Aims and Scope
The Journal of Global Positioning Systems is an official publication of the International Association of Chinese Professionals in Global Positioning Systems (CPGPS). It is published twice a year, in June and December. The Journal will have both print version (ISSN 1446-3156) and CD-ROM version (ISSN 1446-3164), which can be accessed through the CPGPS website at http://www.cpgps.org/journal/journal.html. Whilst CPGPS owns all the copyright of all text material published in the Journal, the authors are responsible for the views and statement expressed in their papers and articles. Neither the authors, the editors nor CPGPS can accept any legal responsibility for the contents published in the journal.

Editor-in-Chief
Jinling Wang
The University of New South Wales, Sydney, Australia
jinling.wang@unsw.edu.au

Editorial Board
RuiZhi Chen
Finnish Geodetic Institute, Finland

Wu Chen
Hong Kong Polytechnic University, Hong Kong

Dorota Grejner-Brzezinska
Ohio State University, United States

Ren Da
Bell Labs/Lucent Technologies, Inc., United States

C.D. de Jong
Delft University of Technology, The Netherlands

Hans-Jürgen Euler
Leica Geo-systems, Switzerland

Yanming Feng
Queensland University of Technology, Australia

Yang Gao
University of Calgary, Canada

Shaowei Han
Thales Navigation, United States

Changdon Kee
Seoul National University, Korea

Hansjoerg Kutterer
Deutsches Geodätisches Forschungsinstitut, Germany

Jiancheng Li
Wuhan University, China

Esmond Mok
Hong Kong Polytechnic University, Hong Kong

J.F. Galera Monico
Departamento de Cartografia, FCT/LINEP, Brazil

Günther Retscher
Vienna University of Technology, Austria

Gethin Roberts
University of Nottingham, United Kingdom

Rock Santerre
Laval University, Canada

Bruno Scherzinger
Applanix Corporation, Canada

Salah Sukkarieh
The University of Sydney, Australia

Todd Walter
Stanford University, United States

Lambert Wanninger
Ingenieurherr Wanninger, Germany

Caijun Xu
Wuhan University, China

Guochang Xu
GeoForschungsZentrum (GFZ) Potsdam, Germany

Ming Yang
National Cheng Kung University, Taiwan

Kefei Zhang
RMIT University, Australia

Guoqing Zhou
Old Dominion University, United States

Editorial Advisory Board
Junyong Chen
National Bureau of Surveying and Mapping, China

Yongqi Chen
Hong Kong Polytechnic University, Hong Kong

Paul Cross
University College London, United Kingdom

Guenter Hein
University FAF Munich, Germany

Gerard Lachapelle
University of Calgary, Canada

Jingnan Liu
Wuhan University, China

Keith D. McDonald
NaviTech, United States

Chris Rizos
The University of New South Wales, Australia

Peter J.G. Teunissen
Delft University of Technology, The Netherlands

Sien-Chong Wu
Jet Propulsion Laboratory, NASA, United States

Yilin Zhao
Motorola, United States

IT Support Team
Satellite Navigation & Positioning Group
The University of New South Wales, Australia

Publication and Copyright
The Journal of Global Positioning Systems is an official publication of the International Association of Chinese Professionals in Global Positioning Systems (CPGPS). It is published twice a year, in June and December. The Journal will have both print version (ISSN 1446-3156) and CD-ROM version (ISSN 1446-3164), which can be accessed through the CPGPS website at http://www.cpgps.org/journal/journal.html. Whilst CPGPS owns all the copyright of all text material published in the Journal, the authors are responsible for the views and statement expressed in their papers and articles. Neither the authors, the editors nor CPGPS can accept any legal responsibility for the contents published in the journal.

Subscriptions and Advertising
Membership dues in the CPGPS include the subscription to the Journal during the period of membership. Subscriptions from non-members and advertising inquiries should be directed to:

CPGPS Headquarters
Department of Geomatics Engineering
The University of Calgary
Calgary, Alberta, Canada T2N 1N4
Fax: +1(403) 284-1980
E-mail: journal@cpgps.org

© CPGPS, 2002. All the rights reserved.
## Table of Contents

Guenter W. Hein and Thomas Pany .......................................................... 73

**GPS Attitude Determination Reliability Performance Improvement Using Low Cost Receivers**  
Chaochao Wang and Gérard Lachapelle ................................................... 85

**Joint Treatment of Random Variability and Imprecision in GPS Data Analysis**  
Hansjörg Kutterer ..................................................................................... 96

**Short-Arc Batch Estimation for GPS-based Onboard Spacecraft Orbit Determination**  
Ning Zhou, Yanming Feng ......................................................................... 106

**Internet-based GPS VRS RTK Positioning with a Multiple Reference Station Network**  
G.R. Hu, V. H.S. Khoo, P. C. Goh, C. L. Law ........................................... 113

### Discussion Forum

**A General Criterion of Integer Ambiguity Search**  
Guochang Xu ............................................................................................ 122

### Experts Forum

**Multiple Reference Station Approach: Overview and Current Research**  
Gerard Lachapelle and Paul Alves .......................................................... 133

**Virtual Reference Station Systems**  
Herbert Landau, Ulrich Vollath, and Xiaoming Chen ............................... 137

**Network RTK Research and Implementation: A Geodetic Perspective**  
Chris Rizos .............................................................................................. 144

**Introduction to the Wide Area Augmentation System**  
Todd Walter ............................................................................................. 151

Guenter W. Hein and Thomas Pany
Institute of Geodesy and Navigation, University FAF Munich

Received: 23 December 2002 / Accepted: 24 December 2002

Abstract. This paper starts with a brief discussion of the Galileo project status and with a description of the present Galileo architecture (space segment, ground segment, user segment). It focuses on explaining special features compared to the American GPS system. The presentation of the user segment comprises a discussion of the actual Galileo signal structure. The Galileo carrier frequency, modulation scheme and data rate of all 10 navigation signals are described as well as parameters of the search and rescue service. The navigation signals are used to realize three types of open services, the safety of life service, two types of commercial services and the public regulated service. The signal performance in terms of the pseudorange code error due to thermal noise and multipath is discussed as well as interference to and from other radionavigation services broadcasting in the E5 and E6 frequency band. The interoperability and compatibility of Galileo and GPS is realized by a properly chosen signal structures in E5a/L5 and E2-L1-E1 and compatible geodetic and time reference frames. Some new results on reciprocal GPS/Galileo signal degradation due to signal overlay are presented as well as basic requirements on the Galileo code sequences.

Key words: GPS, Galileo, Signal Design, European

1 Introduction

Based on the communication of the European Commission of February 9, 1999 the satellite navigation system Galileo is presently under development in Europe (European Commission, 1999). The goal on the one hand is to achieve independence of and also an effective supplementation to the GPS. On the other hand the aim is to considerably improve Europe's capability to gain and preserve an important share of the world market for satellite navigation and related applications and services. The strategic and commercial importance of Galileo for Europe is out of discussion. This has repeatedly been confirmed at the European Council conferences in Cologne, Feira, Nice, Stockholm and at last in Laaken on December 14 and 15, 2001 (Belgian EU Presidency, 2001).

2 Status

The Galileo project is carried out in co-operation by bodies of the European Union (EU) and the European Space Agency (ESA). In principle it is to be realised in three phases: project definition, development and implementation. The fundamental decision for the realization Galileo was made by the council of the European ministers of transport at March 26, 2002. According to the present planning the development and validation phase should cover the period 2002-2005, the implementation phase 2006-2007, and the operational phase could start in 2008 (Fig. 1).

3 Architecture

The main characteristics of the Galileo system architecture can be summarised as follows (Weber et al., 2001):

- Independence of other satellite navigation systems
- Interoperability with GPS (GLONASS)
- Service concept (open, commercial, safety critical, regulated)
- Implementation of an Integrity Service (inside/outside Europe)
- Independence between Integrity Service and Galileo control System (GCS)
Global services (SAR, and referred to navigation data related services)
- Global location and time dissemination on the basis of a global constellation
- Regional components (Monitor and uplink stations)

Integration with regional systems (e.g., EGNOS)
- Integration with local (differential etc.) systems
- Compatibility with future mobile radio networks (UMTS)
It can be seen from Fig. 2 that the main extension of Galileo compared to GPS consists in the implementation of a global / regional segment for integrity monitoring. The goal is to assist the safety critical aircraft navigation (landing approach CAT I) and to locate and guide railway trains (Train control).

**Space Segment**

The space segment of Galileo is intended to consist of a total 30 Mean Earth Orbiting (MEO) satellites configured as walker 27/3/1 (+ 3 replacement satellites) constellation (Benedicto et al., 2000), i.e. distributed over three orbital planes (Fig. 3). The altitude is 23616 km, and the inclination is 56°. The satellite design (Fig. 4) is based on already carried out precursor programs (e.g. GLOBALSTAR) including critical payload technologies, which are developed in accompanying ESA programs. The Galileo satellite has a mass of 625 kg, generates a primary power of 1500 W and belongs with dimensions of 2.7 x 1.2 x 1.1 m³; to the category mini-satellites. The satellite comprises all standard systems for orbit and attitude control, thermal control, etc. Unlike GPS, also Laser retro-reflectors will be integrated in order to assist the orbit determination by satellite Laser ranging.

The navigation payload is the heart of the Galileo satellite. The payload is a regenerative transponder with modern digital and semiconductor technology applied to the essential subsystems. It consists of atomic clocks (Clock Monitoring and Control Unit), the signal generator (Navigation Signal Generation Unit) with CPU, the frequency generator (FPGU), the output amplifier (Solid State Power Amplifier) and the L-band antenna sub-system. As atomic clocks two Rubidium standards (5 \(10^{-15}\) over 100 s) and two space-borne H-Masers (5 \(10^{-14}\) over 10000 s) are to be used.

**Ground Segment**

As already outlined, the Galileo ground segment comprises the control segment for operation as well as orbit and time determination (GCS or Ground Control Segment) and the system for integrity monitoring (IDS or Integrity Determination System).

The number of elements in the GCS and the IDS are under further investigation in the present definition phase. The GCS will consist of about 12-15 reference stations, 5 up-link stations and two control centres. The IDS for Europe will include 16-20 monitor stations, three up-link stations for integrity data and two central stations for integrity computations. In the European area the integration with the EGNOS ground segment plays an important role.

**User Segment**

Like with GPS the Galileo user segment consists of all users on land, on water, in the air and in space. Fig. 5 shows the shares of various application on the European GNSS market as predicted for 2005 (volume 8 billion EURO).

Tab. 1 displays the requirements on the Galileo performance parameters (elevation mask, accuracy, coverage, availability, integrity) as posed by two basically different applications. The requirements for
safety critical applications are identical with the aviation specifications for the precise landing approach of CAT I. In case of the requirements from the mass market it is important that these apply for elevation masks above 25°. This accounts for the specific conditions of land navigation in urban areas (obstructions, multi-path).

In case of the requirements from the mass market it is important that these apply for elevation masks above 25°. This accounts for the specific conditions of land navigation in urban areas (obstructions, multi-path).

<table>
<thead>
<tr>
<th>Tab. 1 Selected requirements on Galileo</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter</td>
</tr>
<tr>
<td>elevation mask</td>
</tr>
<tr>
<td>accuracy (95 %)</td>
</tr>
<tr>
<td>coverage</td>
</tr>
<tr>
<td>availability</td>
</tr>
<tr>
<td>integrity</td>
</tr>
</tbody>
</table>

4 The Galileo Frequency And Signal Baseline

A tentative Galileo frequency and signal plan was presented at the ION GPS-2001 (Hein et al, 2001) which became meanwhile the baseline for the development of Europe’s satellite navigation system. Over the last months several modifications took place leading to a refined signal structure. The main changes and add-ons are described in the following and after that the complete signal structure will be outlined.

Recent Developments

In the lower L-band (i.e. E5a and E5b) the central frequency for E5b was moved to 1207.140 MHz in order to minimize possible interference from the Joint Tactical Information Distribution System (JTIDS) and the Multifunctional Information Distribution System (MIDS). All signals on E5a and E5b are using chip rates of 10 Mcps. The modulation for that band is still being optimized with the possibility to process very wideband signals by jointly using the E5a and E5b bands. This joint use of the bands has the potential to offer enormous accuracy for precise positioning with a low multipath. Data rates have also been fixed.

In the middle (i.e. E6) and upper (i.e. E2-L1-E1) L-band data and chip rates were also defined as well as Search and Rescue (SAR) up- and downlink frequencies.

Extensive interference considerations took place in E5a/E5b concerning Distance Measuring Equipment (DME), the Tactical Air Navigation System (TACAN) and the Galileo overlay on GPS L5; in E6 concerning the mutual interference to/from radars and in E2-L1-E1 frequencies with regard to the Galileo overlay on GPS L1.

The EC Signal Task Force and ESA have refined criteria for the code selection and have as well formulated the requirements on each frequency. Reference codes have been selected allowing initial assessments. Parallel investigations are on-going addressing alternate solutions for the Galileo codes and targeting improved performances, see e.g. (Pratt, 2002).

Frequencies and Signals

Galileo will provide 10 navigation signals in Right Hand Circular Polarization (RHCP) in the frequency ranges 1164-1215 MHz (E5a and E5b), 1215-1300 MHz (E6) and 1559-1592 MHz (E2-L1-E1\(^1\)), which are part of the Radio Navigation Satellite Service (RNSS) allocation. An overview is shown in Fig. 7, indicating the type of modulation, the chip rate and the data rate for each signal. The carrier frequencies, as well as the frequency bands

\(^1\) The frequency band E2-L1-E1 is sometimes denoted as L1 for convenience.
that are common to GPS or to GLONASS are also highlighted.

All the Galileo satellites will share the same nominal frequency, making use of Code Division Multiple Access (CDMA) compatible with the GPS approach.

Six signals, including three data-less channels, so-called pilot tones (ranging codes not modulated by data), are accessible to all Galileo Users on the E5a, E5b and L1 carrier frequencies for Open Services (OS) and Safety-of-life Services (SoL). Two signals on E6 with encrypted ranging codes, including one data-less channel are accessible only to some dedicated users that gain access through a given Commercial Service (CS) provider. Finally, two signals (one in E6 band and one in E2-L1-E1 band) with encrypted ranging codes and data are accessible to authorized users of the Public Regulated Service (PRS).

Fig. 7 Galileo frequency spectrum

<table>
<thead>
<tr>
<th>freq. Bands</th>
<th>E5a</th>
<th>E5b</th>
<th>E6</th>
<th>E2-L1-E1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Channel</td>
<td>I</td>
<td>Q</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>mod. type</td>
<td>I Q</td>
<td>I Q</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>chip rates</td>
<td>Mcps</td>
<td>Mcps</td>
<td>Mcps</td>
<td>Mcps</td>
</tr>
<tr>
<td>symbol rates</td>
<td>sps</td>
<td>N/A</td>
<td>250 sps</td>
<td>N/A</td>
</tr>
<tr>
<td>user min. received power at 10° elevation</td>
<td>DBW</td>
<td>DBW</td>
<td>DBW</td>
<td>DBW</td>
</tr>
</tbody>
</table>

(*) In case of separate modulation of E5a and E5b signals

Tab. 2 Main Galileo navigation signal parameters

---

2 Quadrature Phase Shift Keying
3 Binary Phase Shift Keying
A \( \frac{1}{2} \) rate Viterbi convolutional coding scheme is used for all the transmitted signals.

Four different types of data are carried by the different Galileo signals:

- **OS data**, which are transmitted on the E5a, E5b and E2-L1-E1 carrier frequencies. OS data are accessible to all users and include mainly navigation data and SAR data.
- **CS data** transmitted on the E5b, E6 and E2-L1-E1 carriers. All CS data are encrypted and are provided by some service providers that interface with the Galileo Control Centre. Access to those commercial data is provided directly to the users by the service providers.
- **SoL data** that include mainly integrity and Signal in Space Accuracy (SISA) data. Access to the integrity data may be controlled.
- **PRS data**, transmitted on E6 and L1 carrier frequencies.

A synthesis of the data mapping on Galileo signals is provided in Tab. 2.

**Modulation Schemes**

Given the frequency plan defined earlier and the target services based on the Galileo signals, the type of modulation of the various Galileo carriers are resulting from a compromise between the following criteria:

- Minimization of the implementation losses in the Galileo satellites, making use of the current state of the art of the related equipments.
- Maximization of the power efficiency in the Galileo satellites.
- Minimization of the level of interference induced by the Galileo signals in GPS receivers.
- Optimization of the performance and associated complexity of future Galileo user receivers.

The modulation chosen for each of the Galileo carrier frequency is presented in the following subsections. For the E5 band in particular, the trade-off analysis is on going between two alternate solutions that will be both described.

The main modulation parameters for Galileo signals are summarized on the Tab. 2. The following notation is used:

- \( S_X^Y(t) \) is the rectangular subcarrier on the Y channel in the X frequency band.
- \( m \) is a modulation index, associated to the modified Hexaphase modulation.

**Modulation of the E5 Carrier**

The modulation of E5 will be done according to one of the following schemes:

A. Two QPSK(10) signals will be generated coherently and transmitted through two separate wideband channels on E5a and E5b respectively. The two separate E5a and E5b signals will be amplified separately and combined in RF through an output multiplexer (OMUX) before transmission at the 1176.45 MHz and 1207.14 MHz respective carrier frequencies.

B. One single wideband signal generated following a modified BOC(15,10)\(^4\) modulation called AltBOC(15,10) modulation. This signal is then amplified through a very wideband amplifier before transmission at the 1191.795 MHz carrier frequencies.

In case A the E5 signal can be written as:

\[
S_E(t) = \left( C_{E5a}^I(t) D_{E5a}^I(t) \cos(2\pi F_{E5a} t) - C_{E5a}^Q(t) \sin(2\pi F_{E5a} t) \right)
+ \left( C_{E5b}^I(t) D_{E5b}^I(t) \cos(2\pi F_{E5b} t) - C_{E5b}^Q(t) \sin(2\pi F_{E5b} t) \right)
\]

The modulation in case B is a new modulation concept which main interest is that it combines the two signals (E5a and E5b) in a composite constant envelope signal which can then be injected through a very wideband channel. This wideband signal then can then be exploited in the receivers.

A detailed description of the AltBOC modulation can be found in (Ries et al., 2002b).

Implementation trade-offs and performance comparison between the processing of the very wideband BOC(15,10)-like signal and the joint processing of two separate QPSK signals of 10 Mcps on E5a and E5b is on-going.

**Modulation of the E6 Carrier**

The E6 signal contains three channels that are transmitted at the same E6 carrier frequency. The multiplexing scheme between the three carriers is a major point under consideration today, which shall be carefully optimized. This optimization process shall take into account payload

\(^4\) BOC(\( f_s, f_c \)), denotes a Binary Offset Carrier modulation with a subcarrier frequency \( f_s \) and a code rate \( f_c \).
and receivers implementation complexity and associated performances (including compatibility aspects).

The investigated solutions are time multiplexing and a modified Hexaphase modulation (so-called Interplex modulation). The modified Hexaphase is taken as baseline but the final selection process is on going between those two potential solutions. A QPSK signal resulting from the combination of two channels is phase modulated with the third channel. The modulation index \( m \) is used to set the relative power between the three channels.

Using a Hexaphase modulation, the E6 signal can be written:

\[
S_{E6}(t) = -[C^a_{E6}(t)D^a_{E6}(t)\cos(m(t) + C^a_{E6}(t)\sin(m(t))) + \cos(2\pi F_{E6}(t))]
\]

To be consistent with the relative powers required between the three channels, a value of \( m = 0.6155 \) has been chosen for the modulation index.

**Modulation of the E2-L1-E1 Carrier**

In the same way than the E6 signal, the L1 signal contains three channels that are transmitted at the same L1 carrier frequency using a modified Hexaphase modulation. Time multiplexing is also being analyzed.

The E2-L1-E1 signal, using a Hexaphase modulation, can be written:

\[
S_{E2L1E1}(t) = \left[ C^b_{E2L1E1}(t)D^b_{E2L1E1}(t)\cos(m(t) + \sin(m(t)))\right] \cos(2\pi F_{E2}(t))
\]

The same modulation index of \( m = 0.6155 \) is used.

**5 Galileo Spreading Codes**

The pseudo random noise (PRN) code sequences used for the Galileo navigation signals determine important properties of the system. Therefore a careful selection of Galileo code design parameters is necessary. These parameters include the code length and its relation to the data rate and the auto- and cross-correlation properties of the code sequences. The performance of the Galileo codes is also given by the cold start acquisition time.

A first set of reference codes is being retained that offer a good compromise between acquisition time and protection against interference. These codes are based on shift-registered codes, which will be generated on-board.

The reference ranging codes are constructed tiered codes, consisting in a short duration primary code modulated by a long duration secondary code. The resulting code then has an equivalent duration equal to the one of the long duration secondary codes. The primary codes are based on classical gold codes with register length up to 25. The secondary codes are given by predefined sequences of length up to a 100.

Further alternative codes are presently investigated (Pratt, 2002) and flexibility in the on-board implementation is being considered to foresee the generation of other types of codes.

**Code Length**

The code length for Galileo channels carrying a navigation data message shall fit within one symbol in order to have no code ambiguity. The resulting code lengths are shown in Tab. 3.

<table>
<thead>
<tr>
<th>channels</th>
<th>types of data</th>
<th>code sequence duration</th>
<th>primary code length</th>
<th>secondary code length</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5a</td>
<td>OS</td>
<td>20 ms</td>
<td>10230</td>
<td>20</td>
</tr>
<tr>
<td>E5aQ</td>
<td>no data</td>
<td>100 ms</td>
<td>10230</td>
<td>100</td>
</tr>
<tr>
<td>E5b</td>
<td>OS/CS/SoL</td>
<td>4 ms</td>
<td>10230</td>
<td>4</td>
</tr>
<tr>
<td>E5bQ</td>
<td>no data</td>
<td>100 ms</td>
<td>10230</td>
<td>100</td>
</tr>
<tr>
<td>E6a</td>
<td>PRS</td>
<td>TBD</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>E6b</td>
<td>CS</td>
<td>1 ms</td>
<td>8184</td>
<td>-</td>
</tr>
<tr>
<td>E6c</td>
<td>no data</td>
<td>100 ms</td>
<td>10230</td>
<td>50</td>
</tr>
<tr>
<td>L1A</td>
<td>PRS</td>
<td>TBD</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L1B</td>
<td>OS/CS/SoL</td>
<td>4 ms</td>
<td>8184</td>
<td>-</td>
</tr>
<tr>
<td>L1C</td>
<td>OS/CS/SoL</td>
<td>4 ms</td>
<td>8124</td>
<td>25</td>
</tr>
</tbody>
</table>

For the data-less channels, the basic approach is to consider long codes of 20 ms length. Alternate solutions are however being investigated. The first one is to follow a GPS L5 approach consisting of a short code of 1 ms length equally long to the code in quadrature. The second one is to have a much longer code, which could have duration of 0.7 s as in the case of the L2 civil signal. Especially in the case of E5a and E5b it would be useful to determine the data-less code length by analyzing the susceptibility against local interference.

**Auto- and Cross-Correlation Properties**

The cross-correlation properties (interference) are partly determined by the actual code sequences as will be discussed below. Especially for E5a careful code selection is necessary because at this frequency band Galileo and GPS use the same modulation scheme and code rate.
Acquisition Time

Acquisition time is highly dependent on the applied receiver acquisition technique, but generally 30-50 s for cold acquisition time is envisaged for simple receivers on the E5 signals. For the CS on E6 a acquisition time of 30 s is planned if it is considered as a single frequency product. If not, there will be no specific requirement of the E6 acquisition time. Similar consideration applies for the E2-L1-E1 signal. Again it should be stressed that acquisition time performance is highly dependent on affordable receiver complexity.

Encryption

Simple, inexpensive code encryption, which can be removed on request from the ground, is foreseen for the encrypted CS. Code encryption should be realized as a technique controlling the access of code and data without too much constraints and efforts on the user segment. The removal of the encryption should not create a legacy mantle in the user segment and the complexity of the encryption should be a result of a trade-off of market analysis and adequate protection needed for securing those markets.

Service Mapping on Signals

The data carriers will be assigned to provide the following service categories which are summarized in Tab. 4.

<table>
<thead>
<tr>
<th>Id</th>
<th>OS SF</th>
<th>OS DF</th>
<th>OS IA</th>
<th>SoL</th>
<th>CS VA</th>
<th>CS MC</th>
<th>PRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>E5aLQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E5bLQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E6bC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L1bC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The OS signals would use unencrypted ranging codes and unencrypted navigation data messages on the E5 and E2-L1-E1 carriers. A single frequency (SF) receiver uses signals E2-L1-E1B and E2-L1-E1C; and might receive the GPS C/A code signal on L1. A dual frequency (DF) receiver uses additionally signal E5aI and E5aQ and potentially the GPS L5 signal. Improved accuracy (IA) receivers result by using additionally signal E5bI and E5bQ.

The SoL service would use the OS ranging codes and navigation data messages on all E5 and E2-L1-E1 carriers.

The Value Added (VA) CS signals would use the OS ranging codes and navigation data messages on the signal E2-L1-E1B and E2-L1-E1C and additional CS encrypted data messages and ranging codes on the signal E6a and E6c. The Multi Carrier (MC) Differential Application CS could use in addition the OS ranging codes and navigation data messages on the signal E5a and E5b.

The PRS signals would use the encrypted PRS ranging codes and navigation data messages on the E6 and E2-L1-E1 carriers, represented by signals E6a and E2-L1-E1a.

6 Search And Rescue

The SAR distress messages (from distress emitting beacons to SAR operators), will be detected by the Galileo satellites in the 406-406.1 MHz band and then broadcasted to the dedicated receiving ground stations in the 1544-1545 MHz band, called L6 (below the E2 navigation band and reserved for the emergency services). The SAR data, from SAR operators to distress emitting beacons, will be used for alert acknowledgement and coordination of rescue teams and will be embedded in the OS data of the signal transmitted in the E2-L1-E1 carrier frequency.

7 SOME PERFORMANCE PARAMETERS

Overall performance evaluation of Galileo signals is currently investigated. A major difference of Galileo signals to the currently emitted GPS signals is the BOC (resp. AltBOC) modulation scheme and the large bandwidth employed for most of the signals.

An important parameter in this context is the pseudorange code measurement error due to thermal noise. Reference source not found. shows the Cramer-Rao lower bound (Spilker, 1996) for this value of all Galileo signals and the GPS C/A and L5 signal. A receiver DLL bandwidth of 1 Hz is assumed and a value of –205 dBWs is used to convert the minimum received power to a typical carrier to noise density value. The power of the of the processed signals in one frequency and service (i.e. data and pilot channels) are combined.

From Tab. 5 it is evident that BOC signals exhibit low pseudorange code measurement errors because the power...
spectral density is located at the lower and upper boundary of the frequency spectrum and not at the center as it is for BPSK or QPSK signals.

This also implies that the autocorrelation function of BOC signals shows several peaks and dedicated algorithms must be implemented in the receiver to track the correct (central) peak. Tracking of BOC signals is discussed in Betz, 1999 and Pany et al. 2002).

Large signal bandwidths allow the use of a very narrow correlator spacing. Low thermal noise and low code multipath are the resulting benefits. Code multipath envelopes differ significantly if BOC and BPSK signals are compared as shown in Fig. 8 and Fig. 9. For these figures a coherent early minus late code discriminator is used. A common discriminator spacing of $d=1/14$ is chosen to allow for visual comparisons of all signals and to track the central peak of the BOC(14,2) signal. The multipath signal is -3 dB weaker than the direct signal. Note that typical multipath amplitudes are in the range between -7 and -10 dB.

If E5a and E5b are tracked coherently, this results in an extremely low code tracking error due to thermal noise (cf. 3rd line of Error! Reference source not found.) and good multipath mitigation performance. If the E5a and E5b are tracked separately (non-coherently) as QPSK(10) signals and combined after correlation (i.e. averaging of E5a and E5b pseudoranges) the performance gain is much less (cf. 2nd line of Tab. 5).

