Cesarone et al. (1999) also define the navigation observations available to the Mars Network satellites. They consist of 1-way and 2-way Doppler observations, measured in 1 minute samples, and 1-way and 2-way ranging. The clock available on each of the satellites is also described having a performance of 10^{-14} s/s over 60 s. This is interpreted in this thesis as an Allan standard deviation of 10^{-14} s/s for intervals of 60 s (Allan et al., 1997). This is consistent with the observed behaviour of a rubidium oscillator (Petovello & Lachapelle, 1999).

A companion paper (Ely et al., 1999), presented at the same conference describes the details of the constellation design process. Several possible constellations were considered and compared in order to select a candidate constellation that would provide optimal communication and navigation performance. Ely et al. (1999) use Mean Response Time (MRT) as a navigation figure of merit, which they define as the average time required to collect sufficient observations to compute a user's position to a prescribed accuracy. Detailed navigation assumptions are then given, including

1. 2-Way Doppler measurement uncertainty of 0.5 mm/s at 60 s (1 σ).
2. 2-Way range measurement uncertainty of 1 m (1 σ).
3. User clock fractional frequency stability of 10^{-10} s/s over 60 s.
4. Satellite clock fractional frequency stability of 10^{-14} s/s over 60 s.

5. Lander elevation mask angle of 15°.
6. User position uncertainty requirement of 10 m (1σ).
7. Orbit Errors: 2 m radial, 7 m along track, 7 m across track.
8. Atmospheric errors and other error sources are neglected.

Ely et al. (1999) also predict that the actual range noise will be 10 cm (1σ) and present an optimistic launch schedule that includes a prototype microsatellite being launched in 2003 followed by production launches in 2005, 2007 and 2009, each carrying two satellites.

Finally, Hastrup et al. (1999) describe the Mars Network satellite in detail, including the proposed receiver architecture and the physical characteristics of the satellites. They also describe several satellite-user communications scenarios.

Several other Mars Network related papers were presented in 1999 by various JPL researchers. (Matousek et al., 1999) describes the Mars Micromission Architecture. It consists of a generic mission platform that can be launched as a secondary payload on the CNES Ariane-5. The Mars Network microsatellites are a specific application of this architecture.

Hart et al. (1999) assesses the constellation presented in Ely et al. (1999) using a genetic algorithm, concluding that the existing design is optimal based on the criteria described in Hastrup et al. (1999). Wu et al. (1999) considers the effect of using the first three Mars Network satellites to position the Netlanders landers instead of using the Mars Express satellite, based on the same tracking assumptions for Netlanders that are described in Barriot et al. (2001). Bell & Ely (1999) present results largely presented in Ely et al. (1999). They also list several areas of continued work, including the assessment of autonomous navigation of the satellites, cross-link ranging between satellites and the integration of areostationary satellites.

Two papers were presented in March 2000 at the IEEE Aerospace conference. Bell et al. (2000) summarizes the results of the first three papers from 1999, while Kuo (2000) describes the data handling of the communication portion of the network. Some of the results of the above are published in Edwards et al. (2001) completing the first phase of research conducted by JPL on the Mars Network proposal.

More recent research has focused on the development of the Mars Network transceiver for use on future science orbiters while the plans for a dedicated communications and navigation constellation have been reclassified as a proposed mission as opposed to a planned mission. Ely et al. (2003) present simulated positioning results for a variety of scenarios using only the Mars Reconnaissance Orbiter, which, they report, will carry a prototype of the Mars Network transceiver. Finally, Hastrup et al. (2003), review the development of the Mars Network concept. This paper also states that the exact method of range observation has not yet been determined, even though a specification had already been set for this function.

Several other publications discuss the communications aspects of the Mars Network exclusively. In general, they discuss internet-type protocols that might be used for data transfer to and from Mars and other locations in the solar system. As these do not relate to the navigation performance of the network, they are not reviewed here.

There have been relatively few mentions of the Mars Network proposal by other (non JPL) authors. Mendillo et al. (2003) mention the proposal in a paper about the Mars ionosphere that is discussed in more detail in Section 5.4.3. The Mars Network is also mentioned by LeMaster & Rock (2002) in the context of a surface-based GPS pseudolite navigation system for Mars being developed by Stanford University. The Network is also the subject of graduate course project discussing autonomous navigation for Mars landing spacecraft (Journey-Kaler, 2002). Finally, the author of this thesis has presented preliminary results at a meeting of the Institute of Navigation. All of the results in this paper are included in greater detail in Chapters 4, 5, and 6 of this thesis.

2.4.2 Mars Network Constellation

The proposed Mars Network constellation design consists of six small satellites in 800 km altitude circular orbits. Two of the satellites are in near equatorial orbits, while the other four are in near polar orbits. All six are retrograde orbits having inclinations greater than 90 degrees. The zero epoch Keplerian elements for the constellation are shown in Table 2.4 below. JPL developed this design by balancing the communication and navigation requirements of the dual use constellation. For example, orbital radius was optimized between higher orbiting satellites that would provide better navigation services, but would be less useful for communication due to power limitations. Two sub-constellations were chosen to provide enhanced equatorial coverage (using the two equatorial satellites) while maintaining global coverage using the four near-polar satellites. This configuration also provides varied pass geometry to equatorial users that would not be available with only polar or only equatorial satellites.

2.4.3 Mars Network Transceiver and Signals

The proposed signal design is described in detail in Hastrup et al. (1999). The satellites will be equipped with long range X-band (8 GHz) radios for 2-way Doppler tracking and communication with Earth. Satellite-to-satellite and satellite-to-user

Table 2.4: Zero-epoch Keplerian elements of the Mars Network constellation. Initial values for semi-major axis, a, inclination, i, right ascension of the ascending node, Ω, and mean anomaly, M, are shown.

Satellite #	a (m)	i (°)	Ω (°)	M (°)
1	4196190	172	0	0
2	4196190	172	0	180
3	4196190	111	0	0
4	4196190	111	90	90
5	4196190	111	180	180
6	4196190	111	270	270

tracking and communication will be accomplished using a dual band Frequency Division Multiple Access (FDMA) scheme where 435 to 442 MHz and 400 to 405 MHz channels are used (to allow for simultaneous transmission and reception to and from multiple users). Each of the channels will have a bandwidth of at least 2.1 MHz. The prototype transceiver, which has been named Electra, is described in a user interface document (Matousek, 2002a) and an information sheet for Scout Mission proposals provided by JPL (Matousek, 2002b). Data will be bi-phase modulated and each satellite will be capable of generating a Pseudo-random Noise (PRN) code though this will not initially be implemented. The specified ranging precision using this kind of signal is either 10 cm or 1 m using a 10 second sample (Hastrup et al., 1999), while the range rate precision is 0.5 mm/s using a 1 minute sample (Cesarone et al., 1999).

The antenna used by the Mars Network satellites is described in Edwards et al. (2001) as being an omnidirectional (0 dB gain) antenna, but that it will be in a nadir pointing configuration. In this thesis, it is assumed that the antenna is in fact a nadir pointing hemispherical antenna, and does not receive signals from above the horizon. Edwards et al. (2001) also describe the typical power of a UHF proximity radio system as being 10 Watts. It is assumed that all satellites and landers will carry compatible Mars Network transceivers.

2.4.4 Observables

The Mars Network transceivers will be capable of making one and two-way range and Doppler observations.

Two-way Doppler

In a two-way Doppler observation, an RF signal is sent from a transceiver to a remote transponder which then rebroadcasts the signal back to the transceiver. Since it is impossible to transmit and receive on the same frequency, the transponder first coherently multiplies the received signal by a turn-around ratio $T_{1,2}$. When the signal returns to the transceiver, it will be Doppler shifted by an amount corresponding to the two way range-rate. The Doppler shift is not observed directly, instead the received signal, f_r, is mixed with the reference signal multiplied by the turnaround ratio, f_{ref}, to give a beat frequency of $f_r - f_{\text{ref}}$. The integrated Doppler count, N, can then be obtained by integrating the beat frequency, or counting the zero-crossings of the mixed signal.

$$N = \int_{t_1}^{t_2} (f_r - f_{\text{ref}}) dt \tag{2.8}$$

where t_1 and t_2 are the start and end times of the counting interval. If we assume the reference frequency is constant over the counting interval, then f_{ref} may be removed from the integral

$$N = \int_{t_1}^{t_2} f_r dt - f_{\text{ref}}(t_2 - t_1) \tag{2.9}$$

The final result of a two-way Doppler measurement is an average range rate over the observation interval which is obtained by converting the two-way integrated Doppler over the interval into a one-way range rate by dividing the Doppler count by the interval, $t_2 - t_1$, multiplying by the wavelength of the reference frequency, c/ref, and finally dividing by two. The average range rate is

$$\bar{\dot{\rho}} = -\frac{1}{2}\frac{N}{t_2 - t1}\frac{c}{f_{\mathrm{ref}}} \tag{2.10}$$

Two-way Doppler derived range rate accuracies are typically quoted in terms of mm/s after a particular tracking interval.

One-Way Doppler

One-way Doppler observations are made similarly to two-way Doppler, except that the observation is made on a one-way signal. Because in one-way mode, the transmitter and receiver are separate, both units must have stable oscillators. The receiver compares the received frequency to a local standard. In one-way mode the frequency offset between the local frequency and the received frequency must either be estimated or known beforehand.

Two-Way Range

The two way range measurement, ρ, is derived from the average of the uplink and downlink travel times, τ_u and τ_d:

$$\rho = \frac{1}{2c}(\tau_u + \tau_d) \tag{2.11}$$

where it is assumed that the turn around time is small and known. It can be shown from a first order Taylor expansion that the two-way range is almost identical to the geometric distance between the transceiver and the satellite at the time half way between the transmit and receive times.

While travel time may conceptually describe the Mars Network range observable, a method of precisely measuring this time is required. The Mars Network principal investigators have not yet specified how the range observable will be realized, though it will likely consist of a PRN sequence that can be correlated in the receiver with a replica to compute a delay much the same as C/A-code ranging is done in the GPS system.

One-Way Range

A one-way range, or pseudorange, is obtained by measuring the difference between transmit time, measured on the transmitter's clock, and received time measured on the receiver's clock. It is called a pseudorange because it is the range biased by the difference between the two clocks. In order to use one-way ranging, either the two clocks must be synchronized in advance and be stable to a required accuracy over time, or the relative clock offset must be estimated.

Combination of One-Way and Two-Way Observations

The simultaneous use of one-way and two-way range and Doppler observations between Mars Network elements (satellites and landers) will allow for direct observation of the relative clock offsets and clock drifts (frequency offsets) between elements. For example, a satellite that sends and receives a two-way range observation of a lander while simultaneously tracking a one-way range will have both a range and a pseudorange observation allowing for direct observation of the clock difference between the two elements. This capability will allow for the easy comparison of clocks and the establishment of a Mars-based time scale. Observations between Mars Network elements and Earth will allow the direct observation of the Mars time scale with respect to atomic time on Earth.

2.5 Comparison of the Mars Network with other Global Navigation Satellite Systems

The proposed Mars Network draws from experience on Earth with various past and present radio-navigation systems. In many respects, The Mars Network is similar to the U.S. Navy Satellite System, TRANSIT. TRANSIT was deployed in the early 1960s and remained operational until the end of 1996. It consisted of between three and six polar orbiting satellites transmitting continuous signals at 150 and 400 MHz modulated with an ephemeris message consisting of a set of modified Keplerian elements (Stansell, 1978). Though the Mars Network also consists of low elevation near polar orbiting satellites, and uses a 400 MHz signal, it differs from TRANSIT in two major ways. Unlike TRANSIT, the Mars Network will be capable of making two-way observations and will support range observations. However, it will not be a continuously operating system, making it impossible for a Mars Network users to observe TRANSIT-like long Doppler arcs.

The United States Global Positioning System (GPS) is presently the only operational global navigation satellite system. The system is widely used, has been extensively studied, and is the subject of many textbooks (Kaplan, 1996; Hofmann-Wellenhof et al., 1997; Misra & Enge, 2001). The GPS constellation nominally consists of 24 satellites arranged in six orbital planes designed to provide continuous worldwide instantaneous positioning coverage. Each satellite transmits continuous signals on two channels, L1 and L2. Code-division multiple access (CDMA) using pseudo-random noise (PRN) codes is used to enable multiple satellites to transmit on the same frequency, to spread the spectrum of the signal as a defense against jamming, and to provide a method of making pseudorange observations through correlation of the received signal with replica codes. The Mars Network will likely also use PRN codes for ranging, but otherwise has little in common with GPS in terms of constellation or signal design.

The Mars Network design also draws on some of the experience gained through the French DORIS, and German PRARE satellite tracking systems. Particularly the use of one and two-way signals. Finally, frequency division multiple access (FDMA) is currently used by the Russian GNSS, GLONASS, and will be used by future GPS satellites for satellite-to-satellite crosslink ranging. Table 2.5 summarizes and compares the Mars Network and the above mentioned satellite navigation systems.

Table 2.5: Comparison of the Mars Network to other satellite navigation systems. Values for TRANSIT from Stansell (1978), GPS from Spilker Jr (1996a), GLONASS from Daly & Misra (1996), Galileo from O'Keefe et al. (2002), DORIS and PRARE from Montenbruck & Gill (2000).

System	Mars Network	TRANSIT	GPS	GLONASS	Galileo	DORIS	PRARE
# of Satellites	6	3-6	24 (28)	24 (6)	30	n/a	n/a
Altitude	800 km	1075 km	20200 km	19100	23000	n/a	n/a
Orbital Period	137 min	107 min	718 min	675 min	835 min	n/a	n/a
Orbit Planes	6	3-6	6	3	3	n/a	n/a
Inclination	172° 111°	90°	55°	65°	54°	n/a	n/a
Frequencies	400-435 MHz	150 MHz 400 MHz	1575 MHz 1227 MHz 1176 MHz	1600 MHz 1250 MHz	1575 MHz 1207 MHz 1176 MHz	401 MHz 2036 MHz	2248 MHz 8489 MHz (down) 7225 MHz (up)
Signal	Intermittent	Continuous	Continuous	Continuous	Continuous	Intermittent	Intermittent
Multiple Access	FDMA	none	CDMA	FDMA	CDMA	none	CDMA
Observables							
Range	1 & 2-way		1-way	1-way	1-way		2-way
Doppler	1 & 2-way	1-way	1-way	1-way	1-way	1-way	2-way
Phase			1-way	1-way	1-way		
Orbit Representation	tbd	Modified Keplerian	Modified Keplerian	State Vector	tbd	n/a	n/a

Chapter 3

Mathematical Methodology

This chapter will discuss four different mathematical topics that are subsequently used in this thesis. It begins with a discussion of performance measures for global navigation satellite systems. This is followed by a section discussing various orbit representation schemes and prediction methods. Two estimation techniques, Least Squares and Kalman Filtering, are briefly reviewed. The chapter then concludes with a discussion of various strategies for real-time and post-mission estimation of parameters when attempting to position either a single user or a network of users.

3.1 Performance Measures for Global Navigation Satellite Systems

Availability, accuracy, and reliability are often used as measures to quantify the performance of a navigation system. Even in the context of navigation, these terms have multiple definitions. To avoid confusion, each measure is introduced and discussed in this section.

3.1.1 Availability

From a generic radionavigation standpoint, availability is often used to refer to the percentage of the time a particular system is able to provide the user with navigation solutions (FRP, 2001). This type of availability will be referred to as system availability in this thesis. In the context of GNSS, availability is more often used to refer to the number of satellites or other ranging signals available to the user at any given time. The satellites in the Mars Network are in relatively low orbits, so often no satellites, or only one satellite is in view, or available, to users on the surface. To account for this, two definitions of availability will be used. Availability, or instantaneous availability, will be used to denote the number of transmitters in view to a receiver at a particular observation epoch. This is the standard definition of availability used in GNSS research. Cumulative availability will be used to refer to the total number of observations made by a particular stationary user over a period of time. While instantaneous availability depends only on transmitter-receiver geometry at a particular instant, cumulative availability depends both on the geometry and the observation rate. The concept of a satellite pass is introduced as a third measure of availability to represent cumulative availability in a way that is not dependent on the observation rate. For example, a single satellite in view for 15 minutes to a user making observations every minute would result in an availability of one, a cumulative availability of 15, and be counted as one satellite pass.

3.1.2 Accuracy

Accuracy is a measure of how close the navigation solution provided by the system is to the users true location and velocity. The estimated accuracy of a user with a state vector $\boldsymbol{x}$ is most generally represented by the covariance matrix of the estimated states, $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$. In a least squares solution, the covariance matrix of the estimated states is obtained by mapping the covariance matrix of the observations, $\boldsymbol{C}_{\ell}$, from observation space into state space

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}} = (\boldsymbol{A}^T \boldsymbol{C}_{\ell}^{-1} \boldsymbol{A})^{-1} \tag{3.1}$$

where the design matrix, $\boldsymbol{A}$, is the matrix of partial differentials of the observation vector, $\boldsymbol{\ell}$, with respect to the state vector, $\hat{\boldsymbol{x}}$.

$\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ can be used to define an ellipsoidal confidence region in state space. However, often a more simplified measure of accuracy is desired. One approach is to simply ignore the off-diagonal terms of $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ and use the vector sum of the diagonal terms as a measure of accuracy. For example, when the state vector consists of a three-dimensional position, Mean Radial Spherical Error (MRSE) is defined as

$$\text{MRSE} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \tag{3.2}$$

where σ_x^2, σ_y^2, and σ_z^2 are the diagonal elements of $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$. MRSE is a spherical confidence region that approximates the actual confidence ellipsoid. Clearly this approximation is only accurate provided that the off-diagonal elements of $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ are small. The two-dimensional analog to MRSE is called distance root mean squared (DRMS).

A further simplification can be made in the case that all of the observations have similar, or unknown, accuracies, $\sigma_{\boldsymbol{\ell}}$. If $\boldsymbol{C}_{\ell}$ is replaced by $\sigma_{\boldsymbol{\ell}}^2$ times the identity matrix in equation 3.1

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}} = (\boldsymbol{A}^T(\sigma_{\boldsymbol{\ell}}^2\boldsymbol{I})^{-1}\boldsymbol{A})^{-1} \tag{3.3}$$

$\sigma_{\boldsymbol{\ell}}^2$ can be factored out

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}} = \sigma_{\boldsymbol{\ell}}^2(\boldsymbol{A}^T\boldsymbol{A})^{-1} \tag{3.4}$$

The accuracy can now be decomposed in two scalar quantities: User Equivalent Range Error (UERE) and Dilution of Precision (DOP) by replacing both $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ and $(\boldsymbol{A}_T\boldsymbol{A})^{-1}$ by their traces and taking the square root of both sides.

$$\sqrt{\text{Trace}(\boldsymbol{C}_{\hat{\boldsymbol{x}}})} = \sigma_{\boldsymbol{\ell}}\sqrt{\text{Trace}((\boldsymbol{A}^T\boldsymbol{A})^{-1})} \tag{3.5}$$

For a three-dimensional position, $\sqrt{\text{Trace}(\boldsymbol{C}_{\hat{\boldsymbol{x}}})}$ is the MRSE, $\sigma_{\boldsymbol{\ell}}$ is the UERE, and

$\sqrt{\text{Trace}((\boldsymbol{A}^T\boldsymbol{A})^{-1})}$ is called the position dilution of precision (PDOP).

$$\text{MRSE} = \text{UERE} \times \text{PDOP} \tag{3.6}$$

UERE is a number describing all of the system and user errors as they affect a single error in one user measured range. DOP is the geometry-dependent quantity that maps the UERE (an error in observation space) into user accuracy (in position space).

The advantage of using dilution of precision as an accuracy measure is that it depends only on the positioning geometry described in the design matrix, and thus the geometry of a positioning system can be assessed without having to know the value of the ranging accuracy of the the system.

Dilution of precision was developed as a figure of merit for the Global Positioning System and its standard definition is for the four-by-four (x,y,z, and clock offset) cofactor matrix of the pseudorange point positioning problem

$$\boldsymbol{Q_x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1} \tag{3.7}$$

The square root of the trace of $\boldsymbol{Q_x}$ is generally referred to as Geometric Dilution of Precision (GDOP) while other dilution of precision measures can be formed using only one, two, or three of the diagonal elements. Further DOP definitions and discussion can be found in any introductory satellite navigation text book, for example Hofmann-Wellenhof et al. (1997) or Misra & Enge (2001).

The definition of DOP requires that $\boldsymbol{A}^T\boldsymbol{A}$ be invertable, which is equivalent to requiring that sufficient observations are available for a solution. In GNSS applications, DOP is usually evaluated as a function of the satellite geometry at a particular epoch. In the case of the Mars Network, it becomes necessary to introduce the concept of cumulative dilution of precision, that is the effect of the geometry of the satellites observed over time. The same mathematical definition of DOP is used, but in this case, the design matrix, $\boldsymbol{A}$, contains observations from multiple epochs.

Though the use of DOP as a figure of merit is generally limited to GNSS, it can equally well be employed in other radionavigation systems. MacNicol & Raquet (2002) have used DOP as a figure of merit to evaluate a number of proposed satellite navigation constellations for the Moon and Ma (2003) has applied DOP as a figure of merit to integrated GPS and network-based cellular telephone positioning. Depending on the system, the definition of dilution of precision has to be modified to correspond to the states being estimated. For example, if two-way ranging is used, and a clock state is not being estimated, the cofactor matrix would be a three-by-three matrix and only PDOP would be defined. Dilution of precision is used extensively to evaluate the positioning geometry of the Mars Network in Chapter 4. In all cases, care is taken to describe the precise definition of dilution of precision being used in each case.

3.1.3 Reliability

While estimated accuracy, represented by $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ or DOP, is a useful measure of the expected positioning performance, it is only valid if it is assumed that all of the observations are normally distributed about their mean values. Unfortunately, observations often contain biases or blunders, that can only be properly dealt with using statistical reliability theory. As with availability and accuracy, there are multiple definitions of reliability. The United States Federal Radionavigation Plan defines reliability as one minus the probability of system failure (FRP, 2001). In this thesis, reliability is used to refer to the ability to detect blunders in the measurements and to estimate the effects of undetected blunders on the navigation solution.

Reliability theory was first developed in the context of terrestrial geodetic network adjustment (Baarda, 1967, 1968) and has since been applied extensively to satellite navigation applications, particularly marine navigation and aviation. A brief overview of statistical reliability theory is given below. More detailed discussions of reliability theory can be found in Leick (2004) and Koch (1999).

Reliability can be subdivided into internal reliability and external reliability. Internal reliability refers to the ability of the system to detect a blunder through the statistical testing of the least squares residuals on an epoch-to-epoch basis. The smallest such blunder is called the marginally detectable blunder (MDB). The external reliability of a system is quantified by size of the error in the navigation solution that is caused by a marginally detectable blunder.

In order to detect a blunder in an observation using a least squares approach, statistical testing is conducted on the least squares residuals. In least squares estimation, it is assumed that the residuals are normally distributed. If a blunder is present in an observation, its residual will be biased, but will remain normally distributed. Note that residuals, and hence redundancy in the observations, are required. The least squares residuals are given by

$$\boldsymbol{r} = \boldsymbol{C_r}\boldsymbol{C}_\ell^{-1}\boldsymbol{w} = \boldsymbol{R}\boldsymbol{w} \tag{3.8}$$

where $\boldsymbol{C_r}$ is the covariance matrix of the residuals, $\boldsymbol{C}_\ell$ is the covariance matrix of the observations, and $\boldsymbol{w}$ is the misclosure vector. $\boldsymbol{R}$ is the redundancy matrix and the redundancy of an observation can be expressed by its redundancy number:

$$R_i = \{\boldsymbol{C_r}\boldsymbol{C}_\ell^{-1}\}_{ii} \tag{3.9}$$

which is the i^{th} diagonal element of $\boldsymbol{R}$. The covariance of the residuals is equal to the covariance of the observations minus the covariance of the estimated state, $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$, mapped into observation space by the design matrix $\boldsymbol{A}$.

$$\boldsymbol{C_r} = \boldsymbol{C_\ell} - \boldsymbol{A}\boldsymbol{C_{\hat{x}}}\boldsymbol{A}^T \tag{3.10}$$

$\boldsymbol{C_r}$ is always less than or equal to $\boldsymbol{C_\ell}$, meaning that R_i is always between 0 and 1. A redundancy number of 1 would mean that the observation is completely redundant and thus easily monitored for blunders while a redundancy number of 0 indicates that the navigation solution depends completely on this observation making the identification of a blunder impossible.

To detect a blunder, each residual can be statistically tested where the null hypothesis, H_0, is that the residual is unbiased while the alternative hypothesis, H_a, is that the residual is biased. When such a test is performed, two types of errors may occur. If a good observation is rejected, a Type I error occurs. The probability of this is commonly denoted by α. A Type II error occurs when a blunder is accepted into the solution. The probability of committing a Type II error is denoted by β. Choosing values of α and β determine a bias or non-centrality parameter of H_a that is denoted by δ_0. The relationship between α, β, δ_0, and the probability distributions of the two hypothesis is shown in Figure 3.1.

The marginally detectable blunder for observation i,$|\nabla_i|$, can then be obtained by multiplying δ_0 by the standard deviation of the residual and dividing by the redundancy number.

$$|\nabla_i| = \frac{\delta_0\sqrt{\boldsymbol{C_r}_{ii}}}{R_i} \tag{3.11}$$

Since each residual has a different standard deviation and each observation has a different redundancy, each observation has a different MDB. Assuming only one blunder occurs in a given measurement epoch, the maximum effect of one undetected blunder can be determined by evaluating the effect of each marginally detectable blunder on

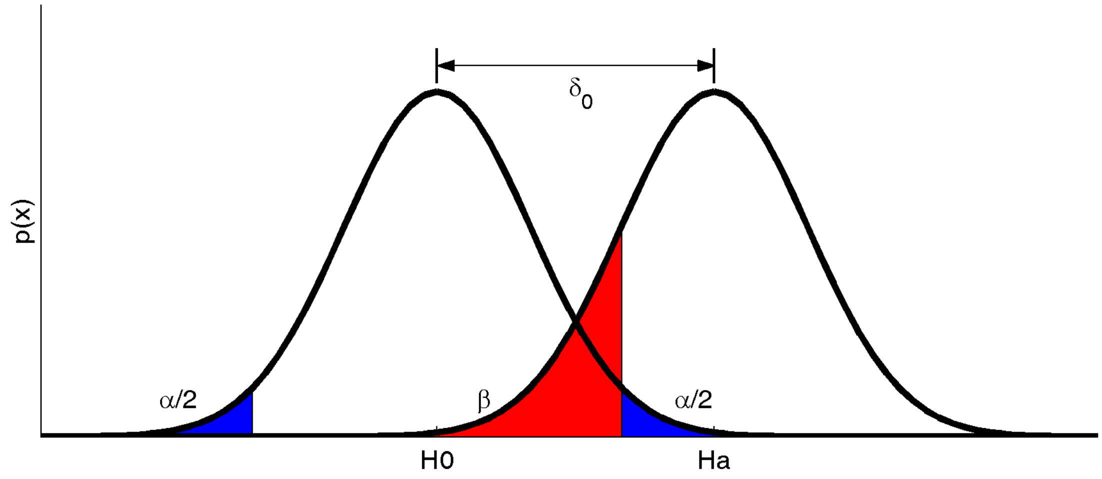

Figure 3.1: Hypothesis testing and the relationship between Type I error, Type II error, and the noncentrality parameter. α is represented by the blue area in the tails of the H_0 probability distribution, while β is the area in red on the left of the H_a probability distribution.

the navigation solution.

$$\boldsymbol{\Delta x} = -(\boldsymbol{A}^T\boldsymbol{C}_\ell^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{C}_\ell^{-1}\nabla \tag{3.12}$$

where ∇ is a column vector of zeros except for the i^{th} row which contains the marginally detectable blunder of the i^{th} observation.

The above theory is restricted theoretically to the least squares solutions where a maximum of one biased observation, or blunder, is present. Also, it is assumed that the estimation is taking place in a least squares batch estimation, where all of the observations and residuals are available for testing. In Chapter 4 of this thesis, this method is used and reliability is quantified by the position error (PE) due to one marginally detectable blunder. This is obtained by evaluating the three-dimensional magnitude of $\boldsymbol{\Delta x}$ for each observation's marginally detectable blunder and retaining only the largest value.

In civil aviation literature, the position error due to one marginally detectable blunder (PE) is called the protection level (PL). This is because a system with a given PL is protected against errors larger than this level going undetected. In principle, if the PL exceeds a specified maximum value, called the protection limit, the system should inform the user not to rely on the navigation solution.

Statistical reliability theory only provides a bound on the size of the position error due to one marginally detectable blunder. The theory says nothing about the probability of such a blunder actually occurring and going undetected. The ability of a navigation system to perform an operation without any type of failure is called continuity (FRP, 2001). To evaluate the continuity of a system, all of its individual components have to be tested to determine a probability of failure for each.

Redundant observations are rarely, if ever, available for a ground user given the Mars Network constellation design. However, reliability, like dilution of precision, can also be calculated in a cumulative sense by assuming the user is storing observations and performing a least squares batch solution using all of the previous observations and the current epoch observation.

Another option would be to compute the reliability of the sequential or filtered solutions instead of the batch solution. The reliability of sequential or filtered solutions is closely related to the reliability of a batch least squares solution. The theoretical development was initially given by Teunissen (1990) and is also presented in Koch (1999). Its application to navigation systems has most recently been discussed in Petovello (2003).

In least squares estimation, blunders are detected by testing the measurement residuals after a least squares solution is obtained. If a blunder is found, it is removed and the least squares solution is recomputed. In sequential least squares and Kalman filtering, an a priori estimate of the state and state covariance is available, making it possible to test the predicted residual before the new observation enters the filter and negatively impacts the quality of the state vector. Testing is conducted by comparing the predicted residuals, $\boldsymbol{\nu}$, with their estimated variance, just as residual testing in the least squares case compares the residuals to their corresponding estimated variances. The predicted residuals are also called the innovation sequence and are obtained by subtracting the predicted observations from the observation vector.

$$\boldsymbol{\nu} = \boldsymbol{\ell} - A\hat{\boldsymbol{x}}^{(-)} \tag{3.13}$$

The covariance of the innovation sequence can be shown to be

$$\boldsymbol{C}_{\boldsymbol{\nu}} = \boldsymbol{A}^T \boldsymbol{C}_{\hat{\boldsymbol{x}}} \boldsymbol{A} + \boldsymbol{C}_{\ell} \tag{3.14}$$

The main disadvantage of using reliability testing in a sequential adjustment or Kalman filter is that observations from previous epochs are no longer directly available at the present epoch. It is conceivable that earlier observations contained blunders that were, at the time, undetectable. In a batch estimation scheme, all of the observations and their respective residuals are available for testing, while in a filter, it is not possible to recover from the effects of a previously undetected blunder.

3.2 Satellite Orbit Representation and Prediction Methods

In this section, satellite orbital motion is discussed. Several methods of representing orbital motion are presented, and general analytical and numerical methods for predicting satellite motion are discussed. The details of modeling Mars Network satellite orbits by numerical integration are discussed in Chapter 5.

3.2.1 Equations of Motion in Rectangular Coordinates

A satellite, like any object, is subject to Newton's laws of motion. Specifically, the second law states that a particle of mass m will accelerate when subject to a force $\boldsymbol{F}$,

$$\boldsymbol{F} = m\ddot{\boldsymbol{r}} \tag{3.15}$$

where $\ddot{\boldsymbol{r}}$ is the acceleration. This second order vector differential equation can be solved as an initial value problem provided that six constants of integration are available, i.e an initial state vector consisting of the position and velocity at the initial time. All that is required in addition to the initial state, is a expression representing the force as a function of the position, velocity, and time.

The simplest case is that of a satellite affected only by gravitation orbiting a much larger point mass or spherically symmetric planet, with the coordinate system centred at the centre of the planet. In this case the force depends only on the position and can be described by Newton's law of gravitation

$$\boldsymbol{F}(\boldsymbol{r}) = -G\frac{mM}{|r|^3}\boldsymbol{r} \tag{3.16}$$

where G is the gravitational constant, M is the mass of the planet, and $\boldsymbol{r}$ is the vector from the centre of the planet to the satellite. Substituting this force into equation 3.15 and eliminating m results in

$$\ddot{\boldsymbol{r}} = -G\frac{M}{|r|^3}\boldsymbol{r} \tag{3.17}$$

which is the standard equation of motion for a satellite orbiting a spherically symmetric planet. The analytical solution of this equation is the standard Keplerian orbital motion, which with appropriate initial conditions will result in the satellite following an elliptical orbit with the planet located at one of the foci of the ellipse.

Kepler originally developed an empirical solution of this problem by trying to describe the motion of the planets about the Sun. His description consisted of three laws. First, that orbits were elliptical, second, that the area swept out by the radius vector over time is constant, and third, that the square of the orbital period is proportional to the cube of the radius. The problem first solved analytically in the 17th century by Issac Newton with the application of his law of gravitation. Keplerian orbital motion has several interesting characteristics. The central force always acts in the opposite direction of the radial vector between the central body and the satellite. This constrains the orbit to a fixed plane since there is never any acceleration perpendicular to the plane. Furthermore, the cross product between the position and velocity vectors of the satellite can be shown to be a constant due to the fact that the acceleration vector and radius vector are parallel.

$$\boldsymbol{r} \times \ddot{\boldsymbol{r}} = \boldsymbol{r} \times \ddot{\boldsymbol{r}} + \dot{\boldsymbol{r}} \times \dot{\boldsymbol{r}} = \frac{d}{dt}(\boldsymbol{r} \times \dot{\boldsymbol{r}}) = 0 \tag{3.18}$$

Since the time derivative is zero, $\boldsymbol{r} \times \dot{\boldsymbol{r}}$ is a constant. This implies Kepler's second law since the integral of a constant over time is a constant.

Instead of using an equation of motion, and six initial conditions, elliptical orbital motion can be represented by six Keplerian elements. Before considering more complicated forces acting on a satellite, the Keplerian Element orbital representation will be discussed.

3.2.2 Keplerian Elements

An ellipitical orbit of a satellite about a spherical planet can be uniquely defined by six Keplerian elements: semi-major axis a, eccentricity e, inclination i, right-ascension of the ascending Ω, argument of perigee ω, and mean anomaly M. The first two elements define the shape of the orbital ellipse, the next three angles give its orientation in space, and the mean anomaly represents the location of the satellite along its orbit. There are several equivalent representations of the final element, including true anomaly ν, eccentric anomaly, E, and time of last perigee passing. The relationship between Ω, i, ω, and ν is shown in Figure 3.2. The mathematical relationship between the true, eccentric, and mean anomalies is discussed in Appendix D.

In most cases, the Keplerian elements of a satellite are interchangeable with the satellite's state vector. Conversion formulas between the Keplerian elements $(a, e, i, \Omega, \omega, M)$ and the state vector $[x, y, z, v_x, v_y, v_z]^T$ are given in Appendix D and can be found in any introductory astrodynamics textbook, for example Roy (1995) or Montenbruck & Gill (2000). In the case of a perfectly circular orbit, the perigee point is undefined, making both the argument of perigee and the mean anomaly impossible to define. Likewise, for non-inclined orbits, the right ascension of the ascending node is undefined.

For a satellite orbiting a spherical planet, a, e, i, Ω, and ω are constants and the mean anomaly increases at a uniform rate n, called the mean motion.

$$n = \sqrt{\frac{GM}{a^3}} \tag{3.19}$$

However, if additional forces are acting on the satellite, the Keplerian analytical solution no longer applies and other methods must be used to solve for the state vector of the satellite. It is still possible to convert between the state vector and the corresponding Keplerian elements, but the Keplerian elements will instead be time varying, or osculating, elements.

3.2.3 Perturbed Keplerian Motion

In the case that the additional forces acting on the satellite are small with respect to the central gravitational force, it is possible to represent the equation of motion of

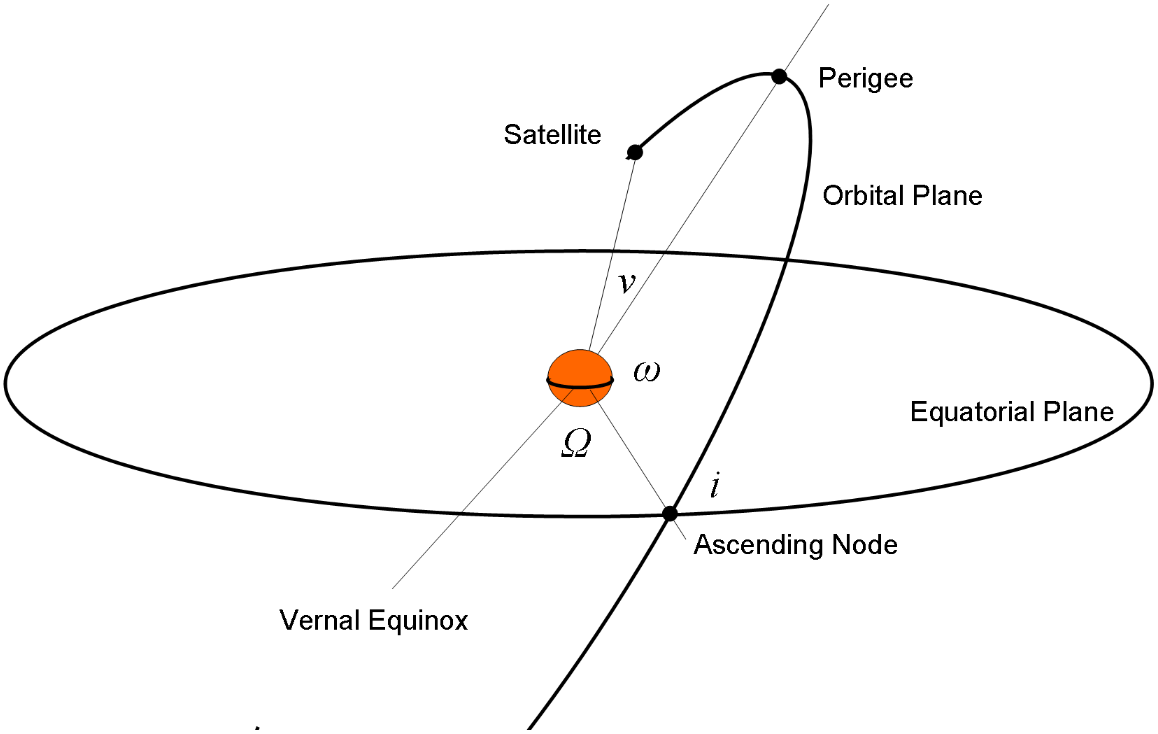

Figure 3.2: Diagram showing the relationship between the four angular Keplerian elements. Ω is the angle in the orbital plane between the vernal equinox and the ascending node, i is the angle between the equatorial and orbital planes, ω is the angle in the orbital plane between the ascending node and the perigee point, and ν is the angle in the orbital plane between the perigee point and the planet. The remaining elements, a and e, define the shape of the orbital ellipse.

the satellite as

$$\ddot{\boldsymbol{r}} = -G\frac{M}{|r|^3}\boldsymbol{r} + \boldsymbol{k} \tag{3.20}$$

where $\boldsymbol{k}$ is the perturbing acceleration vector. Seeber (1993) identifies five different types of perturbing accelerations, including accelerations due to higher order terms in the gravity field of the planet, accelerations due to other celestial bodies, planet and ocean tide effects, atmospheric drag, and radiation pressure effects. Seeber (1993) further describes two major methods for determining satellite motion when perturbing accelerations are present. The first is numerical integration which is discussed in Section 3.2.5 below. The second approach, which he calls analytical integration, involves solving for the time varying Keplerian elements.

The solution, originally formulated by Lagrange, is presented in Seeber (1993), and in more detail in Kaula (1966), and is not repeated here. The general idea, however, is to determine time derivatives of the Keplerian elements as functions of the elements themselves, and of the disturbing potential, or potential due to the disturbing accelerations. Provided that the disturbing potential can be expressed in terms of the orbital elements, it is possible to analytically integrate and solve for the time varying Keplerian elements. Depending on the form of the integral, it is then possible to see if a particular perturbing force causes secular or periodic variation in a particular Keplerian element.

For most Earth and Mars orbiting satellites, the effect of the oblateness of the planet is several orders of magnitude larger than any of the other perturbing accelerations and results in secular perturbations to the right ascension of the ascending node, the argument of perigee, and the mean anomaly.

The three oblateness perturbation equations (adapted from Seeber (1993)) are

$$\frac{d\Omega}{dt} = C_{20}\frac{3na_p^2}{2a^2(1-e^2)^2}\cos i \tag{3.21}$$

$$\frac{d\omega}{dt} = C_{20}\frac{3na_p^2}{4a^2(1-e^2)^2}(1-5\cos^2 i) \tag{3.22}$$

$$\frac{dM}{dt} = n - C_{20}\frac{3na_p^2}{4a^2(1-e^2)^{\frac{3}{2}}}(3\cos^2 i - 1) \tag{3.23}$$

where n is the mean motion defined in equation 3.19, a_p is the equatorial radius of the planet, and C_{20} is the oblateness coefficient of the planet's gravity model. In general, zonal terms of the gravity field give rise to secular drift terms in satellite orbits and long term periodic perturbation, while short term perturbations are caused by tesseral terms in the gravity field (Seeber, 1993).

3.2.4 Modified Keplerian Element Orbit Representations

Several orbit representations have been developed based the the fact that higher order gravity terms tend to cause secular and periodic variations in the Keplerian elements of a satellite. The most common of these for satellite navigation are the GPS ephemeris and almanac formats. Both of these include the six Keplerian elements (or variations of them) and the secular drift rate of the ascending node, $\dot{\Omega}$. The ephemeris message includes further secular rates as well as periodic terms to account for the resonance of the GPS satellite orbits with the rotation rate of the Earth.

