Vol. 7, No. 1, 2008

Journal of Global Positioning Systems

ISSN 1446-3156 (Print Version) ISSN 1446-3164 (CD-ROM Version)

International Association of Chinese Professionals in Global Positioning Systems (CPGPS)

Journal of Global Positioning Systems

Aims and Scope

The Journal of Global Positioning Systems is a peer-reviewed international journal for the publication of original research papers, review articles, invited contributions, also including selected papers presented at non-referred conferences and symposiums. Articles should present discussions of technologies and applications of any positioning systems, including Global Navigation Satellite Systems (GNSS) and their various augmentations and local components, indoor positioning, and inertial navigation systems. Articles presenting advances of other related areas, such as wireless communications, intelligent vehicle systems, sensor networks, spatial information and geosciences are also welcome. Short research and technical notes, book reviews, lecture series and commercial advertisements can be included. Specific questions about the suitability of prospective manuscripts may be directed to the Editor-in-Chief.

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 is available in 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.php. Whilst CPGPS owns all the copyright of all text material published in the Journal, the authors are responsible for the views and statements expressed in their articles. Neither the authors, the editors nor CPGPS can accept any legal responsibility for the contents published in the journal.

Journal of Global Positioning Systems

Vol. 7, No. 1, 2008

Table of Contents

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 1-8

Journal of Global Positioning Systems
2008
© CPGPS

GPS RTK Performance Characteristics and Analysis

Yanming Feng
Faculty of Information Technology, Queensland University of Technology, Australia

Jinling Wang
School of Surveying and Spatial Information Systems, The University of New South Wales, Australia

Abstract

Global Navigation Satellite Systems (GNSS) provide various types of positioning state solutions, such as single point positioning (SPP), precise point positioning (PPP), differential GPS (DGPS) and real time kinematic (RTK) solutions. These solutions are obtained involving different data types, receivers, samples, serving different classes of users. Previous studies on performance characteristics have mainly focused on SPP solutions for safety-of-life navigation applications. This paper defines various useful performance characteristics for carrier phase Ambiguity Resolution (AR) and Position Estimation (PE) solutions in the RTK context. These parameters, including base-rover distance, time-to-first fix (TTFF), AR reliability, RTK accuracy, availability and integrity, etc, effectively represent the performance of a commercial RTK system and can be used to evaluate RTK systems and algorithms, and processing strategies through extensive experimental results. Statistical results from extensive field experiments were obtained using a commercial RTK system, demonstrating convincing overall system performance in different perspectives. Experimental results from three baselines were also analysed using a version of research-oriented RTK software, showing that AR performance improvement of using Wide-lane (WL) and Narrow-lane (NL) signals with respect to the original L1 and L2 signals when the baselines exceed 20 kilometres.

Key words: GNSS, real time kinematic (RTK) positioning, performance characteristics, ambiguity resolution (AR), RTK integrity.

1. Introduction

Global Navigation Satellite Systems (GNSS) positioning may be classified into several different types, depending on (i) the types of measurements used in the positioning estimation, (ii) the data epochs or data arcs required to create a set of solutions, and (iii) the number of receivers involved in the positioning operations. Single Point Positioning (SPP) produces navigation solutions with pseudorange measurements from a single receiver and a single epoch. Precise Point Positioning (PPP) solutions are obtained using both code and phase measurements from a single receiver, but a period of observations, e.g., tens of minutes to hours, regardless of kinematic or static user applications. Differential GPS (DGPS) solutions are based on code measurements from a single epoch as well, but using the differential corrections from a reference station or network. Real Time Kinematic (RTK) positioning makes use of carrier phase measurements in the differential positioning mode, ideally, from a single epoch. Practically, multiple epochs or a short period of observations are often involved to achieve reliable AR, while the RTK solutions are derived from the current epochs.

Different characteristics are required to evaluate various GNSS solutions, to address positioning performance requirements for various applications. Code based SPP and DGPS navigation are the simplest and most robust positioning modes, but, evaluation of code based navigation solutions has been a quite involved problem. The parameters of accuracy, availability, continuity and integrity are defined to evaluate the performance of navigation solutions in aviation navigation (Langley, 1999). For instance, availability is an instantaneous performance characteristic defined as a percentage of time during which the service is available at a certain accuracy. Integrity relates to the level of trust that can be placed in the information provided by the navigation system. It includes the ability of the navigation system to provide timely and valid warnings to users when the system must not be used for the intended operation or phase of flight. GPS does not provide integrity information to users.

In the context of integrity, three parameters, integrity risk, time to alert and alarm-limit are defined. Furthermore, various methods for monitoring the

integrity of GPS SPP solutions have been proposed (i) external monitoring, which relies on a number of ground stations, where a faulty individual satellite is identified and a warning is sent to users within the time-to-alert required. The typical example is the Wide Area Augmentation Systems (WAAS) (Enge et al, 1996; Walter, 2002), (ii) Receiver Autonomous Integrity Monitoring (RAIM) (Brown, 1996), which is applicable within a user receiver to autonomously determine system integrity. The method attempts to detect the existence of faulty measurements and identification of unhealthy satellites.

RTK positioning is a much more complicated and vulnerable process, aiming to achieve the accuracy as high as centimetres with as few as possible data epochs in real time for any user kinematics. Therefore a greater care has to be taken of to characterise the performance and to address the concerns of liability-critical professional positioning users, such as surveying, data acquisitions, machine automation in precision agriculture, mining and construction and future safety-related vehicle navigation. In many applications, users are concerned about not only accuracy, but also availability and integrity of the solutions. For instance, in an open cut mine, the cost of every hour of RTK service outages to the productions would reach the level of one million (Higgins, 2007).

However, RTK performance characteristics are much less studied and understood than those of the SPP solutions. This paper presents a systematic review for RTK performance characteristics and then evaluates the GNSS ambiguity resolution (AR) and RTK performance with GPS measurements in terms of the various RTK performance parameters, which may not necessarily be suitable for real time quality control purposes. In the following sections, we first present the definitions of various performance parameters for AR and RTK solutions in order to comprehensively evaluate performance of a RTK system. Next, we outline the linear equations for AR and Position Estimation (PE) with the specific WL and NL signals, to conceptually demonstrate the dependence of performance on the models and algorithms. In the forth section, we will first examine the statistical results for the different performance parameters of a commercial RTK system, HD-RTK2^{TM}, according to its extensive data sets of different baselines. Utilising the research version of the QUT-RTK software, we then compare AR performance improvement of the WL and NL signals with respect to the use of the original L1 and L2 signals. In this analysis, three 24-h RINEX data sets over the baselines of 21, 56 and 74 km will be analysed. Finally, the major results of RTK performance characterisations and extensive numerical analyses are outlined.

2. Performance Characteristics of a RTK System

A RTK system consists of a continuous operating reference station network and data links between a network server and reference stations and between the server and user-terminals. The reference network comprises a minimum of one reference station and a network server with a data processing facility. Data links set up between the network server and user receivers provide or deliver the differential corrections to user-terminals. The user terminal is generally equipped with a GNSS RTK receiver and a communication device and a user control/interface unit where RTK solutions are integrated or interfaced with a particular application. To completely assess the performance of a RTK system, the following parameters should be considered (Feng & Wang, 2007):

Base-rover distance, which is the maximum radius of circle coverage, where a signal base station can serve effectively, allowing the users to receive the RTCM messages within certain latency and obtain its RTK solutions epoch-by-epoch. A relevant concept is the inter-station distance in the network-based RTK case. As shown in Fig. 1, the base-rover distance D is approximately equivalent to 0.5774 times of the inter-station distance S. For instance in the Virtual Reference Station (VRS) system where the maximum inter-station spacing is S=70 km, the equivalent maximum base-rover distance is about 40 km. The distance limitation is mainly caused by the strong dependence of the ionospheric biases on the separation of two receivers. The next distance-dependent error factor is the residual tropospheric errors after modelling corrections. The effect of broadcast orbital errors is relatively small and may be ignored. The system performance is considered more desirable if a longer base-rover distance is allowed.

Timeliness of RTCM message, which is defined as the ***time latency*** of the latest RTCM message available for users with respect to the user time instant at which the user states are needed to compute. Users will need to predict the ranging corrections to the most current time instant when the user-terminal produce RTK solutions. This latency is the sum of delays caused by data processing at the base station/network centre and data transmissions from base stations to network centre, and messages from the network server to users, typically one to a few seconds. This parameter is obtainable from statistical results for a given operational environment and communication links.
Another related parameter is the ***communication rates***, for instance, 1Hz, or 5Hz and 10Hz. The higher communication rates are required for higher position update rates and control of position accuracy.

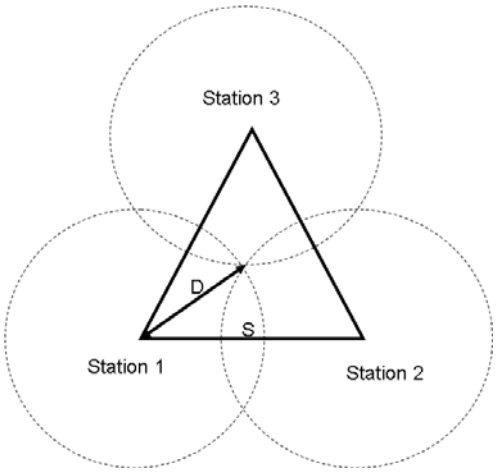

Fig. 1 Relation of Inter-station distance S and equivalent base-rover distance D=0.5774S

The above two parameters are used to evaluate the performance at the system level. At the user terminal, performance of a RTK system may be evaluated using the following characteristics, which may vary with the system performance parameters.

Time To First-Fix (TTFF). This is referred to the time period required to resolve or fix sufficient integer ambiguities of the linear equation system, then perform position estimation.. In some literature, this parameter is known as "Time To Ambiguity-Fix (TTAF)" or "initialisation time" of the RTK system. However, TTAF is more suitable for more general situations where all the integer parameters are resolved and fixed independently at each epoch, involving measurements from single or multiple most recent epochs. It is most desirable if the RTK system always fix the ambiguity integers for all the double-differenced phase measurements of the current epoch in the linear equation system, to minimise the discontinuity of the RTK solutions after any phase breaks. However, one can also define the TTAF based on the partial ambiguity resolution (PAR) concept developed by Teunissen (1999). The question is wether the DD phase measurements with partially resolved ambiguities are sufficient for PE to support the RTK services.

AR fixed rate. This instantaneous performance characteristic is defined as the fixed rates of the integer estimation results. The system may be unavailable for AR, when the geometry is too week, or satellites in view are too few, or the effects of various errors are too strong. AR fixed rate can be calculated by the ratio of the total number of fixed DD integers to the total number of DD integers over a continuously operating session. The fixed integers are these that have passed the validation tests in the integer search process. The problem is that the validation process may include incorrect integers and exclude correct integers. This parameters can be provided by a RTK processing unit.

AR reliability (AR success rate). This is defined as the percentage of the total correctly fixed DD integers over the total number of DD integers. In some studies, the AR success rate was implicitly defined as the total number of epochs, when all the ambiguity integers are correctly fixed, with respect to the total number of epochs of the data session.

The other concern is the performance of RTK positioning with the ambiguities-resolved double-differenced (DD) phase measurements. To this end, the RTK position estimation is similar to the code-based SPP solutions. Therefore, we can similarly introduce the SPP performance parameters to define the performance of the RTK solutions:

RTK accuracy. This is defined as the degree of conformance of an estimated RTK position at a given time to a defined reference coordinate value (or 'true' value) which is obtained from an independent approach, preferably at higher level of accuracy. As usual, the RTK accuracy can be specified versus the rover-base distances, for instance, σ=1.0 cm + 0.5 ppm.

RTK availability (in term of accuracy). This is defined as the percentage of time during which the RTK solutions are available at a certain accuracy using the ambiguity-fixed and/or ambiguity-float phase measurements.

RTK availability (in term of AR reliability). This characteristic is be defined as the percentage of time, of which PE is based on all the phase measurements whose integers have been correctly fixed at each epoch, assuming all the ambiguity-fixed solutions will give required accuracy.

RTK integrity. This relates to the confidential level that can be placed in the information provided by the RTK system. It includes the ability of the RTK navigation system to provide timely and valid warnings to users when the system must not be used for the intended operation. For instance, the RTK system with the integrity capacity can inform users when the actual positional errors of the RTK solutions have exceeded Horizontal/Vertical Protection Levels (HPL/VPL) within a certain Time-To-Alert (TTA) period at a given Integrity Risk (IR). RTK Integrity Risk is defined as the probability that the system claims its normal operational status while actually being in an abnormal status, e.g., the ambiguities being incorrectly fixed and positional errors having exceeded the given HPL.

RTK continuity. This is defined as RTK availability over a certain operational period and conditions. Both

TTAF and AR availability will affect the RTK continuity. This parameter is provided to address user requirement for the tolerable service down-time over a certain operational period, such as 24 hours and 7 days. For instance in mining and civil construction, user tolerable down-time is about 1 to 2 minutes per day, corresponding a 99.9% of continuity requirement of services (Positioning one consulting, 2008).

Those parameters may be either all or selectively used to evaluate performance of a RTK system, although variations and modifications to these definitions are still possible. Of these parameters, the base-rover distance, time to ambiguity fix, AR reliability, RTK availability and RTK accuracy may be of most concerns to most professional users. The concept of RTK integrity is also important from the liability-critical users' perspective. In farming machine automation applications, the HPL is about 10 centimetres while for civil construction machine automation, the requirement for VPL would be as high as a few centimetres ((Positioning one consulting, 2008). However, few existing commercial RTK systems provide sufficient performance parameters in their specifications. Obviously, more detailed performance information would indeed help the users choose a desirable RTK system to meet the performance requirements, including integrity requirements.

3. Linear equations for ambiguity resolution with wide-lane and narrow-lane phase measurements

For a typical single-base RTK problem, the standard linearised observation equations for the n×1 double-differenced pseudoranges P_1 (or C/A) and P_2, carrier phases L_1 and L_2 can be written as follows (Misra and Enge, 2004)

$$\begin{bmatrix}\delta P_1\\ \delta P_2\\ \delta L_1\\ \delta L_2\end{bmatrix}=\begin{bmatrix}P_1-\rho\\ P_2-\rho\\ L_1-\rho\\ L_2-\rho\end{bmatrix}=\begin{bmatrix}A\\ A\\ A\\ A\end{bmatrix}\delta X-\begin{bmatrix}0 & 0\\ 0 & 0\\ \lambda_1 I & 0\\ 0 & \lambda_2 I\end{bmatrix}\begin{bmatrix}N_1\\ N_2\end{bmatrix}+\begin{bmatrix}\varepsilon_{P1}\\ \varepsilon_{P2}\\ \varepsilon_{L1}\\ \varepsilon_{L2}\end{bmatrix} \quad (1)$$

where ρ is the n×1 computed double-differenced range vector; A is the n×3 observational matrix; δX is the 3×1 user state vector; the carrier phases L_1 and L_2; have the wavelengths λ_1 and λ_2 and the integer ambiguities N_1 and N_2 respectively; ε_{p1}, $\varepsilon_{p2,}$ ε_{L1} and ε_{L2} are the noise vectors for the respective measurement vectors δP_1, δP_2 δL_1 and δL_2.

Feng and Rizos (2007) and Feng (2008) suggested the use of the Wide-Lane (WL) $L_{(1,-1)}$ and Narrow-Lane (NL) $L_{(4,-3)}$ instead, in order to minimise the effects of the larger ionospheric errors for AR over longer ranges.

$$P_{(1,1)}= (f_1P_1+f_2P_2)/(f_1+f_2); \quad (2)$$

$$L_{(1,-1)} =(f_1L_1-f_2L_2)/(f_1-f_2); \quad (3)$$

$$L_{(4,-3)} =(4f_1L_1-3f_2L_2)/(4f_1-3f_2); \quad (4)$$

where f_1 and f_2 stand for the frequencies for L1 and L2 carrier respectively. In (2) to (4), the subscript (i, j) represents the integer values of the coefficients of the combined measurements. $L_{(1,-1)}$ and $L_{(4,-3)}$ have the wavelengths of $\lambda_{(1,-1)}$=86.2 cm, and $\lambda_{(4,-3)}$=11.45 cm, respectively. As a result, we have the following linear equations,

$$\begin{bmatrix}\delta P_{1,1}\\ \delta L_{(4,-3)}\\ \delta L_{(1,-1)}\end{bmatrix}=\begin{bmatrix}A\\ A\\ A\end{bmatrix}\delta X-\begin{bmatrix}0 & 0\\ \lambda_{(4,-3)}I & -\lambda_{(4,-3)}I\\ \lambda_{(1,-1)}I & -\lambda_{(1,-1)}I\end{bmatrix}\begin{bmatrix}N_1\\ N_2\end{bmatrix}+\begin{bmatrix}\varepsilon_{P(1,1)}\\ \varepsilon_{L(4,-3)}\\ \varepsilon_{L(1,-1)}\end{bmatrix} \quad (5)$$

where $\delta P_{(1,1)}$, $\delta L_{(4,-3)}$ and $\delta L_{(1,-1)}$ are the residual vectors between the observed and computed range vector ρ or the n×1 double-differenced $P_{(1,1)}$, $L_{(1,-1)}$ and $L_{(4,-3)}$ measurement vectors; I is the n×n identity matrix; $\varepsilon_{p1,1}$, $\varepsilon_{L(1,-1)}$, $\varepsilon_{L(4,-3)}$, are the noise vectors for $P_{(1,1)}$, $L_{(1,-1)}$ and $L_{(4,-3)}$ measurements, respectively. n=k-1 where k is the number of satellites used in computation.

For convenience, Equations (1) and (2) are rewritten as follows,

$$\delta Y_j = A_j \delta X_j + B_j N_j + \varepsilon_j \quad (6)$$

$$E(\varepsilon_j) = 0; Var(\varepsilon_j) = \sigma^2 W_j^{-1} \quad (7)$$

where the subscript j represents the jth epoch; δY is the m×1 observation vector; A is the m×3 matrix; δX is the 3×1 state vector; and B is the m×p matrix, N is the p×1 ambiguity vector; in which m=3(k-1) and p=2(k-1) with dual-frequency GPS measurements.

The above equations are provided for each measurement epoch, implying that the state and ambiguity parameters are estimated and fixed with the measurements at the current epoch only, which yields desirable kinematic position solutions without imposing the assumptions of phase measurement continuity and sample intervals. Modelling is always the first key process for a RTK system. This includes the processes of combining and differencing measurements, imposing constraints such as known coordinates and integers, applying ionosphere and troposphere corrections etc. The stochastic models (7) give the statistical knowledge or assumptions on the residual errors and measurement noises, such as zero

mean white noise, correlations between the measurements of different epochs (Wang, 2000; Wang et al., 2002)

The next key process is to complete AR and PE following one of the AR methods, such as the Least-squares ambiguity decorrelation adjustment (LAMBDA) method (Teunissen, 1995; Teunissen et al., 1997), Minima Search (LMS) method (Pratt et al., 1997). Any improved version of the methods will be an additional advantage. The process basically consists of a Least-Squares estimator and an Integer Search Engine. The estimator provides initial real values for both state parameters and ambiguity parameters and their covariance matrix for the AR integer search, and then final PE after ambiguities are fixed to their correct integer values. The integer search engine performs a statistical search over the potential ambiguity candidates to find and validate the best set of integer candidates.

Implementation of efficient multipath mitigation approaches, quality control and quality assurance procedures in the above modeling and estimation processing is also important. It is fairly the case that the success of a RTK system depends on detailed processing techniques. Some software systems implement a more efficient integer search algorithm, whilst others are superior in deterministic and/or stochastic modeling. The most successful AR software takes good care of the detailed elements, this being especially true in the current GPS system, where only L1 and L2 carriers are available for AR.

4. Performance Analysis of Experimental RTK Solutions

This section will provide numeral analysis results for the performance of RTK solutions obtained from a commercial RTK system and the new algorithms described in Section 3, according to the performance characteristics defined in Section 2.

4.1 HD-RTK2^{TM} Performance

A commercial RTK system, HD-RTK2^{TM}, developed by HandyNav Inc (HandyNav, 2005), was provided to the first author for this performance analysis, so that we can demonstrate some of the performance parameters and how the different parameters are obtained through experiments and are related to each other. Using two NovAtel OEM4 receivers, the GPS testing data were collected for seven static baselines over 2.5 to 31 kilometres in Brisbane and processed using the HD-RTK2TM software.

Fig. 2 shows the TTFF (seconds) versus the baselines in km in the above tests. The RTK standard deviation (STD) accuracy in horizontal and vertical direction is illustrated in Fig. 3, confirming the formal horizontal and vertical accuracy within 1cm+0.5ppm and 2cm+1ppm respectively.

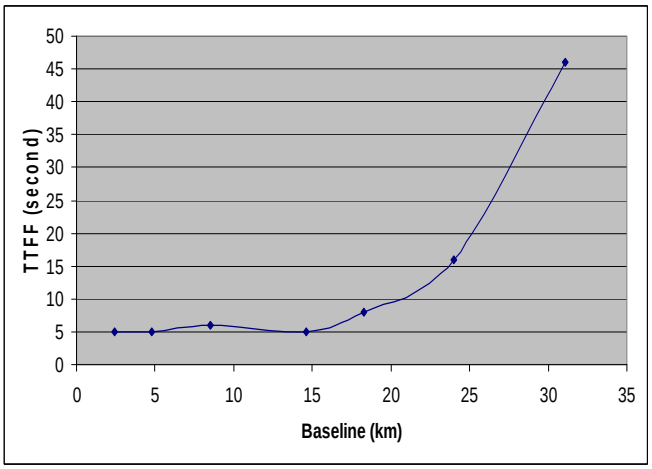

Fig. 2 TTFF (seconds) plotted against baseline lengths

Thanks to HandyNav (2005), we were also able to examine AR availability and AR reliability against different baselines. The GPS data sets were obtained from 16 static baselines over 2 to 45 kilometres from 60 to 400 hours. Each data set was processed separately every 300 seconds, producing over 250,000 sets of results and solutions. Therefore the AR availability and reliability results can more definitively represent the performance characteristics of the tested RTK system.

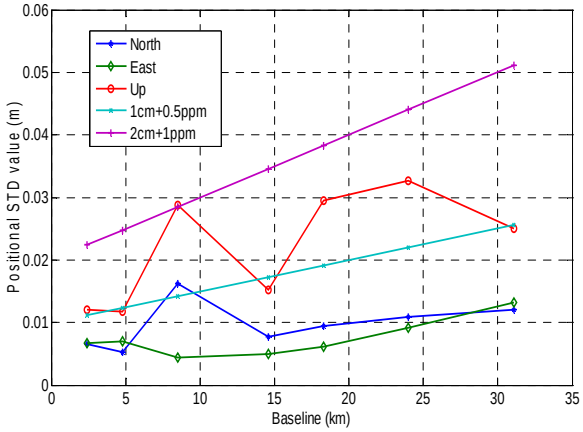

Fig. 3 RTK positioning (STD) accuracy vs base-rover distances

Fig. 4 illustrates the AR availability varying with baseline lengths, while Fig. 5 plots the AR reliability against the baseline length, showing the difference and similarity between the two indicators. We see that AR reliability is not necessarily worse than AR availability. But, in general, the longer the baseline is, the lower the AR availability and AR reliability are.