8 Recent Results Of Interference Studies

The use of the frequency range 960-1215 MHz, containing the lower L-band E5a and E5b, by aeronautical radionavigation services is reserved on a worldwide basis to airborne electronic aids to air navigation and any directly associated ground-based facilities and, on a primary basis, to radionavigation satellite services. This multiple allocation causes interference, which has to be assessed carefully to allow the usage of GPS/Galileo navigation signals for safety critical applications.

Discussion on interference assessment of DME/TACAN, JTIDS/MIDS and radar out of band radiation over L5, E5a and E5b have been conducted since several years. Interference due to these ground-based sources increases with altitude since more interfering signals are received. The sensitive parameter in this context is the acquisition threshold having limited margins to cope with interference of 5.8 dB for GPS L5, 4.8 dB for E5a and 3.3 dB for E5b. Tracking threshold and data demodulation threshold values are a few dB higher. A standard time domain pulse blanking receiver and advanced signal processing is assumed to be used (Hegarty et al., 2000). It should be noted that in contrast to the US, Europe does not plan at present to re-allocate certain DMEs to circumvent this problem.
9 Compatibility/interoperability Of Galileo-GPS

Galileo shall be designed and developed using time, geodesy and signal structure standards interoperable and compatible with civil GPS and its augmentations.

Compatibility is in this context understood as the assurance that Galileo or GPS will not degrade the stand-alone service of the other system. Interoperability is the ability for the combined use of both GNSS to improve upon accuracy, integrity, availability and reliability through the use of a single common receiver design.

Signal-in-Space

The Galileo/GPS interoperability is realized by a partial frequency overlap with different signal structures and/or different code sequences. At E5a (resp. L5) and E2-L1-E1 (resp. L1) Galileo and GPS signals are broadcasted using identical carrier frequencies. At L1 spectral separation of GPS and Galileo signals is given by the different modulation schemes. This allows jamming of civil signals without affecting GPS M-code or the Galileo PRS service.

Using the same center frequencies drastically simplifies receiver frontend design at the cost of mutual interference of both systems. This so-called inter-system interference adds to the interference of navigation signals belonging to the same system, called intra-system interference. Only the sum of both types of interference is relevant for determining the receiver performance.

Interference has been described in (Hein et al., 2001, de Mateo et al., 2002 and Ries et al., 2002a) and a brief overview plus update shall be given in the following. For details we refer to (Godet et al., 2002), where satellite orbital parameters, antenna diagrams, user locations, signal characteristics are described. It can be shown that the $C/N_0$ degradation of GPS C/A code signals due to Galileo BOC(2,2) signals is never above 0.2 dB over the world at any time. For the International Space Station it is 0.22 dB. The maximum $C/N_0$ degradation as a function of geographical coordinates is shown in Fig. 10.

The maximum GPS C/A code intra-system interference computed is below 2.7 dB. This represents the maximum self-interference that GPS C/A codes are currently suffering and explains that GPS C/A real power is about 3 dB above specifications.

The maximum inter-system interference (0.2 dB) cannot occur at the same time nor at the same space than the maximum intra-system interference. Conversely, the maximum intra-system interference is reached when the inter-system interference is minimal.

The maximum total (intra- plus inter-system interference) is shown to be slightly above 2.7 dB, which yields a degradation of current GPS C/A code worst case link budget by only 0.05 dB\(^6\).

It should be noted that C/A degradation due to other Galileo signals is much less than for the BOC(2,2) signal (Hein et al., 2001). Therefore, there is a high confidence that no GPS user will be affected by the Galileo signal overlay on L1.

GPS L5 signal $C/N_0$ degradation due to Galileo E5a as a function of geographical coordinates is shown in Error! Reference source not found.. Galileo signal degradation due to GPS signals has also been investigated and a summary is shown in Tab. 6.

From Tab. 6 it is evident that reciprocal interference levels are very low on L1. They are more significant in E5a/L5. We noted in the last section that DME

---

\(^6\) By modifying the GPS constellation (number of satellites and power), this value can go up to 0.08 dB, cf. (Godet et al., 2002)
interference of E5a and L5 signal leaves only a small margin to civil aviation users at high altitudes, especially over Europe where no DME reallocation is planned. Therefore GPS degradation on Galileo in E5a must be carefully assessed in future work.

<table>
<thead>
<tr>
<th>Tab. 6 Reciprocal level of interference (worst case link budget degradation / inter-system $C/N_0$ degradation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency band</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>L1</td>
</tr>
<tr>
<td>E5a/L5</td>
</tr>
</tbody>
</table>

**Geodetic Coordinate Reference Frame**

For the Galileo coordinate reference system international civilian standards will be adopted. However, for various reasons the realization of the Galileo coordinate and time reference frame should be based on stations and clocks different from those of GPS. These reasons include independence and vulnerability of both systems, allowing one system to act as a backup solution for the other.

The Galileo Terrestrial Reference Frame (GTRF) shall be in practical terms an independent realization of the International Terrestrial Reference System (ITRS) established by the Central Bureau of the International Earth Rotation Service (IERS).

The ITRF is based on a set of station coordinates and velocities derived from observations of VLBI, LLR, SLR, GPS and DORIS. A reduction of the individual coordinates to a common reference epoch considering their station velocity models is performed using fixed plate motion models or estimated velocity fields.

GPS uses WGS84 as coordinate reference frame, practically also a realization of the ITRS, realized by the coordinates of the GPS control stations. The differences between WGS84 and the GTRF are expected to be only a few cm.

This implies for the interoperability of both GNSS systems that the WGS84 and GTRF will be identical within the accuracy of both realizations (i.e. coordinate reference frames are compatible). This accuracy is sufficient for navigation and most other user requirements and the remaining discrepancies in the 2 cm level are only of interest for research in geosciences. Transformation parameters can be provided by a Galileo external Geodetic Reference Service Provider – if needed at all. At the moment it is not foreseen to put such information in the navigation data message.

A coordinate reference frame has to be accomplished by an Earth’s gravity model. For example, the WGS84 uses a spherical harmonic expansion of the gravity potential up to the order and degree 360. For Galileo a similar model must be considered. In that context the European satellite gravity missions GOCE and CHAMP as well as the American mission GRACE are of importance.

**Time Reference Frame**

The Galileo System Time (GST) shall be a continuous coordinate time scale steered towards the International Atomic Time (TAI) with an offset of less then 33 ns. The GST limits, expressed as a time offset relative to TAI, 95% of the time over any yearly time interval, should be 50 ns. The difference between GST and TAI and between GST and UTC(Pred) shall be broadcasted to the users via the signal-in-space of each service.

The offset of the GST with respect to the GPS system time is monitored in the Galileo ground segment and the offset is eventually broadcasted to the user.

The offset might also be estimated in the user receiver with very high accuracy by spending just one satellite observation – the accuracy is (probably) higher than that one (eventually) broadcasted. Thus, broadcasting might be not necessary for the general navigation user.

**Interoperability Summary**

The Galileo system follows international recommendations for steering of its time and coordinate references (UTC and ITRS). This itself enables a possible high level of interoperability in case GPS follows the same, very reasonable, rules.

**Acknowledgements**

The article is based on work of the European Commission Signal Task Force (Hein et al., 2002). The underlying investigations were supported by many European national space agencies like e.g. Centre National d’Etudes Spatiales (France), Deutsches Zentrum für Luft- und Raumfahrt (DLR, Germany) and Defence Science and Technology Laboratory (United Kingdom) Their support and contribution is acknowledged.

**References**


GPS Attitude Determination Reliability Performance Improvement Using Low Cost Receivers

Chaochao Wang and Gérard Lachapelle
Department of Geomatics Engineering, University of Calgary
Received: 13 October 2002 / Accepted: 4 December 2002

Abstract. This paper describes different methods to improve reliability of attitude estimation using low cost GPS receivers. Previous work has shown that low cost receiver attitude determination systems are more susceptible to measurement errors, such as multipath, phase center offsets, and cycle slips. In some cases, these error sources lead to severely erroneous attitude estimates and/or to a lower availability. The reliability control in the attitude determination becomes imperative to users, as most attitude applications require a high level of reliability.

The three methods tested herein to improve reliability are the use of a high data rate, fixed angular constraints, and a quality control algorithm implemented with a Kalman filter. The use of high rate measurements improves error detection as well as ambiguity fixing time. Fixed angular constraints in a multi-antenna attitude system is effective to reject incorrect solutions during the ambiguity resolution phase of the process. Utilizing a Kalman filter with a high data rate, e.g. 10 Hz, not only increases reliability through an increase of information, but also can improve accuracy and availability. The simultaneous utilization of the above methods significantly improves reliability, as demonstrated through a series of hardware simulations and field tests. The low cost receiver type selected is the CMC Allstar receiver equipped with a commercially available low cost antenna.

Finally, the use of statistically reliability measures, namely internal and external reliability measures, shows the inherent limitations of a low cost system and the need to either use better antennas and/or external aiding in the form of low cost sensors.

Key words: GPS, Attitude Determination, Low Cost Receiver

1 Introduction

Multi-antenna GPS systems provide a high accuracy attitude solution without error drift over time [e.g., Lu 1994]. The performance of GPS attitude determination is a function of receiver firmware, satellite geometry, antenna carrier phase stability, multipath rejection ability and inter-antenna distances. With advances in GPS receiver technology, low cost receivers equipped with phase lock loops that output precisely time-synchronized carrier phase measurements are now available on the market. The use of this grade of GPS receiver for attitude determination has proven feasible [e.g., Hoyle et al. 2002]. However, it has been found that multipath, antenna phase center offsets and cycle slips are major error sources that mitigate the performance of low cost receiver attitude solutions. In worst-case scenarios, these errors severely affect the integrated carrier phase measurements and lead to incorrect attitude estimates in attitude determination. Therefore, the reliability of attitude estimation becomes a major issue.

The objective of this paper is to investigate three methodologies to improve the reliability performance of attitude determination using low cost receivers. Three different schemes, namely the use of high rate carrier phase measurements, fixed angular constraints and a Kalman filter with a statistical quality control system, are used interactively to improve reliability. These schemes are implemented in a high performance, open architecture attitude determination software, namely HEADRT™, for testing [e.g., Hoyle et al 2002]. The performances of different methods are examined both in hardware simulation mode and under field static and kinematic conditions.
2 GPS Attitude Determination

By definition, attitude is the rotation of a specific frame with respect to a reference frame, which is well defined in space. In the case of a multi-antenna system, this specific frame is usually referred to as the antenna body frame, while the local level frame is selected as the reference frame. Once the antenna vector in the local level frame is precisely determined, the three Euler attitude angles in the rotation matrix can be estimated using Equation 1.

\[
\begin{pmatrix}
    x^b \\
    y^b \\
    z^b
\end{pmatrix} = R^b_2(r)R^b_1(p)R^b_3(h)\begin{pmatrix}
    x^ll \\
    y^ll \\
    z^ll
\end{pmatrix}
\]

where

- \( h, p, r \) denote heading, pitch and roll
- \( x, y, z \) are the coordinates of the antenna vector
- superscript \( b \) represents the body frame
- superscript \( ll \) stands for the local level frame

The GPS receivers determine the inter-antenna vectors firstly in a Conventional Terrestrial frame, namely WGS-84. The carrier phase measurements have to be used as observables in this application since the attitude determination system requires high precision relative positioning between the antennas. In the general case that independent (non-dedicated) receivers are used and each receiver has a separate oscillator, the double differencing combinations are formed so that not only the clock errors but also the line biases caused by the different cable lengths can be removed. Without clock and line bias errors, the carrier phase double difference observation equation is expressed as

\[
\nabla \Delta \Phi = \nabla \Delta \rho + \lambda \nabla \Delta N + \nabla \Delta d_i + \nabla \Delta d_{\text{ion}} + \nabla \Delta d_{\text{trop}} + \nabla \Delta e_{\text{mult}} + \nabla \Delta e_{\text{ras}} + \nabla \Delta e_{\text{ant}}
\]

Because of the short inter-antenna distance (generally less than 20 m), the spatially correlated orbital and atmospheric errors virtually cancel out from the equation. The errors sources remaining here are only multipath, antenna phase center offsets and carrier receiver noise, provided that the double difference integer ambiguities are correctly solved.

3 Reliability Problems Using Low Cost Receivers

Previous research has shown the advantages and limitations of using low cost receivers such as the CMC Electronic Allstar, for attitude determination [e.g. Hoyle et al 2002]. Without multipath and antenna phase centre errors need to be introduced, thereby allowing a performance analysis of the receiver noise and tracking loops. Under field conditions, the low cost receiver is more likely to suffer from carrier phase multipath and antenna phase instabilities. In practice each of these two error sources range from a few mm to 1 cm (although higher values are possible). In some severe cases, the two error sources, coupled with cycle slips, significantly deteriorate the carrier phase measurements and the wrong double difference ambiguities could be produced from the ambiguity resolution. The incorrect ambiguities eventually lead to the erroneous attitude estimates, which impair the reliability of the whole attitude determination system. In order to improve the overall attitude performance, some measures should be taken to enhance the reliability of attitude determination using low cost receivers.

4 Attitude Determination Algorithm

In the HEADRT+™ software, the attitude determination estimation process is carried out in two phases. The first phase determines the correct double difference carrier phase ambiguities for the antenna vector(s). After the coordinate transformation from WGS-84 into the local level frame, the attitude parameters are estimated from the vector components with corresponding variance-covariance matrix in the second phase [e.g. Lu 1994].

The ambiguity resolution used in the software is based on the Least Squares Ambiguity Search Technique (LSAST) [Hatch 1991]. This method has the advantages of a small number of candidate ambiguity combinations and high computational efficiency. Given that the vector lengths are small, this technique is effective for this purpose. The ambiguity search region is defined as a sphere with the...
radius of the inter-antenna distance(s). After forming all possible ambiguity combinations, different discrimination tests are conducted to isolate the correct ambiguity set using the fixed antenna distance(s) and some other statistical properties [e.g., Hoyle et al. 2002]. Two statistical tests are involved in the ambiguity resolution, namely the ratio test and the Chi-square test. The underlying assumption of a ratio test is that the residuals of the correct ambiguities should be significantly smaller than those of the incorrect ones. Only if the ratio of the two smallest residual quadratic forms is greater than a preset threshold (normally 2.5 to 4), is the potential ambiguity set with the smallest quadratic form accepted as the correct ambiguity set. For true ambiguities, it is assumed that the double difference residuals are normally distributed, and the sum of the quadratic forms follows the Chi-square distribution, with the degree of freedom being the redundant measurement number. Therefore, a Chi-square test based on the residuals is conducted to verify the double difference ambiguities in the software.

Once the inter-antenna vector ambiguities are fixed, the inter-antenna vector components are transformed from WGS-84 earth-fixed frame into local level coordinates and the attitude parameters are computed from an implicit least squares estimation. Currently, no dynamic constraints of the platforms are implemented in the filtering process to permit an epoch-by-epoch assessment of attitude estimation under any dynamics, subject to the availability of unbiased receiver carrier phase).

5 High Data Rate

As the CMC Allstar receiver can output raw time synchronized carrier phase measurements up to 10 Hz, it allows for high data rate processing in HEADRT+™, both for ambiguity resolution and attitude estimation. The higher data rate can benefit the ambiguity resolution process due to the high availability of phase measurements. Also, platform dynamics can be predicted for short time intervals in many applications and outlier estimation in the antenna vector lengths can be easily detected and further rejected using filtering of the high rate measurements. In this section, only the effect of the high data rate on ambiguity resolution will be investigated. The impact of the high data rate on Kalman filter estimation will be discussed in the sequel.

In order to evaluate the performance of ambiguity resolution, the time to fixed ambiguities is utilized. In this test, two receivers were used, both for the hardware simulator and field test. In the latter case, two AT575-104 low cost antennas were used. The hardware simulation test was done using a Spirent STR-4760 simulator. As no errors were simulated, the only remaining error present was receiver measurement noise. The field test was conducted on the roof of Engineering building at the University of Calgary. The inter-antenna distances were about 1 m in both tests. The data was collected at a 10-Hz rate. The double difference ambiguities were intentionally reset every 120 seconds during the data processing to gather enough trials for a meaningful analysis. The Minimum Time To Ambiguity Fix (MTTAF) was set to 1 epoch and the fixing ratio was set to 3 in HEADRT+™.

Fig. 1 shows the ambiguity fixing times for the case of the hardware simulation test. Without multipath and antenna phase center offset, the integer ambiguities were successfully determined within a single epoch (1 s or 0.1 s) during each trial, demonstrating that the CMC receiver measurement noise is not a significant factor that is affecting ambiguity resolution performance.

The corresponding static field test statistics are shown in Fig. 2. With the existence of multipath and antenna phase center errors, 19.6 % of the ambiguities were fixed in one second with 1 Hz data. The integer ambiguities were fixed in 5 seconds 84.9 % of the time. Meanwhile, with 10 Hz measurement rate, the corresponding values were 89.4 % and 93.4 % respectively. The time required to fix the ambiguity could be significantly reduced using high data rate in this case during some trials. However, there are cases where the fixing time was larger than 60 second. This was related to the presence of time-correlated multipath and antenna phase center offset errors. The high data measurement is less effective to these errors.

Tab. 1 shows that the probability of resolving correct ambiguities for the field test is 93% for 1 Hz data and
96% for 10 Hz data. With the higher rate measurements, the ambiguity resolution reliability can thus be only slightly improved. Even though the incorrect ambiguities were selected occasionally, they can be easily rejected in the attitude software either by improving the MTTAF parameter in ambiguity resolution or by the reliability control in the attitude estimation phase.

\[
\theta^E = \cos^{-1}\frac{\tilde{b}_{AB} \cdot \tilde{b}_{AC}}{||\tilde{b}_{AB}|| \cdot ||\tilde{b}_{AC}||} = \cos^{-1}\frac{\Delta E_{AB} \Delta E_{AC} + \Delta N_{AB} \Delta N_{AC} + \Delta V_{AB} \Delta V_{AC}}{b_{AB} \cdot b_{AC}} \tag{3}
\]

The numerical value of the angular tolerance \(\delta\) in (4) depends on the inter-antenna distance and the quality of phase measurements, which are a function of measurement noise, multipath and phase centre stability. In the case of antenna vector lengths of 1-2 m and a moderate carrier phase measurement quality, a 5-degree tolerance is appropriate to detect the wrong ambiguities. If at least four antennas are used in the attitude determination system and only one vector ambiguity is wrong, this erroneous ambiguity combination can be detected and identified by checking all the angles between the inter antenna vectors.

6 Fixed Angular Constraint Scheme

If one can assume that the antennas are mounted on a rigid platform, then their relative positions are fixed regardless of the platform motion. The full antenna frame geometry is known a priori and appropriate constraints can be used in the ambiguity resolution process to take advantage of this knowledge. Many geometric constraints have been brought forward for the ambiguity resolution in multi-antenna GPS attitude determination system. [El-Mowafy 1994, Euler and Hill 1995] In this research, the fixed angle between the antenna vectors, as well as the vector length, was employed to verify the double difference ambiguities.

The implementation of the angular constraint is straightforward. First, the fixed planar angles \(\theta\) between antenna vector pairs could either be measured a priori or calculated using the antenna coordinates in the body frame. Once the integer ambiguities of the antenna vector pairs have been determined, the angle between the pairs can be directly computed using the antenna vector coordinates in the local level frame:

\[
\theta^E = \cos^{-1}\frac{\tilde{b}_{AB} \cdot \tilde{b}_{AC}}{||\tilde{b}_{AB}|| \cdot ||\tilde{b}_{AC}||} = \cos^{-1}\frac{\Delta E_{AB} \Delta E_{AC} + \Delta N_{AB} \Delta N_{AC} + \Delta V_{AB} \Delta V_{AC}}{b_{AB} \cdot b_{AC}} \tag{3}
\]

where

- \(\theta^E\) is the estimated angle between the two antenna vectors
- \(\tilde{b}_{AB}, \tilde{b}_{AC}\) are the antenna vectors in local level frame
- \(b_{AB}, b_{AC}\) are the lengths of the antenna vectors
- \(\Delta E, \Delta N, \Delta V\) are three components of antenna vector in east, north, and vertical directions
- subscripts \(A, B, C\) represent the primary antenna and two secondary antennas

Then, the estimated angle \(\theta^E\) is compared with the known angle \(\theta\). If the ambiguities of two inter-antenna vectors are correctly solved, the two angles should be consistent within a certain tolerance.

\[
|\theta - \theta^E| < \delta \tag{4}
\]

A hardware simulation test was conducted with the 4760 simulator to investigate the validity of the angular constraint scheme. An antenna body frame was simulated using inter-antenna distances of 1 m. The angles between the antenna vectors were intentionally set to 90 degrees in this test. The initial parameters used in the software were

- Fixing ratio = 3
- MTTAF = 1 epoch

Fig. 3 shows the satellite’s azimuth and elevation DOPs during the test. At GPS time 216932 s, the loss of SV27 signal in one of the secondary receivers caused the failure of the Chi-square test and the re-initialization of the double difference ambiguity for the corresponding inter-antenna vector. Unfortunately, the wrong ambiguity was determined due to the short MTTAF. When SV27 was re-acquired by the receiver, an incorrect ambiguity was first
found, with the true ambiguity obtained afterwards. The effect of this error on the inter-antenna vector solutions during this period is shown in Fig. 4. The inter-antenna length components are obviously incorrect. However, the length itself was corrected solved and testing of the solution with that known length failed to detect the incorrect solution in this case.

Since no quality control procedure was performed in the least squares attitude estimation, the erroneous inter-antenna vector solutions inevitably led to the wrong attitude parameters. The error effects on the attitude component estimates are shown in Fig. 5.

After the angular constraint scheme was implemented in the software, the wrong ambiguity was easily detected and the erroneous vector solution was successfully detected and excluded from the attitude estimation. As shown in Fig. 6, the correct attitude components were estimated in the least squares solution using the other two inter-antenna vectors. The small shift in the attitude estimates is due to the exclusion of SV27 and the resulting slight change of satellite geometry. The mean and rms agreements in heading, pitch and roll are 2.1, 0.3, -0.2, 2.8, 3.4 and 3.3 arcmins, respectively.

By employing the angle consistency check in the ambiguity resolution, some incorrect ambiguity solutions can be effectively rejected, which significantly improves the reliability of multi-antenna attitude determination.

7 Kalman Filter Estimation

Kalman filtering estimation provides a recursive method for the determination of attitude components through a predicting and updating process. The general formulas in Kalman filtering can be written as [Brown & Hwang 1992]

\[
    z_k = H_k \cdot x_k + v_k \\
    x_k = \phi_k \cdot x_{k-1} + w_k
\]

where

- \( z_k \) is the measurement vector at time \( k \)
- \( H_k \) is the design matrix
- \( x_k \) is state vector at time \( k \)
- \( v_k \) is the measurement noise with covariance \( R \)
- \( \phi_k \) is the transition matrix
- \( w_k \) is the process noise with covariance \( Q \)

In attitude determination using vector components, the “measurements” are the antenna vector components in the local level frame. The design matrix is the partial derivative of the rotation matrix with respect to the state vector in Equation 1.
The state vector here includes the three Euler attitude parameters and their angular rates:

$$x = \begin{pmatrix} \psi & \theta & \phi & \dot{\psi} & \dot{\theta} & \dot{\phi} \end{pmatrix}^T$$

(8)

The transition matrix $\Phi$ and the process noise can be expressed as follows

$$\begin{bmatrix}
1 & 0 & 0 & dt & 0 & 0 \\
0 & 1 & 0 & 0 & dt & 0 \\
0 & 0 & 1 & 0 & 0 & dt \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

(9)

$$Q = \begin{bmatrix}
0 & 0 & 0 & \sigma^2 \dot{\psi} & 0 & 0 \\
0 & 0 & 0 & 0 & \sigma^2 \dot{\theta} & 0 \\
0 & 0 & 0 & \sigma^2 \dot{\phi} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

(10)

The numerical values of the angular rate variances in (10) represent the tightness of the dynamic constraint of the Kalman filter. In vehicular attitude determination, the sigma of the angular rate in the $Q$ matrix is empirically selected as 2 degrees per epoch in 10 Hz sampling in the present case. Intuitively, one realizes that the effectiveness of the filter in detecting incorrect solutions, thus improving reliability, will depend on our a priori knowledge of the vehicle dynamic and of the measurement rate.

Using this model, the attitude parameters and their angular rates can be correctly estimated in the Kalman filter as long as all the measurements are free of errors.

As previously mentioned, the measurements used in the Kalman filter are the inter-antenna vector solutions after ambiguity resolution. In the case that the wrong ambiguity is determined, these “quasi-measurements” are in error and the attitude estimates calculated from the Kalman filter may deviate from the truth. In order to reject the incorrect inter-antenna vector solutions from the Kalman filter and improve the reliability of the attitude estimates, a quality control system based on the filter innovation sequences is introduced herein.

The innovation sequence is the difference between the actual system output and the predicted output based on the predicted state (see Equation 11). [Teunissen & Salzman 1988].

$$V_k = z_k - f(\hat{x}_k)$$

(11)

Under normal conditions, the innovation sequence is a zero-mean Gaussian white noise sequence with known variance. In the presence of erroneous measurements, such assumptions are no longer valid, and the innovation sequence deviates from its zero mean and white noise properties. Thus some statistical tests can be conducted to detect and identify outliers or faults in the measurements.

Firstly, an overall model test is conducted to detect the errors in the measurement vector. The test statistics in this global test are given as

$$T_k = v_k^T C^{-1}_k v_k \sim \chi^2(m,0)$$

(12)

where

- $m$ is the number of observations taken at time $k$.
- $C_{v_i}$ is the covariance matrix of the innovation and
- $\chi^2$ is the Chi-square probability with a significance level of $\alpha$.

If the global test is rejected, the system error can be identified with the one-dimensional local slippage test.

$$w_i = \frac{i - 0.5}{\sqrt{C_{v_i}^{-1}}} \sim N(0,1)$$

(13)

where

$$l_i = (0,...,0,1,0,...,0)^T$$

for $i = 1,...,m$

The significant level $\alpha$ in the local test is suggested to be 0.999, which leads to a boundary value of 3.29. Thus the i-th measurement is flagged for rejection when

$$|w_i| > 3.29$$

(14)

When implementing statistical tests to identify outliers in the measurements, two types of errors may be made, as shown in Fig. 7. The first type (Type I) is rejecting a good measurement. The probability associated with this type error is denoted by $\alpha$. If a bad measurement is accepted by the test, a Type II error occurs. The probability of a Type II error is expressed as $\beta$.

![Fig. 7 Type I/II Errors](image)

Given the probability values of Type I and Type II errors, the Minimum Detectable Blunder (MDB) can be calculated as the ability to detect errors in the system as
\[ |\nabla z_i| = \frac{\delta_0}{\sqrt{I_i^T C^{-1} I_i}} \]  

(15)

where \( \delta_0 \) is a function of \( \alpha \) and \( \beta \) (see Fig. 8).

In GPS kinematic applications, \( \alpha \) and \( \beta \) are commonly selected to be 0.001 and 0.2 respectively and \( \delta_0 \) is then 4.13.

In the presence of strong multipath, the identification test (Equation 15) may be too sensitive and will sometimes lead to a false alarm. In order to alleviate this problem, a further step was introduced by comparing the innovations with the MDB. If the innovation is larger than the MDB, the measurement is identified erroneous, otherwise it is considered a false alarm.

The modified Kalman-filter-based attitude determination software was tested with the data collected with the Spirent 4760 hardware simulator using four CMC receivers. A vehicle trajectory was simulated in this test and the antenna configuration is shown in Figure 8. The maximum attitude changes were about 20 degree/s in heading and several degrees per second in pitch. The true attitude during the test is plotted in Fig. 9.