As the secular drift of the ascending node is usually the largest perturbation effect on Earth orbiting satellites, the almanac format is often used as an approximate representation for satellites. Mars is also an oblate planet and this representation was used by Ely et al. (1999) to simulate the time varying satellite coverage of the Mars Network. The almanac approximation may be sufficient for low accuracy modeling, but it is insufficient for high accuracy prediction of satellite motion. This is particularly true for satellites that are subject to higher order gravitational forces, tidal forces, and non-conservative forces such as radiation pressure and drag. Though many efforts have been made to deal with these forces analytically, the most common and efficient way to model satellite motion subject to these forces is to use numerical techniques.

3.2.5 Numerical Integration

The motion of a satellite is governed by either a system of six first order differential equations or a system of three second-order differential equations. The integration, or solution of these equations, requires six integration constants. The problem can be treated as either an initial value, or boundary value problem. The boundary value problem is discussed in detail in Beutler & Rothacher (2002). In the case of the initial value problem, the six constants of integration can be either the initial state vector of the satellite, or the corresponding Keplerian elements of the satellite. Analytical solutions for the simplest cases have already been discussed. For all others, numerical integration of the equations of motion must be used.

There are two major methods of numerical integration for orbits: Cowell's Method and Encke's Method (Seeber, 1993). The first directly determines the satellite state, $[x, y, z, v_x, v_y, v_z]^T$, while the second numerically determines perturbations to the Keplerian parameters. In this thesis, Cowell's method is used for its conceptual simplicity. In the past, Cowell's method has been criticized for having two main disadvantages: increased computational time and the possibility of increased round off error. These two disadvantages are of less importance today with the development of faster computers and the use of double precision arithmetic. Two general methods of numerical integration will now be discussed: single step methods and multistep methods.

Single Step Methods

The simplest single step method to solve a first order differential equation by numerical integration is known as Euler's Method (Cheney & Kincaid, 1994). A first order initial value problem of the form

$$\dot{x} = f(t_0, x(t)),\ x(t_0) = x_0 \tag{3.24}$$

may be solved for a time h in the future by Taylor expansion of $x(t)$ about t_0.

$$x(t_0 + h) = x(t_0) + h\dot{x}(t_0) + \frac{h^2}{2}(t_0^2) + \ldots \tag{3.25}$$

Keeping only the first two terms and substituting equation 3.24 for $\dot{x}(t_0)$ gives

$$x(t_0 + h) = x(t_0) + hf(t_0, x(t_0)) \tag{3.26}$$

which is a formula that can predict $x(t)$ provided a functional form of the first derivative of x is known. This prediction of x is called an Euler step. Further h sized steps may then be performed to propagate the solution into the future. The method is a single step method because for each single step, only the result from the previous step is required. Multistep methods, which are discussed below store the results of previous steps and use them in subsequent steps. Clearly, the ability of Euler's method to correctly solve an initial boundary problem depends on the size of the step and the linearity of the function $x(t)$. If $x(t)$ is not very linear, either a very small step size will be required or higher order terms in the Taylor expansion will be required. Unfortunately this requires that higher derivatives of $x(t)$ be computed.

Runge-Kutta methods are a family of higher order single step methods that avoid the computation of higher order derivatives. The fourth-order Runge-Kutta method (RK4) is probably the most common single step method for solving initial value problems. Instead of evaluating the slope $f(t, x(t))$, only at $t = t_0 + h$, RK4 uses a weighted mean of four slopes that approximates the solution that would have been obtained by using terms in the Taylor expansion up to order h^4. The RK4 method is suggested for the prediction of satellite orbits by Buffett (1985), though he does not implement the method.

Multistep Methods

In contrast to single step methods, which approximate the function by an approximate curve originating from a single point, multistep methods approximate the function to be integrated by a polynomial interpolating previous solutions as well as extrapolating forward in time. In principle, multistep methods can increase the effective order and the allowable step size without having to increase the number of function evaluations. The cost for this increased efficiency is the requirement to store previous solutions. A detailed introduction to multistep methods for orbit prediction applications is presented in Montenbruck & Gill (2000). The orbit simulations using numerical integration presented in Chapter 5 makes use of an Adams-Bashforth-Moulton predictor-corrector multistep method called DE, originally developed by Shampine & Gordon (1975). Montenbruck & Gill (2000) recommend this method of numerical integration for all purpose high-precision orbit prediction applications. They also provide a C++ language implementation of the algorithm with a license for non-commercial use. This implementation has been used in this thesis and will be discussed in greater detail in Chapter 5.

3.3 Estimation Procedures

3.3.1 Least Squares

The simplest linear estimation problem involves the estimation of a set of parameters or states, $\boldsymbol{x}$, using a set of observations, $\boldsymbol{\ell}$ that are corrupted by measurement noise $\boldsymbol{v}$. The observations and states are linearly related by

$$\boldsymbol{\ell} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v} \tag{3.27}$$

where $\mathbf{A}$ is called the design matrix. The design matrix is a Jacobian matrix of the form

$$\boldsymbol{A} = \frac{\partial \boldsymbol{\ell}}{\partial \boldsymbol{x}} \tag{3.28}$$

and serves as a linear transformation between observation and state space.

Typically, there are more observations than there are unknown parameters and many writers call this an overdetermined problem. However, it is in fact an underdetermined problem since the noise vector, $\boldsymbol{v}$, is also unknown. To find a unique solution, a condition must be applied that allows for one unique solution.

The weighted least squares solution of this problem is the estimate of the parameters, $\hat{\boldsymbol{x}}$, that minimizes the weighted sum of squares of the residuals. The residuals, $\hat{\boldsymbol{r}}$, are the differences between the observations and the estimated state mapped into observation space.

$$\min\{[\boldsymbol{\ell} - \boldsymbol{A}\hat{\boldsymbol{x}}]^T \boldsymbol{C}_{\ell}^{-1} [\boldsymbol{\ell} - \boldsymbol{A}\hat{\boldsymbol{x}}]\} \tag{3.29}$$

The vector $\boldsymbol{A}\hat{\boldsymbol{x}}$ is often referred to as the adjusted observations, $\hat{\boldsymbol{\ell}}$ (Krakiwsky, 1990). The inverse of the covariance of the observations, $\boldsymbol{C}_{\ell}$, is used to define the relative weights between the observations.

The solution to this minimization problem is

$$\hat{\boldsymbol{x}} = -(\boldsymbol{A}^T \boldsymbol{C}_\ell^{-1} \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{C}_\ell^{-1} \boldsymbol{\ell} \tag{3.30}$$

The covariance of the estimated states, $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ is given by

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}} = (\boldsymbol{A}^T \boldsymbol{C}_\ell \boldsymbol{A})^{-1} \tag{3.31}$$

which conceptually is the mapping of the covariance matrix of the observations, $\boldsymbol{C}_\ell$, into state space through the design matrix $\boldsymbol{A}$.

The above formulation represents the simplest case of linear least squares estimation of time invariant parameters. In general, the relationship between the observations and the parameters may not be linear and the estimated parameters may not be constant over time. Standard least squares methods for linearizing non-linear observations and dealing with time varying parameters have been developed and are not reproduced here. However, these methods are the basis for similar methods in Kalman Filtering that will be discussed in the following section.

3.3.2 Kalman Filtering

The Kalman filter is a recursive estimation algorithm for predicting time evolution of linear systems and updating the state with observations. It differs from least squares estimation in that additional information about the system is included in a dynamic model of the system.

A linear system may be described by a state vector $\boldsymbol{x}$ and the linear system of differential equations

$$\dot{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{x} \tag{3.32}$$

where the system of equations is described by the dynamics matrix, $\boldsymbol{F}$. Given a set of initial conditions, this system of differential equations should describe the system for all time. The purpose of a Kalman filter is to both predict the future value of the state vector and its covariance matrix based on this system of differential equations, and to allow for the inclusion of observations in the solution.

A Kalman filter consists of two distinct processes, prediction and updating. The prediction step is required to predict the value of the state vector from one epoch to the next. In the standard discrete-time formulation, the prediction step for the state vector from time $k-1$ to time k can be expressed by

$$\hat{\boldsymbol{x}}_k^{(-)} = \boldsymbol{\Phi}_{k,k-1}\hat{\boldsymbol{x}}^{(+)}{}_{k-1} \tag{3.33}$$

where $\hat{\boldsymbol{x}}^{(+)}{}_{k-1}$ is the previous value of the state vector, $\hat{\boldsymbol{x}}_k^{(-)}$ is the predicted value of the state vector, and $\boldsymbol{\Phi}_{k,k-1}$ is the transition matrix. Prediction of the covariance matrix is also accomplished using the transition matrix and through the addition of process noise, described by the matrix $\boldsymbol{Q}_{k,k-1}$.

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}}^{(-)}{}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{C}_{\hat{\boldsymbol{x}}}^{(+)}{}_{k-1}\boldsymbol{\Phi}_{k,k-1}^T + \boldsymbol{Q}_{k,k-1} \tag{3.34}$$

The transition matrix, $\boldsymbol{\Phi}_{k,k-1}$, is a Jacobian matrix describing the differential relationship between the state at time $k-1$ and the state at time k and is related to the dynamics matrix, $\boldsymbol{F}$, by solving the dynamics equations

$$\dot{\boldsymbol{x}} = \boldsymbol{F}\boldsymbol{x} \tag{3.35}$$

If $\boldsymbol{F}$ is a constant over the time of interest, this system of differential equations has the solution

$$\boldsymbol{x}(t_k) = e^{\boldsymbol{F}\Delta t}\boldsymbol{x}(t_{k-1}) \tag{3.36}$$

where $\Delta t = t_k - t_{k-1}$. Using the matrix exponential power series expansion

$$e^{\boldsymbol{F}\Delta t} = \boldsymbol{I} + \boldsymbol{F}\cdot\Delta t + \frac{1}{2!}\boldsymbol{F}^2\cdot\Delta t^2 + \ldots \tag{3.37}$$

and retaining only the first two terms gives the standard result for the transition matrix given a constant dynamics matrix

$$\boldsymbol{\Phi} = \boldsymbol{I} + \boldsymbol{F} \cdot \Delta t \tag{3.38}$$

The prediction process of a Kalman filter is actually a method for solving an initial value problem. If the linear approximation of $\boldsymbol{\Phi}_{k,k-1}$ from equation 3.38 is used, the prediction step of a Kalman Filter is equivalent to Euler's method of numerical solution of differential equations. This method is described in more detail in Section 3.2.5.

The important feature of the Kalman filter is its ability to incorporate new observations using a Kalman update step. The update step follows a prediction step up to the time of the new observations. The updated state vector $\hat{\boldsymbol{x}}_k^{(+)}$ is obtained from the predicted state vector $\hat{\boldsymbol{x}}_k^{(-)}$ and the observation vector $\boldsymbol{\ell}$ by

$$\hat{\boldsymbol{x}}_k^{(+)} = \hat{\boldsymbol{x}}_k^{(-)} + \boldsymbol{K}_k(\boldsymbol{\ell}_k - \boldsymbol{A}_k\hat{\boldsymbol{x}}_k^{(-)}) \tag{3.39}$$

where $\boldsymbol{A}_k$ is the design matrix and $\boldsymbol{K}_k$ is called the Kalman gain matrix.

$$\boldsymbol{K}_k = \boldsymbol{C}^{(-)}_{\hat{\boldsymbol{x}}}{}_k \boldsymbol{A}^T [\boldsymbol{A}\boldsymbol{C}_{\hat{\boldsymbol{x}}}{}^{(-)}{}_k \boldsymbol{A} + \boldsymbol{C}_\ell]^{-1} \tag{3.40}$$

The covariance matrix of the states is also updated.

$$\boldsymbol{C}^{(+)}_{\hat{\boldsymbol{x}}}{}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{A})\boldsymbol{C}^{(-)}_{\hat{\boldsymbol{x}}}{}_k \tag{3.41}$$

A further advantage of the Kalman filter is in its independent formulation of the prediction and update steps. If observations are available, an update can be performed, if not, the prediction part of the algorithm can continue until new observations are made.

Linearized Kalman Filter

The linearized Kalman Filter is a direct extension of the Kalman filter for non-linear systems. Instead of estimating the state vector $\boldsymbol{x}$, a nominal or reference trajectory, $\boldsymbol{x}_{\mathrm{ref}}$, is introduced and the error state, or deviation from the reference trajectory, is estimated.

$$\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_{\mathrm{ref}} \tag{3.42}$$

The linearized Kalman filter is based on the assumption that the reference trajectory is close enough to the true trajectory that small variations from the reference trajectory are linear quantities. Following the derivation given in Grewal & Andrews (1993), the linearized Kalman filter prediction equations corresponding to equations 3.33 and 3.34 are

$$\widehat{\delta\boldsymbol{x}}_k^{(-)} = \boldsymbol{\Phi}_{k,k-1}\widehat{\delta\boldsymbol{x}}^{(+)}{}_{k-1} \tag{3.43}$$

and

$$\boldsymbol{C}^{(-)}_{\hat{\boldsymbol{x}}}{}_k = \boldsymbol{\Phi}_{k,k-1}\boldsymbol{C}^{(+)}_{\hat{\boldsymbol{x}}}{}_{k-1}\boldsymbol{\Phi}^T_{k,k-1} + \boldsymbol{Q}_{k,k-1} \tag{3.44}$$

and the update equations corresponding to equations 3.39, 3.40, and 3.41 are

$$\hat{\boldsymbol{x}}_k^{(+)} = \hat{\boldsymbol{x}}_k^{(-)} + \boldsymbol{K}_k(\boldsymbol{\ell}_k - \boldsymbol{A}_k\hat{\boldsymbol{x}}_{\mathrm{ref}} - \boldsymbol{A}_k\hat{\boldsymbol{x}}_k^{(-)}) \tag{3.45}$$

$$\boldsymbol{K}_k = \boldsymbol{C}^{(-)}_{\hat{\boldsymbol{x}}}{}_k\boldsymbol{A}^T[\boldsymbol{A}\boldsymbol{C}_{\hat{\boldsymbol{x}}}{}^{(-)}{}_k\boldsymbol{A} + \boldsymbol{C}_\ell]^{-1} \tag{3.46}$$

and

$$\boldsymbol{C}^{(+)}_{\hat{\boldsymbol{x}}}{}_k = (\boldsymbol{I} - \boldsymbol{K}_k\boldsymbol{A})\boldsymbol{C}^{(-)}_{\hat{\boldsymbol{x}}}{}_k \tag{3.47}$$

There are two major algebraic differences between a linear and a linearized Kalman filter. First, the error state is now being estimated, and second, the observation input in the update step is now the difference between the observation, $\boldsymbol{\ell}_k$, and sum of the reference and error state vectors mapped into observation space with the design matrix, $\boldsymbol{A}$.

Extended Kalman Filter

The linearized Kalman filter is sufficient for dealing with non-linear systems provided that the true trajectory remains close to the reference trajectory. However, if the true trajectory deviates too far from the reference trajectory, the linearity assumption may no longer be valid. The strategy of the extended Kalman filter is to redefine the reference trajectory by adopting the current error state as the new initial state. With a new initial state, a new reference trajectory may be computed. In this respect, the extended Kalman filter is more robust than the linearized filter. However, the redefinition of the reference trajectory may be computationally expensive and difficult to accomplish in real time. On the other hand, in a linearized Kalman filter, the reference trajectory may be computed offline, as well as the transition and design matrices, if the observation intervals and corresponding geometry are known in advance.

Challenges of Kalman Filtering for Orbit Determination

Applications of Kalman filtering to positioning and navigation are well known with the method being particularly popular in kinematic carrier phase differential GPS processing and inertial navigation. Kalman filtering is also commonly used in precise orbit determination (POD). This usually consists of determining the orbit of a low earth orbiting satellite carrying a dual-frequency GPS receiver using very high rate GPS positions as the filter input. Using this technique, it is possible to obtain centimetre level orbits in near real time (Montenbruck & Gill, 2000).

A more challenging problem is the use of sparse observations to determine the orbit of a satellite. Generally this is done by relying heavily on a dynamics model of the satellite motion and incorporating the effects of observations as they become available.

Kalman filtering for orbit determination is also further complicated by the need to properly define the transition matrix as a function of the dynamics model. In most GPS, INS, or POD applications, the observation rate is fast enough that transition times are relatively short. In these cases, the matrix exponential approximation of the transition matrix (equation 3.38) is sufficient. When observations are sparse and the dynamics matrix, $\boldsymbol{F}$, is not constant, the transition matrix must be obtained more rigorously so that it will accurately reflect the dynamics model over longer transition times.

A derivation of the transition matrix in terms of the dynamics model is given in Appendix A. Kalman filters must also be carefully tuned to ensure the correct balance between observation variance and process noise. These issues will be discussed further with the results of Kalman filtering for positioning and orbit improvement in Chapter 6.

3.4 Network Processing Strategies

The Mars Network and its users can be thought of as a time varying geodetic network where one or more of the network points are being estimated depending on the particular positioning scenario. Positioning results will depend on the strategies used to process the observations and how covariances between observations and states are handled. Of particular interest is how to process observations between two users who are attempting to estimate or improve their own state vectors.

This problem is considered in the context of Earth-based GNSS inter-satellite ranging by Wolf (2000). He proposes three different processing strategies: centralized post-processing, pairwise processing, and autonomous processing. Each of these strategies will now be discussed in the context of the Mars Network.

3.4.1 Centralized Processing

In a centralized post-processing scheme, all observations from all elements in the network are collected at a central location and processed together. A state vector containing the state variables of each element is estimated based on all available observations. The resulting state estimate and state covariance includes all covariances between the different elements in the network. This is clearly the optimal estimation method, however, its implementation is not suited for real-time applications since all of the observations must first be collected before being processed. This strategy can be implemented both in a near real-time Kalman filter (depending on the time required to collect the observations) or as a post-processed least squares solution. This will likely be the first strategy used with the Mars Network, with all the data analysis taking place post-mission on Earth.

3.4.2 Pairwise Processing

A somewhat sub-optimal approach would be for two network elements that observe each other to estimate their state vectors together, taking into account the covariance between their state vectors. One problem with this approach is deciding what happens with the covariance information after the update. If the network elements make observations of other elements before again observing each other, the covariance between their state vectors will no longer be valid. More likely, this information would simply be discarded after an observation session between the two elements.

3.4.3 Fully Autonomous Processing

Finally, fully autonomous processing is accomplished when one network element observes other elements and estimates its state vector while holding the other elements fixed. All that is required is that the other elements transmit their state vectors, and possibly state covariance matrices. The observing network element can then estimate its position based on the observation of the other elements. Fully autonomous processing is the strategy employed in single point GPS processing and is clearly the most suited for real-time applications as very little data needs to be transmitted between network elements. In this strategy, all covariance between elements is ignored. Wolf (2000) demonstrates the obvious sub-optimality of ignoring the covariance between the state vectors of different elements, but does not implement any of the three real-time processing strategies.

One further advantage of this approach is that it is easy to generalize to include network elements that are not making observations and improving their state estimates. In the scenarios presented in Chapter 6, all of the elements in the network (satellites and landers) are assumed to be capable of making observations, though in certain scenarios, only some elements might be estimating their positions.

Chapter 4

Theoretical Navigation Performance of the Mars Network

In this chapter, the performance of the Mars Network for positioning users on the surface of the planet is evaluated using the figures of merit defined in Section 3.1.

4.1 Simulation Scenarios

4.1.1 Constellation Design

The theoretical navigation performance of a GNSS can be evaluated based on the satellite-user geometry using assumptions about the quality of the observations. Since it is not necessary to simulate actual observations, the precise positions of the satellites are not required and the almanac orbit representation, discussed in Section 3.2.4, can be used. To create almanac records for each of the satellites, the constellation design Keplerian elements from Table 2.4 were used. The constellation design specifies an altitude, inclination, right-ascension, and mean anomaly for each satellite. The semi-major axis was obtained from the altitude by adding the equatorial radius of Mars from Table 2.1. The design also specifies circular orbits, which are both difficult to achieve in practice and are not well described by Keplerian elements. Each satellite was assigned an eccentricity of 0.0001, and an initial argument of perigee of 0°. The secular drift rate of the right ascension of the ascending node was calculated using equation 3.21. The almanac parameters for each of the Mars network satellites are shown in Table 4.1.

Table 4.1: Mars Network satellite almanac parameters.

Satellite #	a (m)	e	i (°)	Ω (°)	ω (°)	M (°)	$\dot{\Omega}$ (°/day)
1	4196190	0.0001	172	0	0	0	7.2
2	4196190	0.0001	172	0	0	180	7.2
3	4196190	0.0001	111	0	0	0	2.6
4	4196190	0.0001	111	90	0	90	2.6
5	4196190	0.0001	111	180	0	180	2.6
6	4196190	0.0001	111	270	0	270	2.6

The initial orientation of the constellation with respect to the surface of the planet is not defined in any documents describing the Mars Network that are discussed in Section 2.4. In this thesis it is assumed that the the right ascensions are measured with respect to the Mars vernal equinox. The orientation of the constellation with respect to the surface will then depend on the rotation parameters of the planet. The choice of initial orientation is clearly arbitrary and will have no impact on the assessed performance of the system provided that the assessment is conducted over all longitudes. In order to make use of low accuracy ephemerides of Phobos and Deimos in Chapter 5, all of the simulations in this thesis are conducted as near as possible to the J2000 standard epoch. For simplicity, the simulations described in this section begin at 0h00 UT on January 1, 2000 (12 hours before J2000), though this configuration could represent start times on any other date since the initial rotation of the planet repeats once every Mars sidereal day.

The ground tracks of the Mars Network satellites for the first four hours of the simulation are shown in Figure 4.1. At the initial epoch, three of the satellites are co-located at one point near the equator. Satellite 2 is located opposite these three, and the other two satellites are near the north and south poles respectively.

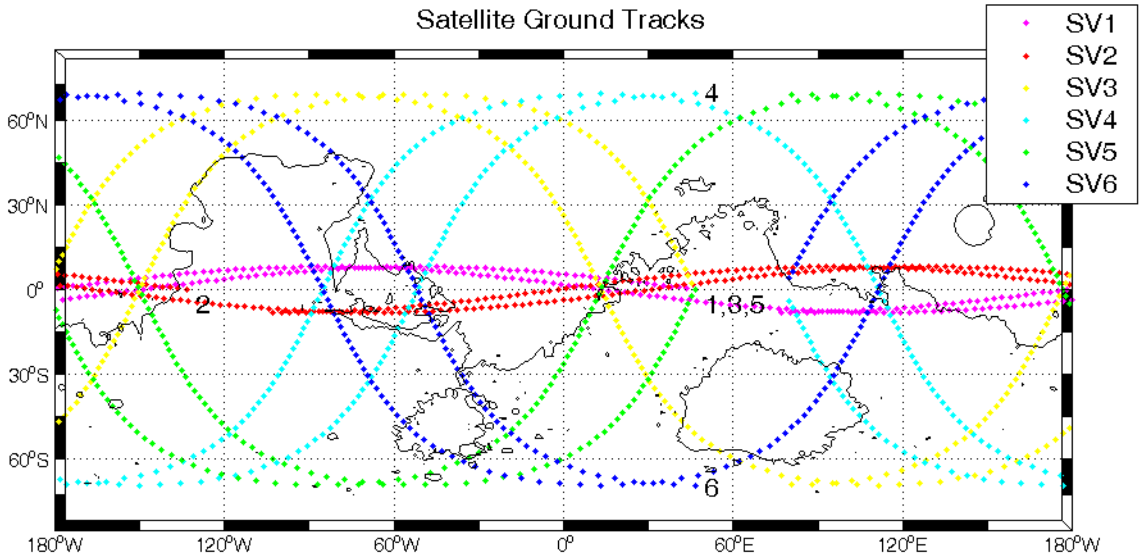

Figure 4.1: Ground tracks of the Mars Network satellites for 4 hours after the initial epoch. The satellites are in retrograde orbits (moving from east to west). The initial positions of each satellite are indicated on the map by the numbers 1 to 6.

4.1.2 Observation Scenarios

In the simulations described in this section, it is assumed that only the two-way range observable is used. Users on the surface are assumed to be stationary and are estimating their three dimensional position vector, $[x, y, z]^T$. The users are both attempting to compute their instantaneous positions, when sufficient observations are available, and are accumulating observations over time. Each observation contributes one row to the design matrix, $\boldsymbol{A}$, which has the form

$$\boldsymbol{A}_i = \left[-\frac{x_s - x_r}{\rho} \quad -\frac{y_s - y_r}{\rho} \quad -\frac{z_s - z_r}{\rho} \right] \tag{4.1}$$

where the s and r subscripted coordinates correspond to the satellite and receiver respectively and ρ is the range between the satellite and receiver.

Simulated users are positioned at five degree intervals in latitude and longitude around the planet. Each user attempts to observe all passing satellites once every three minutes. This observation interval is adopted from Ely et al. (1999) where it is stated that the Mars Network satellites will not provide continuous signals due to power limitations. The user receiver elevation cutoff angle is varied between the horizon and twenty degrees in five degree intervals.

4.1.3 Ground User Pass Geometry

The region on the ground where a satellite is visible is called a coverage circle. The radius of this circle depends both on the satellite altitude and the elevation cutoff angle being used by the receiver. (In principle, the size of the circle also depends on the beam width of the satellite antenna, but here it is assumed that the beam includes the entire geometric coverage circle.) The radius of the coverage circle as a function of the user elevation mask angle is shown in Table 4.2.

4.2 Availability Results

4.2.1 Instantaneous Availability

Figure 4.2 shows the instantaneous availability, or number of satellites in view, at the start of the simulation, after 15 minutes, after 30 minutes, and after 45 minutes, for a user with a 15 degree elevation mask. This figure shows an interesting effect

Table 4.2: Radius of Mars Network satellite coverage circle as a function of elevation mask angle.

Elevation Angle	Beamwidth	Coverage Circle Radius	Coverage Circle Radius
0°	54.0°	36.0°	2132 km
5°	53.7°	31.3°	1853 km
10°	52.8°	27.2°	1609 km
15°	51.4°	23.6°	1398 km
20°	49.5°	20.5°	1214 km

of the Mars Network constellation design. At the initial epoch, one of the two near-equatorial satellites is co-located with two of the near polar satellites. The other two near-polar satellites are at their most northerly and southerly latitudes respectively. For a very brief period, there are three satellites in view for users in this part of the equatorial region. However, 15 minutes later, the two near-polar orbiting satellites have ascended and descended, leaving only one satellite in view for most users, with smaller regions with two satellites in view and an even smaller region with three in view. After 30 minutes, the polar satellites have switched places and there are again three satellites in view at another location along the equator.

This initial phasing, with regions with three satellites simultaneously in view near the equator, will not last as the ascending nodes of the two near-equatorial satellites will drift at a different rate than those of the near-polar satellites. However, there will still be intermittent two-satellite coverage near the equator provided by pairs to near-polar satellites.

Figure 4.3 shows the percentage of time, over the course of one sol, that two or more satellites are visible to a user with a 15 degree elevation mask. With two satellites visible, instantaneous two dimensional positioning should be possible. Clearly, this capability is only available to near equatorial users, and then only briefly (less than 10% of the time). Using lower elevation angles, the percentage of the time, and the range of equatorial latitudes, for which instantaneous positioning is possible is increased, but it is still only available to low latitude users.

4.2.2 Satellite Passes

The number of satellite passes can be used as a representation of satellite availability and is used by Cesarone et al. (1999) as one measure to assess the communications

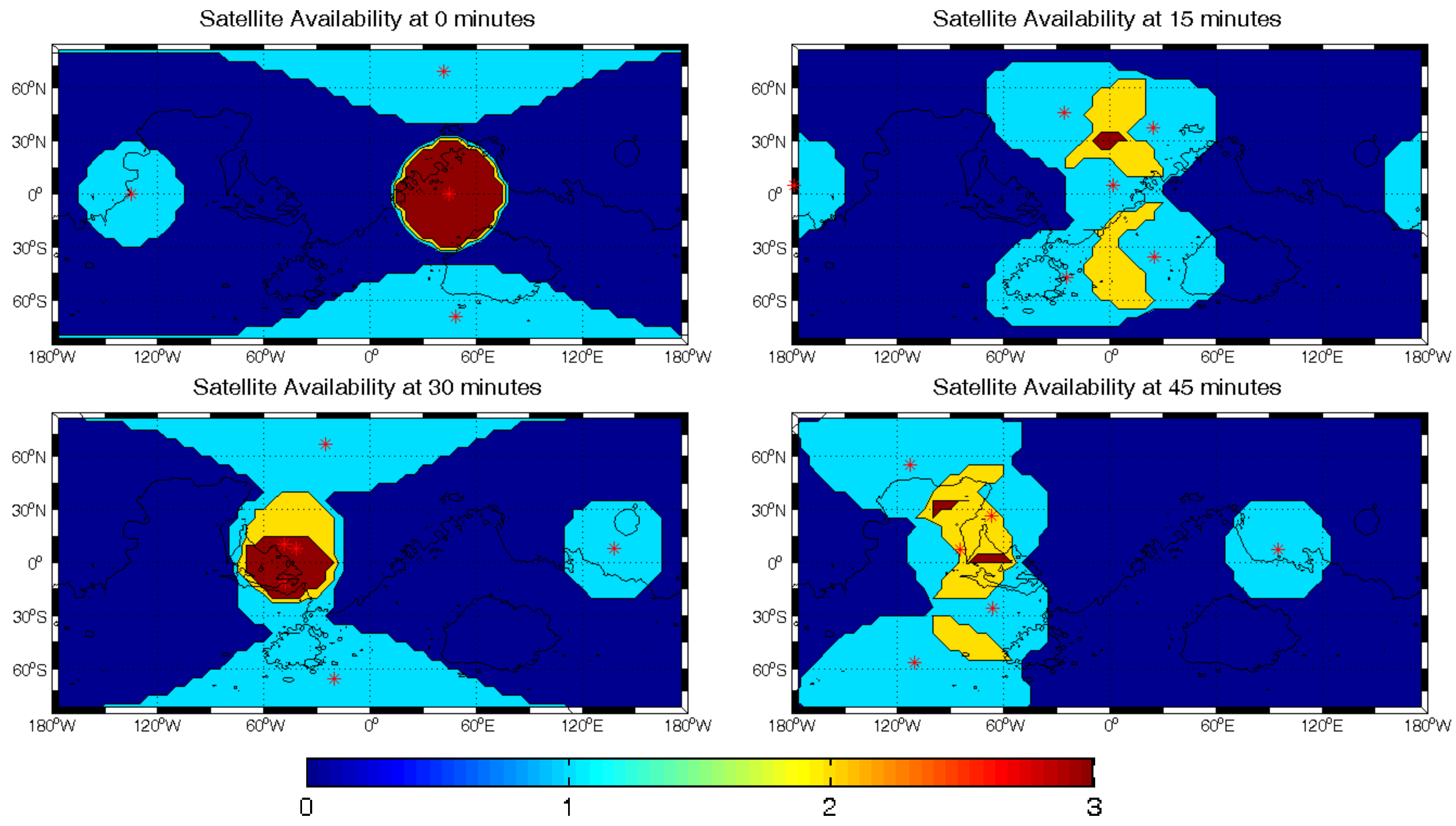

Figure 4.2: Instantaneous coverage at 3 epochs over approximately one quarter of an orbital period.

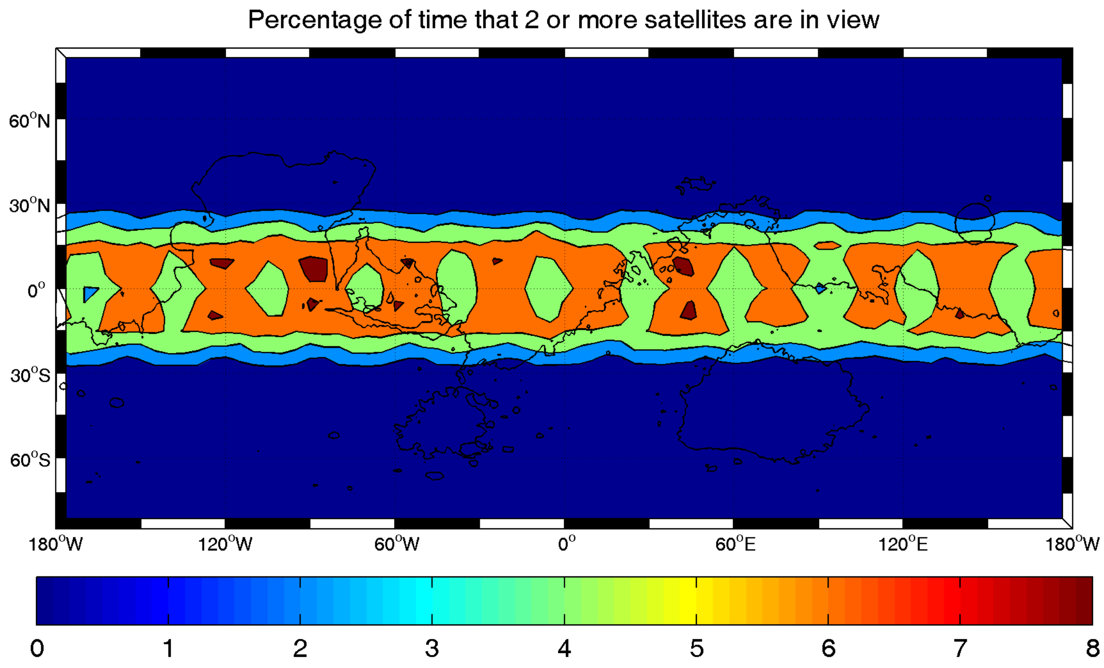

Figure 4.3: Map of the percentage of the time that 2 or more satellites are in view for a user with a 15° elevation mask.

capability of the Network. Here it is used to assess cumulative availability in a way that is not dependent on the observation rate. Figure 4.4 shows the minimum, mean, and maximum number of satellite passes per sol as a function of user latitude for the entire constellation, and for a single near-equatorial and a single near-polar satellite. Each satellite makes 11 orbits of the planet per sol. Every location on Mars receives between 14 and 44 satellite passes per sol with the best coverage available to low latitude and polar users. The two near equatorial satellites are visible on every pass to all users located below 15 degrees latitude, and are not visible at all to users above 35 degrees. The four polar satellites are visible to users at all latitudes, though the number of return visits per sol increases from around three per sol at the equator to 11 per sol at the poles. Users at 35 degrees have the poorest overall coverage, averaging 17 passes per sol. The results in Figure 4.4 assume the user is employing a 15° elevation

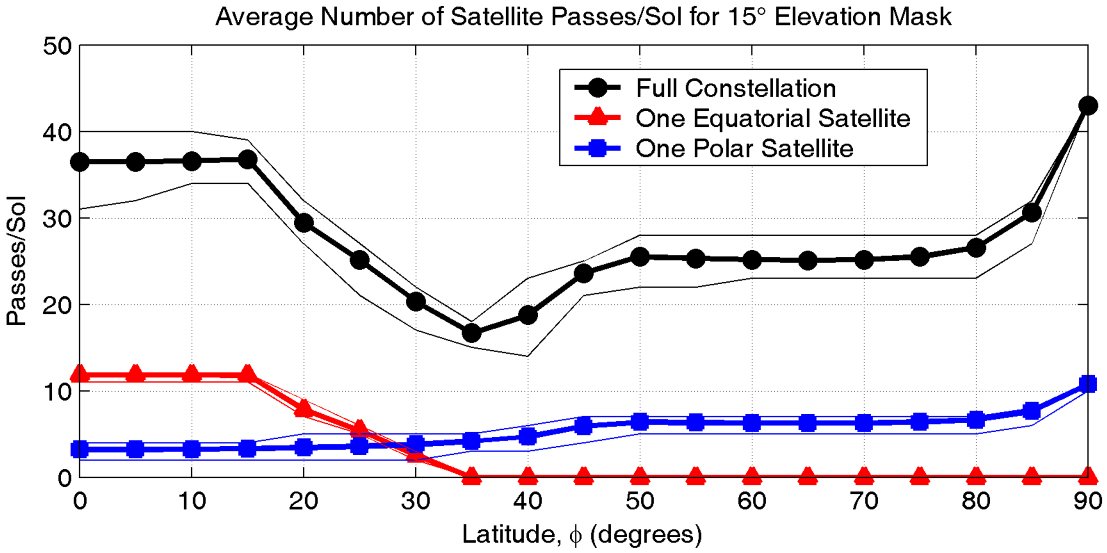

Figure 4.4: Mean number of satellite passes per sol as a function of latitude for the six satellite constellation, a single equatorial satellite only, and a single polar satellite only. In this case a 15 degree elevation mask is used. Maximum and minimum values are shown as the thin lines above and below the mean values.

mask which was the assumption used by Ely et al. (1999) to design the constellation. In GPS processing, a 15° degree elevation mask is a relatively conservative value aimed at reducing the effects of multipath and atmospheric errors. On Mars however, the atmospheric effects will be much smaller than on Earth so perhaps elevation angles below 15° should be considered. Expected levels of atmospheric errors on Mars are discussed in Section 5.4. Figures 4.5, 4.6, 4.7, and 4.8 repeat the results displayed in Figure 4.4 for user elevation masks of 0°, 5°, 10°, and 20° respectively.

As expected, the number of passes per sol increases as the elevation mask is decreased. In all cases but one, the smallest number of passes per sol occurs at the lowest latitude where the equatorial satellites can no longer be observed. The exception to this rule is the 20 degree elevation mask case, where the four near-polar satellites are no longer observed by users at the poles. The lack of coverage at the poles in this case is easily explained. The maximum latitude, $\phi_{\max}$, where a satellite

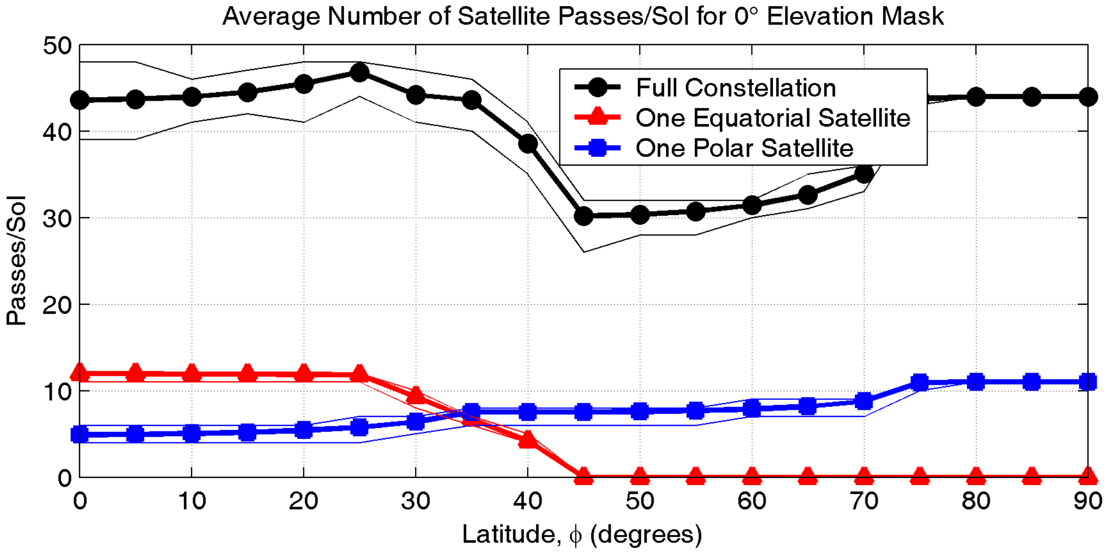

Figure 4.5: Mean number of satellite passes per sol as a function of latitude for a user with a 0° elevation mask.

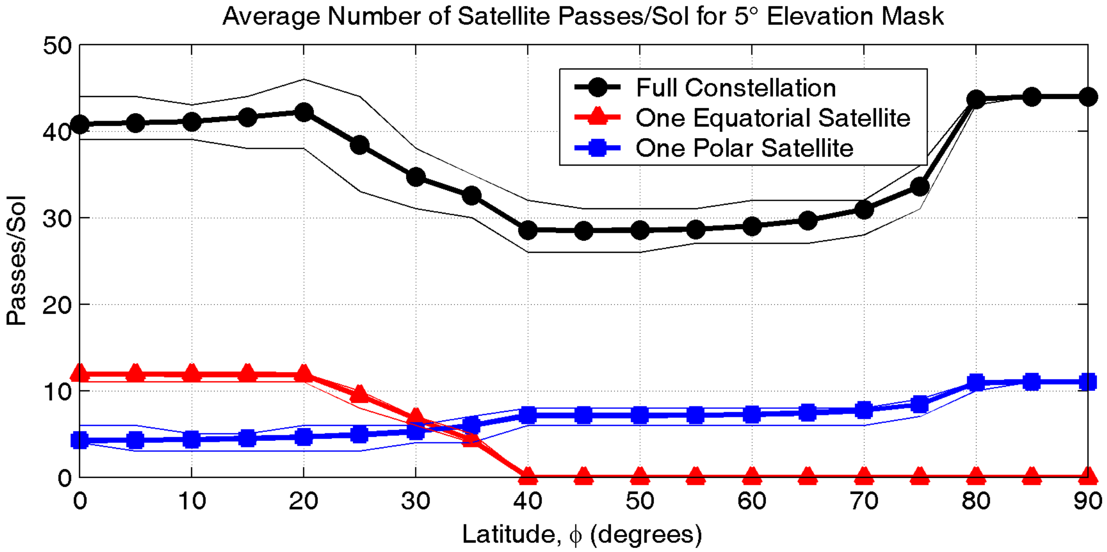

Figure 4.6: Mean number of satellite passes per sol as a function of latitude for a user with a 5° elevation mask.