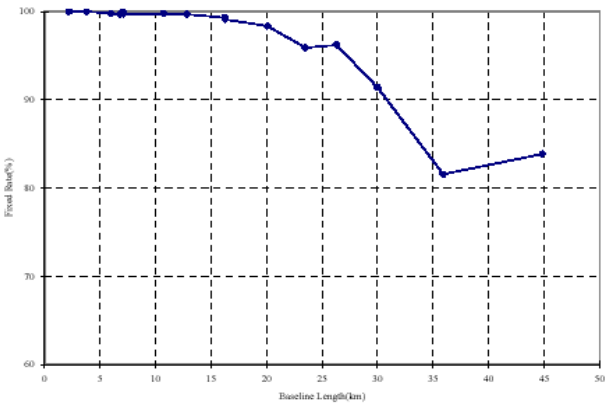

Fig. 4 AR availability variation Vs base-rover distances

In summary, with these extensive experimental results, we conclude that the HD-RTM2^{TM} system can provide instant RTK solutions for distance of up to 20 km, and ambiguity-fixed solutions for distance of up to 50km. The AR reliability of the fixed solutions is above 99% for 20km baselines and 98% for 50 km baselines. The position accuracy for integer-fixed solutions is 1cm+0.5ppm (horizontal) and 2cm+1pmm (vertical). It is believed that these performance specifications would be more convincing to users than the specifications given in the most commercial RTK systems

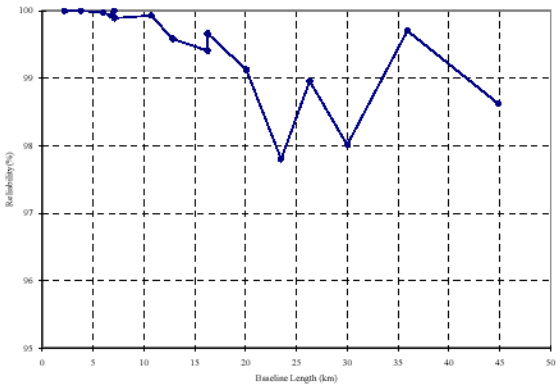

Fig. 5 AR reliability variation Vs base-rover distances.

4.2 Performance Analysis of Proposed RTK Models and Algorithms

Using the research version of the QUT-RTK software, we are able to test performance of different models, algorithms and statistical conditions and processing strategies in terms of various characteristics defined in Section 2. In this context, we examine the performance advantages of the model (5), using the WL $L_{(1,-1)}$ and NL $L_{(4,-3)}$, with respect to the model (1), using L1 and L2 signals directly. Theoretically, the AR with the model (5) may perform better than the model (1) when over longer baselines where the effects of ionospheric delay is minimized with respect to the wavelengths, as demonstrated in Feng and Rizos (2007) and Feng (2008). We now examine the AR and positioning estimation (PE) performance using data sets of three different baselines. In principle, this advantage may be more evident for longer baseline and when the AR is performed instantly with measurements from single epochs. Hence, the following experimental results will focus on a few key performance parameters such as AR reliability and RTK availability and RTK accuracy, based on single-epoch ambiguity resolution.

Three 24-h GPS data sets were collected on 1 January 2007 from US Continuously Operating Reference Stations network (http://www.ngs.noaa.gov/CORS). All three baselines are South-North directions, sampled at 15 second intervals.

Tables 1 and 2 summarise the performance statistical results obtained with the models (1) and (2), respectively, for three baselines of 21, 56 and 74 km. It is noted that the AR reliability and RTK availability of the model (1) are evidently higher than those of the model (2), especially for the longer baselines. It is important to note that the above performance results are obtained purposely to reflect the benefits of a new algorithm under the same circumstance and do not represent the potential performance of the RTK system in use. In fact, we have intentionally removed some modelling process like known integer constraints and advanced stochastic modelling procedures, which could change the AR and RTK performance results. On the other hand, the proposed characteristics may also be effective to assess different stochastic models and particular processing strategies, similarly through extensive numerical studies.

Table 1 Performance results of the model (5), with WL $L_{(1,-1)}$ and NL $L_{(4,-3)}$ observables for three baselines

	P478-P474 21km	P473-P478 56km	P473-474 74km
Total No of epochs	5760	5760	5758
Number of epochs with wrong integers	82	727	1309
Number of DD phase measurements	72562	72452	72310
Number of wrong WL/NL integers	12/180	134/2487	143/4668
AR success rates for WL/NL signals Overall(average)	99.97% 99.50% 99.74%	99.63% 93.13% 96.38%	99.60% 87.09% 93.35%
RTK availability (correct integer-fix)	98.58%	87.38%	77.27%
RTK availability (0.025,0.025,0.05cm)	98.44% 99.79% 94.53%	84.81% 90.73% 90.90%	68.41% 84.00% 77.56%
RTK accuracy (STD in North East and Up (m)	0.008m 0.006m 0.025m	0.009m 0.007m 0.020m	0.009m 0.008m 0.030m

5. Concluding remarks

The paper has contributed to definitions of various AR and RTK performance characteristics, including Base-Rover distance, Timeliness of RTCM messages, Time-to-first fix, AR availability, AR reliability, RTK availability and RTK accuracy etc. These characteristics enable a more comprehensive assessment to the performance of a RTK system or some particular RTK models and algorithms.

The above performance characteristics may be provided as technical specifications of commercial RTK systems to users, who can easily decide which products to use to meet their performance requirements. On the other hand, users and suppliers can test a commercial RTK system against the given parameters through extensive field experiments under various service and observational conditions

Table 2 Performance results of the model (1), with L1 and L2 observables for three baselines

	P478-P474 21km	P473-P478 56km	P473-474 74km
Number of epochs	5760	5760	5758
Number of epochs with wrong integers	123	1226	1877
Number of DD phase measurements	72562	72452	72310
Number of wrong WL/NL integers	75/319	1158/4548	1671/7107
AR success rate WL/NL Overall (average)	99.79% 99.12% 99.32%	96.80% 87.45% 92.12%	95.38% 80.34% 87.86%
RTK availability	97.86%	78.66%	67.40%
RTK availability (0.025,0.025,0.05cm)	98.14% 99.10% 94.32%	79.62% 84.17% 78.92%	58.98% 75.22% 65.18%
RTK accuracy (STD in North East and Up (m)	0,008m 0.006m 0.025m	0.008m 0.007m 0.020m	0.009m 0.008m 0.031m

The above concepts and inter-relationship of the different parameters have been preliminarily demonstrated using the commercial HD-RTK2^{TM} RTK system. Statistical results from the GPS data collected for 16 static baselines over 2 to 45 kilometres and for 60 to 400 hours have confirmed its convincing performance of the system from different perspectives. Experimental results from 3 dual-frequency data sets have been analysed using the research version of the QUT-RTK software, showing AR performance improvement of using WL and NL observables with respect to the original L1 and L2 observables when the baselines exceed 20 kilometres.

References

Brown R G (1996) ***Receiver Autonomous Integrity Monitoring***, Global Positioning System: Theory and Applications, Vol II, published by American Institute of Aeronautics and Astronautics, Washington D. C, 1996, pp143-165.

Feng Y (2008) ***GNSS Three Carrier Ambiguity Resolution Using Ionosphere-reduced Virtual*** Signals, Journal of Geodesy, http://dx.doi.org/10.1007/s00190-008-0209-x.

Feng Y, Wang J (2007), ***Exploring GNSS RTK Performance Benefits with GPS and Virtual Galileo Measurements***, Proceedings of ION-NTM 2007, pp 218-226, San Diego, CA.

Higgins, M, (2007) ***GNSS, CORS and Positioning Infrastructure: Business and the Future***, Presentation at FIG Regional Conference Costa Rica November 2007.

HandyNav (2005) ***HandyNav RTK Testing Report,*** provided for licensed users, 2005.

Langley, R. B (1999) ***The Integrity of GPS***, GPS World Vo1 10, No 3, March 1999, pp60-63

Enge, P, Van Dierendonck, A J (1996) ***Wide Area Augmentation System,*** in Global Positioning System: Theory and Applications, Vol II, published by American Institute of Aeronautics and Astronautics, Washington D. C, 1996, pp117-142

Mirsa P, Enge P (2004) ***Global Positioning Systems, Signals, Measurements and Performance***, Ganga-Jamuna Press, pp 227-254.

Pratt, M., B Burke, and P Misra (1997) ***Single-Epoch Integer Ambiguity Resolution with GPS Carrier Phase: Initial Results***, Navigation, Vol 32, No 4, Pp386-400.

Positioning One consulting (2008), ***User Needs Analysis***, Technical Report for Cooperative Research Centre for Spatial Information, July 2008.

Teunissen, P.J.G. (1995) ***The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation***. Journal of Geodesy, Vol. 70, No. 1-2, pp. 65-82.

Teunissen, P.J.G, P J De Jonge and C.C. J.M Tiberious (1997). ***Performance of the LAMBDA Method for fast GPS Ambiguity Resolution***, Navigation, Vol 44, No 3, pp373-383.

Teunissen PJG, Joosten P, Tiberius C.C.J.M (1999) ***Geometry-free ambiguity success rates in case of partial fixing.*** Proceedings of ION-NTM 1999, pp 201-207, Jan 25-27, San Diego CA

Walter, T. (2002) ***Introduction to the Wide Area Augmentation System***, Journal of Global Positioning Systems 1(2), pp151-153.

Wang J. (2000) ***Stochastic modelling for RTK GPS/Glonass positioning***, Navigation, Journal of the US Institute of Navigation, 46(4), 297-305

Wang J., C. Satirapod & C. Rizos (2002) ***Stochastic assessment of GPS carrier phase measurements for precise static relative positioning***, Journal of Geodesy, 76(2), 95-104

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 9-17

Journal of Global Positioning Systems
2008
© CPGPS

A Pseudolite-Based Maritime Navigation System: Concept through to Demonstration

Dr. Charles S. Dixon and Mr. Russell G. Morrison
EADS Astrium, Anchorage Road, Portsmouth, PO3 5PU, United Kingdom

Abstract

GPS alone will be unable to meet emerging performance requirements for maritime applications with respect to service robustness, accuracy, integrity and availability. Even when Galileo (or indeed other nascent GNSS) eventually become operational there will be performance gaps. In particular, identified applications in port areas (for example automated docking) and in inland waterways, have very stringent performance requirements.

EADS Astrium Ltd has developed a Test Environment for a Maritime Navigation System in order to perform demonstrations for the maritime community. This includes a number of Transmit Stations mounted on existing structures and terrain at accurately known locations overlooking a navigation zone (such as a maritime harbour). These continuously transmit replica GNSS signals. Another key component of the Navigation System is the single Monitoring and Control Station. This monitors each transmitter's signal, calculates clock corrections and provides each Transmit Station with the contents of its navigation messages, emulating the function of the Galileo central control facility. It also controls and monitors any demonstration signal scenarios.

This paper presents details of Astrium's Pseudolite-based Navigation System that is the foundation of a Maritime Test Environment. As well as design details, the paper presents outcomes from a public demonstration undertaken in Oban, on the West Coast of Scotland in March 2008.

Key words: Pseudolite, Maritime navigation system, Transition station, GNSS

1. Introduction

The EC 6th Framework Project "MARUSE" was led by Kongsberg Seatex AS and managed by the GNSS Supervisory Authority. MARUSE has been described in a number of earlier publications and the interested reader is referred to those [1, 2, 3, 4]. This paper is based upon an earlier publication [5] but additionally presents specifics and findings from demonstrations undertaken in Oban, on the West Coast of Scotland in March 2008.

Maritime User Requirements are presented based on an analysis of emerging user needs conducted as part of project MARUSE. The Maritime Test Environment, completed by EADS Astrium, is then described in detail. This includes the critical Transmit stations and the Monitor and Control Station. Signal transmissions are also described and an outline of the demonstrations undertaken.

EADS Astrium the industrial Prime Contractor for development of the Navigation Payload of the Galileo Satellites. Navigation payload efforts at EADS Astrium are now focused on the development of four "In Orbit Validation" (IOV) spacecraft payloads that will prove all significant elements of the Galileo system. In parallel, development of ground-based GNSS transmitters, known as pseudolites has been taking place to augment GNSS coverage for technically challenging or difficult domains. Such domains include the infamous "urban canyon" and similar domains where GNSS signals may be partially or completely blocked by buildings or other obstructions.

A particular domain of interest is Maritime, where augmentation of GNSS coverage for certain harbour operations is seen as important to support efficient future operations, including automated harbour approach and automated docking. Emerging performance requirements for maritime applications with respect to accuracy,

integrity and availability require more from a GNSS System than can be supplied by GPS alone. Once it is fully operational, Galileo will greatly enhance performance. In order to guarantee uninterrupted high-accuracy and high-integrity navigation, further augmentation to GNSS may be necessary, for example through the use of pseudolites.

2. Maritime User Requirements

Emerging performance requirements for maritime applications with respect to accuracy, integrity, availability, and continuity are rather demanding, and indeed are considered beyond the capabilities of a GNSS System to support. Identified applications in port areas (for example automated docking) and in inland waterways, have very stringent performance requirements as shown in Table 1. The figures are based on an analysis prepared by the General Lighthouse Authorities of the United Kingdom and Ireland, contracting through the Corporation of Trinity House within Project MARUSE [6].

Table 1 Maritime Application Requirements (extract from [6])

	Navigation Phase			
	Ocean	**Coastal**	**Port**	**Inland Waterways**
Requirement	Ship-to-ship coordination	AtoN Management[1]	Automatic Docking	Law Enforcement
Accuracy	H: 10 m V: N/A	H: 1 m V: 1 m	H: 0.1 m V: 0.1 m	H: 1 m V: 1 m
Availability	99.8%	99.8%	99.8%	99.8%
Continuity	99.97%	N/A	99.97%	99.97%
Integrity	Yes	Yes	Yes	Yes

3. MARUSE Maritime Test Environment

EADS Astrium has developed a transportable Test Environment in which to perform navigation demonstrations as part of the MARUSE demonstrations for the maritime community. This includes:

- A number of Transmit Stations mounted on existing structures (cranes/gantries, berths, buildings, hills) at accurately known locations overlooking the demonstration zone (several kilometres in extent). These continuously transmit Galileo-like signals and also have the capability to transmit pulsed pseudolite signals.
- A single Monitoring and Control Station with line of sight visibility of each Transmit Station. This monitors each transmitter's signal, calculates clock corrections and provides each Transmit Station with the contents of its navigation messages, emulating the function of the real Galileo central control facility. It also controls and monitors demonstration signal scenarios.

The Test Environment is designed to facilitate demonstration of Galileo differentiators, such as accuracy, integrity, and availability in an environment that is realistic for the maritime application groups as considered in Table 1. In Figure 1 the configuration of the Test Environment is illustrated. This includes several synchronised Transmit Stations, a Monitor and Control Station, and the WLAN interconnecting them.

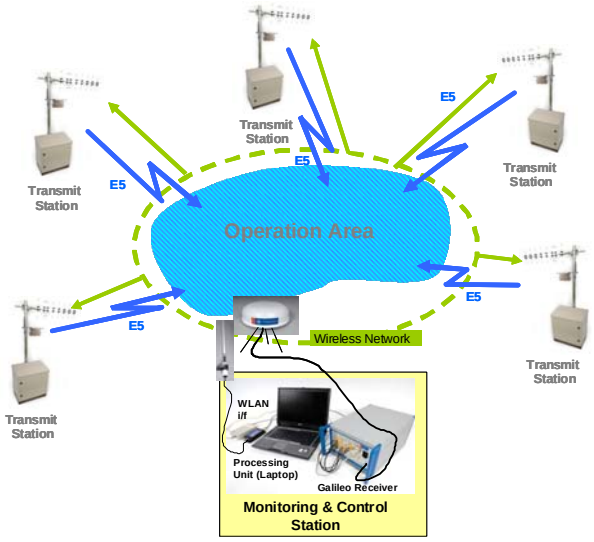

Fig. 1 Maritime Navigation Test Environment Configuration

A navigation solution may be obtained at a user receiver from a mixture of GPS and Pseudolite or Galileo signals by introducing a second time variable (i.e. local receiver clock offset from both GPS time and Galileo time). With the Test Environment Master Clock standing in for Galileo Time, there is an advantage in its synchronisation to GPS Time (i.e. removal of the second time variable in the receiver solution) but this is not strictly necessary for user navigation. A minimum of 5 signal measurements are required to obtain a user navigation solution with

[1] AtoN Management means management of Aids to Navigation (placement of buoys, etc.)

unsynchronised System Times, or 4 measurements with synchronised System Times. The Test Environment's Master Clock will for demonstration purposes be synchronised to GPS time.

The demonstration environment requires, in addition to the Transmitters, a minimum of two GNSS receivers, one for on-vessel applications and one to provide references to the M&C facility. These receivers will use conventional GPS L1 transmissions as well as the Test Environment signals.

The chosen Test Environment signal has characteristics as close to Galileo E5 Alt-BOC signal as possible. The navigation message content and receiver functionality have been adapted to cope with the Transmit Stations being stationary, rather than in a MEO orbit. In addition the Transmit Stations will be mounted at low elevations. This will limit the accuracy of any height determination and may result in multipath effects that must be handled by the receiver. The close transmit-receive range will result in larger than normal received power level variations although this will be minimised by care in setting the output power levels, and in positioning of the transmitters in relation to the demonstration zone.

For any service provision demonstration to be realistic, navigation solutions obtained from the Test Environment must contain errors representative of those expected from the true Galileo system. Operational Galileo errors are expected to originate primarily from spacecraft orbit and clock prediction errors (ephemeris), and from ionosphere and troposphere propagation variation not corrected by the use of two frequencies or imperfectly modelled. Orbit errors are effectively eliminated from the test environment transmissions through precise survey of each of the transmitter locations. Residual uncertainties should be at the centimetre level. In addition, propagation errors will be minimal because of the small spatial extent of the covered area. Clock errors will however be somewhat larger than their Galileo equivalents since the Test Environment uses ovenised crystal oscillators (OCXO) instead of the atomic clocks used in GPS (Rubidium and Caesium) and Galileo (Hydrogen Maser and Rubidium). The update rate for clock corrections parameters in the test environment system will therefore be necessarily higher than for GNSS systems. A baseline rate of once per 50 seconds has been established, this value maximising correspondence with the Galileo message frame structure [6] on the E5a and E5b carriers.

All of the types of error discussed above lead only to bias and uncertainty in pseudorange measurements at user level, so their origin does not alter their impact. The combined impact of such errors can thus be simulated by the M&C facility introducing deliberate errors in the clock correction parameters sent to each transmitter for inclusion in the navigation data. The testbed can also emulate changes in which "Spacecraft" are visible by changing the PRN code used by a particular transmitter. The M&C facility can similarly emulate "Spacecraft" degradation or outage.

3.1 Transmit Stations

The Transmit Stations are a variation of Pseudolite technology. The Transmit Stations generate and broadcast navigation signals based on those planned for use by the future Galileo System.

Transmit Stations are compact and self-contained, and produce continuous GNSS satellite-like transmissions at a source level that is preset and maintained constant for the duration of any demonstration. Transmitters can either work independently, or can be synchronised in groups to work together as a single, coherent system. The design places few restrictions on the location of their deployment, ensuring that demonstration planners have the freedom to create optimal signal environments for maritime or other sector application demonstrations.

Figure 2 illustrates the Transmit Station architecture. It comprises four distinct parts: a timing and frequency generation section, a radio frequency (RF) section, a real-time firmware section including an FPGA device and control and a "soft real-time" software section hosted on a central processing unit (CPU).

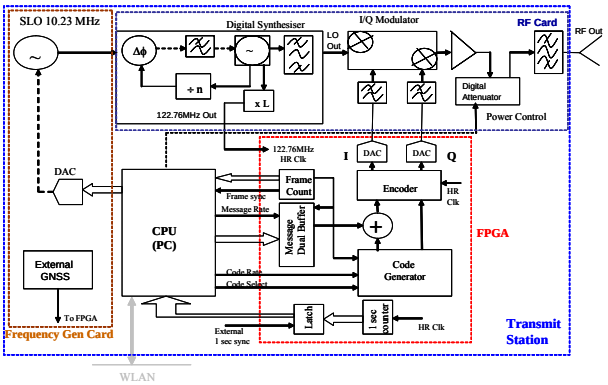

Fig. 2 Transmit Station Architecture

Each of the Transmit Stations within the Demonstration Test Environment is designed to perform the following functions:

- Accept commands from remote source (M&C Station) via wireless network, and continue to operate usefully in absence of external commands;
- Maintain stable local timing reference, and synchronise to external standard when available;
- Generate coherent broadcast carrier frequency;
- Generate PRN spread spectrum code and modulate carrier;
- Generate navigation message and modulate the PRN frames;
- Amplify, filter and broadcast resultant Galileo-like navigation signal at required source level;
- Restrict out-of-band transmissions;
- Maintain physical security and protect system from external environment;
- Be capable of working from internal power supply;
- Record status data.

Hardware Implementation

To minimise risk, commercial off-the-shelf (COTS) hardware was used where possible. The RF circuitry is based on an existing Astrium modulator/upconverer, with modified filters to meet the required RF performance.This module has a cPCI format, which dictated the form of the other equipments. The CPU and FPGA are PMC modules, mounted on a dual PMC/cPCI carrier. The FPGA module has a dual high-speed DAC interface card fitted. All are commercial modules, selected for performance and minimum power requirements.

The modules are mounted in a commercial 2-slot 19-inch cPCI rack, with a 12v dc power supply. An internal Ethernet LAN provides system connectivity. Figure 3 illustrates the Transmit Station main electronics package as realised in hardware.

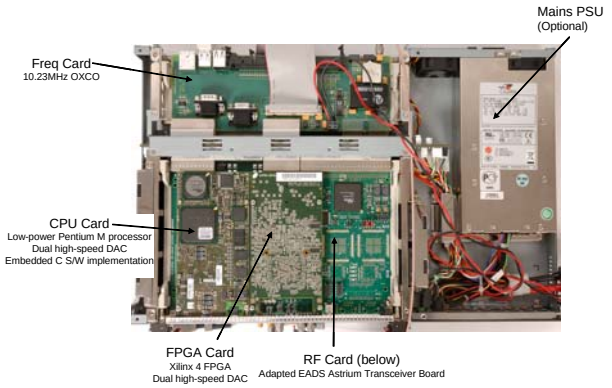

Fig. 3 Transmit Station Main Electronics

Two boards were designed specifically for the EADS Astrium Pseudolite. These connect to the rear of the cPCI rack, using existing connections. They contain the oven-controlled crystal oscillator (OCXO), plus circuitry to adjust its frequency; test interfaces and interconnection between the RF board and the processor. Various test/debug interfaces are also fitted.

The designated "master" pseudolite also has an EGNOS receiver fitted to one of the boards, to provide a low-jitter 1pps reference for oscillator frequency control.

External to the 19-inch rack are the wireless LAN components (RF Modem and Ethernet Bridge) and power-switching components (relays and switches for manual control, plus an iBoot switch controlled by the Ethernet Bridge).

The system is powered by a commercial 12v lead-acid battery, and optionally by a mains power supply.

The electrical equipment is housed in a weatherproof cabinet, providing protection to IP66 (powerful water jets).

The RF Navigation antenna is a corner reflector mono-pole, and the WLAN communications antenna is a Yagi. Both are mounted externally.

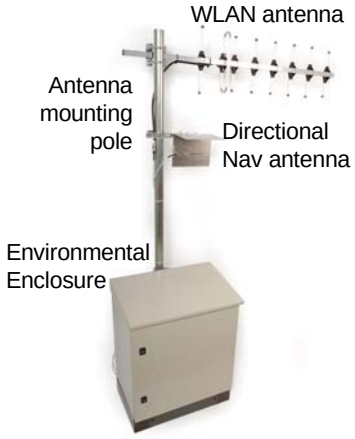

Fig. 4 Transmit Station Physical Realisation

Pseudolite RF Transmitter

The Pseudolite transmitter is based on a single stage single sideband up-conversion process and attractive, particularly in comparison with alternative double sideband schemes incorporating filters and frequently comprising multiple stages. Although of greater

complexity such schemes are well tried and can consistently deliver high spectral performance.

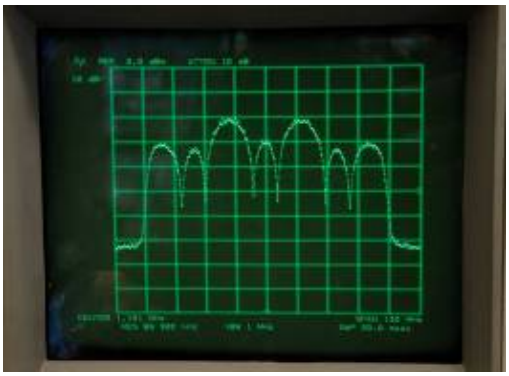

Fig. 5 PSL RF Transmitter Output Spectrum

The SSB output is generated by means of a single chip complex up-converter. This device incorporates two single stage double sideband suppressed carrier mixers whose local oscillators are in phase quadrature. The baseband input signals, also in phase quadrature, are up-converted by these mixers. Their outputs are combined to produce either an upper or lower sideband signal depending on the lead/lag relationship between the baseband inputs. The remainder of the circuitry is concerned with the creation and control of these baseband signals which will maximize the up-conversion process.