The results, summarized in Fig. 10 and Table 2, show that the Kalman filter method did not work well with tight dynamic constraints using a 1-Hz data rate, as overshooting effects occur. With a 10-Hz data rate, the performance of the filter is excellent, the attitude parameter estimates being slightly better than those of the least squares estimates.

![Fig. 8 Simulated vehicle test antenna configuration](image)

![Fig. 9 True attitude parameters during hardware simulator test](image)

![Fig. 10 Attitude estimate errors using different estimation methods](image)

![Tab. 2 RMS – Kalman filter versus least-squares](image)

<table>
<thead>
<tr>
<th></th>
<th>Heading</th>
<th>Pitch</th>
<th>Roll</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Hz LS</td>
<td>3.9’</td>
<td>11.8’</td>
<td>9.9’</td>
</tr>
<tr>
<td>10 Hz KF</td>
<td>3.9</td>
<td>9.8</td>
<td>7.9</td>
</tr>
<tr>
<td>1 Hz KF1</td>
<td>31.9</td>
<td>11.8</td>
<td>9.0</td>
</tr>
</tbody>
</table>

In order to test the performance of cycle slip detection using the quality control method implemented by the Kalman filter, 80 cycle slips were introduced in the carrier phase measurements on different receivers with a magnitude ranging from 1 to 8 cycles. Using the traditional phase prediction detection and inter-antenna length consistency check, all the cycle slips but one were either detected or recovered. The remaining cycle slip was removed when the Kalman filter was used, as shown in Fig. 12.
8 Field Test And Result Analysis

A kinematic field test was carried out using two grades of GPS receiver. The high grade system consisted of two NovAtel Beeline receivers and four NovAtel 501 antennas, while the low grade system consisted of four CMC Electronic Allstar receivers and four AT575-70 antennas.

The NovAtel Beeline receiver is a high performance bi-antenna receiver for 2-D attitude determination. The NovAtel 501 antenna has very good antenna phase center stability. The AT575-70 active antenna is a small size low-cost (5 cm in diameter) OEM antenna often used with CMC Allstar receivers. Two antenna frames were mounted with similar geometry on the roof of a minivan, to create the mobile platform in this test, as shown in Figure 12. The antenna configuration used here was the same as in the above simulation test (Fig. 9). The raw GPS measurements from two attitude systems were logged with laptops in a 10-Hz rate.

The azimuth and elevation DOP and the number of satellites tracked are shown in Fig. 13. During the test, the number of satellite tracked was mostly around six to seven, except in some cases where there was heavy foliage near the road, and the satellite numbers dropped to five or less.

Fig. 14 show the residuals of double difference pairs at every epoch in the inter-antenna vector solutions. These residuals represent the overall effect of measurement errors, including multipath and antenna phase center errors, assuming that the double difference ambiguities are correctly solved. The average RMS values are given in Tab. 3. The CMC units have much larger double difference residuals since their carrier phase measurements are more affected by multipath and antenna phase center errors than those of the Beeline units. Note that the CMC antenna phase centre errors are not calibrated. It is not realistic to do so for such low cost antennas that are likely to include unit-to-unit variations.

The three Euler attitude parameter estimates using the Beeline units are shown in Fig. 15. The blue dots are the
least squares attitude estimates and their 3-sigma standard deviation envelopes, while the red dots are the corresponding estimates from the Kalman filter with the quality control method turned on.

<table>
<thead>
<tr>
<th>Inter-antenna Vector</th>
<th>Beeline Rcvrs</th>
<th>CMC Rcvrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>

Tab. 3 Residuals RMS (mm) for Beeline and CMC receivers

Using least squares estimation, wrong attitude parameter estimates were output when heavy satellite blockages occurred. The reason for this is that the least squares estimation was severely affected by incorrect vector solutions in such circumstances. Once the base satellite is lost, the double difference ambiguities have to be resolved at the next epoch. As is known, ambiguity resolution performance is highly correlated to the number of visible satellites and their geometry. In a heavy signal blockage area with strong multipath and phase center variations, ambiguity resolution is more likely to result in an incorrect solution, which leads to erroneous attitude parameter estimates. These incorrect estimates can however be easily identified by inspecting the 3-sigma standard deviation envelopes.

Once a Kalman filtering with the quality control method and the angular constraints are implemented, the wrong inter-antenna vector solutions are detected and excluded from the solution. This eliminates erroneous attitude parameters from the output. The Kalman filter 3-sigma standard deviation envelopes are slightly smaller than those from the least squares method due to the filter constraints. As can be seen in the figures, the standard deviation improvement is more significant in pitch and roll than in heading. This is because the pitch and roll dynamics were lower than those of heading.
slips, erroneous inter-antenna vector solutions were more frequently determined. When more incorrect solutions were rejected by the Kalman filter, the availability of attitude estimates degraded due to the reduction of correct “quasi-observables” compared with the result from the least squares method. The lowered number of vector solutions involved in attitude estimation, coupled with the larger carrier phase errors, caused large variations in the estimation accuracy of the Kalman filter.

The estimated attitude differences between the Beeline and CMC systems are shown in Fig. 17, and the corresponding statistics are summarized in Table 4. The estimated differences are mostly within 1.5 degrees in heading and 3 degrees in pitch and roll. The largest differences occur during periods of poor satellite geometry.

![Fig. 17 Attitude differences between the Beeline and CMC systems](image1)

<table>
<thead>
<tr>
<th>Difference</th>
<th>Heading</th>
<th>Pitch</th>
<th>Roll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.68</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RMS</td>
<td>0.94</td>
<td>2.26</td>
<td>2.17</td>
</tr>
<tr>
<td>Max(abs)</td>
<td>4.56</td>
<td>8.64</td>
<td>9.44</td>
</tr>
</tbody>
</table>

Figure 18 shows the external statistical reliability of the two systems. External reliability is the impact of the maximum measurement errors that could occur and go undetected, on the attitude estimates, for two systems. This reliability measure is a function of the quality of carrier phase measurements and of the redundancy numbers in the Kalman filters. The external reliability of the Beeline system is fairly consistent during the test except during times of poor geometry. The corresponding reliability of the CMC system is much poorer due to the higher multipath and antenna phase center errors. It is important to note that the estimated attitude differences in Figure 23 are within the reliability numbers of Figure 24 and 25. Thus, one can conclude that the CMC units have reached their limit in term of accuracy performance, if one assumes that the choice of antennas is limited to current low cost units. In order to increase attitude component estimation performance, higher performance, but more expensive antennas could be used. The use of long inter-antenna distances would also improve accuracy. Aiding with external sensors is the other alternative.

![Fig. 18 External reliability](image2)

### 9 Conclusions

Multipath, cycle slips and antenna phase center instability are major error sources limiting the reliability of standalone GPS-based attitude determination with low cost receivers. Even if a required level of accuracy can be achieved with a given multiple receiver configuration, reliability becomes a major issue. It has been demonstrated herein that the use of angular constraints and a Kalman filter with a high data rate are effective to significantly improve reliability. However, the use of statistical reliability analysis has also shown the limitations of the above techniques.

Another technique is currently being assessed to improve reliability and error detection, namely the use of low cost rate gyro integrated with the antenna assembly in various configurations. Given a GPS data rate of 10 Hz, such low cost rate gyro should still be useful for short term prediction between the GPS measurements, smoothing, error detection and enhancement of availability. Early results indicate these are possible enhancements indeed.
References


Joint Treatment of Random Variability and Imprecision in GPS Data Analysis

Hansjörg Kutterer
DGFI, Marstallplatz 8, D-80539 Munich, Germany
e-mail: kutterer@dgfi.badw.de; Tel: +49(89)23031-214; Fax: +49(89)23031-240
Received: 27 November 2002 / Accepted: 17 December 2002

Abstract. In the geodetic applications of the Global Positioning System (GPS) various types of data uncertainty are relevant. The most prominent ones are random variability (stochasticity) and imprecision. Stochasticity is caused by uncontrollable effects during the observation process. Imprecision is due to remaining systematic deviations between data and model due to imperfect knowledge or just for practical reasons. Depending on the particular application either stochasticity or imprecision may dominate the uncertainty budget. For the joint treatment of stochasticity and imprecision two main problems have to be solved. First, the imprecision of the original data has to be modelled in an adequate way. Then this imprecision has to be transferred to the quantities of interest. Fuzzy data analysis offers a proper mathematical theory to handle both problems. The main outcome is confidence regions for estimated parameters which are superposed by the effects of data imprecision. In the paper two applications are considered in a general way: the resolution of the phase ambiguity parameters and the estimation of point positions. The paper concludes with numerical examples for ambiguity resolution.

Key words: Fuzzy data analysis, imprecision, fuzzy confidence regions, GPS, ambiguity resolution

1 Introduction

Today, the Global Positioning System (GPS) is intensively used in geodetic applications as it is efficient and easy to access. The GPS consists of nominally 24 satellites on six orbital planes. It supplies the broadcast transmission of one-way microwave signals on two frequencies from the satellites to the individual ground stations. The 3D position of the GPS ground antenna and the receiver clock offset can be determined by simultaneously observing the signals of at least four satellites. This yields the satellite-receiver distances either directly using the code observations or indirectly via the (carrier wave) phase observations. For the second type of observations, the ambiguity parameters have to be determined. For further reading on GPS and on ambiguity resolution techniques see, e.g., Hofmann-Wellenhof et al. (1997), Parkinson and Spilker (1996), Teunissen and Kleusberg (1998).

GPS observations are biased by a variety of physical effects which have to be considered and handled in data processing. There are mainly three groups of causes. The most important one is due to the propagation of the signals. As the path of the GPS signals leads through the complete atmosphere, ionospherically and tropospherically caused travel-time delays have to be taken into account. They are superposed by multipath effects due to signal reflections in the vicinity of the tracking GPS antenna. The second group comprises all satellite effects like, e.g., signal transmission delays, satellite clock errors, satellite orbit errors, and satellite antenna offsets. Station and receiver effects like, e.g., signal reception delays and receiver clock errors belong to the third group. In addition, the GPS data processing results show characteristics due to the software and the operator.

Several techniques can be applied to reduce or eliminate most of the systematic effects such as the use of correction models with fixed or free parameters or of linear combinations of the GPS observations such as double differencing. Longer-term periodic signals such as diurnal ones can be weakened if the observation time is sufficiently long. However, such effects can not be eliminated completely due to the imperfect knowledge.
and the approximate character of the models in use, respectively. Hence, the uncertainty due to remaining systematic effects (imprecision) must be taken into account in addition to the random variability (stochasticity) of the observations. Kutterer (2001a, 2002) gives a general discussion of uncertainty in geodetic data analysis. Imprecision is particularly relevant in case of long distances between the GPS sites or very short observation intervals as in neither case it is possible to completely describe and remove systematic effects. This paper can be seen as an extension and generalization of the results given by Kutterer (2001b).

Fuzzy data analysis (Bandemer and Näther, 1992; Viertl, 1996) has proven to be an adequate mathematical tool to handle imprecision. Moreover, the combination of methods from stochastic and fuzzy theory allows the extension of classical geodetic data analysis to account for the effects due to superposed imprecision. In the following, the basics of fuzzy data analysis are presented. Two alternatives for the definition of fuzzy vectors are discussed. If the classical formulas of statistics are fuzzified by means of Zadeh’s extension principle, stochasticity and imprecision can be treated simultaneously. Thus, imprecise confidence regions for the ambiguity parameters and for the point positions can be defined and discussed. At the end of the paper the results of simulation studies are given. They illustrate the applicability of the theory and quantify the impact of imprecision.

2 Basics of fuzzy data analysis

Fuzzy-theory was initiated by Zadeh (1965) in order to extend classical set theory by describing the degree (of membership) that a certain element belongs to a set. In classical set theory the membership degrees are either 1 (is element) or 0 (is not element). In fuzzy set theory the range of membership degree is [0,1]. Thus, a fuzzy set is defined as

\[
A = \{(x, m_A(x)) | x \in X\}, m_A : X \rightarrow [0,1].
\]  

The degree of membership is given by the membership function which is denoted by \(m_A(x)\). X is a classical set such as the set \(R\) of the real numbers. Important notions are the support of a fuzzy set (classical set with positive degrees of membership), the height of a fuzzy set (maximum membership degree), the core of a fuzzy set (the classical set with membership degree equal to 1), and the \(\alpha\)-cut of a fuzzy set (classical set with membership degree greater equal \(\alpha\) \(\in [0,1]\)). For further reading see standard references on fuzzy data analysis such as Bandemer and Näther (1992) or Viertl (1996).

The most important operation in fuzzy-theory is the intersection of fuzzy sets. It is defined through the resulting membership function

\[
m_{A \cap B} = \min(m_A, m_B)
\]

This definition is mostly used. Other consistent extensions of the classical intersection operator are available. See, e.g., Dubois and Prade (1980).

Fuzzy numbers can be defined based on fuzzy sets. A fuzzy number is a fuzzy set with a single element core and compact \(\alpha\)-cuts. The L-fuzzy numbers defined by Dubois and Prade (1980) are widely used. They are exclusively considered in this paper. Their membership function is given by a strictly decreasing non-negative reference function \(L\) with [0,1] as the range of values.

\[
m_X(x) = \begin{cases} 
L\left(\frac{x_m - x}{x_s}\right), & x_1 \leq x < x_m \\
L\left(\frac{x - x_m}{x_s}\right), & x_u \geq x \geq x_m \\
0, & \text{else}
\end{cases}
\]

Due to the single element core, \(L(0) = 1\). For a graphical sketch of a L-fuzzy number with a linear reference function see Fig. 1. Formally, it can be represented by \(\tilde{X} = (x_m, x_s)_L\). The mean point is denoted by \(x_m\). The spread \(x_s\) serves as a scale factor. In practice, a typical membership function vanishes outside the interval given by the lower bound \(x_s\) and the upper bound \(x_u\).

![Fig. 1 L-fuzzy number with linear membership function (triangular fuzzy number)](image)
allows the generalization of functions with real arguments to functions with fuzzy arguments. For L-fuzzy numbers,

\[
\mathbf{B} = \mathfrak{g}(\mathbf{A}_1, \ldots, \mathbf{A}_n) \iff \\
m_\mathbf{B}(y) = \sup_{(x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n} \min \{ m_{\mathbf{A}_1}(x_1), \ldots, m_{\mathbf{A}_n}(x_n) \} \quad \forall \ y \in Y
\]

(4)

The extended arithmetic rules are, e.g.,

\[
\begin{align*}
\mathbf{X} + \mathbf{Y} &= (x_m + y_m, x_s + y_s)_L & \text{Addition} \\
\mathbf{X} - \mathbf{Y} &= (x_m - y_m, x_s + y_s)_L & \text{Subtraction} \\
a\mathbf{X} &= (a x_m, a x_s)_L & \text{Multiplication by a real number}
\end{align*}
\]

(5a, b, c)

The type of the reference function is preserved. The arithmetic operations can be carried out simply based on the mean points and the spreads. Please note that subtraction is not the inverse of addition. In fuzzy data analysis the spreads are regarded as measures of fuzziness or imprecision, respectively. Obviously, they are just added (linear propagation) in contrast to the addition of variances (quadratic propagation of the standard deviations) according to the Gaussian law of error propagation.

There are several possibilities to combine fuzzy numbers to a vector; see, e.g. Viertl (1996), Kutterer (2002). The mostly used way is to build a fuzzy vector by the minimum rule, i.e. using the minimum operator according to Eq. (2). In the 2D case this reads as

\[
m_\mathbf{Z}(z) = \min(m_\mathbf{X}(x), m_\mathbf{Y}(y))
\]

(6)

For a graphical representation (linear reference function L as in Fig. 1) see Fig. 2. Such fuzzy vectors are called non-interactive (independent components).

A linear mapping of a fuzzy vector by the minimum rule can be approximated by the tightest inclusion

\[
m_{\mathbf{FZ}}(z) = (\mathbf{F} z_m, \mathbf{F} | z_s)_L
\]

(7)

The function \( h \) is monotonously decreasing and non-negative with \( h(0)=1 \). The spreads and the interaction of the components are quantified in the positive definite uncertainty matrix \( \mathbf{U} \). Interaction is principally present due to the quadratic form which is the argument of \( h \). Fuzzy vectors according to Eq. (8) are called fuzzy vectors of elliptic type. See Fig. 3 for a graphical representation; the function \( h \) of non-negative real arguments \( p \) and the matrix \( \mathbf{U} \) are chosen as

\[
h(p) = \max \left(1 - p^{1/2}, 0\right), \quad \mathbf{U} = \begin{bmatrix} x_s^2 & 0 \\ 0 & y_s^2 \end{bmatrix}
\]

(8)

Linear mappings of fuzzy vectors of elliptic type are given in closed form by

\[
m_{\mathbf{Y,FZ}}(y) = h \left( (y - y_m)^T \left( \mathbf{F} \mathbf{U} \mathbf{F}^T \right)^{-1} (y - y_m) \right)
\]

(9)

The simplicity and closeness of Eqs. (8) and (9) is in contrast to the problems of motivating and formulating interactive (i.e. fuzzy-theoretically dependent) components. The equivalence with the Gaussian error propagation (variance propagation law) is obvious. But it has to be kept in mind that the interpretation is different since the membership functions must not be confused with the independently defined density functions of probability theory. Nevertheless, a quadratic propagation
of spreads is available by means of fuzzy vectors of elliptic type; see Eq. (9).

3 Modeling and propagation of data imprecision

Fuzziness and imprecision are considered as being identical in the following as it is common practice in fuzzy data analysis. Hence, fuzzy data analysis can be applied to handle the impact of observation imprecision on the parameters of interest. As already motivated in Section 1, there are several sources of imprecision in GPS data acquisition and analysis. Hence, both stochasticity and imprecision have to be considered in a general combined approach. Stochasticity is assumed to be superposed by imprecision. This is the basic condition of the extension principle according to Eq. (4).

There are three steps to derive the imprecision of the quantities of interest. First, the imprecision of a single observation has to be described by means of a fuzzy (or imprecise) number. This can be based on a questionnaire to be completed by experts in order to assess the particular application; see, e.g., Kutterer (2002) for details. Second, the fuzzy numbers representing the imprecise observations have to be combined to a fuzzy (or imprecise) vector. This can be based on the two types of fuzzy vectors given in Section 2; see Eq. (6) for the definition of a fuzzy vector by the minimum rule and Eq. (8) for a fuzzy vector of elliptic type. Third, the extension principle according to Eq. (4) has to be applied to the real-valued functional expressions. Here, the least-squares estimator (LSE)

$$\hat{\beta} = \left( X^T W X \right)^{-1} X^T W y$$

of the (deterministic) parameters $\beta$ in a Gauss-Markoff model is considered first. Its variance-covariance matrix (vcm) reads as

$$\Sigma_{\hat{\beta} \hat{\beta}} = \sigma_0^2 \left( X^T W X \right)^{-1}$$

The column-regular \([n \times u]\)-dimensional configuration matrix is denoted by $X$ and the \([n \times n]\)-dimensional regular weight matrix of the observations by $W$. The vector of the observations is represented by $y$. The a priori variance factor is given by $\sigma_0^2$.

The second quantity of interest is the \((1-\gamma)\)-confidence region for the expected value $\mu$ of $\hat{\beta}$ which is given by

$$K_{1-\gamma}(\hat{\beta}) = \left\{ \mu \left| \left( \mu - \hat{\beta} \right)^T \Sigma_{\hat{\beta} \hat{\beta}}^{-1} \left( \mu - \hat{\beta} \right) \leq \chi^2_{n,1-\gamma} \right. \right\}$$

with $\chi^2_{n,1-\gamma}$ the \((1-\gamma)\)-fractile value of the $\chi^2$-distribution with $u$ degrees of freedom.

3.1 Extended least-squares estimator

In the first case (LSE according to Eq. (10)) the extension principle reads as

$$m_{\beta}(\hat{\beta}) = \sup_{\beta \in \mathbb{R}^u} m_{\beta}(y) \quad \forall \hat{\beta} \in \mathbb{R}^u$$

with $m_{\beta}(y)$ the membership function of the vector of the $n$ imprecise observations and $m_{\beta}(\hat{\beta})$ the membership function of the vector of the $u$ imprecise estimated parameters. The use of fuzzy vectors by the minimum rule yields

$$m_{\beta}(\hat{\beta}) = h(\hat{\beta})$$

according to Eq. (7). The use of fuzzy vectors of elliptic type yields

$$m_{\beta}(\hat{\beta}) = h\left( \left( \hat{\beta} - \hat{\beta}_m \right)^T \left( X^T W X \right)^{-1} W X \left( X^T W X \right)^{-1} \left( \hat{\beta} - \hat{\beta}_m \right) \right)$$

according to Eq. (9) with

$$\hat{\beta}_m = \left( X^T W X \right)^{-1} X^T W y_m$$

This can be rewritten as

$$\Sigma_{\hat{\beta} \hat{\beta}} = \sigma_0^2 \left( X^T W X \right)^{-1}$$

The column-regular \([n \times u]\)-dimensional configuration matrix is denoted by $X$ and the \([n \times n]\)-dimensional regular weight matrix of the observations by $W$. The vector of the observations is represented by $y$. The a priori variance factor is given by $\sigma_0^2$.

The second quantity of interest is the \((1-\gamma)\)-confidence region for the expected value $\mu$ of $\hat{\beta}$ which is given by

$$K_{1-\gamma}(\hat{\beta}) = \left\{ \mu \left| \left( \mu - \hat{\beta} \right)^T \Sigma_{\hat{\beta} \hat{\beta}}^{-1} \left( \mu - \hat{\beta} \right) \leq \chi^2_{n,1-\gamma} \right. \right\}$$

with $\chi^2_{n,1-\gamma}$ the \((1-\gamma)\)-fractile value of the $\chi^2$-distribution with $u$ degrees of freedom.

3.1 Extended least-squares estimator

In the first case (LSE according to Eq. (10)) the extension principle reads as
respectively; see, e.g., Kutterer (2002). Please note that for both the fuzzy vectors by the minimum rule and the fuzzy vectors of elliptic type the mean point vectors $\beta_m$ are identical with the classical (precise) least-squares estimators. Thus, both presented fuzzy extensions are consistent with the real-valued case.

### 3.2 Extended confidence regions

In case of the confidence regions, see Eq. (12), the extension principle reads as

$$m_{\kappa, \gamma}(\beta) = \sup_{\mu \in K_{\kappa, \gamma}} m_{\kappa}(\mu) \quad \forall \beta \in \mathbb{R}^n \quad (16)$$

Eq. (16) represents a constrained optimization problem. The imprecise confidence region is the solution of this problem. In order to obtain a closed-form expression for the results, fuzzy vectors of elliptic type

$$m_\beta(\mu) = h\left(\left(\mu - \hat{\beta}_m\right)^T U_{\beta \beta}^{-1} (\mu - \hat{\beta}_m)\right) \quad (17)$$

and with the Lagrangian multiplier $\lambda$.

Two special cases can be distinguished: Obviously, as long as $\beta \in K_{\kappa, \gamma}(\hat{\beta}_m)$, there is always a $\mu \in K_{\kappa, \gamma}(\hat{\beta})$ with $\mu = \hat{\beta}_m$ and hence $d^2_{U_{\beta \beta}}(\mu, \hat{\beta}_m) = 0$.

Consequently, the resulting value of the membership function is equal to one. Hence, the obtained classical set corresponds to the confidence region given by Eq. (12).

In all other cases, i.e., $\beta \notin K_{\kappa, \gamma}(\hat{\beta}_m)$, there is $\kappa^2 = \chi^2_{\alpha, \beta} = constant$. Then the problem can be

$$\Phi(\mu) = \left(\mu - \hat{\beta}_m\right)^T U_{\beta \beta}^{-1} (\mu - \hat{\beta}_m) - \lambda \left(\left(\mu - \hat{\beta}\right)^T \Sigma_{\beta \beta}^{-1} (\mu - \hat{\beta}) - \kappa^2\right), \text{ with } \kappa^2 \in \left[0, \chi^2_{\alpha, \beta}\right] \quad (18)$$

understood in a geometrical way as the determination of the distance between the point $\hat{\beta}_m$ and the hyperellipsoid $K_{\kappa, \gamma}(\hat{\beta})$. As now the distance is in any case positive, the corresponding values of the membership function are less than one. Hence, the confidence region according to Eq. (12) is the core (see Section 2) of the extended confidence region.

The objective function given in Eq. (18) now reads as

$$\frac{\partial \Phi(\mu)}{\partial \mu} = 2\left(\mu - \hat{\beta}_m\right)^T U_{\beta \beta}^{-1} \left(\mu - \hat{\beta}_m\right) - 2 \lambda \left(\mu - \hat{\beta}\right)^T \Sigma_{\beta \beta}^{-1} (\mu - \hat{\beta}) = 0$$

and

$$d^2_{U_{\beta \beta}}(\mu, \hat{\beta}_m) = \left(\mu - \hat{\beta}_m\right)^T U_{\beta \beta}^{-1} (\mu - \hat{\beta}_m)$$

under the side condition

$$\mu \in K_{\kappa, \gamma}(\hat{\beta}) = \left\{ \mu \left| \left(\mu - \hat{\beta}\right)^T \Sigma_{\beta \beta}^{-1} (\mu - \hat{\beta}) \leq \kappa^2 \right\} \right.$$  

An equivalent side condition is

$$(\mu - \hat{\beta})^T \Sigma_{\beta \beta}^{-1} (\mu - \hat{\beta}) = \kappa^2, \text{ with } \kappa^2 \in \left[0, \chi^2_{\alpha, \beta}\right]$$

Thus, the objective function to be minimized with respect to (w.r.t.) $\mu$ reads as

$$\Phi(\mu) = \left(\mu - \hat{\beta}_m\right)^T U_{\beta \beta}^{-1} (\mu - \hat{\beta}_m) - \lambda \left(\left(\mu - \hat{\beta}\right)^T \Sigma_{\beta \beta}^{-1} (\mu - \hat{\beta}) - \kappa^2\right), \text{ with } \kappa^2 \in \left[0, \chi^2_{\alpha, \beta}\right] \quad (19)$$

The determination of the stationary point which refers to the minimum requires the differentiation of $\Phi$ w.r.t. $\mu$ which yields

$$\mu - \hat{\beta} = \left(U_{\beta \beta}^{-1} - \lambda \Sigma_{\beta \beta}^{-1}\right) \left(U_{\beta \beta}^{-1} - \lambda \Sigma_{\beta \beta}^{-1}\right)^{-1} U_{\beta \beta}^{-1} \left(\hat{\beta}_m - \hat{\beta}\right) \quad (20)$$

Hence,

$$d^2_{U_{\beta \beta}}(\mu, \hat{\beta}_m) = \left(\mu - \hat{\beta}_m\right)^T U_{\beta \beta}^{-1} (\mu - \hat{\beta}_m)$$

Differentiation of $\Phi$ w.r.t. $\lambda$ yields

$$\left(\mu - \hat{\beta}\right)^T \Sigma_{\beta \beta}^{-1} (\mu - \hat{\beta}) - \kappa^2 = 0.$$
Insertion of Eq. (21) into the last one finally yields the single equation

$$\left(\hat{\beta}_m - \hat{\beta}\right)^T \left(\Sigma_{\beta\beta} - \lambda \ U_{\beta\beta}\right)^{-1} \Sigma_{\beta\beta} \left(\hat{\beta}_m - \hat{\beta}\right) = \kappa^2$$

which is nonlinear w.r.t. the single unknown \(\lambda\). \(\lambda\) can be determined numerically by a common root-finding method. When its actual value is known, Eqs. (20) and (21) yield

$$\mu - \hat{\beta}_m = \lambda \ U_{\beta\beta} \Sigma_{\beta\beta}^{-1} \left(\mu - \hat{\beta}\right) = \lambda \ U_{\beta\beta} \Sigma_{\beta\beta}^{-1} \left(U_{\beta\beta}^{-1} - \lambda \ \Sigma_{\beta\beta}^{-1}\right) U_{\beta\beta} \left(\hat{\beta}_m - \hat{\beta}\right)$$

(22)

Finally the membership function of the imprecise confidence region is obtained regarding Eq. (16) as

$$m_{\gamma_1} \left(\hat{\beta}\right) = \left\{ \begin{array}{ll}
1, & \hat{\beta} \in K_{1-\gamma} \left(\hat{\beta}_m\right) \\
\text{h} \left(\left(\mu - \hat{\beta}_m\right)^T \left(U_{\beta\beta}^{-1} - \lambda \ \Sigma_{\beta\beta}^{-1}\right) \left(\mu - \hat{\beta}_m\right)\right), & \hat{\beta} \not\in K_{1-\gamma} \left(\hat{\beta}_m\right) \end{array} \right.$$  

(23)

with \(\left(\mu - \hat{\beta}_m\right)\) as given in Eq. (22).