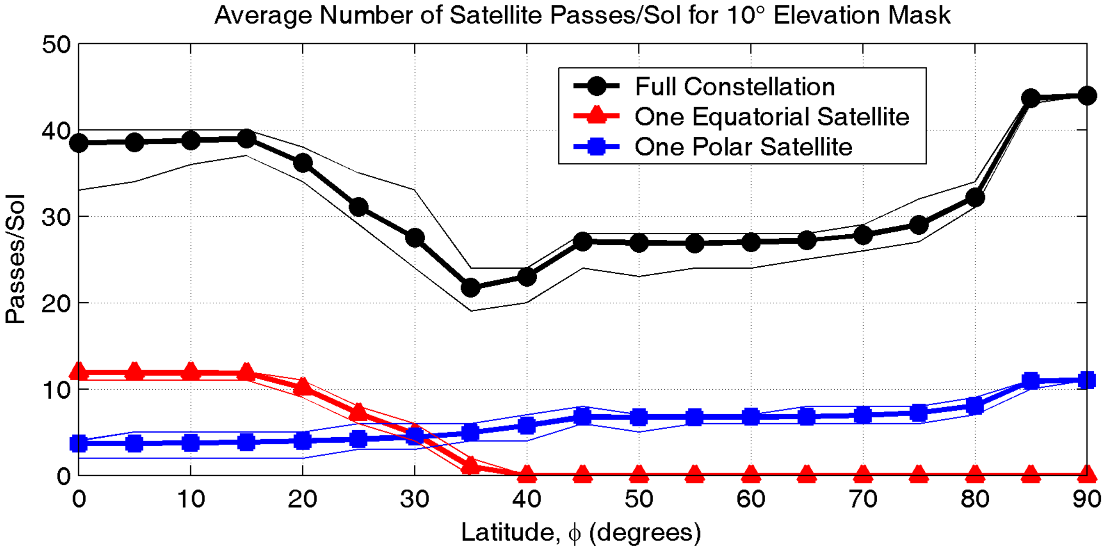

Figure 4.7: Mean number of satellite passes per sol as a function of latitude for a user with a 10° elevation mask.

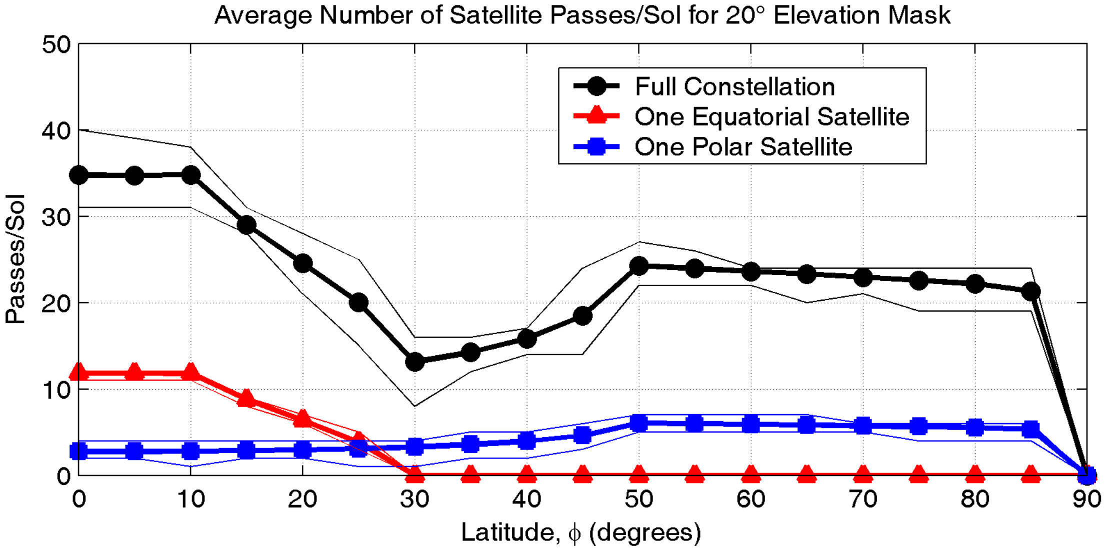

Figure 4.8: Mean number of satellite passes per sol as a function of latitude for a user with a 20° elevation mask.

is visible can be obtained by adding the radius of the coverage circle, r_{cc}, to the maximum latitude of the ground track. The maximum latitude of the ground track is given by the inclination of the orbital plane, or in this case 180° minus the inclination since the satellites are in retrograde orbits.

$$\begin{aligned} \phi_{\max} &= r_{\mathrm{cc}} + (180^\circ - i) \qquad (4.2) \\ &= 20.5^\circ + (180^\circ - 60^\circ) \\ &= 89.5^\circ \end{aligned}$$

This formula can also be used to determine the maximum latitude where the near equatorial satellites are visible.

4.2.3 Cumulative Availability

The mean number of observations per sol, assuming an observation rate of one observation every three minutes, as a function of latitude and user elevation mask angle is shown in Figure 4.9. Comparing this figure to Figures 4.4 to 4.8 shows that a typical pass results in four to seven observations, i.e. a typical pass lasts between 12 and 21 minutes with the pass length being generally longer when lower elevation masks are employed.

4.3 Accuracy Results

4.3.1 Instantaneous Dilution of Precision

The ability of the Mars Network to provide instantaneous position solutions is very limited as was shown in Figure 4.2 previously. Most of the relatively infrequent intervals where instantaneous three dimensional positioning is possible occur when three satellites cross paths above the user. In these cases, positioning geometry will

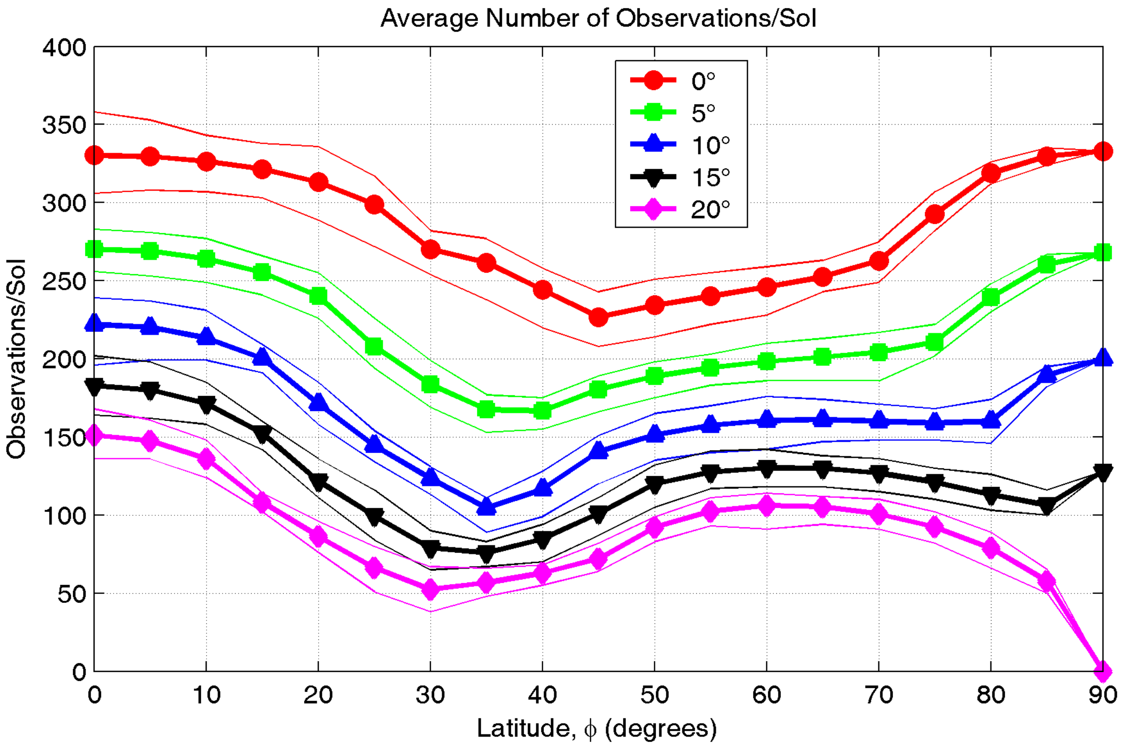

Figure 4.9: Average total number of observations during one sol as a function of latitude. A 3 minute observation interval is assumed.

be very poor. Even when using a 0° elevation mask, there are only very few epochs when 3 sufficiently separated satellites can be tracked to allow for instantaneous three dimensional positioning. Figure 4.10 is a map showing the percentage of the time that the instantaneous position dilution of precision is less than 100 for users with a 0° elevation mask.

4.3.2 Cumulative Dilution of Precision

Availability and pass statistics provide some indication of the global coverage of a satellite constellation, but do not assess the quality of the geometry of the navigation solution or provide an estimate of achievable positioning accuracy. For this, cumulative dilution of precision is used. Cumulative PDOP is computed by considering the contribution of all of a user's observations since the beginning of the simulation, assuming the user is making two-way range observations of all satellites available above a given elevation angle. Figures 4.11, 4.12, 4.13, and 4.14 show cumulative PDOP value after 1, 2, 3, and 4 hours respectively. The values plotted on these figures are obtained by averaging the cumulative PDOPs of all of the simulated users at a given latitude 1, 2, 3, and 4 hours into the simulation. In all four of the figures, there are ranges of latitudes for which the average PDOP value is off the scale. This indicates that at least one of the simulated users at that latitude has yet to compute a position at all due to a lack of observations. This is particularly evident for high elevation masks and short observation times. For example, in Figure 4.11, there is at least one user at at each latitude between 55° and 70° that has not computed a position at all, even when using a 0 degree elevation mask.

An alternative method to quantify cumulative PDOP performance is presented in Figures 4.15 and 4.16. Here the average times required for a user to collect sufficient

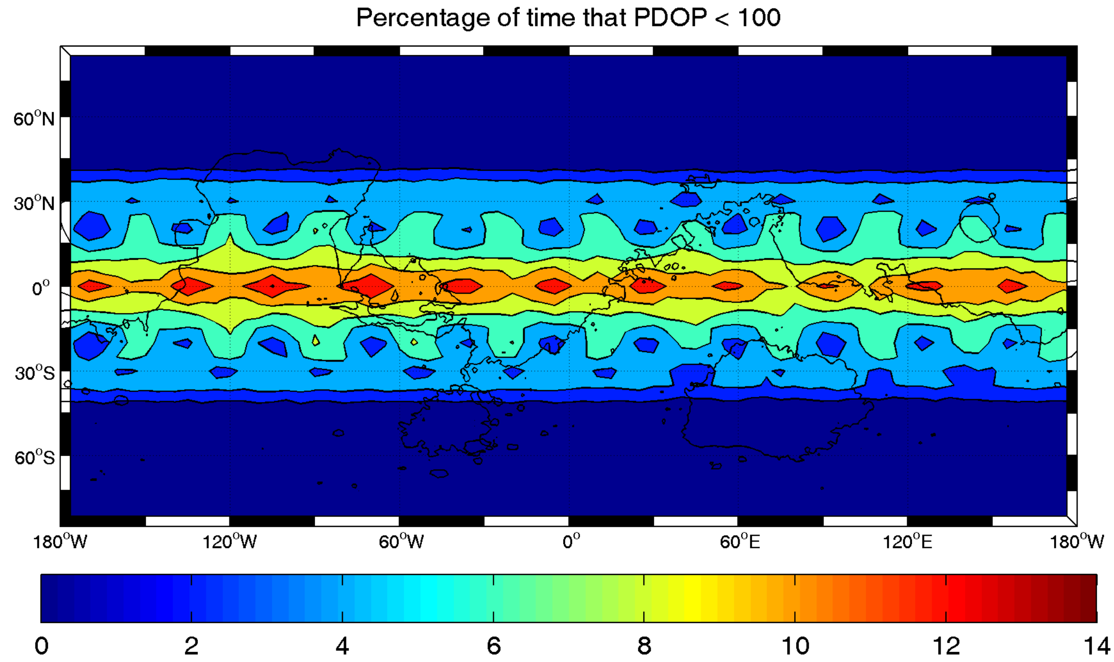

Figure 4.10: Map of the percentage of the time that PDOP is less than 100.

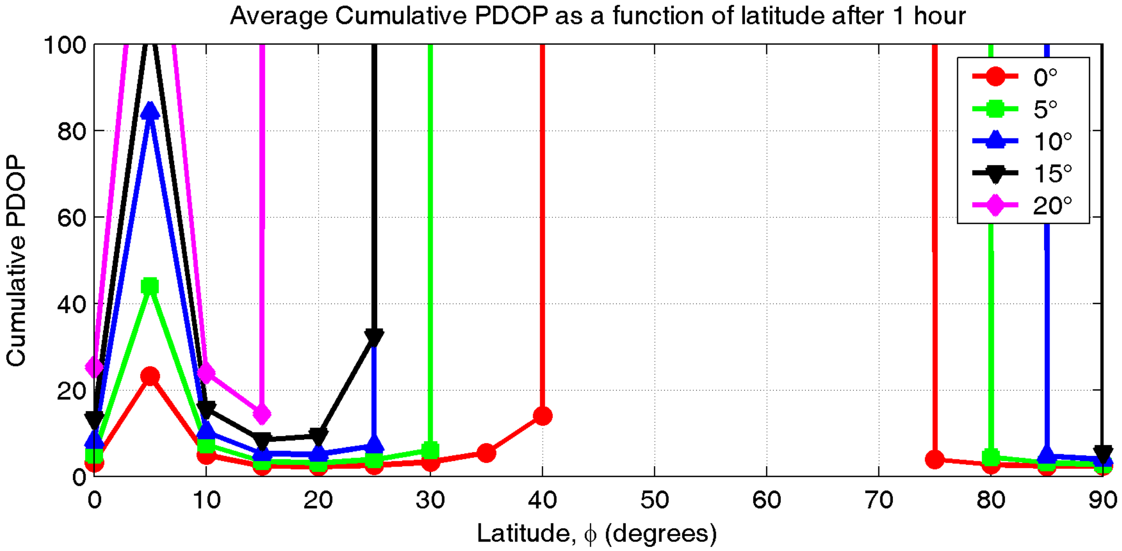

Figure 4.11: Cumulative dilution of precision as a function of latitude after 1 hour.

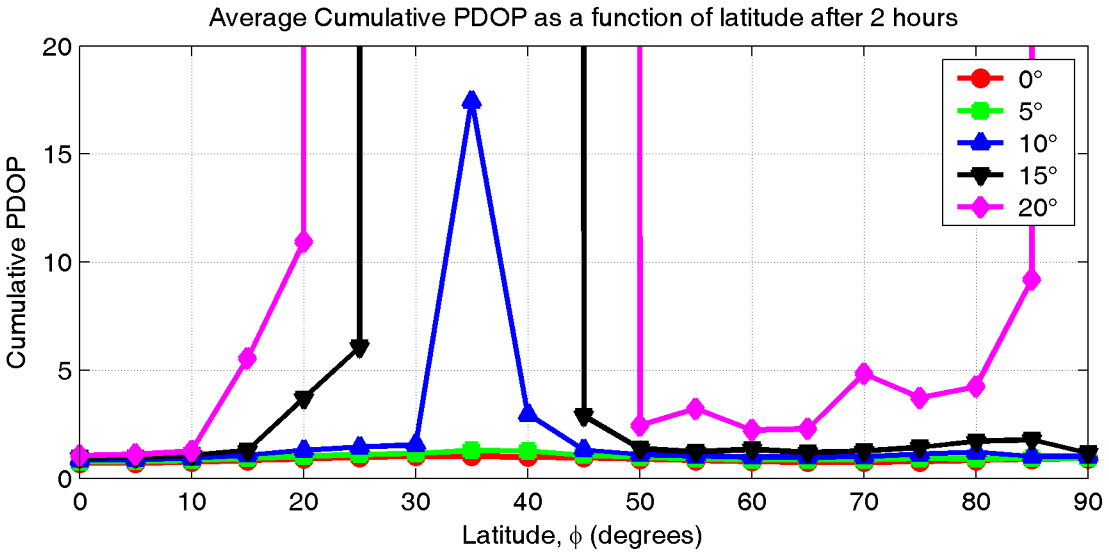

Figure 4.12: Cumulative dilution of precision as a function of latitude after 2 hours.

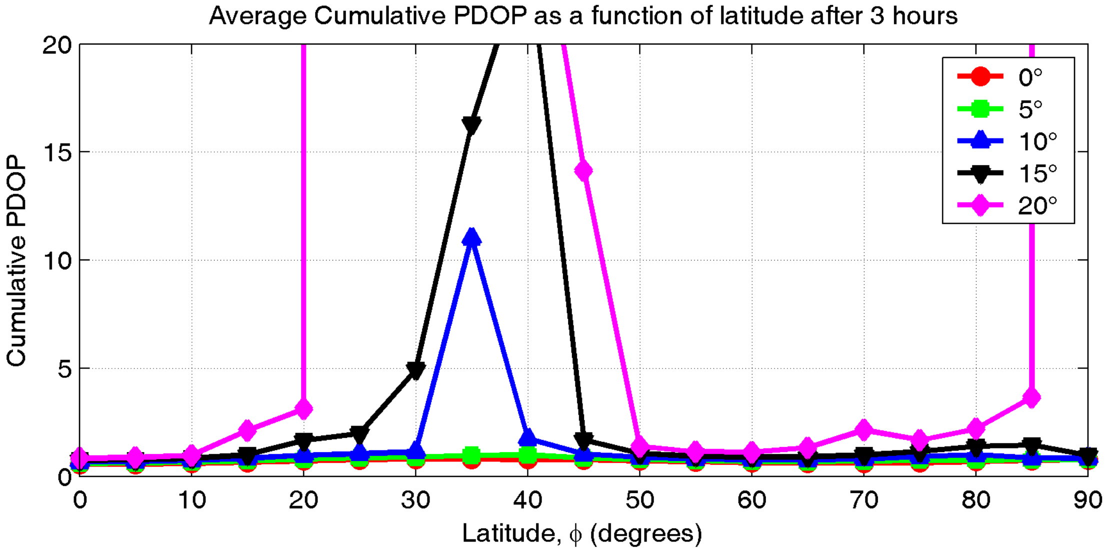

Figure 4.13: Cumulative dilution of precision as a function of latitude after 3 hours.

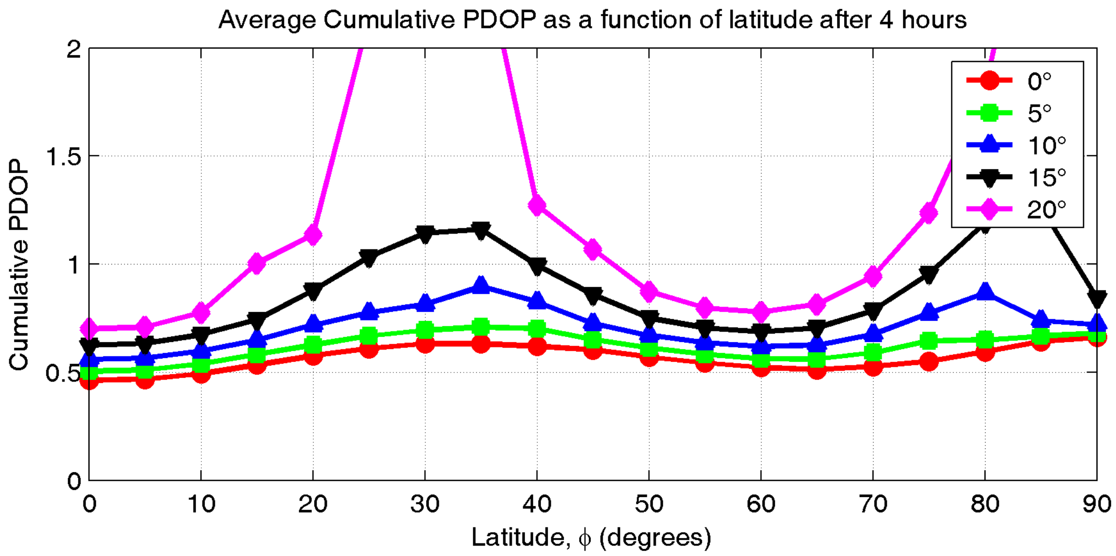

Figure 4.14: Cumulative dilution of precision as a function of latitude after 4 hours.

observations to obtain a cumulative PDOP of 100 and 10 is shown.

These results, along with a specified user equivalent range error, can be interpreted analogously to the mean time to fix statistic presented in Ely et al. (1999). For example, if a fix is defined as an MRSE of 100 m, and the UERE is specified as 10 m, then the time required to obtain a PDOP of 10 is equivalent to the time required to obtain a position with an MRSE of 100 m.

Similar to the number of passes and number of observations, users at 35 degrees latitude are the poorest served, requiring on average 60 minutes to reach a PDOP of 100 and 90 minutes to reach a PDOP of 10 when using a 15 degree elevation mask. This figure also shows the benefit of varied pass geometry on the time required to achieve a desired positioning accuracy. Polar users, which from Figures 4.4 to 4.8, make slightly fewer observations per sol than equatorial users, are able to reach low PDOP levels sooner because the satellite passes they do observe are at right angles to each other compared to the equatorial users who observe frequent East-West passes

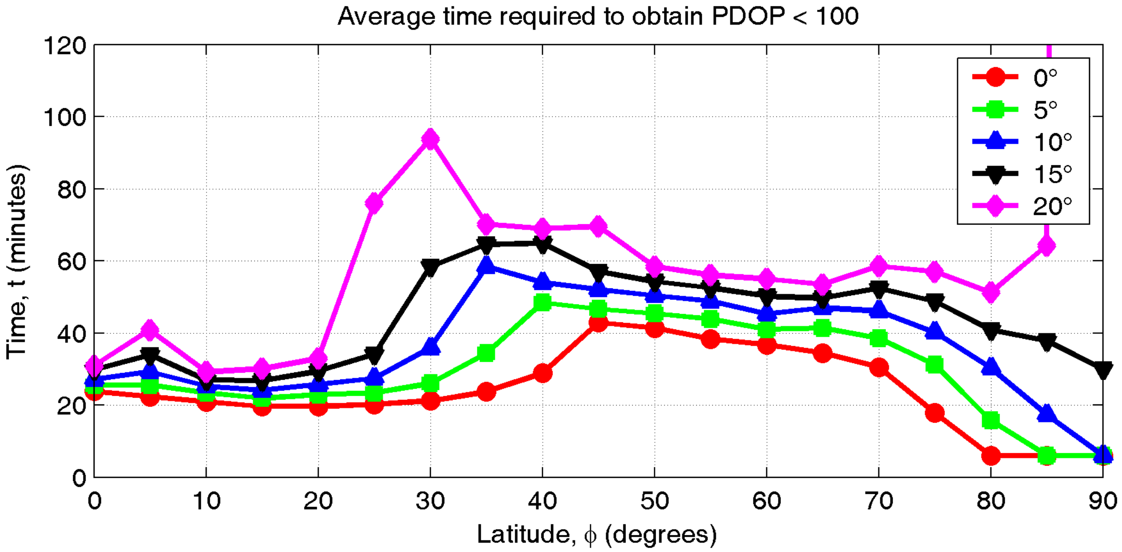

Figure 4.15: Average time required to obtain a cumulative PDOP less than 100 as a function of latitude.

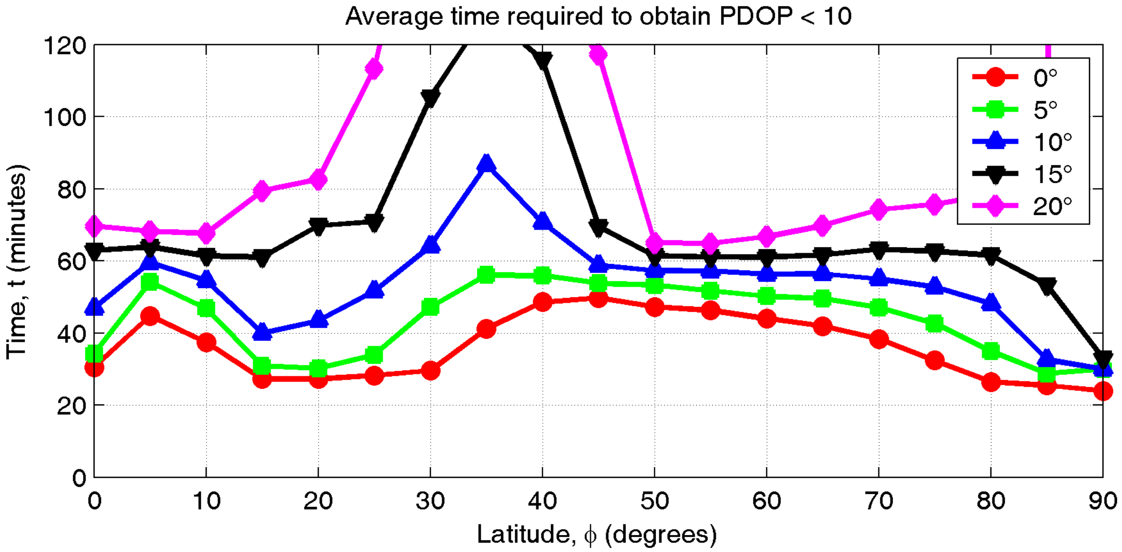

Figure 4.16: Average time required to obtain a cumulative PDOP less than 10 as a function of latitude.

of the near equatorial satellites with only occasional North-South passes of near-polar satellites.

4.4 Reliability Results

The fact that no more than three satellites are ever simultaneously in view to a ground user of the Mars Network means that the concept of instantaneous reliability is not applicable in this case. Reliability requires redundancy, and redundant observations are never available at a single epoch. However, it is possible to evaluate the reliability of the Mars Network for a user who collects observations over time in order to compute a solution.

As discussed in Section 3.1.3, reliability theory may be applied to a batch least squares or a sequential solution. The advantage of using a batch solution is that it is possible to detect blunders in all of the observations, not just those of the present epoch. The price for this capability is that all of the observations must be stored, as opposed to a sequential or filtered solution where information from the previous observations is contained in the state estimate and cannot be tested.

External reliability for the Mars Network was assessed by computing the three dimensional position error due to one marginally detectable blunder at a given epoch assuming that a sequential least squares solution method is being used. Sequential least squares is equivalent to Kalman filtering with no transition process and no process noise, which would be a reasonable data processing scenario for positioning a stationary ground user without storing previous observations. For all of the results in this section, a UERE of 10 m is assumed.

Figure 4.17 shows the maximum three dimensional position error (PE) due to one marginally detectable blunder as a function of time for a user with a 0° elevation mask located at the equator. The parameters α and β were set to 0.1% and 10% respectively, resulting in a non-centrality parameter, δ_0, of 4.57.

Note that PE can only be calculated while observations are being made since reliability at a specific time depends on the geometry of the observations at that time. Cumulative dilution of precision on the other hand simply remains constant while no new observations are being made.

The observation geometry dependence of reliability leads to an interesting result. The PE at the end of each satellite pass is consistently smaller than at the beginning of each pass, but in the middle of the pass, the PE is larger. Unlike PDOP, which will always decrease as new observations are added, reliability measures will not always decrease when new observations are added to a sequential adjustment. The marginally detectable blunder of a vertical range observation will be larger than that of a horizontal range observation if the user's vertical position is not as well known as the user's horizontal position. This is typically the case with satellite navigation systems. Another way to explain this effect is to consider observation redundancy. The more redundant an observation, the easier it is to monitor that observation for blunders. The redundancy of an observation is obtained with equation 3.9 and depends on the observation variance, the observation geometry, and the covariance of the estimated state. If the vertical coordinate is less well determined than the horizontal coordinates, then near vertical range observations will be less redundant, and thus less reliable, than those from lower elevations. Note that this effect is partially the result of all of the observations being assigned equal uncertainty in terms of UERE. Assigning all range observations the same variance is a common practice in GPS processing on Earth even though it is well known that lower elevation

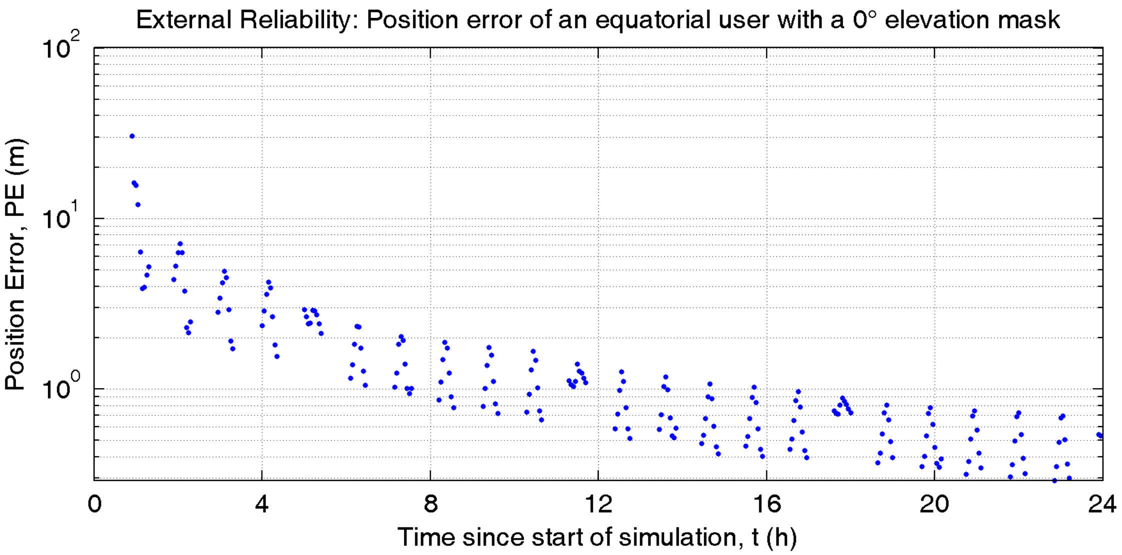

Figure 4.17: Maximum three dimensional position error (PE) due to one marginally detectable blunder for a user with a 0° elevation mask located at 0° N, 0° E making range observations every three minutes. Each satellite pass is visible as an inverted 'v' shape with the highest elevation observations being the least statistically reliable assuming all observations have the same observation variance.

observations are more susceptible to multipath and atmospheric errors.

Because reliability depends on current epoch observation geometry, it is difficult to produce generalized results for the whole planet, similar to those presented above for availability and accuracy.

Figure 4.18 is a map of the the maximum three dimensional position error (PE) due to one marginally detectable blunder after 4 hours of tracking for a user with a 0° elevation mask. For the regions of the planet where at least one satellite is visible, the current value of PE is displayed. For the rest of the planet, the value of PE due to the most recent observation is displayed.

Results of the other elevation masks are summarized in Figure 4.19. Similar to the dilution of precision results in Figure 4.14, external reliability is best for equatorial and polar users while the mid-latitudes are more poorly served.

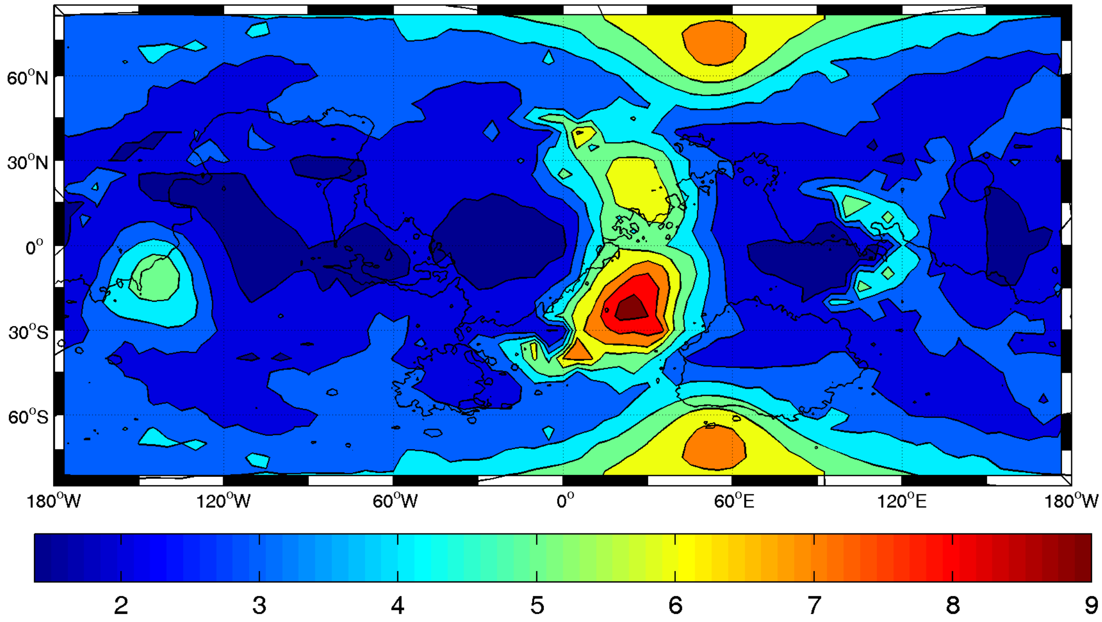

Figure 4.18: External reliability after four hours of tracking expressed by the maximum position error (in metres) due to one marginally detectable blunder in a most recent observation.

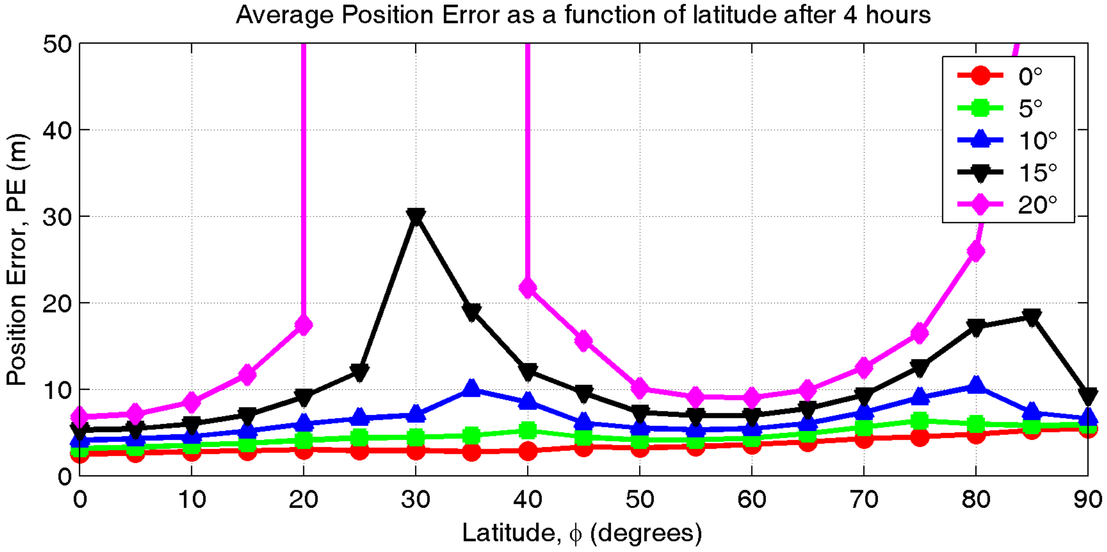

Figure 4.19: Average external reliability as a function of latitude and elevation mask angle after four hours. Similar to the availability and DOP figures earlier in the chapter, this figure shows the average value after 4 hours for all of the simulated users at a given latitude

The values of α and β (0.1% and 10%) that were used to produce the above reliability results were adopted from Ryan (2002) where he uses the reliability to assess Earth-based GNSS in the context of marine navigation. These values of α and β provide a very pessimistic assessment of reliability since they represent a requirement that 99.9% of the good observations are accepted while only 10% of the bad observations are allowed into the solution. A more optimistic estimate of the reliability can be obtained if both α and β are increased, which results in a decreased non-centrality parameter. If α and β are set to 5% and 20%, meaning that 95% of the good observations are accepted along with 20% of the blunders, the non-centrality parameter is reduced to 2.80. The MDB and external reliability measures will be correspondingly reduced. Figure 4.20 compares the average external reliability as a function of latitude after 4 hours for two values of the non-centrality parameter: 4.57 and 2.80. Of course, the results are not surprising, the external reliability measure is simply scaled proportionally to the change in the non-centrality parameter. The marginally detectable blunder depends linearly on the non-centrality parameter, and the external reliability (maximum position error) depends linearly on the marginally detectable blunder. All of the other reliability results in the previous sections could also be scaled similarly.

The cost of obtaining more optimistic reliability results is an increased probability of discarding good observations, which is something that should be avoided in the Mars Network considering how few observations are available to begin with.

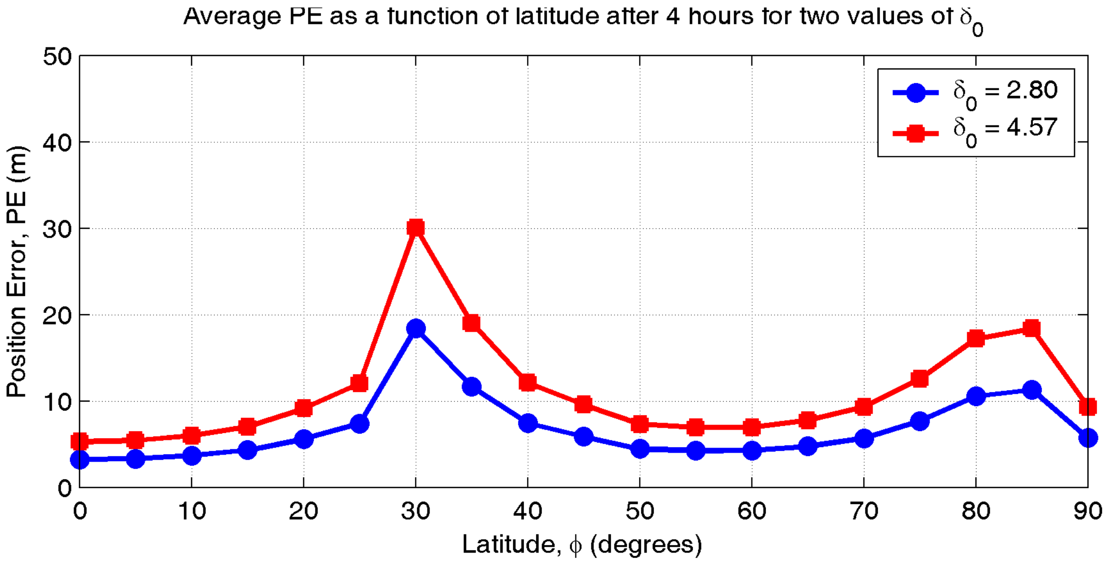

Figure 4.20: Average external reliability as a function of latitude after four hours for two values of δ_0. The red line indicates the position error computed using $\delta_0 = 4.57$. The blue line indicates the position error computed using $\delta_0 = 2.80$.

4.5 Accuracy and Reliability with the Addition of a Height Constraint

The addition of a height constraint will improve both instantaneous and cumulative accuracy and reliability results. The topography of Mars is very well known as a result of the extensive mapping mission conducted by the Mars Orbiting Laser Altimeter on board Mars Global Surveyor. According to Smith et al. (1999), the MOLA topography is accurate in height to within 13 m with a horizontal resolution of approximately 1 degree (or 59 km at the equator). This topography model could be used as a height constraint in order to allow for instantaneous positioning with only two two-way ranges. To quantify the improvement in instantaneous positioning that can be achieved by adding a height constraint, the simulation of instantaneous PDOP described in Section 4.3.1 was repeated with each simulated user employing a height constraint. The constraint was implemented as a pseudo-observation in the local zenith direction with an observation standard deviation of 20 m. Figure 4.21 is a map of the percentage of the time that PDOP is less than 100 for a height constrained user with a 0° elevation mask. Comparing these results to those of 4.10 shows a much larger range of latitudes for which instantaneous positioning is now possible in the case that only two simultaneous observations are required.

The addition of a height constraint does little for instantaneous reliability since there are still only rarely three satellites in view, the minimum required for an overdetermined solution.

4.6 Constellation Design Assessment and Recommendations

The design specifications of the Mars Network require global coverage with minimum variability in coverage as a function of latitude. From the results presented in this chapter it can be seen that these requirements have only been partially met. The system does provide global coverage for users with elevation masks below 20°. However, the mid-latitude regions are poorly served compared to the equatorial and polar regions. One possible solution would be to slightly increase the inclination of the two near-equatorial satellites. However, doing this would reduce the range of latitudes near the equator that are guaranteed coverage with every pass of the two near-equatorial satellites. The coverage variations could also be reduced by increasing the altitude of the satellites and thus increasing the size of their coverage circles. Unfortunately higher satellites require higher power signals and the principal investigators of the Mars Network have already rejected this option based on the need to balance communication and navigation performance with limited power transmitters.