The main characteristics of the Pseudolite RF transmitter are tabled below:

Table 2 (Astrium) Pseudolite RF Transmitter Characteristics

Characteristic	Value	Units
Transmit Power Range	0 to -55	dBm
Programmable Synthesiser	1191.795	MHz
Programmable Attenuation Range	55	dB
Input Frequency Reference (Fref)	10.23	MHz
Synchronised Output Clock Facility (HR Clk)	12x Fref = 122.76	MHz
Output Clock Level	14 ± 1dB	dBm

3.2 Monitor and Control Station

The overall role of the Monitoring & Control (M&C) Station is to coordinate the operation of the Transmit / PSL Stations so that they form a coherent navigation system.

The Monitor and Control Station (M&C) Station within the Demonstration Test Environment is designed to perform a number of key functions.

At the heart of the M&C is a processing element which provides the ability to accept and process data from the GNSS Reference Receiver via a data link. The data being received from the GNSS Reference Receiver will include measurements of Pseudorange, Carrier Phase, Doppler, C/No and lock time; data such as raw Navigation symbols and tracking & receiver status.

The M&C processing element also provides the mechanism for generating commands for dissemination over the communications link to each Transmit Station for passing of navigation parameters and miscellaneous data. Linked with this is the mechanism for accepting and processing acknowledgements and other health & status data from each of the transmit stations received via the communications link.

A User interface is provided to enable inputting of parameters, obtaining system status general system housekeeping as well as functionality for storage of Status data on appropriate media. An example M&C screen is given in Figure 6 below.

The M&C processing element also provides the low-level algorithms for timing within the system. It maintains the master timing reference, by synchronising to external standard. Also included are the algorithms for determining timing offsets (ΔT) to be sent to each Transmit Station. Timing algorithms were described in a previous paper [5].

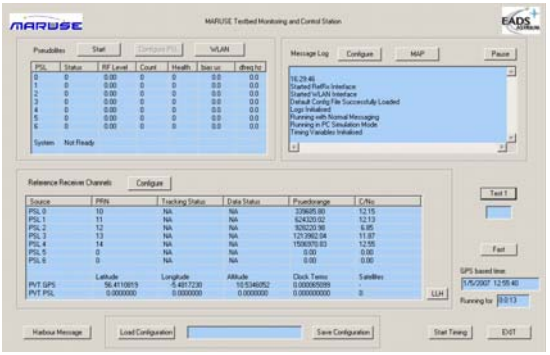

Fig. 6 M&C Screen User Interface

3.3 Galileo Reference Receiver

A Septentrio GeneRx Receiver is used as the Pseudolite Reference Receiver for this work. This unit is capable of tracking GPS satellites as well as the Galileo-like transmissions of the pseudolites. Its main functions are to measure Pseudoranges from the transmit stations, to collect the transmitted Navigation Data, to time-stamp this

data and to relay it to the M&C via a dedicated RS-232 serial link.

3.4 Communications Link

The Communications Link interconnects the M&C Station with each of the Transmit Stations at there remote locations. The communications link utilises medium to long range Wireless Local Area Network (WLAN) capabilities. For the prototype equipment in Project MARUSE, X8200 Radio Modems are used. These RF modems can achieve a range of up to 20km in free space and the RF power can be varied from 5mW to 500mW. Serial data can be transmitted by X8200 with baud rates from 1.2K to 115.2K over distances of 10km to 20km line of sight, and can operate in both license-exempt and licensed bands. Antennas used are 10dB Yagi at the Transmit Stations and 0dB omnidirectional antenna at the M&C.

For an operational system this link type might be appropriate, although for many applications communications links may already exist which could meet the relatively modest communications demands of the system.

4. Test Environment Signals

The Signals transmitted by the Transmit stations are replica Galileo signals based on the E5A and E5B AltBOC-modulated transmissions [7]. Differences from the Galileo transmissions are described below.

In the Test Environment, there will be no ionosphere between the Transmit Stations and the Receivers; it is therefore of limited value to use multiple transmission frequencies in a terrestrial pseudolite environment. The transmissions used for project MARUSE are identical in terms of frequency and modulation to the Galileo transmissions. This comprising two pairs of I&Q components referred to as E5A and E5B, which are AltBOC modulated onto an RF carrier centred at 1191.795 MHz.

The wide band E5 signal shall be generated with the AltBOC modulation of side-band sub-carrier rate $R_{S,E5}$ = $1/T_{S,E5}$ = 15.345 MHz (15 x 1.023 MHz) according to the expression [7]:

$$
\begin{aligned}
s_{E5}(t) = & \frac{1}{2\cdot\sqrt{2}}\cdot\left(e_{E5a-I}(t)+j\cdot e_{E5a-Q}(t)\right)\cdot\left[sc_{E5-S}(t)-j\cdot sc_{E5-S}\left(t-T_{s,E5}/4\right)\right]+ \\
& \frac{1}{2\cdot\sqrt{2}}\cdot\left(e_{E5b-I}(t)+j\cdot e_{E5b-Q}(t)\right)\cdot\left[sc_{E5-S}(t)+j\cdot sc_{E5-S}\left(t-T_{s,E5}/4\right)\right]+ \\
& \frac{1}{2\cdot\sqrt{2}}\cdot\left(\bar{e}_{E5a-I}(t)+j\cdot\bar{e}_{E5a-Q}(t)\right)\cdot\left[sc_{E5-P}(t)-j\cdot sc_{E5-P}\left(t-T_{s,E5}/4\right)\right]+ \\
& \frac{1}{2\cdot\sqrt{2}}\cdot\left(\bar{e}_{E5b-I}(t)+j\cdot\bar{e}_{E5b-Q}(t)\right)\cdot\left[sc_{E5-P}(t)+j\cdot sc_{E5-P}\left(t-T_{s,E5}/4\right)\right] \quad .
\end{aligned}
$$

with the binary signal components e_{E5a-I}, e_{E5a-Q}, e_{E5b-I} and e_{E5b-Q} (consisting of the BPSK modulated ranging codes multiplied with the respective navigation data stream for the data-channels).

The main characteristics of the MARUSE E5 transmitted signal are highlighted in the table 3.

Table 3 Signal Characteristics

Characteristic	**Details**
TX Signal Bandwidth	92 MHz
TX Centre Frequency	1191.795 MHz
Signal	E5 (E5a & E5b)
Service	I/NAV, F/NAV
Modulation	AltBOC
E5A Chip rate	10.230 MChip / s
E5B Chip rate	10.230 MChip / s
E5A-I Symbol Rate	50 symbols / s
E5B-I Symbol Rate	250 symbols / s
E5A-I Code period	20 ms
E5A-I Primary Code length	10230 (chips)
E5A-I Secondary Code length	20 (chips)
E5A-Q Code period	100 ms
E5A-Q Primary Code length	10230 (chips)
E5A-Q Secondary Code length	100 (chips)
E5B-I Code period	4 ms
E5B-I Primary Code length	10230 (chips)
E5B-I Secondary Code length	4 (chips)
E5B-QCode period	100 ms
E5B-Q Primary Code length	10230 (chips)
E5B-Q Secondary Code length	100 (chips)
Adapted for TX Stations:	
Navigation Framing	
Navigation Messages	Almanac
	Ephemeris
	Timing
	GPS Offset
Other Message Types	Low Bit Rate Communications message

5. Oban Demonstrations

Tests and live demonstrations took place in Oban on the West coast of Scotland, on 4th March 2008. Observers from around Europe saw the NLV Pharos undertake a variety of precision manoeuvres in the bay. These

included placement of a new wreck-marking buoy. Crucial to the success of the demonstration were a number of new systems developed under the MARUSE project, including innovative onboard dynamic positioning technologies and an array of EADS Astrium pseudolites which mimicked GALILEO satellite signals. The pseudolites were placed in various locations around Oban Bay, enabling a vessel to use GALILEO-like signals for the first time to navigate in British waters.

The demonstration in Oban was the fourth in a series undertaken by MARUSE. Participants watched the NLV Pharos in real time via closed-circuit television as it manoeuvred using the EADS Astrium pseudolites. The demonstration was hosted by the Research and Radionavigation Directorate of the General Lighthouse Authorities of the UK and Northern Ireland (GLAs), which include the Northern Lighthouse Board, the Commissioners' of Irish Lights and Trinity House.

Major aims of the demonstrations are summarised in Table 4, and these included demonstration of Galileo-based navigation in advance of Galileo deployment, accuracy characterisation of the equipment, and interoperability confirmation, both of GPS and Galileo, and of these systems interoperating with pseudolite augmentation.

Table 4: Demonstration Technical Items

ITEM	DETAIL
Navigation using Galileo-like signals	E5a, E5b, E5 AltBOC Code-phase solution
Accuracy	Required <5m; Desirable ~1m
Interoperability	(a) Galileo + GPS (b) Interoperability + PSL

5.1 Demonstration Scenario

The public demonstration undertaken in Oban was based on the concept of a vessel blocking a busy waterway. It used a combination of Pseudolites, GPS and Virtual Aids to Navigation (AtoN) (making use of AIS and Electronic chart displays (ECDIS)). The demonstration also used NLV Pharos vessel from Northern Lighthouse Board to undertake the physical demonstrations within Oban bay. The scenario for the demonstration was as follows:

- A simulation of a vessel entering the Oban waterway (using AIS to show its position) is shown on an ECDIS electronic chart display.
- The simulated vessel suffers a failure causing it to remain in the waterway as an obstruction
- The position of this vessel is marked by the General Lighthouse Authorities (GLAs) using the virtual AtoN (AIS, ECDIS) and its position is broadcast via AIS to all nearby vessels.
- The NLV Pharos is deployed to physically mark the vessel with a Buoy, over the same location as that provided by the Virtual AtoN, using the enhanced positioning from the EADS Astrium Pseudolites.

The demonstration scenario is shown in following figure:

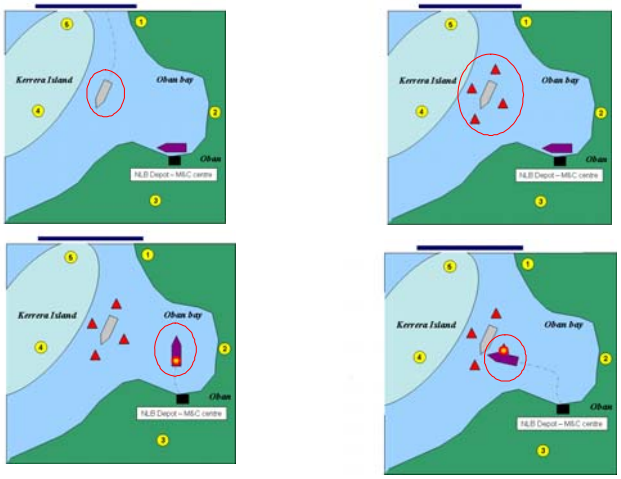

Fig. 7 Demonstration Scenario (courtesy of Dr Alan Grant General Lighthouse Authorities of the United Kingdom and Ireland)

Fig. 8 NLV Pharos deploying a wreck marker buoy during Oban Demonstration

6. Mission accomplished

The MARUSE demonstrations undertaken in Oban, which involved the placement of virtual buoys and the retrieval and placement of a real emergency buoy over a simulated wreck in Oban Bay, were considered a success by the GSA.

The role of the MARUSE pseudolites in this demonstration was to mimic future GALILEO signals, but they can also serve as stand-alone infrastructure elements. "These pseudolites can be very useful in filling gaps in current GNSS services," said MARUSE project manager Stig Erik Christiansen of Norway's Kongsberg Seatex, "especially in problem areas like 'urban canyons' or deep locks in our inland waterways. And, at sea, this kind of technology can help to reduce the potentially terrible risk of collision between tankers, for example, ultimately reducing operating costs by increasing the precision and efficiency of navigation."

7. Conclusions

A Maritime Navigation Test Environment has been created by EADS Astrium and its partners as part of Project MARUSE. This will facilitate testing and demonstration of two distinct benefits to future shipping. Firstly, Galileo differentiators (more visible sources of transmission, higher accuracy, integrity) have been demonstrated in advance of full operation of the Galileo System. This has helped to introduce the benefits of GNSS in general, and Galileo in particular to the Maritime domain, whilst simultaneously soliciting feedback that can be used to modify the final implementation of Galileo.

Secondly, pseudolite operation has been demonstrated. Pseudolites may be appropriate to augment GNSS performance in critical situations or difficult environments. Critical situations include those where safety of life and/or continuity of high-quality system performance are paramount. Difficult environments include situations where GNSS signals may become blocked by local obstructions such as harbour gantry cranes and other local obstructions.

The EADS Astrium approach has been to produce a sophisticated yet low-cost Test Environment within which Galileo receivers and other user equipment and applications can be tested. The Environment is focused on the demands of the maritime domain, but is also flexible and moveable facilitating re-use for other applications and domains.

Demonstrations have been undertaken in Oban on the West Coast of Scotland in spring 2008 as part of Project MARUSE. This used the pseudolite transmission stations and M&C equipment developed by EADS Astrium and described in this article, Receivers from Septentrio, User Terminals from Kongsberg Seatex, and Vessels from Trinity House Lighthouse Authority.

8. Future Work

Future adaptations of the Pseudolite and M&C Equipment for other domains are foreseen. This may include Emergency Management, Rail, Road, and Security-related domains.

The Pseudolite-based Transmit Stations and M&C Equipment have been added to the EADS Astrium product range. These can augment GNSS performance in critical situations or difficult environments. Critical situations include those where safety of life and/or continuity of high-quality system performance are paramount. Difficult environments include situations where GNSS signals may become blocked by local obstructions, and encompass urban canyons and indoor operations among other environments.

Acknowledgements

The authors would like to thank the European GNSS Supervisory Authority (GSA) for their financial support to the MARUSE project; it is largely upon the developments ongoing in that project that this paper is based. It should be pointed out that the views and opinions expressed throughout are those of the authors, and as such do not necessarily represent those of GSA.

The authors wish to acknowledge support and encouragement from the Managements at EADS Astrium. In addition, thanks are due to partners in Project MARUSE, and in particular to Stig Erik Christiansen of Kongsberg Seatex AS, to Dr Alan Grant of Trinity House Lighthouse Service, and to Alain Suskind and Axel Van Den Burg of Septentrio.

The authors would like to acknowledge the generous support of individuals and organisations that helped with, and hosted, the pseudolite equipment in and around Oban. Particular thanks are due to many members of the NLB, to Mrs Aileen Miller, to the very Reverend Donald Mackay and to Argyll & Bute Council.

References

[1] Grant, A.J., N. Ward, S.E. Christiansen, C.S. Dixon, A. Suskind; ***MARUSE Demonstrating the use of Maritime Galileo Pseudolites***. Accepted for ION GNSS 2007 Conference, 25-28 September, 2007 – Fort Worth, TX, USA

[2] Christiansen, S.E., C.Amlacher, G.Burden, C.S. Dixon, A.J. Grant, S Mazzeo, A. Suskind; ***Introduction of Galileo in Maritime and Inland Waterway Applications***. ENC GNSS'07 / TimeNav'07, 29-31 May 2007, Geneva, Switzerland

[3] Dixon, C.S., R.G. Morrison; ***Maritime Demonstration Testbed for Galileo***. ION GNSS 2006 Conference, 26-29 September 2006, Fort Worth TX, USA

[4] Dixon, C.S., R.G. Morrison and S.E. Christiansen; ***Testbed for Galileo Maritime Demonstration***. ENC GNSS 2006 Conference, 7-10 May 2006, Manchester, United Kingdom.

[5] Dixon, C.S., Bolt, M., Helliwell, A., Lau Semedo, P., Morrison, R.G., Parks, M., Whittaker, A., Williams, M. ***Demonstration of a Galileo Based Maritime Navigation System*** RIN NAV 07 30th October – 1st November 2007

[6] ***MARUSE Application Group Consolidation – Project Document: MARUSE-DD06-THL-AGC-R000***, Final Version dated 17th February 2007.

[7] ***Galileo Open Service Signal-In-Space Interface Control Document (SIS ICD).*** GAL OS ICD/D.0 Issued 2006 Draft 0

[8] Dixon, CS and RG Morrison; ***Demonstration of a Galileo-Based Maritime Navigation System: First Demonstration Results***. ENC GNSS 2008 Toulouse, April 2008

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 18-26

Journal of Global Positioning Systems
2008
© CPGPS

On outdoor positioning with Wi-Fi

Binghao Li, Ishrat J. Quader, Andrew G. Dempster
School of Surveying and Spatial Information System, UNSW, Australia

Abstract

Though GPS is the most popular positioning system at present it does not perform well in indoor environments and metropolitan city areas. Wi-Fi positioning has received much attention due to its advantages with respect to indoor positioning and the wide spread of the Wi-Fi access points (APs). Its performance in an outdoor environment is also of interest as a Wi-Fi based positioning system can overcome the shortcomings of GPS. In this paper, the Wi-Fi positioning technologies which can be used in outdoor environments - trilateration and fingerprinting, are discussed. An experiment based on fingerprinting has been carried out in the Sydney CBD area where Wi-Fi APs are densely deployed. The test results show that the Wi-Fi positioning system based on fingerprinting works well for outdoor localization, especially when directional information is utilized.

Key words: Wi-Fi, Outdoor Positioning, Fingerprint

1. Introduction

GPS is the fully functional satellite based positioning system at present. It is widely used in many fields around the world. However it has its shortcomings such as it takes some time to get the first fix; it does not perform well indoors or in urban canyons. GPS receivers need to "see" at least 3 satellites which are relatively well distributed in the sky to calculate its 2D position. Hence in environments where the sky is blocked, positioning becomes difficult; even impossible. In indoor environments, some alternative systems, like active badge (Want et al., 1992), cricket (Priyantha et al., 2000) etc., have been developed for positioning, but they cannot be used widely due to their inherent problems. Outdoors, mobile phone networks (3GPP, 2004), television signals (Eggert and Raquet, 2004) and pseudolites (Barnes et al., 2006) can be utilized. Since 802.11 Wireless LAN (also known as Wi-Fi) technology has been widely utilized, a large number of access points (APs) have been deployed both indoors and outdoors. Wi-Fi positioning technology has attracted much attention from both researchers and companies (Ladd et al., 2002; Youssef and Agrawala, 2005; Li et al., 2005; http://www.ekahua.com/).

Wi-Fi aims to provide local wireless access to fixed network architectures. Its market is growing rapidly as the flexibility, connectivity, mobility, and low cost of this technology meet the needs of consumers. A group of specifications has been ratified by the IEEE 802.11 working group. Of these, 802.11b has become the industry standard. It operates at rates up to 11 Mbps in the 2.4 GHz band, which is the only accepted Industrial, Scientific and Medical band available worldwide (Bing, 2002). The next mainstream Wi-Fi standard is 802.11g which provides optional data rates of up to 54 Mbps, operates in the same band as 802.11b, and requires backward compatibility with 802.11b devices (Geier, 2002).

Obviously, Wi-Fi is not designed or deployed for the purpose of positioning. However, the measurements of signal strength (SS) of the signal transmitted by either AP or station imply the possibility of finding the location of the mobile user (MU). In fact, several SS based techniques have been proposed for location estimation in indoor environments in which Wi-Fi is deployed (Bahl and Padmanabhan, 2000; Li et al., 2005). There are essentially two categories of such techniques. One uses a signal propagation model and information about the geometry of the building to convert SS to a distance measurement. 'Trilateration' can then compute the location of the MU (Li, 2006). This approach is simple to implement; however it does have difficulties in building a sufficiently good model of signal propagation that is adequate for real world applications since so many factors affect the signal propagation. The other category of Wi-Fi positioning is 'Location Fingerprinting'. This class of technique has received more attention recently as it is able to address some of the problems related to non-line-of-sight (NLOS) and multipath propagation (Haldat, 2002). The basis of location fingerprinting is first to establish a database that contains the measurements of wireless signals at some reference points (RPs) in the area of Wi-Fi coverage. Then the location of the MU can be identified by comparing its SS measurements with the reference data (Ladd et al., 2002; Li et al., 2005). The disadvantages of this approach are the database

generation and maintenance requirements. Other methods based on different measurements rather than SS have been proposed, such as utilizing time-of-arrival (TOA), time-difference-of-arrival (TDOA), angle-of-arrival etc. The Wi-Fi positioning solution provided by WhereNet is an example. TDOA measurements are used in the position computation (http://www.wherenet.com). Hence WhereNet cannot use standard Wi-Fi devices.

There are several Wi-Fi positioning systems developed such as 'RADAR' from Microsoft Research (Bahl and Padmanabhan, 2000); 'Horus' from University of Maryland, USA (Youssef and Agrawala, 2005); 'ipos' of IMST GmbH, Kamp-Lintfort, Germany (http://www.centrum21.de/) and 'WPS' from the University of New South Wales, Australia (Wang et al., 2003). Commercial versions of Wi-Fi positioning systems are very rare. The best-known system is the 'Ekahau Positioning Engine (EPE)', offered by Ekahau Inc. (http://www.ekahua.com/products/positioningengine/). 'Pango' is a similar system which is part of innerwireless now (http://www.innerwireless.com/vision-over.asp). All these systems are basically focused on indoor positioning. Skyhook Wireless has developed a metro-area positioning system which can be used for outdoor positioning (http://www.skyhookwireless.com).
While many reports about Wi-Fi positioning in indoor environments are available, its performance outdoors is also of interest. In the following section, the possible outdoor positioning technologies using Wi-Fi are discussed. Then an outdoor test based on fingerprinting is introduced in section 3 and 4. In section 5, the results are reported. Finally, concluding remarks are given.

2. Outdoor Positioning Technologies Using Wi-Fi

In an outdoor environment, both trilateration and fingerprinting can be utilized. Furthermore, other positioning systems could also be valid, such as GPS. Hence the integration of Wi-Fi and GPS (or other systems) is also a choice.

2.1 Trilateration

The trilateration approach is relatively simple. Three base stations (or more) with known coordinates are required (refer to Figure 1). If the distance R from the base station to a MU can be measured, a circle with radius R can be drawn. Circles intersect at one point which is the MU's position. The coordinate of the MU can then be easily calculated. Generally, trilateration is used if the TOA measurement (which can be easily converted to distance) can be obtained, for example in mobile phone positioning. However, the measurements available in Wi-Fi are SS rather than the distance. Hence, the SS should be converted to distance first. So, the trilateration approach consists of two steps: the first step, using a signal propagation model to convert SS to AP-MU separation distance; the second step, least-squares or other methods (such as a geometric method) can be used to compute the location. The first step is the key of this approach.

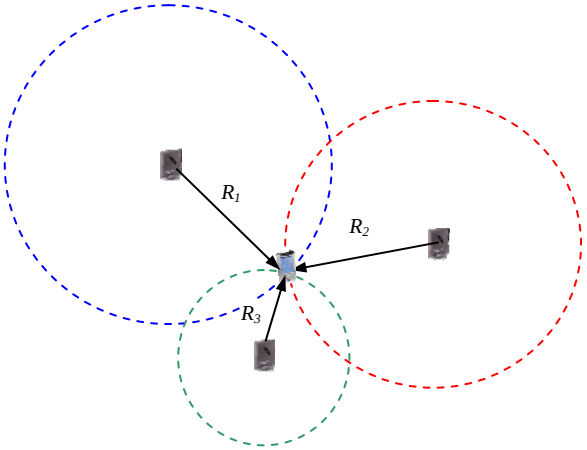

Fig. 1 Trilateration approach using Wi-Fi

Since the environments vary significantly from place to place, if the detail of the environment cannot be obtained easily, the simplest way to find the relationship of SS and AP-MU separation distance is by collecting some SS data at some points with the known coordinates. This means an extra procedure, called a learning procedure, has to be added to the trilateration approach (actually, the learning procedure is almost inevitable no matter what approach is used). The application of this approach in indoor environments has been investigated (Bahl and Padmanabhan, 2000; Li et al., 2005). Skyhook claims their system uses 'triangulation' (more accurately 'trilateration') to determine the MU's position in outdoor (http://www.skyhookwireless.com/whoweare/faq.php).