Please note that the second derivative of \(\Phi\) w.r.t. \(\mu\)

$$\frac{\partial^2 \Phi(\mu)}{\partial \mu^2} = 2 \left(U_{\beta\beta}^{-1} - \lambda \ \Sigma_{\beta\beta}^{-1}\right)$$

has to be positive definite to assure a minimum what can easily be checked.

Eq. (23) describes the extension of classical confidence regions to confidence regions where the originally random-type quantities of interest are superposed by imprecision. It is a significant generalization of the corresponding Eq. (20) in Kutterer (2001b) where \(U_{\beta\beta}\) had to be proportional. Like in the analysis of GPS observations both the phase ambiguity search spaces and the precision of point positions are represented by confidence regions, Eq. (23) plays the key role in any case when imprecision has to be taken into account.

Fig. 4 shows exemplarily for the 2D case the superposition of a classical \((1-\gamma)\) confidence ellipse and an imprecise 2D vector of elliptic type. It is obvious that the superposition of the two quantities does not yield an elliptic quantity. The maximum membership degree is obtained for the mean point of the imprecise vector and the corresponding confidence ellipse. This is only valid for the classical confidence region.

The qualitative difference of the presented results from the common understanding of accuracy is obvious from the extended LSE and the extended confidence regions. Actually, the introduced combined measures of stochasticity and imprecision are closer to the idea of accuracy in practical applications. The classical statistical point of view implies reduction of uncertainty just by repetition of observations. If fuzzy-theory is used to model and handle imprecision this is not possible. The amount of imprecision is kept when observations are repeated. Imprecision can only be reduced outside the particular observation scenario as it is according to common sense.
4 Extended GPS phase ambiguity search spaces

The linearized functional model of GPS code and phase observations reads as

$$ E(y) = X\beta = A\xi + Z\zeta $$

(24)

with the expectation $E(.)$, the real-valued parameters $\xi$ such as coordinates and the integer ambiguity parameters $\zeta$. The matrices $A$ and $Z$ denote the two corresponding components of the configuration matrix $X$. Please note that the vector $y$ comprises both code and phase observations or differences, respectively. For the following there is no need to distinguish between undifferenced and double-differenced observations. The only impact is then on the adequate parametrization. The rows of matrix $Z$ which correspond with the code observations are naturally equal to zero. The vcm or dispersion matrix of $y$ given by

$$ D(y) = \Sigma_{yy} $$

(25)

A real-valued approximation of the integer ambiguity parameters is obtained by a least-squares estimation weighted by $\Sigma_{yy}^{-1}$ as

$$ \hat{\zeta} = F y $$

(26a)

with

$$ F = \left(Z^T \Sigma_{yy}^{-1} \left(\Sigma_{yy} - A \left(A^T \Sigma_{yy}^{-1} A\right)^{-1} A^T\right) \Sigma_{yy}^{-1} Z\right)^{-1} \left(Z^T \Sigma_{yy}^{-1} \left(\Sigma_{yy} - A \left(A^T \Sigma_{yy}^{-1} A\right)^{-1} A^T\right) \Sigma_{yy}^{-1} Z\right)^{-1} $$

(26b)

what leads to

$$ \Sigma_{\hat{\zeta}}^{-1} = F \Sigma_{yy}^{-1} F^T = \left(Z^T \Sigma_{yy}^{-1} \left(\Sigma_{yy} - A \left(A^T \Sigma_{yy}^{-1} A\right)^{-1} A^T\right) \Sigma_{yy}^{-1} Z\right)^{-1} \left(Z^T \Sigma_{yy}^{-1} \left(\Sigma_{yy} - A \left(A^T \Sigma_{yy}^{-1} A\right)^{-1} A^T\right) \Sigma_{yy}^{-1} Z\right)^{-1} $$

(27)

for the vcm of the real-valued estimates of the integer ambiguity parameters. Consequently, the corresponding $(1-\gamma)$-confidence hyperellipsoid regarding Eq. (12) reads as

$$ K_{f,1-\gamma}(\hat{\zeta}) = \left\{ \mu \in \mathbb{R}^f : \mu - \hat{\zeta} \Sigma_{\hat{\zeta}}^{-1} \left(\mu - \hat{\zeta}\right)^T \leq \chi_{f,1-\gamma}^2 \right\} $$

(28)

with $f$ denoting the number of ambiguity parameters; see Kutterer (2001b). The confidence region given in Eq. (28) can be set up based on code observations only. It serves as a search space for the integer ambiguity parameters.

The methods proposed in literature for ambiguity resolution differ mainly in the strategy how to identify the „correct“ ambiguity parameter. In any case they depend and rely on the adequateness of the models given in Eqs. (24) and (25). In particular, the functional model has to be accurate in the meaning that the existing errors are only assignable to the observations and that they are all and exclusively random. However, this does not hold in general. This assumption is certainly not suitable for short observation times or real-time applications and for long baselines. Hence, the imprecision of the observations has to be assessed and modelled as mentioned above. Then it has to be superposed to the search space by applying the procedure shown in Section 3, in particular by using Eq. (23).

As the extended search space is obviously enlarged, more candidate vectors have to be taken into account for ambiguity resolution. If a rounding procedure is applied such as the LAMBDA method (Teunissen and Kleusberg, 1998), there is no change for the integer-estimated ambiguity vector. However, due to the increased number of candidates the separability of the best and the second-best solution may be reduced which leads to more reliable results. In all other methods like, e.g., On-The-Fly algorithms (Abidin, 1993; Leinen, 2001), the degree of imprecision given by the membership function of the extended search space offers additional information for the validity of the solution.

5 Extended error measures for GPS site positions

Imprecise $(1-\gamma)$-confidence regions (ellipses and ellipsoids, respectively) for the 3D positions of GPS sites can be given in analogy to the ambiguity resolution presented in Section 4. As soon as the ambiguity parameters are known (and fixed, respectively), the phase observations can be used as highly precise distance observations. The functional model according to Eq. (23) simplifies to

$$ E(\bar{y}) = A\xi, \text{ with } \bar{y} = y - Z\zeta $$

(29)

but the stochastic model represented by the vcm

$$ D(\bar{y}) = \Sigma_{\bar{y}\bar{y}} = \Sigma_{yy} $$

(30)

is unchanged because the introduced ambiguities are considered as exact. Least-squares estimation of the remaining real-valued unknown parameters like, e.g., position coordinates or tropospheric parameters, yields

$$ \hat{\xi} = \left(A^T \Sigma_{yy}^{-1} A\right)^{-1} A^T \Sigma_{yy}^{-1} \bar{y} $$

(31)
with the corresponding vcm

$$
\Sigma_{\xi} = (A^T \Sigma_{\gamma}^{-1} A)^{-1} \tag{32}
$$

The submatrix for a set of parameters such as the coordinates of a particular point is obtained by means of a selection matrix like, e.g.,

$$
S_i = \begin{bmatrix} 0_{3 \times 3i-1} & I_3 & 0_{3 \times (n-3i)} \end{bmatrix} \tag{33}
$$

Hence, for the position of the $i^{th}$ point it is

$$
\hat{\xi}_i = S_i (A^T \Sigma_{\tau\gamma}^{-1} A)^{-1} A^T \Sigma_{\tau\gamma}^{-1} \bar{y} \tag{34}
$$

and

$$
\Sigma_{\xi i} = S_i (A^T \Sigma_{\tau\gamma}^{-1} A)^{-1} S_i^T \tag{35}
$$

Its classical $(1-\gamma)$-confidence ellipsoid reads as

$$
K_{1-\gamma}(\xi_i) = \left\{ \mu \left| \begin{array}{c} \mu - \hat{\xi}_i \\ \Sigma_{\xi i} \end{array} \right|^T \begin{array}{c} \mu - \hat{\xi}_i \\ \Sigma_{\xi i} \end{array} \leq \chi^2_{3,1-\gamma} \right\} \tag{36}
$$

The corresponding imprecise confidence ellipsoid is obtained by means of the procedure given in Section 3, mainly using Eq. (23).

### 6. Examples

In the following, the impact of the proposed superposition of stochasticity and imprecision on the size of the search space is shown exemplarily. The main idea is to extend the classical search space by scaling the semi-axes of the respective confidence hyperellipsoid so that the result is the tightest inclusion of the support of the imprecise confidence region. From a practical point of view this is an important first step to consider imprecise observations. Below, several GPS real-time scenarios are simulated and discussed. This section is organized as follows. First, the general configuration and the estimation procedure are given. Second, the modeling of the imprecise observations and the derivation of the imprecise vector of the ambiguity parameters are described. Third, the resulting scaling factors for the classical search spaces are compiled and discussed. The results of the simulation runs were derived by means of the procedure which was described in Section 3 and which led to Eq. (23).

The scenarios are based on the nominal GPS configuration with 24 satellites which was simulated according to orbital elements published by Parkinson and Spilker (1996). The respective solutions are based on single epoch observations to all visible satellites. The number of satellites was controlled by means of an elevation mask: If $n$ satellites were visible and $m < n$ satellites were considered, those $n-m$ satellites with lowest elevation were dropped. The $(1-\gamma)$-confidence hyperellipsoids are based on a code-only approximate position. The standard deviation of the code observations was chosen as $0.3$ m in order to obtain realistic magnitudes. The integer ambiguity parameters were approximated by means of a least-squares adjustment (float solution) as it is common practice in GPS data analysis. From Fig. 4 it is already clear that the resulting imprecise confidence regions are no hyperellipsoids but more complex quantities. They are fuzzy supersets of the classical precise hyperellipsoids.

The actual ratio of stochasticity and imprecision of the observations depends on the respective configuration. It is part of the complete uncertainty budget. The relevant types of uncertainty can be described and quantified by experts in several ways using a detailed questionnaire: A sensitivity analysis of the applied correction models for example gives insight in critical parameters, site-dependent effects such as multipath can be assessed by studying the local situation, and extensive controlled variations of the observation configuration indicate the magnitude of external effects.

In the following examples some illustrative values were chosen for the amount of imprecision. The imprecise vector of the real-valued approximation of the ambiguity parameters was represented as an imprecise vector of elliptic type. It was deduced from its range of values (convex polyhedron) which is directly computable from the imprecision of the observations in the considered cases because of the relatively low dimension of the parameter space. This polyhedron was then enclosed by the tightest possible hyperellipsoid in order to define the support of an imprecise vector of elliptic type. A linear reference function was chosen as in Figures 1 and 3 and Eq. (3), respectively.

For the first simulation runs the imprecision of all code observations was introduced as $0.03$ m (10% of the value of the standard deviation) what is a very restrained assumption. Table 1 shows the scaling factors of the semi-axes of the classical search space which were obtained for GPS observation sites in three different latitudes: equatorial region (Latitude $= 0^\circ$), mean latitudes ($45^\circ$) and the poles ($90^\circ$). It is obvious that the scaling factor depends only slightly on the configuration - less on the latitude and more on the number of satellites. The latter is due to the fact that the amount of imprecision increases by the number of observations. There is no significant dependence of the results on the time of observation. By taking an average value of $1.25$ one can state that the assumed imprecision of 10% requires an extension of the search space by 25%.
The quality of the approximation of the actual imprecise search space by scaling the classical (precise) one becomes poorer with increasing importance of imprecision. In such cases the individual scaling factors for the respective semi-axes can differ significantly. Fig. 4 illustrates this: The imprecise confidence ellipse is principally obtained by superposing two ellipses; however, the resulting quantity is not elliptic and cannot be represented uniquely by an ellipse. Nevertheless, if the maximum values for the scaling factors are taken the inclusion property is kept in any case.

The examples indicate that the ratio of stochasticity and imprecision plays a leading role in the extension of the classical search space in order to take imprecision into account. Hence, the traditional procedure of ambiguity resolution is inadequate when imprecision dominates the uncertainty budget. A rule-of-thumb for practitioners reads as follows: The search space needs to be extended even in the case of low imprecision. When the observations are likely to be imprecise at least in the same magnitude as stochasticity the semi-axes of the search space should be lengthened by a factor of at least 3. In this way the quality of the validation of the resolved ambiguity vector can be improved. Some remarks on this topic were also made at the end of Section 4.

7 Conclusions and outlook

As imprecision has to be considered in a variety of geodetic applications of the GPS the joint treatment of stochasticity and imprecision in GPS data analysis is important. Imprecision is an independent type of uncertainty and in general it can not be reduced or transformed to stochasticity. Hence, the common modeling as given in Eqs. (24) and (25) is incomplete because it does not take imprecision into account. Fuzzy-theory allows to distinguish strictly between these two types of uncertainty. For this reason it is suitable to handle both stochasticity and imprecision. Moreover, it allows to control the type of propagation of imprecision from the observations to the parameters; see Eq. (7) for linear propagation and Eq. (9) for quadratic propagation, respectively. In both cases the classical least-squares estimator is kept as mean point of the resulting fuzzy set.

The benefit of the joint treatment of stochasticity and imprecision for the GPS community is two-fold. On the one hand the resolution of the phase ambiguity parameters can be improved by extending the classical search space by simple scaling. This is a first step to more reliable results in real-time GPS. On the other hand the quality of the point positions determined by means of GPS can be described more thoroughly. It is well known that their formal precision is too optimistic. Imprecise confidence regions can objectify the common measures of precision and accuracy since imprecision cannot be reduced by repeated observations.

There are some issues which are worthwhile for further studies. First, the uncertainty budget of GPS observation configurations has to be evaluated thoroughly for the practical application of the presented approach. Thus, a look-up table for observation imprecision could be worked out for typical configurations. Second, the notion of imprecision was based here on L-fuzzy numbers which imply identical left and right spreads; see standard references on fuzzy-theory. In this way the knowledge of possible asymmetries in remaining systematic effects could be modeled what would lead directly to biases in least-squares estimation which have to be taken into account. Third, it is up to now not sufficiently understood how imprecision propagates in practice from the observations to the parameters of interest. There could be more possibilities than the linear and the quadratic propagation which were considered in this paper.

<table>
<thead>
<tr>
<th>% of std.dev. → # of sat. ↓</th>
<th>0.1σ_{code}</th>
<th>0.2σ_{code}</th>
<th>0.5σ_{code}</th>
<th>1.0σ_{code}</th>
<th>2.0σ_{code}</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.21</td>
<td>1.41</td>
<td>2.03</td>
<td>3.06</td>
<td>5.12</td>
</tr>
<tr>
<td>5</td>
<td>1.23</td>
<td>1.47</td>
<td>2.17</td>
<td>3.33</td>
<td>5.65</td>
</tr>
<tr>
<td>6</td>
<td>1.25</td>
<td>1.51</td>
<td>2.26</td>
<td>3.52</td>
<td>6.05</td>
</tr>
<tr>
<td>7</td>
<td>1.26</td>
<td>1.52</td>
<td>2.31</td>
<td>3.61</td>
<td>6.23</td>
</tr>
<tr>
<td>8</td>
<td>1.27</td>
<td>1.54</td>
<td>2.35</td>
<td>3.71</td>
<td>6.41</td>
</tr>
</tbody>
</table>

Tab. 1 Maximum scaling factor for the semi-axes of the classical (precise) ambiguity search space in the real-time case (single epoch observations) to take imprecision into account. The standard deviations of the code observations equal 0.3 m, the imprecision equals 0.03 m.

In further simulations runs the ratio of imprecision and stochasticity (in terms of standard deviations) was varied. Table 2 shows the results which were derived for the GPS observation site with latitude = 45° for different ratios. A linear dependence of the scaling factor on the chosen ratio can be found. In case of identical magnitudes of stochasticity and imprecision (ratio=1) the semi-axes of the search spaces need to be increased by a factor significantly larger than 3.

<table>
<thead>
<tr>
<th># of sat. ↓</th>
<th>0°</th>
<th>45°</th>
<th>90°</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1.20</td>
<td>1.21</td>
<td>1.20</td>
</tr>
<tr>
<td>5</td>
<td>1.23</td>
<td>1.23</td>
<td>1.23</td>
</tr>
<tr>
<td>6</td>
<td>1.25</td>
<td>1.25</td>
<td>1.25</td>
</tr>
<tr>
<td>7</td>
<td>1.25</td>
<td>1.26</td>
<td>1.27</td>
</tr>
<tr>
<td>8</td>
<td>1.27</td>
<td>1.27</td>
<td>1.27</td>
</tr>
<tr>
<td>9</td>
<td>1.29</td>
<td>___</td>
<td>1.28</td>
</tr>
</tbody>
</table>

Tab. 2 Maximum scaling factor for the semi-axes of the classical (precise) ambiguity search space in the real-time case (single epoch observations) to take imprecision into account. The standard deviations of the code observations equals 0.3 m, the imprecision is varied.
References


Short-Arc Batch Estimation for GPS-Based Onboard Spacecraft Orbit Determination

Ning Zhou, Yanming Feng
Cooperative Research Centre for Satellite Systems, Queensland University of Technology, Australia

Received: 4 October 2002 / Accepted: 20 January 2003

Abstract. In dynamic orbit determination, the problem is that a batch estimator assumes use of more sophisticated models for both force and observation models, dealing with large amounts of observations. As a result, the computational workload may not be acceptable for onboard orbit determination. In this paper, the short-arc batch estimation is experimentally studied in order to address both estimation robustness and computational problems in GPS-based onboard orbit determination. The technical basis for the batch estimation will be outlined. The experimental results from three 96-hour data sets collected from Topex/Poseidon (T/P), SAC-C and CHAMP missions are presented. These results have demonstrated that use of shorter data arcs allow for simplifications of both orbit physical and observational models, while achieving a 3D RMS orbit accuracy of meter level consistently.

Key words: GPS, Onboard orbit determination, Batch estimation, Low Earth Orbiter (LEO).

1 Introduction

Onboard accurate orbit determination is a fundamental step towards autonomous satellite operation and navigation. Onboard stand-alone GPS navigation solutions are as accurate in low earth orbit as on the ground: currently a RMS positional accuracy of 10 to 20 meters achievable with zero Selective Availability (SA), using the civilian broadcast GPS signals. A satellite orbit is highly predictable with initial states. However, accumulation of orbit force errors may cause orbit solutions to fail. An orbit filter will make use of observations along the orbit to correct force model parameters and provide improved orbit solutions.

Particularly, an orbit improvement procedure is of interest in the following circumstances:

- A higher orbit accuracy, for instance, of meter level, is needed to satisfy advanced space engineering applications, including satellite flying formation and docking, etc (Bertiger et al, 1998). In addition, a filtered orbit can lead to a more accurate predicting orbit.
- Continuous orbit information is required, but GPS navigation solutions are only available at discrete time epochs, especially when onboard GPS operates intermittently. For instance, the Australia Federation satellite –FedSat- operates 2-by-10 minutes in each orbit period, because of the restriction of on-board power supply (Feng, 1999).
- GPS-based onboard navigation solutions cannot be provided regularly as the number of GPS satellites in view are sometimes fewer than four. An example is the satellite flying in Geostationary (GEO) orbits, where GPS signals from an average of one to two GPS satellites are tracked from space by the down-looking antenna (Mehlen et al, 2001, Yunck, 1996).
- There always exist orbital modeling errors, which sometimes grow beyond the GPS observation uncertainty. Filtering techniques will correct or reduce effects of these modeling errors.

In order to address these problems, the paper presents a robust filtering strategy for onboard spacecraft orbit determination, which allows use of variable data intervals for filtering updates to achieve optimal overall orbit estimation accuracy and solution stability. Solution stability is defined as solution convergence with respect to the epoch state (Feng et al, 1997). Kalman filtering requires a long data arc to reach the convergent solution. However, dynamic model errors may be accumulated rapidly in the long nominal orbit and the batch least square over the long orbit incurs heavy computational
burden, which may be unacceptable for the spacecraft onboard processing environment.

After a brief description of the batch estimation algorithms, the paper presents extensive experimental results from three Low Earth Orbiter (LEO) missions. The experimental studies include both commission and omission errors in an attempt to arrive at a realistic error estimate. The data analysis will focus on effects of orbit dynamic models, GPS measurement quality and the performance of batch orbit filtering solutions with different lengths of data arcs.

2 Theoretical Basis

From the point of view of celestial dynamics, the differential equation of motion of a satellite could be expressed in this form:

\[
\ddot{\mathbf{r}} = -\frac{\mathbf{r}}{r^3} \left( \frac{GM}{r^2} \mathbf{r} - \mathbf{F}(1, \mathbf{r}, \mathbf{r}) \right)
\]

where:
- \( \ddot{\mathbf{r}} \) is the satellite acceleration vector
- \( \mathbf{r} \) is the satellite position vector
- \( GM \) is the product of the gravitational constant G and earth mass M
- \( \frac{GM}{r^2} \) is the acceleration force due to the central body of the earth
- \( \mathbf{F} \) is a function of the spacecraft state and time, it represents all the perturbation forces acting on the satellite

The perturbed forces acting on the spacecraft include non-spherically and inhomogeneous mass distribution within the Earth (central body); the third celestial bodies (sun, moon etc), earth and oceanic tides; the atmospheric drag, solar radiation pressure and geomagnetic effects, etc. Simplification of force models is necessary in the onboard processing environment. However, for low earth orbiters (LEO), special care has to be taken to minimise the effects of the remaining modeling errors of the atmospheric drag force, in order to achieve the required orbit accuracy.

The explicit term for the acceleration due to the atmospheric drag can be presented as

\[
F_D = -\frac{1}{2} \left( \frac{C_DS}{m} \right) \rho VV
\]

where \( C_D \) is the drag coefficient, \( S/m \) is the ratio of spacecraft effective area to its mass; \( \rho \) is the atmospheric mass density at the current location of the spacecraft; \( V \) is the velocity vector relative to the kinetic atmosphere; \( v \) and \( v_a \) are the geocentric velocity vectors of the satellite and atmosphere. It is obvious that \( F_D \) depends on parameters \( C_D \), \( S/m \), \( v_a \), and the distribution of atmospheric mass density. The difficulty is that all the three quantities have uncertainties:

- the drag coefficient \( C_D \) is an empirical number,
- the ratio \( S/m \) varies due to the attitude variation of the satellite traveling along its path;
- the rotating velocity of the kinetic atmosphere varies from 0.8 to 1.4 for the orbits between 200km to 1200 km;

The mass density \( \rho \) for the air particles responds sensitively to the solar activity, season, longitude, latitude, local time and magnetic storm conditions. The widely referred models include those in the CIRA (Cooperative Institute for Research in the Atmosphere) series, such as CIRA-61, CIRA-65, CIRA 72, and CIRA 86; those in the Jacchia series, such as J-65, J-71 and J-77. There are also MSIS83, MSIS86, MSIS 90 (Hedin, 1991) and Drag Temperature Model (DTM) (Barlier at al.1977, Bruinsma and Thuiller, 2000). To allow for easy autonomous onboard processing, we use a simplified model for the calculation of the upper atmospheric density (Liu, 2000):

\[
\rho = \rho_0 \left( 1 + \frac{\mu}{2} \left( \frac{r - \sigma}{H_0} \right)^2 \exp\left[ -\left( \frac{r - \sigma}{H_0} \right) \right] \right)
\]

In this equation, \( \rho_0 \) is the density of the Earth’s atmosphere at a reference point with the altitude \( H_0 \); \( r \) is the altitude of the spacecraft; \( \mu = 0.10 \), \( \sigma \) is the distance between the centre of the Earth and the reference point. In the batch estimation, the value of \( B = -\frac{1}{2} \left( \frac{C_DS}{m} \right) \) is considered as a constant over a short data arc (eg, a few to several hours), or as a function of the arc length:

\[
B = B_0 + B_1 (t - t_0)
\]

to be estimated together with the orbit state parameters.

Satellite orbit determination has two distinct procedures: orbit integration and orbit improvement. Orbit integration yields a nominal orbit trajectory while orbit improvement estimates the epoch state with all the measurements collected over the data arc in a batch estimation manner. Generally, numerical methods of varying complexity are applied for propagating the state vector between its update intervals, which are of minutes, hours or days. There are many numerical methods to solve the differential equation, such as RK (F), Adams and Cowell methods. An efficient method of orbit integration, called the Integral Equation (IE) method, has been developed in our research efforts. The numerical solution of the
Integral Equation is theoretically equivalent to that of a differential equation for a motion of a satellite, but the algorithm of Integral Equation is simpler, and can be easily implemented for onboard processing. The state solution can be summarized as follows (Feng, 2001):

\[
X_x(t) = \Phi(t, t_0)X_x(t_0) + \int_{t_0}^{t} \Phi(t, \tau)F[r, X_x(\tau)]d\tau
\]

(5)

Where \( X \) is the state vector (position and velocity), \( \Phi(t, t_0) \) is the state transition matrix from \( t_0 \) to \( t \), and \( F[r, X_x(\tau)] \) is the state vector at the initial epoch \( t_0 \) which can be in the beginning, middle or end of the data arc:

\[
\Delta X(t) = \Phi_x(t, t_0)\Delta X(t_0) + \Phi_\mu(t, t_0)\Delta \mu
\]

(6)

where, \( \Delta X(t) \) is the 6-by-1 state vector; \( \Delta \mu \) is a physical parameter vector related to solar radiation pressure and/or atmospheric drag coefficients, depending on the orbits and data arcs. It was these state transition matrices that make the numerical solution of the integral equation comparatively simple.

To estimate the initial epoch state of the orbit, we need to establish a state equation that relates the state derivations of the current epoch \( t \) to the initial epoch \( t_0 \), which can be in the beginning, middle or end of the data arc:

\[
\Delta X = \Phi x(t, t_0)\Delta X(t_0) + \Phi \mu(t, t_0)\Delta \mu
\]

(6)

where, \( \Delta X \) is a 6-by-1 state vector; \( \Delta \mu \) is a physical parameter vector related to solar radiation pressure and/or atmospheric drag coefficients, depending on the orbits and data arcs. It was these state transition matrices that make the numerical solution of the integral equation comparatively simple.

To estimate the initial epoch state of the orbit, we need to establish a state equation that relates the state derivations of the current epoch \( t \) to the initial epoch \( t_0 \), which can be in the beginning, middle or end of the data arc:

\[
\Delta X = \Phi x(t, t_0)\Delta X(t_0) + \Phi \mu(t, t_0)\Delta \mu
\]

(6)

where, \( \Delta X \) is a 6-by-1 state vector; \( \Delta \mu \) is a physical parameter vector related to solar radiation pressure and/or atmospheric drag coefficients, depending on the orbits and data arcs. It was these state transition matrices that make the numerical solution of the integral equation comparatively simple.

To estimate the initial epoch state of the orbit, we need to establish a state equation that relates the state derivations of the current epoch \( t \) to the initial epoch \( t_0 \), which can be in the beginning, middle or end of the data arc:

\[
\Delta X = \Phi x(t, t_0)\Delta X(t_0) + \Phi \mu(t, t_0)\Delta \mu
\]

(6)

where, \( \Delta X \) is a 6-by-1 state vector; \( \Delta \mu \) is a physical parameter vector related to solar radiation pressure and/or atmospheric drag coefficients, depending on the orbits and data arcs. It was these state transition matrices that make the numerical solution of the integral equation comparatively simple.