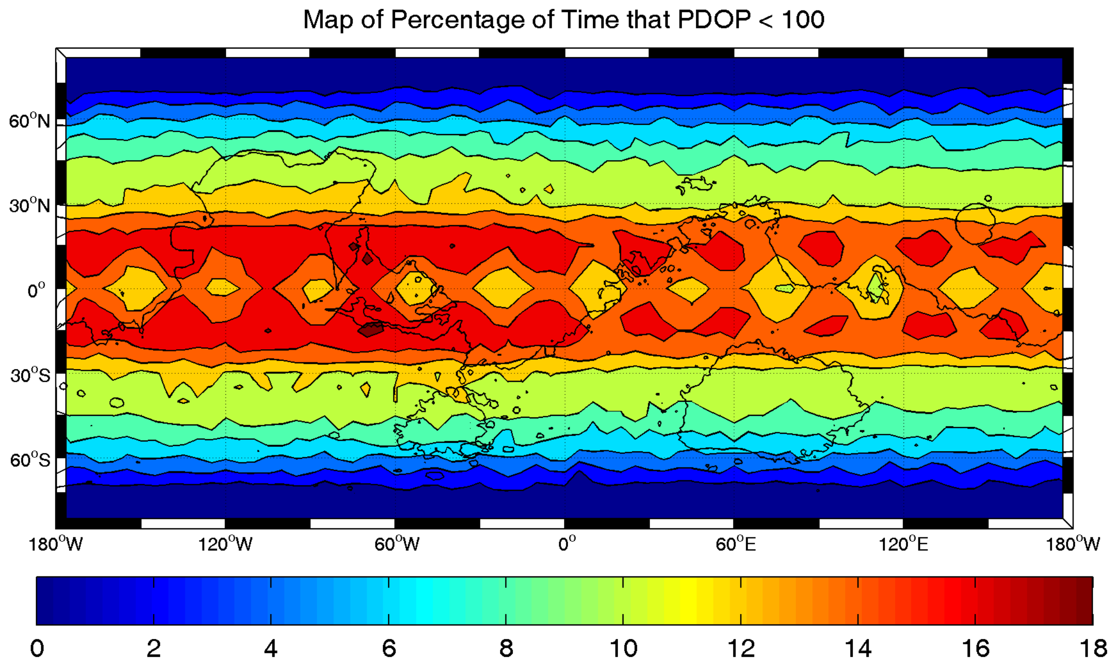

Figure 4.21: Map of the percentage of the time that PDOP is less than 100 for users employing a height constraint.

While instantaneous positioning has never been a goal of the Mars Network, it is still disappointing that the system, as proposed, is unable to provide this service for more users more of the time. Again, increasing the altitude of the satellites would help, but ultimately more satellites would be required which will not be feasible until there are more potential users on Mars.

Chapter 5

Numerical Constellation and Observation Simulations

The Keplerian plus secular drift approximation is a useful tool for modelling the performance of a satellite constellation over time, however, it is insufficient for producing simulated orbital errors when simulated observations are required. This is particularly true in the case of the Mars Network where the satellites are in relatively low orbits and higher order gravity perturbations are an important effect. In this chapter, a detailed numerical simulation of the satellite orbits will be presented. The goal of this simulation is to produce realistic trajectories of the Mars Network satellites and ground users. These trajectories will then be used along with observation and observation error models to produce a set of simulated observations with realistic error effects. This observation set will then be used to test a number of positioning algorithms and scenarios in Chapter 6.

5.1 Force Models for Mars Network Satellites

In this section, models for the most significant forces acting on the Mars Network satellites will be presented. These include: gravity, solar radiation pressure, and solar and Mars moon tidal forces. Less significant forces and affects that were not modelled, such as planetary albedo, other third body forces, and relativistic effects will also be discussed. In each case, the particular force is modelled so that the acceleration of the satellite due to that force may be included in the numerical integration of each satellite's equations of motion

$$\ddot{\boldsymbol{r}}(t) = f(\boldsymbol{r}, \dot{\boldsymbol{r}}, t) \tag{5.1}$$

which can be expressed as the system of equations

$$\dot{\boldsymbol{r}}(t) = \boldsymbol{v}(t) \tag{5.2}$$

$$\dot{\boldsymbol{v}}(t) = f(\boldsymbol{r}, \boldsymbol{v}, t) \tag{5.3}$$

where $\boldsymbol{r}$ and $\boldsymbol{v}$ are the position and velocity vectors respectively, and $f(\boldsymbol{r}, \boldsymbol{v}, t)$ is the sum of all the accelerations experienced by the satellite. Figures 5.1 and 5.2 show four hour-time series of the accelerations experienced by Mars Network satellites 1 and 3 due to each of the forces described below.

5.1.1 Gravity

Gravitational acceleration was computed as a function of position following the method outlined in Montenbruck & Gill (2000). The normalized coefficients from the GMM2B model described in Section 2.3.3, $\bar{C}_{nm}$ and $\bar{S}_{nm}$, were first unnormalized and then the acceleration was computed by evaluating the gradient of the gravitational potential

$$\ddot{r} = \nabla \frac{GM}{r} \sum_{n=0}^{n_{\max}} \sum_{m=0}^{n} \frac{R^n}{r^n} P_{nm}(\sin\phi)(C_{nm}\cos(m\lambda) + S_{nm}\sin(m\lambda)) \tag{5.4}$$

where $n_{\max}$ is the maximum degree of the gravity model, in this case 80, and P_{nm} are the Legendre polynomials. The gradient of the potential was evaluated using the Legendre polynomial recursion formula and partial acceleration equations from Cunningham (1970). This particular recursion is recommended by Montenbruck & Gill (2000) over the many other recursion formulas for Legendre polynomials for its numerical stability. Since the numerical integration of the satellite trajectory must be done in an inertial frame, and the gravity field is attached to the planet, the position vector is first transformed from the MCI frame to the MCMF frame and resulting acceleration vector is then transformed back to the MCI frame using the rotation, precession and nutation models described in Section 2.3.1.

5.1.2 Tidal Forces due to Mars' Moons and the Sun

The third body, or tidal, effects of the Sun, Phobos and Deimos are computed as

$$\ddot{\boldsymbol{r}} = GM\left(\frac{\boldsymbol{s}-\boldsymbol{r}}{|\boldsymbol{s}-\boldsymbol{r}|^3} - \frac{\boldsymbol{s}}{|\boldsymbol{s}|^3}\right) \tag{5.5}$$

where M is the mass of the third body, $\boldsymbol{r}$ is MCI position vector of the satellite and $\boldsymbol{s}$ is the MCI position vector of the third body. The second term is due to the fact that we want to know the acceleration in the Mars-centred inertial frame, not the frame where the planet is also accelerating due to the effect of the third body. The effects of third bodies other than the Sun, Phobos, and Deimos were neglected as discussed in Section 5.1.5.

This calculation requires Mars-centred ephemerides for the Sun, Phobos, and Deimos. A solar ephemeris was obtained using formulas given by Allison & McEwan (2000). Ephemerides for Phobos and Deimos were taken from Chapront-Touzé (1990).

Mars-Centered Solar Ephemeris

Allison & McEwan (2000) have developed Mars-centred solar ephemeris and solar timing formulas. These involve determining the location of the sun in a Mars-centred inertial frame by computing the position of Mars in a Sun-centred ecliptic inertial frame (defined by the plane of Mars' orbit and the Mars vernal equinox) and reversing the sign of the resulting vector. The resulting vector is then rotated into the

Mars equatorial inertial frame giving a position of the Sun in Mars-centred inertial coordinates.

Ephemerides for Phobos and Deimos

Chapront-Touzé (1990) has developed accurate semi-analytical theories for the orbital motion of the Martian moons, called ESAPHO and ESADE, for Phobos and Deimos respectively. In addition to complex ephemeris theory series expansions, Chapront-Touzé (1990) also provides state vectors (positions and velocities) for each moon, in the MCI reference frame, for the J2000 standard epoch. These values are reproduced in Table 5.1. Provided simulations are performed near the J2000 standard epoch, a Keplerian approximation of the ephemerides of Phobos and Deimos is sufficient to approximate the effect of Phobos and Deimos on the Mars Network constellation. The equivalent Keplerian elements to the state vectors in Table 5.1 are given in Table 5.2.

Table 5.1: Mars-centred inertial positions and velocities of Phobos and Deimos at the J2000 epoch (from Chapront-Touzé (1990)).

	Phobos	Deimos
x	-9412920 m	-19175920 m
y	1384400 m	-13500900 m
z	126000 m	-84860 m
v_x	-301.264 ms^{-1}	778.082 ms^{-1}
v_y	-2084.265 ms^{-1}	-1104.906 ms^{-1}
v_z	27.561 ms^{-1}	-20.940 ms^{-1}

5.1.3 Solar Radiation Pressure

Solar radiation pressure, $P_\odot$, is directly proportional to the solar flux.

$$P_\odot = \frac{\Phi}{c} \tag{5.6}$$

Table 5.2: Keplerian elements of Phobos and Deimos at the J2000 standard epoch.

	Phobos	Deimos
a	9378630 m	23457870 m
e	0.0147	0.00025
i	1.067 °	0.912 °
Ω	126.365 °	22.004 °
ω	216.394 °	-156.550 °
M	189.141 °	349.701 °
Period	27575.5 s (0.32 days)	109080.5 s (1.27 days)

According to King (2001), the average solar flux in the Mars orbit is 589.2 Wm^{-2} resulting in a solar radiation pressure of

$$P_{\odot} \approx 1.96 \times 10^{-6}\text{Nm}^{-2} \tag{5.7}$$

If it is assumed that the cross-section of the satellite, A, faces the sun at all times, the acceleration due to solar radiation pressure can be computed as

$$\ddot{\boldsymbol{r}} = -P_{\odot} C_R \frac{A}{m} \tag{5.8}$$

where C_R is the radiation pressure coefficient and m is the mass of the satellite. C_R was set to 1.21, which is a standard value for a typical solar panel (van der Ha & Modi, 1977). The mass of a Mars Network satellite was set to 73 kg based on information from Matousek et al. (1999) indicating that the total mass at launch of two Mars Network satellites would be 220 kg, and that two thirds of the launch mass would be propellant. An arbitrary cross-section area of 1 m^2 was assumed, which is consistent with a micro-satellite.

Unlike the gravitational and tidal accelerations, solar radiation pressure acceleration is not a continuous effect since it is not present when the satellite is in the shadow of the planet. To determine if solar radiation pressure is present, a cylindrical eclipse model was used. A more sophisticated model would account for the reduced solar radiation pressure during partial eclipses of the Sun, for example, when the Sun is rising or setting over Mars from the point of view of the satellite.

5.1.4 Drag

The drag force can be computed as a function of the satellite mass, m, drag coefficient, C_D, satellite cross-sectional area, A, atmospheric density, ρ, and satellite speed, v_r relative to the atmosphere.

$$\ddot{r} = -\frac{1}{2} C_D \frac{A}{m} \rho v_r^2. \tag{5.9}$$

In the case of the Mars Network satellites, this effect will be negligible due to the very small atmospheric density at 800 km altitude. Unfortunately, most publications related to Mars atmosphere do not describe its density above altitudes in the 100 km range. According to Keating et al. (1998), the density of the Mars atmosphere at an altitude of 135 km is approximately 10^{-9} kgm^{-3} and decreases approximately one order of magnitude every 15 km. At a height of 800 km, this would result in a density of 10^{-44} kgm^{-3}. At this level of density, the drag acceleration on a Mars network satellite would be on the order of 10^{-37} ms^{-2}. Consequently, the effect of drag on Mars Network satellites is assumed to be zero.

5.1.5 Higher Order Effects

Relativistic Effects

In this thesis, relativistic effects have been neglected. General relativity affects the rate of passage of time through an effect known as gravitational redshift. In the case of GPS, the clocks on the satellites run at a rate of approximately 45 parts per billion faster than clocks located on the geoid (Ashby & Spilker Jr, 1996). In the GPS satellites, this effect has been mitigated by applying a constant frequency offset

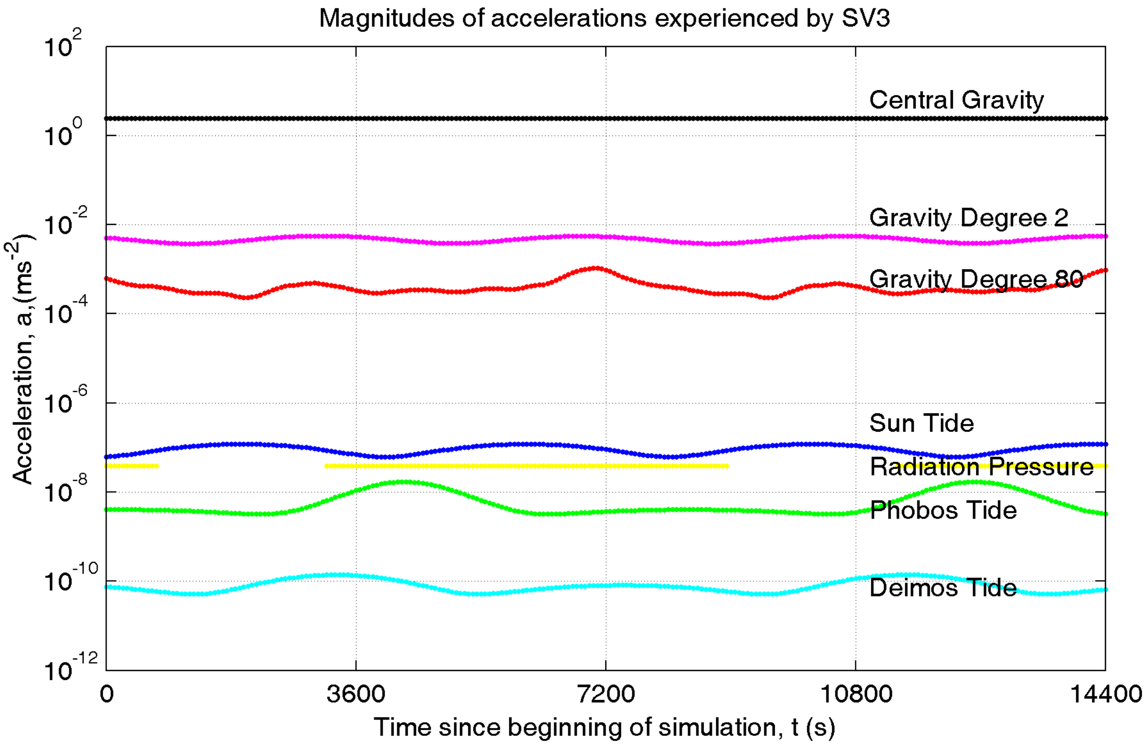

Figure 5.1: Accelerations experienced by Mars Network satellite 1 (a near-equatorial satellite) over a four hour period. Note the discontinuous solar radiation pressure acceleration.

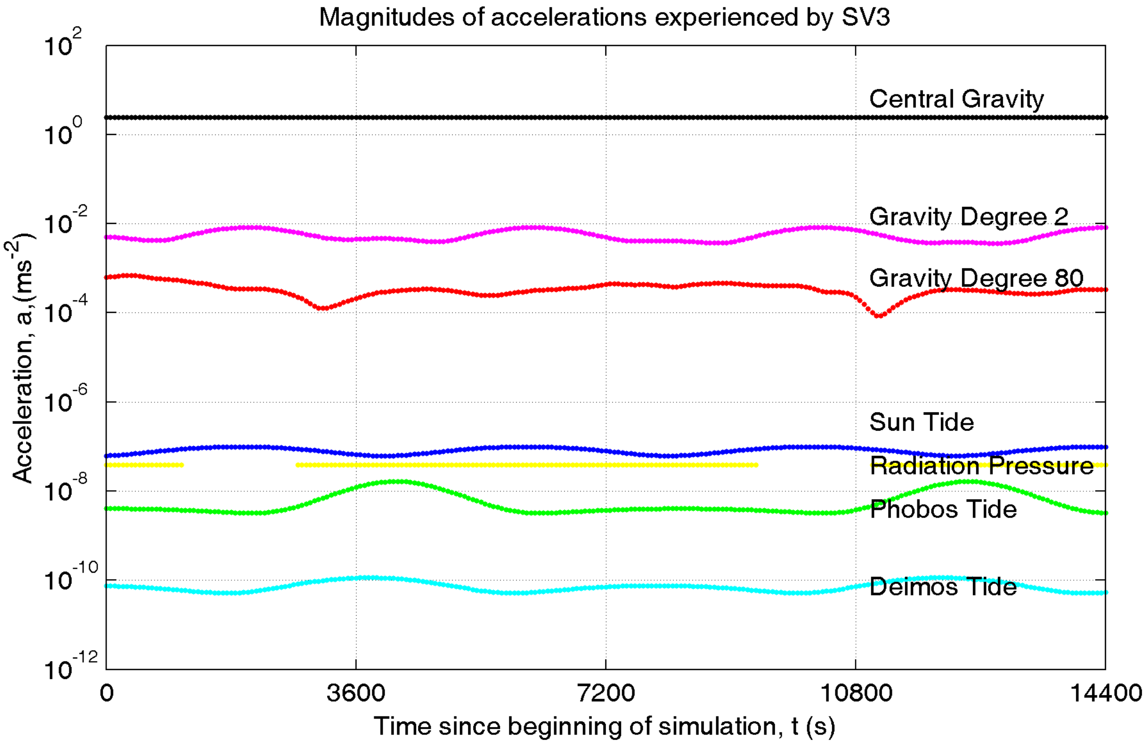

Figure 5.2: Accelerations experienced by Mars Network satellite 3 (a near-polar satellite) over a four hour period. Note the greater variability in acceleration from gravity terms of order two due to the more greatly varying latitude of this satellite.

to the satellite clocks to make them appear to be running at the nominal rate when observed from Earth. In this thesis, it is assumed that a similar procedure has been applied such that the clocks orbiting Mars will appear to run at the same rate as those on the surface.

General relativity also describes the gravitational bending of satellite navigation signals. According to Montenbruck & Gill (2000), a rough estimate of the size of this general relativistic effect on ranges can be given by the Schwarzschild radius

$$r = \frac{2GM}{c^2} \tag{5.10}$$

where G is the universal gravitational constant and M is the mass of the planet. r is approximately 1 mm in the case of the vicinity of Mars.

Second order time dilation effects also occur when satellites in eccentric orbits move within the gravitational potential of a planet. In GPS, a standard correction is applied by the user. For the Mars Network, this effect has not been modelled as it is assumed that the user will apply the correction.

In satellite navigation literature, the Sagnac effect is often considered when discussing relativistic effects. In fact, it is not a relativistic effect, but arises when calculations are made in a rotating reference frame. When viewed from an inertial frame, the Sagnac effect is simply the result of a stationary (in the rotating frame) receiver moving in the inertial frame in the interval between the transmit and receive times. (Ashby & Spilker Jr, 1996). This effect is avoided in this simulation by computing the range (difference between transmit and receive times) in the inertial frame.

Other Higher Order Effects

All other higher order effects have been neglected. These include Albedo (radiation pressure reflected from the planet), solid tide effects (time variations of gravity due to tidal deformation of the planet), transmit radiation pressure, and tidal effects due to other bodies in the solar system. For example, the effect of Jupiter, at its closest approach to Mars is several orders of magnitude less than the tidal effects of Phobos and Deimos.

It should be emphasized that the purpose of modelling the forces acting on the Mars Network satellites is to generate somewhat more realistic trajectories for the satellites compared to simplifications such as the Keplerian or Almanac representations. The goal is not to perfectly model the satellite trajectory, but to model typical orbital behaviour to provide a realistic orbital error when processing simulated observations. If the goal were to perfectly model the trajectories, the higher order effects would not be neglected, and models for the more important effects would be more sophisticated. This type of very precise modelling is required when attempting to estimate an exact orbit for scientific purposes. For example, Lemoine et al. (2001) model the orbit of Mars Global Surveyor using a detailed physical model of the spacecraft accounting for solar radiation and albedo pressure on different surfaces as well as all tidal effects from every body in the solar system. In their case, this is required as they are trying to reconstruct exact orbital arcs of MGS from sparse earth-based Doppler tracking data while simultaneously estimating the gravity field of the planet.

5.2 Generation of “Truth” Trajectories for Each Satellite

Using the time and position dependent force models described in the previous section, the trajectories of the Mars Network satellites can be simulated by numerical integration of their equations of motion. The motion of each of the satellites is an initial value problem, as all of the forces acting on the satellites are functions of the satellites present state only. Numerical integration was done using the variable-order variable-stepsize multistep algorithm DE developed by Shampine & Gordon (1975) and recommended for orbit determination problems by Montenbruck & Gill (2000). The input of the algorithm is the initial state (position and velocity), and a pointer to a function that computes the acceleration vector as a function of the present position, velocity, and time. This function then calls functions that compute the various forces acting on the satellite, adds up all of these forces, and returns the force vector divided by the satellite mass. Inertial and planet-fixed trajectories of each satellite are then output every 60 seconds for the duration of the simulation. Mars-fixed trajectories are computed by evaluating the rotation matrices (described in Section 2.3) that relate the inertial and Mars-fixed frames. Inertial trajectories for stationary landers are also computed so that simulated observations between landers and satellites can be constructed in the inertial frame.

5.3 Observation Models

Unlike GPS or TRANSIT, the Mars Network proposal does not specify continuously transmitting satellites. It does however provide for intermittent one and two-way ranging between satellites and users and between the satellites themselves.

5.3.1 One-Way Range

A one-way range observation is made when one transceiver receives the ranging broadcast of another. To model this observation, it is assumed that each transceiver in the network begins a transmission once every 3 minutes. The geometric range traveled is calculated by interpolating the simulated inertial coordinates of the transmitter and receiver at transmit and receive times respectively. The transmit time occurs on three minute intervals and the receive time (which depends on the range) is computed iteratively.

5.3.2 One-Way Range Rate

The one way range-rate is computed similarly to the one-way range. To compute the range rate, the velocity vectors of the transmitter and receiver, at transmit time and receiver time, are projected onto the range vector and added to give the instantaneous line-of-sight range rate.

5.3.3 One-Way Integrated Doppler Count

The integrated Doppler count, or phase observation, is the observation most often used for position determination in past and present Mars applications, for example, tracking the Mars Global Surveyor from Earth. However, integrated Doppler count requires continuous tracking in order to be a useful observable. Since the Mars Network proposal does not include continuously transmitting satellites, this observable has not been considered.

5.3.4 Two-Way Range

The two-way range observation is modelled assuming that the transmitter begins a transmission at an even 3 minute interval (to be received as a one-way range by the receiver). The receiver then coherently retransmits the signal at its receive time on a different channel. This signal is then received by the first transmitter. This model requires two positions to be computed iteratively. The turnaround latency is assumed to be zero for simplicity. In practice the turnaround latency is a small constant delay that would be determined for each channel on each transceiver before launch.

5.3.5 Two-Way Range Rate

The two-way range rate, or two-way Doppler frequency is modelled by determining the line-of-sight velocities of the transmitter at transmit time, the receiver at receive time and the transmitter at its receive time. Again, the turnaround latency is assumed to be zero and the coherent multiplication factors are assumed to be known.

5.4 Error Models

At this point, precise range and range rates between network elements have been computed. To make these ranges and range rates into realistic simulated observations, errors must be modelled and added. In Earth-based satellite navigation, the standard error sources are due to atmospheric effects, clock effects, orbital error, multipath, and noise. In this section each of these effects is considered in the context of Mars and a simple model for each effect is presented.

5.4.1 Orbital Error

In this case, it is not necessary to simulate orbital error by adding an error term to simulated ranges and range rates. This is because the ranges and range rates have been determined using a detailed model of the trajectory. This was not the case in some previous GNSS simulations (Luo, 2001; Alves, 2001), where ad hoc orbital error terms needed to be added to account for the fact that perfect circular or Keplerian satellite trajectories had been assumed.

In this thesis, orbital error, as a user effect, will arise when a user calculates a satellite position using a simplified representation of the orbit. For example, a user might have GPS almanac-like information about the satellites, or a previous position and velocity state and a simplified force model. The size of the orbital error will depend on the quality, detail, and age of the user's information about the satellite orbits. Figure 5.3 shows the magnitude of the position difference between the most sophisticated trajectory (the truth trajectory generated in Section 5.2) and less sophisticated models over a period of one day for Mars Network Satellite 1. Two simplified models are shown. First, the result of neglecting the tidal and radiation pressure terms, and second, the effect of also truncating the gravity model to degree and order 20. Further truncation of the gravity model results in more rapidly increasing orbital errors. In all cases, an exact knowledge of the satellite state at time = 0 is assumed. In general, the more simplified the orbit representation, the more quickly it will become unreliable. For example, the GPS ephemeris message is updated every two to four hours to keep the resulting orbital errors below specified limits.

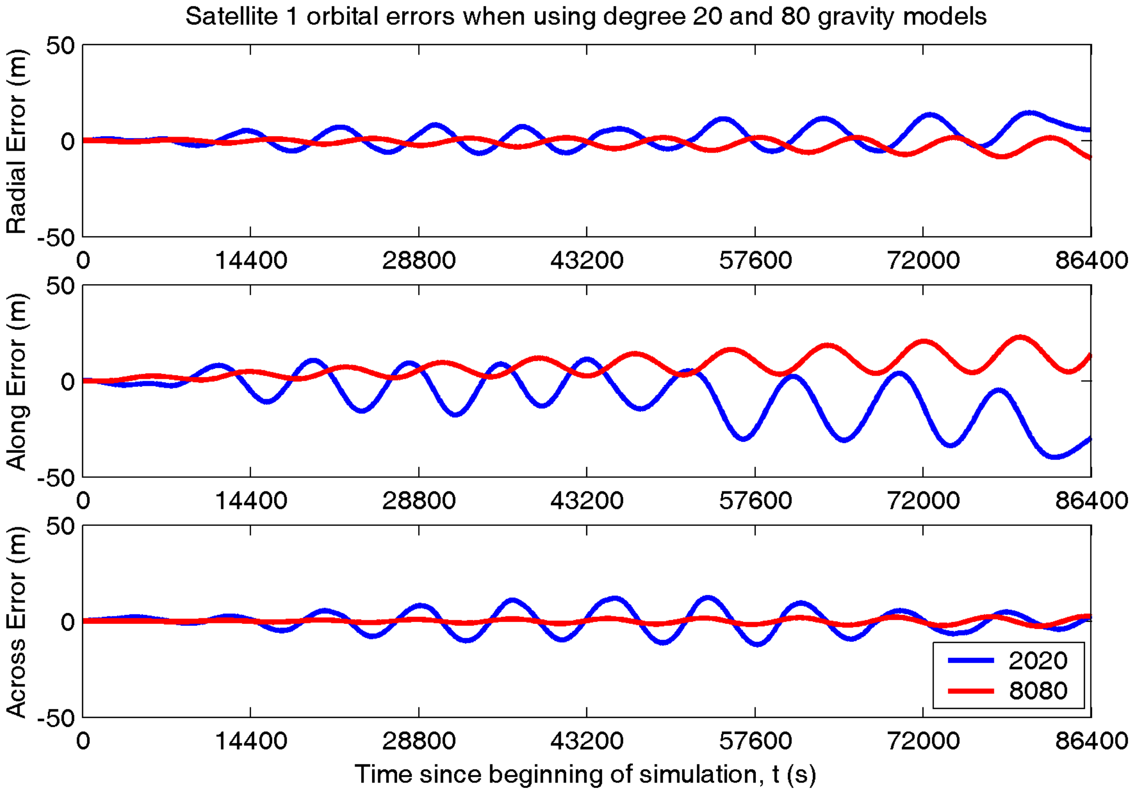

Figure 5.3: Magnitude of orbit error of Mars Network satellite 1.

5.4.2 Troposphere Error

Signals between Mars Network satellites and ground users may be affected both by the ionosphere and troposphere of Mars depending on the geometry of the signal path. Figure 5.4 illustrates the possible signal paths, including satellite-to-lander signals and satellite-to-satellite signals. Satellite-to-lander signals will be affected by both the Mars ionosphere and troposphere while the effect on satellite-to-satellite signals depends on the geometry. A model for simulating tropospheric delay is presented in this section. The ionospheric effect is discussed in Section 5.4.3.

In order to model the atmospheric delay of radio waves travelling in the atmosphere of Mars, a model of the refractivity of the atmosphere is required. Refractivity, N, is defined as the fractional change in refractive index, n, in parts-per-million.

$$N = (n - 1) \times 10^6 \tag{5.11}$$

This problem was addressed for the atmosphere of Earth by researchers working in the field of radar and microwave communication in the late 1940s and 1950s. Essen & Froome (1951), and many other researchers, published formulas for the refractivity of air as a function of temperature and the partial pressures of constituent gases of the Earth's atmosphere. Their development may be followed to produce a similar formula for the refractivity of the atmosphere of Mars.

To first order, Essen & Froome (1951) model the constituents of Earth's atmosphere by assuming they obey the ideal gas law. Based on this assumption, the refractivity as a function of pressure and temperature may be extrapolated from laboratory conditions as follows:

$$N = N_0 \frac{273.15}{1013} \frac{p}{T} \tag{5.12}$$

where N_0 is the refractivity at 1013 HPa pressure and 273.15 K temperature, and p

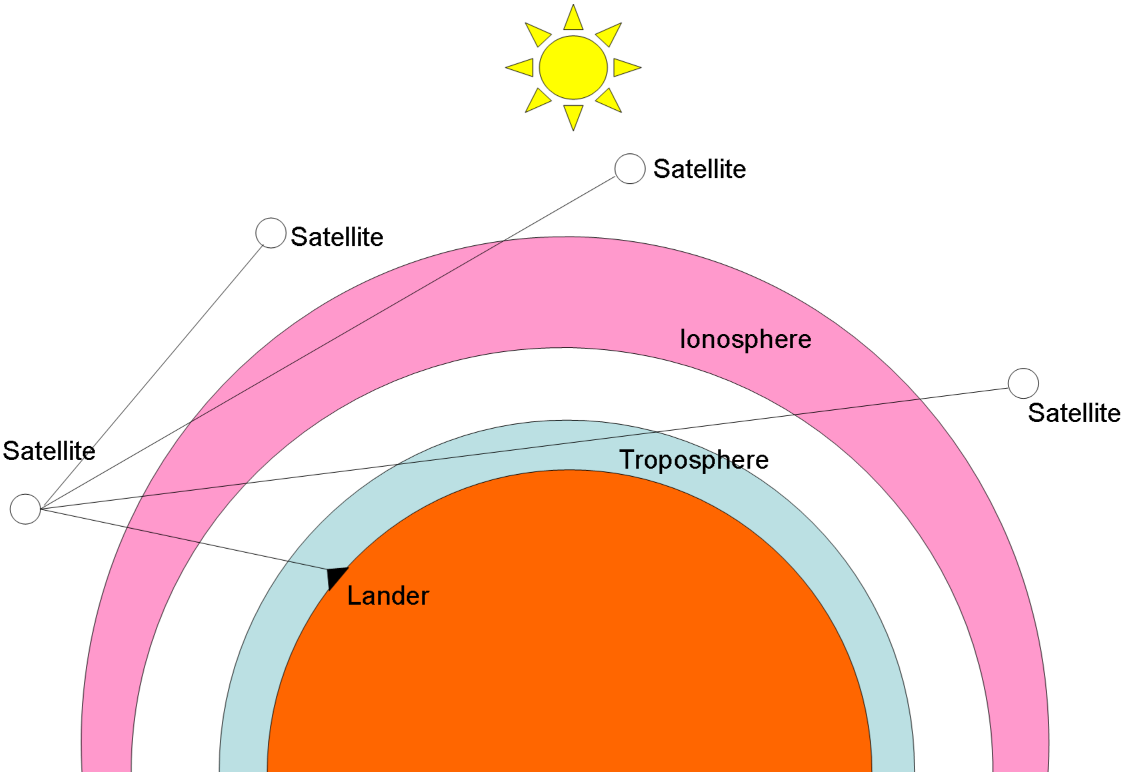

Figure 5.4: Satellite-to-satellite and Satellite-to-lander signal paths. Paths of signals involving the satellite at the left of the figure are shown. Signals may be affected by the ionosphere and the troposphere, depending on the geometry. The troposphere model developed in this section is constant everywhere on the planet while the ionosphere model depends on position of the elevation angle of the Sun. This figure is not to scale and ray bending is not shown.

and T are the pressure and temperature respectively. Assuming that the constituents the atmosphere of Mars may be treated as ideal gases, the refractivity of a Mars atmosphere at 273.15 K and 1013 HPa may be computed by forming a weighted average of the refractivities of each of the components. This value can then be extrapolated using equation 5.12 to arbitrary pressures and temperatures (i.e. pressures and temperatures that occur on Mars). Essen & Froome (1951) report refractivities for all the components of Earth's atmosphere, both from their own work and values published by others. Figure 5.5 shows the refractivity values they report for carbon dioxide, argon, nitrogen and oxygen, the four major components of the atmosphere of Mars. Essen & Froome (1951) and the other researchers they cite used different experimental methods at different frequencies, and produced results with varying error estimates. In this thesis, the refractivity of each gas at 400 MHz (the nominal frequency of the Mars Network), was determined performing a weighted linear fit to the observed refractivities and evaluating the linear fit at 400 MHz. The linear fits are shown as green lines and the 400 MHz values are shown as red dots in Figure 5.5. These results are also summarized in Table 5.3. The weighted average refractivity is 483 ppm, resulting in the following expression to compute the refractivity as a function of temperature and pressure.

$$N = 483.0 \times \frac{273.15}{1013}\frac{p}{T} = 130.2\frac{p}{T}; \tag{5.13}$$

To apply this formula, models for temperature and pressure on Mars are required. There are several published model atmospheres for Mars. For example, Benson (2000) has developed a model by fitting curves to profiles obtained by radio occultation techniques using the MGS orbiter. Other observation based models are presented in tables and figures in Lodders & Fegley Jr. (1998) and further more complicated

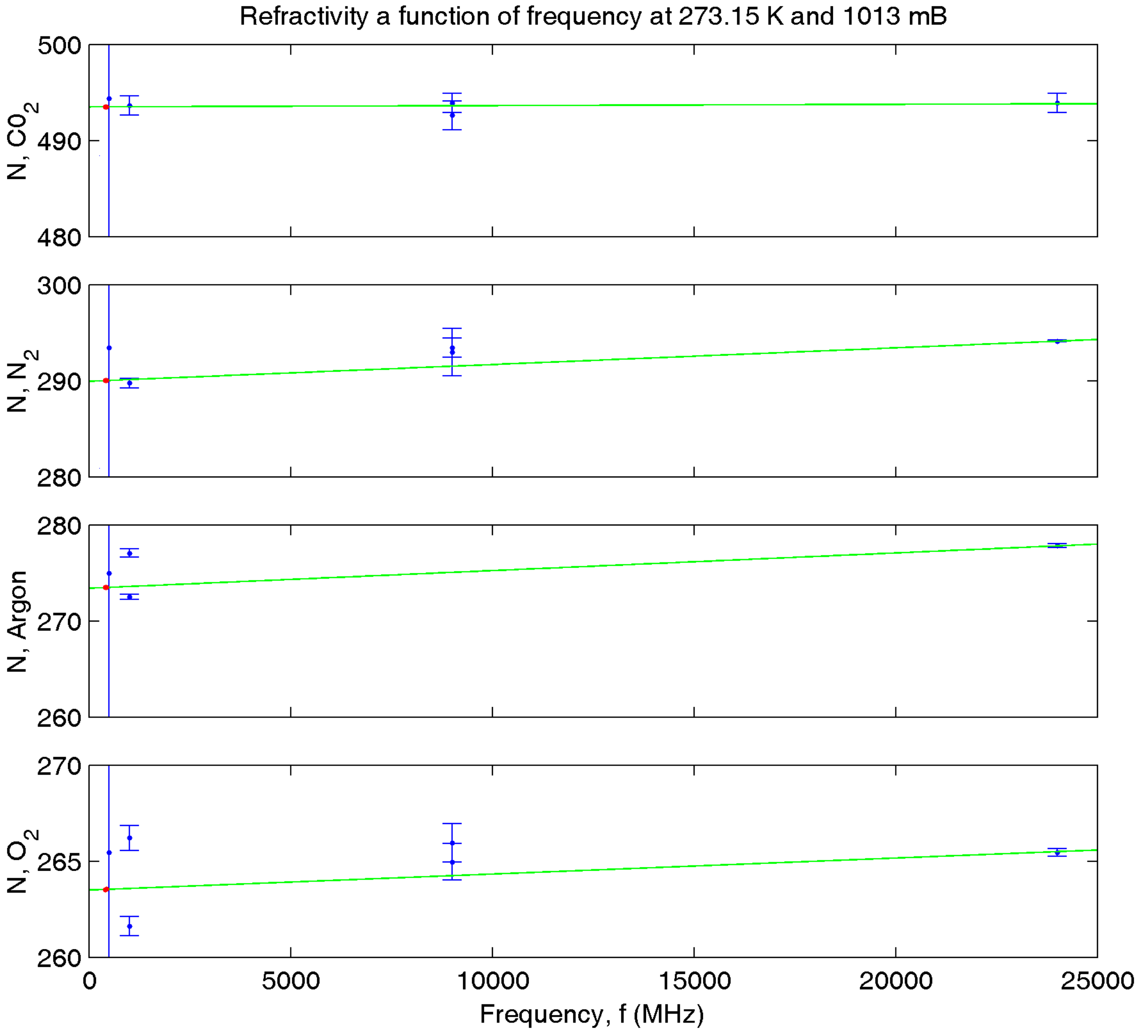

Figure 5.5: Refractivitities of carbon dioxide ($C0_2$), nitrogen (N_2), oxygen (O_2), and argon (Ar) at standard temperature and pressure, as reported by Essen & Froome (1951). The red dot in each subplot is the result of linear fit to the experimental results evaluated at 400 MHz, the nominal frequency of the Mars Network transceivers.

Table 5.3: Refractivities of the constituents of the Mars atmosphere at 400 MHz at 273.15 K and 1013 mB. The relative abundances, also shown in Table 2.2 are from King (2001) and do not add up to 100%. The missing 0.25% are other trace gases and it is assumed that they will not significantly affect the overall refractivity which is primary due to CO_2.

Gas	Refractivity	Abundance
Carbon Dioxide (CO_2)	493.5	95.32%
Nitrogen (N_2)	290.1	2.70%
Argon (Ar)	273.5	1.60%
Oxygen (O_2)	263.5	0.13%
Weighted Average	**483.0**	**99.75%**

models based on numerical modelling of the atmosphere have also been developed (Justus et al., 2002). In order to model the refractivity of the Mars atmosphere, a simplified version of the model atmosphere profiles from Lodders & Fegley Jr. (1998) is used. Figure 5.6 shows the temperature, pressure and resulting refractivity (calculated using equation 5.13). The simplified temperature profile was obtained by fitting a constant lapse rate to the temperatures from Lodders & Fegley Jr. (1998) up to 70 km altitude. The pressure profile is obtained from the surface pressure and the generally accepted assumption of a scale height of 11.0 km (King, 2001). With a model for refractivity in place, it is now possible to model tropospheric delays. Two possible approaches were considered. The first would be to determine the zenith delay and use a mapping function to map this value to various elevation angles. This is the standard method used in Earth-based GNSS, however, it was not selected for this work for two reasons. First, a suitable mapping function must be selected. The flat planet mapping function, $1/\sin(E)$, is poor choice for low elevation satellites, and many of the other standard mapping functions used on Earth contain empirical constants that make them unsuitable for use on Mars (at least until real observations are made and new constants are determined). The second problem with the mapped

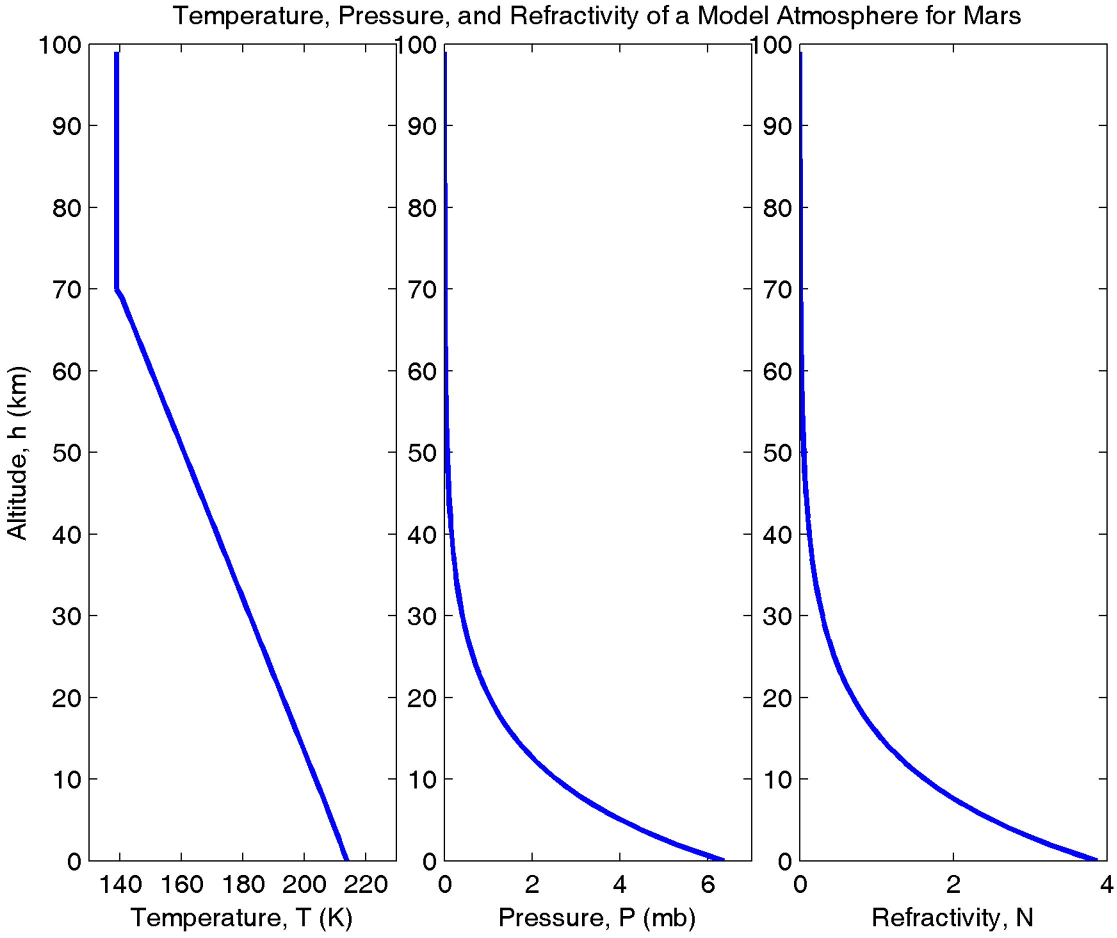

Figure 5.6: Temperature, pressure, and refractivity profiles for a standard Mars atmosphere.

zenith delay approach is that it is not suitable for determining the tropospheric delay on inter-satellite ranges.