There are two types of errors associated with this approach: signal propagation model error and NLOS error (for details refer to Li et al. (2004)). Those two errors are not trivial. Hence the estimate of the MU's position may be very inaccurate. However, if the NLOS contaminated measurement can be detected and removed and the propagation model can be well chosen, the trilateration approach may be a good choice.

2.2 Fingerprinting

The word 'fingerprint' here denotes the location-sensitive parameters of measured radio signals: in the case of Wi-Fi this is SS. Similar to human fingerprints, the fingerprint of a specific place can be used to identify the location. The key idea of the fingerprinting approach is to map location-sensitive parameters of measured radio signals in areas of interest.

Location fingerprinting has two phases: 'training' and 'positioning' (Li et al., 2005). The objective of the training phase is to build a fingerprint database. In order to generate the database, RPs must first be carefully selected. Generally, the data acquired are the SSs measured by the MU. Locating a MU at one RP, the received SSs of all the APs are measured. From such measurements the characteristic features of that RP are determined, and are then recorded in the database. This process is repeated at another RP, and so forth until all RPs are visited. In the positioning phase, the MU measures the SS at a place where it requires its position. The measurements are compared with the data in the database using an appropriate search/matching algorithm. The outcome is the likeliest location of the MU. The whole process is illustrated in Figure 2.

The fingerprinting approach has been accepted as an effective method for Wi-Fi positioning, despite having some disadvantages. There are in fact two ways to estimate the unknown location. The simpler is the deterministic method (Bahl and Padmanabhan, 2000; Saha et al., 2003; Li et al., 2006). The average SS, which is taken over several measurements, of each Wi-Fi AP measured at each RP is used to create the fingerprint database. Since the variation of the SS measured at each point is large, in order to achieve more accurate results, the probabilistic approach (Roos et al., 2002; Li et al., 2006) has also been developed. Unfortunately, the distribution of the SS is non-Gaussian. Even worse, it varies at different locations, and at the same location when the orientation of the antenna changes (Ladd et al., 2002; Li et al., 2007). Hence many measurements are necessary, and this takes more time to generate the SS distribution at each RP. Furthermore, this increases the database size and the computational burden. However, the establishment of the location fingerprint database is an essential prerequisite. To achieve a good estimate of user location, the more RPs, or in other words, the smaller the granularity, the better. And since the measured SS is affected by so many factors, the variation of the received SS at each point can be as large as 10dB to 15dB. Therefore, the more measurements obtained at each point the better. However, more RPs and more measurements mean that the training phase is a significant task in terms of labour and time.

Many researchers have noticed the impact of the MU's orientation - the direction in which the receiver's antenna is pointing. When the user changes orientation, the received SS can change significantly (Ladd et al., 2002; Li et al., 2005). Xiang et al. (2004) and Li et al. (2007) investigated the methods using directional information to improve the positioning accuracy and estimate the MU's direction in an indoor environment.

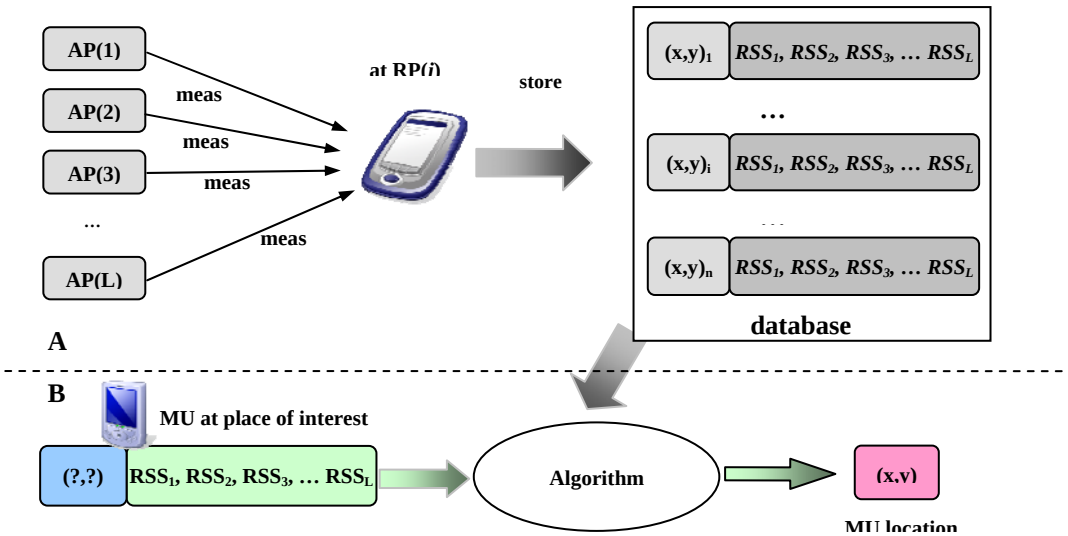

Fig. 2 Two phases of fingerprinting approach, A) training B) positioning

2.3 Wi-Fi plus GPS (Other Sensors)

Outdoors, the probability of receiving a line-of-sight signal from at least one GPS satellite is quite high. It is possible to integrate Wi-Fi and GPS to estimate the MU's position. However, as mentioned previously, Wi-Fi cannot provide the TOA measurement. Using a model can convert SS to AP-MU distance, but this distance is very inaccurate. The fingerprinting approach cannot provide range measurements. So, how to integrate Wi-Fi and GPS is an interesting problem to be investigated. Integration of Wi-Fi with other sensors is also of great interest.

3. Test bed and Equipment

To investigate the performance of Wi-Fi positioning in an outdoor environment, the Sydney CBD was chosen to carry out the test. This area is well serviced with Wi-Fi signals. The test area has a dimension of about 500m by 800m. More than 1300 APs can be detected in this area. Some of them are deployed by telecommunication companies such as Telstra as part of their fixed infrastructure. Some APs are established by private users for personal use. The chosen area has a typical urban setup with tall buildings and towers blocking the sky. Figure 3 shows the test area.

Fig. 3 Sydney CBD area where test was conducted (www.airviewonline.com.au)

During the experiment researchers used a Compaq iPAQ 3970 personal digital assistant (PDA) running the Pocket PC 2002 operating system (http://www.compaq.com). A Wireless card from Lucent Technology Wi-Fi Orinoco Wireless Golden Card (http://www.orinocowireless.com) has also been used. For comparison purpose, a Garmin eTrex GPS receiver has been utilized to collect GPS data (Figure 4).

Fig. 4 Equipment used in the test (Wi-Fi card, PDA and GPS receiver)

The software used to collect and preliminarily process SS data is NetStumbler and a pruned version for the PDA called MiniStumbler (http://www.netstumbler.com). Screenshots of the user interface are shown in Figure 5.

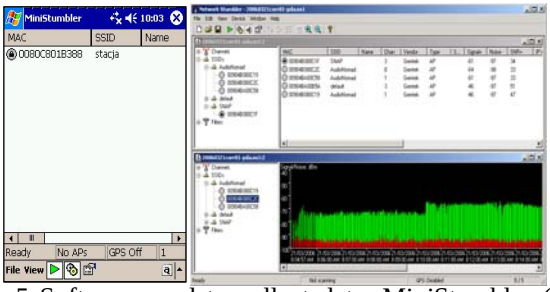

Fig. 5 Software used to collect data, MiniStumbler (for PDA); software used to preliminarily process the data, NetStumbler (for PC)

4. Methodology and Data Collection

As introduced in section 2, the trilateration approach is the simplest to use in outdoor positioning. However, apart from the difficulties which have been discussed previously, there is another issue that must be considered seriously – the coordinate of the APs. It is not easy to obtain this information. The reality is the operators of these fixed infrastructures won't give the coordinates free (or if they do not want this information to be public at all) while the owners of the private APs do not even know the coordinates. To detect these APs (mainly the APs in fixed infrastructure) and measure their coordinates becomes the preliminary requirement for use of the trilateration

approach. Skyhook's major contribution is to collect this information and their data are in confidence. So, the simple approach becomes hard to carry out. Furthermore, it is very unlikely to obtain better results than using another method. Hence the trilateration approach was not considered for this test.

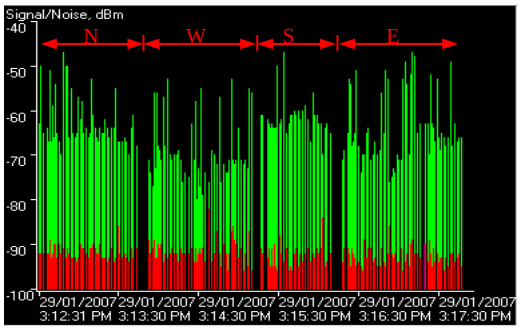

Fig. 6 The correlation between the received SS and the orientation of the MU

AP coordinates are not required in the fingerprint approach. Moreover all of the APs (both the APs as fixed infrastructure and the private users' APs) signal can be utilised for positioning as the private user's APs are unlikely to move far away inside the shop or apartment. Similar to the indoor environment, a direct correlation between the received SS and the orientation of the MU can be observed. Figure 6 gives an example. The data were collected at the south-west corner of the intersection of George Street and Martin Place. The AP is a Telstra CBD hot spot with the MAC of 0011209C1BC0. It implies that directional information may also be able to be used to improve the positioning accuracy and possibly estimate the MU's orientation. Figure 7 depicts the two ways to generate the database: the traditional way and the way to utilize the directional information (for details refer to Li et al. (2007)).

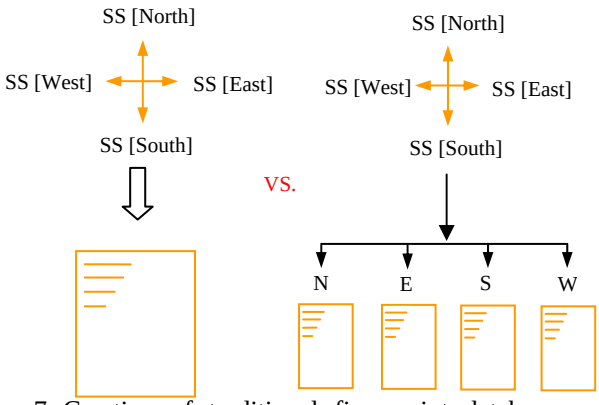

Fig. 7 Creation of traditional fingerprint database vs direction-based fingerprint database (four directions)

In total, 172 RPs and 23 test points (TPs) are evenly distributed in the test area (see Figure 8). At each RP, the data were collected as follows:

- Set the PDA, facing north.
- Open the MiniStumbler and commence data collection of RSS for around 60 seconds, log it into a file.
- Orientate the device to face east by rotating 90 degrees clockwise and collect data for around 60 seconds, and again log it into a file.

Repeat data collection procedure for the south and west orientations.

Fig. 8 The RPs and TPs in CBD area of Sydney (image from Google earth)

The data collected at 10 of the test points consists of directional information and the rest only have the data collected facing one direction. Hence there are effectively 53 TPs.

In this test, the simple deterministic method was used. The 'Nearest Neighbour' (NN) algorithm (Bahl and Padmanabhan, 2000; Li et al., 2005) was applied because of its simplicity and reasonable level of accuracy. The basic idea is to calculate the distance (in signal space) between the observed set of SS measurements [ss_1, ss_2, ss_3 …… ss_n] and the SS measurements recorded in the database [SS_1, SS_2, SS_3…… SS_n]. The distance between these two vectors can be stated as

$$L_q = \left(\sum_{i=1}^{n} |s_i - S_i|^q \right)^{\frac{1}{q}}$$

Manhattan (q = 1) and Euclidean distance (q = 2) are the most common distant measurements. The RP which can provide the smallest signal distance is the nearest neighbour. The estimate of the MU's position is the position of the nearest neighbour.

In an indoor test, the test area is small, so the number of the APs is small and most of the time the MU can receive the signals from most of the APs, the NN works fine. However, in this outdoor test, with more than a thousand APs, we noticed that the NN could not always find the 'true' nearest neighbour. Figure 9 gives an example. The SSs detected at RP1, RP8 and TP1 are listed in the table. If the SS from a certain AP can not be detected, -100 dBm is nominally "recorded". Obviously, RP1 is more likely the nearest neighbour than RP8. However, the calculated signal distance shows that RP8 is the nearest neighbour. So applying NN directly has a problem. To solve this problem, the concept of 'candidates' for nearest neighbour was introduced. Before NN is applied, the candidates are selected from all the RPs. The candidates are the RPs which can 'hear' the signal from similar APs to those that can also be heard at the specific TP. In the previous example, at TP1, AP1 to AP3 were detected; at RP1, AP1 to AP6 were detected; at RP8, AP1 and AP7 were detected. The same AP(s) which appeared at RP1 and TP1 are AP1 to AP3 while at RP8 and TP1 is AP1 only. If the similarity (the proportion of matching APs) is set to 1/3, RP1 is a candidate while RP8 is not considered. Using the candidate method has two advantages: one is that the 'true' nearest neighbour can be found; the other is the calculation is sped up. Since fingerprinting is a pattern matching procedure, pattern matching algorithms can be applied. Using candidates is just a simple attempt to find a reliable algorithm and more investigation is needed.

	AP1	AP2	AP3	AP4	AP5	AP6	AP7
RP1	***-60***	***-60***	***-60***	***-60***	***-60***	***-60***	-100
RP8	***-90***	-100	-100	-100	-100	-100	***-85***
TP1	***-70***	***-70***	***-70***	-100	-100	-100	-100

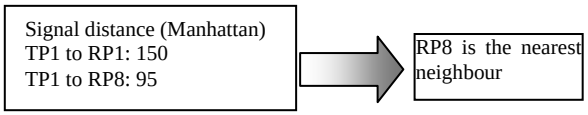

Fig. 9 An example of finding the wrong nearest neighbour

5. Test Results and Analysis

The test results are analyzed and reported in this section.

5.1 Using GPS only for positioning

The test area has a typical urban setting and the number of visible satellites are limited by tall buildings, towers etc. GPS needs at least 3 satellites to calculate a position (2D) and this requirement is not met in most of the TPs. Figure 10 shows the number of visible satellites from all the TPs. Thus this indicates the situation if GPS was used to calculate position.

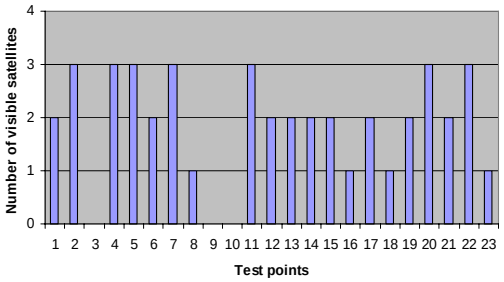

Fig. 10 Number of visible satellites from TPs

It is clearly noticeable that four or more satellites are not visible from any of the TPs. 3 satellites are visible from 7 TPs (about 30% of all the TPs) and even fewer satellites are visible from the rest of the TPs. No satellites were visible at 3 of the TPs. Obviously, at most of the TPs the number of visible satellites is not sufficient and thus a position could not be calculated. Furthermore, in the TPs where position can be calculated with 3 visible satellites (using a fixed altitude, for instance), the geometric distribution of the satellites is bad, i.e. the dilution of positioning (DOP) values are quite large (Kaplan, 1996). The consequence is that the error is quite high. Another disadvantage is that the time to calculate the first position fix is not short. Therefore, some other techniques of positioning are required to achieve satisfactory level of accuracy and speed.

5.2 Positioning based on the traditional database

As discussed in section 4, the proportion of matching APs is important to estimate the MU's position. If the proportion is too small, the problem described in section 4 may still exist; while if the proportion is too large, the 'true' nearest neighbour can be wrongly eliminated. Data were processed matching one fourth of APs, then one third and half of the APs. After choosing the candidates, the NN algorithm was applied. Figure 11 compares the results. The best result is found matching half of all the APs. These results are quite preliminary and further investigation is needed.

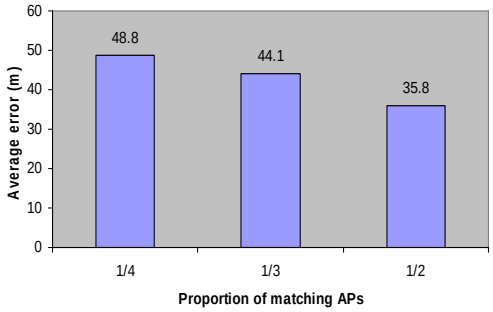

Fig. 11 Number of visible satellites from TPs

5.3 Using directional information

Table 1 summarises the results of using both traditional database and the directional database for each TP. It is clear that using the directional fingerprint database can provide more accurate estimates of position. The average error drops from 35.8m (traditional approach) to 23.5m (direction-based approach). The position estimates of all the test points except TP36 are better (or at least the same) using the directional SS information. However, estimating the MU's direction is difficult. In this test, only about 30% of the orientation estimates are correct (slight better than random). Further investigation is required.

Table 1 The positioning error using the directional fingerprint database vs traditional fingerprint database (units are metres)

	Directional database	Traditional database		Directional database	Traditional database		Directional database	Traditional database
TP1	14.7	14.7	TP19	16.0	16.0	TP37	10.0	52.4
TP2	14.7	14.7	TP20	25.7	25.7	TP38	10.0	10.0
TP3	14.7	25.3	TP21	45.3	97.7	TP39	10.0	52.4
TP4	14.7	51.2	TP22	16.9	97.7	TP40	52.4	52.4
TP5	25.7	33.6	TP23	97.7	97.7	TP41	3.4	3.4
TP6	25.7	33.6	TP24	45.3	97.7	TP42	11.4	11.4
TP7	25.7	33.6	TP25	25.2	25.2	TP43	30.8	30.8
TP8	33.6	33.6	TP26	25.2	63.5	TP44	15.1	15.1
TP9	29.6	29.6	TP27	25.2	25.2	TP45	60.9	60.9
TP10	17.3	17.3	TP28	25.2	63.5	TP46	23.3	50.7
TP11	29.6	29.6	TP29	5.8	76.4	TP47	5.2	5.2
TP12	17.3	17.3	TP30	5.8	5.8	TP48	7.5	7.5
TP13	27.2	27.2	TP31	5.8	5.8	TP49	39.0	85.1
TP14	26.8	27.2	TP32	5.8	5.8	TP50	21.5	35.7
TP15	26.8	82.2	TP33	18.4	18.4	TP51	32.1	32.1
TP16	27.2	27.2	TP34	18.4	18.4	TP52	23.7	46.9
TP17	16.0	16.0	TP35	18.4	18.4	TP53	10.9	38.7
TP18	16.0	16.0	TP36	50.7	18.4	Mean	*23.5*	*35.8*

6. Concluding Remarks

In this paper, the positioning technologies for outdoor positioning using Wi-Fi are discussed. An outdoor test is carried out in the Sydney CBD. In such an outdoor environment the Wi-Fi signal generally is much stronger than in indoor environments, but there are many pedestrians and cars. Hence, signal propagation is more complex. The test results show that fingerprinting works well for outdoor positioning, with errors in the tens of meters. Using the direction-based fingerprint approach can improve the performance. However, the orientation results do not show the same level of success as the indoor experiment (Li et al., 2007) and require further investigation.

As discussed through the previous sections, further works are needed for several issues, such as:

- How to decide the percentage of APs matching;
- New algorithm for outdoor positioning (pattern matching);
- The way to integrate Wi-Fi and GPS.

References

3GPP (2004): ***Location Services (LCS)***; Functional description; Stage 2, version 8.9.0., 3GPP TS 03.71, <http://www.3gpp.org/ftp/Specs/html-info/0371.htm>.

Bahl P.; Padmanabhan V.N. (2000): ***RADAR: An in-building RF-based user location and tracking system***, IEEE Infocom 2000, Tel Aviv, Israel, 26-30 March, vol. 2, pp. 775-784.

Barnes J.; Rizos C.; Kanli M.; Pahwa A. (2006): ***A positioning technology for classically difficult GNSS environments from Locata***, IEEE/ION PLANS, San Diego, California, 25-27 April, 715-721.

Bing B. (2000): ***Wireless Local Area Networks: Abridged edition of High-Speed Wireless ATM and LANs***, Artech House Publishes, Boston.

Eggert R.J.; Raquet J.F. (2004): ***Evaluating the navigation potential of the NTSC analog television broadcast signal***, ION GNSS 17th International Technical Meeting of the Satellite Division, Long Beach, California, USA, 21-24 September, pp. 2436-2446.

Geier J. (2002): ***802.11 Alphabet Soup***, <http://www.wi-fiplanet.com/tutorials/article.php/10724_1439551_1>.

Haldat E. (2002): ***Fingerprinting based-technique for positioning***, <http://www.telecomlab.oulu.fi/home/Radiotekniikka/materiaali/fingerprinting-report.pdf>.

Kaplan E.D. (ed.) (1996): ***Understanding GPS: principles and applications***, Artech House, Boston.

Ladd A.M.; Bekris K.E.; Rudys A.; Marceau G.; Kavraki L.E.; Dan, S. (2002): ***Robotics-based location sensing using wireless Ethernet***, Eighth ACM Int. Conf. on Mobile Computing & Networking (MOBICOM), Atlanta, Georgia, USA, 23-28 September, pp. 227-238.

Li B.; Kam J.; Lui J.; Dempster A.G. (2007): ***Use of directional information in Wireless LAN based indoor positioning***, IGNSS2007 Symp. on GPS/GNSS, Sydney, Australia, 4-6 December, paper 49, CD-ROM procs.

Li B.; Rizos C.; Lee H.K. (2004): ***Utilizing kriging to generate a NLOS error correction map for network based mobile positioning***, Int. Symp. on GPS/GNSS, Sydney, Australia, 6-8 December, paper 179, CD-ROM procs.

Li B.; Salter J.; Dempster A.G.; Rizos C. (2006): ***Indoor Positioning Techniques Based on Wireless LAN***, First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications, Sydney, Australia, 13-16 March, paper 113.

Li B.; Wang Y.; Lee, H.K.; Dempster A.G.; Rizos C. (2005): ***A new method for yielding a database of location fingerprints in WLAN***, IEE Proc. Communications, vol. 152, no. 5, pp. 580-586.

Li, B. (2006): ***Terrestrial mobile user positioning using TDOA and fingerprinting techniques***, PhD thesis, School of Surveying & Spatial Information Systems, University of New South Wales, Sydney, Australia.

Priyantha N.B.; Chakraborty A.; Balakrishnan H. (2000): ***The cricket location-support system***, Proc. 6th ACM International Conference on Mobile Computing and Networking, Boston, USA, 06-11 August, pp.32-43.

Roos, T., Myllymaki, P., Tirri, H., Misikangas, P. and Sievanan, J. (2002): ***A probabilistic approach to WLAN user location estimation***, International Journal of Wireless Information Networks, vol. 9, no. 3, pp. 155-164.

Saha S.; Chaudhuri K.; Sanghi D.; Bhagwat P. (2003): ***Location determination of a mobile device using IEEE 802.11b access point signals***, IEEE Wireless Communications & Networking Conference (WCNC), New Orleans, Louisiana, USA, 16-20 March, vol.3, pp.1987-1992.

Wang Y; Jia X; Lee H.K.; Li G.Y. (2003): ***An indoor wireless positioning system based on WLAN infrastructure***, 6th Int. Symp. on Satellite Navigation Technology Including Mobile Positioning & Location Services, Melbourne, Australia, 22-25 July, CD-ROM proc., paper 54.

Want R.; Hopper A.; Falcão V.; Gibbons, J. (1992): ***The Active Badge Location System***, ACM Trans. Information Systems, vol. 10, no. 1, pp. 91-102.

Xiang, Z., Song, S., Chen, J., Wang, H., Huang, J., Gao, X. (2004): ***A wireless LAN-based indoor positioning technology***, IBM J. Res & Dev, Vol 48 no. 5/6, September/November 2004.

Youssef M.; Agrawala A. (2005): ***The Horus WLAN location determination system***, International Conference On Mobile Systems, Applications And Services, Seattle, Washington, USA, 06-08 June, Session: Location, pp. 205-218.