To estimate the initial epoch state of the orbit, we need to establish a state equation that relates the state derivations of the current epoch \( t \) to the initial epoch \( t_0 \), which can be in the beginning, middle or end of the data arc:

\[
\Delta X = \Phi x(t, t_0)\Delta X(t_0) + \Phi \mu(t, t_0)\Delta \mu
\]

(6)

where, \( \Delta X \) is a 6-by-1 state vector; \( \Delta \mu \) is a physical parameter vector related to solar radiation pressure and/or atmospheric drag coefficients, depending on the orbits and data arcs. It was these state transition matrices that make the numerical solution of the integral equation comparatively simple.
for onboard orbit determination. In this research effort, we test short-arc batch estimation strategies to address both orbit accuracy and computational burden problems for onboard orbit determination with GPS code measurements.

3 Experimental Results

The purpose of the experimental studies is to evaluate the performance of the proposed batch estimation strategies for onboard orbit determination, against different data arc lengths. Experimental results are obtained from three LEO missions: TOPEX/Poseidon, SAC-C and CHAMP. Their orbit altitudes are 1340km, 700km and 450km respectively.

Topex/Poseidon (T/P) is a joint project between the National Aeronautics and Space Administration (NASA) and the French Space Agency, Centre National d’Etudes Spatiales (CNES). The T/P satellite carries a 6-channel Motorola Monarch Receiver, which is capable of collecting dual-frequency (L1/L2) data when the GPS anti-spoofing (AS) function is inactive.

CHAMP was launched in July 2000 into a circular orbit of 450 kilometres to support geoscientific and atmospheric research; the mission is managed by GFZ in Germany. The GPS payload consists of a BlakJack receiver with 3 antennas, the facing-up antenna provides data for precise orbit determination services, the down facing one for GPS altimeter and the limb antenna for atmospheric sounding (Kuang, 2001).

SAC-C is an international cooperative mission between NASA and the Argentine Commission on Space Activities (CONAE). SAC-C provides multi-spectral imaging of terrestrial and coastal environments. It carries a TurboRogue III GPS and four high gain antennas developed by the JPL. It is capable of automatically acquiring selected GPS transmissions that are refracted by the Earth’s atmosphere and reflected from the Earth’s surface.

3.1 Measurement Quality and Single Point Positioning Errors

To which extent the orbit solution can be improved by using the batch estimation or filtering procedure depends on not only the estimation models and algorithms, but also the quality of actual measurements. In the discussion below, we present evaluation results for the measurement accuracy and single point positioning solutions (i.e., navigation solutions).

Evaluation of code measurement noise level is based on the following equations:

\[
P_{\text{M}}(t) = [P(t+1) - P(t)] - \lambda [L_{\text{1}}(t+1) - L_{\text{1}}(t)]
\]

\[
P_{\text{CM}}(t) = [P_{\text{C}}(t+1) - P_{\text{C}}(t)] - \lambda [L_{\text{C}}(t+1) - L_{\text{C}}(t)]
\]

\[t = 1, 2, 3, \ldots\] (12)

where \( P_{\text{C}} \) is ionosphere-corrected code measurements, \( \lambda \) is the wavelength of L1 frequency (1575.42MHz). \( P_{\text{M}} \) and \( P_{\text{CM}} \) mainly contain receiver noise and multipath errors. The standard deviations of the observations \( P_{\text{1}} \) and \( P_{\text{C}} \) are given as:

\[
\sigma_{P_{\text{1}}} = \sqrt{\frac{\sigma_{P_{\text{M}}}^2}{2}} \quad \sigma_{P_{\text{C}}} = \sqrt{\frac{\sigma_{P_{\text{CM}}}^2}{2}}
\] (13)

Due to possible variation of atmospheric conditions between epochs, \( \sigma_{P_{\text{1}}} \) is a conservative estimate of the standard deviation for the measurements \( P_{\text{1}} \).

Tab. 1 provides a summary of the three sets of GPS flight data. As mentioned above, all data are SA free. Tab. 2 summarizes the RMS values for the three data sets against elevation angle. It is observed that the GPS data with elevation angle below 10 degrees are much noisier than those with higher elevation angles. This is particularly true for CHAMP and SAC-C data sets. Nevertheless, the noise levels of \( P_{\text{1}} \) code measurements in the three data sets are still normal. They are 52cm, 32cm and 34cm respectively, showing a consistent data quality.

| Tab. 1 Summary of GPS data sets |
|---------------------|-----------------|-----------------|
| Satellite          | CHAMP           | SAC-C           | T/P             |
| Start date:         | 13/02/2002      | 14/02/2002      | 09/10/2001      |
| Data arc length: (hour) | 96              | 96              | 96              |
| Data Type:          | P1, P2          | P1, P2          | P1              |
| Sample Rate         | 10 seconds      | 10 seconds      | 5 minutes       |
| Effective measurements | 33,992 epochs | 34,496 epochs | 1,125 epochs |
|                      | 243,762 observables | 198,249 observables | 4,580 observables |

| Tab. 2 Standard deviation of \( P_{\text{1}} \) measurements for T/P, CHAMP and SAC-C flight data |
|---------------------|-----------------|-----------------|
| Satellite | Standard Deviation | All data | Elev <10 | 10 < Elev <25 | Elev >25 |
| CHAMP          | Stddev (cm) | 52.6 | 93.0 | 63.9 | 41.7 |
|                | % | 4.7% | 25.6% | 69.7% |
| SAC-C          | Stddev (cm) | 32.0 | 73.8 | 46.3 | 17.4 |
|                | % | - | 4.2% | 28.0% | 68.3% |
| T/P            | Stddev (cm) | 34.2 | 38.0 | 38.5 | 32.9 |
|                | % | - | 1.7% | 18.9% | 79.4% |

The single point positioning (SPP) solutions for CHAMP and SAC-C data were performed using \( P_{\text{1}} \) code measurements. The differences between SPP solutions and JPL’s POD solutions were obtained for all the data points where there are 4 or more satellites in view. Fig. 2 illustrates the 3D RMS positional accuracy with the...
CHAMP data set, plotted against the GDOP values and visibility of GPS satellites. It is clearly seen that there are indeed quite a few data points where only 2 or 3 satellites are visible. With sufficient satellites, GDOP values are evidently worse than those normally experienced on ground. As a consequence, onboard SPP solutions are frequently corrupted, with many cases where the 3D RMS positional uncertainty exceeds 100 meters with SA-free. This fact again shows the importance of on-board orbit improvement procedure to overcome the solution outages.

3.2 Batch Estimation Results

Batch estimation processing is performed with the above-mentioned data sets. We first present results with given atmospheric drag coefficient and Solar Radiation Pressure parameter (the default value of the model or estimated from somewhere else), where only 6 state variables are estimated over each data arc. For the whole orbit of 96 hours, the estimation process proceeds with six choices of data intervals: 1h, 2h, 6h, 12h, 24h and 48h. Figure 3 illustrates the 3D RMS orbit errors of the 96h SAC-C orbit, obtained with three data arc options: 1h, 2h and 6h. Figure 4 summarises the overall 3D RMS orbit errors resulted from each data set. Figure 5 compares the batch estimation results from the SAC-C data sets, using 2-h data intervals, with the SPP solutions.

Next, we present results with the atmospheric drag coefficient and Solar Radiation Pressure coefficient estimated along with the six state variables. Figure 6 illustrates the 3D RMS orbit errors of the SAC-C orbit again. Figure 7 summarises the overall 3D RMS orbit errors under different filter strategies for each data set.
Batch estimation with a data arc of either too short (e.g., less than 1 hour) or too long (e.g., example, over 24 hours) produces poorer filtering results. In general, a data arc of one to four orbit periods appears sufficient for orbit estimation with the state equations with six-state parameters along with atmospheric drag and solar radiation pressure parameters. If these physical parameters are estimated simultaneously, better results can be achieved with longer data arcs for the tested LEO orbits.

4 Conclusions

A dynamic approach is necessary to onboard orbit determination at different altitudes for achieving meter-level orbit accuracy and providing continuous orbit solutions in the circumstances where there are spare samples and/or fewer GPS observations. Our research efforts have been made to test the simple and robust dynamic method—short-arc batch estimation—in order to address both orbit accuracy and computational burden problems for onboard orbit determination with GPS code measurements.

The experimental results from three 4-day data sets from Topex/Poseidon, SAC-C and CHAMP missions have demonstrated that use of shorter data arcs allows for simplifications of both physical and observational models. With a data arc of a few hours, the batch estimation procedure that estimates drag coefficient and solar radiation pressure along with the six-state parameters achieves a 3D RMS orbit accuracy of meter level consistently with GPS code measurements.

From these figures, we have the following observations:

- With a data arc of as short as 2 hours, batch estimation for the tested orbits achieves a 3D RMS orbit accuracy of meter level consistently with GPS code measurements. It is noteworthy that the orbit periods for T/P, SAC-C and CHAMP are about 112, 99 and 94 minutes respectively.
- The data arcs required to achieve the best batch estimation results strongly depend on the omissions and commissions of orbit dynamic models and state equations, as well as the orbit altitudes. For instance, for the SAC-C and CHAMP orbits, the best batch estimates are achieved with the data arcs of 6 and 2 hours, respectively, when the six-state parameters are estimated and the atmospheric drag and solar radiation pressure coefficients are fixed to the known. If these two additional physical parameters are estimated together with the state parameters, the data arcs for the best orbit filtering results for these two orbits are extended to 12 hours and 6 hours respectively.

Acknowledgements

This work was carried out in the Cooperative Research Centre for Satellite System with financial support from the Commonwealth of Australia through the Cooperative Research Centre program and Queensland State Government.

References


Internet-based GPS VRS RTK Positioning with a Multiple Reference Station Network

G.R. Hu, V. H.S. Khoo, P. C. Goh, C. L. Law

GPS Centre, College of Engineering, Nanyang Technological University, 639798, Singapore
e-mail: egrhu@ntu.edu.sg; Tel: +65-67904105; Fax: +65-67906059

Received: 27 November 2002 / Accepted: 27 December 2002

Abstract. Multiple reference station networks have been established for high precision applications in many countries worldwide. However, real-time application is still a difficult task in practice. Virtual reference station (VRS) concept is an efficient method of transmitting corrections to the network users for RTK positioning. Today’s challenge for VRS RTK positioning lies in adapting advance wireless communication technologies for real time corrections. With the availability of GPRS technology, an Internet-based VRS RTK positioning infrastructure via GPRS has been developed and tested. This paper discusses the VRS data delivery mechanism, and gives an overview on VRS data generation for RTK positioning. Field test results are presented to evaluate the performance of the proposed system. The results demonstrate that Internet-based VRS RTK positioning can be achieved to better than 4 centimeters accuracy in horizontal position. Height accuracy is in the range of 1 to 6 centimeters.

Key words: GPS, Accuracy, Networks, Errors.

1 Introduction

GPS Real Time Kinematic (RTK) positioning is becoming increasingly important for many precise GPS applications - surveying, construction, precision farming and high accuracy Geographic Information System (GIS). Traditionally, a user receiver requires a reference station within 10 kilometers to ensure centimeter level accuracy. Moreover, multiple reference station networks have been installed in many countries to overcome the limitations of standard RTK systems.

During the past few years, different approaches have been tested to take advantage of the availability of multiple reference stations. The focus of most of the research has been on modeling the spatial behavior of the distance-dependent errors (especially ionospheric errors). However, little research has been conducted on the distribution of these corrections to potential GPS users located within, and surrounding the network coverage area in real time. This is an integral part of GPS RTK positioning and it must be adequately addressed before a practical realization of the multiple reference station network is implemented (Fotopoulos et al., 2001).

Recently, the use of VRS (Virtual Reference Station) concept has been proposed by many research groups as a more feasible approach for relaying network correction information to the network RTK users (see, Wanninger, 1997; Marel, 1998; Vollath et al., 2000; Cannon et al., 2001; Euler et al., 2001). This approach does not require an actual physical reference station. Instead, it allows the user to access data of a non-existent virtual reference station at any location within the network coverage area. Also, the VRS approach is more flexible in terms of permitting users to use their current receivers and software without involving any special software to manage simultaneously corrections from a series of reference stations. With VRS, users within the reference station network can operate consistently at greater distances without degrading accuracy. It is necessary, however, to provide a reliable data communication link for the transmission of VRS data from a control center to a user receiver.

There are several ways of transferring the GPS data for RTK positioning. Hiromune et al. (2000) have investigated GPS RTK positioning using TV audio data broadcast and evaluated its validity. Global differential corrections are already available currently for DGPS applications over the open Internet via a TCP server.
running at JPL (http://gipsy.jpl.nasa.gov/igdg, 2002; Muellerschon et al., 2000). For certain wireless services, there are frequency and power restrictions which regulate the use of such data transmission devices. GSM, a widely available public service, can be used as a distribution channel for the VRS data (Vollath et al., 2000). Using GSM as the communication link is unfortunately relatively expensive. From an operational point of view, cost is a very important consideration (Petrovski et al., 2001). This cost may be reduced using GPRS (General Packet Radio Service).

Every wireless communication mode has its pros and cons but must in general be able to support VRS RTK data link with low data latencies, good mobile performance, inexpensive user equipment, and nationwide coverage. The objective of this paper is to demonstrate a new method referred to as Internet-based VRS RTK positioning via GPRS.

This paper describes the VRS data generation approach. The Internet-based VRS data communication method is also discussed and field test results are included to evaluate the performance of the proposed system.

2 Overview of VRS Data Generation Approach

An overview of the VRS data generation approach is given in this section. In order to create data of a virtual reference station from the observations of the network of real reference stations, several processing steps have to be performed. The first step is to resolve the double-differenced carrier phase ambiguities between stations in the network.

The double-difference carrier phase observable between stations can be expressed as (e.g. Hofmann-Wellenhof et al., 1994; Leick 1995; Teunissen and Kleusberg 1998; Han, 1997)

\[
\Delta \nabla = \alpha \nabla + \Delta \Delta \nabla + \Delta \Delta \nabla - \alpha \nabla \Delta
\]

After the double-differenced ambiguities associated with the reference stations have been fixed to their correct values, the so-called correction terms for the atmospheric biases and other errors can be generated from the residuals in L1 and L2 carrier phase measurements on a satellite-by-satellite, epoch-by-epoch basis. The purpose of the corrections is to reduce the influence of the spatially correlated errors. This means that when the corrections are applied to the raw code and phase observations of the user, the influence from the atmospheric errors and other errors will be reduced or eliminated. This results in improved positioning performance. There are many methods to formulate corrections for the user (see, for example, Wanninger, 1995; Wübbena et al., 1996; Han & Rizos, 1996; Gao et al., 1997; Raquet 1998). Dai et al. (2001) showed that the performance of all the methods is similar. A distance based linear interpolation algorithm was adopted here (Wübbena et al., 1996; Han & Rizos, 1996; Gao et al., 1997). The advantage of this method for real-time use is

Examine Eq. (1), it can be seen that the double difference ambiguities between the reference stations must be known, along with precise coordinates for the reference stations. The coordinates of the reference stations may be provided by the local survey authority, in the case of a permanent regional reference network. Alternatively, they may be obtained through a static survey of each station over long periods using the traditional long-range static positioning procedure. However, even with precisely known coordinates, it is not easy to fix ambiguities between reference stations of the network in real-time.

In order to support RTK positioning, the double-differenced integer ambiguities between reference stations of the network must be resolved in real time, because these ambiguities should be recalculated instantaneously in case of cycle slips or new satellites arise or long data gap occurs. Here, a residual-based adaptive Kalman filter is proposed to resolve the double-differenced carrier phase ambiguities between stations of the network with real-time capability. This is based on data of previous epochs and not just from the current one (Chen et al. 2000, Hu et al. 2002). In order to help the network ambiguity resolution process, the orbital error de-correlation can be reduced or eliminated using the IGS UltraRapid Predicted Orbit (igu) instead of broadcast orbits. The precise ephemeris can be obtained from the International GPS Service (IGS) centres which includes one day’s predicted ephemeris (Roulston et al., 2000).

After the double-differenced ambiguities associated with the reference stations have been fixed to their correct values, the so-called correction terms for the atmospheric biases and other errors can be generated from the residuals in L1 and L2 carrier phase measurements on a satellite-by-satellite, epoch-by-epoch basis. The purpose of the corrections is to reduce the influence of the spatially correlated errors. This means that when the corrections are applied to the raw code and phase observations of the user, the influence from the atmospheric errors and other errors will be reduced or eliminated. This results in improved positioning performance. There are many methods to formulate corrections for the user (see, for example, Wanninger, 1995; Wübbena et al., 1996; Han & Rizos, 1996; Gao et al., 1997; Raquet 1998). Dai et al. (2001) showed that the performance of all the methods is similar. A distance based linear interpolation algorithm was adopted here (Wübbena et al., 1996; Han & Rizos, 1996; Gao et al., 1997). The advantage of this method for real-time use is
that interpolations are performed on an epoch-by-epoch
and satellite-by-satellite basis. The advantage of this
model is to eliminate orbit bias because the coefficients
are derived in the way that orbit bias can be removed.
The ionospheric delay and the tropospheric delay can also
be reduced to the same degree that the epoch-by-epoch
and satellite-by-satellite ionosphere and the troposphere
models can do. Multipath and measurements noises can
be reduced if the user receiver is located within the figure
formed by the reference stations so that the coefficients
are less than one. Otherwise the multipath and noise may
be amplified because the coefficients might be larger than
one (Han, 1997). In order to apply this method in real-
time, a modified 2-dimensional linear interpolation model
which uses the user’s horizontal coordinates as
parameters is employed. A more detailed description of
the approach is given in Chen et al. (2000) and Hu et al.
(2002).

In the next step, VRS data for user receivers are
generated as required. In order to generate VRS data as
though there was a reference station at the user’s location,
the user’s approximate position and the user’s position
relative to this VRS, the carrier phase and pseudorange
observations from master reference station has to be
place geometrically improved and improved by applying the
corrections of the network according to user’s
approximate position, i.e. VRS position. User’s
approximate position can be obtained by absolute GPS-
code-positioning or output in NMEA format.

\[
\rho_{x}^{s}(t) = \|x^{s} - x^{r}\| 
\]

(2)

\[
\rho_{x}^{v}(t) = \|x^{s} - x^{v}\| 
\]

(3)

The change in the geometric range
\[
\Delta \rho_{x}^{s} = \rho_{x}^{s}(t) - \rho_{x}^{s}(t) 
\]
can be applied to all observations to
displace the carrier phase and pseudorange observations
from master reference station to VRS position. After
geometric corrections are applied to master reference
station raw data, the corrections of the network are used
to displace VRS data. A standard troposphere model,
such as Saastamoinen model (Saastamoinen, 1973), can
be used to correct for tropospheric delay effects. Then the
VRS data is generated as RTCM or CMR (Compact
Measurement Record) format (Talbot, 1996) and
delivered to the user.

As described above, the VRS data generation strategy
consists of three steps. The first step is to resolve the
double-differenced carrier phase ambiguities between
stations of the network in real-time. The second step is to
generate the corrections on a satellite-by-satellite, epoch-
by-epoch basis for user according to user’s approximate
position. The third step is to form VRS data as RTCM or
CMR messages by applying the corrections to the data of
the master reference station and displaced to the location
of the VRS for users. The VRS data is then transmitted to
the user where it is fed into the receiver as single
reference station RTK corrections. The receiver can then
use these corrections just as it would in the single
reference station RTK approach. VRS data
communication link issues are discussed in the following
section.

3 VRS Data Communication Issues

An efficient communication link is critical for VRS RTK
positioning since timely transfer of the VRS data is
required for such a system. The communication link must
provide reliability for data transfer and should not cause
any significant transmission time delay. It is also desired
that the links are available without much restriction in
order to cover wide range of users. To date, there are
several practical modes of distribution of the VRS data to
users in real-time, such as GSM and Internet (Hada et al.,
1999; Vollath et al., 2000; Liu and Gao, 2001; Ko et al.,
2001).

Fig. 1 GPRS data transfer latency on 8 July 2002

Given a developed mobile phone network, the most
straightforward delivery mechanism is GSM, the low risk
approach. A cellular phone allows bi-directional
communication between a user and the VRS data control
center. Therefore the user can transmit its approximate
position to the control center. Further advantages are that
there is no need to apply for radio frequencies, and
reduced installation cost. One of the disadvantages of
GSM is the limited number of parallel users imposed on
the control center because each GSM line can only
support one user. Another major drawback of GSM is
cost as the user needs to be logged on constantly while
using the VRS RTK service. In Singapore, this costs
about 6-7 US$ per hour. This cost will be significantly
reduced with GPRS (down to 2-3 US$ per hour). More importantly, GPRS provides an “always-on” service. GPRS provides a stable and reliable connection with latencies less than 1 second. Fig. 1 shows a sample series of the data transfer delay via GPRS tested on 8 July 2002 in Singapore. For details of the test method see, for example, Liu et al. (2001). Latencies observed are usually in the order of a few hundred milliseconds in the test. Occasional latency of up to one or two seconds is not critical for GPS RTK positioning.

There is also a possibility to use the Internet as a data link between the control center and the user (Hada et al., 1999; Liu and Gao, 2001; Ko et al., 2001). Internet is the worldwide computer network system and becomes more important and common communication method mainly because of its fast data rate and availability to unlimited users. The Internet uses a bi-directional communication. Therefore, the user can also send its approximate position to the control center and request their demand to the control center, and the control center can provide VRS data for each user's demand according to the approximate position of user.

With the increased capacity of the Internet, and the recent developments in communication technology, especially of GPRS, the open Internet is a reliable choice to return GPS VRS data for RTK positioning via a GPRS. GPRS can send and receive information directly from the Internet, support the transfer of the VRS data through the Internet in a favorable way. Therefore, an Internet-based GPS VRS RTK system was developed in Nanyang Technological University. This system uses the GPRS as the communication link between the control center and user stations.

Software at control center server has been written to generate and broadcast VRS data via Internet, which is assigned with a global IP address. The software opens the ports and waits for user connection. There is a login procedure required to gain access to the data. The software handles incoming login and generates the VRS data for the user according to user's approximate position.

For user side, a client software was written in the PocketPC environment. The user is equipped with either a single or dual frequency receiver capable of performing RTK positioning and is connected to a GPRS-enabled PDA (Pocket PC only). The PDA establishes the communications with the receiver through its serial port. The user can connect to the server of the control center through the Internet via GPRS at his location without any extra equipment. Figure 2 shows the architecture for linking the control center with the user receiver using GPRS.

4 Analysis of Internet-based VRS Performance for RTK Positioning

In order to validate the viability and performance of Internet-based VRS RTK positioning using above method, the first field test was carried out at different locations in Singapore in July and August 2002, based on a prototype reference station infrastructure (Singapore Integrated Multiple Reference Station Network (SIMRSN)). SIMRSN consists of five reference stations located at four corners and in the middle of Singapore with dual-frequency GPS receivers, as shown in Fig.3. Some of the features of the SIMRSN are described in Chen et al. 2000, or Hu et al. 2002.

Raw observations at each reference station are relayed to a control center using leased lines. The user is equipped with a RTK receiver and a Pocket PC with GPRS, as shown in Fig. 2. He uses the client software to log in the server of the control center through Internet via GPRS, and sends his approximate position to the SIMRSN control center. The control center software automatically receives the user’s approximate position and selects the nearest reference station to the user as the master reference station. The raw data from that reference station is then geometrically displaced and improved by applying the corrections of the network according to the approximate position of the user. This is formed as VRS data, then transmitted as RTCM or CMR messages to the user receiver through the serial port of Pocket PC via GPRS at a data rate of 9,600 bps. The receiver can then
use these messages just as it would be in the single reference station RTK approach.

Five test points with different distances to the master reference station (NYPC) were used for testing the new communication method, as shown in Fig. 3.

To graphically ascertain the distribution of the computed errors, histograms of the horizontal and vertical position accuracy are plotted in Fig. 5 and Fig. 6. These figures present the absolute as well as the cumulative frequency for the horizontal and vertical position accuracy (external accuracy). It can be seen that 99% horizontal (2D) position accuracy is below 3 centimeters, and 99% vertical position accuracy is below 6 centimeters.

Fig. 3 Singapore Integrated Multiple Reference Station Network (SIMRSN) and Internet-based VRS RTK test stations

4.1 Results at NTU Test Site

The test station identified as NTU is taken as the first example, 19.1 kilometers away from the master reference station NYPC. A Leica SR530 dual-frequency receiver was used together with an AT502 antenna. This test was conducted during the time span 00:49-07:29 UTC (08:49-15:29 local time) on 31 July 2002 (DOY 212, GPS week 1177) at the rooftop of the N1 building located at Nanyang Technological University. The real-time positions during the test were continuously recorded by logging NMEA data output by one port of receiver. The surveying mode used was RTK and log data. Hence, the ground truth position could be computed from the logged long time raw data. Fig. 4 shows a representative sample of the differences (“true errors”) in the northing, easting and height components in the local datum SVY95 to the ground truth position. The standard deviations (inner accuracy) of castig, northing and height components are 0.012m, 0.010m and 0.048m respectively. As expected, the standard deviation in horizontal position is a factor of 2 to 3 better than in height. The reason for significant offset in the height component is caused by the residual tropospheric delay.

Besides the accuracy, a crucial factor in the operational use of RTK GPS is the speed with which it can initialize (i.e. solve for integer ambiguities). This is expressed as TTF (Time to Fix) or TTFA (Time To Fix Ambiguities) value (Edwards et al., 1999; Wübbena et al. 2001). The TTF of RTK refers to the observation time it requires to resolve integer ambiguities in real-time after re-initialization. A program was written to monitor the status of the receiver’s positioning mode. Five seconds after each successful fixing – or 3 minutes at the latest, if no fixing occurred – the receiver’s RTK engine was completely reset. Hence, an analysis of TTF was possible. For the NTU station, sample time series of the TTF for RTK vs. number of satellites and HDOP value is given in Fig. 7. As can be seen, excluding those requiring more than 3 minutes to initialize, the TTF for VRS RTK was within one minute at most of time during the test, with the average of 40 seconds.

Fig. 4 “true errors” in the northing, easting and height for NTU station

Fig. 5 Analysis of Horizontal (2D) position accuracy for NTU station

Fig. 6 Analysis of Vertical position accuracy for NTU station
4.2 Other Test Station Results

The accuracy and initialization time were analyzed at all the test stations. In the previous section, the results of the NTU test station were discussed. In this section, the five test stations were compared to evaluate the performance of the Internet-based VRS RTK positioning at different locations in Singapore, as summarized in Tab. 1. Note that similar accuracy and initialization time can be achieved during these tests. The worst performance was recorded at ECP station, about 3 kilometers outside the network, and 12.3 kilometers away from the master reference station. ECP test station uses Trimble 5700 dual-frequency receiver and Zephyr geodetic antenna, was conducted on 8 August 2002 (DOY 220, GPS week 1178), approximately between 02:48 and 07:13 UTC (10:48-15:13 local time). The position of the test station has been determined approximately 4636 times, providing a large test sample with independent ambiguity resolutions during the test period.