In order to avoid both of these problems, tropospheric delay was estimated by tracing through a series of concentric spherical shells. Each shell was one kilometre thick with a constant refractivity, and the refractivity was assumed to be zero above 100 km altitude. The direct geometric path length, the curved path length, and the optical path length can then be compared to determine the tropospheric delay.

Figure 5.7 shows the resulting tropospheric delay (optical path length minus direct geometric path length) as a function of elevation angle for a user on the surface. Using this method, it is possible to model changes in path length due both to optical path delay and ray bending. In this case, effect of ray bending causes a maximum range error of approximately 2 mm for 0 degree elevation signals which is more than an order of magnitude smaller than the effect due to the delay. This is similar to the result observed with GPS signals on Earth where ray bending accounts for a maximum of 10 cm of 2 m to 6 m effect (Spilker Jr, 1996b).

This same method can be used to model the tropospheric delay for an inter-satellite range by ray tracing between the point of closest approach of the inter-satellite vector to the planet and each of the satellites. The tropospheric delay for an inter-satellite signal as a function of altitude of the point of closest approach is shown in Figure 5.8.

5.4.3 Ionosphere Error

Previous attempts to simulate ionospheric effects on GNSS signals for Earth have relied on the availability of detailed ionospheric models or observations. Luo (2001) used the global ionospheric maps produced by the Center of Orbit Determination

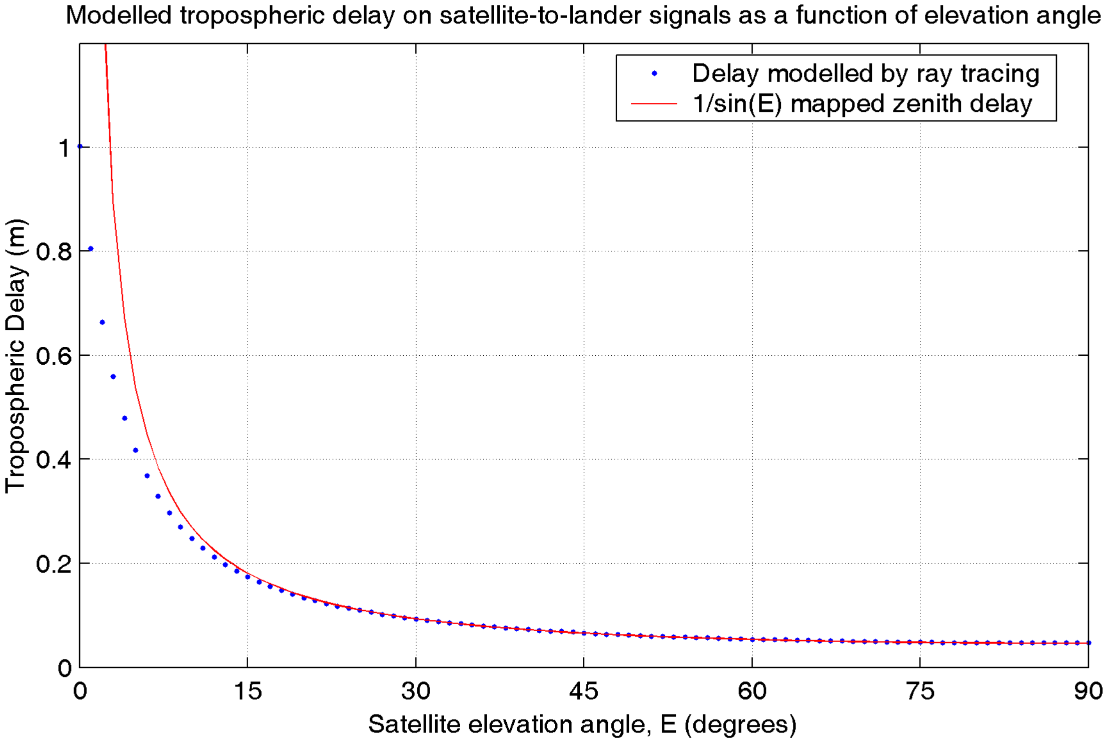

Figure 5.7: Tropospheric delay as a function of elevation angle for a user on the surface. The dashed line is the zenith value mapped using the $\frac{1}{\sin(E)}$ mapping function which agrees well with the ray tracing result for satellite above 15° elevation angles.

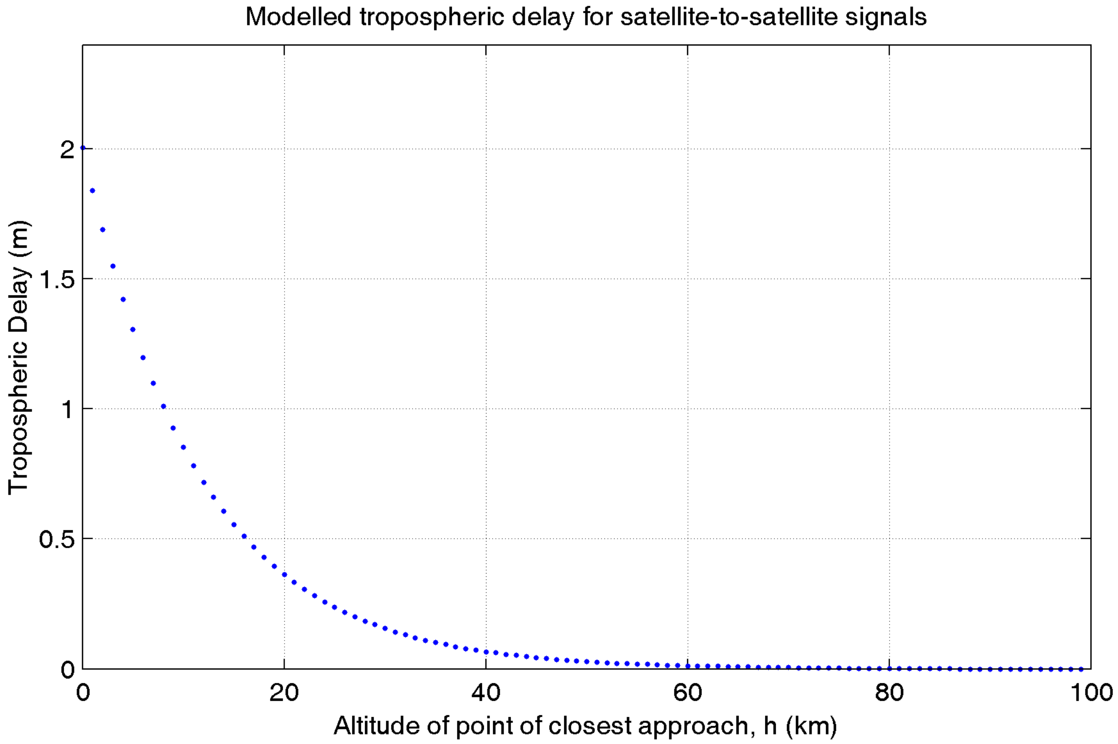

Figure 5.8: Tropospheric delay for an inter-satellite range as a function of the altitude of the point of closest approach of the inter-satellite vector to the planet. For a closest approach of 0 km, the effect is twice the value for a lander observation of a 0 degree elevation satellite observation.

in Europe (CODE) to provide a realistic global TEC model. Alves (2001) further refines this model by adding random time dependent spatially correlated components to simulate small regional scale variations in TEC. Both Luo (2001) and Alves (2001) calculate a simulated ionospheric ranging error by assuming the ionosphere to be a thin shell, computing the ionospheric pierce point, evaluating the ionospheric delay at the pierce point, and finally applying a mapping function based on the elevation angle of the signal path with respect to the model ionospheric shell.

The effect of the Mars ionosphere on ranging signals has only recently begun to be investigated. Mendillo et al. (2003) have investigated electron density profiles observed with the Mars Global Surveyor radio science experiment. They show that the Mars ionosphere exists at altitudes between 100 and 200 km with a peak electron density at around 135 km with a secondary peak at 110 km. The estimated Total Electron Content (TEC) in the region of the Mars Global Surveyor observations is approximately 0.7 TECU, or 0.7×10^{16} el m^{-2}. At GPS frequencies, this would represent a very small range error, however, the proposed operating frequencies of the Mars Network are in the 400 MHz band and are thus more affected by the ionosphere. A typical TEC value for Mars of 0.7 TECU would cause a range delay of

$$\Delta R = 40.3 \frac{\mathrm{TEC}}{f^2} = 40.3 \frac{0.7 \times 10^{16}}{(400 \times 10^6)^2} = 1.76\,\mathrm{m} \tag{5.14}$$

Mendillo et al. (2003) have developed a local (solar) time and latitude dependent model of TEC for Mars for which they present results (but not the model itself) for two latitudes, 24° and 67° N, averaged over the week of 24-31 December 1998 (near the Mars summer solstice). Their results are shown in Figure 5.9.

They do not report results for other latitudes or seasons as the focus of their paper is to compare their model results to radio occultation observations made of the Mars

Global Surveyor orbiter. However, a crude global model can be generated from these results by following the suggestion of Ho et al. (2002) that TEC will be primarily a function of solar elevation angle. Figure 5.10 shows the TEC values from Figure 5.9 plotted as a function of solar elevation angle. Figure 5.9 also shows the results of various functional fits to this data. Mendillo et al. (2003) suggest peak electron density is proportional to $\sin(E)^{0.55}$. Assuming that TEC is proportional to peak electron density, this function should provide a good fit. However, a simple quadratic function appears to better fit the limited number of data points they report. This model has been adopted for TEC as a function of solar elevation angle.

$$\begin{aligned} \mathrm{TEC}(E) &= -0.0001E^2 + 0.0150E + 0.2852 \quad , \quad E > -15^\circ \\ &= 0 \quad , \quad E < -15^\circ \end{aligned} \tag{5.15}$$

This model can be converted into a model of TEC as a function of position and time using the Mars-centred solar ephemeris discussed in Section 5.1.2.

With a model for TEC as a function of time and position, it is now possible to model the ionospheric error. For satellite to lander ranges, the effect is modelled by assuming the ionosphere to be concentrated in a thin shell at height 135 km. Equation 5.15 is evaluated at the ionospheric pierce point and the slant delay is then obtained by mapping the zenith value by $1\sin(E)$, where E is the elevation angle of the lander

Table 5.4: Functions fitted to TEC as a function of solar elevation angle. The third column is the standard deviation of the residuals of the fit. By this measure of goodness of fit, the quadratic function best fits the very limited amount of data.

Function Name	Functional Form	σ_r (TECU)
Linear	$0.0083 + 0.3363E$	0.0802
Quadratic	$0.2852 + 0.0150E - 0.0001E^2$	0.0243
Sin(E) + Constant	$0.2969 + 0.6521\sin(E)$	0.0424
Square root of Sin(E))	$0.1436 + 0.7358\sqrt{\sin(E)}$	0.0302

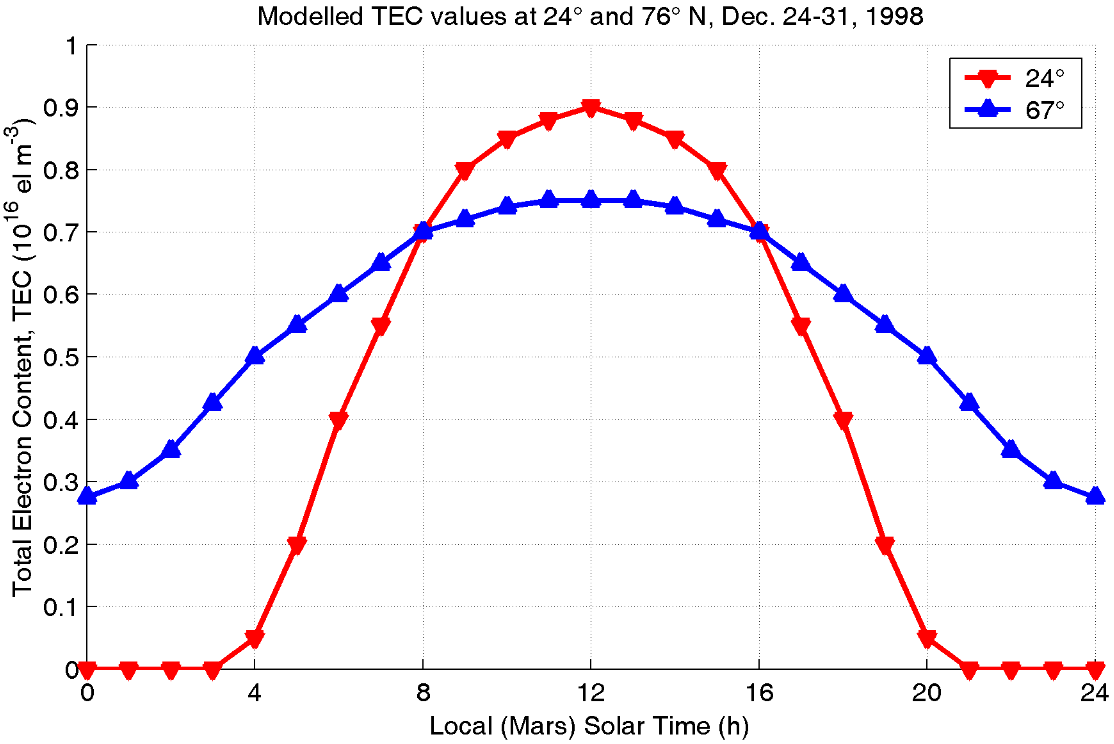

Figure 5.9: Average modelled total electron content as a function of local (Mars) solar time, for two locations at 24° and 67° N latitude during the week of December 24–31,1998, as reported in Mendillo et al. (2003).

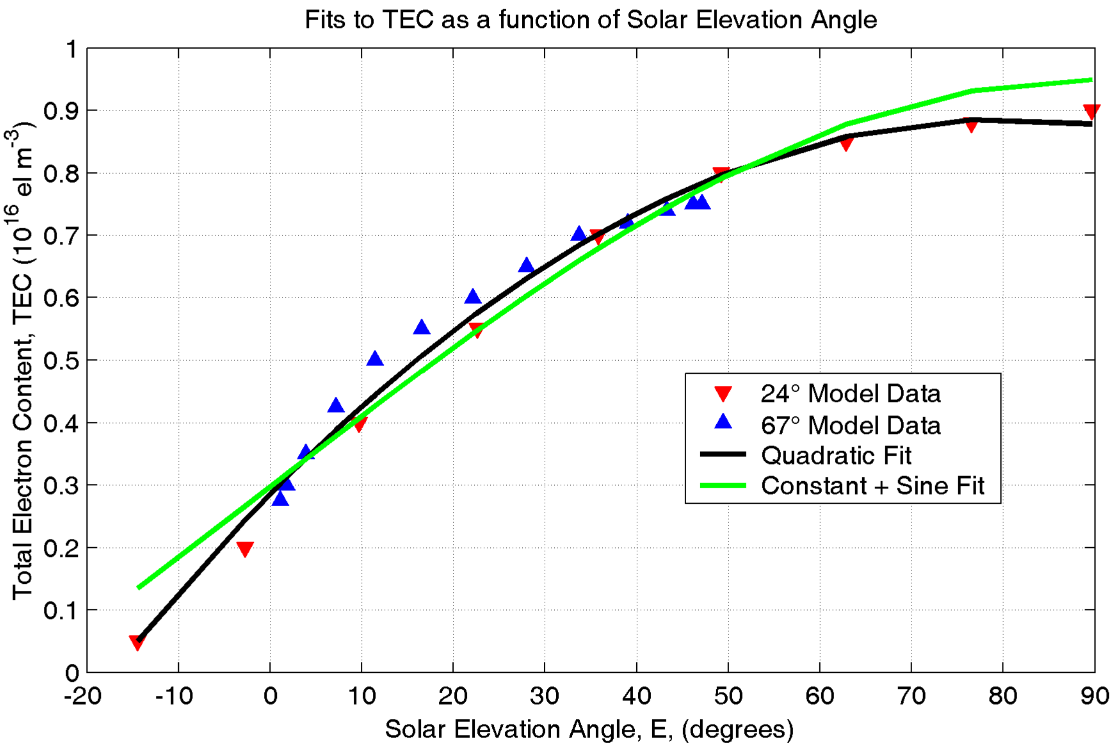

Figure 5.10: TEC values from Figure 5.9 as a function of solar elevation angle. The quadratic and sine + constant fits are also shown.

to satellite vector with respect to the ionospheric shell. Modelled ionospheric delay for a satellite-to-lander range as a function of elevation angle is shown in Figure 5.11.

Modelling the ionospheric effect on a satellite-to-satellite signal is a more complicated procedure. The simplest approximation is to use the same model as for the satellite-to-ground signal path, only in this case there may be two, one, or no intersections between the signal path and the ionospheric shell. This approach would clearly not provide a realistic delay for signals passing above the altitude of the ionospheric shell and would result in an infinite value for signals that are tangent to the shell.

To account for this effect, the same ray-tracing procedure used to compute the tropospheric error is applied to the ionospheric case. First, the point of closest ap-

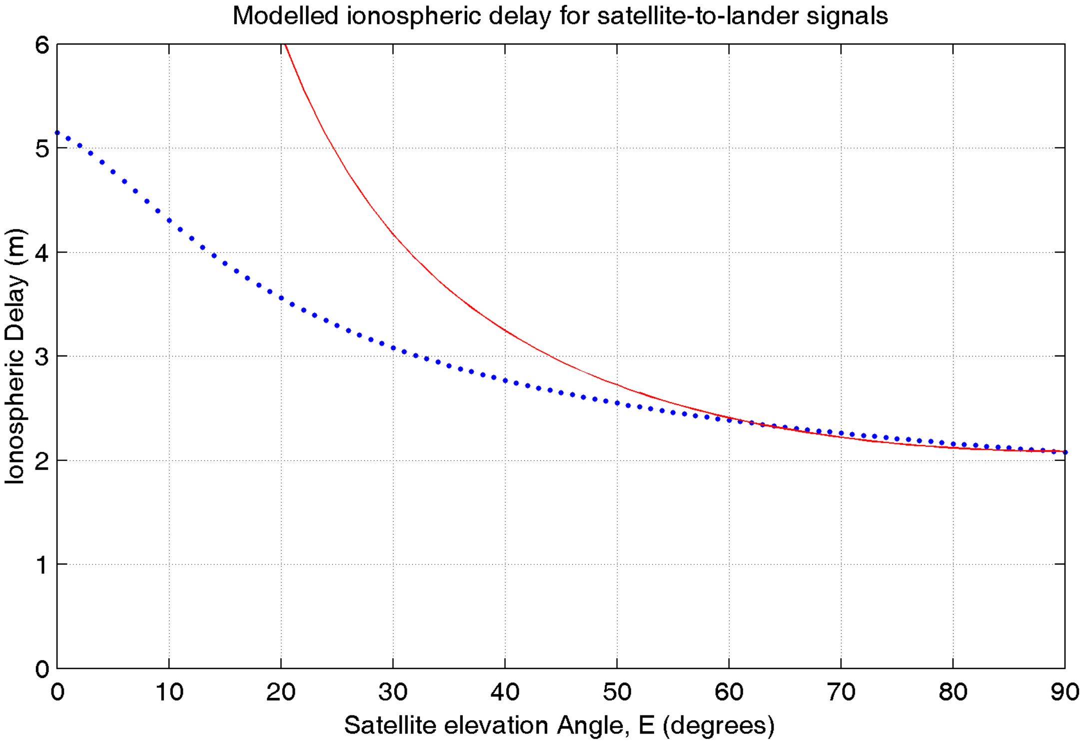

Figure 5.11: Modelled ionospheric delay for a satellite to lander range as a function of elevation angle. The red line indicates the value that would have been obtained by mapping the zenith value by $\frac{1}{\sin(E)}$ instead of computing the value of TEC, and applying the mapping function, at the ionospheric shell pierce point. This is a worst case scenario with the lander located at the sub-solar point. Actual modelled values will vary, depending on the elevation of the Sun and of the signal path at the ionospheric shell pierce point.

proach to the planet of the inter-satellite vector is computed. Then ray tracing is performed in both directions through a five layer approximation of the Mars ionosphere based on profiles of Mars' ionosphere electron density shown in Mendillo et al. (2003). The figure showing these profiles is reproduced in Figure 5.12. The relative electron densities of the five layer approximation are shown in Figure 5.13. If the assumption is made that all profiles have this same relative shape, then electron density and refractivity in each of the five layers in the model can be computed as a function of TEC.

Unlike the troposphere refractivity model, which depends only on height, the ionospheric model is dependent on horizontal position as well. To account for this, the TEC is evaluated at each pierce point of the inter-satellite vector when the closest approach distance is less than 135 km. Ray tracing is then done in each direction using two different five layer profiles. In the case that the inter-satellite vector does not pass through the ionospheric shell, TEC is calculated once at the point of closest approach. In this case, the total delay will be twice the delay computed by ray tracing in one direction from the point of closest approach. Figure 5.14 shows the ionospheric delay on an inter-satellite range as a function of the altitude of the point of closest approach for the worst case scenario where the point of closest approach is the sub-solar point. As expected, the modelled value for a closest approach altitude of zero corresponds to twice the zero elevation modelled ionospheric delay for a satellite-to-lander signal shown in Figure 5.11.

5.4.4 Clock Errors

Simulating the clock errors in a GNSS is a difficult task. In this thesis, the Mars Network clocks have been modelled as errorless. The rationale for omitting the clock

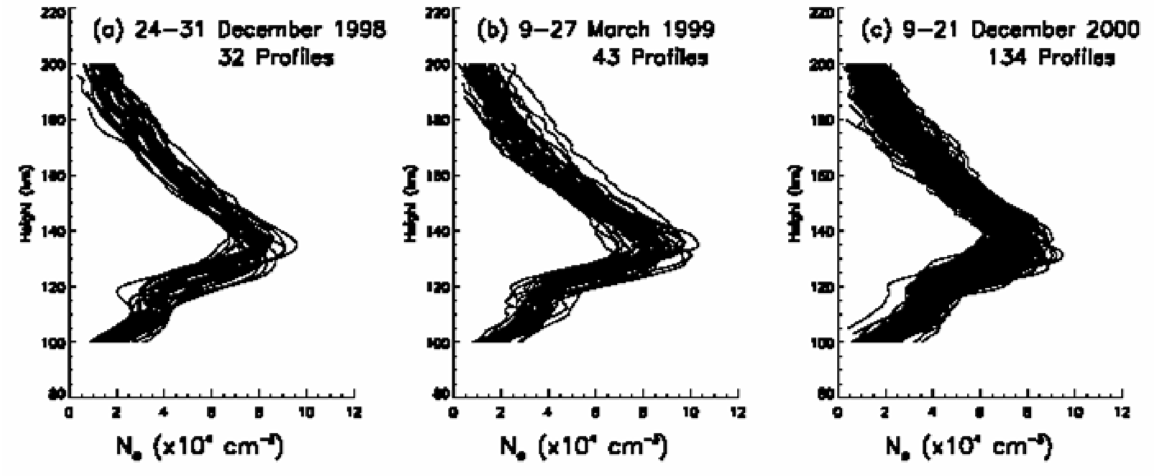

Figure 5.12: Profiles of Mars ionosphere electron density obtained by the Mars Global Surveyor Radio Science team (from Mendillo et al. (2003)).

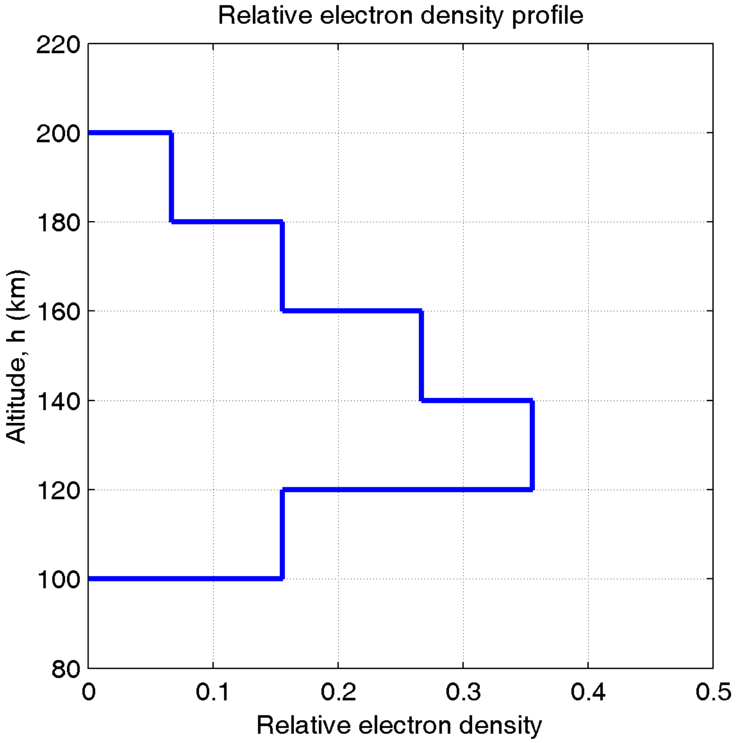

Figure 5.13: Relative electron density profile assumed to model electron density as a function of zenith TEC.

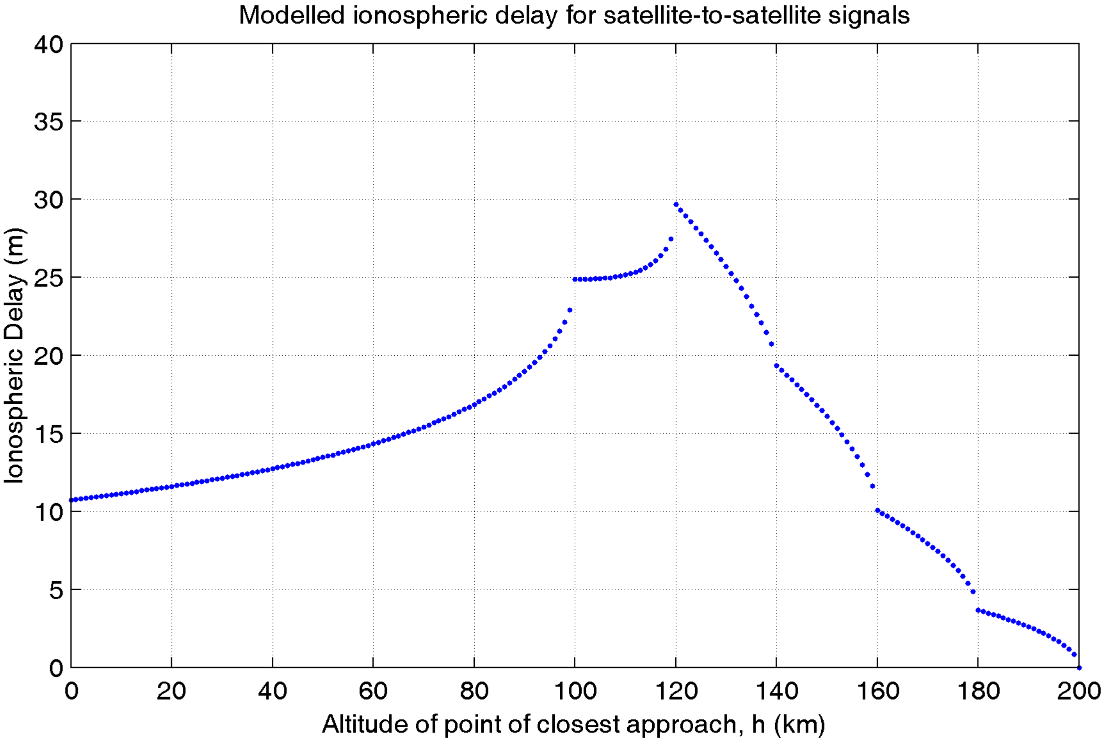

Figure 5.14: Ionospheric delay on an inter-satellite range as a function of altitude of the point of closest approach. In this example, the point of closest approach is the sub-solar point. The discontinuities are a result of the discrete electron density profile used to model refractivity in five concentric 10 km thick shells between 100 and 200 km. This is the worst case scenario, actual modelled values will depend on the TEC corresponding to the geometry of the inter-satellite range.

error is as follows: each Mars network satellite will carry a precise clock, probably a rubidium gas cell time standard. It is further assumed that a similar quality time standard may be included on any landers. Each will be capable of making one-way and two-way range observations. A one-way and a two-way range observation made simultaneously will allow for the direct solution of the relative clock offset between the two users. Furthermore, Doppler tracking from Earth will allow for direct observation of clock behaviour. The combination of these two methods should allow both for the establishment of a Mars-based time scale and for real-time clock steering and modelling of the clock errors of the Mars Network satellites. This simplification will only affect the validity of simulated positioning results if the clock errors are not included at all as states in the positioning solution. In Chapter 6, it is assumed that the clock offset and clock drift rates are estimated independently of the position and velocity states by comparing the one-way and two-way observations. The two-way observations are then used to determine position and velocity. It is also assumed that clock offsets will not be large enough to affect the predicted satellite positions. This is a reasonable assumption given the quality of clock proposed for the Mars Network elements and the fact that the one and two-way observations will be compared first to determine the relative clock offset, before the satellite positions are predicted and the two-way observations are used for positioning. An alternative approach would be to use both one-way and two-way observations and simultaneously estimate the clock states with the position and velocity.

5.4.5 Multipath

GNSS multipath is typically modelled by assuming a particular transmitter, reflector and receiver configuration in conjunction with a particular signal structure. Unfortunately the exact signal structure to be used for ranging purposes has not been specified (Hastrup et al., 2003). One possibility is that the receiver would obtain a range by measuring the time of arrival by correlation of a predefined sequence of data that would be transmitted at some predefined interval, lasting on the order of seconds. The fastest data rate according to Hastrup et al. (1999) would be 2048 kbps, which would imply a bit length of 147 m. 1024 kbps is also listed as a supported data rate. This is the same rate as the C/A code on GPS which has been extensively studied. Based on the assumption that a GPS multipath model is valid in this case, ground user multipath error has been simulated using the simplified GPS multipath model described by Luo (2001) that assumes a single specular reflecting plane with a randomly varying reflectivity is located below the receiving antenna. This is a reasonable model for a typical Mars lander and landing location. Generally, landers are landed in flat regions (where it is safer to land) and antennas are usually located on the top of the lander. Figure 5.15 illustrates this configuration.

The path delay, assuming the transmitter is far enough from the receiver to assume that the incoming signals are parallel, can be obtained by solving the right angle triangle formed between the direct signal, the reflected signal, and the line perpendicular to the direct signal that passes through the reflection point, noting that the angle between the direct and reflected signals at the antenna is twice the satellite elevation angle, E. The multipath delay is the difference between the reflected signal side of this triangle, and the direct signal side. The ratio of the reflected signal side to the direct signal side is $\cos(2E)$. If this length of the reflected side is denoted by r, then the difference between the two, or delay, is given by

$$l_{\text{delay}} = r(1 - \cos(2E)) \tag{5.16}$$

The length of the reflected side is also related to the height of the antenna, h, by

$$h = r \sin E \tag{5.17}$$

substituting for r in equation 5.16 gives

$$l_{\text{delay}} = \frac{h}{\sin E}(1 - \cos(2E)) \tag{5.18}$$

which can be reduced to

$$l_{\text{delay}} = 2h \sin E \tag{5.19}$$

using the trigonometric identity

$$1 - \cos E = 2 \sin^2 E \tag{5.20}$$

This model is a simplification of a more complex model developed by Ray (2000) and expanded for multiple reflectors by Ryan (2002). It computes the code signal multipath as a function of the direct and reflected signal amplitudes, the multipath path delay, and the receiver correlator spacing as a function of the signal chip (or bit) rate. The phase multipath is computed as a function of the the reflected signal amplitude, phase wavelength, and multipath path delay.

5.4.6 Noise

Receiver noise has been added to the simulated observations based on the 10 cm (1σ) range uncertainty and 0.5 mm/s (1σ) range rate uncertainty specification given in Ely et al. (1999) and Cesarone et al. (1999). Both range and range rate noise was modelled as random normal noise.

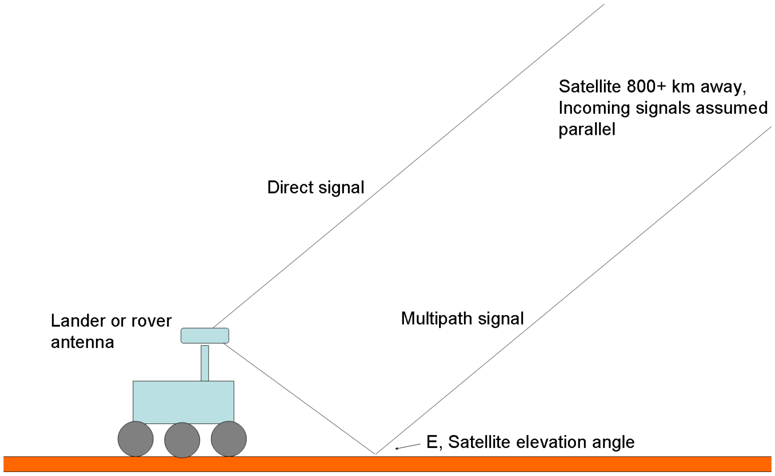

Figure 5.15: Single reflector multipath scenario for a lander.

5.5 Application of Modelled Errors

In the previous section, methods for simulating six types of errors affecting GNSS signals were developed. This section will briefly describe how each of these modelled errors should be applied to the ranges and range rates computed from theoretical trajectories in Section 5.3.

5.5.1 Uncorrelated Errors

The application of uncorrelated errors (multipath and noise) does not require much special consideration. Noise can simply be added to the range and range rate and multipath can be estimated as a function of transmitter-reflector-receiver geometry and added as a bias to range observations. In this regard, multipath will be temporally correlated when the geometry is slowly changing.

5.5.2 Correlated Errors

The application of models of spatially and temporally correlated error sources requires extra care, both in making sure that the correlation properties of the errors are preserved and in understanding how these errors might be handled in user software.

As described in Section 5.4.1, there is no need to explicitly add orbital error to the simulated observations as it is implicit in the trajectory computation described in section 5.2. However, the two atmospheric errors must be carefully considered. For the clock errors, there are two possible approaches. The first is to assume that all the clocks in all of the elements in the system are directly controlled by two-way ranging and simply assume that the clock offsets are always known; a small random value can then be added to each one-way observation to represent small uncertainties in the clock synchronization. The second approach requires that each element in the network have a modelled clock state that drifts as a function of time. This (larger) clock offset would then have to be added to all one-way ranges and likewise, the clock drift would be added to all one-way range rate observations, though two-way ranges and range rates would only be affected by the changes in the one clock over the very short two-way transmission times.

There are several possible strategies for dealing with the two atmospheric error sources. The first strategy is to assume that the modelled error values are realistic values and simply apply them to the theoretical ranges and range rates. This approach is used by Luo (2001) and Alves (2001) to simulate these errors for Earth-based observers of GPS and Galileo satellites. The difficulty with this method is that the same models for tropospheric and ionospheric error are often used by processing algorithms to mitigate the errors. When processing real observations, the application of these models will leave a residual error due to imperfections in the model. However, with simulated data, no residual error will remain if the same models are used to generate and compensate for the atmospheric effects. To account for this, many GNSS software simulators will add ad hoc random errors which will remain after models are applied by processing software to remove the errors.

A second strategy is to assume that ionospheric and tropospheric error models will not be employed by the user. If this is the case, then the modelled errors can be applied to the simulated data as is and used to evaluate the performance of the user algorithm in the presence of the errors. A final approach is to assume the user has access to the same models, but then limit the model accuracies by adding random effects that are a function of the modelled error values. Wolf (2000) takes this approach in his GNSS software simulator where he assumes that the ionospheric and troposphere models are 50% and 80% accurate respectively. He then adds random normal atmospheric errors that have standard deviations of 0.5 and 0.2 times the modelled values and simply excludes the correction step in his processing algorithm. In Chapter 6, the third strategy described above will be employed to evaluate positioning performance in the case where atmospheric models are not used to mitigate atmospheric errors, and in the case where models are used.

Finally, extra care must taken in simulating GNSS error effects if the observations are to be differenced by the user. The models used must ensure that there is the correct level of spatial and temporal correlation between observations so that differencing to reduce this effect will yield realistic results. The Mars Network design documents (Ely et al., 1999; Hastrup et al., 1999) do not discuss differencing, and none of the scenarios in this thesis will consider it, though it could be applied in the case of a rover positioning itself with respect to a nearby lander.

5.5.3 Mars Network User Error Budget

Based on the models described in the Chapter, and the specifications of the Mars Network given by Ely et al. (1999) and Cesarone et al. (1999), it is possible to construct a generalized error budget for a Mars Network user. A list of error sources affecting range observations, and a range of anticipated values for each error source are listed in Table 5.5. These values will have to be verified experimentally once the Mars Network transceivers are in use on Mars.

The orbital error will be the main source of error in all cases. Second to that is the ionospheric effect, which should be able to be reduced by applying some kind of latitude and solar elevation angle dependent model. The model presented in Section 5.4.3 could be used for this in practice, analogously to how the Earth ionosphere model that is broadcast by GPS is used by some single point positioning users. However, a more sophisticated model, based on more observations of the ionosphere of Mars would be preferable. Due to the very thin nature of the neutral atmosphere of Mars, the tropospheric effect is negligible. Clock errors are effectively removed by two-way observations. Multipath has been modelled after GPS multipath, the actual characteristics will depend on how the range observation is implemented in the Mars Network transceiver.

Table 5.5: Error budget for a Mars Network user.

Error Source	Range of Values (1 σ)
Orbits	5–20 m (will depend on how well they are modelled)
Ionosphere	2–6 m (day) 0 m (night)
Troposphere	0.05 m (zenith)
Clocks	0 (assuming 2-way observations)
Multipath	0.01–10 m
Receiver Noise	0.1 - 1 m

Chapter 6

Positioning and Orbit Determination Results using Simulated Observations

In this chapter positioning and orbit determination results are presented using simulated data obtained from the simulation procedures described in Chapter 5.

6.1 Simulated Data Set

One and two way range and Doppler observations were generated for all possible line-of-sight links between the six Mars Network satellites and eight simulated landers located at positions listed in Table 6.1. The lander locations represent two actual missions (MER 1 and MER 2), two proposed missions (Netlanders and Phoenix), and one failed mission (Polar Lander). They were chosen for this simulation for their distribution around the planet (two polar, 3 mid-latitude, 3 equatorial) and because they represent places on Mars where present and future missions actually plan to land.

The simulated data set is one day long with a three minute observation interval. This is the same observation convention used by Ely et al. (1999) and used in Chapter 4.

6.2 Fully Autonomous Processing Strategy

The fully autonomous processing strategy, introduced in Section 3.4.3, presents a good option for real time operations as there is no need to centrally collect all of the observations. The specifics of this processing strategy will now be discussed. Positioning and orbit determination results using this strategy are presented in the sections below.

Table 6.1: Locations of landers used to generate simulated satellite-to-lander observations.

Lander Name	Latitude (°)	Longitude (°)
Netlander 1	27.0 N	130.0 W
Netlander 2	12.5 S	160.0 W
Netlander 3	35.0 N	70.0 W
Netlander 4	32.5 S	85.0 E
MER 1	14.5 S	175.0 E
MER 2	2.0 S	7.18 W
Polar Lander	76.0 S	175.0 E
Phoenix	67.5 N	120.0 W

6.2.1 Assumptions and Conventions

The purpose of processing Mars Network observations is to determine the positions and velocities of each satellite and lander. Each satellite or lander, hereafter referred to as a network element, has a state vector and state covariance matrix. For the i^{th} element, these are denoted as $\boldsymbol{x}_i$ and $\boldsymbol{C}_{\hat{\boldsymbol{x}}i}$, where

$$\boldsymbol{x}_i = [x, y, z, v_x, v_y, v_z]_i^T \tag{6.1}$$

In addition, each element has a dynamics model $f_i(\boldsymbol{x}, t)$ that describes the predicted evolution of the state vector over time. For satellites, the dynamics model describes motion in a truncated Mars gravity model consisting of spherical harmonic coefficients to degree and order 20. For landers, the model is that they are stationary on the surface. Using these reference models, each element is capable of predicting its state vector and state covariance matrix forward in time.

Furthermore, each element is capable of making two-way range and Doppler observations and assimilating these observations into its state vector using a Kalman filter update step. It is assumed that one-way observations are also made, but they are used only to independently estimate the relative clock offset and clock drifts between pairs of elements (by differencing with the two-way observations) and are not included in the element's positioning solution.

6.2.2 Implementation

Both landers and orbiters are implemented as linearized Kalman filters, meaning that the state vector is broken down into a reference trajectory and an error state. The reference trajectory is determined by the deterministic dynamics model while the error state is estimated from observations and predicted forward using the transition matrix.

When a contact between two elements occurs, the elements exchange their predicted state vectors and covariance matrices. The element making and processing the observation assumes that the position and velocity of the remote network element is described by its most current state vector and its covariance matrix predicted forward to the appropriate time (transmit time for a one-way observation, and turnaround time for a two-way observation).

In most cases, only one other network element is in view. As a result, observations must be filtered one at a time. When range and range rate observations are made simultaneously, they are processed as a pair.