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 27-34

Journal of Global Positioning Systems
2008
© CPGPS

Designing and Implementing a RFID-based Indoor Guidance System

C. C. Chang
Department of Applied Geomatics, Ching Yun University, Taiwan.

P. C. Lou and H. Y. Chen
Department of Information Management, Yu Da University, Taiwan.

Abstract

Most people are not familiar with the indoor environment because most interior spaces are similar, and as such do not arouse the interest of most visitors. Although the GPS combined with the GIS has been broadly applied to many navigation applications, it might be still insufficient in an indoor environment where GPS signals are often severely obstructed. To meet the indoor requirements, the technique of RFID (Radio Frequency IDentification) was utilized to play an important locating role in the implementation of an indoor guidance system. The RFID tags, containing 1 KB capacity divided by 64 blocks, were adopted to accommodate the spatial-related information for working with the shortest routing for the system. By selecting the start and end points in the operation, the guidance system can suggest a direct, shortest distance, path. The locations of the passing points were identified and obtained by retrieving the spatial-related data from the tags next to the moving RFID reader. The guidance messages, including suggested path, modified path and moving directions, can be immediately presented to the users and arrive at their destination via the system's graphic and voice interfaces. This guidance system has been comprehensively tested for its operation functions and was evaluated by a group of users, indicating that the average time for way finding in an indoor guidance trial can be efficiently reduced by 50%. This prototype for an indoor guidance system is expected to be capable of working on a portable device, such as a PDA or mobile phone, thereby extending its practical application.

Keywords: Guidance System, RFID, Indoor Locating, Shortest Routing Operation.

1 Introduction

More and more location-related information is applied with today's computer and communication technologies in order to provide so-called location-based services (LBS) to make our lives safer and more convenient (Williams and Christensen, 2005). Previous LBS activities have been primarily used in many commercial applications, such as personal navigation devices, safety and security services, vehicle tracking and fleet management. Generally speaking, LBS can be appropriately grouped by their application domain, e.g. entertainment applications, e-commerce applications, emergency applications, or mobility applications (Abwerzger et al., 2007).

The tour guide service is identified as one of the important applications of LBS. In this operating scenario, tourists or visitors arriving at an unfamiliar place can utilize a portable platform with mobile positioning capability to download necessary system and information files. Then, they can input their points of interest and obtain the required routing information on how to travel/move to those points. It is obvious that a mobile positioning technique that can instantly provide the locations of the visitor to the information system is a core component of such service (Küpper, 2005).

Among the location performance parameters, i.e. availability, accuracy, reliability, integrity and continuity, the ubiquitous availability both outdoors and indoors is a key factor to create a large number of applications for LBS (Lachapelle, 2004). Nowadays, GPS is the most promising and most popular technology in three-dimensional positioning. However, the reception of GPS signals inside most buildings is not reliable for positioning. Some auxiliary techniques, such as GPS pseudolite and Assisted-GPS, have been developed to more effectively obtain a position in a harsh environment (Barnes, et al., 2003; Bryant, 2005; Chang et al., 2005), but the cost of establishing those infrastructures for practical operation is still high.

In recent years, the indoor guidance system has shown its importance in many places, e.g. museums, hospitals, airport terminals and exhibition halls among others (Chou et al., 2004; Föckler et al., 2005; Lay, 2005). In such a

large indoor environment, people may not be easy to find their way to the points of interest as they generally are confused by the very similar space designs in those places. To meet both technical and cost requirements, this study proposed and tested an indoor guidance system based on the RFID. The RFID is mainly operated for indoor locating, applying the steps of (1) storing accurate location information and other necessary data into the tags; (2) installing RFID tags on paths in a building; and (3) connecting an RFID reader to the guidance system for data retrieval. The guidance system can, therefore, receive the location information from the RFID reader, provide the shortest routing operation, and work with the graphic and vocal interfaces to guide the users to their designated points.

2 Indoor Locating Techniques

A guidance system can be defined as a tool for providing the location and routing information, such as direction and distance between one's present location and one's desired destination, using vocal or graphic functions. While executing a guidance mission, a full system service includes the determination of the standing point, finding the route and direction to the destination, record of the completed trajectory, and the expectations of time and distance for the rest of the journey.

It is necessary that a guidance message is correctly provided at some "decision" points for all the waypoints. For an indoor guidance, a sensor attached to the person or moving object to locate the standing points by receiving the relevant spatial information from the responder at the decision point is an effective way of carrying out this mission. Previous development for indoor locating includes the use of GPS, bluetooth, infrared and WiFi etc. (Sun et al., 2005; Wang et al., 2005; Peng et al., 2006; Retscher, 2007). However, the RFID technique operates with many advantages, such as inexpensive, lightweight, tag re-writable, wireless sensing without line-of-sight, with a wide range of operation, a high communication speed, friendly to RF propagation, large data capacity, easy programming, and less dependence on RF links to the external data sources, and therefore has been gaining many interests in locating application (Finkenzeller, 2003; Chon et al., 2004; Chou et al., 2004; Miller, 2006). The performances of some existing indoor locating techniques are compared and summarised in Table 1.

Table 1 Comparison of indoor locating techniques

Performance	GPS	Bluetooth	Infrared	RFID
Accuracy	Low	Medium	High	High
Signal Error Ratio	High	Low	Low	Low
Power Consuming	High	High	Low	Low
Penetration	Bad	Good	Bad	Good

It is evident from Table 1 that the performance of the RFID system has shown its potential for indoor locating application. Briefly, the RFID system consists of three components, tags, reader and software including driver, middleware and application. The tags, also known as the transponders, can store the ID and additional information for any proposed application. The RFID reader is capable of writing/retrieving information into/from the tags. Through the communication between the antennas in both reader and tag, the application software is functioning to show the embedded information to the system users. The RFID system can be categorised by either an active or a passive system, depending if the tags are equipped with an internal battery or not. The RFID can also be grouped by frequency range, such as low (100-500 kHz), high (10-15 MHz), ultra high (850-950 MHz) and microwave (2.4-5.8 GHz) frequency.

3 System Development

The system's architecture and the tag's data contents were both designed for the purpose of developing an indoor guidance system based on RFID for locating. The hardware, software, interface and routing operation were selected and programmed for implementation in this guidance system.

3.1 System Architecture

The RFID-based indoor guidance system is constructed as shown in Fig. 1. The core operating functions rely on the information system developed in this study. This information system is composed of four basic functions, i.e. system control, locating, routing and graphic/voice guidance.

After setting up data-storing tags along the indoor paths in advance, the system begins its work by selecting a point of interest. The information system then activates the RFID reader to retrieve the spatial information from the nearest tags based on the control function, and locates the position of the user on a map using the system's locating function. Based on the shortest path for any decision point as suggested by the system's routing function, the guidance system operates its graphic and voice functions to guide the user to the destination.

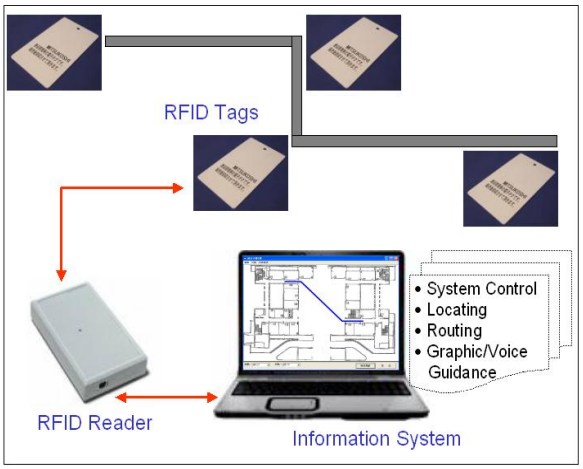

Fig. 1 System architecture

3.2 Software and Hardware

The indoor guidance system was built on a laptop computer equipped with Microsoft Windows XP, Intel Pentium(R) 4 mobile CPU (1.6 GHz) and 256 MB RAM. The system was programmed using Microsoft Visual Basic 6.0, which corresponds with the zlg500B.dll function library for developing the control command for the RFID reader. In addition, the voice interface was based on Microsoft Speech SDK 5.1.

A relatively low cost device was considered at this stage in the development, therefore a passive type of RFID was used. The system utilized a Philips MiFare RC 500 RFID reader as the locating sensor. This type of RFID is based on the RS-232 communication protocol, 13.56 MHz frequency, ISO14443A standard and a maximum 10 cm communication range. It should be noted that an active RFID device can be alternatively adopted to increase the contact range between the reader and the tags in order to expand the feasibility of this guidance system.

The tag used in combination with the RFID reader is a MiFare S50, which has a 1 KB memory. This type of tag consists of 16 sector sets and 4 block sets. A total of 64 data blocks (16 x 4) can be defined to contain all the location and spatial information required for indoor guidance.

3.3 Routing Operation

Since the indoor environment is usually quite regular, the decision points defined as the interest points and turning points on the routes can be given a node code together with its location coordinate and the spatial relationship to the adjacent nodes. The RFID tags, embedded with the location-based information and attached at all the decision points, are used to provide the necessary information for the routing operation to carry out the indoor guiding.

In this study, the routes connecting any two adjacent nodes are generally identified as an “available” path. However, the routes were also possibly defined as an “unavailable” path if the two nodes are not able to pass or are not next to each other. For the routes marked “available”, the spatial-related information for the two adjacent nodes are managed by a database and written into the tags affixed at the decision points. On the other hand, the information for the “unavailable” path is not stored by the tags and is certainly not provided for routing. To further realise the application of the “available” path information, a scenario of more than one path between the two nodes is shown in Fig. 2 and its routing operation is explained.

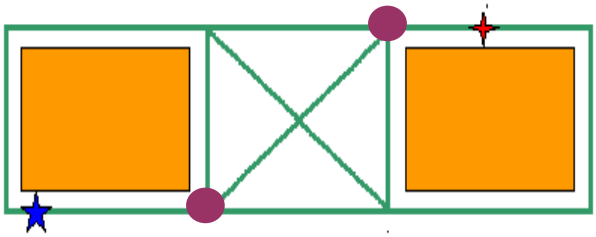

Fig. 2 Scenario of “available” paths between two nodes (The symbols of cross, star and circle represent the start, end and way points, respectively)

When the start and end points are set up, any possible routes pre-defined as the “available” paths are selected for the routing operation. For the scenario shown in Fig. 2, the system can select one of the “available” paths with the shortest distance from the tag’s spatial database. The routing operation can automatically provide all the waypoints along the suggested route, even though the user may not follow the suggestion and take an alternative route to reach the end point. This routing operation is easy to program and requires the least amount of processing time. The practical operation adopted by the system for routing is listed as follows:

1. System and Device Initiation
2. Select Case Function
3. Case Locating:
4. RF Sensing; Read Block C1
5. Case Way Finding:
6. RF Sensing; Read Block Coordinate & Node
7. Until Node = Destination or Stop Character
8. End Select
9. If Stop Character < > True Then
10. Guide Location or Path via Speech & GUI
11. Else: Show Message
12. End If

It is evident from the above-mentioned operation that RFID function programs must be loaded to initiate and command the RFID device. When carrying out a locating function, the RFID reader retrieves tag information to obtain the location. Once the routing command is made, the system reads more spatial information from the tag to determine the shortest path for the user to arrive at the desired point of interest. At the same time, the graphic and vocal interfaces are activated to give more advice to the system users if needed.

3.4 Tag Contents

Considering the memory constraints of a tag, it would be almost impossible to store all the required spatial information in a tag. However, some useful information, such as the coordinates and codes of the decision points, can be embedded into the tags for application. The plan coordinate components (x, y) for a local frame are measured and stored in the tags for each decision point. This data set of coordinates allows to easily show locations on the map, calculate the distance and identify the spatial relationship between any two points. On the other hand, the code information given to each point is a simple type of data occupying less memory in a tag, but it must be maintained by a database for establishing its location and spatial relationship with other tags. Since the tags applied are multi-blocks for data storage, the spatial information including the coordinate, node code and spatial relationship between the two adjacent nodes were decided to be embedded into the tags (see Fig. 3).

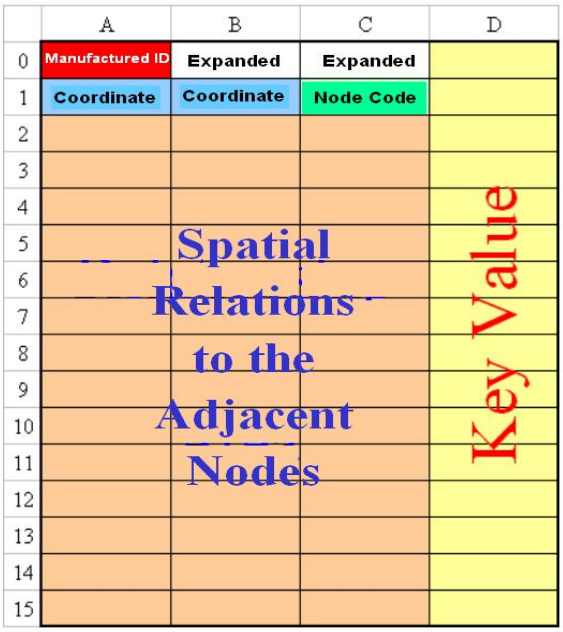

Fig. 3 Data contents stored in a multi-block tag

The data storage of a multi-block tag is defined by a read-only A0 block for the manufactured ID, the entire D blocks of key values for access control, and the B0 as well as the C0 blocks for function expansion. The coordinate components of (x, y) and the node code are stored at A1, B1 and C1 block, respectively, for a location designated as a decision point for guidance. The rest of the blocks, from A2, B2, C2 to A15, B15, C15, are used to describe the spatial relationship between that point and other adjacent nodes (see Table 2 for an example).

Table 2: Numerical example for a spatial relationship

Character	1	2	3	4	5	6	7	8
Value	4	0	0	0	0	1	1	2
Value	5	0	0	0	0	1	1	5

The tags used by the guidance system contain 64 (16 x 4) blocks for the entire data storage, in which a total of 52 (14 x 3) blocks are available to store the spatial relationship with this tag's location. For each block, the data capacity is 8 bits. As seen in the first row of Table 2, the 1st character 4 represents an adjacent node located at this tag's right-hand direction with node code 112 is stored in the 6th to 8th characters. Another node with code 115, also located at the right-hand side and next to code 112, has a value of 5 at the 1st character of row 2. This relation means that code 112 and code 115 are sequential on the same route and in the same direction as this tag. It should also be noted that the 2nd through the 5th characters are reserved with 0000 for any further definitions, such as the values for expressing the different floors or buildings. The direction index shown in the first value of a row, as shown in Table 3, can be designated for voice guidance.

Table 3: Direction index designed for the spatial relationship

Direction	To Next Node	To Node	Sequential
Keep Going	0		1
Turn Right Front	2		3
Turn Right	4		5
Turn Right Rear	6		7
Make U Turn	8		9
Turn Left Rear	A		B
Turn Left	C		D
Turn Left Front	E		F

4. System Demonstration

The guidance system developed in this study can be applied to control the RFID device, locate a position, suggest the shortest path and provide both graphic and vocal guidance messages. The system was tested on the ground floor of the administration building at Yu-Da University. The system operation is demonstrated and described as follows:

(1) Select a building and floor from the scroll menu on the top left to display its plan map on the system.

(2) Select a start point from the scroll menu on the bottom left most for this mission.

(3) Select an end point also on the bottom left for this mission.

(4) Press function key on the bottom right to carry on routing operation and provide the shortest path for this mission (see Fig. 4).

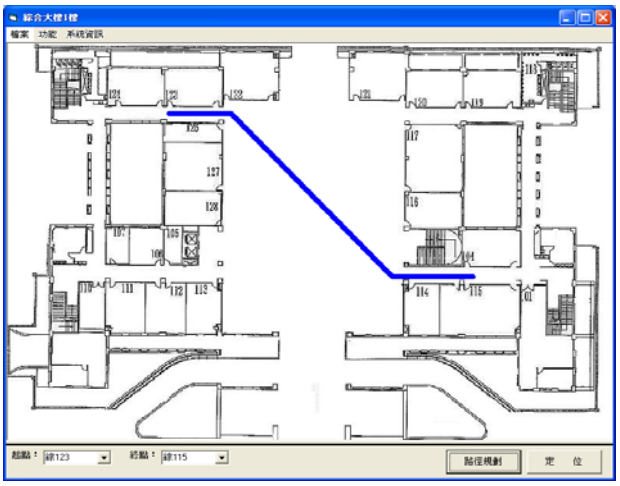

Fig. 4 Routing for the shortest path

(5) Press function key on the bottom right most to initiate RFID reader to locate the position of the user.

(6) For waypoint guidance, the system shows the route traveled and the user's present location at the waypoint with a different colour (see circles in Fig. 5 and Fig. 6).

(7) On the final section of the route, the waypoint colour is changed again to show that the user, using the RFID-based guidance system, has followed the suggested path and has arrived at his/her destination (see circle in Fig. 7).

(8) For a user not following the shortest path suggestion, a routing modification function has been developed and is provided by the system. As seen in Fig. 8, the user is located at the starting point on the suggested route. The system's graphic page soon shows a new routing path, see Fig. 9, when it is detected by the system that any waypoint is found to be different from the one suggested.

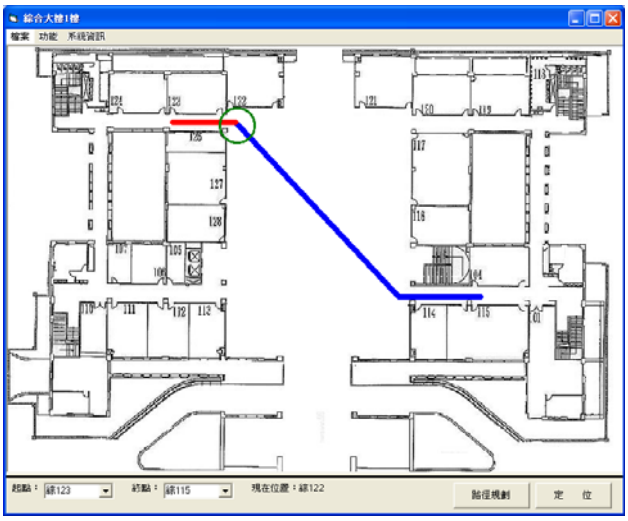

Fig. 5 Guidance at the start

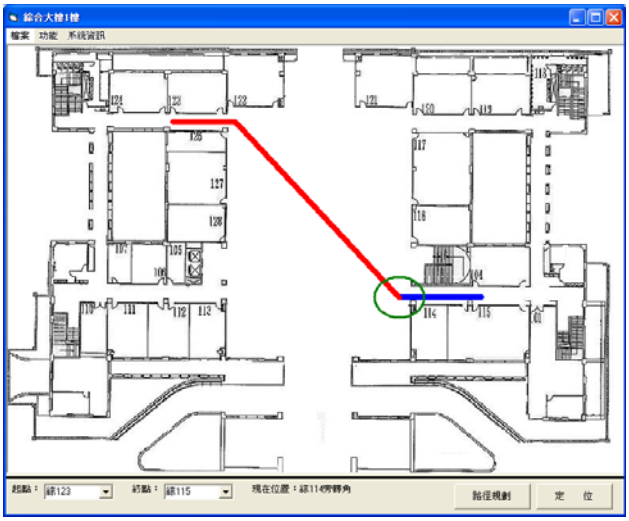

Fig. 6 Waypoint guidance

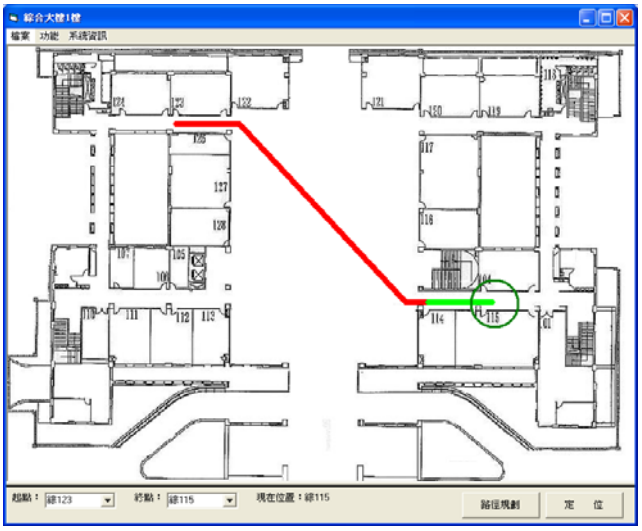

Fig. 7 Guidance at the end

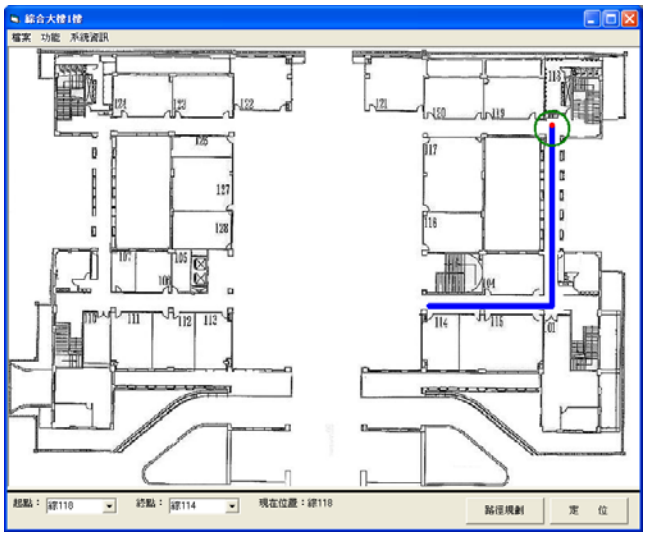

Fig. 8 Locating at the start of the suggested path

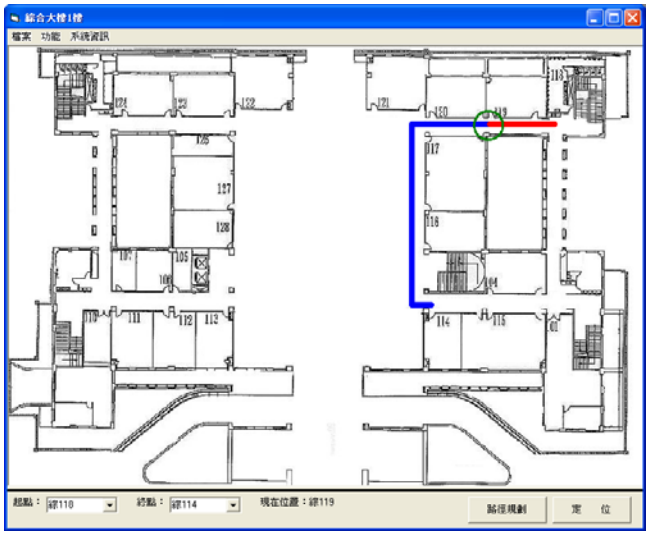

Fig. 9 Waypoint guidance after routing modification

(9) After routing modification, the user can follow both the graphic and the voice guidance to move along the updated path to reach the destination (see Fig. 10).

5. Efficiency Tests

To determine the efficiency of using this RFID-based indoor guidance system, an evaluation was carried out by inviting 10 visitors to find their way from the same starting point to the same destination. The time needed by each participant was recorded and used to evaluate the system efficiency.

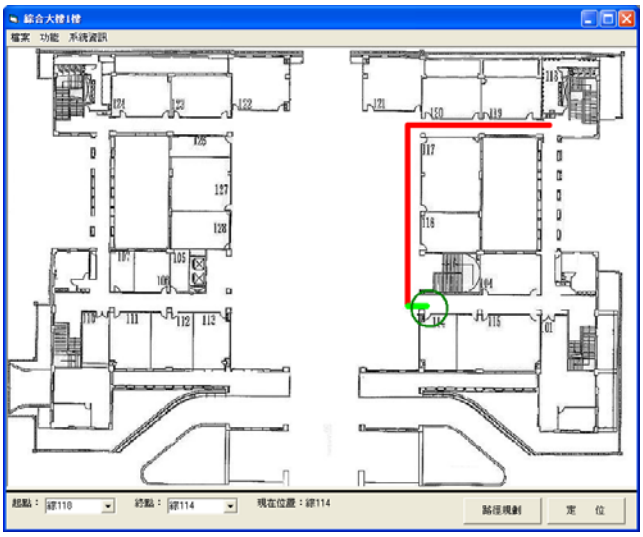

Fig. 10 Guidance at the end after routing modification

The 10 participants were divided into two equal groups. Five participants used the guidance system to find their way, and the other five did not use any auxiliary device for guidance. A total of 5 sections of trial paths were selected from the indoor environment for the tests. The average times needed in 5 trial paths for the 5 participants in each group, non-system user and system user, are listed in Table 4. The efficiency of the proposed guiding system, based on the ratio of time saved by the system user over the time needed by the non-system user, is also given in Table 4. The comparison of the time difference for each trial path is shown in Fig. 11.