<table>
<thead>
<tr>
<th>Test Station</th>
<th>Distance from Master Reference Station (km)</th>
<th>Standard Deviation (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Northing</td>
<td>Easting</td>
</tr>
<tr>
<td>NTU</td>
<td>19.1</td>
<td>0.010</td>
</tr>
<tr>
<td>CCK</td>
<td>11.1</td>
<td>0.016</td>
</tr>
<tr>
<td>TOR</td>
<td>9.7</td>
<td>0.014</td>
</tr>
<tr>
<td>SEL</td>
<td>6.3</td>
<td>0.022</td>
</tr>
<tr>
<td>ECP</td>
<td>12.3</td>
<td>0.012</td>
</tr>
</tbody>
</table>

The horizontal position scatter and the true position of the ECP station is displayed in Fig. 8. The standard deviation for the northing, easting and height components are 12 millimeters, 16 millimeters and 51 millimeters, respectively. Fig. 9 and Fig. 10 show the histograms of the horizontal and vertical position accuracy for ECP station. These figures present that 91% horizontal (2D) position accuracy is below 3 centimeters, and 92% vertical position accuracy is below 6 centimeters.
Acknowledgements

The authors would like to thank Dr. Chen Xiaoming and Mr. Teo Swee Tiong who help to develop the software used in SIMRSN. Also thanks to Prof. Chris Rizos and his research group SNAP of University of New South Wales for their technical support.

References


5 Concluding Remarks

Precise RTK positioning over longer distances requires a network of GPS reference stations. An Internet-based VRS RTK positioning infrastructure via GPRS has been developed and introduced in Singapore by Nanyang Technological University, Singapore in collaboration with the University of New South Wales, Australia. Using the active multiple reference station network in Singapore, field tests at different locations in Singapore confirmed that better than 4 centimeters horizontal accuracy and 1-6 centimeters vertical accuracy can be achieved, with initialization time less than 3 minutes. It is worthy to be noted that it is difficult to do standard RTK positioning in the tropics where turbulent atmospheric conditions result in rapid changes in ionospheric and tropospheric errors even over short distances.

Internet-based VRS RTK positioning system is a real-time, centimeter level service for construction, rapid surveying, and GIS. Some navigation applications like automatic parking and ship docking also require centimeter level accuracy. It can also easily be used in various related application areas including Intelligent Transportation System (ITS), emergency system, Automatic Vehicle Location (AVL), and personal navigation system. With better wireless communications being introduced, VRS RTK positioning will see more usage soon.

Concluding Remarks

Precise RTK positioning over longer distances requires a network of GPS reference stations. An Internet-based VRS RTK positioning infrastructure via GPRS has been developed and introduced in Singapore by Nanyang Technological University, Singapore in collaboration with the University of New South Wales, Australia. Using the active multiple reference station network in Singapore, field tests at different locations in Singapore confirmed that better than 4 centimeters horizontal accuracy and 1-6 centimeters vertical accuracy can be achieved, with initialization time less than 3 minutes. It is worthy to be noted that it is difficult to do standard RTK positioning in the tropics where turbulent atmospheric conditions result in rapid changes in ionospheric and tropospheric errors even over short distances.

Internet-based VRS RTK positioning system is a real-time, centimeter level service for construction, rapid surveying, and GIS. Some navigation applications like automatic parking and ship docking also require centimeter level accuracy. It can also easily be used in various related application areas including Intelligent Transportation System (ITS), emergency system, Automatic Vehicle Location (AVL), and personal navigation system. With better wireless communications being introduced, VRS RTK positioning will see more usage soon.

5 Concluding Remarks

Precise RTK positioning over longer distances requires a network of GPS reference stations. An Internet-based VRS RTK positioning infrastructure via GPRS has been developed and introduced in Singapore by Nanyang Technological University, Singapore in collaboration with the University of New South Wales, Australia. Using the active multiple reference station network in Singapore, field tests at different locations in Singapore confirmed that better than 4 centimeters horizontal accuracy and 1-6 centimeters vertical accuracy can be achieved, with initialization time less than 3 minutes. It is worthy to be noted that it is difficult to do standard RTK positioning in the tropics where turbulent atmospheric conditions result in rapid changes in ionospheric and tropospheric errors even over short distances.

Internet-based VRS RTK positioning system is a real-time, centimeter level service for construction, rapid surveying, and GIS. Some navigation applications like automatic parking and ship docking also require centimeter level accuracy. It can also easily be used in various related application areas including Intelligent Transportation System (ITS), emergency system, Automatic Vehicle Location (AVL), and personal navigation system. With better wireless communications being introduced, VRS RTK positioning will see more usage soon.


Discussion Forum

“Experts Forum” is a regular column in this Journal featuring discussions on recent advances in global satellite positioning systems and their applications. Experts in various fields are welcome to contribute a short article to briefly describe their research directions and current activities, present recent results or identify remaining problems, freely expressing ideas and visions for future development. In this issue, Drs Tom Yunck and Cinzia Zuffada of Jet Propulsion Laboratory (JPL), Drs Peter Schwintzer and Christoph Reigber of GeoForschungsZentrum Postdam, and Dr Elizabeth Essex of La Trobe University, will review the scientific applications of GPS technology in earth sciences. The topics include atmospheric radio occultation, ocean remote sensing with GPS, contributions of GPS to global gravity field recovery, and space and ground based GPS ionospheric sensing.

The column is coordinated by Dr Yanming Feng of Queensland University of Technology, who appreciates your contribution to this column, along with your comments or ideas for topics for future issues (y.feng@qut.edu.au).
Abstract. A general criterion for integer ambiguity searching is derived in this paper. The criterion takes into account not only the residuals caused by ambiguity parameter changing, but also the residuals caused by coordinates changing through ambiguity fixing. The search can be carried out in a coordinate domain, in an ambiguity domain or in both domains. The three searching scenarios are theoretically equivalent. The optimality and uniqueness properties of the proposed criterion are also discussed. A numerical explanation of the general criterion is outlined. The theoretical relationship between the general criterion and the commonly used least squares ambiguity search (LSAS) criterion is derived in an equivalent case in detail. It shows that the LSAS criterion is just one of the terms of the equivalent criterion. Numerical examples are given to illustrate the behaviour of the two components of the equivalent criterion.

Key words: Integer Ambiguity Searching Criterion

1 Introduction

It is well-known that the ambiguity resolution is a key problem which has to be solved in GPS static and kinematic precise positioning. Some well-derived ambiguity fixing and searching algorithms have been published during the last decade. These methods can be generally classified as four types. The first type includes Remondi's static initialisation approach (Remondi 1984; Hofmann-Wellenhof et al. 1997; Wang et al. 1988), which requires a static survey time to solve the ambiguity unknowns after any complete loss of lock. Normally, the results are good enough to take a round up ambiguity fixing. The second type includes the so-called phase-code combined methods (Han & Rizos 1995, 1997; Sjoeberg 1998, 1999); the phase and code have to be used in the derivation as if they have the same precision, and in case of anti-spoofing (AS), the C/A code has to be used. A search process is still needed in this case. The third type is the so-called ambiguity function method (Remondi 1984; Hofmann-Wellenhof et al. 1997); its search domain is a geometric one. The fourth type includes approaches, their search domain is only in domain of ambiguity, including some optimal algorithms to reduce the search area and to accelerate the search process (Euler & Landau 1992; Teunissen 1995; Leick 1995; Han & Rizos 1995, 1997). Because of the statistic character of validation criteria, sometimes no valid result is obtained at the end of the search processes.

The effort to develop the KSGSoft (Kinematic/Static GPS Software) at the GeoForschungsZentrum (GFZ) Potsdam began at the beginning of 1994 due to the requirement of kinematic GPS positioning in aerogravimetry applications (Xu et al. 1998, 1999). An optimal ambiguity resolution method is needed in order to implement it into the software; however, selecting the published algorithms has turned out to be a difficult task. This has led to the independent development of this so-called integer ambiguity search method (cf. Xu 2003). It turns out to be a very promising algorithm; the search domain could be in the domain of coordinate or ambiguity or both, and it is reliable and fast. Using this general criterion, an optimal ambiguity vector can be searched for and found out. The searched result is the optimal one under the least squares principle and integer ambiguity property.

The theoretical background of this method is the well-known conditional least squares adjustment and will be outlined below in the section 2. The well-known least squares ambiguity search (LSAS) criterion is derived in section 3. An analogue derivation of using coordinate condition is outlined in section 4. A general criterion is presented in section 5. Properties of the general criterion are discussed in section 6. The relationship between the general criterion and the least squares ambiguity search criterion is derived in an equivalent case in section 7.
Numerical examples are given in section 8. Conclusions and comments are given in the last section.

2 Conditional Least Squares Adjustment

The principle of least squares adjustment with condition equations can be summarised as below (Cui et al. 1982; Leick 1995; Gotthardt 1978; Xu 2003):

1). Linearised observation equation system can be represented by:

\[ V = L - AX, \quad P \] (1)

where

\( L \): observation vector of dimension \( m \),
\( A \): coefficient matrix of dimension \( m \times n \),
\( X \): unknown vector of dimension \( n \),
\( V \): residual vector of dimension \( m \),
\( n \): number of unknowns,
\( m \): number of observations,
\( P \): symmetric and quadratic weight matrix of dimension \( m \times m \).

2). The condition equation system can be written as:

\[ CX - W = 0 \] (2)

where

\( C \): coefficient matrix of dimension \( r \times n \),
\( W \): constant vector of dimension \( r \),
\( r \): number of conditions.

3). The least squares criterion for solving the observation equations with condition equations is well-known as:

\[ V^T PV = \min \] (3)

where

\( V^T \): the transpose of the related vector \( V \).

4). The solution of the conditional problem (1) and (2) under the least squares principle (3) is then:

\[ X = (A^T PA)^{-1}(A^T PL) - (A^T PA)^{-1}C^T K \] (4)

and

\[ K = (CQc)^{-1}(CQW_j - W) \] (5)

where \( A^T, C^T \) are the transpose matrices of \( A, C \), the superscript \( ^{-1} \) is an inversion operator, \( Q = (A^T PA)^{-1} \), \( K \) is a gain vector (of dimension \( r \)), index \( c \) is used to denote the variables related to the conditional solution, and \( W_j = A^T PL \).

5). The accuracies of the solutions are then:

\[ p[i] = s_d \sqrt{Q_{c[i][i]}} \] (6)

where \( i \) is the element index of a vector or a matrix, \( \sqrt{()} \) is the square root operator, \( s_d \) is the standard deviation (or sigma) of unit weight, \( p[i] \) is the \( i \)-th element of the precision vector, \( Q_{c[i][i]} \) is the \( i \)-th diagonal element of the quadratic matrix \( Q_c \), and

\[ Q_j = Q - QC^T QjCQ \] (7)

\[ Q_2 = (CQC^j)^{-1} \] (8)

\[ s_d = \sqrt{(V^T PV)/(m-n)} \] if \( m > n \) (9)

6). For recursive convenience, \( (V^T PV)_c \) can be calculated by using:

\[ (V^T PV)_c = L^T PL - (A^T PL)^T X_0 - WTK \] (10)

Above are the complete formulas of conditional least squares adjustment. The application of such an algorithm for the purpose of integer ambiguity search will be further discussed in later sections.

3 Integer Ambiguity Search in Ambiguity Domain

GPS observation equations can be represented with (1). Considering the case without conditions (2), i.e., \( C = 0 \) and \( W = 0 \), the above equations are the same as the results of normal least squares adjustment. So the least squares solution of (1) is

\[ X_0 = Q(A^T PL) = QW_j \] (11)

and

\[ (V^T PV)_0 = L^T PL - (A^T PL)^T X_0 \] (12)

\[ s_d = \sqrt{(V^T PV)/(m-n)} \] if \( m > n \) (13)

\[ p[i] = s_d \sqrt{Q_{c[i][i]}} \] (14)

Where index \( _0 \) is used for convenience to denote the variables related to the normal least squares solution without conditions. \( X_0 \) is the complete unknown vector including coordinates and ambiguities and is called a float solution later on. Solution \( X_0 \) is the optimal one under least squares principle. However, because of the observation and model errors as well as method limitations, float solution \( X_0 \) may not be exactly the right one, e.g. the ambiguity parameters are real numbers and do not fit to the integer property. Therefore one sometimes needs to search for a solution, say \( X \), which not only fulfils some special conditions, but also meanwhile keeps the deviation of the solution as small as possible (minimum). This can be represented by

\[ V_x^T PV_x = \min \] (15)

In (15) the \( V_x \) is the residuals vector in case of solution \( X \). For simplification, let:
Simplifying (19), one gets:

\[ X = \begin{pmatrix} Y \\ N \end{pmatrix}, \quad Q = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \]

where \( Y \) is the coordinate vector, \( N \) is the ambiguity vector (generally, a real vector). To use the conditional adjustment algorithm for integer ambiguity searching in ambiguity domain, the condition shall be selected as \( N = W \), here \( W \) is, of course, an integer vector. Generally, letting \( C = (0, E) \), then condition (2) turns out to be:

\[ N = W \] (17)

Using definitions of \( C \) and \( Q \), one has:

\[ CQ = \begin{pmatrix} Q_{21} \\ Q_{22} \end{pmatrix} \]

\[ CQC^T = Q_{22} \]

The float solution is denoted as

\[ X_0 = \begin{pmatrix} Y_0 \\ N_0 \end{pmatrix} = \begin{pmatrix} Q_{11}W_{11} + Q_{12}W_{12} \\ Q_{21}W_{11} + Q_{22}W_{12} \end{pmatrix} \]

where \( X_0 \) is the solution of (1) without condition (17). The gain vector \( KN \) can be computed by:

\[ K_N = (Q_{22})^{-1}(CQW_1 - W) = (Q_{22})^{-1}(N_0 - W) \] (18)

So under the condition (17), the conditional least squares solution (4) can be written as:

\[ X_c = \begin{pmatrix} Y_c \\ N_c \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} W_{11} \\ W_{12} - K_N \end{pmatrix} \]

\[ = \begin{pmatrix} Y_0 \\ N_0 \end{pmatrix} - \begin{pmatrix} Q_{12} \\ Q_{22} \end{pmatrix} K_N \] (19)

Simplifying (19), one gets:

\[ Y_c = Y_0 - Q_{12}K_N \] (20)

and

\[ N_c = N_0 - Q_{22}K_N = N_0 - Q_{22}(Q_{22})^{-1}(N_0 - W) = W \] (21)

The precision computing formulas under condition (17) can be derived as below:

\[ Q_c = Q - QC^T (Q_{22})^{-1} CQ \]

\[ = \begin{pmatrix} Q_{11} - Q_{12}(Q_{22})^{-1}Q_{21} & 0 \\ 0 & 0 \end{pmatrix} \] (22)

\[ (V^TPV)_c = (V^TPV)_0 + (N_0 - W)^T (Q_{22})^{-1} (N_0 - W) \] (23)

where \((V^TPV)_0\) is the value obtained without condition (17). The second term on the right-hand side of (23) is the often-used least squares ambiguity search criterion, (cf. e.g. Teunissen 1995; Euler & Landau 1992; Hofmann-Wellenhof et al. 1997), which can be expressed as

\[ \delta(dN) = (N_0 - N)^T (Q_{22})^{-1} (N_0 - N) \] (24)

It indicates that any ambiguity fixing will cause an enlargement of the standard deviation. However, one may also notice that here only the enlargement of the standard deviation caused by ambiguity parameter changing has been considered. Any ambiguity fixing will lead to a related coordinate changing (cf. (20)). Furthermore, the condition (17) does not really exist. Ambiguities are integers, however, they are unknowns. The formula to compute the accuracy vector of the ambiguity does not exist too, because the ambiguity condition is considered exactly known in conditional adjustment (cf. Xu 2003).

### 4 Standard Deviation Enlargement Caused by Coordinate Changing

Analogous to above discussion, the condition could be selected as \( Y = W \), here \( W \) is a coordinate vector. Generally, letting \( C = (E, 0) \), where \( E \) is an identity matrix with dimension of \( r \times r \), \( C \) has dimension \( r \times n \), condition (2) turns out to be:

\[ Y = W \] (25)

Using definitions of \( C \) and \( Q \), one has:

\[ CQ = \begin{pmatrix} Q_{11} \\ Q_{12} \end{pmatrix} \]

\[ CQC^T = Q_{11} \]

The float solution is denoted by

\[ X_0 = \begin{pmatrix} Y_0 \\ N_0 \end{pmatrix} = \begin{pmatrix} Q_{11}W_{11} + Q_{12}W_{12} \\ Q_{21}W_{11} + Q_{22}W_{12} \end{pmatrix} \]

where \( X_0 \) is the solution of (1) without condition (25). The gain vector \( K_Y \) can be computed by using (5):

\[ K_Y = (Q_{11})^{-1}(CQW_1 - W) = (Q_{11})^{-1}(Y_0 - W) \] (26)

So under the condition (25), the conditional least squares solution (4) can be written as:

\[ X_c = \begin{pmatrix} Y_c \\ N_c \end{pmatrix} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix} \begin{pmatrix} W_{11} - K_Y \\ W_{12} \end{pmatrix} \]

\[ = \begin{pmatrix} Y_0 \\ N_0 \end{pmatrix} - \begin{pmatrix} Q_{11} \\ Q_{21} \end{pmatrix} K_Y \] (27)

Simplifying (27), one gets:

\[ Y_c = Y_0 - Q_{11}K_Y \]

\[ = Y_0 - Q_{11}(Q_{11})^{-1}(Y_0 - W) = W \] (28)
and
\[ N_c = N_0 - Q_{23} K_Y \] (29)
For any given constant coordinate vector \( W \), an ambiguity vector \( N_c \) can be found out (or computed). In such a case, \( N_c \) is a float vector and \( Y_1 \) is exactly the same as that given in condition (25). If the \( X_0 \) is a correct one, the computed \( N_c \) should be very close to an integer vector under the assumptions made at the beginning. The searched integer ambiguity vector is then \( \text{Fix}(N_c) \), where \( \text{Fix}(\cdot) \) is a round up function for rounding up a real number to its nearest integer number. A more detailed discussion on the use of the rounding function to the \( \text{Fix}() \) is discussed in next section.

\[ \delta_{c} = (Y_0 - Y)^T(Q_{11})^{-1}(Y_0 - Y) \] (32)
It is obvious that such an effect has to be taken into account in the ambiguity fixing. This will be further discussed in next section.

5 Integer Ambiguity Search in Coordinate and Ambiguity Domains

Even the to be fixed solution is an unknown vector, however, in order to see the enlargement of the standard deviation caused by the fixed solution, the condition could be selected as \( X = W \), here \( W \) consists of two subvectors (coordinate and ambiguity parameter related subvectors). And only the ambiguity parameter related subvector is an integer one. Letting \( C = E \), condition (2) is then:
\[ X = W \] (33)
One has:
\[ CQ = CQC^T = Q \]
Denote \( X_0 = QW_1 \); here \( X_0 \) is the solution of (1) without condition (33). The gain \( K \) can be computed by:
\[ K = Q^{-1}(CQW_1 - W) = Q^{-1}(X_0 - W) \] (34)
So under the condition (33), the conditional least squares solution (4) can be written as:
\[ X_c = X_0 - QK = X_0 - QQ^{-1}(X_0 - W) = W \] (35)
Precision computing formulas under condition (33) can be derived as below:
\[ Q_c = 0 \] (36)
\[ (V^TPV)_c = (V^TPV)_0 + (X_0 - W)^TQ^{-1}Q^{-1}(X_0 - W) \] (37)
where \( (V^TPV)_0 \) is the value obtained without condition (33).

The second term on the right side of (37) can be used as a general criterion for integer ambiguity search, i.e.:
\[ \delta = (X_0 - X)^TQ^{-1}(X_0 - X) \] (38)
It indicates the enlargement of the standard deviation caused by fixed solution \( X \). A minimum value of (38) is equivalent to a minimum value of \( (V^TPV)_c \). Therefore an optimal fixed solution has to be searched for so that (38) has the minimum value. To be noticed is that the minimum value of (38) is not a minimization process, but just a searching process to find out the optimal \( X \). (38) has obviously a more general form than the least squares ambiguity search criterion (24) does.

In all above three derivations, to be noticed is that in the precision vector the condition related elements are not defined. This is because in the conditional adjustment conditions are considered exactly known. However, in integer ambiguity searching, to be tested candidates (e.g. integer ambiguity) are indeed not exactly known or say, known with uncertainty (float solution with its precision). The uncertainty of the computed ambiguity and selected coordinate vectors (related to the searching in coordinate domain), and the uncertainty of the computed coordinate and selected ambiguity vectors (related to the searching in ambiguity domain), as well as the uncertainty of the selected vector in both domains (related to the searching in both domains) should be taken into account in any cases. Therefore (38) is a more reasonable criterion and should be used generally in ambiguity searching no matter in which domain the search will be made. Under such criterion, the deviation of the result vector \( X \) related to the float vector \( X_0 \) is homogenously considered. For computing the precision of the searched \( X \), the formulas of least squares adjustment shall be further used, and meanwhile the enlarged residuals shall be taken into account by
\[ p[1] = s_d \cdot \text{sqrt}(Q[1][1]) \] (39)
\[ s_d = \text{sqrt}((V^TPV)/(m-n)) \quad \text{if } m > n \] (40)
\[ (V^TPV)_c = (V^TPV)_0 + \delta \] (41)
In other words, the original $Q$ matrix and $(V^T PV)_0$ of the least squares problem (1) are further used. The $\delta$ has the function of enlarging the standard deviation. The formulas of (38), (39−41) are partly derived from the conditional adjustment, however, the formulas have nothing to do with the conditions. Searching for a minimum $\delta$ leads to a minimum of $s_2$ and therefore the best precision vector $p[i]$. The geometric explanation of here proposed integer ambiguity searching criterion is discussed in section 6.

The general criterion of (38) is used for all three searching scenarios, where $X_0$ is the float solution, $Q$ is the inversion of the complete normal matrix of (1). $X$ is the selected candidate vector in case of searching in both coordinate and ambiguity domains. In case of searching in coordinate domain, $X$ consists of the selected sub-vector of $Y_c$ in (28) and the computed sub-vector of $\text{Fix}(N_c)$ in (29), i.e.:

$$X = \begin{pmatrix} Y_c \\ \text{Fix}(N_c) \end{pmatrix}$$

(42)

The reason why the $\text{Fix}(N_c)$ is used here will be discussed theoretically in next section. In the case of searching in ambiguity domain, $X$ consists of the selected sub-vector of $N_c$ in (21) and the computed coordinate sub-vector $Y_c$ in (20), i.e.:

$$X = \begin{pmatrix} Y_c \\ N_c \end{pmatrix}$$

(43)

6 Properties of the General Criterion

1). Equivalence of the Three Searching Processes

To be emphasised is that the same searching criterion (38) and the same formulas of precision estimation (39−41) are used in the three integer ambiguity search scenarios. And the same normal equations of (1) is used to compute the vector $N_c$ using selected $Y_c$ or to compute the $Y_c$ using selected $N_c$ if necessary. The three searching processes indeed deal with the same problem, just as different ways of searching are used.

Suppose by searching in ambiguity domain, the vector $X = (Y_c, N_c)^T$ is found so that $\delta$ reaches the minimum, where $N_c$ is selected integer sub-vector and $Y_c$ is the computed one. In the case of searching in coordinate domain, if the selected coordinate sub-vector $Y$ is exactly the same as $Y_c$, then integer sub-vector $N$ obtained by computation should be exactly the same as $N_c$. Taking the computing errors into account, the computed $N$ could be a real vector, however, the errors must be very small and the rounding vector $\text{Fix}(N)$ must be the same as $N_c$. (This is also the reason why the rounding function is used for the computed vector $N_c$ in the case of search in coordinate domain). We see now the same results will be obtained theoretically in the both searching cases. Therefore, the searching methods in coordinate domain or in ambiguity domain are theoretically equivalent.

Suppose by searching in ambiguity domain, again, the vector $(Y_c, N_c)^T$ is obtained. And in the case of searching in both coordinate and ambiguity domains, a candidate vector $X = (Y, N)^T$ is selected so that $\delta$ reaches the minimum, where $N$ is selected integer sub-vector and $Y$ is selected coordinate vector. Because of the optimality and uniqueness properties of the vector $X$ in (38) (please refer to 2, which is discussed next), here selected $(Y, N)^T$ must be equal to $(Y_c, N_c)^T$. So the theoretical equivalency of the three searching processes is confirmed.

In practice, it could be difficult to have a selected $Y$ that exactly equals the computed $Y_c$ (computed by searching in ambiguity domain). However, it is always possible to get a $Y$ that is as close as required to $Y_c$ by selecting smaller search steps.

2). Optimality and Uniqueness Properties

The float solution $X_0$ is the optimal and unique solution of (1) under the least squares principle. Using the integer ambiguity search criterion (38), analogously, the searched vector $X$ is the optimal solution of (1) under the least squares principle and integer ambiguity properties. A minimum of $\delta$ in (38) will lead to a minimum of $(V^T PV)_0$ in (41). The uniqueness property is obvious. If $X_1$ and $X_2$ are such that $\delta(X_1) = \delta(X_2) = \min.$, or $\delta(X_1) – \delta(X_2) = 0$, then by using (38), one may assume that $X_1$ must be equal to $X_2$.

3). Geometric Explanation of the General Criterion

Geometrically, $\delta = (X_0 − X)^T Q^{-1}(X_0 − X)$ is the “distance” between the vector $X$ and float vector $X_0$. The distance contributed to enlarge the standard deviation $s_2$ (cf. (40)). Ambiguity searching is then the search for the vector, which own the integer ambiguity property and has the minimum distance to the float vector.

In the next section, the relationship between above proposed general criterion and the common used least squares ambiguity search criterion (derived in §3) will be discussed.

7 Relationship Between the Two Criteria

We are going to prove theoretically that LSAS criterion (24) is just one of the terms of an equivalent criterion of the general criterion (38) as follows.

The normal equation of (1) can be denoted by (use notation of (16)):
Setting \( Y \) into (46), one gets a normal equation related to the second block of unknowns:

\[
M_2N = B_2
\]

where

\[
M_2 = M_{22} - M_{21}(M_{11})^{-1}M_{12}
\]

(48)

\[
B_2 = W_{12} - M_{21}(M_{11})^{-1}W_{11}
\]

(49)

Similarly, from (46), one has

\[
N = (M_{22})^{-1}(W_{12} - M_{21}Y)
\]

(50)

Setting \( N \) into (45), one gets a normal equation related to the first block of unknowns:

\[
M_1Y = B_1
\]

where

\[
M_1 = M_{11} - M_{12}(M_{22})^{-1}M_{21}
\]

(52)

\[
B_1 = W_{11} - M_{12}(M_{22})^{-1}W_{12}
\]

(53)

Then the normal equation of (44) can be written by combining (51) and (47) as

\[
\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \begin{pmatrix} Y \\ N \end{pmatrix} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}
\]

(54)

(44) and (54) are two equivalent normal equations, therefore the integer ambiguity search using (44) or (54) are also equivalent. Using the notation (16), the normal equation of (1) is \( MX = W \), and the general criterion is (38). Because of \( M = Q^T \), (38) is the same as: \( (X_0 - X)^T M (X_0 - X) \). So for the normal equation of (54), the related general criterion (38) turns out to be (put the diagonal \( M \) into above formula!):

\[
\delta_1 = \begin{pmatrix} Y_0 - Y \\ N_0 - N \end{pmatrix}^T \begin{pmatrix} M_{11} & 0 \\ 0 & M_{22} \end{pmatrix} \begin{pmatrix} Y_0 - Y \\ N_0 - N \end{pmatrix}
\]

or

\[
\delta_1 = (Y_0 - Y)^T M_1(Y_0 - Y) + (N_0 - N)^T M_2(N_0 - N)
\]

(55)

It has to be emphasised that search criterion (55) is equivalent to the criterion (38), however, they are not identical, or generally, \( \delta \neq \delta_1 \). Furthermore, denote

\[
\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}^{-1} = \begin{pmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{pmatrix}
\]

(56)

where (cf. e.g. Cui et al. 1982; Leick 1995; Gotthardt 1978)

\[
Q_{11} = (M_{11} - M_{12}(M_{22})^{-1}M_{12})^{-1}
\]

(57)

\[
Q_{22} = (M_{22} - M_{21}(M_{11})^{-1}M_{12})^{-1}
\]

(58)

\[
Q_{12} = (M_{11})^{-1/2}(-M_{12}Q_{22})
\]

(59)

\[
Q_{21} = (M_{22})^{-1/2}(-M_{21}Q_{11})
\]

(60)

then after comparing (57) and (58) with (52) and (48) one has

\[
Q_{11} = (M_1)^{-1}, \quad Q_{22} = (M_2)^{-1}
\]

or

\[
M_1 = (Q_{11})^{-1}, \quad M_2 = (Q_{22})^{-1}
\]

(61)

Then (55) turns out to be

\[
\delta_1 = (Y_0 - Y)^T (Q_{11})^{-1/2}(Y_0 - Y) + (N_0 - N)^T (Q_{22})^{-1/2}(N_0 - N)
\]

(62)

Note that the second term on the right-hand side of (62) is exactly the same as the criterion of the least squares ambiguity search (24). In other words, the criterion of least squares ambiguity search is just one term of the equivalent criterion (62) (q.e.d.).