Two elements are shown in Figure 6.1. Each has a state vector, covariance, and a dynamics model for itself, and the ability to predict the state of the other element based on the state vector it receives. Each element is capable of making and using observations.

The observation variance is computed by adding the range and range rate measurement variances, σ_ρ^2 and $\sigma_{\dot{\rho}}^2$, to the position and velocity state variances of the remote element projected onto the line-of-sight using the design matrix $\boldsymbol{A}$.

$$\boldsymbol{C}_\ell = \begin{bmatrix} \sigma_\rho^2 & 0 \\ 0 & \sigma_{\dot{\rho}}^2 \end{bmatrix} + \boldsymbol{A}\boldsymbol{C}_{\hat{\boldsymbol{x}}}^{(-)}\boldsymbol{A}^T \tag{6.2}$$

This allows the local receiver to properly weight observations made of both well and poorly determined remote elements. The local element then performs a Kalman update step to obtain a new state vector and state covariance.

One advantage of this approach is that since each element is operating independently, it is not necessary for every element to be making observations and be updating its state. Some of the elements can be assigned as fixed, where they are simply predicting their states and covariances based on their dynamics models. Each element may also be set to ignore certain elements and each element can set its own elevation mask angle.

This implementation allows for easy testing of different scenarios, including lander positioning with fixed orbits, orbit determination with fixed landers, and many other combinations. Results from several of these scenarios are presented in the following sections.

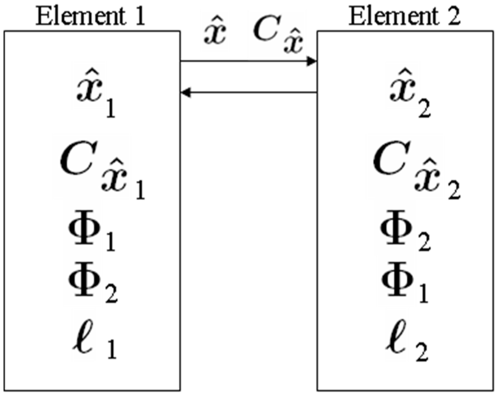

Figure 6.1: Schematic of two Mars Network elements. Each element has a state vector, covariance matrix and transition matrix. The two element exchange state vectors and covariance matrices and each is capable making observations and of predicting the state of the other element.

6.2.3 Implementation Details: Transition Matrix and Process Noise

A suitable transition matrix is required for both the satellites and the landers. For the landers, the result is trivial since they are stationary. The transition matrix is identity. For the orbiters, the transition matrix must be evaluated along the reference trajectory using numerical integration. This procedure, along with a derivation of the differential equation for the transition matrix, is described in greater detail in Appendix A. Because a linearized Kalman filter has been used, the transition matrix can be computed offline. If an extended Kalman filter were used, the transition matrix would have to be recomputed every time the reference trajectory is updated.

Process noise in a Kalman filter serves two functions. First, it represents the uncertainty in the dynamics model and provides a mathematical representation for the fact that the uncertainty of the state vector determined by prediction alone should grow over time. Its second function is to limit the influence of prior observations by not allowing the state covariance to converge to too small of a value.

For stationary landers, there is no need for process noise as the dynamics model of a lander is very well understood. A lander is simply stationary on the surface of the planet. For satellites, the issue of process noise is more complicated.

The effect of stochastic modelling error at any given time can be described as a set of zero mean random variables with a covariance matrix $\boldsymbol{Q}$. The effect of process noise over time is then obtained by integrating $\boldsymbol{Q}$ with the transition matrix from time $k-1$ to time k.

$$\boldsymbol{Q}_{k,k-1} = \int_{t_{k-1}}^{t_k} \boldsymbol{\Phi}(t_k, t)\boldsymbol{Q}(t)\boldsymbol{\Phi}^T(t_k, t)dt \tag{6.3}$$

Unfortunately $\boldsymbol{Q}$ is commonly used in Kalman filtering literature to both represent the covariance of the process noise and the integral. In this thesis the integral is subscripted with the interval $k, k-1$. Often $\boldsymbol{Q}$ is referred to as the spectral density matrix, while $\boldsymbol{Q}_{k,k-1}$ is called the process noise matrix.

In most discrete Kalman filter implementations, an expression for the process noise is obtained by assuming that only the highest order derivatives in the state vector are not fully modelled by the dynamics matrix. $\boldsymbol{Q}$ then only has a small number of non-zero elements along the diagonal. In these same applications, the dynamics matrix is normally not time dependent and a simple expression for the transition matrix can be obtained. In these cases, the integral in equation 6.3 is usually solved analytically.

The most common example of this is the constant velocity model, where the state vector consists of a position and velocity and it is assumed that the dynamics of the position state are perfectly modelled and the velocity state is subject to a stochastic error q. For the one-dimensional case, the constant velocity transition matrix is

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{6.4}$$

The spectral density matrix is

$$\boldsymbol{Q} = \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix} \tag{6.5}$$

and

$$\begin{aligned} \boldsymbol{Q}_{k,k-1} &= \int_{t_{k-1}}^{t_k} \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & q \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \Delta t & 1 \end{bmatrix} dt & (6.6) \\ &= \begin{bmatrix} \frac{q\Delta t^3}{3} & \frac{q\Delta t^2}{2} \\ \frac{q\Delta t^2}{2} & q\Delta t \end{bmatrix} & (6.7) \end{aligned}$$

For orbit determination problems, the transition matrix is not constant and it is not sufficient to assume that mis-modelling only affects the highest order derivative in the state vector. The problem is further complicated by the form of the transition matrix for orbital motion. In the constant velocity model, a small error in velocity will cause an error in position, but a small position error has no effect on the velocity estimate. For orbital motion, this is not the case since the future position depends on the velocity and the future velocity depends on the position.

This makes the selection of appropriate process noise much more complicated since noise added to the position will propagate into the velocity and will then propagate back into the position.

Different authors have different recommendations. Yunck (1996) suggests that error in the dynamics model be handled by estimating additional force error parameters implemented as first order Gauss-Markov processes. This way, it is possible to estimate unmodelled forces when sufficient observations are available, but the effect of these forces will decay to zero in the absence of observations. Montenbruck & Gill (2000) devote less than two pages to the discussion of process noise and in the end recommend that process noise be added by hand as required to “find a proper balance between process and measurement noise to ensure optimum filter performance".

In the results presented in this chapter, process noise has been neglected. In all of the scenarios, the available observations are very sparse and there is no risk of the user's position solution converging to a wrong value and becoming impossible to correct with new observations. In Section 6.4.1 below, where landers are being positioned with fixed satellites, the opposite occurs and the user's position is overly sensitive to erroneous measurements caused by improperly modelled satellite orbits. This effect could be reduced by adding process noise to the fixed satellites. An alternative approach is to allow the satellites to estimate their orbital errors based on observations of fixed landers. This is demonstrated in Section 6.5 and applied to improve lander positioning in Section 6.6.

6.3 Selection of Simulation Parameters and their Effects on the Results

In this thesis, the Mars Network constellation is evaluated as designed. Even without considering variations in the constellation design, there are many different parameters to consider when simulating the Mars Network and each parameter has many possible values. Some examples of variable parameters include the initial position errors of the network elements, the gravity model used by the user, and the level of the observation errors. No attempt has been made to present results from all possible combinations of values of these parameters. Instead a selected number representative positioning and orbit improvement scenarios are presented beginning in Section 6.4. However, it is useful to consider the sensitivity of the results to changes in the simulation parameters. This will now be discussed.

6.3.1 Effect of Initial State Error

Each of the elements in the network begins with an a priori state vector containing its position and velocity at the beginning of the scenario. This initial value may differ from the network element's true value and this difference will have an effect on the results. The effect of small variations in the initial value of the state on future values of the state can be assessed by considering the state transition matrix of the element. For example, an initial position error for a lander positioning itself making observations of satellites will result an increased time to converge to the true position. For an orbiter, An initial position or velocity error can lead to both secular and periodic position errors depending on the size and direction of the initial bias. For example, an initial across track position error will result in a varying across track error that repeats every orbit.

Initial position errors are more harmful if they affect network elements that are being used to position other elements. For example, a lander being used as ground control that has a 100 m position error can cause up to 100 m biases in range observations made to that lander, depending on the observation geometry. A satellite making multiple passes of this lander will then miscalculate its trajectory based on these biased observations.

6.3.2 Effect of Orbital Error Level and other Observation Errors

Independent of errors in the initial state vector of the network elements are errors in the model used to predict the orbits. In general, the more sophisticated the model, the smaller the orbital errors, and resulting observation biases, will be. This is similar for the other observation error sources. A user with a more sophisticated ionospheric model will have smaller residual ionospheric errors. The effect of orbital and observation errors is most pronounced for biases and in general the larger the bias, the larger the bias in the resulting user solution.

6.3.3 Effect of Filter Covariances

The ability of a least squares adjustment or Kalman filter to arrive at the correct solution is directly affected by the covariance information given to the filter. Particularly the relative weight between the initial covariance of the states and the variance of the observations. Adjusting these will change the balance between the dynamics model and the new observations. In the scenarios described in the rest of this chapter, realistic values for the initial state covariance and observation variance are used. Increasing the relative weight of the observation variance would cause faster convergence, but could lead to convergence to the wrong value if biased observations are present. Conversely, increasing the relative weight of the initial state estimate will make the result more closely follow the dynamics model.

Results of several scenarios will now be presented that illustrate the major positioning and orbit improvement capabilities of the Mars Network. It is important to keep in mind that, while these are only a small set of examples, the results of many other scenarios can be inferred by considering the relationships, between simulation parameters and results, that have been discussed in this section.

6.4 Positioning Landers with Satellite Orbits Fixed

6.4.1 Description of Scenario

The first scenario to be considered is the simplest case that assumes that the satellite orbits are modelled by the network users with a degree and order 20 gravity model.

However, the simulated observations are generated using the complete force model. The difference between the simplified and complete model trajectories is the orbital error. The orbital error as a function of time for each satellite is similar to that shown for Satellite 1 (the blue lines) in Figure 5.3. At the beginning of the simulation, each satellite is in its true position, but over time the orbital error increases as the degree and order 20 gravity model diverges from the true force model used to generate the simulated observations.

In this scenario, the initial positions and position variances of the landers were based on the final approach landing ellipse of the Mars Pathfinder Lander. The dimensions of Pathfinder's landing ellipse were approximately 15 km by 8.4 km, with the semi-major axis oriented in an approximately east-west direction (Vaughan, 1997). Each lander was assigned a position covariance matrix representing an ellipsoid with 1 σ semi-axes of 15 km, 8.4 km and 100 m, oriented east, north, and up respectively. The 100 m vertical value effectively serves as a height constraint, forcing the satellite observations to first improve the horizontal coordinates before affecting the height coordinate. In practice, the initial value of this height constraint could be increased or decreased depending on the quality and variability of the topography in the particular landing region. However, unlike instantaneous positioning with a height constraint discussed in Section 4.5, the height constraint in this case is only applied once in the form of the initial covariance of the position.

Each of the landers was arbitrarily assigned a starting location 8 km east, 4 km north, and 20 m above its true location in order to test the the ability of each lander to make observations and converge to its true position. Results for several scenarios using these initial conditions are presented below.

6.4.2 Results

First, lander positioning performance was tested assuming that only two-way range observations were being made. The velocity state of each lander was fixed to zero in the MCMF frame and neither the clock offset nor the clock drift were estimated.

The satellite state vectors were held completely fixed to the dynamics model. No process noise was added. The covariance matrix of each satellite was set to zero and the range observation standard deviation was set to 10 m in accordance with the user error budget given in Table 5.5. Not providing the orbiters with covariance matrices and not adding process noise ignores the fact that satellite orbit error will grow as a function of time. This is similar to point positioning with GPS, in that the ephemeris information is assumed to be perfect and any orbital errors are accounted for statistically in the observation variance.

Figure 6.2 shows the time series of position errors (estimated position - true position) for Netlander 1. The figure shows a relatively fast convergence to a reasonable position solution within the first two hours, followed by continued convergence. However, as the simulation progresses, the orbital errors become larger and further observations begin to degrade the position solution. The step-like shape of each of the position time series is due to the fact that the position estimate only changes during a satellite pass, when observations are actually being made.

Figures 6.3, 6.4 and 6.5 show similar results for Netlander 2, MER 1, and Phoenix. Results for MER 2 and Polar Lander are very similar to those of MER 1 and Phoenix. RMS position errors computed over four hour intervals are tabulated in Table 6.2.

In general, the equatorial landers have the best results. Netlander 2 and MER 1 and 2, for example, converge to accurate positions quickly and stay there for the entire test. The equatorial landers have the advantage of having a large number of satellite passes with varying geometries. The mid-latitude and polar landers are less well positioned and tend to diverge from their true positions as the orbital errors increase. Figure 6.6 contains azimuth/elevation angle skyplots for each of the four landers whose positions are shown in Figures 6.2 to 6.5. Clearly MER 1 and Netlander 2 have the most varied pass geometry. Phoenix only observes satellites in two-thirds of the sky and Netlander 1 has varied pass geometry, but not very many observations due to its mid-latitude location.

Figure 6.7 shows the innovation sequence, or predicted residual, of each observation. In a properly modelled filter, the innovation sequence should be a normally distributed quantity, however, because the user predicts the satellite orbits with an incorrect model, the orbital errors grow over time and so does the innovation sequence. This is particularly apparent in the second half of the simulation as the position solutions of some of the landers begin to diverge.

The total number of observations and corresponding cumulative position dilution of precision (PDOP) for Netlander 1, Netlander 2, MER 1 and Phoenix are shown in Figure 6.7.

6.4.3 Reliability Testing

The scenario described above was repeated with reliability testing implemented whereby any observations with predicted residuals in excess of 3.3 times the standard deviation of the predicted residual were rejected. This corresponds to the value of α of 0.001 that was used for reliability assessment of the constellation in Section 4.4.

Figures 6.9 and 6.10 show the east, north, and up position error components of Netlander 1 and Phoenix for the case that reliability testing is employed compared to the case where it is not. The most notable result is the reduction in divergence of

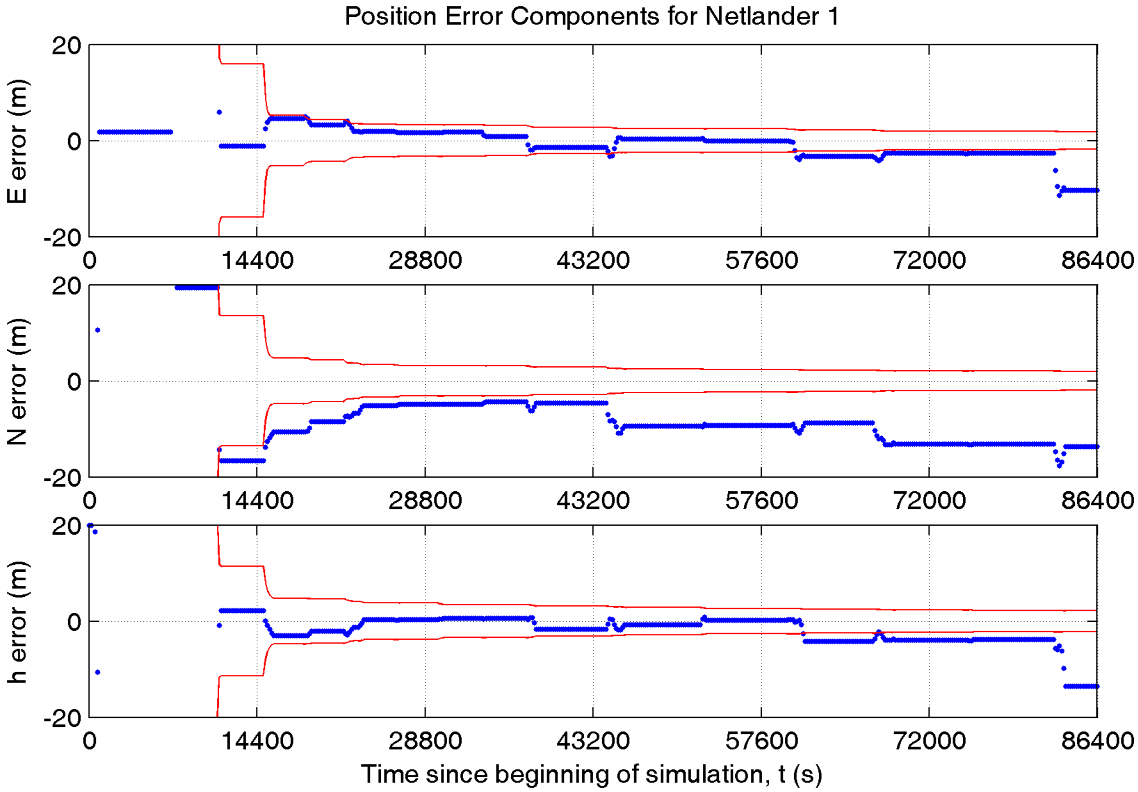

Figure 6.2: East, north, and up position errors for Netlander 1.

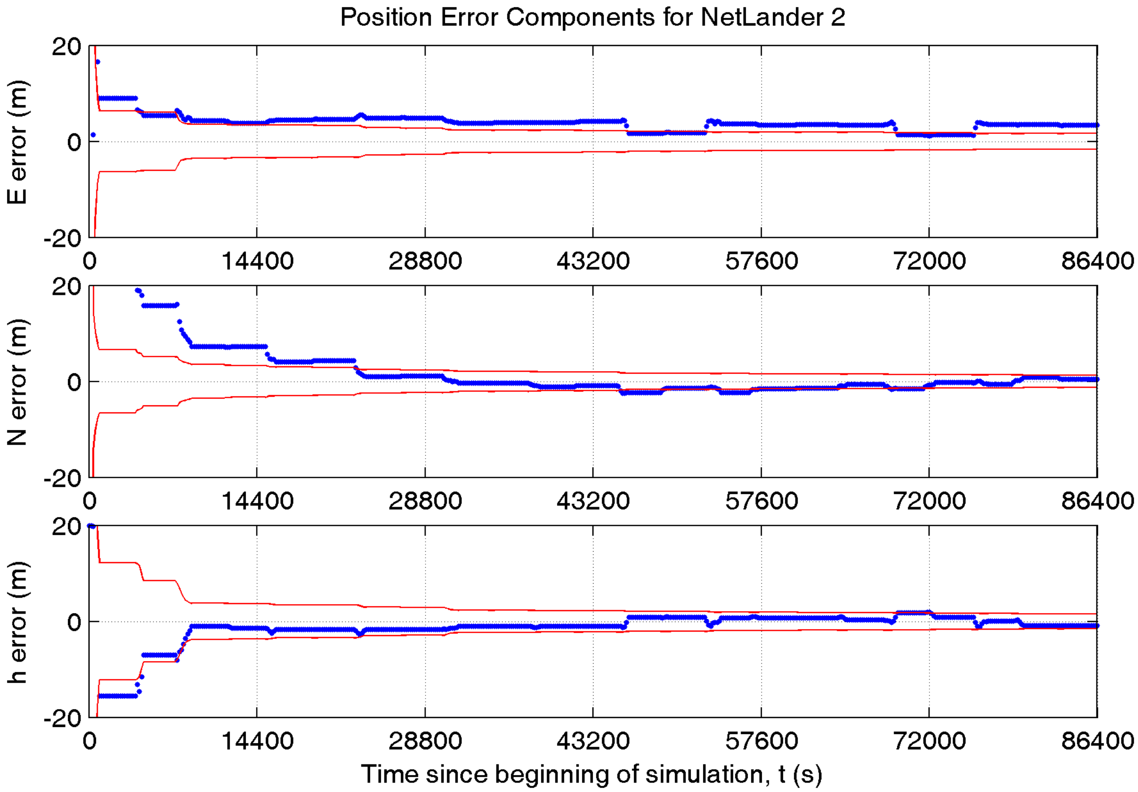

Figure 6.3: East, north, and up position errors for Netlander 2.

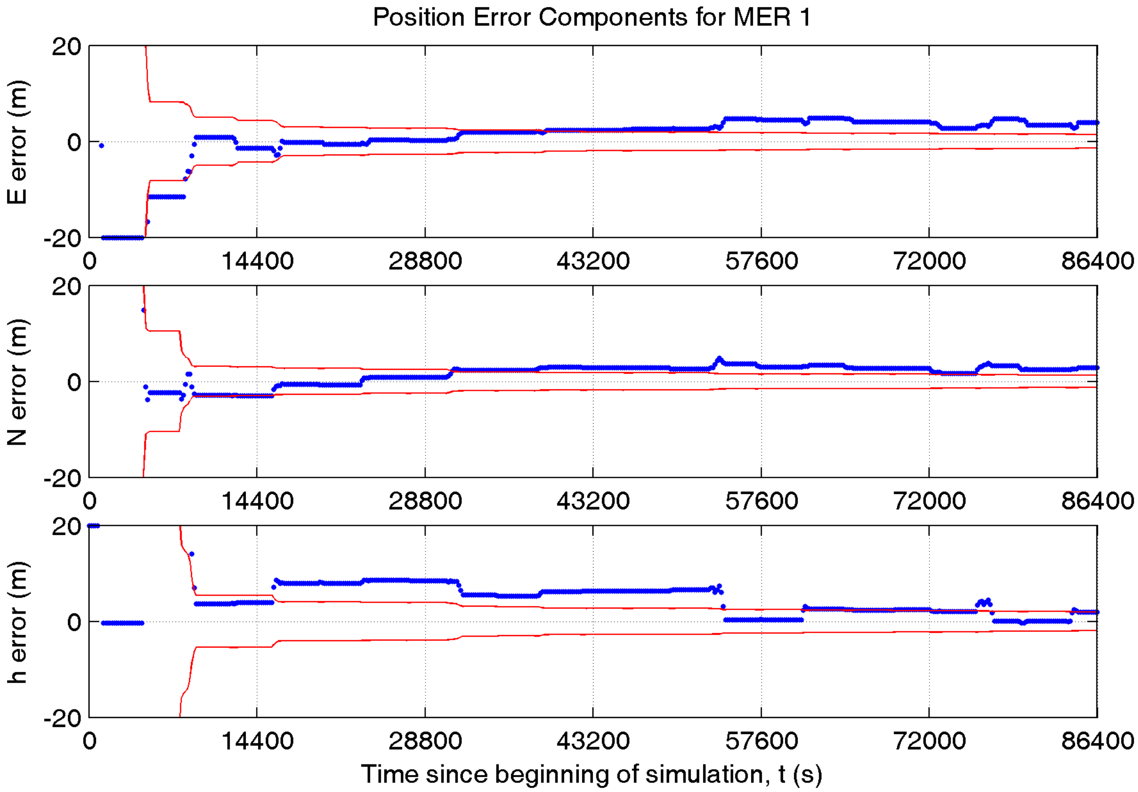

Figure 6.4: East, north, and up position errors for MER 1.

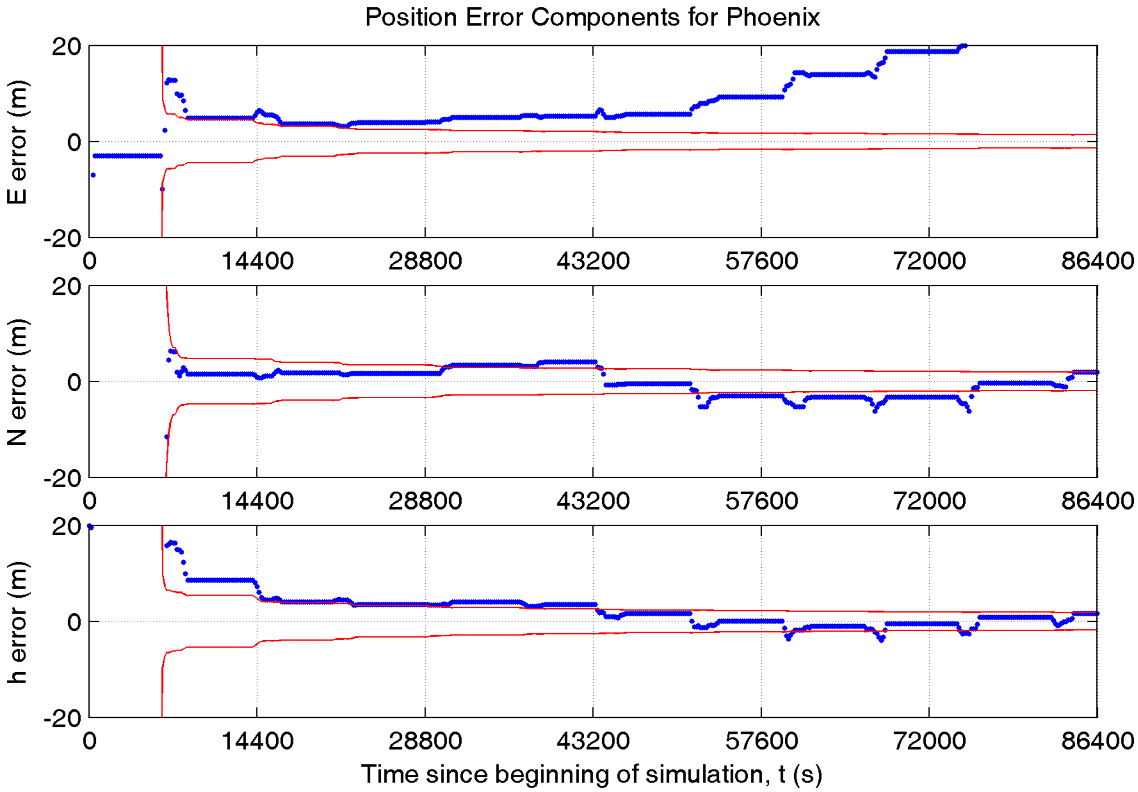

Figure 6.5: East, north, and up position errors for Phoenix.

Table 6.2: Lander RMS position errors in metres computed over four hour intervals for landers beginning with large horizontal position errors making two-way range observations to Mars Network satellites every three minutes.

Time	1–4 h	4–8 h	8–12 h	12–16 h	16–20 h	20–24 h
Netlander 1						
E	1503.294	3.128	1.447	0.787	2.802	5.523
N	800.474	8.679	4.703	8.997	10.440	13.484
h	22.548	1.919	1.038	0.834	3.428	6.906
Netlander 2						
E	1264.932	4.643	4.054	3.111	3.188	3.059
N	632.659	3.691	0.755	1.813	1.275	0.558
h	14.395	1.742	1.185	0.831	0.940	0.824
Netlander 3						
E	4816.714	4.332	3.752	5.418	6.992	7.396
N	2454.254	2.294	2.116	2.966	5.305	8.530
h	16.540	6.281	1.323	3.721	6.899	12.861
Netlander 4						
E	3582.167	1.889	2.392	5.891	6.107	7.774
N	1805.025	5.974	3.072	2.419	0.683	3.150
h	70.916	2.858	1.442	1.592	1.032	3.762
MER 1						
E	2208.738	0.758	1.880	3.166	4.382	3.637
N	1000.543	1.179	2.379	3.085	2.964	2.600
h	22.794	7.870	6.384	5.678	2.125	1.823
MER 2						
E	2037.320	0.797	1.338	1.480	1.006	0.653
N	973.653	1.219	2.703	4.040	1.070	1.024
h	39.223	2.437	2.450	0.174	1.091	1.965
Polar Lander						
E	907.220	3.069	4.337	6.794	8.815	13.161
N	503.020	5.056	1.106	2.570	3.722	2.425
h	43.909	6.698	1.802	6.883	10.787	9.160
Phoenix						
E	917.808	4.104	5.0468	7.092	14.833	23.103
N	697.034	1.675	3.4256	2.349	3.636	2.123
h	20.785	4.086	3.7242	1.399	1.394	1.132

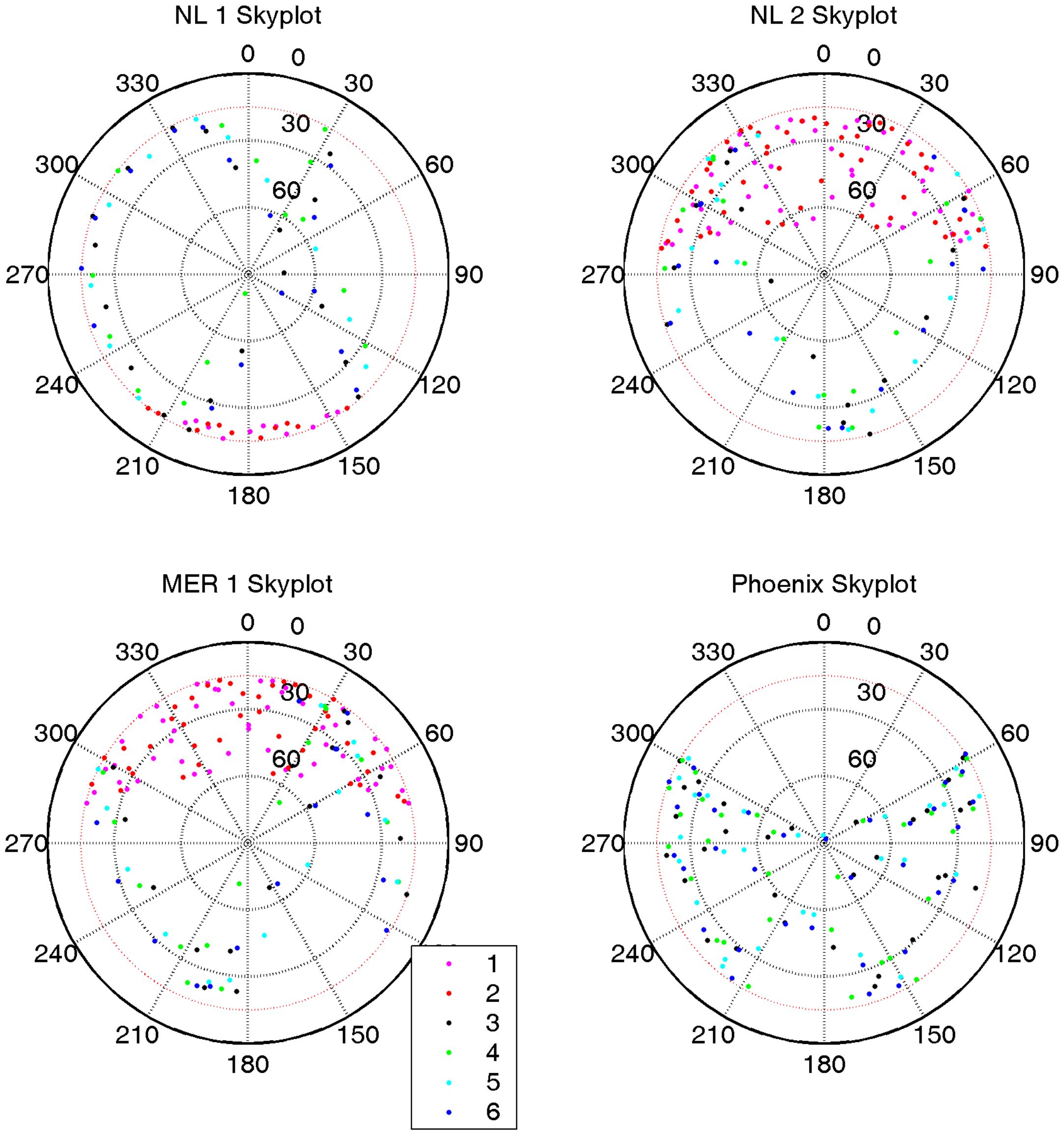

Figure 6.6: Azimuth and elevation of observations made by Netlander 1, Netlander 2, MER 1, and Phoenix.

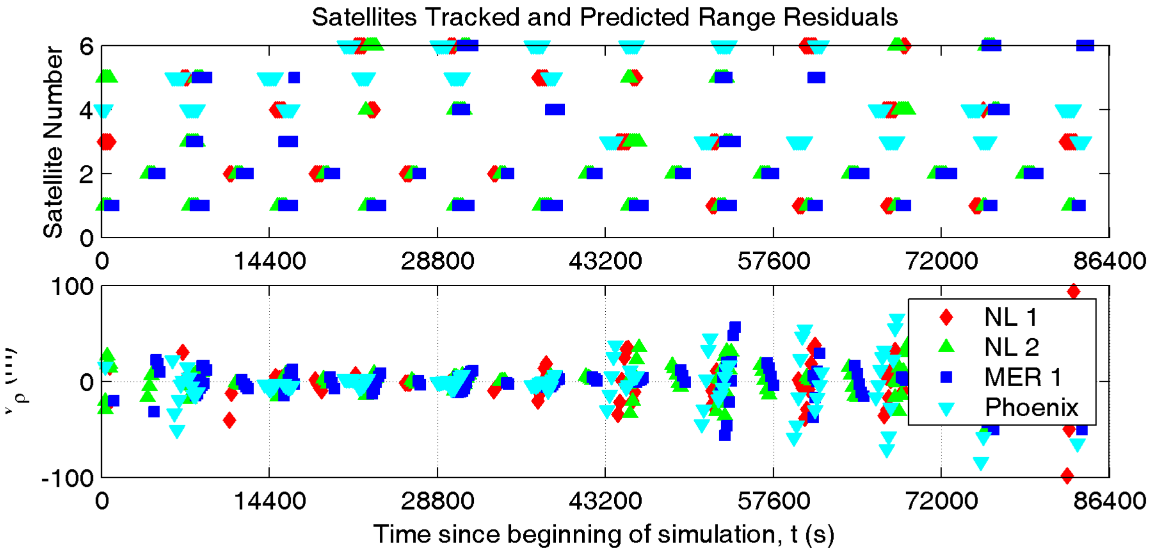

Figure 6.7: Satellites tracked and predicted range residuals (innovation sequences) for Netlander 1, Netlander 2, MER 1 and Phoenix.

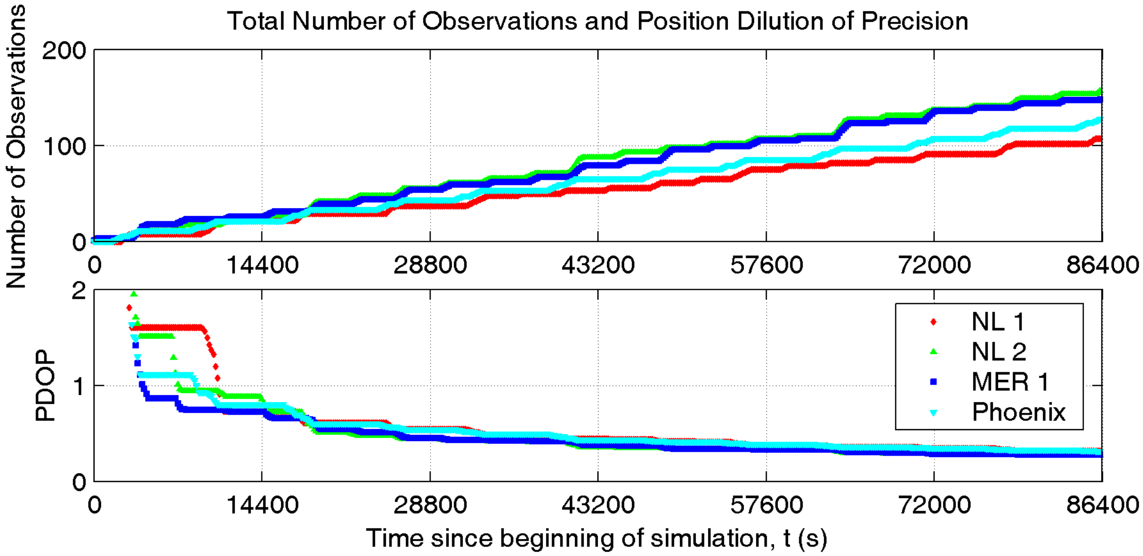

Figure 6.8: Total number of observations and cumulative position dilution of precision for Netlander 1, Netlander 2, MER 1 and Phoenix.

the position from the true value near the end of each simulation. This occurs because the biased observations resulting from increased orbital errors are not allowed into the position solution. Results for the six other landers are similar.

6.4.4 Discussion of Orbit Fixed Results

Clearly holding the orbits completely fixed to their reference trajectories is not an acceptable solution, especially in the long term. Without reliability testing, position errors are corrupted by growing orbital errors. With reliability testing, observations are rejected leaving few observations for positioning. In order to deal keep the orbital errors at a reasonable level, regular updates of the orbits would be required, most likely from Earth-based tracking, unless the users were equipped to compute the reference trajectories with the most sophisticated models possible. Another possible solution is to add appropriate process noise to the orbiter covariance matrices so that observations made later in the simulation would be less influential on the solution than those made earlier. This is not an acceptable long term solution however, since the landers would effectively stop making new observations of the satellites after the predicted satellite orbital errors become too large. A more obvious solution is to use the landers to track the satellites and improve the satellite orbits.

6.5 Improving Satellite Orbits with Landers Fixed

In Chapter 1, the possibility of using landers as ground control was introduced. In this section, the eight landers listed in Table 6.1 are evaluated for their ability to provide ground control. Specifically, the ability of the four Netlanders, the four other landers, and all eight landers, to improve the orbits of the Mars Network satellites is compared. In this scenario, there is no initial position error for the landers, though uncertainties

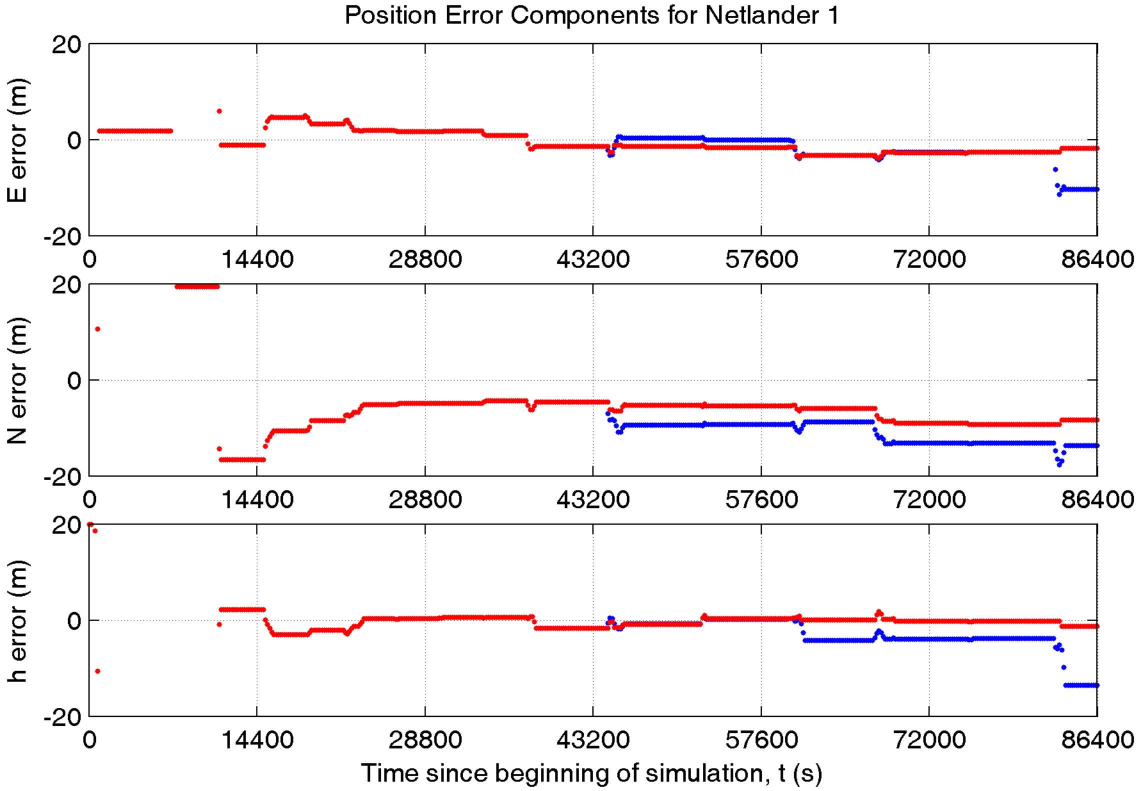

Figure 6.9: East, north, and up position errors for Netlander 1 with and without reliability testing. Results with reliability testing are shown in red, without in blue.