Table 4: Average way finding time for the two test groups

Trial Path	Non-System User (sec)	System User (sec)	Reduction (sec)	Efficiency (%)
1	118	45	73	62%
2	92	41	51	55%
3	112	48	64	57%
4	102	37	65	64%
5	51	45	6	12%
Average	95	43	52	50%

It is evident from Table 4 that the system users required less time in each trial path to find the way to their destination. The average time required for each of the 5 trial paths could be reduced from 95 seconds to 43 seconds when the RFID-based guidance system was applied. The efficiency indicator showed that the time required could be significantly reduced by 50%, with the use of the proposed guidance system. Fig. 11 also shows that the system users consistently took a similar amount of time to find their destination in each trial path, whereas the non-system users tented to show a slightly downward

trend in the time required to arrive at their destination as they worked their way through the 5 trial paths.

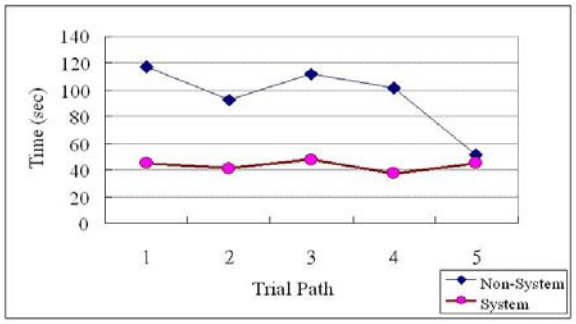

Fig. 11 Comparison of way finding time for the five trial paths

To further investigate the reasons for taking more time in way findings for the non-system users, their feedbacks were collected and summarised as follows:

(1) It was not able to clearly remember what routes I had already passed in this unfamiliar indoor environment;
(2) It was not easy to judge the direction in which I was going in the beginning of the trials, with the result that I repeated several routes;
(3) There usually was no spatial indication provided at the decision points, making it difficult to find the shortest paths to where I needed to go;
(4) It was getting used to this indoor environment towards the end of the trails so that it became easier to my way to the destination;
(5) The destinations could be inferred from the room numbers in a sequence to save time in the way findings.

It should be noted that this guidance system was only implemented in a small and simple space measuring around 100 m x 100 m, with regular room and route patterns. It is expected that the system efficiency can be further improved for more complicated or larger space.

6. Conclusions and Suggestions

This paper proposed, developed and tested a low cost indoor guidance system based on RFID and information technology. The advantages of using RFID for indoor locating, the architecture for the system, the software and hardware used for the development, the routing operation applied for way findings, the spatial information defined for the multi-block tags, the graphic/voice functions used for the system demonstration as well as the system efficiency tests were all carried out and described in this paper.

The indoor guidance system using an RFID reader along with the tags embedded with the spatial information was proven to be feasible in the preliminary phase. However, many issues need to be further investigated, such as (1) adopting a higher frequency or an active type of RFID device to increase the contact range to 1-3 m, based on an anti-collision design for tag data retrieval; (2) coupling a low cost orientation sensor to indicate the direction in a more automatic and continual manner, instead of manually pre-storing that information in the tags; (3) developing a more comprehensive design for tag content and display interfaces to accommodate a multi-floor indoor guidance system c/w a video play function; and (4) expanding this prototype of indoor guidance system to work in a portable device, such as a PDA or mobile phone, for more practical applications.

Acknowledgements

This work was supported by the National Science Council of the ROC under grant no. NSC 96-2221-E-231-028.

References

Abwerzger G., Wasle E., Fridh M., Lem O., Hanley J., Hanson P., Jeannnot M. and Claverotte L. (2007) ***GNSS and Location-based Services***. Inside GNSS, 2(4): 54-63.

Barnes J., Rizos C., Wang J., Small D., Voigt G. and Gambale N. (2003) ***High Precision Indoor and Outdoor Positioning Using LocataNet***. Journal of Global Positioning System, 2(2): 73-82.

Bryant R. (2005) ***Assisted GPS - Using Cellular Telephone Networks for GPS Anywhere***. GPS World, May 2005, 40-45.

Chang C.C., Lou P. C. and Ke P. J. (2005) ***Simulation Tests for Indoor Positioning with Real and Virtual GPS Pseudolite Observations***. International Symposium on GPS/GNSS 2005, Hong Kong.

Chon H.D., Jun S., Jung H. and An S. W. (2004) ***Using RFID for Accurate Positioning***. Journal of Global Positioning System, 3(1-2): 32-39.

Chou L.D., Wu C.H., Ho S.P., Lee C.C. and Chen J.M. (2004) ***Requirement Analysis and Implementation of Palm-based Multimedia Museum Guide Systems***. IEEE 18th International Conference on Advanced Information Networking and Application, Taiwan.

Finkenzeller K. (2003) ***RFID Handbook: Fundamentals and Applications in Contactless Smart Cards and Identification***. John Wiley & Sons Ltd.

Föckler P., Zeidler T., Brombach B., Bruns E. and Bimber O. (2005) ***PhoneGuide: Museum Guidance Supported by On-device Object Recognition on Mobile Phones***. ACM 4th International Conference on Mobile and Ubiquitous Multimedia, New Zealand.

Küpper A. (2005) ***Location-based Services: Fundamentals and Operation***. John Wiley & Sons Ltd.

Lachapelle G. (2004) ***GNSS Indoor Location Technologies***. Journal of Global Positioning System, 3(1-2): 2-11.

Lay D.S. (2005) ***The Design and the Evaluation of the Museum Digital Guide and Learning System***. National Computer Symposium, Taiwan.

Miller L.E. (2006) ***Indoor Navigation for First Responders: A Feasibility Study***. National Institute of Standards and Technology.

Peng Y.J., Yeh S.C. and Wu Y.H. (2006) ***A Research of WiFi Positioning Technology for Indoor Automatic Guide Systems***. MCU International Academic Conference, Taiwan.

Retscher G. (2007) ***Test and Integration of Location Sensors for a Multi-sensor Personal Navigator***. The Journal of Navigation, 60: 107-117.

Sun G., Chen J., Guo W. and Liu K.J. (2005) ***Signal Processing Techniques in Network-aided Positioning***. IEEE Signal Processing Magazine, June 2005, 12-23.

Wang C.L., Chiou Y.S. and Yeh S.H. (2005) ***An Indoor Location Scheme based on Wireless Local Area Networks***. IEEE Consumer Communications and Networking Conference, Las Vegas.

Williams D.H. and Christensen G. (2005) ***The Definitive Guide to: Mobile Positioning & Location Management***. Mind Commerce.

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 35-45

Journal of Global Positioning Systems
2008
© CPGPS

Investigation of Instantaneous Carrier Phase Ambiguity Resolution with the GPS/GALILEO Combination using the General Ambiguity Search Criterion

D. B. Morujão
Faculdade de Ciências da Universidade de Lisboa, Portugal

V. B. Mendes
Instituto Dom Luís (IDL), Laboratório de Tectonofísica e Tectónica Experimental (LATTEX), Faculdade de Ciências da Universidade de Lisboa, Portugal

Abstract

The success of high-precision Global Navigation Satellite Systems (GNSS) kinematic positioning depends partly on the ability to resolve the integer phase ambiguities. In this paper, we propose a new algorithm for instantaneous kinematic ambiguity resolution, for present and modernised GPS and for GALILEO. This approach - the GEneral Criterion Cascading Ambiguity Resolution (GECCAR) - selects the integer set of ambiguities using the General Ambiguity Search Criterion (GASC). Simulation runs have shown that single-epoch L1/E1 frequency ambiguity resolution was possible 99% of the considered epochs, when the three frequencies from both systems were used together. This new approach shows an improvement in the selection of the correct set of ambiguities when compared with the selection made by the Integer Least Squares Criterion (ILSC). We conclude that the GECCAR approach is a very promising algorithm for instantaneous ambiguity resolution.

Key words: GNSS, Ambiguity Resolution, Ambiguity Search Criterion

1 Introduction

Global Navigation Satellite Systems (GNSS) allow for high-precision positioning when the carrier phase ambiguities are correctly estimated. Several ambiguity resolution algorithms have been published during the past decades, but the majority of them was developed for application with the two-frequency GPS and is not well suitable for the modernised GPS and the GALILEO. One exception is the LAMBDA (Least-squares AMBiguity Decorrelation Adjustment) method (Teunissen, 1993), widely used for single and dual-frequency GPS processing (Jong *et al*., 1999). Among the ambiguity resolution methods developed for the three frequency systems we can refer the CIR (Cascade Integer Resolution) (Jung, 1999; Jung *et al*., 2000), the TCAR (Three-Carrier Ambiguity Resolution) (Forsell *et al*., 1997), the ITCAR (Integrated Three Carrier Ambiguity Resolution) (Vollath *et al*., 1998) and the Geometry-based Cascading Ambiguity Resolution (Zhang *et al*., 2003) methods. Usually, they were tested in static positioning or in kinematic positioning with short baselines. Xu (2006) presents another approach on ambiguity resolution, based on Voronoi Cells.

In general, the ambiguity resolution methods use the Integer Least Squares Criterion (ILSC) to select the integer ambiguities. This paper presents a new instantaneous ambiguity resolution algorithm, the GEneral Criterion Cascading Ambiguity Resolution (GECCAR), which is based on the General Ambiguity Search Criterion (GASC) and that can be applied to present and modernised GPS and to GALILEO.

The main objective was to create an efficient ambiguity resolution algorithm that enables the GNSS kinematic positioning for medium and long distances between the rover and the reference stations. In order to prove the effectiveness of this algorithm, results obtained using real GPS data and simulated data from the modernised GPS-only system, from the GALILEO-only system and from both systems will be described. In addition, results obtained when the ILSC was used instead of the GASC will also be presented, to make possible an adequate comparison between both criteria solutions.

2 Observation Model and Ambiguity Resolution

The general GNSS observation model, in which code and carrier phase observables are included, can be written in the following system of observation equations:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{v} \tag{1}$$

with $\boldsymbol{A} = \left[\boldsymbol{A}_a \mid \boldsymbol{A}_b\right]$

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{a} \\ - \\ \boldsymbol{b} \end{bmatrix}$$

where $\boldsymbol{y}$ is the observation vector (m×1), $\boldsymbol{x}$ is the complete vector of unknown parameters ((n+t)×1) (comprising the integer double-difference carrier phase ambiguities, $\boldsymbol{a}$ (n×1), and the coordinates and other unknown parameters, $\boldsymbol{b}$ (t×1)), $\boldsymbol{A}$ is the design matrix (m×(n+t)) for the full unknown vector, which can be expressed in terms of the design matrices $\boldsymbol{A}_a$ (m×n) and $\boldsymbol{A}_b$ (m×t), corresponding to vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively, and $\boldsymbol{v}$ is the measurement noise vector (m×1).

In general, the estimation criterion for solving the linear system of equations (Eq. (1)) is based on the least squares principle, which satisfies the condition:

$$\min_{\boldsymbol{a},\boldsymbol{b}} \left\| \boldsymbol{y} - \boldsymbol{A}_a\boldsymbol{a} - \boldsymbol{A}_b\boldsymbol{b} \right\|^2_{\boldsymbol{Q}_y}, \tag{2}$$

where $\|\|^2_{\boldsymbol{Q}_y} = (.)^T \boldsymbol{Q}_y^{-1}(.)$ and $\boldsymbol{Q}_y$ is the variance-covariance matrix of the double-difference observables. The solution of the condition expressed by Eq. (2) would be an ordinary unconstrained least squares problem if all the parameters were allowed to range through the space of real numbers (Teunissen, 1998); however, there is no standard technique to solve Eq. (2) taking into account that the ambiguities must take integer values. This problem is known as the Integer Least Squares problem (Teunissen, 1993). Its solution is based on the orthogonal decomposition of Eq. (2):

$$\left\|\boldsymbol{y} - \boldsymbol{A}_a\boldsymbol{a} - \boldsymbol{A}_b\boldsymbol{b}\right\|^2_{\boldsymbol{Q}_y} = \left\|\hat{\boldsymbol{v}}\right\|^2_{\boldsymbol{Q}_y} + \left\|\hat{\boldsymbol{a}} - \boldsymbol{a}\right\|^2_{\boldsymbol{Q}_{\hat{a}}} + \left\|\hat{\boldsymbol{b}}(\boldsymbol{a}) - \boldsymbol{b}\right\|^2_{\boldsymbol{Q}_{\hat{b}(a)}} \tag{3}$$

where $\hat{\boldsymbol{v}}$ is the unconstrained least squares residual vector, $\hat{\boldsymbol{b}}(\boldsymbol{a})$ is the least squares estimate of $\boldsymbol{b}$ conditioned on $\boldsymbol{a}$, and $\boldsymbol{Q}_{\hat{b}(a)}$ is the corresponding variance-covariance matrix. As a consequence, the minimisation of the problem in Eq. (2), taking into account the integer constraint of the ambiguities, can be solved in three steps. In the first step one disregards the integer constrains on the ambiguities and performs a standard least squares adjustment. As a result, one obtains real-value estimates of $\boldsymbol{a}$ and $\boldsymbol{b}$ ($\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$), referred to as the float solution, together with their variance-covariance matrix, $\hat{\boldsymbol{Q}}$:

$$\hat{\boldsymbol{x}} = \begin{bmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{bmatrix} = \left(\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{y} \tag{4}$$

$$\hat{\boldsymbol{Q}} = \begin{bmatrix} \boldsymbol{Q}_{\hat{a}} & \boldsymbol{Q}_{\hat{a}\hat{b}} \\ \boldsymbol{Q}_{\hat{b}\hat{a}} & \boldsymbol{Q}_{\hat{b}} \end{bmatrix} = \left(\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A}\right)^{-1} \tag{5}$$

where $\boldsymbol{P} = \boldsymbol{Q}_y^{-1}$. In the second step, using the float estimation of the ambiguities, $\hat{\boldsymbol{a}}$, one estimates the corresponding integer ambiguities, usually by minimising the second term of the right-hand side of Eq. (3) and setting the last term to zero (Teunissen, 2003; Verhagen, 2004). In this way, the criterion used to estimate the integer ambiguities, $\breve{\boldsymbol{a}}$, called as the *Integer Least Squares Criterion* (ILSC), is:

$$\breve{\boldsymbol{a}}_{ILS} = \min_{\boldsymbol{a}\in Z^n} \left\|\hat{\boldsymbol{a}} - \boldsymbol{a}\right\|^2_{\boldsymbol{Q}_{\hat{a}}}. \tag{6}$$

This second step is called the ambiguity resolution process. Finally, in the third step, one solves for the last term, and the fixed solution of the remaining parameters is:

$$\breve{\boldsymbol{b}} = \hat{\boldsymbol{b}} - \boldsymbol{Q}_{\hat{b}\hat{a}}\boldsymbol{Q}_{\hat{a}}^{-1}\left(\hat{\boldsymbol{a}} - \breve{\boldsymbol{a}}\right). \tag{7}$$

Teunissen (1998) stressed the importance to consider the question whether the estimated set of integer ambiguities, $\breve{\boldsymbol{a}}$, should be accepted or not if it is of poor quality. That conducted to the need of perform a validation of the estimated integer ambiguities. Various validation procedures, based on several statistical tests, have been proposed in the past, in order to validate the estimated integer ambiguities. Some of them give satisfying results and are widely used. However, those statistical tests are based on incorrect assumptions and lack a theoretical basis (Verhagen, 2004). One of the most popular validation procedures is the ratio test in which the statistic is the ratio of the second minimum (*sec min*) quadratic form of the residuals to the minimum (*min*) quadratic form of the residuals. Using this test, if

$$\frac{\left(\hat{\boldsymbol{v}}^T\boldsymbol{P}\hat{\boldsymbol{v}}\right)_{\sec min}}{\left(\hat{\boldsymbol{v}}^T\boldsymbol{P}\hat{\boldsymbol{v}}\right)_{min}} > RT$$

where the *RT* is an empirical critical value (the value 2.0 has been suggested by Euler and Landau (1992) and Wei and Schwarz (1995) and the value 1.44 was proposed by Tiberius *et al.* (1997)), the ambiguity set that generates the minimum quadratic form of the residuals is validated as the correct solution. Otherwise, there is no integer solution for the ambiguities, for the data set used. Although the ratio test is considered a useful validation procedure and is commonly used, with different critical values, it has a problem in the statistical assumption that

the least-squares residuals of the best solution and the second-best solution are independent (Teunissen, 1998).
Xu (2002) proposed a new criterion to select the integer ambiguity set of carrier phase among the search area, based on the fact that the ILSC (Eq. (6)) was not generally optimal. This criterion, called the *General Ambiguity Search Criterion* (GASC), has the form:

$$\min_{\boldsymbol{x}} \left\| \hat{\boldsymbol{x}} - \boldsymbol{x} \right\|_{\boldsymbol{Q}_{\hat{x}}}^2 , \tag{8}$$

where $\boldsymbol{x} = \left[\boldsymbol{a}\ \boldsymbol{b}\right]^T$, $\boldsymbol{a} \in Z^n$, $\boldsymbol{b} \in R^t$, and $\hat{\boldsymbol{x}} = \left[\hat{\boldsymbol{a}}\ \hat{\boldsymbol{b}}\right]^T$. This new criterion was developed based on the fact that the third term of the right-hand side of Eq. (3) can not be set to zero (Xu, 2004). Unlike the ILSC, the GASC takes into account the residuals for all the unknowns and not only for the ambiguities, providing a unique and optimal solution under the least squares principle and under the condition of integer ambiguities (Xu, 2007). Therefore, with the given data, following the GASC, there is no need to validate the estimated integer ambiguities, as the estimated set of integer ambiguities is the best one that can be reached.

By diagonalising the normal equations related to the observation equation system (Eq. (1)), which have the ambiguity parameters separated from the other unknowns, a criterion equivalent to Eq. (8), known as *Equivalent General Criterion* (EGC), can be written (Xu, 2004):

$$\min_{\boldsymbol{a}\in Z^n,\, \boldsymbol{b}\in R^t} \left(\left(\hat{\boldsymbol{a}} - \boldsymbol{a}\right)^T \boldsymbol{Q}_{\hat{a}}^{-1} \left(\boldsymbol{a} - \boldsymbol{a}\right) + \left(\hat{\boldsymbol{b}} - \boldsymbol{b}\right)^T \boldsymbol{Q}_{\boldsymbol{b}}^{-1} \left(\boldsymbol{b} - \boldsymbol{b}\right) \right). \tag{9}$$

For convenience, by denoting

$$da = \left(\hat{\boldsymbol{a}} - \boldsymbol{a}\right)^T \boldsymbol{Q}_{\hat{a}}^{-1} \left(\hat{\boldsymbol{a}} - \boldsymbol{a}\right) \tag{10}$$

and

$$db = \left(\hat{\boldsymbol{b}} - \boldsymbol{b}\right)^T \boldsymbol{Q}_{\hat{b}}^{-1} \left(\hat{\boldsymbol{b}} - \boldsymbol{b}\right), \tag{11}$$

Eq. (9) may be written as

$$\min_{\boldsymbol{a}\in Z^n,\, \boldsymbol{b}\in R^t} \left(da + db\right), \tag{12}$$

where *da* represents an enlarging of the residuals due to the ambiguity change caused by ambiguity fixing, and *db* represents an enlarging of the residuals due to the coordinates change caused by ambiguity fixing.

The EGC provides a way to show the relationship between the ILSC and the GASC: the function to be minimised using the ILSC is just one of the terms of the EGC (Xu, 2002). Thus, minimising *da* is exactly the same as the ILSC. When the results obtained using Eq. (6) are different from those obtained with Eq. (9), the results from the search using Eq. (6) are only sub-optimal, due to the optimality and uniqueness property of Eq. (9) (Xu, 2007).

3 A New Ambiguity Resolution Algorithm

There are two approaches to resolve the ambiguities – a single-epoch (or instantaneous) approach and a multi-epoch approach. In the single-epoch approach the observations from each epoch are processed independently, whereas the multi-epoch approach uses many sequential observations together. When the observations are significantly contaminated by biases, such as multipath, residual atmospheric effects, and satellite errors, it is more difficult to instantaneously resolve the ambiguities correctly. A multi-epoch approach using a recursive implementation is the most reliable way for dealing with the problem. However, the instantaneous ambiguity resolution has several advantages - it is resistant to negative effects of cycle slips or a loss of lock and the changes in the constellation of the tracked satellites do not introduce additional complications to the data processing.

In order to instantaneous resolve the ambiguities for GPS and for GALILEO, a new ambiguity resolution procedure based on the General Criterion was developed and implemented: the GEneral Criterion Cascading Ambiguity Resolution (GECCAR). The GECCAR uses:

1. a cascading procedure;
2. an *a priori* transformation to decorrelate the ambiguities;
3. a search algorithm, where each ambiguity is constrained with the values of previously selected ambiguities;
4. the General Ambiguity Search Criterion or the Equivalent General Criterion for integer ambiguity selection.

The cascading procedure for the three-frequency systems was introduced by Forsell *et al*. (1997), who developed the TCAR (Three-Carrier Ambiguity Resolution) method for the GALILEO carrier phase ambiguity resolution, and by Jung (1999), who suggested the CIR (Cascade Integer Resolution) method for the modernised GPS. Although proposed for different systems, both methods are based on the idea of widelaning to take advantage of the stepwise-improved precision in carrier phase ranges from the longest wavelength to the shortest wavelength. Both methods are geometry-free, instantaneous integer ambiguity resolution methods, using integer rounding. Other approaches have been developed based on this principle, as the ITCAR (Integrated Three Carrier

Ambiguity Resolution) (Vollath *et al*., 1998) and the Geometry-based Cascading Ambiguity Resolution (Zhang *et al*., 2003) methods.

The linear combinations used in the implemented algorithm are based on the set of frequencies established for modernised GPS and for GALILEO (see Table 1) and are presented in Table 2.

The full GECCAR procedure consists of three steps. In the first step, the EWL (Extra Wide Lane) ambiguities are estimated using the most precise pseudorange available and the EWL phase combination as observables. In the second step, with the ranges based on the results obtained in the first step, the WL (Wide Lane), or the ML (Medium Lane), ambiguities are estimated using the most precise pseudorange and the WL, or the ML, phase combination as observables. Finally, in the third step, the L1/E1 ambiguities are estimated – the observables used are the L1/E1, L2/E5b and L5/E5a carrier phases, and the unknown ambiguities are just the L1/E1 ambiguities as the L2/E5b and L5/E5a ambiguities may be written in function of L1/E1, ML, WL and EWL ambiguities and the ML, WL and EWL ambiguities have been estimated in the second and first steps, respectively.

Table 1 Frequencies established for the GALILEO and the modernised GPS.

System	Signal	Frequency (MHz)
GALILEO	E1	1575.42
	E5a	1176.45
	E5b	1207.14
GPS	L1	1575.42
	L2	1227.60
	L5	1176.45

Table 2 Modernised GPS and GALILEO frequency combinations.