It should be emphasised that the consistency between the coordinate sub-vector \( Y \) and ambiguity sub-vector \( N \) is implicitly used by the proof. Therefore (62) is only valid if the \( Y \) and \( N \) are consistent each other. The first term on the right-hand side of (62) is the same as the (32), which indicates an enlargement of the standard deviation due to the coordinate change caused by ambiguity fixing.

Now, it is obvious that

1) only if one may lead from a minimum value of (24)

\[
\delta(dN) = (N_0 - N)^T (Q_{22})^{-1/2}(N_0 - N)
\]

(63)
to get a minimum value of (62)

\[
\delta_1 = (Y_0 - Y)^T (Q_{11})^{-1/2}(Y_0 - Y) + (N_0 - N)^T (Q_{22})^{-1/2}(N_0 - N)
\]

(64)

then the least squares ambiguity search is equivalent to the general method proposed in §5. However, such a generality does not exist. Therefore, the LSAS criterion is generally not equivalent to the criterion (64) (which is equivalent to the general criterion (38)). Furthermore, using (20) and (18) one has

\[
Y_0 - Y = Q_{12}(Q_{22})^{-1/2}(N_0 - N)
\]

(65)

Putting (65) into (64), one has

\[
\delta_1 = (N_0 - N)^T \cdot \{ (Q_{22})^{-1}[E + Q_{21}(Q_{11})^{-1}Q_{12}(Q_{22})^{-1}] \} \cdot (N_0 - N)
\]

(66)
One may see clearly now the differences between the two criteria (63) and (66).

2). If one may not lead from a minimum value of (63) to get a minimum value of (64), then the least squares ambiguity search may not find the optimal results in view point of the criterion (62). In this case, only criterion (62) reaches a minimum with a unique and optimal vector $X$.

3). The coordinate change due to the ambiguity fixing has not been taken into account in the least squares ambiguity search criterion.

A by-product of above derivation is that we have now a criterion (62) which is equivalent to the criterion (38). By computing the precision vector of (39)–(41), the $\delta$ has to be computed using (38), because the $\delta$ is not equal $\delta_1$ in general.

8 Numerical Examples of General Criterion and LSAS Criterion

Several numerical examples are given here to illustrate the behaviour of the two terms of the criterion. For convenience, we denote the first and second terms of the right-hand side of (62) as $\delta(dY)$ and $\delta(dN)$ respectively. $\delta_1 = \delta(dY) + \delta(dN)$ is the equivalent criterion of the general criterion and is denoted as $\delta$ (total). The term $\delta(dN)$ is the LSAS criterion. Of course, the search is made in the ambiguity domain. Analogues, the general criterion is also used for search in the ambiguity domain. The search area is determined by the precision vector of the float solution. All possible candidates are tested one by one, and the related $\delta_1$ are compared to each other to find out the minimum.

In the first example, precise orbits and dual-frequency GPS data of 15 April 1999 at station Brst (N 48.3805°, E 355.5034°) and Hers (N 50.8673°, E 0.3363°) are used. Session length is 4 hours. The total search candidate number is 1020. Results of the two sigma components are illustrated as 2-D graphics with the 1st axis of search number and the 2nd axis of sigma in Fig. 1. The red and blue lines represent $\delta(dY)$ and $\delta(dN)$, respectively. $\delta(dY)$ reaches the minimum at the search number 237, and $\delta(dN)$ at 769. $\delta$ (total) is plotted in Fig. 2, and it shows that the general criterion reaches the minimum at the search number 493. For more detail, a part of the results are listed in the Tab. 1.

<table>
<thead>
<tr>
<th>Search No.</th>
<th>$\delta(dN)$</th>
<th>$\delta(dY)$</th>
<th>$\delta$(total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>237</td>
<td>183.0937</td>
<td>97.8046</td>
<td>280.8984</td>
</tr>
<tr>
<td>493</td>
<td>181.7359</td>
<td>97.9494</td>
<td>279.6853</td>
</tr>
<tr>
<td>769</td>
<td>93.3593</td>
<td>315.2760</td>
<td>408.6353</td>
</tr>
<tr>
<td>771</td>
<td>96.0678</td>
<td>343.5736</td>
<td>439.6414</td>
</tr>
</tbody>
</table>

Fig. 1 Two components of the general ambiguity search criterion
Fig. 2 General ambiguity search criterion

Fig. 3 Example of general ambiguity search criterion
The $\delta(dN)$ reaches the second minimum at search No. 771. This example shows that the minimum of $\delta(dN)$ may not lead to the minimum of total sigma, because the related $\delta(dY)$ is large. If the sigma ratio criterion is used in this case, the LSAS method will reject the found minimum and explain that no significant ambiguity fixing can be made. However, because of the uniqueness principle of the general criterion, the search reaches the total minimum uniquely.

The second example is very similar to the first one. The sigmas of the search process are plotted in Figure 3, where $\delta(dY)$ is much smaller than $\delta(dN)$. $\delta(dN)$ reaches the minimum at the search number 5 and $\delta(dY)$ at 171. $\delta(total)$ reaches the minimum at the search number 129. The total 11 ambiguity parameters are fixed and listed in Table 2. Two ambiguity fixings have just one cycle difference at the 6th ambiguity parameter. The related coordinate solutions after the ambiguity fixings are listed in Table 3. The coordinate differences at component $x$ and $z$ are about 5 mm. Even the results are very similar, however, two criteria do give different results.

The float solution is the optimal solution of the GPS problem under the least squares minimum principle. Using the general criterion, the searched solution is the optimal solution under the least squares minimum principle and under the condition of integer ambiguities. However, the ambiguity searching criterion is just a statistic criterion. Statistic correctness does not guarantee correctness in all applications. Ambiguity fixing only takes into account the coordinate change due to the ambiguity fixing. Numerical examples shown that, a minimum $\delta(dN)$ may have a relatively large $\delta(dY)$, and therefore a minimum $\delta(dN)$ may not guarantee a minimum $\delta(total)$.

In the third example, real GPS data of 3 October 1997 at station Faim (N 38.5295°, E 328.8715°) and Flor (N 39.4493°, E 328.8715°) are used. The sigmas of the search process are listed in Table 4. Both $\delta(dN)$ and $\delta(total)$ reach the minimum at the search number 5. This indicates that the LSAS criterion may sometimes reach the same result as that of the equivalent criterion being used.

### 9 Conclusions and Comments

1). Conclusions

A general criterion of integer ambiguity search is proposed in this paper. The search can be carried out in a coordinate domain, in an ambiguity domain or in both domains. The criterion takes the both coordinate and ambiguity residuals into account. The equivalency of the three searching processes are proved theoretically. The searched result is optimal and unique under the least squares minimum principle and under the condition of integer ambiguities. The criterion has a clear numerical explanation. The theoretical relationship between the general criterion and the common used least squares ambiguity search (LSAS) criterion is derived in detail. It shows that the LSAS criterion is just one of the terms of the equivalent criterion of the general criterion (does not take into account the coordinate change due to the ambiguity fixing). Numerical examples shown that, a minimum $\delta(dN)$ may have a relatively large $\delta(dY)$, and therefore a minimum $\delta(dN)$ may not guarantee a minimum $\delta(total)$.

2). Comments

The float solution is the optimal solution of the GPS problem under the least squares minimum principle. Using the general criterion, the searched solution is the optimal solution under the least squares minimum principle and under the condition of integer ambiguities. However, the ambiguity searching criterion is just a statistic criterion. Statistic correctness does not guarantee correctness in all applications. Ambiguity fixing only makes sense when the GPS observables are good enough and the data processing models are accurate enough.

### References


Gotthardt, E. (1978): Einfuehrung in die Ausgleichungsrechnung, Herbert Wichmann Verlag Karlsruhe


Experts Forum

“Experts Forum” is a regular column in this Journal featuring discussions on recent advances in global satellite positioning systems and their applications. Experts in various fields are welcome to contribute a short article to briefly describe their research directions and current activities, present recent results or identify remaining problems, freely expressing ideas and visions for future development. In this issue, Drs Tom Yunck and Cinzia Zuffada of Jet Propulsion Laboratory (JPL), Drs Peter Schwintzer and Christoph Reigber of GeoForschungsZentrum Postdam, and Dr Elizabeth Essex of La Trobe University, will review the scientific applications of GPS technology in earth sciences. The topics include atmospheric radio occultation, ocean remote sensing with GPS, contributions of GPS to global gravity field recovery, and space and ground based GPS ionospheric sensing.

The column is coordinated by Dr Yanming Feng of Queensland University of Technology, who appreciates your contribution to this column, along with your comments or ideas for topics for future issues (y.feng@qut.edu.au).
Multiple Reference Station Approach: Overview and Current Research

G. Lachapelle And P. Alves
University Of Calgary
Received: 2 December 2002 / Accepted: 9 December 2002

Introduction

The availability of high precision reliable three-dimensional positioning tools has opened the door for many applications. GPS is being used by many industries for positioning and location-based services, many of which require a high level of precision. These applications include 3-D marine navigation in constricted waterways, aircraft positioning and guidance, land navigation, and water vapor estimation to name a few. Inland marine transportation, for example, is limited by the height positioning accuracy. If the accuracy is increased, then more cargo can be carried while ensuring clearance under the keel. GPS technological advances will assist these industries to be more competitive and cost effective, as well as as creating innovative location-based services.

The accuracy of the system dictates the range of its use; in other words, the more accurately users can determine their positions, the more applications that will utilize that accuracy level. The highest accuracy positions are obtained only after the estimation of the carrier phase ambiguities. The ability to resolve these biases is limited primarily by the correlated errors due to the troposphere, the ionosphere, and satellite orbits. The use of differential techniques, whereby a station with known coordinates is used as a reference station, has proven successful in the removal or reduction of these errors. The largest drawback of this approach, however, is the relatively short distance that must be maintained between the two receivers.

In an attempt to overcome these shortcomings, methods based on the use of networks of GPS reference stations are being developed to further reduce the effect of correlated errors and thus improve positioning accuracy. These networks of reference stations can be used to measure the correlated errors for a region and predict their effects (through advanced interpolation methods) spatially and temporally within the network. This process (Network RTK) can reduce the effects of the correlated errors much more than the single reference station approach allowing for reference stations to be spaced much further apart to cover a larger service area than the traditional approach and still maintain the same level of performance.

Network RTK is comprised of three main processes: network correction computation, correction interpolation, and correction transmission. The network correction computation uses the network reference stations to precisely estimate the differential correlated errors for the region. This is usually accomplished using carrier phase observations with fixed ambiguities between the network stations. Thus ambiguity fixing between these stations is a major part of this process. The second process interpolates these network corrections to determine the effects of the correlated errors at the user’s position. The third is the generation of virtual reference station (VRS) measurements to relay the corrections to the rover receiver for use with standard RTK software.
Measured Network ErrorS

The first step of Network RTK is to measure the errors at the reference stations. In most cases, the errors are measured as the difference between the carrier phase observations (with fixed ambiguities) and the range, which is calculated using the known reference station coordinates. These errors can be measured in terms of the raw L1 and L2 carrier phase observations or a linear combination of the L1 and L2 observations. Linear combinations are used to isolate the various error sources to take advantage of their unique characteristics.

Interpolation of the Measured Network ErrorS

The interpolation of the correlated errors to the location of the user receiver assumes a stochastic and physical relationship between the errors. For example, all interpolation methods result in the closest reference stations having the most influence over the predicted value because a close reference station is more likely to experience the same error conditions as the rover receiver than one which is further away.

Raquet [1998] proposed a method of interpolating the observed errors between the reference stations to a user’s position anywhere in the network. In this method an exterior process determines the carrier-phase integer ambiguities between the reference stations. These ambiguities are then used to estimate the differential errors between the reference stations. The measured errors are interpolated to a user in the network using a linear least-squares prediction method. Covariance functions that represent the stochastic behavior of the errors must be determined at the outset. This method has been implemented in a functional real-time system and provides good improvement in post-mission and real-time [Cannon et al 2001a, 2001b; Fortes et al 2000a, 2000b, 2001; Alves et al 2001; Raquet et al 1998; Raquet 1998; Zhang1999a, Zhang & Lachapelle 2001].

Wanninger [1999], Vollath et al [2000a] and Wübbena et al [2001a] discuss a slightly different interpolation scheme where only the surrounding three stations are used to predict the corrections at the user. In this simpler model, a plane is fit to the error estimates at the three surrounding stations. This plane represents the two-dimensional differential errors within the three-station triangle. This method has also proven to provide good positioning results under a quiet ionosphere and with a relatively high reference station density [Wanninger 1999; Wübbena et al 2001a, 2001b; Vollath et al 2000a, 2000b, 2001, 2002].

Virtual Reference Station Calculation

Once the corrections for the rover are determined, they still need to be transmitted to the user receiver in a suitable format. The traditional single baseline approach has a large impact on this process because most off-the-shelf receivers do not yet have a method of accepting network corrections. To compensate, many network RTK systems create virtual reference stations. A virtual reference station is a collection of corrected data from one reference receiver (in the network) that has been corrected for a local area within the network. This data is usually geometrically translated to be close to the region for which it is corrected. The rover receiver can then accept the virtual reference station data as a single reference station. This process is described in Fotopoulos [2000].

In general, the virtual reference station approach creates a “reference station” for use with standard off-the-shelf receivers that do not have the capability of accepting network corrections. There are many downfalls to this approach: The rover receiver will interpret the virtual reference station as a single reference station, which may cause the rover to use a processing scheme that is not optimal [Townsend et al 2000]. In most cases, the rover receiver will optimize the processing scheme based on the distance between its position and that of the reference station. In the case of a virtual reference station the position of the reference station is arbitrary because it is based on a network of stations. A solution would be to have the service provider ensure that the VRS is an appropriate distance away from the user to optimize the processing scheme but this is not always possible with multiple users. This requires the service provider to know the approximate position of the user. In this case, the rover would be required to send its position, via NMEA messages, to the processing control center to insure that the interpolation is calculated for the correct position and to position the VRS appropriately. The complex two-way communications network required is another drawback [Euler et al 2001]. A final downfall of this method is that it does not comply with the RTCM standard because the standard does not allow for the reference station data to be corrected for atmospheric or orbit errors [Townsend et al 2000].

Future of Multiple Reference Station Approaches

Current recommendations for the future of the virtual reference station approach deals mainly with standardization of network RTK messages. Once standard RTCM network corrections are selected then network corrections can be feed directly into the rover receiver without the need for a virtual reference station. Townsend [2000] proposes a grid based correction scheme where
the corrections for various points on an irregular grid are passed to the rover receiver. The rover receiver can then use any interpolation scheme to calculate and apply the corrections. The grid points could contain only reference stations or reference stations and predicted errors, which were determined through interpolation.

Euler et al. [2001] proposes a similar scheme where the corrections for a master reference station and the coordinates for a master reference station are given along with correction and coordinate differences relative to the master station. In this scheme the rover receiver has the option of interpolating the corrections for its location or simply reconstructing the observations for a single reference station. This gives receiver manufacturers the liberty to implement any interpolation scheme they feel is best.

Although multiple reference station RTK methods have proven to be effective in test networks, operational deployment remains complex and high ionospheric activity levels have limited their advantages during the past few years during which the methods have been tested. Serious reliability issues remain. However the introduction of a 2nd and 3rd authorized civilian activity levels have limited their advantages during the past few years during which the methods have been tested. Serious reliability issues remain. However the introduction of a 2nd and 3rd authorized civilian.

Bibliography


Virtual Reference Station Systems

Herbert Landau, Ulrich Vollath, Xiaoming Chen
Trimble Terrasat GmbH

Biography

Dr. Herbert Landau, is Managing Director of Trimble Terrasat and Director of GPS Algorithms and Infrastructure software in Trimble’s G&E Division. He received his Ph.D. in Geodesy from the University FAF Munich, Germany in 1988. He has many years of experience in GPS and has been involved in a large variety of GPS and GLONASS developments for high precision positioning systems and applications.

Dr. Ulrich Vollath, received a Ph.D. in Computer Science from the Munich University of Technology (TUM) in 1993. He is Senior Manager of the department of kinematic positioning and real-time systems of Trimble Terrasat.

Dr. Xiaoming Chen, is a software engineer at Trimble Terrasat. He holds a PhD in Geodesy from Wuhan Technical University of Surveying and Mapping.

Introduction

High accuracy Real-Time Kinematic Positioning with GPS is one of today’s most widely used surveying techniques. But, the effects of the ionosphere and troposphere, which create systematic errors in the raw data, restrict its use. In practice, these mean that the distance between a rover (mobile) receiver and its reference station has to be quite short in order to work efficiently.

In some countries GPS reference station networks exist, and provide data to individual users. For RTK, due to the need for short distances between reference and rover, the networks need to be very dense. Although of sufficient density for good DGPS, some national networks are just not dense enough to provide complete coverage for RTK. There are gaps in the coverage. The situation is worse during periods of high solar activity, such as in the first few years of the new Millennium, since these periods have extremely high atmospheric disturbance.

The use of a network of reference stations instead of a single reference station allows to model the systematic errors in the region and thus provides the possibility of an error reduction [1], [2], [3], [4], [5]. This allows a user not only to increase the distance at which the rover receiver is located from the reference, it also increases the reliability of the system and reduces the RTK initialization time. The concept can be used not only to set-up new networks, but also to improve the performance of old, established networks. The network error correction terms can be transmitted to the rover in two principle modes:

1. A Virtual Reference station mode as described below. This mode requires bi-directional communication. The basic advantage of this mode is that it makes use of existing RTCM and CMR standards implemented in all major geodetic rover receivers and thus is compatible with existing hardware.

2. A broadcast mode, in which the error corrections due to atmospheric and orbit effects are transmitted in a special format, which requires changes of rover receiver hardware or additional hardware to convert the non-standard format to a standard RTCM data stream before used by the rover.

In the following we will describe in detail the Virtual Reference Station idea first and then comment the current status of broadcast format implementations.

The Virtual Reference Station Concept

The “Virtual Reference Station” concept is based on having a network of GPS reference stations continuously
connected via data links to a control center. A computer at the control center continuously gathers the information from all receivers, and creates a living database of Regional Area Corrections.

These are used to create a Virtual Reference Station, situated only a few meters from where any rover is situated, together with the raw data, which would have come from it. The rover interprets and uses the data just as if it has come from a real reference station. The resulting performance improvement of RTK is dramatic.

The implementation of the VRS idea into a functional system solution follows the following principles. First we need a number of reference stations (at least three), which are connected to the network server via some communication links.

The GPS rover sends its approximate position to the control center that is running GPSNet. It does this by using a mobile phone data link, such as GSM, to send a standard NMEA position string called GGA. This format was chosen because it is available on most receivers.

The control center will accept the position, and responds by sending RTCM correction data to the rover. As soon as it is received, the rover will compute a high quality DGPS solution, and update its position. The rover then sends its new position to the control center.

The network server will now calculate new RTCM corrections so that they appear to be coming from a station right beside the rover. It sends them back out on the mobile phone data link (e.g. GSM). The DGPS solution is accurate to +/-1 meter, which is good enough to ensure that the atmospheric and ephemeris distortions, modeled for the entire reference station network, are applied correctly.

This technique of creating raw reference station data for a new, invisible, unoccupied station is what gives the concept its name, “The Virtual Reference Station Concept” [2], [3]. Using the technique, it is possible to perform highly improved RTK positioning within the entire station network.

**Network RTK With A Broadcast Format**

As pointed out above the VRS technique is only working properly with bi-directional communication like GSM/GPRS etc. For radio solutions a one-directional broadcast solution with a special format is required. Unfortunately there is currently no internationally standardized network format, which could be used for this. The RTCM committee is currently discussing and evaluating several proposals for a network broadcast format, but no receiver manufacturer has currently implemented such a format [6], [7]. The SAPOS committee in Germany had decided to standardize on a very simple linear description of error components also called area network corrections (FKP). This proposal is now one of the standard formats used in Germany in addition to the VRS technique. The FKP format is using the RTCM 2.3 message 59 description to come up with a special message including the linear correction parameters to approximate the error behavior in the neighborhood of a physical reference station. This implementation of a broadcast message has its limitations. It will serve as a good solution for a moving rover in the neighborhood of a station. However, if the rover moves too far from the station the errors will increase and the system will need to change the reference
station to a station nearby the actual rover position. Currently the use of the SAPOS FKP RTCM 59 message was only implemented by a few receiver manufacturers, i.e. it has not yet been accepted as a quasi standard (Trimble has implemented support for this message in the 5700 and 5800 rovers).

Server Software

Each reference station is equipped with a receiver, antenna, power supply and modem through which it communicates to the control center.

The computer at the control center, which runs a network server software like Trimble’s GPSNet is the nerve center of the concept. While connected to all the receivers in the network, it performs several major tasks including:

- Raw data import and quality check
- RINEX and compact RINEX data storage
- Antenna phase center corrections (relative and absolute models supported)
- Modeling and Estimation of systematic errors
- Generation of data to create a virtual position for the rover receiver
- Generation of an RTCM data stream for the virtual position
- Transmission of RTCM data to the rover in the field
- Generation of the SAPOS FKP broadcast network correction stream

The network server software also performs a continuous computation of the following parameters by analyzing code and carrier phase observations:

- Multipath errors
- Ionospheric errors
- Tropospheric errors
- Ephemeris errors
- Carrier phase ambiguities for L1 and L2

When performing this task the software makes use of the full network information, rather than using only a local subset of the reference station network.

Using the computed parameters, the server will recompute all GPS data, interpolating to match the position of the rover, which may be at any location within the reference station network. By doing that, the systematic errors for RTK are reduced considerably.

Error Interpolation

When interpolating the errors for the Virtual Reference Station within the server software the errors are interpolated from the residuals of the surrounding reference stations based on a special interpolation technique. In Trimble’s GPSNet we are using a weighted linear approximation approach and a least squares collocation approach. The interpolation technique allows to interpolate (user 1) but also extrapolate (user 2) (Fig. 4).

GPSNet Software Set-up

The GPSNet software is provided in three major modules:

1. The basic module GPSNet allows the data collection from reference stations, detection and correction of cycle slips, QA/QC analysis and raw data storage in RINEX format. It also provides a module for generation of RTCM messages to mobile users.
2. DGPSNet is an extension of the basic module GPSNet allowing the reduction of error sources like atmospheric and orbital effects and multipath. It is targeted for DGPS users and includes the generation of RTCM messages 1,2,3,9 supporting L1 C/A code differential positioning. The precision of these solutions ranges from sub-meter to 2-3 decimeters depending on the type of mobile receiver used.
3. RTKNet is the Trimble VRS solution for RTK users and is an extension to the basic GPSNet module. RTKNet provides the same kind of formats like DGPSNet plus the RTCM messages 18, 19,20,21 and 22, 23, 24 and the CMR/ CMR+ formats.

Typical Vrs Hardware Set-up

The control center is continuously communicating with the reference stations and receives raw data with an update rate of 1 Hz. It also controls the receivers.
Different methods can be used to transfer the data from the remote stations to the control center.

- Continuous analog or digital modem lines may be used. This method requires a modem at the reference station site and in the control center. The modem in the control center may be connected to the PC serial port directly using RS232, but if several reference stations are to be connected, a router such as CISCO 3640 can be used. The router will forward the data via Local Area Network to the control center computer. It is extendable with respect to the number of supported lines, and this allows an almost unlimited number of data lines to be available. The PC running GPSNet receives its data via IP protocol from the CISCO router. The remote stations are identified via IP port numbers.

- Frame Relay connections may be used. Although these are not always available in telephone networks, it may be the best transfer method to choose, especially over longer distances. In this case, the reference station requires a converter for the RS232 data stream. In the case where only Frame Relay is used, a more simple router such as CISCO 2500 can be used in the control center. In this configuration each remote station has its own IP address and the router at the center only translates from Frame Relay to LAN and vice versa.

- The data may be also transferred via the Internet using a DSL or other access. In that case the serial interface protocol of the receiver has to be converted to a TCP/IP protocol. This can be achieved by a local PC or a comserver.

In each case, the data can come directly from the receiver, or via GPSBase, Trimble’s autonomous reference station software.

The RTCM message is transferred to the rover via a mobile phone network such as GSM, or other digital mobile phone services like GPRS. An alternative communication medium is CDPD in the US.

A dial access server such as a CISCO AS 5300 handles the incoming calls. This allows all rovers to use the same phone number to connect to the system. The access server will provide an IP connection to the GPSNet software resulting in both its registration within the system, and the creation of RTCM data for that rover. The access server can be extended to handle increased volume of parallel incoming calls.

Typical Field Set-up Procedure

On a typical field session, the following set-up procedure is performed.

1. After starting the local receiver in real-time positioning mode, the user dials into the Virtual Reference Station Network service via a mobile phone capable of data transmission. This is normally done using one central phone number available for a whole state or country.
2. When the caller is successfully authenticated, the local receiver sends a navigation solution of its current position as a rough position estimate to the computing center.
3. After receiving the rough position estimate, the computing center creates a Virtual Reference Station at this location.
4. A continuous data stream of reference data generated for the Virtual Reference Station position is sent to the field user receiver. This can be done in RTCM or other real-time formats like CMR2.

Operating Reference Station Networks

The presented concepts are implemented in a number of reference station network installations since up to 4 years on a worldwide basis. Many experiences with operation and performance were made during this time. The following table gives an overview of some existing installations of Virtual Reference Station Networks by Trimble.

<table>
<thead>
<tr>
<th>Network</th>
<th>Location</th>
<th>Stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAPOS Bavaria</td>
<td>Germany</td>
<td>44</td>
</tr>
<tr>
<td>SAPOS Hessen</td>
<td>Germany</td>
<td>28</td>
</tr>
<tr>
<td>SAPOS BaWü</td>
<td>Germany</td>
<td>36</td>
</tr>
<tr>
<td>SAPOS Thüringen</td>
<td>Germany</td>
<td>24</td>
</tr>
<tr>
<td>SAPOS Sachsen</td>
<td>Germany</td>
<td>19</td>
</tr>
<tr>
<td>SAPOS NRW</td>
<td>Germany</td>
<td>34</td>
</tr>
</tbody>
</table>

Fig. 5 Typical VRS system set-up
ASCOS/Ruhrgas Germany 120
BKG Germany 20
Swissat Switzerland 23
Swiss Topo Switzerland 29
BEV Austria 5
LE34 Denmark 26
OC GIS Vlandeeren Belgium 40
SWEPOS Sweden 56
Tampere Finland 4
Lennen Puhelin Finland 8
Statens Kartverk Norway 6
Bysat Cz Czech Rep. 4
Shenzhen China 4
NGDS Japan 200
Jenoba Japan 200
NCDOT USA 5
MnDOT USA 6
QBR Queensland Australia 5
Bay Area Network USA 10
Colorado Network USA 10
Christchurch NZ. Nealand 6
Munich Test Network Germany 42

Sample Networks

In the following we are showing some examples of network installations in Europe and Asia. For example, Denmark is using a Trimble solution with Trimble reference station receivers and Trimble network server software GPSNet. RTCM corrections are transmitted to the users in the field via GSM.