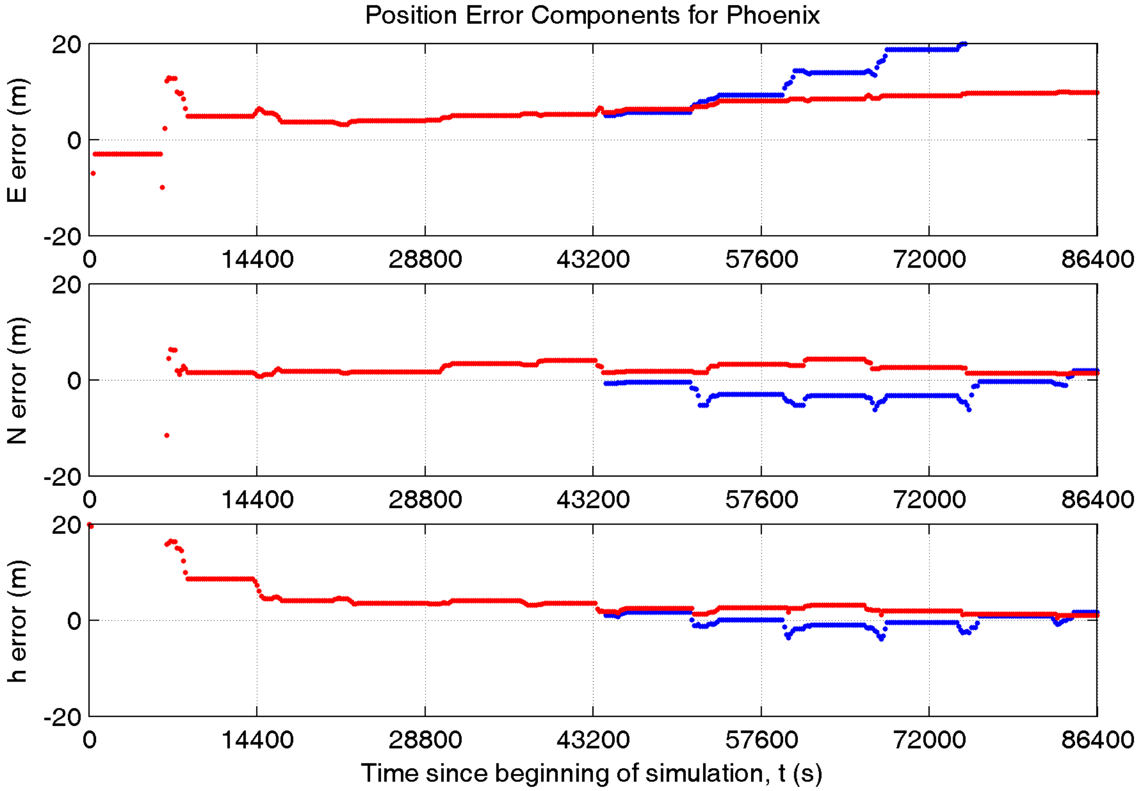

Figure 6.10: East, north, and up position errors for Phoenix with and without reliability testing. Results with reliability testing are shown in red, without in blue.

in their positions are accounted for in the range uncertainty. The satellites make two way range and range rate observations of the landers and attempt to estimate their position and velocity error states. A 15 ° lander elevation mask is used and inter-satellite observations are not used. Ranges are assumed to have a 10 m standard deviation, which includes both ranging uncertainty and lander position uncertainty. Range rates are assigned a 10 mm/s standard deviation.

Figure 6.11 is a time series of the estimated position error states of Satellite 1 as a function of time for all three ground control scenarios (Netlanders, other landers, all landers). The radial, along track, and across track components of the error estimate are given in the frame of the reference trajectory. For comparison, the difference between the true satellite trajectory and the reference trajectory is also shown. Clearly, the satellite trajectory is being improved by each of the lander configurations, though the along track direction results are poorer than the other two directions.

The two lower subplots in Figure 6.11 show the range and range rate innovation sequences (or predicted residuals). They are clearly not normally distributed as structure can be seen in every satellite pass, however, the range innovation sequences are not nearly as large as those in Figure 6.7 for the case where landers were positioning with the orbits held fixed. This is a direct result of the orbiters adjusting their own positions to better fit the observations.

The actual position errors (i.e. estimated position - true position, as opposed to the filter error state estimates) are shown in Figure 6.12. Results for Satellite 2, which is also in a near-equatorial orbit are similar. Figures 6.13 and 6.12 show the similar results for Satellite 3. The three other polar satellites behave similarly.

In all cases, the along track component is the more poorly controlled. This is due to the observation geometry. Most satellite-to-lander ranges and range rates are the

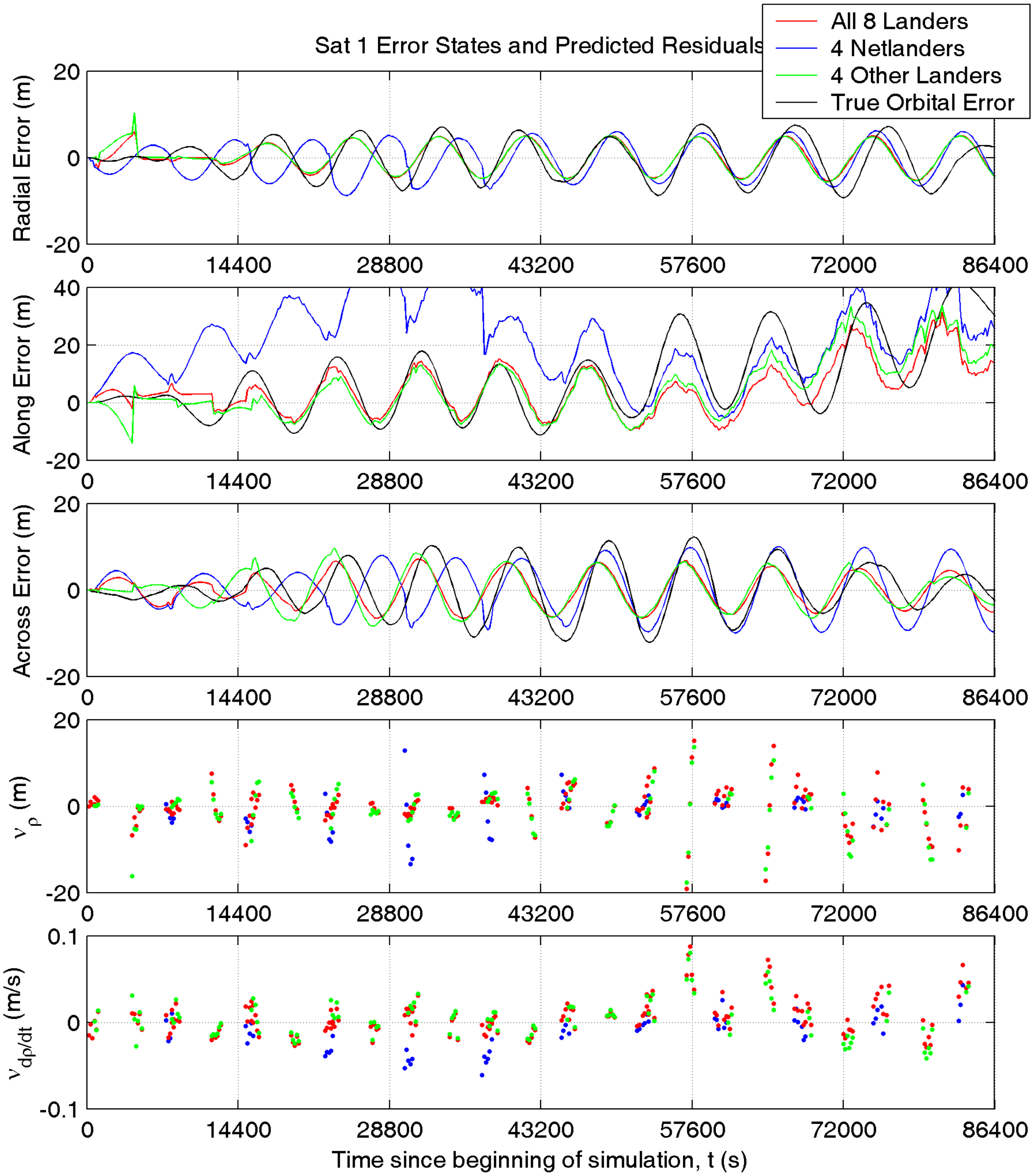

Figure 6.11: Estimated radial, along track, and across track position error states of satellite 1 for three different configurations of lander ground control. The true orbital error is also shown. The range and range rate predicted residuals for each observation are shown in the lower two subplots.

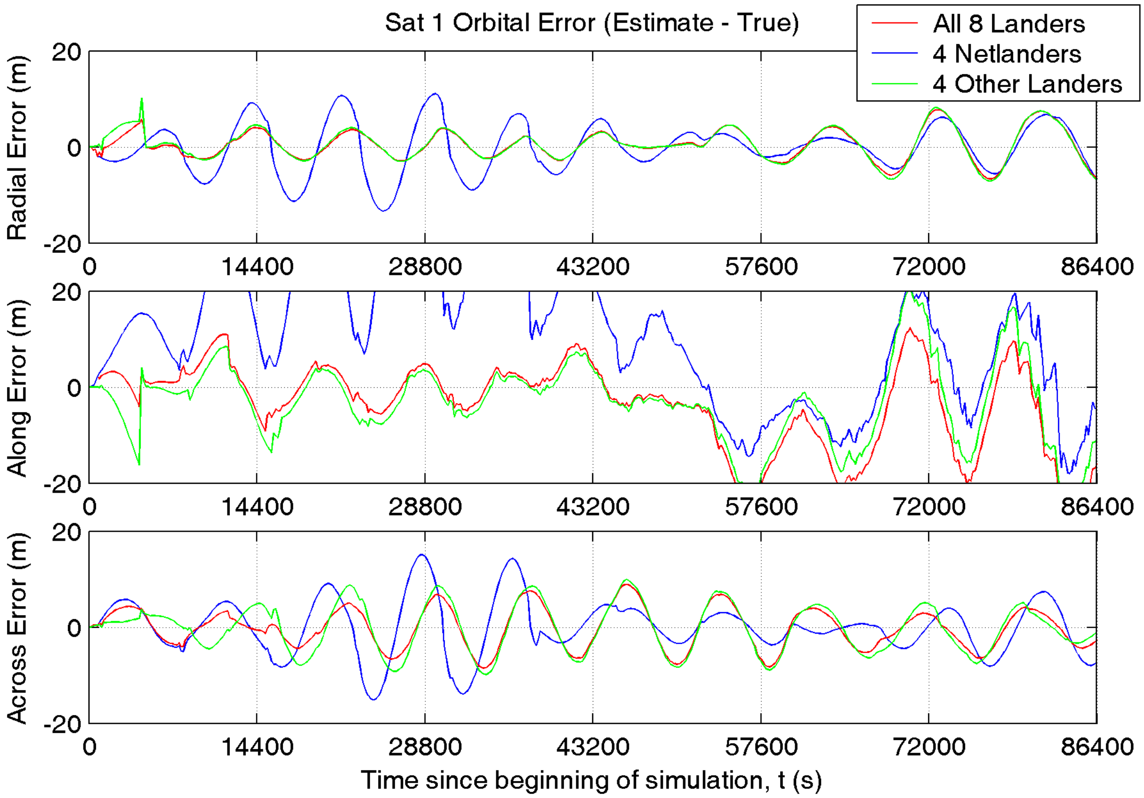

Figure 6.12: Estimated orbital error of Satellite 1 for three different configurations of lander ground control. The plots show the radial, along track and across track differences between the true satellite trajectory and the estimated satellite position in the frame of the true satellite trajectory.

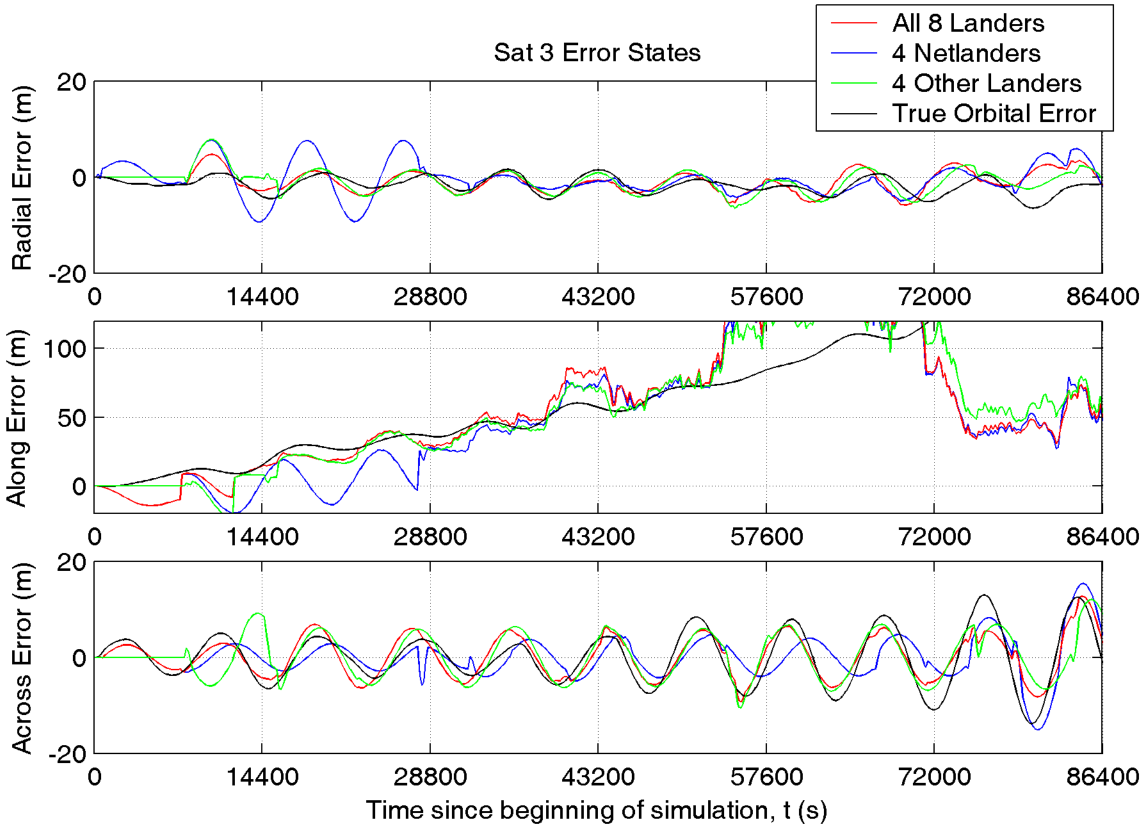

Figure 6.13: Estimated radial, along track, and across track position error states of satellite 3 for three different configurations of lander ground control. The true orbital error is also shown.

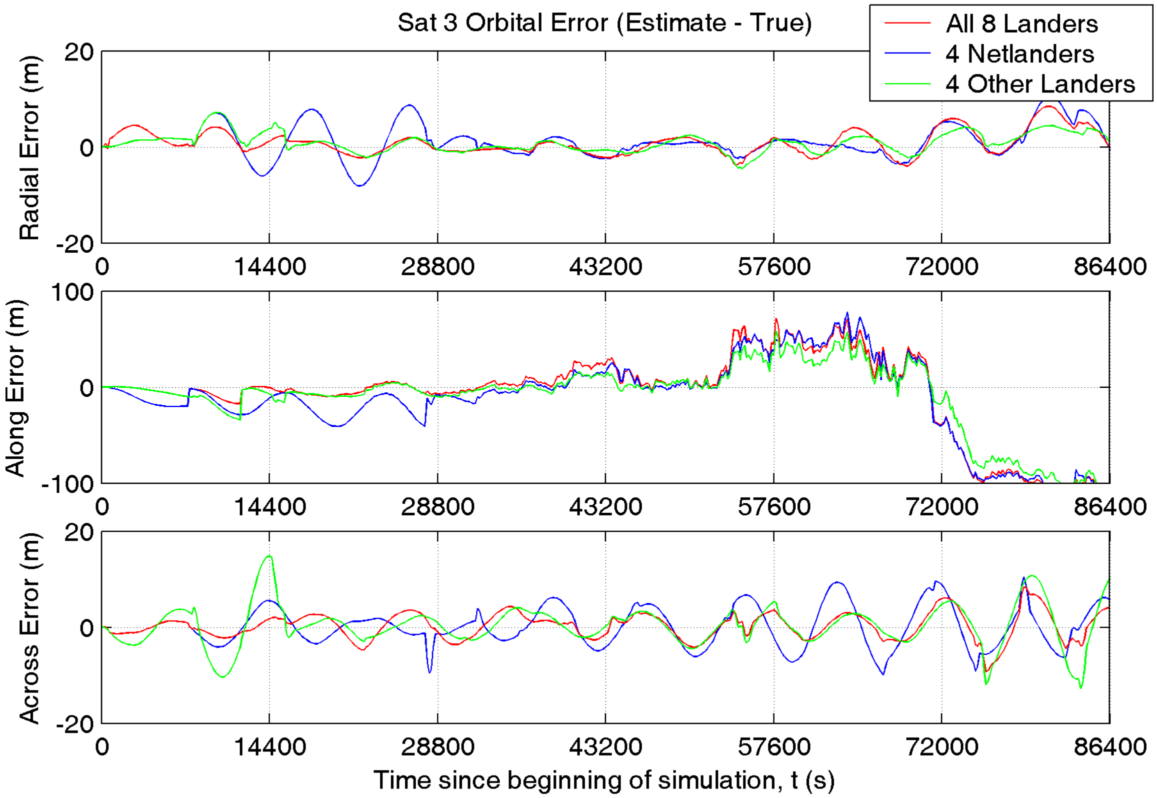

Figure 6.14: Estimated orbital error of satellite 3 for three different configurations of lander ground control. The plots show the radial, along track and across track differences between the true satellite trajectory and the estimated satellite position in the frame of the true satellite trajectory.

radial and across track directions. Observations that are mainly in the along track direction are less frequent as they only occur when the satellite is rising or setting.

One possibility to improve the along track performance would be to include observations to landers that were rejected by the 15° lander elevation cutoff angle. Figure 6.15 shows the position error states for Satellite 1 for both 15° and 0° lander elevation angles. In places, the along track error improves, and in others it is worse. There are similar variabilities in the across track error, though the inclusion of lower elevation observations has little effect on the radial error. This is likely due to the fact that allowing lower elevation observations not only allows the orbiter to track possibly one or two more observations at it flies over a lander, but also allows it to track landers that appear on the horizon on either side. In both cases, the low elevation observations are more affected by the ionospheric error, so their inclusion may be doing more harm, in the form of biased observations, than good.

The fact that limited observations from a small number of ground control stations can correct simplified orbits also demonstrates the potential geodetic use of the Mars Network. Since higher order gravity and other effects are observed in this scenario, the next logical step would be to use these improved orbits to solve for the underlying gravity coefficients. This would most appropriately be done in a batch least squares adjustment with the largest possible set of observations and the most sophisticated models available for the various forces acting on the satellites. This was exactly the procedure employed by Lemoine et al. (2001) using Earth-tracked orbital arcs as input to determine the GMM2B gravity model for Mars.

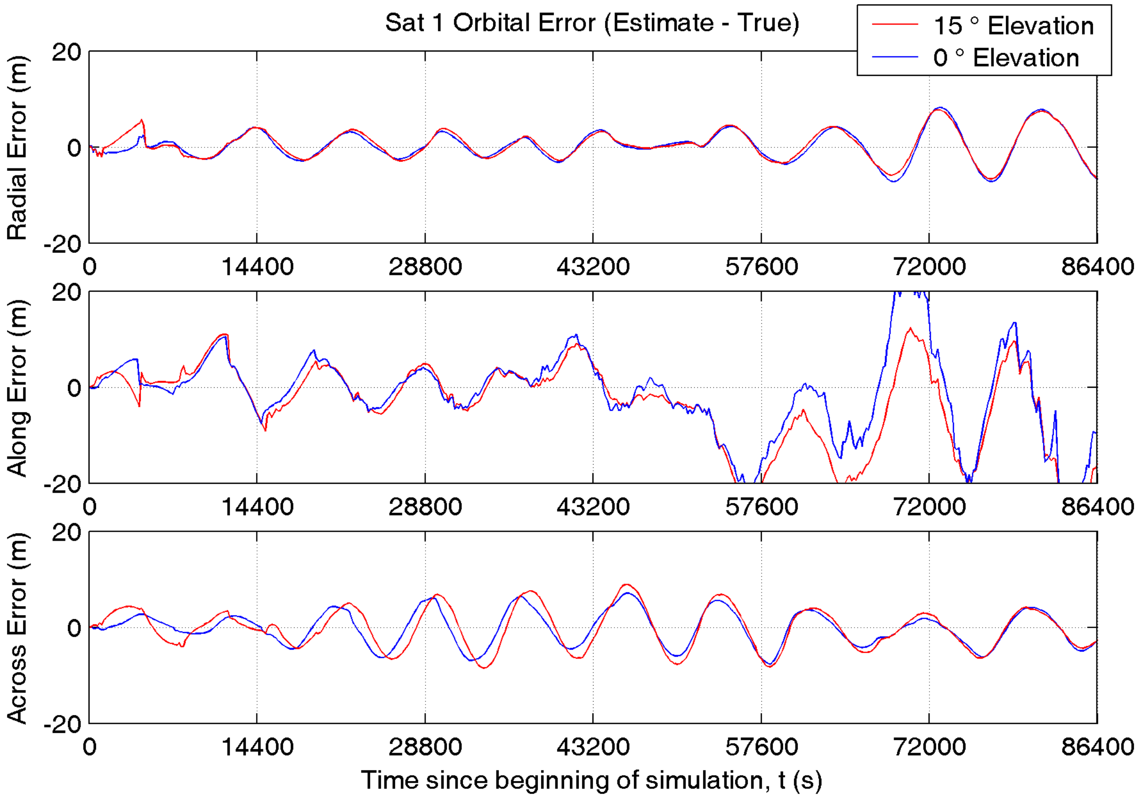

Figure 6.15: Orbit error components of satellite 1 when tracking landers with 15° and 0° elevation cutoff angles.

6.6 Simultaneously Improving Orbits and Positioning Landers

In Section 6.4.1 the satellite orbits limited the ability to position landers. In section 6.5 fixed landers were shown to improve the satellite orbits. It should be possible to use a subset of the landers as ground control and to both improve the satellite orbits and obtain better lander position solutions.

This is demonstrated in the following scenario. The four Netlanders were selected as ground control and fixed to their true positions. The satellites then used observations of the fixed ground stations to improve their orbits and the remaining four landers used observations of the six satellites to determine their positions. In this scenario, the satellites did not make use of observations to the free landers to determine their orbits. The initial conditions for the orbiters were the same as those used in Section 6.5 and the initial conditions of the four landers to be positioned were the the same as those used in Section 6.4.1 (i.e lander positions initially biased, but inside their respective landing ellipses).

Position error components for two of the landers (MER 1 and Phoenix) are shown in Figures 6.16 and 6.17. Because the initial states of the landers, and the observations that they used, were identical to the lander positioning scenario in Section 6.4.1, the results of these two scenarios can be directly compared to see the effect of improving the orbits using the four Netlanders as ground control.

The level of position error is not reduced in general, but there is a noticeable improvement in the second half of the test for Phoenix where the large divergence in the east coordinate is no longer present.

This is an encouraging result as it demonstrates the advantage that using some

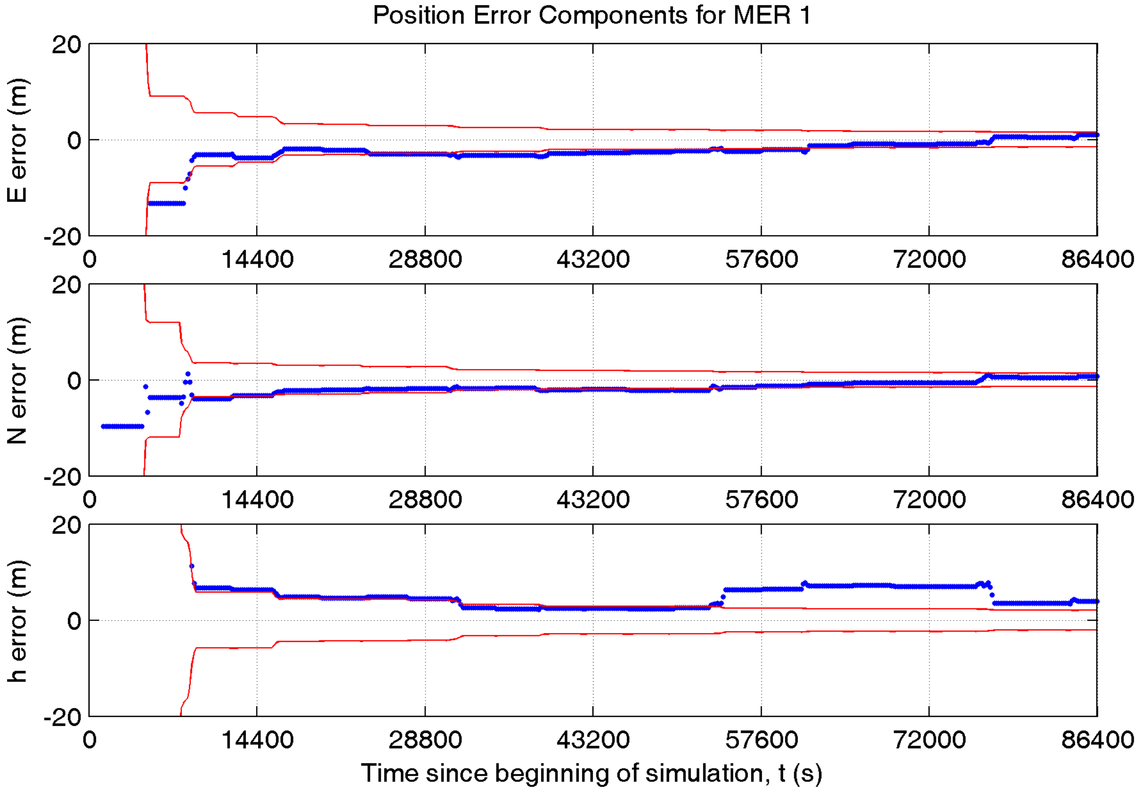

Figure 6.16: Position errors of MER 1 with satellite orbits estimated using Netlanders as ground control.

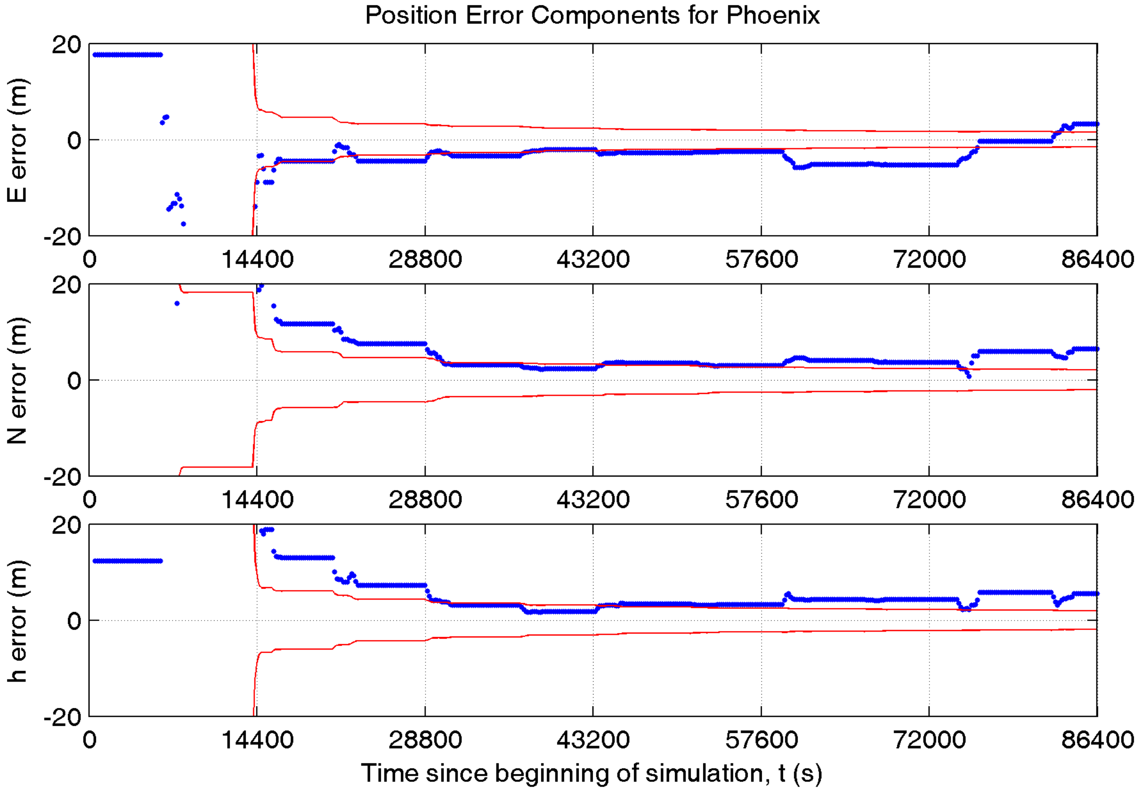

Figure 6.17: Position errors of Phoenix lander with satellite orbits estimated using Netlanders as ground control.

form of ground control can have on the ability to position other landers. In a sense, this is a form of dynamic differential positioning, where instead of just differencing between a base station and and a remote station to reduce the orbital error, as is the case in differential GPS processing, the base station observation is used to directly reduce the orbital error on the satellite. This allows for the advantages of differential positioning without the need for the satellite to be simultaneously visible to a base station and the remote station.

6.7 Use of Inter-Satellite Observations

In Chapter 5, inter-satellite as well as satellite-to-lander observations were simulated. The usefulness of inter-satellite ranges and inter-satellite range-rates will now be examined in two simulated scenarios.

The first scenario is to test the effect of inter-satellite ranges on the orbit determination of one satellite assuming that all of the other satellites are fixed to their reference trajectories. In this case, tracking of fixed landers is also employed and the results are compared to the case where orbits are improved using lander tracking only.

One other scenario involving inter-satellite ranges was tested. In this case, all of the satellites are estimating their states with all of the landers held fixed.

Estimated position error states and range and range rate predicted residuals for Satellite 1 for both of the above scenarios are shown in Figure 6.18. There are many more inter-satellite observations than there are satellite-to-lander observations; however, as can be seen from the figure, these observations do not noticeably improve the results. In the case where the other satellite orbits are fixed, this is because all of the inter-satellite ranges are affected by the orbital errors of the other satellites, which

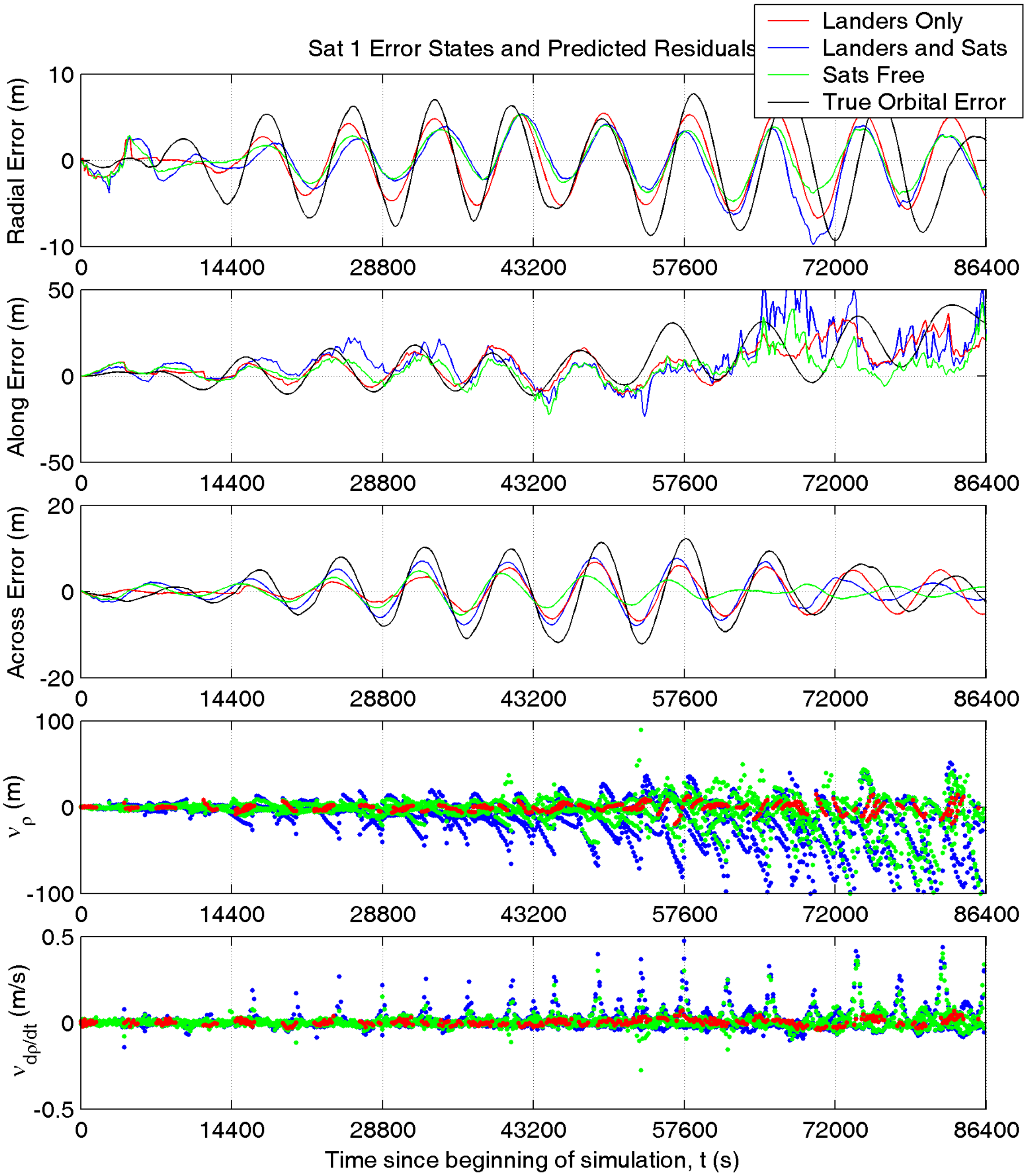

Figure 6.18: Satellite 1 position error states and range and range rate residuals obtained from tracking fixed landers with and without the use of inter-satellite observations to the other satellites. Two cases are shown. The blue lines correspond to the case where the other five satellites are fixed to their reference trajectories. The green lines show the case where the other satellites are estimating their own orbits using observations from landers and inter-satellite observations.

are not being estimated or controlled in this test. This can be seen in the magnitudes of the predicted residuals, which increase as a function of time as the orbital errors of the other satellites grow. The largest effect on the position solution of the satellite is seen in the along track direction, which is not surprising as it is the component of the position error state that is the least well determined from the lander observations.

In the case where the other satellites are free and are estimating their orbital errors, the magnitude of the residuals near the end of the simulation is generally smaller, indicating that the orbits of the other satellites have somewhat improved, at least in terms of fitting the observations. However, the end result in this case is not closer to the true orbit than the case where the satellite only uses observations from landers.

These results suggest that the fully autonomous processing scheme is not well suited for cases when multiple elements have poorly determined state vectors, i.e. satellites observing satellites whose state vectors contain large orbital errors.

6.8 Discussion

6.8.1 Simultaneous Lander and Satellite Positioning

The initial goal of the autonomous processing strategy presented in this chapter was to have every element in the network (satellites and landers) positioning itself using observations of every other element in the network. In principle, with reasonably correct initial conditions and properly weighted observations and errors, it should be possible to have every element free to improve its position.

However, all of the results presented in this chapter have relied on having some of the network element state vectors fixed to provide control. Though this is not required in theory, since the initial state estimates of each element ensure that the network is not datum deficient, in practice it is difficult to avoid diverging solutions if fixed control of some form (ie. fixed landers or satellites) is not provided. The main limitation is the fact that the satellites are moving and small errors in their positions and velocities propagate into secular and periodic errors over time. The problem is compounded by the relatively small number of observations, sometimes only three or four per pass, which make it difficult to average out observation errors.

It would be exceedingly difficult to simply deploy the satellites and landers and let them all position themselves. This particular scenario has been attempted with little success and is not reported in detail. In practice, the "all elements free" scenario only seems to work well if the initial position errors and corresponding variances of each of the elements are very small. Of course, if the initial position errors and variances are small enough, it is equivalent to having fixed control.

6.8.2 A Practical Strategy for Mars Network Implementation

Assuming that the Mars Network and a set of landers were simultaneously sent to Mars in the future, a more practical scenario for establishing a reference frame and a GNSS for Mars would probably consist of the following steps. First, initial positions and orbits would be determined with Earth-based tracking of the satellites along with simultaneous lander positioning on the surface. Earth would provide the control for the satellites and the satellites would position the landers. Once the lander positions were sufficiently well determined, they could be fixed and the landers could then be used to provide control for the satellites. At this point, Earth-based tracking would be switched off and the coordinates of the landing sites would define the reference frame for Mars.

It should also be noted that in a real-world application, the most accurate possible representation of the orbits would be used. In this chapter, the users were assumed to have a degree and order 20 gravity model, and nothing else to model the trajectories of the satellites. This was done to see if the network would be able to estimate the orbital errors. In reality, the orbits would be modelled in the most sophisticated way possible given the computer power available to the user. Orbital error would still occur due to imperfections in the model, but it would hopefully be smaller and not grow over time as much as the error from using degree and order 20 gravity model.

6.8.3 Centralized Processing and Batch Estimation

Apart from brief mentions in Chapter 3, so far only autonomous processing using sequential least squares or Kalman filtering has been discussed. There are other approaches that could be used. Specifically filtering using centralized processing, and batch least squares. In both of these methods, the covariances between the state vectors of different network elements would be properly considered, as opposed to the autonomous strategy where it is ignored. A further advantage of batch least squares processing is that divergence is not an issue since all of the observations are considered together to find the overall best fitting solution of all of the orbits and all of the lander positions. The disadvantage of the batch solution is that it cannot be used in real time. In practice, the batch solution would be used to precisely position landers post-mission, while the autonomous filtering method presented in this chapter could be used for real time applications. This is exactly analogous to the current practice in GPS data processing on Earth.

6.8.4 Evaluation of Overall Performance

From the few scenarios considered in this chapter, it is possible to make some conclusions about the performance of the Mars Network constellation. Overall, the performance of the Mars Network for positioning landers when the orbits are fixed is very good, provided that the model of the orbits is reasonably close to the true trajectories. As was predicted in Chapter 4, the positioning performance of the Network for a particular lander will depend highly on the lander's latitude, with mid-latitude landers being the most poorly served.

The ability of a small number of landers to effectively control orbital errors over the short term has also been demonstrated with promising initial results. The ability to use these improved orbits to determine improved positions of other landers is also a very significant result.

The usefulness of inter-satellite signals with the autonomous processing scheme is somewhat questionable, particularly if one or more satellites are constrained to their reference trajectories and those reference trajectories are in error.

Only one processing strategy was discussed in this chapter, and then only a limited number of scenarios were demonstrated. There are many other observation scenarios and processing options that could, and should be tested. These will be discussed in the Future Work section of Chapter 7.

Chapter 7

Conclusions and Recommendations

7.1 Conclusions

The main results of Chapters 4, 5, and 6 are summarized below.

In Chapter 4, the Mars Network constellation design was evaluated using figures of merit commonly used to assess Earth-based GNSS. The six satellite Mars Network constellation will provide excellent navigation coverage for stationary equatorial and polar users in terms of availability, and accuracy. This was clearly the intent of its designers, since the equatorial and polar regions are of particular interest. Unfortunately, the constellation, as designed, provides poorest coverage in the 30 to 40 degree latitude region. The Network will also provide very limited instantaneous positioning capabilities to equatorial users and higher latitude users if a height constraint is applied. Unfortunately there are never redundant observations, so instantaneous testing of the residuals for reliability purposes will not be possible. Blunder detection is possible for a lander that makes many observations over time.

Chapter 5 presents a detailed simulation of the Mars Network satellite orbits, observables, and errors. In previous studies, the Mars Network principal investigators neglected atmospheric errors when simulating the performance of the system. In Chapter 5, two very simple atmospheric error models were presented. A simple Mars neutral atmospheric error model produces zenith delay of only 5 cm, while an extremely simplified ionospheric model results in a maximum zenith delay of 2.3 m. Clearly, neglecting the neutral atmosphere delay is a valid assumption, but the ionospheric effects require more study.

Position domain results using simulated observations and an autonomous processing strategy were demonstrated in Chapter 6. The ability of the Mars Network to position surface elements to within 10 metres over several hours was confirmed. However the accuracy of these position estimates is most affected by the quality of the satellite orbits.

Simulations presented in Chapter 6 were pessimistic in that the reference trajectory was the degree and order 20 gravity model. In practice, the most complicated model available should be used if the computing resources are available to the user. In the case that a truncated model must be used, the results from Chapter 6 show that it is possible to obtain reasonably accurate positions on the surface for several hours provided good, externally generated, initial positions for the satellites are available.

Also in Chapter 6, the ability of a small set of landers to improve the orbits and in turn improve positioning for other landers was demonstrated. This is a very promising result that warrants further study.

7.2 Recommendations

Based on the results in this thesis, the following recommendations can be made.

The Mars Network, as designed, provides very good, but uneven coverage. An increase in the inclination of the two near equatorial satellites would help provide more even coverage in the 30 to 40 degree latitude range.

Also, higher orbits would results in larger coverage circles and longer satellite passes. Increasing the altitude would have to be weighed against the increased power requirements. A higher orbit would also be less affected by higher order gravity terms, which would be an advantage for positioning applications, but would mean less geodetic information could be gathered from observations of the satellites.

The use of inter-satellite links must be very carefully considered, particularly if either of the satellites has a poorly estimated position or the signal path is traveling through the ionosphere. If either of these situations applies and is not properly modelled, the inter-satellite observations will not improve the orbit estimation.

A second frequency would be a useful addition, both for making ionospheric correction and observations of the ionosphere itself.

The inclusion of landers in the network will be essential if the desired result is a self contained navigation system for Mars. If there is no ground control, it will have to be provided from Earth via tracking with the Deep Space Network. DSN observation time is limited and expensive making Mars-based ground control a more attractive option. All future Mars landers are potential ground control stations, but the ideal ground control station would preferably be long lasting and would likely require more durable electronic components and nuclear power. The Viking landers are a good model. They were expensive, heavy, and nuclear powered, but they lasted years, compared to months for Mars Pathfinder and the MER rovers.

7.3 Future Work

There are several areas to continue the work started in this thesis.

The effects of other types of observations could also be included, for example other navigation sensors on the landers. This would be particularly important in the case of a slow moving rover where some other means of navigation would be required between satellite passes. While compasses and barometers would not be useful in Mars' weak magnetic field and thin atmosphere, inertial sensors could be used.