Lane	System	Linear Combination	Wave length (m)
EWL	GPS	L2-L5	5.861
	GALILEO	E5b-E5a	9.765
WL	GPS	L1-L2	0.862
	GALILEO	E1-E5b	0.814
ML	GPS	L1-L5	0.751
	GALILEO	E1-E5a	0.751

In each step, we follow the decorrelation process proposed by Teunissen (1993). Before estimating the ambiguities as integers, the float ambiguities, $\hat{\boldsymbol{a}}$, are transformed into an equivalent, but less correlated set of ambiguities, $\hat{\boldsymbol{z}}$, and the corresponding variance-covariance matrix, $\boldsymbol{Q}_{\hat{a}}$, is transformed into the variance-covariance matrix, $\boldsymbol{Q}_{\hat{z}}$, using the Z-transformation:

$$\hat{\boldsymbol{z}} = \boldsymbol{Z}^T \hat{\boldsymbol{a}}; \quad \boldsymbol{Q}_{\hat{z}} = \boldsymbol{Z}^T \boldsymbol{Q}_{\hat{a}} \boldsymbol{Z} .$$

This **Z**-matrix needs to fulfil certain requirements, in order to be admissible, as the integer nature of the ambiguities should be maintained. Therefore, all entries in the **Z**-matrix should be integers and the inverse of the **Z**-matrix should exist and its entries should be integers as well (Teunissen, 1994). The Z-transformation should aim at maximum possible decorrelation of the ambiguities to make the search algorithm more efficient.

After the decorrelation, the search is performed, over the search space region, in order to estimate the correct values of the ambiguity parameter vector. This is done by constraining each ambiguity candidate on the values of the previously selected ambiguities. In order to select the correct ambiguity set, $\breve{\boldsymbol{a}}$, using the Equivalent General Criterion, it should be done, for each candidate $\boldsymbol{a}$

- the computation of *da*;
- the computation of $\boldsymbol{b}$ and *db*;
- the computation of *da*+*db*;
- the selection of $\breve{\boldsymbol{a}}$ that minimises *da*+*db*.

4 Simulation Description

With the purpose of examining the effectiveness of the proposed algorithm for instantaneous ambiguity resolution, for modernised GPS and for GALILEO, a software-based GNSS data simulator was developed in C++. This software simulates the 27 satellite ephemerides for GALILEO, using the parameters given by Zandbergen *et al.* (2004). The GPS satellite positions were based on the final IGS (International GNSS Service) orbits. It was considered that the GST (GALILEO System Time) is synchronised with the GPS time and that the WGS84 is the reference frame for both systems.

The simulator generates pseudorange and carrier phase observables, for the three frequencies anticipated for the modernised GPS and for the GALILEO, by the addition of the errors inherent to double-difference observation (ionospheric and tropospheric delays, multipath error and receiver noise) to the true ranges. The errors are added to the non differenced true ranges between each station and each observed satellite. In the case of the carrier phase observables, an integer ambiguity value was established for every pair station/satellite. All the other sources of error affecting GNSS positioning ambience were not taken into account, as the contribution of those errors is insignificant when the double-difference model is used at the processing software.

The ionospheric errors were generated using Global Ionospheric Maps (GIMs) from the IGS. These maps are produced in a daily basis, and distributed in the IONospheric EXchange (IONEX) format, and represent

the Vertical Total Electron Content (VTEC), each 2 hours, at grids with a resolution of 2.5º latitude by 5º longitude. These maps define the global trend of the ionospheric errors in a set of coefficients of a spherical harmonic expansion. This spherical harmonic expansion is then used to define a grid of VTEC values on a two dimensional ionospheric shell above the area of interest. These grid values are interpolated to the pierce point of the observation and the interpolated values are multiplied by an elevation mapping function to give the final ionospheric error (Alves, 2001). As the same IONEX maps were used when the data was processed, a Non Modelled Ionospheric Residual (NMIR) was added to the ionospheric errors, based on the root-mean-square (RMS) maps associated with the GIMs. IGS also provides daily RMS data files with the variance values of VTEC.

Generally, the hydrostatic component of the tropospheric delay in the zenith direction is about 2.3 m at sea-level and represents about 90% of the total tropospheric delay. As found by Mendes (1999), the hydrostatic component of the tropospheric delay can be modeled to sub-millimeter accuracy with the use of prediction models such as Saastamoinen (1973). The highly variable non-hydrostatic tropospheric delay can only be modeled to an accuracy of a few centimeters in the zenith direction. Further error is introduced when the zenith tropospheric delay is mapped to the elevation angle of the satellite with the use of a mapping function. At the simulation, the tropospheric delays were generated using the Saastamoinen (1973) zenith delay models, combined with the Global Mapping Functions (GMFs) (Boehm *et al.*, 2006), to model the elevation dependence of the zenith tropospheric delay. As the same model was used when the data was processed, an error of 5% of the calculated tropospheric delay was inserted into the non differenced simulated observables.

Regarding the receiver noise, 0.01 cycles were added for all the carrier phase observables for both GPS and GALILEO. For the code pseudoranges, the values considered are shown in Table 3. For the multipath error, a value of 0.10 m was established for all the pseudoranges and a value of 0.025 cycles was set for all the carrier phases, for both GPS and GALILEO.

Table 3 Simulated receiver noise.

	Signal	Receiver Noise
GPS Code	L1	0.40 m
	L2	0.30 m
	L5	0.20 m
GALILEO Code	E1	0.25 m
	E5b	0.30 m
	E5a	0.20 m
Phase	All	0.01 cycles

5 Tests and Results

The aim of the tests was the evaluation of the proposed instantaneous ambiguity resolution algorithm - GECCAR – for modernised GPS and for GALILEO, in kinematic positioning, for different ionospheric residual levels and different frequency scenarios. With this purpose, software to process GNSS data was developed in C++. The implemented model for positioning was a geometry-based model with double-difference observables. The existence of common frequencies between GPS and GALILEO was not taken into account when forming the double-difference equations.

The performance of the proposed algorithm was evaluated in terms of the *Percentage of Correct Instantaneous Ambiguity Resolution* (PCIAR) at L1 (for GPS) or E1 (for GALILEO) frequency. This is not a probabilistic value, as the ambiguity resolution success rate generally used. The correct ambiguity values are previously known, so the estimated ambiguity values may be compared with the true ones. The PCIAR is, then, the value that is calculated by dividing the number of epochs where all the ambiguities were correctly fixed by the total number of epochs of the whole data set.

Several experiments have been carried out using different real and simulated observation sessions. The simulated example presented below is a representative subset of all the experiments carried out and its results illustrate the results obtained with all the experiments.

In order to evaluate the proposed algorithm with actual dual-frequency GPS data, results from a test using real data are shown before presenting the results with simulated data. The GPS real data used in this test was collected during a flight in Central Europe, in May 12th 2005, between 12:00 and 14:10, at a rate of 1 Hz. During the flight, the distance between the rover and the reference antennas varied between 200 m and 280 km, approximately.

The GPS real data was processed, using the software described above, for two different scenarios - GPS2, using L1 and L2 GPS real data, and GPS1, using L1-only GPS real data. With a minimum elevation angle of 15º, between 4 and 8 satellites were available. The ionospheric errors were modelated using GIMs. The tropospheric delays were modelated using the Saastamoinen zenith delay models, combined with the GMFs. At the GPS2 scenario, the cascading ambiguity resolution process comprised two steps: it began with the WL ambiguity estimation and then estimated the L1 ambiguity set. When using the GPS1 scenario, the GECCAR degenerated on a scheme with just one step, processing the L1 pseudoranges and carrier phase observables.

With the goal of evaluating the advantage that the GASC has on the proposed ambiguity resolution algorithm, the observations were processed two more times, using the same software but replacing the GASC by the ILSC at the ambiguity resolution function (with RT=1.44 and RT= 2.0), for each scenario. Table 4 lists the PCIAR values obtained for the two scenarios. The second column shows the PCIAR values when the GASC was used and the third and the fourth columns show the PCIAR values using the ILSC, with RT=1.44 and RT=2.0, respectively. The advantage of using the GASC, instead of the ILSC, is clear for the two scenarios - the PCIAR values reached on all the runs when the ILSC was used are smaller than those obtained when the GASC was used.

Table 4 PCIAR values for the real data tests.

SCENARIO	GASC	ILSC (1.44)	ILSC (2.0)
GPS2	97.1	94.3	92.3
GPS1	62.8	57.1	50.3

Figure 1 illustrates the faults in the estimation of the correct ambiguity set, as a function of the distance between the rover and the reference station, for the GPS2 scenario, using either the GASC or the ILSC (with RT=1.44 and RT=2.0). Table 5 shows the analysis of the partial PCIAR values, in function of the distance, for the GPS2 scenario. Each line corresponds to an interval of 10 km distance, along the trajectory, when the partial PCIAR value is different from 100, at that interval, in one of the runs.

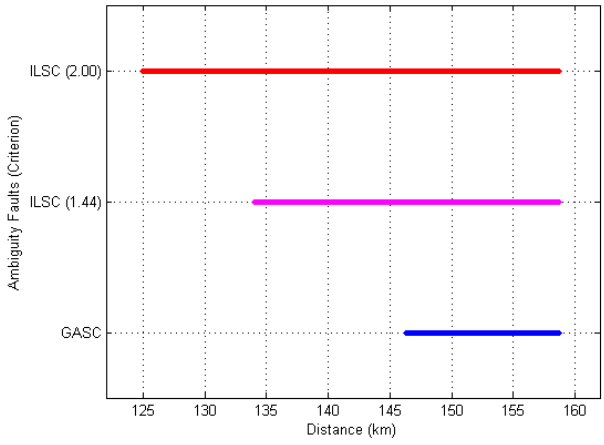

Fig. 1 Faults in ambiguity resolution, using the GPS2 scenario.

Table 5 Partial PCIAR values for the GPS2 scenario.

Criterion Distance (km)	GASC	ILSC (1.44)	ILSC (2.0)
120 – 130	100.0	100.0	49.4
130 – 140	100.0	39.0	0.0
140 – 150	61.9	0.0	0.0
150 – 160	14.1	14.1	14.1

Note that, due to the way that the atmospheric errors were handled, the ambiguity faults are not function of the distance between the rover and the reference stations; instead, the errors are dependent of the spatial errors of the models. It is clear the advantage of using the cascading scheme for instantaneous ambiguity resolution, even if it comprises only two steps.

For the simulated tests, twelve different scenarios, presented in Table 6, were established. At the GPS3 scenario, that uses simulated L1, L2 and L5 modernised GPS data, the cascading ambiguity resolution process comprises the three steps described above. It begins with the estimation of the EWL ambiguities, then estimates the WL ambiguities and finally estimates the L1 ambiguity set. The GPS2A scenario uses simulated GPS data at L1 and L2 frequencies and the cascading ambiguity resolution process comprises just two steps to estimate the WL ambiguities and the L1 ambiguities. The GPS2B scenario uses simulated L1 and L5 modernised GPS data, and the cascading ambiguity resolution process is identical to the process used with the GPS2A scenario replacing the WL ambiguities by the ML ambiguities. When using the GPS1A scenario, the GECCAR degenerates on a scheme with just one step that processes the L1 pseudoranges and carrier phase observables. The GAL3 scenario contains E1, E5a and E5b simulated data from GALILEO and the cascading process comprises the three steps as the GPS3 scenario. The cascading ambiguity resolution at the GAL2A scenario, which utilises E1 and E5b GALILEO simulated data, is similar to the GPS2A scenario. The GAL2B scenario, where the simulated GALILEO data from E1 and E5a frequencies is used, has the same two steps as the GPS2A scenario. The GAL1A scenario, which uses the E1 frequency data, comprises just one step, as the GPS1A scenario. The GNSS3 scenario uses simulated data from GPS and GALILEO systems at L1, L2, L5, E1, E5a and E5b frequencies. The cascading ambiguity resolution process consists of three steps that estimate firstly the EWL ambiguities for each system, then the WL ambiguities for each system and finally the L1/E1 ambiguities. The L1, L2, E1 and E5b frequency data from GPS and GALILEO are used at the GNSS2A scenario. The two steps of the cascading procedure consist of the WL ambiguity estimation followed by the L1/E1 ambiguity estimation. The GNSS2B scenario uses L1, L5, E1 and E5a frequency data from GPS and GALILEO and the ML ambiguities are estimated before the L1/E1 ambiguity estimation. The GNSS1A scenario comprises just one step to process the L1/E1 data.

The observation session related to the results presented below was simulated based on an aircraft trajectory, at the North Atlantic zone, from (37º 44' N, 25º 40' W) to (39º 27' N, 31º 08' W), on May 12th 2005, between 12h 00m and 14h 10m. For each GPS and GALILEO satellite

above the horizon, pseudorange and phase observables were simulated with a 1 Hz data rate, for the three frequencies. The largest distance between the rover and the reference station was ~512 km. Two levels of NMIR, based on the RMS IONEX maps, were generated: a low level (with 1 RMS) and a medium level (with 2 RMS). The use of RMS maps enables to create an ionospheric error consistent and close to the real situation. The minimum and maximum values concerning the effect of the NMIR on the L1/E1 double difference phase observables, for the different observed satellites, corresponding to 1 RMS, varied between –22.3 cm and 25.4 cm.

Table 6 Simulated data test scenarios.

Scenario	GNSS Type	Phase Frequencies
GPS3	GPS	L1, L2, L5
GPS2A	GPS	L1, L2
GPS2B	GPS	L1, L5
GPS1A	GPS	L1
GAL3	GALILEO	E1, E5a, E5b
GAL2A	GALILEO	E1, E5b
GAL2B	GALILEO	E1, E5a
GAL1A	GALILEO	E1
GNSS3	GPS + GALILEO	L1, L2, L5 + E1, E5a, E5b
GNSS2A	GPS + GALILEO	L1, L2 + E1, E5b
GNSS2B	GPS + GALILEO	L1, L5 + E1, E5a
GNSS1A	GPS + GALILEO	L1 + E1

The simulated data was independently processed twenty four times, using the software described above, for the twelve scenarios and the two levels of NMIR. Using a mask angle of 15º, between 6 and 8 GPS satellites and between 5 and 7 GALILEO satellites were used. With the goal of evaluating the advantage that the GASC has on the proposed ambiguity resolution algorithm, the observations were processed two more times, using the same software but replacing the GASC by the ILSC at the ambiguity resolution function (with RT=1.44 and RT= 2.0), for each scenario and each level of NMIR, using also a 15º mask angle. Table 7 lists the PCIAR values obtained for all the scenarios, using either the GASC or the ILSC, for the two levels of NMIR. It should be pointed out that all the PCIAR values presented are related to the L1/E1 ambiguities. At the steps where the EWL, WL or ML ambiguities were estimated, the corresponding PCIAR values were always 100.

As at the test of real data, the advantage of using the cascading scheme for instantaneous ambiguity resolution is also evident. The results obtained with the GPS1A, the GAL1A and GNSS1A scenarios are always worse than the results obtained with the related scenarios that use two or three frequencies. The merit of using three frequencies is more patent for the medium level of NMIR. Also, the advantage of using the GASC, instead of the ILSC, is clearer for the medium level of NMIR. The PCIAR values reached on all the runs when the ILSC was used are smaller, generally, than those obtained when the GASC was used. For the GNSS scenarios, with a suitable choice of the observed satellites based on its elevation, achievable due to the large number of observed GNSS satellites (between 7 and 12 satellites were available), it was possible to get PCIAR values above 97.7 when two or three frequencies were used together with the GASC, for both levels of NMIR.

Figures 2, 3 and 4 show the values of the distance between the rover and the reference stations related to the epochs where there is a fault in the estimation of the correct ambiguity set, using either the GASC or the ILSC (with RT=1.44 and RT=2.0), obtained when using, respectively, the GPS3, GAL3 and GNSS3 scenarios with the medium NMIR level.

Table 7 PCIAR values for the simulated data tests.

SCENARIO	NMIR LOW			NMIR MEDIUM		
	GASC	ILSC (1.44)	ILSC (2.00)	GASC	ILSC (1.44)	ILSC (2.00)
GPS3	95.8	93.2	89.8	88.5	85.8	82.4
GPS2A	95.4	92.7	89.4	86.8	85.5	82.0
GPS2B	95.3	90.2	85.5	86.2	81.3	73.9
GPS1A	55.2	47.0	28.7	20.2	14.1	10.3
GAL3	100	96.9	93.4	85.3	77.7	66.5
GAL2A	100	88.0	85.1	82.2	66.7	53.7
GAL2B	98.0	85.6	82.8	80.8	55.5	52.0
GAL1A	56.3	37.6	27.7	21.2	16.0	12.2
GNSS3	99.1	99.0	97.0	99.1	96.6	91.2
GNSS2A	99.0	99.0	97.0	98.2	95.7	86.3
GNSS2B	98.7	98.3	96.5	97.7	93.0	80.4
GNSS1A	42.4	33.3	18.7	21.1	15.1	6.2

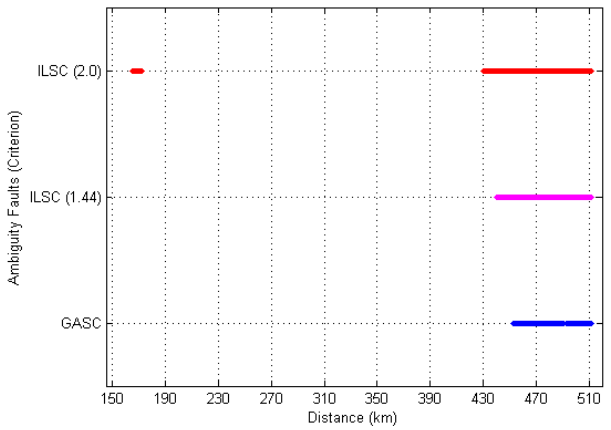

Fig. 2 Faults in ambiguity resolution for the GPS3 scenario, with the medium NMIR level.

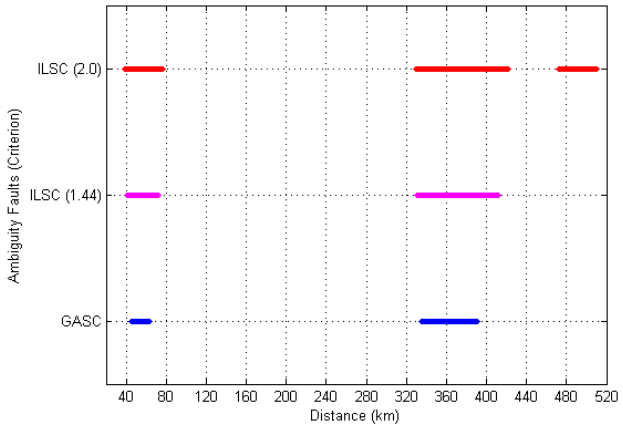

Fig. 3 Faults in ambiguity resolution for the GAL3 scenario, with the medium NMIR level.

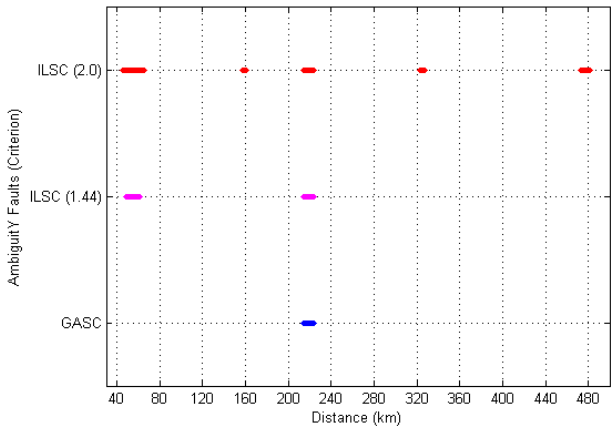

Fig. 4 Faults in ambiguity resolution for the GNSS3 scenario, with the medium NMIR level.

Due to the way that the atmospheric errors were simulated, the ambiguity faults do not increase as the distance between the rover and the reference stations is growing. Tables 8, 9 and 10 list the PCIAR values for each interval of 10 km distance between the rover and the reference stations, where the PCIAR is not 100 for one of the cases (GASC, ILSC with RT=1.44 and ILSC with RT=2), obtained when using the conditions presented in Figures 2, 3 and 4, respectively.

Table 8 Partial PCIAR values for the GPS3 scenario, with the medium NMIR level.

Criterion Distance (km)	GASC	ILSC (1.44)	ILSC (2.0)
160 - 170	100.0	100.0	44.7
170 - 180	100.0	100.0	79.0
430 - 440	100.0	98.1	2.0
440 - 450	100.0	0.0	0.0
450 - 460	22.1	0.0	0.0
460 - 470	0.0	0.0	0.0
470 - 480	0.0	0.0	0.0
480 - 490	0.0	0.0	0.0
490 - 500	12.3	0.0	0.0
500 - 512	0.0	0.0	0.0

Table 9 Partial PCIAR values for the GAL3 scenario, with the medium NMIR level.

Criterion Distance (km)	GASC	ILSC (1.44)	ILSC (2.0)
30 – 40	100.0	98.7	86.8
40 - 50	48.3	0.0	0.0
50 - 60	0.0	0.0	0.0
60 - 70	67.6	0.0	0.0
70 - 80	100.0	70.9	25.2
80 - 90	100.0	57.9	0.0
90 - 100	100.0	100.0	0.0
320 – 330	100.0	100.0	88.2
330 - 340	50.3	5.9	0.0
340 - 350	0.0	0.0	0.0
350 - 360	0.0	0.0	0.0
360 - 370	0.0	0.0	0.0
370 - 380	0.0	0.0	0.0
380 - 390	0.0	0.0	0.0
390 - 400	83.8	0.0	0.0
400 - 410	100.0	0.0	0.0
410 - 420	100.0	83.1	0.0
420 - 430	100.0	100.0	85.0
470 - 480	100.0	100.0	12.3
480 - 490	100.0	100.0	0.0
490 - 500	100.0	100.0	0.0
500 - 512	100.0	100.0	0.0

Table 10 Partial PCIAR values for the GNSS3 scenario, with the medium NMIR level.

Criterion Distance (km)	GASC	ILSC (1.44)	ILSC (2.0)
40 - 50	100.0	81.5	44.4
50 - 60	100.0	0.0	0.00
60 - 70	100.0	89.4	48.3
150 - 160	100.0	100.0	66.2
160 - 170	100.0	100.0	87.5
210 - 220	71.1	70.4	70.4
220 - 230	82.2	82.2	81.6
320 - 330	100.0	100.0	47.1
470 - 480	100.0	100.0	20.8
480 - 490	100.0	100.0	79.9

In the following, taken from the above processed data, three examples are given to illustrate the behaviour of the two terms of (12). It should be remembered that the sum of these two terms, denoted by *Total*, is the quantity that is used by the GASC to select the correct ambiguity set, and that *da* is the quantity that is used to select the correct set of ambiguities when the ILSC is used.

Tables 11 and 12 are related to the epoch 1010, when data from the GALILEO-only system was used and 6 satellites were observed. Table 11 presents the numbers of the ambiguity set candidates that generated the smallest value for *Total* and the smallest and second-smallest values for *da*. When using the GASC, the selection made was the third candidate as it was the one that generated the smallest *Total*. The choice made by the ILSC was the first one, as it produced the smallest value for *da*. If the ratio test were used, in this case, with a critical value 1.44 or 2.0, the ILSC would reject the found minimum. Table 12 shows the values differences between the true ambiguities and the selected candidate set. The GASC selected the right set of ambiguities. The ILSC choice gave a wrong value for the ambiguities (two of the ambiguities differ one cycle from the true value) but, as this set was not validated, there was no solution, given by the ILSC, for that epoch.

Table 11 Ambiguity candidates, for epoch 1010, using the GAL3 scenario, for medium NMIR.

Candidates	*da*	*db*	*Total*
01	1.720	1.406	3.126
02	2.068	1.694	3.762
03	2.072	0.228	2.300

Table 12 Search result, for epoch 1010, using the GAL3 scenario, for medium NMIR.

Ambiguities	1	2	3	4	5
ILSC selection – 1st candidate	-1	0	1	0	0
GASC solution – 3rd candidate	0	0	0	0	0

Tables 13 and 14 show the results obtained with data from the epoch 1080, also when data from the GALILEO-only system was used and 6 satellites were observed. In this case the ILSC generated the correct set of integer ambiguities but that set was not validated by the ratio test (using RT=1.44 or RT=2.0) and a solution was not given for that epoch. The GASC, with the same candidate vector, because its uniqueness principle, reached the total minimum uniquely and presented the correct solution.

Table 13 Ambiguity candidates, for epoch 1080, using the GAL3 scenario, for medium NMIR.

Candidates	*da*	*db*	*Total*
01	1.300	0.165	1.465
02	1.616	1.330	2.946

Table 14 Search result, for epoch 1080, using the GAL3 scenario, for medium NMIR.