Another Trimble network was set-up in Switzerland by the Swisstopo organization; again Trimble receivers are used and have proven to provide excellent performance since the start of the VRS system in 1999.

In the German SAPOS system the largest area is covered by the Trimble VRS networking solution. Currently six major states have decided to use the Trimble solution. Besides these public German organizations a private organization has started to provide a Germany-wide service. This company ASCOS is belonging to the Ruhrgas AG, the largest gas provider in Germany, and is currently using a VRS solution from Trimble to operate 120 reference stations.

Ascos/Ruhrgas is primarily targeting the utility management market and wants to serve all companies doing pipeline and cable GIS surveying. Their business plan is based on the investigation that we have more than four million kilometers of gas, drinking water, wastewater, phone of electricity lines, which require surveys. ASCOS estimates that approximately 3% of these lines have to be surveyed every year.

Currently two independent and competing organizations are operating a 200-station network with Trimble VRS software. The Japanese Geographical Survey Institute...
As a test network, a part around Munich, Germany of the BLVA network of the land surveying authorities in Bavaria has been used. This test-bed is continuously operated for GPSNet software testing and development. The network consists of 7 stations, each station has a dual-frequency GPS receiver and is permanently connected to the Trimble Terrasat office via leased data lines. The network configuration is shown in Fig. 10 including the inter-station baseline lengths.

The Trimble Network RTK software [Trimble, 2002] can be operated in either a VRS or Broadcast mode. The broadcast mode was used for the tests described below.

Four concurrent tests were run over a 40-hour period to evaluate the performance of RTK in single-baseline and network modes. Four Trimble 5700 receivers were connected to the same antenna at the Trimble Terrasat office at Höhenkirchen. Standard single-baseline data was fed from Munich and Toelz into two 5700 receivers, thus giving rise to 16km and 32km baselines, respectively. Network corrections were input to the other two 5700 receivers at Höhenkirchen. One set of corrections was derived from the entire network, while the second correction stream was created without the nearest network station - Munich.

**Test Results**

**Real-time Performance Test**

**Test Description**

![Fig. 11 Cumulative probability of ambiguity resolution for single-base and network modes over 16km and 32km baselines.](image)

Fig. 11 illustrates the cumulative time-to-initialize for the 4 different 5700 receivers. The results of network corrected data provide very noticeable benefits for the time-to-initialize statistics. The 16km baseline with network corrections exhibits the sharpest elbow in figure 15. In other words, the majority of initializations occur within a short period of time. The network-corrected 16km baseline results are comparable to those regularly achieved with single-baseline RTK on lines less than 10km where ionospheric biases are typically small. The 32km baseline with network corrections has the next-best
performance. The single-base RTK results for the 16 and
32km lines exhibit the worst initialization times.

Bibliography


Network RTK Research and Implementation - A Geodetic Perspective

C. Rizos
School of Surveying and Spatial Information Systems, The University of New South Wales, Sydney NSW 2052 AUSTRALIA
e-mail: mailto:c.rizos@unsw.edu.au

Received: 3 January 2003 / Accepted: 10 January 2003

Background: Why Complicate Matters?

The standard mode of precise differential positioning is for one reference receiver to be located at a base station whose coordinates are known, while the second receiver's coordinates are determined relative to this reference receiver. This is the principle underlying pseudo-range-based differential GPS (or DGPS for short) techniques. However, for high precision applications, the use of carrier phase data must be used, but comes at a cost in terms of overall system complexity because the measurements are ambiguous, requiring that ambiguity resolution algorithms be incorporated as an integral part of the data processing software. Such high accuracy techniques are the result of progressive R&D innovations, which have been subsequently implemented by the GPS manufacturers in their top-of-the-line “GPS surveying” products. Over the last decade or so several significant developments have resulted in this high accuracy performance also being available in real-time -- that is, in the field, immediately following the making of measurements, and after the data from the reference receiver has been transmitted to the (second) field receiver for processing via some sort of data communication links (e.g., VHF or UHF radio, cellular telephone, FM radio sub-carrier or satellite com link). Real-time precise positioning is even possible when the GPS receiver is in motion. These systems are commonly referred to as RTK systems (“real-time-kinematic”), and make feasible the use of GPS-RTK for many time-critical applications such as engineering surveying, GPS-guided earthworks/excavations, machine control and other high precision navigation applications.

The limitation of single base RTK is the distance between reference receiver and the rover receiver due to distance-dependent biases such as orbit error, and ionospheric and tropospheric signal refraction. This has restricted the inter-receiver distance to 10km or less. On the other hand, Wide Area Differential GPS (WADGPS) and the Wide Area Augmentation System (WAAS) use a network of master and monitor stations spread over a wide geographic area, and because the measurement biases can be modelled and corrected for, the positioning accuracy will be almost independent of the inter-receiver distance (or baseline length). However, these are pseudo-range based systems intended to deliver accuracies at a metre level. Continuously operating reference stations have been deployed globally to support very high accuracy geodetic applications for well over a decade. How can GPS surveying take advantage of such developments in geodesy and global navigation? The answer is to take advantage of multiple reference station networks, in such implementations as Network RTK.

Network RTK is a centimetre-accuracy, real-time, carrier phase-based positioning technique capable of operating over inter-receiver distances up to many tens of kilometres (the distance between a rover and the closest reference station receiver) with equivalent performance to current single base RTK systems (operating over much...
shorter baselines). The reference stations must be deployed in a dense enough pattern to model distance-dependent errors to such an accuracy that residual double-differenced carrier phase observable errors can be ignored in the context of rapid ambiguity resolution. Network RTK is therefore the logical outcome of the continuous search for a GPS positioning technique that challenges the current constraints of cm-accuracy, high productivity, carrier phase-based positioning.

Network-based Positioning: The Geodetic Perspective

All GPS-based positioning techniques operate under a set of constraints (Rizos, 2002). These constraints may be baseline length, attainable accuracy, assured reliability, geometrical strength, signal availability, time-to-solution, instrumentation, operational modes, and so on. GPS product designers must develop systems (comprising hardware, software and field procedures) that are optimised for a certain target user market, by addressing only those constraints that are crucial to the most common user scenarios. For example, current single base RTK systems are capable of high performance when measured in terms of such parameters as accuracy, time-to-solution (i.e. speed of ambiguity resolution after signal interruption), utility (due to the generation of real-time solutions), flexibility (being able to be used in static and kinematic applications), ease-of-use, and cost-effectiveness. As a result the sale of RTK systems is booming. However, the authors believe that the 10km baseline constraint will increasingly become an issue.

RTK GPS users will demand an infrastructure of base stations to support them, in much the same way that DGPS users have for many years been able to take advantage of free-to-air or fee-based differential services. However, it is generally unrealistic to deploy reference receivers across a country, or even just within a city, at such a density that all users are within 10km of a reference receiver transmitting RTK messages. Network RTK techniques use base station separations of several tens of kilometres, hence requiring fewer reference receivers. This significantly reduces the infrastructure investment required. The development of Network RTK can viewed from three distinct perspectives:

1. The evolution of the high productivity GPS Surveying technique in order to preserve single base RTK performance, but to permit much greater GPS inter-receiver distances. The change from single base to multi-base allows for the empirical modelling of the distance-dependent measurement biases. It is this modelling (and the transmission of ‘corrections’ for the normally unaccounted for biases) that overcomes the distance constraint, with no requirement for an upgrade to the user equipment software. The same GPS surveying user functionality is preserved.

2. The use of sparse networks of base stations is the basis of WADGPS and WAAS positioning techniques (Lachapelle et al., 2002). Data from the base station network are sent to a central computing facility, and empirical models of the distance-dependent biases are generated in the form of ‘corrections’ (which may be in the form of proprietary messages, or an industry standard RTCM or WAAS message type). These corrections are transmitted to user across a wide geographic area (most commonly via satellite communication links). However, because such Augmented GPS Navigation techniques use pseudo-range data, and the separation of the base stations is typically many hundreds to several thousands of kilometres, sub-decimetre level positioning accuracy is unattainable. The evolution to Network RTK would require a significant improvement in accuracy, through the use of carrier phase data and a much denser deployment of reference receivers.

3. GPS Geodesy has evolved since the early 1980s into a powerful, ultra precise positioning technique that is used for a range of applications, including the definition of the fundamental geodetic framework and the measurement for tectonic motion. GPS Geodesy uses a multi-receiver data processing methodology in which all measurement biases, no matter how small, are carefully accounted for in the functional and stochastic models of the double-differenced carrier phase observables. Continuously operating reference station (CORS) networks have been established around the world to support a range of geodetic applications. Positioning accuracy at the few parts-per-billion (ppb) are now routinely obtained, using sophisticated data processing algorithms in packages such as the Bernese software (Rothacher & Mervart, 1996). (One ppb is equivalent to 1mm relative accuracy over a baseline one thousand km in length.) Clearly GPS Geodesy could evolve into the Network RTK technique, if receivers were permitted to be in motion and data processing could be undertaken in real-time. Both of these are significant challenges. Furthermore, the Network RTK strategy could be used to densify high precision CORS networks for certain geodetic applications.

The author describes below several developments in GPS Geodesy that could be viewed as being predecessors to the development of the Network RTK concept. In fact, network-based positioning techniques have been an interest of geodesists for some time. During the past decade the International Association of Geodesy (IAG) has established several Special Study Groups (SSG) to research several topics concerned with permanent GPS networks. In 1999 the IAG established SSG1.179 “Wide Area Modelling for Precise Satellite Positioning”. The Chair. SSG1.179, Dr. Shaowei Han, will report to the
IAG at the next General Assembly in 2003, in Sapporo (Japan).

Kinematic Geodesy: An Evolutionary 'Deadend'?

Colombo et al. (1995) describes an experiment in which a moving vessel in Sydney Harbour (Australia) was positioned to sub-decimetre accuracy relative to several GPS reference receivers deployed at distances up to 1000km from the mobile receiver. This was a dramatic new GPS Geodesy technique that challenged the requirement that geodetic accuracy over long inter-receiver distances was only possible for a static receiver that was collecting carrier phase data over many hours. It was indeed a geodetic technique because: (a) all measurement biases were accounted for in the functional model, (b) sub-part-per-million relative accuracy was obtained, and (c) a simultaneous multi-receiver solution was performed. The first author coined the expression “kinematic geodesy” to describe this technique. In 1995 an Australian Research Council grant was obtained to support graduate studies into high precision, long- and medium-range, kinematic GPS positioning, as reported in Han (1997).

The data processing algorithm used by Colombo et al. (1995) was particularly innovative, consisting of a partitioned Kalman filter that estimated the slow-changing biases such as due to satellite orbit error and atmospheric effects, at the same time generating epoch-by-epoch kinematic coordinate solutions for the mobile receiver, using carrier phase data from several reference receivers (as well as the mobile receiver). The observation biases were carefully modelled, as in the ‘standard’ geodetic methodology used by GAMIT and the Bernese software, and 3-D accuracy is of the order of 3-5cm, for any length of baseline. During the last ten years, this technique has been used in Australia, Denmark, Japan, Spain, The Netherlands and the U.S. for applications as diverse as sea buoys, boats, aircraft, trucks, and altimetric satellites. Recent publications reporting on “kinematic geodesy” projects include Colombo et al. (2000, 2001, 2002).

However, the promise shown by this technique has not led to its widespread adoption by geodesists. Nevertheless this technique can lay claim to having demonstrated, for the first time, the feasibility of carrier phase-based positioning of a moving platform over very long baselines. Amongst its shortcomings are the simultaneous analysis of all GPS data (from reference receivers and the mobile receiver), and the difficulty in implementing this technique in real-time.

Low-cost Deformation Monitoring: The Utility Of Mixed Networks

Deformation monitoring of structures (such as bridges, buildings, etc.) and ground monumentation (in volcanic, ground subsidence and geological faulting zones) are ideal geodetic applications of GPS (Rizos et al., 1997). To keep the cost of such monitoring systems low, single-frequency GPS receivers are often used (see, e.g., the theses by Chen, 2001; Roberts, 2002). However, data from single-frequency GPS receivers cannot be corrected for ionospheric delay, as is the case with dual-frequency data. Therefore a combination of single- and dual-frequency instrumentation in a mixed-mode network is a feasible methodology for ensuring high accuracy coordinate results using a large number of static receivers must be deployed permanently across a region experiencing deformation, while keeping hardware costs as low as possible. This is possible by augmenting the single-frequency receivers with a small number of dual-frequency receivers surrounding the zone of deformation. The primary function of this fiducial network is to generate empirical ‘correction’ terms to the double-differenced phase observables within the deformation monitoring network. This research was funded by the Australian Research Council (1999-2001), in address the need for a low-cost Indonesian volcano monitoring system.

This methodology has been tested in many networks, and results reported in a large number of papers, including Rizos et al. (2000a, 2000b), Chen et al. (2001). Dai et al. (2001) extended this methodology to include integrated GPS/GLONASS reference receiver networks. This methodology can address geodetic applications where a CORS network of geodetic quality GPS receivers exists. Furthermore, this data processing strategy is identical to what we now know as the Network RTK, or multiple reference station, class of techniques. That is, there are three distinct processes: reference station network data processing to generate ‘corrections’, correction of double-differenced phase data involving user receiver(s), and (static or kinematic) baseline processing using the corrected GPS phase observables. It is this separation of processes that sets this class of techniques apart from the conventional multi-station geodetic technique, and the ‘kinematic geodesy” approach described earlier. The extension of this methodology to operate in real-time, though an engineering challenge, is relatively straightforward.

Network RTK Issues: Theoretical & Practical Challenges

Many investigators have contributed to the definition of the appropriate functional and stochastic models for
medium-range and long-range GPS/GLONASS survey-type positioning (as opposed to geodetic techniques) using CORS networks. Research has addressed topics such as: multipath mitigation algorithms, troposphere model refinement, regional ionosphere modelling algorithms, phase centre calibration, and orbit bias modelling. The authors would be unable to do justice to all contributions in this review paper and refer the reader review papers such as Rizos & Han (2002). Although most of these research topics are of general interest to precise GPS positioning, several are explicitly related to the processing of CORS network data in order to generate the empirical ‘correction’ data that must be transmitted to users in Network RTK type implementations. Some of these topics include: rapid ambiguity resolution for the network receivers, validation of the ambiguities so resolved, the nature of the model for the distance-dependent biases across the CORS network, the method of interpolation of the corrections for the user-base station baseline, and the format for the transmitted ‘correction’ data.

After the double-differenced ambiguities associated with the reference station receivers have been fixed to their correct values, the double-differenced GPS/GLONASS residuals can be generated. The spatially correlated errors to be interpolated could be the pseudo-range and carrier phase residuals for the L1 and/or L2 frequencies, or other linear combinations. One core issue for multi-reference receiver techniques is how to interpolate the distance-dependent biases generated from the reference station network for the user's location? Over the past few years, in order to interpolate (or model) the distance-dependent residual biases, several interpolation methods have been proposed. They include the Linear Combination Model (Han & Rizos, 1996; Han, 1997), the Distance-Based Linear Interpolation Method (Gao & Li, 1998), the Linear Interpolation Method (Wanner, 1995), the Low-Order Surface Model (Wübbena et al., 1996; Fotopoulos & Cannon, 2001), and the Least Squares Collocation Method (Raquet, 1998; Marell, 1998). The theoretical and numerical comparison of the various interpolation algorithms has been made by Dai et al. (2003), and there is no obviously ‘superior’ technique. The essential common formula has been identified: all use n-1 coefficients and the n-1 independent ‘correction terms’ generated from a n reference station network to form a linear combination that mitigates spatially correlated biases at user stations.

While theoretical and numerical studies have contributed to the development of the Network RTK class of techniques, there are a host of ‘practical’ issues that must be addressed in order to implement a RTK service that operates ‘24/7’. For example, the Network RTK system needs a data management system and a data communication system. It needs to manage corrections generated in real-time, the raw measurement data, multipath template for each reference stations (for multipath mitigation), precise/predicted IGS orbits, etc. There are two aspects to the data communication system: (a) between the master control station (MCS - where all the calculations are undertaken) and the various reference stations, and (b) communication between the MCS and users. Furthermore, from the Network RTK implementation point of view, there are three possible architectures: (1) generation of the Virtual Reference Station (VRS) and its corrections, (2) generating and broadcasting Network RTK corrections, or (3) broadcasting raw data for all the reference stations. The debate about the ‘best’ architecture is still raging, and it is likely that combinations of some or all may be implemented, with the appropriate RTCM/RTK messages being defined. However, research into all aspects of Network RTK, theoretical and practical, is difficult to undertake in universities because of the expense of establishing and operating ‘test networks’.  

Singapore Integrated Multiple Reference Station Network

Due to the complexity (and cost) involved in establishing fully functioning receiver networks, the data links and the data processing/management servers at the master control station (MCS), there have been comparatively few university-based Network RTK systems established to support research. During the last few years, to the best of the authors’ knowledge, only the Singapore Integrated Multiple Reference Station Network (SIMRSN) has been operating both as a research facility and an operational Network RTK that can be used by surveyors. The SIMRSN is a joint research and development initiative between the Surveying and Mapping Laboratory, of the Nanyang Technological University (NTU), Singapore (http://gis.ntu.edu.sg/generaterix/index.htm), the Satellite Navigation and Positioning group, of the University of New South Wales (UNSW), Australia (http://www.gmat.unsw.edu.au/snapwork/singapore.htm), and the Singapore Land Authority (SLA). In Singapore the project was funded by the National Science and Technology Board (1998-2001), while in Australia it was funded by the Australian Research Council (1999-2001).

The SIMRSN consists of five continuously operating reference stations (tracking satellites 24 hours a day), connected by high speed data lines to the MCS at NTU (Figure 1). It is a high quality and multi-functional network designed to serve the various needs of real-time precise positioning, such as surveying, civil engineering, precise navigation, road pricing etc. The SIMRSN also serves off-line non real-time users via the Internet. The inter-receiver distances are of the order of several tens of kilometres at most. However, tests conducted in 2001
have shown that even a network with such comparatively short baselines had difficulty in modelling the disturbed ionosphere in equatorial regions, during the last solar maximum period of the 11 year sunspot cycle (Hu et al., 2002a; 2002b; 2002c).

![Fig. 1 The Singapore Integrated Multiple Reference Station Network - supporting R&D into Network RTK and other network-based positioning concepts.](image)

### Concluding Remarks

Network RTK is best implemented by a service provider, an organisation that operates the receiver network infrastructure, the necessary data communication links and the MCS facility. This is a radically different scheme to the standard single base RTK where the GPS Surveyor owns and operates all of the equipment. At present there are very few continuously operating Network RTK systems. However, with the likely upgrade of CORS networks around the world to offer RTK services over the next few years, the author believes that there will be a boom in Network RTK implementations.

There is currently only one commercial product, the Trimble VRS (Vollath et al., 2002), although the Leica company has also developed a Network RTK system (Euler et al., 2001). A number of test networks have been operating in Europe, the U.S., Australia, New Zealand, China and Japan. However, a unique university-led Network RTK system has been operating in Singapore for a number of years. Australian and Singaporean researchers have gained invaluable insight into the challenges of operating such an infrastructure on a ‘24/7’ basis. It is intended to mirror this facility in Sydney during 2003, supporting independent research into Network RTK algorithms, products, operational issues, and business models, carried out outside North America, Europe and east Asia.

The ‘roots’ of Network RTK can be found in geodesy, surveying and precise navigation. Each sub-discipline can claim some credit for the development of the Network RTK concept. The author in this paper has emphasised the geodetic perspective, and shown how geodetic methodology and applications were a driver for multi-reference receiver techniques that ultimately led to the development of Network RTK. The paper has also highlighted the contributions of Australian and Singaporean researchers to the development and implementation of the data processing algorithms, and associated data management and communication systems, that underpin the totally university-developed Network RTK service.

### Bibliography


Chen X., S. Han, C. Rizos & P.C. Goh (2000) Improving real-time positioning efficiency using the Singapore Integrated Multiple Reference Station Network (SIMRSN), 13th Int. Tech. Meeting of the Satellite Div. of...
the U.S. Institute of Navigation, Salt Lake City, Utah, 19-22 September, 9-18.


Rizos C., S. Han, H.Y. Chen & P.C. Goh (1999) Continuously operating GPS reference station networks: New algorithms and applications of carrier phase-based medium-range, static and kinematic positioning, in “Quo vadis geodesy…?”, special publication to celebrate Prof. Erik W. Grafarend’s 60th birthday, Dept. of Geodesy & Geoinformatics, University of Stuttgart, ISSN 0933-2839, 367-378.


Introduction to the Wide Area Augmentation System

Todd Walter
Stanford University
e-mail: twalter@Stanford.edu

Received: 28 October 2002 / Accepted: 4 November 2002

Biography

Dr. Todd Walter received his B. S. in physics from Rensselaer Polytechnic Institute and his Ph.D. in 1993 from Stanford University. He is currently a Senior Research Engineer at Stanford University. He is a member of the WAAS Integrity Performance Panel (WIPP) focused on the implementation of WAAS and the development of its later stages. Key contributions include: early prototype development proving the feasibility of WAAS, significant contribution to MOPS design and validation, co-editing of the Institute of Navigation's book of papers about WAAS and its European and Japanese counterparts, and design of ionospheric algorithms for WAAS. He was the co-recipient of the 2001 ION early achievement award.

The United States Federal Aviation Administration (FAA) is on the verge of implementing an entirely new navigation system. The Wide Area Augmentation System (WAAS) will monitor and correct the ranging signals from the GPS constellation of satellites. Most importantly, WAAS will provide a certified level of integrity. The corrections will improve the vertical accuracy of the system from ten or more meters to just one or two. It is the integrity, however, that will open the doors for widespread aviation use. Integrity is the ability to bound the remaining errors on the signal. Thus, the user will have a rigid limit on the magnitude of their position error. Although GPS does have a strong track record, it has not yet been generally approved for the most demanding of aviation applications: precision approach. GPS was not designed for this application and lacks the necessary real-time monitoring. Currently, aviation uses a technique exploiting redundant satellites, to provide sufficient integrity. This technique requires good geometry and offers confidence bounds, termed protection levels, measured in hundreds of meters. As such it is suitable for en route flying and non-precision approach. However, precision approach, which brings airplanes within a few hundred feet of the ground, has more stringent needs. Here the protection level must be measured in tens of meters. WAAS will enable aircraft to conduct several forms of precision approach.

WAAS grew out of several concepts that originated in the 1980s. The first, the GPS Integrity Channel (GIC), proposed using ground monitoring to determine GPS satellite integrity. The integrity information was to be broadcast to users via a communication satellite. Separately, researchers examined the concept of Wide Area Differential GPS (WADGPS). This was a real-time extension to earlier work for surveying that uses a network of widely distributed reference stations to improve the accuracy of GPS position fixes. These ideas were combined, in the early 1990s, with the recognition that the communication downlink could be accomplished with a GPS-like signal at the same frequency. This final design would provide differential corrections (WADGPS), integrity (GIC), and an additional ranging source to supplement GPS. In 1994, the FAA awarded a contract to create this system.

WAAS measures observed ranges to the satellites from its reference stations. These raw observables are transmitted to master stations via a terrestrial communication network. Each master station processes the measurements to create corrections for satellite clock error, satellite ephemeris error, and ionospheric path delay. More importantly, it determines bounds for the uncertainty in the remaining errors. These corrections and bounds are sent to a ground uplink station that transmits them to geostationary satellites. The geostationary satellites act as “bent-pipes” and reradiate the signal down toward Earth. These signals are received along with the GPS signals and used to formulate position estimates.
The geostationary signal is very similar to the GPS signal. It provides an additional ranging source as well as carrying differential corrections and confidence bounds. In order to achieve these two goals, the signal must have a very low data rate. Otherwise, the required signal power would interfere with the GPS satellites. The chosen data rate is only 250 bits per second. This low bandwidth must carry all of the differential corrections and confidence bounds applicable over a broad geographic range.

Although WAAS is not yet certified, there is an operational test signal that has been available since August of 2000. Already there are several commercial WAAS-capable receivers. Because it now incorporates all of the integrity monitoring algorithms that will exist at commissioning, this experimental signal has a much higher level of integrity than the typical differential system. It is finding widespread use in agricultural and many other applications, as more and more non-aviation users are discovering the benefits of this free signal. It offers the great advantage that the differential corrections come in the same antenna that the GPS signals do. There are no additional antennas to install, no local reference stations to set up, and no additional communication channels to maintain. Just turn it on and use it anywhere in the United States. Similar systems are also being developed in Europe, Japan, India, and China.

The main challenge for WAAS is the certification of the system’s integrity. It is not enough that the system has performed flawlessly in the past. We must be confident that it will continue to operate fault-free in the future. This is accomplished by investigating the possible fault modes of the system for different operating conditions. System performance in the face of worst-case faults is analyzed. Confidence bounds are generated based on this very conservative analysis. The result is large confidence bounds that we are certain contain the worst undetected error modes. While this makes the system safe, the resulting confidence bounds in the position domain, termed Vertical Protection Level (VPL) and Horizontal Protection Level (HPL), range from 20 meters upward, while the accuracy remains better than a few meters. This conservatism is necessary at the early stages of the system. As we gain more understanding of the different fault-modes and their likelihood of occurrence we will be able to lower these protection levels, while having little impact on the underlying accuracy. Smaller protection levels will translate into higher availability and lower approach minima.

The WAAS test signal-in-space has attracted numerous users for its improved accuracy. However, these users have encountered some shortcomings in its performance. Chief among these is the limited geostationary satellites’ coverage. Currently, there are only two geostationary satellites and both are relatively low to the horizon for most of the U.S. While these are sufficient for aviation users with unobstructed views to the sky, they are problematic for terrestrial users. Very few parts of the country can see both satellites. If the only geostationary satellite above your horizon is obstructed, you will no longer receive differential corrections. The long-term plan is to have redundant coverage everywhere within the U.S. and have satellites at higher elevation angles. Although users may encounter limited accessibility today, the FAA intends to utilize additional geostationary satellites beginning in 2005. The final configuration will likely be four geostationary satellites. While they offer the advantage of always being in view, geostationary satellites have the drawback that they can only be in a certain part of the sky. For the United States, these satellites will always be to the south and below 50 degrees elevation angle for all but the southernmost parts of the country. However, the additional geostationary satellites at more optimal locations should provide dramatic improvements for nearly all users.

Two other limitations of the system are uncertainty in ionospheric delay and susceptibility to interference. Bounding the possible errors in the ionosphere, for single frequency users, is currently the dominant source of uncertainty for WAAS. Furthermore, the ionosphere above the U.S. is better-studied and less variable than most of the rest of the world. Thus, it will pose an even greater threat to availability in other regions. Vulnerability to interference has been the topic of several reports and results from our use of an extremely low power signal on a single frequency. Fortunately, GPS modernization addresses both problems. Two civil frequencies in protected spectrum bands allow users to directly measure their own ionosphere. This allows smaller uncertainties and greater availability. Furthermore, the aircraft has a reversionary mode if either frequency is lost since it can use the existing WAAS grid of ionospheric corrections. Additionally the modernized signals will have more power making them more resistant to interference. Current research focuses on further mitigation of these issues.

WAAS will provide numerous aviation benefits when it is declared operational in mid to late 2003. There will be enhanced runway capability as all usable runway ends can have some level of service. How low one can fly depends on local terrain and obstacles as well as airport infrastructure (paint and lighting). The increased accuracy throughout the service region will eventually allow closer separations with the same or greater level of safety as today’s equipment. Additionally, the instrumentation required to provide and use WAAS is substantially less expensive than today’s suite of equipment. In 2003, the first level of service, LPV, will be offered. This will allow aircraft to come as low as 250 feet above the ground before transitioning to visual guidance. Over time the level of service is expected to improve. In later phases better availability and lower decision heights will become...
possible. To learn more about this exciting system please go to http://gps.faa.gov/ Programs/WAAS/waas.htm or http://waas.stanford.edu.

**Further Reading**