Each of the models used to create the simulated observations in Chapter 5 could be improved (with the exception perhaps of the gravity model). The ionospheric error simulation should be improved to better represent the actual behaviour of the Mars ionosphere. This will not be possible until more information about the Mars ionosphere is available. Improvements in these models would lead to slightly more realistic simulated observations, however this would not greatly affect on the results presented in Chapter 6 since the orbital errors are by far the largest error source.

The linearized Kalman filter solutions presented in Chapter 6 could be compared to extended Kalman filter solutions to verify if the linearization about the reference trajectory of the satellites is a valid assumption.

Finally, perhaps the most obvious future task is to obtain real data and verify the simulated results. This will only be possible when the Mars Network and other lander missions are launched.

The Mars Network and Netlanders are both very ambitious proposals. Unfortunately there is no indication if and when either of these proposed systems might actually be deployed. The most recent publications by the Mars Network principal investigators (Ely et al., 2003) indicate that they are focusing on developing the Mars Network transceiver and it appears that there are no longer any concrete plans to deploy the six satellite network proposed in 1999. However, they do plan to demonstrate the technology by positioning the Mars Exploration Rovers using the Mars Odyssey Orbiter, all of which are equipped with a prototype receiver.

Even if all future science orbiters carry the Mars Network transceiver, as is currently the plan, they will not be able to provide the same kind of positioning performance as the six satellite constellation. Most Mars science orbiters to date have been remote sensing satellites in low altitude polar orbits. This is a good configuration for imaging, but it provides very poor pass geometry (north-south ground tracks only) and a small coverage circle.

In this thesis, the abilities of the Mars Network constellation to accurately and quickly position landers have been demonstrated. Hopefully the Mars Network, or something like it will be deployed soon, so that future Mars missions will be able to benefit from the positioning and communications services that it will provide.

References

Acuna, M., J. Connerney, N. Ness, R. Lin, D. Mitchell, C. Carlson, J. McFadden, K. Anderson, H. Reme, C. Mazell, D. Vignes, P. Wasilewski, & P. Cloutier (1999) Global Distribution of Crustal Magnetization Discoverd by the Mars Global Surveyor MAG/ER Experiment. *Science*, 284, 790–793.
URL `http://mgs-mager.gsfc.nasa.gov/publications/science_284_5415_acuna/science_284_5415_acuna.pdf`

Allan, D., N. Ashby, & C. Hodge (1997) The Science of Timekeeping. Technical Report Application Note 1289, Hewlett Packard.

Allison, M. & M. McEwan (2000) A post-Pathfinder evaluation of areocentric solar corrdinates with improved timing recipes for Mars seasonal/dirunal climate studies. *Planetary and Space Science*, 48, 215–235.

Alves, P. (2001) The Effect of Galileo on Carrier Phase Ambiguity Resolution. *Proceedings of the Satellite Division of the Institute of Navigation 14th International Technical Meeting (September 11-14, 2001, Salt Lake City, Utah)*, 2086–2095.

Ashby, N. & J. J. Spilker Jr (1996) *Introduction to Relativistic Effects on the Global Positioning System*, Volume 1, *Global Positioning System: Theory and Applications*, Chapter 18, 623–697. American Institute of Aeronautics and Astronautics, Washington.

Baarda, W. (1967) Statistical Concepts in Geodesy. *Netherlands Geodetic Commission*, 4(2).

Baarda, W. (1968) A Testing Procedure for use in Geodetic Networks. *Netherlands Geodetic Commission*, 5(2).

Barriot, J.-P., V. Dehant, W. Folkner, J.-C. Cerisier, A. Ribes, J. Benoist, T. V. Hoolst, P. Defraigne, R. Warnant, R. Preston, L. Romans, S. Wu, & A. Wernik (2001) The Netlander Ionosphere and Geodesy Experiment. *Advances in Space Research*, 28(8), 1137–1249.

Bell, D., R. Cesarone, T. Ely, C. Edwards, & S. Townes (2000) Mars Network: A Mars Orbiting Communications and Navigation Satellite Constellation. *2000 IEEE Aerospace Conference (March 18-25, 2000, Big Sky, Montana)*, 75–87.

Bell, D. & T. Ely (1999) Constellation Design of the Mars Network. *1999 International Symposium on Space Communications and Navigation Technologies (September 21-23, 1999, Pasadena, California).*

Benson, T. (2000) Mars Atmosphere Model - Metric Units. Webpage. URL `http://www.grc.nasa.gov/WWW/K-12/airplane/atmosmrm.html`

Beutler, G. & M. Rothacher (2002) Advanced Aspects of Satellite Positioning. Distinguished International Lectures Series ENGO 699.80, The University of Calgary.

Bouguillon, S. & J. Souchay (1999) Precise modeling of the precession-nutation of Mars. *Astronomy and Astrophysics*, 245, 282–297.

Buffett, B. A. (1985) *A Short-Arc Orbit Determination for the Global Positioning System*. Master's thesis, University of Calgary.

Cesarone, R., R. Hastrup, D. Bell, D. Lyons, & K. Nelson (1999) Architectural Design for a Mars Communications and Navigational Orbital Infrastructure. *AAS/AIAA Astrodynamics Specialist Conference (August 16-19, 1999, Girdwood, Alaska).*

Chapront-Touzé, M. (1990) Orbits of the Martian satellites from ESAPHO and ESADE theories. *Astronomy and Astrophysics*, 240, 159–172.

Cheney, W. & D. Kincaid (1994) *Numerical Mathematics and Computing*. Brooks/Cole Publishing Company, New York, 3 Edition.

Crider, D., M. Acuña, J. Connerney, D. Mitchell, R. Lin, P. Cloutier, H. Rème, C. Mazelle, D. Brain, N. Ness, & S. Bauer (2001) 3D Multiscale Mass Loaded MKD Simulations of the Solar Wind Interaction with Mars. *Advances in Space Research*, 27(11), 1831–1836.

Cunningham, L. (1970) On the Computation of the Spherical Harmonic Terms needed during the Numerical Integration of the Orbital Motion of an Artificial Satellite. *Celestial Mechanics*, 2, 207–216.

Daly, P. & P. Misra (1996) *GPS and Global Navigation Satellite System (GLONASS)*, Volume 2, *Global Positioning System: Theory and Applications*, Chapter 9, 243–272. American Institute of Aeronautics and Astronautics, Washington.

Dehant, V., T. V. Hoolst, P. Defraigne, R. Warnant, & F. Rooskeek (1994) NEIGE: NETlander Ionosphere and Geodesy Experiment. *Space Scientific Research in Belgium / Volume 2 Space Sciences Part 2 Chapter 6 Planetary Geodesy*, 173–184. URL `http://www.busoc.be/general/spacesciences/VDehantTVanHoolst.pdf`

Duxbury, T., R. Kirk, B. Archinal, & G. Neumann (2002) Mars Geodesy/Cartography working group recommendations on Mars cartographic constants and coordinate systems. *Proceedings of the Symposium on Geopotential Theory, Processing and Applications (2000, Ottawa, Ontario)*, 4 pages.

Edwards, C., J. Adams, D. Bell, R. Cesarone, R. DePaula, J. Durning, T. Ely, R. Leung, C. McGraw, & S. Rosell (2001) Strategies for Telecommunicaionts and Navigation in Support of Mars Exploration. *Acta Astronautica*, 48(5–12), 661–668.

Ely, T., R. Anderson, Y. Bar-Sever, D. Bell, J. Guinn, M. Jah, P. Kallemeyn, E. Levene, L. Romans, & S. Wu (1999) Mars Network Constellation Design Drivers and Strategies. *AAS/AIAA Astrodynamics Specialist Conference (August 16-19, 1999, Girdwood, Alaska)*.

Ely, T., J. Guinn, & E. Quintanilla (2003) Navigation Services of the Mars Network. *ION 59th Annual Meeting/CIGTF 22nd Guidance Test Symposium (June 23-25, 2003, Albuquerque, New Mexico)*, 427–434.

Essen, L. & K. D. Froome (1951) The Refractive Indices and Dielectric Constants of Air and its Principal Consitituents and 24,000 Mc/s. *The Proceedings of the Physical Society, Section B*, 64,Part 10(382B), 863–875.

Folkner, W. M., C. F. Yoder, D. N. Yuan, E. M. Standish, & R. Preston (1997) Interior Structure and Seasonal Mass Redistribution of Mars from Radio Tracking of Mars Pathfinder. *Science*, 278, 1749–1751.

FRP (2001) *2001 Federal Radionavigation Plan*. The United States Department of Defence and Department of Transportation, Springfield, Virginia.
URL `http://www.navcen.uscg.gov/pubs/frp2001/FRP2001.pdf`

Gauchez, D. & J. Souchay (2000) Analysis and Comparison of Different Estimations of Mars' Polar Motion. *Earth, Moon, and Planets*, 84, 33–51.

Grewal, M. S. & A. P. Andrews (1993) *Kalman Filtering Theory and Practice*. Prentice Hall, Englewood Cliffs, New Jersey.

Hart, M. J., E. A. Williams, & E. Campbell (1999) Mars Navigation / Communication Satellite Constellation Design Using a Genetic Algorithm. *1999 International Symposium on Space Communications and Navigation Technologies (September 21-23, 1999, Pasadena, California)*.

Hastrup, R., D. Bell, R. Cesarone, C. Edwards, T. Ely, J. Guinn, S. Rosell, J. Srinivasan, & S. Townes (2003) Mars Network for enabling low-cost missions. *Acta Astronautica*, 52, 227–235.

Hastrup, R., R. Cesarone, J. Srinivasan, & D. Morabito (1999) Mars Comm/Nav MicroSat Network. *13th Annual AIAA/USU Conference on Small Satellites, (August 23-26, 1999, Logan, Utah)*, 1–11.

Ho, C., N. Gloshan, & A. Kliore (2002) Radio Wave Propagation Handbook for Communication on and Around Mars. Technical Report JPL Publication 02-5, NASA Jet Propulsion Laboratory, Califoria Institute of Technology, Pasadena, California.

Hofmann-Wellenhof, B., H. Lictenegger, & J. Collins (1997) *GPS Theory and Practice*. Springer-Verlag, Wien - New York.

Hoolst, T. V., V. Dehant, & P. Defraigne (2000) Chandler wobble and Free Core Nutation for Mars. *Planetary and Space Science*, 48, 1145–1151.

Jones, B. W. (1999) *Discovering the Solar System*. John Wiley and Sons, Chichester, West Sussex, England.

Journey-Kaler (2002) Autonomous Navigation for Mars Landing Spacecraft Using a Proposed Martian Satellite Network. Colarado Center for Astrodynamics Research, ASEN5050 Project Webpage.
URL `http://www-ccar.colorado.edu/asen5050/projects/projects_2002/journey-kaler/`

JPL (2000) Mars Network Images. webpage.
URL `http://eis.jpl.nasa.gov/marsnet/images/images.html`

JPL (2004a) Beyond 2009. webpage.
URL `http://marsprogram.jpl.nasa.gov/missions/future/futureMissions.html`

JPL (2004b) Daily Updates - March 3, 2004 Spirit Status for sol 59. webpage.
URL `http://www.jpl.nasa.gov/mer2004/daily/3-3-04.cfm`

Justus, C., B. James, S. Bougher, A. Bridger, R. H. adn J.R. Murphy, & S. Engel (2002) MARS-GRAM 2000: A Mars Atmospheric Model for Engineering Applications. *Advances in Space Research*, 29(2), 193–202.

Kaplan, E. (1996) *Understanding GPS: Principles and Applications*. Artech House Publishers, Boston.

Kaula, W. M. (1966) *Theory of Satellite Geodesy*. Blaisdell Publishing Company, Toronto.

Keating, G., S. Bougher, R. Zurek, R. Tolsen, G. Cancro, S. Noll, J. Parker, T. Schellenberg, R. Shane, B. Wilkerson, J. Murphy, J. Hollingworth, R. Haberle, M. Joshi, J. Pearl, B. Conrath, M. Smith, R. Clancy, R. Blanchard, R. Wilmoth, D. Rault, T. Martin, D. Lyons, R. Esposito, M. Johnson, C. Whetzel, C. Justus, & J. Babicke (1998) The Structure of the Upper Atmosphere of Mars: In Situ Accelerometer Measurements from Mars Global Surveyor. *Science*, 279, 1672–1676.

King, J. H. (2001) Mars Fact Sheet. Webpage.
URL `http://mssdc.gsfc.nasa.gov/planetary/factsheet/marsfact.html`

Koch, K. R. (1999) *Parameter Estimation and Hypothesis Testing in Linear Models*. Springer-Verlag, New York.

Krakiwsky, E. J. (1990) *The Method of Least Squares: A Synthesis of Advances*, Volume 10003 of *UCGE Reports*. University of Calgary, Department of Geomatics Engineering, Calgary.

Kuo, N. (2000) Mars Network Operations Concept. *Proceedings of the 2000 IEEE Aerospace Conference, (March 18-25, 2000, Big Sky, Montana)*, Volume 2, 209–216.

Leick, A. (2004) *GPS Satellite Surveying*. John Wiley and Sons, Inc., Hoboken, New Jersey, 3 Edition.

LeMaster, E. A. & S. M. Rock (2002) An Improved Solution Algorithm for Self-Calibrating Pseudolite Arrays. *Proceedings of the Institute of Navigation 2002 National Technical Meeting (January 28-30, 2002, San Diego, California)*, 562–572.

Lemoine, F. G., D. E. Smith, D. Rowlands, M. Zuber, G. A. Neumann, D. S. Chinn, & D. Pavlis (2001) An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor. *Journal of Geophysical Research*, 106(E10), 23359–23376.
URL `http://bowie.gsfc.nasa.gov/926/MARS/GMM2B.html`

Lodders, K. & B. Fegley Jr. (1998) *The Planetary Scientists Companion*. Oxford University Press, New York.

Luo, N. (2001) *Precise Relative Positioning of Multiple Moving Platforms Using GPS Carrier Phase Observables*. Ph.D. thesis, The University of Calgary.
URL `http://www.geomatics.ucalgary.ca/Papers/Thesis/GL/00.20147.NLuo.pdf`

Ma, C. (2003) Integration of GPS and Cellular Networks to Improve Wireless Location Performance. *Proceedings of the Satellite Division of the Institute of Navigation 16th International Technical Meeting (September 9-12, 2003, Portland, Oregon)*, 1585–1596.

MacNicol, J. H. & J. F. Raquet (2002) A Study of Satellite Positioning Tehcniques for Use On and Around the Moon. *Proceedings of the Institute of Navigation 58th Annual Meeting, (June 24-26, 2002, Albuquerque, New Mexico)*, 506–515.

Matousek, S. (2002a) Electra Mars Proximity Link Communications and Navigation Payload Description. Webpage.

Matousek, S. (2002b) Mars Relay Description for Scout 2007 Proposals. Webpage.
URL `http://centuari.larc.nasa.gov/mars/marslib.html`

Matousek, S., K. Leschly, B. Gershman, & J. Reimer (1999) Mars Micromissions. *13th Annual AIAA/USU Conference on Small Satellites (August 23-26, 1999, Logan, Utah).*

Mendillo, M., X. Pi, S. Smith, C. Martinis, J. Wilson, & D. Hinson (2003) Ionospheric Effects upon a Satellite Navigation System at Mars. *submitted to Radio Science*.
URL `http://sirius.bu.edu/pub/preprints/mars3.pdf`

Misra, P. & P. Enge (2001) *Global Positioning System: Signals, Measurements, and Performance*. Ganga-Jamuna Press, Lincoln, Massachusetts.

MOLA Science Team (2004) Labeled MOLA Map of Mars. webpage.
URL `http://ltpwww.gsfc.nasa.gov/tharsis/map_lab.html`

Montenbruck, O. & E. Gill (2000) *Satellite Orbits: Models, Methods and Applications*. Springer, Berlin.

Moomaw, B. (2002) Mars Program Facing Collapse. Space Daily webpage news article.
URL `http://www.spacedaily.com/news/mar-general-02d1.html`

O'Keefe, K., S. Ryan, & G. Lachapelle (2002) Global Availability and Reliability Assessment of the GPS and Galileo Global Navigation Satellite Systems. *Canadian Aeronautics and Space Journal*, 48(2), 123–132.

Petovello, M. (2003) *Real-Time Integration of a Tactical-Grade IMU and GPS for High-Accuracy Positioning and Navigation*. Ph.D. thesis, University of Calgary.
URL `http://www.geomatics.ucalgary.ca/links/GradTheses.html`

Petovello, M. & G. Lachapelle (1999) Using Standalone GPS to Evaluate Precise Ocsillator Stability in the Time Domain. *Proceedings of the Satellite Division of the Institute of Navigation 12th International Technical Meeting (September 14-17, 1999, Nashville, Tennessee)*, 519–522.

Ray, J. (2000) *Mitigation of GPS Code and Carrier Phase Multipath Effects using a Multi-Antenna System*. Ph.D. thesis, University of Calgary.
URL `http://www.geomatics.ucalgary.ca/links/GradTheses.html`

Roosbeek, F. (1999) Analytical development of rigid mars nutation and tide generating potential series. *Celestial Mechanics and Dynamical Astronomy*, 75, 287–300.

Roosbeek, F. (2003) personal communication.

Roy, A. E. (1995) *The Foundations of Astrodynamics*. The Macmillan Company, New York.

Ryan, S. (2002) *Augmentation of Marine DGPS for Marine Navigation*. Ph.D. thesis, University of Calgary.
URL `http://www.geomatics.ucalgary.ca/links/GradTheses.html`

Schwarz, K. P. (1998) *Fundamentals of Geodesy*, Volume 10014 of *UCGE Reports*. University of Calgary, Department of Geomatics Engineering, Calgary.

Seeber, G. (1993) *Satellite Geodesy*. Walter de Gruyter, New York.

Shampine, L. F. & M. K. Gordon (1975) *Computer Solution of Ordinary Differential Equations*. Freeman and Company, San Francisco.

Sheehan, W. (1996) *The Planet Mars: A History of Observation and Discovery*. The University of Arizona Press, Tuscon, Arizona.

Shinagawa, H. (2000) Our Current Understanding of the Ionosphere of Mars. *Advances in Space Research*, 26(10), 1599–1608.

Smith, D., M. Zuber, S. Solomon, R. Phillips, J. Head, J. Garvin, W. Banerdt, D. Muhleman, G.H.Pettengill, G. Neumann, F. Lemoine, J. Abshire, O. Aharonson, C. Brown, S. Hauck, A. Ivanov, P. McGovern, H. Zwally, & T. Duxbury (1999) The global topography of Mars and implications for surface evolution. *Science*, 284(5419), 1495–1503.

Spilker Jr, J. J. (1996a) *Satellite Constellation and Geometric Dilution*, Volume 1, *Global Positioning System: Theory and Applications*, Chapter 5, 177–208. American Institute of Aeronautics and Astronautics, Washington.

Spilker Jr, J. J. (1996b) *Tropospheric Effects on GPS*, Volume 1, *Global Positioning System: Theory and Applications*, Chapter 13, 517–546. American Institute of Aeronautics and Astronautics, Washington.

Stansell, T. (1978) *The Transit Navigation Satellite System*. Magnavox Government and Industrial Electronics Company, Torrance, California.

Teunissen, P. J. G. (1990) An Integrity And Quality Control Procedure For Use In Multi Sensor Integration. *Proceedings of the Satellite Division of the Institute of Navigation International Technical Meeting (September 19-21, 1990, Colorado Springs, Colorado)*, 519–522.

van der Ha, J. & V. Modi (1977) Analytical Evaluation of Solar Radiation Induced Perturbations of Space Structures. *Journal of the Astronautical Sciences*, 25(4), 283–306.

Vaughan, R. (1997) Mars Pathfinder Trajectory Data. Webpage.
URL `http://mars.jpl.nasa.gov/MPF/mpfwwwimages/mpffootp.html`

Williams, D. R. (2001) Chronology of Mars Exploration. Webpage.
URL `http://nssdc.gsfc.nasa.gov/planetary/chronology_mars.html`

Wolf, R. (2000) Onboard Autonomous Integrity Monitoring using Intersatellite Links. *Proceedings of the Satellite Division of the Institute of Navigation 13th International Technical Meeting (September 19-22, 2000, Salt Lake City, Utah)*, 1572–1581.
URL `http://www.ifen.com/publications/IONGPS2000_ISL.pdf`

Wu, S., Y. Bar-Sever, S. Nandi, & L. Romans (1999) Netlander Positioning with the Mars Infrastructure Constellation and its Impact on Martian Geodesy. *1999 International Symposium on Space Communications and Navigation Technologies (September 21-23, 1999, Pasadena, California.*
URL `http://descansymposium.jpl.nasa.gov/database/pdf/506_wusc.pdf`

Yuan, D.-H., W. Sjogren, A. S. Konopliv, & A. Kucinskas (2001) Gravity field of Mars: A 75th Degree and Order Model. *Journal of Geophysical Research*, 106(E10), 23377–23401.

Yunck, T. P. (1996) *Orbit Determination*, Volume 2, *Global Positioning System: Theory and Applications*, Chapter 21, 177–208. American Institute of Aeronautics and Astronautics, Washington.

Appendix A

Satellite State Transition Matrix

The dynamics of a system governed by a set of first order differential equations may be described by the set of differential equations

$$\dot{\boldsymbol{x}} = f(t, \boldsymbol{x}) \tag{A.1}$$

where $\boldsymbol{x}$ is the state vector and $f(\boldsymbol{x}, t)$ is a function that depends on time and the state vector. If the system is linear, or linearized, $f(\boldsymbol{x}, t)$ depends linearly on $\boldsymbol{x}$ and may be replaced by the dynamics matrix, $\boldsymbol{F}$

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(t) \cdot \boldsymbol{x}(t) \tag{A.2}$$

This system of equations, and appropriate initial conditions for each of the states, defines the state vector for all time. However, for discrete estimation, a formulation is required that relates the state at one time to the state vector at any other time

$$\boldsymbol{x}_k = \boldsymbol{\Phi}_{k,k-1} \boldsymbol{x}_{k-1} \tag{A.3}$$

where $\boldsymbol{\Phi}$ is the transition matrix from epoch $k-1$ to epoch k. The transition matrix maps the state vector between two times similarly to how the design matrix, $\boldsymbol{A}$, maps between observation and state space. Like the design matrix, $\boldsymbol{\Phi}$ is a Jacobian matrix, i.e. a matrix of partial differentials of one vector with respect to another

$$\boldsymbol{\Phi} = \frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{x}_{k-1}} \tag{A.4}$$

For a state vector consisting of position, and velocity,

$$\boldsymbol{x} = [x, y, z, v_x, v_y, v_z]^T \tag{A.5}$$

the transition matrix would be of the form

$$\frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{x}_{k-1}} = \begin{bmatrix} \frac{\partial x_k}{\partial x_{k-1}} & \frac{\partial x_k}{\partial y_{k-1}} & \cdots & \frac{\partial x_k}{\partial v_{zk}} \\ \frac{\partial y_k}{\partial x_{k-1}} & \frac{\partial y_k}{\partial y_{k-1}} & \cdots & \frac{\partial y_k}{\partial v_{zk}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_{zk}}{\partial x_{k-1}} & \frac{\partial v_{zk}}{\partial y_{k-1}} & \cdots & \frac{\partial v_{zk}}{\partial v_{zk}} \end{bmatrix} \tag{A.6}$$

In most navigation applications, the the dynamics matrix $\mathbf{F}$ is a constant and the transition matrix may be obtained solving the differential equation

$$\dot{\boldsymbol{x}} = \boldsymbol{F} \cdot \boldsymbol{x}(t) \tag{A.7}$$

which has the standard solution

$$\boldsymbol{x}(t_k) = e^{\boldsymbol{F}\Delta t}\boldsymbol{x}(t_{k-1}) \tag{A.8}$$

where $\Delta t = t_k - t_{k-1}$. Using the matrix exponential power series expansion

$$e^{\boldsymbol{F}\Delta t} = \boldsymbol{I} + \boldsymbol{F}\Delta t + \frac{1}{2!}\boldsymbol{F}^2\Delta t^2 + \ldots \tag{A.9}$$

and retaining only the first two terms gives the standard result for the transition matrix given a constant dynamics matrix

$$\boldsymbol{\Phi} = \boldsymbol{I} + \boldsymbol{F}\Delta t \tag{A.10}$$

If the dynamics matrix is not a constant over the transition interval t_k, t_{k-1}, then the above solution does not apply. Instead, the transition matrix must be determined numerically.

The simplest approach, used by Wolf (2000), is to integrate the trajectory from epoch t_{k-1} to epoch t_k twice for each element in the transition matrix, starting with a small perturbation in each direction of each state vector element and then difference the the two results to form a numerical derivative or first difference of each element in the state vector with respect to every other element in the state vector.

In a more rigorous numerical technique, described by Montenbruck & Gill (2000), first order differential equations are derived for each element in the transition matrix. These equations are then integrated along with the equations of motion from epoch t_{k-1} to epoch t_k, with the initial condition $\boldsymbol{\Phi}(t_{k-1}, t_{k-1}) = \boldsymbol{I}$.

The following is a derivation of the first order differential equation of the transition matrix. If follows closely the derivation given by Montenbruck & Gill (2000).

Given that the transition matrix is defined as the Jacobian matrix

$$\boldsymbol{\Phi}(t_k, t_{k-1}) = \frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{x}_{k-1}} \tag{A.11}$$

and given that the dynamic system in question obeys the first order differential equation

$$\frac{d}{dt}\boldsymbol{x}_k = f(t_k, \boldsymbol{x}_k) \tag{A.12}$$

The first order differential equation of the transition matrix may be obtained by taking the partial differential of each side of the dynamics equation for time t_k with respect to the state vector at a previous time, t_{k-1}

$$\frac{\partial}{\partial \boldsymbol{x}_{k-1}} \frac{d}{dt}\boldsymbol{x}_k = \frac{\partial f(t_k, \boldsymbol{x}_k)}{\partial \boldsymbol{x}_{k-1}} \ , \tag{A.13}$$

multiplying the right hand side by the identity matrix $\partial \boldsymbol{x}_k / \partial \boldsymbol{x}_k$

$$\frac{\partial}{\partial \boldsymbol{x}_{k-1}} \frac{d}{dt}\boldsymbol{x}_k = \frac{\partial f(t, \boldsymbol{x}_k)}{\partial \boldsymbol{x}_{k-1}} \cdot \frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{x}_k} \tag{A.14}$$

and rearranging the left and right sides to give

$$\frac{d}{dt} \frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{x}_{k-1}} = \frac{\partial f(t, \boldsymbol{x}_k)}{\partial \boldsymbol{x}_k} \cdot \frac{\partial \boldsymbol{x}_k}{\partial \boldsymbol{x}_{k-1}} \tag{A.15}$$

Substituting for the transition matrix gives

$$\frac{d}{dt}\boldsymbol{\Phi}(t, t_{k-1}) = \frac{\partial f(t, \boldsymbol{x}_k)}{\partial \boldsymbol{x}_k} \cdot \boldsymbol{\Phi}(t, t_{k-1}) \tag{A.16}$$

which is a set of first order differential equations for each element in $\boldsymbol{\Phi}(t_k, t_{k-1})$. In order to integrate this set of equations, $\partial f(t, \boldsymbol{x}_k)/\partial \boldsymbol{x}_k$ must be evaluated along the path of $\boldsymbol{x}$ in state space. Functions for the partial derivatives of the of the dynamics function with respect to state vector elements are required and these 36 differential equations must be integrated along with the 6 differential equations for the state vector itself. Numerically integrating a system of 42 differential equations is computationally expensive. If a linearized Kalman filter is used, the computation of the reference trajectory and corresponding transition matrix can be done offline. The result is a time series of n transition matrices, each mapping the state vector from the initial epoch t_0 to some epoch t_k in the future. The transition matrix between two arbitrary epochs $k-1$ and k is obtained by inverting one of the transition matrices, ie. mapping from the $k-1$ epoch to the 0 epoch and then back to the k epoch.

$$\boldsymbol{\Phi}(t_k, t_{k-1}) = \boldsymbol{\Phi}^{-1}(t_0, t_{k-1})\boldsymbol{\Phi}(t_0, t_k) \tag{A.17}$$

However if an extended Kalman filter is used, both the reference trajectory and the transition matrix must be re-evaluated after every observation update.

Appendix B

Mars Geodetic Formulas

B.1 Mars Rotation

Mars rotation is given by Airy Mean Siderial Time (AMST) which is the hour angle of the Mars mean vernal equinox observed from the Mars prime meridian.

$$\text{AMST} = 313.476 + 350.8919852T \tag{B.1}$$

where T is the time in days since the J2000 epoch. Airy Apparent Siderial time (AAST) is obtained by adding the Mars equation of the equinoxes to AMST. The equation of the equinoxes is obtained, analogously to that of Earth, from the nutation in longitude and the mean obliquity (Schwarz, 1998).

$$\text{AAST} = \text{AMST} + \Delta\psi \cos\epsilon \tag{B.2}$$

Nutation and mean obliquity are discussed below.

B.2 Mars Precession

The precession of Mars was determined by Folkner et al. (1997) to be -7576 mas/year. The precession matrix $\boldsymbol{P}_{t0,t}$ is

$$\boldsymbol{P}_{t0,t} = \boldsymbol{R}_1(-\epsilon_t)\boldsymbol{R}_3(\pi)\boldsymbol{R}_1(\epsilon_{t0}) \tag{B.3}$$

where π is the precession angle, and ϵ_{t0} and ϵ_t are the mean obliquities of Mars at times $t0$ and t respectively. The mean obliquity of Mars is given by Roosbeek (1999) as

$$25.192028020^\circ + 42.995625822"T \tag{B.4}$$

where T is the time since the J2000 epoch in centuries.

B.3 Mars Nutation

The Roosbeek Mars Analytical Nutation Series consists of Fourier series for nutation in longitude and the nutation in obliquity of Mars. The angular arguments are linear combinations of the mean longitudes of Saturn, Jupiter, Mars, Earth, Phobos, and Deimos. Formulas for each term can be found in Roosbeek (1999). Each series is of the form

$$\Delta = \sum_{i=1}^{n}(S_i \sin f_i + C_i \cos f_i) \tag{B.5}$$

where S_i and C_i are the sine and cosine terms given in Table B.1, and f_i is the phase obtained by taking the linear combination of the fundamental arguments given in the first six columns of Table B.1. The period of each f_i is also given in Table B.1. The nutation matrix, $\boldsymbol{N}$, is then obtained from the nutation angles, $\Delta\epsilon$ and $\Delta\psi$ by

$$\boldsymbol{N} = \boldsymbol{R}_1(-\epsilon - \Delta\epsilon)\boldsymbol{R}_3(-\Delta\psi)\boldsymbol{R}_1(\epsilon) \tag{B.6}$$

where epsilon is the mean obliquity of Mars at the time of interest.

Table B.1: Roosbeek Mars nutation series

i	Sa	Ju	Ma	Te	NPh	NDe	T(days)	C_i(mas)	S_i(mas)
	Longitude Terms								
1	0	0	7	0	0	0	98.140	-0.10	0.00
2	0	0	6	0	0	0	114.497	-0.89	0.25
3	0	0	5	0	0	0	137.396	-6.26	-0.89
4	0	0	4	0	0	0	171.745	-34.82	-21.66
5	0	-2	4	0	0	0	186.533	0.00	-0.13
6	0	0	3	0	0	0	228.993	-137.00	-200.03
7	0	-1	3	0	0	0	241.772	0.00	0.16
8	0	-2	3	0	0	0	256.061	0.00	-0.16
9	0	-3	10	-4	0	0	343.309	0.31	0.00
10	0	0	2	0	0	0	343.490	-220.64	-1108.21
11	0	3	-6	4	0	0	343.671	0.28	-0.15
12	0	-2	2	0	0	0	408.217	0.00	-0.23
13	0	0	1	0	0	0	686.980	-282.42	-477.62
14	0	0	-1	1	0	0	779.936	0.00	-0.12
15	0	-1	1	0	0	0	816.435	0.00	0.44
16	0	0	0	0	-1	0	825.641	0.00	9.88
17	0	0	-3	2	0	0	901.985	0.00	-0.11
18	0	-2	1	0	0	0	1006.010	0.21	-0.40
19	0	1	0	0	0	0	4332.590	0.00	0.20
20	0	0	2	-1	0	0	5764.010	0.20	0.19
21	0	0	0	0	0	-1	19998.900	0.00	4.39
22	5	-2	0	0	0	0	322615.000	-0.30	-0.28
23	0	-3	8	-4	0	0	651393.000	0.00	0.73
24	-3	0	19	-10	0	0	2341002.000	2.57	0.00
	Obliquity Terms								
1	0	0	6	0	0	0	114.497	-0.12	-0.42
2	0	0	5	0	0	0	137.396	0.43	-2.93
3	0	0	4	0	0	0	171.745	10.21	-16.19
4	0	0	3	0	0	0	228.993	93.51	-62.65
5	0	-3	10	-4	0	0	343.309	0.00	0.14
6	0	0	2	0	0	0	343.490	507.40	-88.41
7	0	3	-6	4	0	0	343.671	0.00	0.13
8	0	0	1	0	0	0	686.980	-47.68	-11.94
9	0	0	0	0	-1	0	825.641	4.20	0.00
10	0	0	0	0	0	-1	19998.900	1.86	0.00

Appendix C

Description of Computer Programs

This appendix describes the software that was developed for the research described in this thesis. Two types of software were developed, simulation software and processing software. The two software development stages were done separately in order to avoid common logic errors between the two. The simulation software includes a positioning and navigation geometry and reliability simulator, a dynamical satellite orbit simulator, a GNSS observation simulator and various other utility programs. The processing software consists a single program developed to implement the autonomous filtering strategy described in Chapter 6.

C.1 Simulation Software

C.1.1 Geometry Simulator

The results in Chapter 4 were obtained with a geometry simulator program developed specifically for Mars. The program works by calculating the positions of all satellites and simulated users and then computing the relevant design matrices and observations covariance matrices for each user at each simulated observation time.

These values are then used to evaluate the availability, accuracy, and reliability parameters for that user at that epoch. It is based on SimGNSS1 (formerly known as GGPlan4.6), a single point user GNSS preplanning software package developed at the University of Calgary by Thom Morley and Sam Ryan (Ryan, 2002), however it has been completely rewritten to make it possible to use multiple sources of satellite position information, to more easily and efficiently accommodate large grids of users, and to output results in a format that is more easily imported into MatLab for graphical interpretation. Unlike SimGNSS1, it is also capable of computing cumulative values of DOP and reliability parameters for static users using a sequential least squares solution.

The input to the program is a YUMA format almanac file containing the initial positions of the satellites, and an option file containing input parameters such as the observation rate, user equivalent range error, elevation mask, and the positions of the simulated users.

C.1.2 Numerical Orbit Simulator

The second piece of simulation software is a numerical orbit simulator, used to produce simulated satellite orbits in Chapter 5. The program takes an initial state vector for each satellite and uses functions describing the acceleration as a function of position, velocity, and time to numerically integrate the orbit of each satellite over time. This software contains functions that calculate all of the geodetic quantities of Mars described in Chapter 2, as well all of the modelled forces described in Chapter 5.

The input to this software is an initial position file containing either the planet-centred inertial coordinates or Keplerian elements of each satellite, as well as a mass, reflection coefficient, and drag coefficient for each satellite. An option file is used to set other parameters such as start and end time, output rate, and whether or not to simulate certain forces.

The output of this module consists of three files for each satellite `CTprnx.txt`, `CIprnx.txt`, and `Keprnx.txt`, where `x` is the unique number (ie. PRN) assigned to each satellite. The first two files contain the planet-centred planet-fixed and planet-centred inertial positions and velocities where each row of each file is of the format $(t, x, y, z, v_x, v_y, v_z)$. The third file contains the instantaneous Keplerian elements of the satellite where each row is of the format $(t, a, e, i, \Omega, \omega, M)$.

C.1.3 Observation and Error Source Simulator

The final simulation tool is a Mars GNSS observation simulator that has been developed using some of the functions of the software package SimGNSS2 developed by Ning Luo, Paul Alves and Olivier Julien (Luo, 2001; Alves, 2001). It is designed to simulate observations between satellites and landers by reading in the inertial position output files of the orbit simulator (or any other source) for each satellite and similar files for each lander and calculating the true range and range rate between each element and then applying error models to generate error-affected range and Doppler observations. The error models described in Chapter 5 are implemented in the program. The multipath model was taken from SimGNSS2 while the other models were developed specifically for this application. The output of this program consists of files names `Obsxxyy.txt` where `xx` is the number of the receiving user/satellite and `yy` is the number of the sending user/satellite.

C.1.4 Subroutines and Functions Borrowed from Other Authors

The numerical integration functions, specifically the class SAT_DE, used in the orbit simulator were taken from Montenbruck & Gill (2000), in which the authors have granted limited license for individual non-commercial use of their software. The multipath model in the observation simulator was taken from SimGNSS2 (Luo, 2001).

C.2 Processing Software

The processing software used in Chapter 6 consists of a single program that reads in observation files generated by the Observation Simulator described above. This observation files are used in conjunction with an option file containing initial states, covariances, and other information. The program implements a separate object for each element in the network that contains a Kalman filter and utilities for reading in observations and communicating with other elements in the network. Each element object is able to ask other elements for copies of their state vectors and covariances matrices in order to compute position updates. Each element is also capable of computing a reference trajectory for the satellite based on a degree and order 20 gravity model of Mars. This is accomplished with the same routine as is used in the numerical orbit simulator described above. This is the only function that is used by both the simulator and the processing software, though in the simulator, the gravity model is evaluated to degree and order 80. The program outputs two files for each element, one containing a time series of the estimated state of each element, and one containing a time series of the filter activity of each element: i.e. listing when observations were made, what the geometry of the observation was, and what the predicted residual of the observation was.

Appendix D

Keplerian Elements and Cartesian Coordinates

D.1 Conversion from Cartesian Coordinates to Keplerian Elements

The six Keplerian elements can be written as functions of inertial position and velocity of the satellite, provided that the mass of the central body is known, and the the satellite is not in a perfectly circular or equatorial orbit. The following formulas are taken from Montenbruck & Gill (2000) but are found in any textbook that discusses orbital mechanics.

The inclination, i, and the right ascension of the ascending node, Ω, can be determined from the cross product of the position and velocity vectors.

$$\boldsymbol{h} = \boldsymbol{r} \times \dot{\boldsymbol{r}} \tag{D.1}$$

This is sometimes referred to as the areal velocity vector. This vector is normal to the orbital plane and is a constant quantity in Keplerian motion (resulting in a constant orbital plane, represented by the inclination and right ascension of the ascending node). i and Ω are obtained from the components of $\boldsymbol{h}$

$$i = \tan^{-1}\left(\frac{\sqrt{h_x^2 + h_y^2}}{h_z}\right) \tag{D.2}$$

$$\Omega = \tan^{-1}\left(\frac{h_x}{-h_y}\right) \tag{D.3}$$

The semi-major axis is also obtained from the position and velocity,

$$a = \left(\frac{2}{r} - \frac{v^2}{GM}\right)^{-1} \tag{D.4}$$

where GM is the gravitational constant of the central body. The eccentricity can then be obtained from the areal velocity and the semi-major axis

$$e = \sqrt{1 - \frac{h^2}{GMa}} \tag{D.5}$$

The two remaining Keplerian elements are more complicated. To obtain the mean anomaly, M, the eccentric anomaly, E, must be determined first

$$E = \tan^{-1}\left(\frac{\boldsymbol{r} \cdot \dot{\boldsymbol{r}}/(a^2 n)}{1 - r/a}\right) \tag{D.6}$$

where n is the mean motion, or mean orbital speed in radians per second

$$n = \sqrt{\frac{GM}{a^3}} \tag{D.7}$$

The mean anomaly is given by Kepler's equation

$$M(t) = E(t) - e \sin E(t) \tag{D.8}$$

Finally, the argument of perigee, ω, is obtained from the difference between the argument of latitude, u, and the true anomaly, ν

$$\omega = u - \nu \tag{D.9}$$

where

$$u = \tan^{-1}\left(\frac{z}{-xh_y/h + yh_x/h}\right) \tag{D.10}$$

and

$$\nu = \tan^{-1}\left(\frac{\sqrt{1 - e^2} \sin E}{cosE - e}\right) \tag{D.11}$$

D.2 Conversion from Keplerian Elements to Cartesian Coordinates

There conversion from Keplerian elements to Cartesian coordinates is as follows. First the position vector, $\hat{\boldsymbol{r}}$ of the satellite is determined in the plane of the orbit, where the

x-axis points in the direction of the perigee. The true anomaly is the angle between the x-axis and the position vector so the orbital plane coordinates are

$$\hat{x} = r\cos\nu = a(\cos E - e) \tag{D.12}$$

$$\hat{y} = r\sin\nu = a\sqrt{1-e^2}\sin E \tag{D.13}$$

The eccentric anomaly is obtained from the mean anomaly by evaluating equation D.8 iteratively. Equatorial Cartesian coordinates are then obtained by rotating the position vector from the orbital plane to the planet centred-inertial frame using three rotations.

$$\boldsymbol{r} = R_3(-\Omega)R_1(-i)R_3(-\omega)\hat{\boldsymbol{r}} \tag{D.14}$$

The velocity vector is obtained by determining the velocity in the orbital plane and likewise rotating it into the planet-centred inertial frame. The x and y velocities in the orbital frame are

$$\begin{aligned} \dot{\hat{x}} &= -a\sin E\dot{E} && \text{(D.15)} \\ &= -n\frac{a^2}{r}\sin E \\ \dot{\hat{y}} &= a\sqrt{1-e^2}\cos E\dot{E} && \text{(D.16)} \\ &= \frac{a^2}{r}\sqrt{1-e^2}\cos E \end{aligned}$$