Ambiguities	1	2	3	4	5
ILSC selection – 1st candidate	0	0	0	0	0
GASC solution – 1st candidate	0	0	0	0	0

Tables 15 and 16 describe the selection made by both criteria for the epoch 3440, when data from the GALILEO-only system was used and 6 satellites were observed. Both criteria selected the correct set of integer ambiguities and presented the solution.

Table 15 Ambiguity candidates, for epoch 3440, using the GAL3 scenario, for medium NMIR.

Candidates	*da*	*db*	*Total*
01	0.121	0.068	0.189
02	0.665	0.554	1.219

Table 16 Search result, for epoch 3440, using the GAL3 scenario, for medium NMIR.

Ambiguities	1	2	3	4	5
ILSC solution – 1st candidate	0	0	0	0	0
GASC solution – 1st candidate	0	0	0	0	0

6 Conclusions

The results above show, as similar results from other tests carried out, that the GASC represents a clear improvement in the selection of the correct set of ambiguities. It may be concluded that the GECCAR approach is a very promising algorithm for instantaneous ambiguity resolution. Simulation runs have shown that single-epoch ambiguity resolution was possible 99% of the considered epochs, when the three frequencies from both systems were used together. Benefits from a multi-epochs approach remain to be studied, but a performance gain is expected.

Acknowledgements

The authors would like to thank Dr. Guochang Xu for the many valuable discussions about ambiguity resolution, and the anonymous reviewers for their many constructive comments. D. B. Morujão is supported through *Fundação para a Ciência e a Tecnologia* grant SFRH/BD/28678/ 2006.

References

Alves, P. (2001). ***The effect of Galileo on carrier phase ambiguity resolution.*** *Proceedings of ION GPS 2001*, The 14th International Technical Meeting of the Satellite Division of the Institute of Navigation, Salt Lake City, Utah, USA, September 11-14, pp. 933-944.

Boehm, J., A E. Niell, P. Tregoning, H.Schuh (2006). ***Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data.*** Geoph. Res. Letters, **33**, L07304, DOI 10.1029/2005GL025546.

Euler, H-J., H. Landau (1992). ***Fast GPS Ambiguity Resolution On-The-Fly for Real-Time Applications***. *Proceeding of the 6th International Geodetic Symposium on Satellite Positioning*, Columbus, Ohio, 17-20 March, pp. 650-659.

Forsell, B., M. Martín-Neira, R.A.Harris (1997***). Carrier phase ambiguity resolution in GNSS-2.*** *Proceedings of ION GPS-97*, The 10th International Technical Meeting of the Satellite Division of the Institute of Navigation, Kansas City, Missouri, September 16-19, pp. 1727-1736.

Jong, C. D. de, N. F. Jonkman, P. Joosten (1999***). The benefits of a third frequency to GNSS users***. *Proceedings of NIN workshop: GPS – De heilige grall?*, Nederlands Instituut voor Navigatie, November 5, Delft.

Jung, J. (1999). ***High integrity carrier phase navigation for future LAAS using multiple civilian GPS signals***. *Proceedings of ION GPS-99*, The 12th International Technical Meeting of the Satellite Division of the Institute of Navigation, Nashville, USA, September 14-17, pp. 727-736.

Jung, J., P. Enge, B. Pervan (2000). ***Optimization of cascade integer resolution with three civil frequencies***. *Proceedings of ION GPS 2000*, The 13th International Technical Meeting of the Satellite Division of the Institute of Navigation, Salt Lake City, Utah, USA, September 19-22, pp. 2191-2200.

Mendes, V. B. (1999). ***Modeling the neutral-atmosphere propagation delay in radiometric space techniques***. Ph. D. dissertation, Department of Geodesy and Geomatics Engineering Technical Report Nº. 199, University of New Brunswick, Fredericton, New Brunswick, Canada, 353 pp.

Saastamoinen, I. (1973***) Contribution of the theory of atmospheric refraction.*** In three parts, *Bulletin Géodésique*, No. 105, pp. 279-298, No. 106, pp.383-397, No. 107, pp. 13-14.

Teunissen, P.J.G. (1993). ***Least squares estimation of the integer GPS ambiguities.*** Invited lecture, Section IV: Theory and methodology, IAG General Meeting, Beijing, China, August 1993. Also in LGR-Series n. 6, Delft Geodetic Computing Centre, Delft University of Technology, Delft.

Teunissen, P.J.G. (1994***). The invertible GPS ambiguity transformations.*** Manuscripta Geodaetica, 20 (6): 489-497.

Teunissen, P.J.G. (1998). ***GPS carrier phase ambiguity fixing concepts.*** In: Teunissen, P.J.G., Kleusberg (Eds.), *GPS for Geodesy*, Springer Verlag, Berlin, Chap. 8, pp. 319-388.

Teunissen, P.J.G. (2003)***. Theory of carrier phase ambiguity resolution***. Wuhan University Journal of Natural Sciences, 8(2b), 471-484.

Tiberius, C.C.J.M, P.J.G. Teunissen, and P.J. de Jonge (1997). ***Kinematic GPS: performance and quality control.*** Proceedings of International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation KIS'97, Banff, Canada, June 3-6, pp. 289-299.

Verhagen, S. (2004). ***Integer ambiguity validation: an open problem?*** *GPS Solutions*, 8(1):36-43, DOI 10.1007/ s1029100400875.

Vollath, U., S. Birnbach, H. Landau (1998). ***An analysis of three-carrier ambiguity resolution (TCAR) technique for precise relative positioning in GNSS-2.*** *Proceedings of ION GPS-98*, The 11th International Technical Meeting of the Satellite Division of the Institute of Navigation, Nashville, Tennessee, USA, September 15-18, pp. 417-426.

Wei M. Schwarz K.-P. (1995). ***Fast ambiguity resolution using an integer nonlinear programming method***. *Proceedings of ION GPS-95*, The 8th International Technical Meeting of the Satellite Division of the Institute of Navigation, Palm Springs, 12-15 Sept, pp. 1101-1110.

Xu, G. (2002). ***A General Criterion of Integer Ambiguity Search.*** *Journal of Global Positioning Systems*, Vol. 1, Nº 2, pp.122-131.

Xu, G. (2004***). MFGsoft – Multi-Functional GPS/(Galileo) Software***, (Version of 2004), Scientific Technical Report STR04/17, GeoForschungsZentrum (GFZ), Potsdam.

Xu, G. (2007). ***GPS. Theory, Algorithms and Applications***. Second edition. Springer-Verlag, Berlin, Heidelberg.

Xu, P. (2006). ***Voronoi Cells, Probabilistic Bounds, and Hypothesis Testing in Mixed Integer Linear models***. IEEE Transactions on Information Theory, Vol. 52, Nº 7, pp. 3122-3138, DOI 10.1109/TIT.2006.876356.

Zandbergen, R., S. Dinwiddy, J. Hahn, E. Breeuwer and D. Blonski (2004). ***Galileo orbit selection.*** *Proceedings of ION GNSS 2004*, The 17th International Technical Meeting of the Satellite Division of the Institute of Navigation, September 21-24, Long Beach, CA, pp. 616-623.

Zhang, W., M.E. Cannon, O. Julien, P. Alves (2003). ***Investigation of combined GPS /GALILEO cascading ambiguity resolution schemes.*** *Proceedings of ION GPS/GNSS 2003*, The 16th International Technical Meeting of the Satellite Division of the Institute of Navigation, Portland, OR, USA, September 9-12, pp. 2599-2610.

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 46-61

Journal of Global Positioning Systems
2008

A Novel Architecture for Ultra-Tight HSGPS-INS Integration

Guojiang Gao and Gérard Lachapelle
Positioning, Location and Navigation Group (PLAN)
Department of Geomatics Engineering, University of Calgary, Alberta, Canada

ABSTRACT

Global Positioning System (GPS) currently fulfills the positioning requirements of many applications under Line-Of-Sight (LOS) environments. However, many Location-Based Services (LBS) and navigation applications such as vehicular navigation and personal location require positioning capabilities in environments where LOS is not readily available, e.g., urban areas, indoors and dense forests. Such environments either block the signals completely or attenuate them to a power level that is 10-30 dB lower than the nominal signal power. This renders it impractical for a standard GPS receiver to acquire and maintain signal tracking, which causes discontinuous positioning in such environments.

In order to address the issue of GPS tracking and positioning in degraded signal environments, a novel architecture for ultra-tight integration of a High Sensitivity GPS (HSGPS) receiver with an inertial navigation system (INS) is proposed herein. By enhancing receiver signal tracking loops through the use of optimal estimators and with external aiding, the capabilities of the receiver can be substantially improved. The proposed approach is distinct from the commonly used ultra-tightly coupled GPS/INS approaches and makes use of different tracking enhancement technologies used in typical HSGPS receivers, multi-channel cooperated receivers and the current ultra-tightly coupled GPS/INS methods. Furthermore, the effects of inertial measurement unit (IMU) quality, receiver oscillator noise and coherent integration time on weak signal tracking are also analyzed.

Simulated test results in both static and dynamic testes show that, the designed INS-aided GPS receiver can track the incoming weak GPS signals down to 15 dB-Hz without carrier phase locked, or 25 dB-Hz with carrier phase locked. When there are multiple strong GPS signals in view, the other weak signals can be tracked down to 15 dB-Hz with carrier phase locked.

KEY WORDS: ULTRA-TIGHT INTEGRATION, HSGPS, INS

INTRODUCTION

Standard Global Positioning System (GPS) technologies fulfill the positioning requirements of many applications intended for environments with clear Line-Of-Sight (LOS) to satellites. However, many Location Based Services (LBS) and applications such as vehicular navigation and personal location require positioning capabilities in environments where LOS to satellites is not readily available, e.g., urban areas, indoors and dense forest areas (e.g., Lachapelle et al., 2003). Such environments either completely block the GPS signals or attenuate them to a power level that is 10-30 dB lower than the nominal signal power (van Diggelen and Abraham, 2001). This makes it impractical for a standard receiver to acquire and maintain signal tracking, which causes discontinuous positioning in such environments.

In degraded signal environments, e.g., urban canyons and indoors, positioning availability and accuracy are affected by weak signal power, strong multipath/echo-only signals and receiver dynamics. The characteristics of degraded GPS signal environments are summarized in Table 1. In these environments, signal attenuation and strong specular reflectivity are primary sources of signal degradation. For vehicle navigation in urban areas, multipath/echo-only signals constitute interference sources that change quickly and behave randomly due to vehicle motion (MacGougan, 2003). The signal intensity for personal positioning in indoor environments (e.g., fireman positioning in buildings) is commonly 20-30 dB lower than that found outdoors (Lachapelle, 2007)
To address this issue, High Sensitivity GPS (HSGPS) technologies (Watson et al., 2006), Assisted GPS (AGPS) systems (van Diggelen and Abraham, 2001),

multi-channel co-operated receivers (Zhodzishsky et al., 1998) and cellular network-based solutions (Ma et al., 2007) have been developed. However, these technologies and systems still fail to maintain continuity of positioning with acceptable accuracies, specifically in the indoors. Thus, new receiver technologies have to be explored for enhanced signal acquisition and tracking performance.

Recently, ultra-tight integration of GPS and inertial navigation systems has received considerable attention for this purpose. In an INS-assisted GPS receiver, which is also called ultra-tightly coupled or deeply integrated GPS/INS, an external INS is used to provide receiver dynamics information to allow GPS receiver to do long coherent integration to track weak signals in sight (Soloviev et al., 2004). Measuring receiver dynamics through INS aiding enables the INS-assisted GPS receiver to track an incoming weak signal which is 20-30 dB lower or more than normal and therefore projects a strong light beam into the "indoor darkness" (Beser et al., 2002; Soloviev et al., 2004b; Kreye et al., 2000; Sennott, 1997).

Table 1 Characteristics in Different Operating Environments and Applications

Features	Urban Canyon Vehicle Navigation	Indoor Personal Positioning
Signal Fading	10-30 dB	20-30 dB
Multipath Signal	Strong, high frequency	Strong, low frequency
Platform Dynamics	Moderate	Low
Map Matching	Easy to implement	Maps not readily available
Desired Receiver Size	Moderate	Small

Table 2 (Gao, 2007) summarizes the performance of HSGPS, multi-channel co-operated receivers (also called COOP tracking receivers) and INS-assisted GPS receivers. It shows that an INS-assisted GPS receiver is far superior to the other positioning technologies mentioned above and offers the greatest potential for meeting navigation and positioning requirements under attenuated signals. In INS-assisted GPS receivers, velocity aiding from INS enhances the GPS phase lock loops (PLL), which are the weakest loops in the receiver. Furthermore, full navigation capability, including carrier phase output under attenuated signals, is preserved in INS-assisted GPS receivers. This availability of accurate carrier phase measurements is deemed necessary for many high-accuracy applications.

The paper continues with the discussion on the current ultra-tightly coupled GPS/INS systems and other weak signal tracking technologies, such as HSGPS and multi-channel co-operated receivers. Then, the design of a novel INS-assisted GPS receiver for degraded GPS signal tracking is introduced. The proposed architecture is distinct from current ultra-tightly coupled GPS/INS systems and uses a combination of different tracking technologies like HSGPS or multi-channel co-operated GPS receivers or traditional ultra-tightly coupled GPS/INS approaches. Some system design issues, such as IMU quality requirements, the limits of very long coherent integration time and receiver clock error compensation, are also addressed. Information about the testing tools utilized herein, including an INS simulator developed for this purpose, is provided. Test results and analysis are then presented using simulated data sets to assess the performance of the INS-assisted GPS receiver, followed by conclusions.

CURRENT ULTRA-TIGHTLY COUPLED GPS/INS SYSTEMS

Based on the type of Kalman filter used, ultra-tight integration can be implemented in three different ways (Gao, 2007), namely: (1) loosely coupled Kalman filter-based ultra-tight integration, (2) tightly coupled Kalman filter-based ultra-tight integration, and (3) ultra-tightly coupled Kalman filter-based ultra-tight integration. Figure 1 summarizes the different architectures. Figure 2 shows two different types of architectures for INS-assisted GPS receivers, as proposed by Gautier and Parkinson (2003), Alban et al. (2003) and Gustafson et al. (2000).

The first architecture shown in Figure 2(a) is based on a loosely or tightly coupled integration scheme. All individual DLL (delay lock loops) and PLL are inside the receiver. The Kalman filter utilizes either raw measurements or processed positions and velocities from the GPS receiver to update the INS periodically. The updated INS information is then used to predict the phase and Doppler used as aiding to the receiver. Thus, based on the type of measurements used for updating the INS, these strategies can be classified as loosely coupled Kalman filter-based ultra-tight integration or tightly coupled Kalman filter-based ultra-tight integration. However, in the second architecture shown in Figure 2(b), an ultra-tightly coupled Kalman filter is used in place of conventional in-receiver PLLs and, in some cases, even DLLs. This filter operates on in-phase (I) and quadra-phase (Q) components of the signal directly. This

integration strategy is referred to as ultra-tightly coupled Kalman filter-based ultra-tight integration.

Most of the work done in ultra-tight integration of GPS and INS has focused on the above three integration strategies. Although these new architectures offer flexibility, from the point of view of information theory, it is suggested herein that simply adopting different kinds of Kalman filters will not improve positioning performance of GPS/INS integrated systems significantly. This concept is best illustrated by an analogy of water in a river. Although water looks very different at higher, medium, and lowest points of the river, the volume of water is the same at all these points. This is the case for line-of-sight environments, where a tightly coupled Kalman filter will not provide significant improvements in an integrated system performance as compared to a loosely coupled filter.

There are also some specific limitations in present architectures. The first limitation is that the receiver tracking capability is sensitive to IMU quality. For reliable aiding from INS, a velocity accuracy of 1 cm/s along the LOS direction is required from the INS solution (Soloviev et al., 2004a), which requires a high quality IMU.

Table 2 Performances of Different Positioning Methods under Attenuated Signals

Items	HSGPS	CO-OP Tracking	GPS/INS Ultra-tight integration	Notes
Tracking Sensitivity	Good	Good	Excellent	15-25 dB lower than regular signals for HSGPS and CO-OP tracking; 20-30 dB for ultra-tight integration
Acquisition Sensitivity	Poor	Good	Excellent	Because of long integration time, long Time-To-First Fix (TTFF) for HSGPS; INS measurements and/or information from other tracking channels can be used to speed up the acquisition process in ultra-tight integration and CO-OP tracking, especially in hot start.
Re-Acquisition Capability	Poor	Good	Excellent	Due to long pre-detection integration time (PIT), re-acquisition is time-consuming in HSGPS; Aiding from INS measurements facilitates rapid re-acquisition in ultra-tight integration.
Position Data Up Rate	Low	Low	High	In ultra-tight integration, the data rate can be increased to above 100 Hz using INS aiding, with a Kalman filter running at a low recursive rate.
Positioning Accuracy	Poor	Good	Excellent	In HSGPS, positioning accuracy is degraded by multipath signal and frequency/phase tracking error; In ultra-tight integration, INS solution can help in blunder detection and noise compression (by using long time integration).
Carrier Phase Output	Poor	Good	Excellent	In HSGPS, limited benefit for PLL tracking, and thus difficult to output carrier phase observation; Ultra-tight integration method can output precise phase observation and avoid/reduce cycle slips.
Dynamic Response	Poor	Poor	Excellent	HSGPS is used mainly for low dynamic users. Ultra-tight integration can be used for both low and high dynamic users and thus in both commercial and military applications
Receiver Size	Small	Small	Moderate/Big	For HSGPS, no need for any other hardware; a good size under Ultra-tight integration for MEMS IMU
Power Cost	Low	Low	Moderate/High	For HSGPS, no other hardware required so no additional power cost. For Ultra-tight integration, additional external sensor needed, so more power is required.
Multipath Mitigation	Poor	Good	Excellent	Ultra-tight integrated navigator can detect multipath signals and track weak LOS signals directly in urban areas and indoor environments.

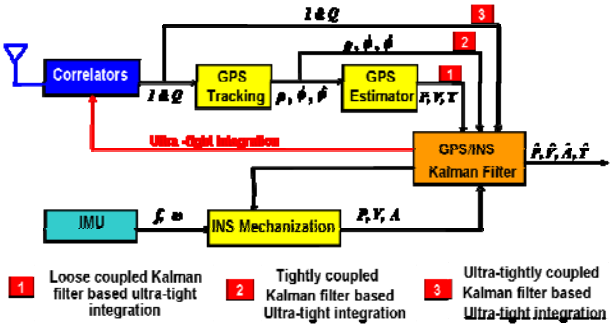

Fig. 1 Ultra-tightly Coupled GPS/INS with Loosely, Tightly or Ultra-tightly Coupled Kalman Filter

The second limitation is that before the integrated system moves to attenuated signal environments, it has to be initialized under LOS environments, which includes GPS-only receiver initial acquisition and INS initial alignment.

Furthermore, the accuracy estimation of a Kalman filter relies on the assumption of correct stochastic modeling of both system measurement errors, which may not be possible in severe urban canyon environments. For vehicle navigation in urban canyons, GPS measurement faults such as those caused by multipath or echo-only signals are very significant. Also, IMU measurement errors may not be compensated effectively, which ultimately degrades the receiver tracking performance.

Small INS velocity errors usually remain after the INS is well aligned with GPS measurements and the vehicle's motion contains sufficient dynamics to make all the INS error states observable. However, in the static case or where the vehicle motion does not contain enough dynamics to ensure observability, the accuracy of the Doppler information from the INS may be of poor quality. A particular concern is the poorly observable states of the azimuth and gyro bias. If the estimates of these states are of poor quality, the velocity component errors provided by a low cost INS in static situations can be very large. It should also be noted that the dynamics that needs to be estimated and compensated by the INS are not limited to the LOS velocity component but to all other components. For example, an erroneous azimuth estimate can result in dynamics errors due to "phase wind-up" effects (Tetewsky and Mullen, 1996; Don et al., 2005).

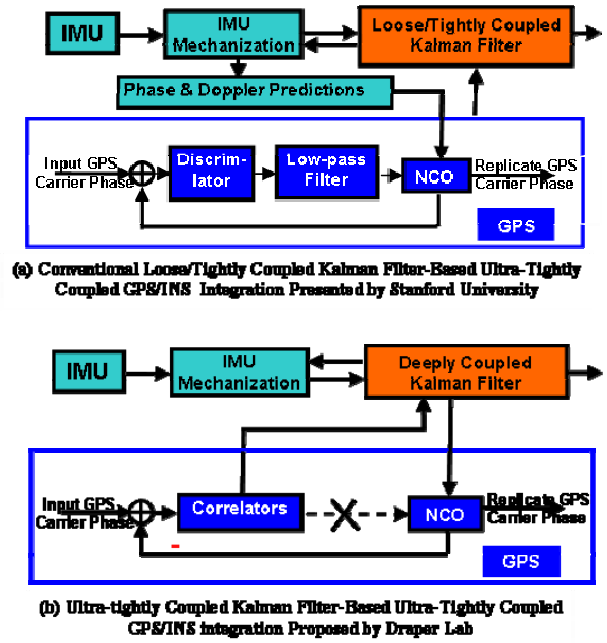

Fig. 2 Two Different Architectures of Current Ultra-tightly Coupled GPS/INS

Furthermore, when a low cost MEMS-based INS is involved in an INS-assisted GPS receiver, because of the very low accuracy and very poor stability of IMU sensors, the INS velocity errors might contain jumps or blunders. These large errors may cause major problems in an INS-assisted GPS receiver. Given the GPS/INS Kalman filter observability issue discussed above, MEMS-based INS/GPS ultra-tight integration is still a major challenge.

DESIGN OF NOVEL INS-ASSISTED GPS RECEIVER

To overcome the limitations of current INS-assisted GPS receivers, a technique based on multi-channel co-operated tracking (COOP tracking) is proposed herein to estimate and track the Doppler prediction errors caused by INS errors. This technique is fully described in Gao (2007). Initial results were presented by Gao and Lachapelle (2006).

The architecture of the proposed integration system is shown in Figure 3. The ultra-tightly integrated system used the software receiver GNSS_SoftRx^{TM} (Ma et al., 2004) as a starting point. The proposed strategy includes three loops. The first loop includes the conventional loosely or tightly coupled Kalman filter, which predicts

the user Doppler based on information from the INS. In this study, a loosely coupled Kalman filter is suggested. This loop provides external Doppler aiding and clock error correction from the INS. In order to decrease the effects of INS positioning errors on receiver Doppler prediction and receiver clock error compensation, a COOP loop is used as the second loop to estimate the carrier Doppler error and receiver clock error caused by INS positioning errors. Thus, the INS and COOP loops operate together in order to provide a nearly perfect reference for receiver dynamics and receiver clock errors. Since receiver dynamics is removed from the signal, long coherent/non-coherent integration time for individual PLLs and DLLs becomes possible, which constitutes the third loop of the proposed architecture. Thus, the PLL/DLL loops track the differences between the incoming and the local signals, which have been compensated by INS and COOP loops.

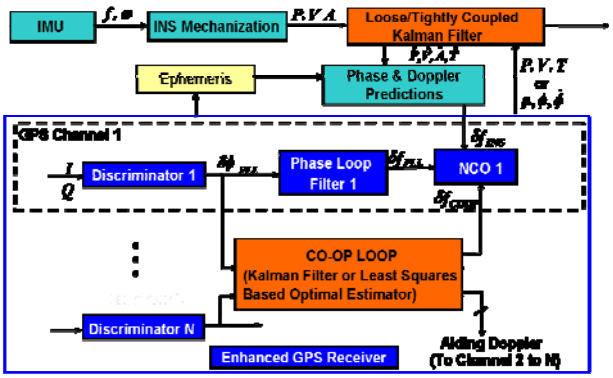

Fig. 3 Proposed Architecture of Ultra-tightly Coupled GPS/INS

The characteristics and advantages of the proposed system can be summarized as follows:

1 The HSGPS receiver optimizes parameterization and employs data wipe off technology for long coherent integration. This structure enables initialization in a weak signal environment.

2 To provide INS aiding to the GPS receiver, a loosely/tightly coupled GPS/INS approach such as that used by Petovello (2003) in the SAINT™ software is suggested to handle INS measurements and provide a corrected INS solution for receiver aiding.

3 The multi-channel cooperated tracking loops track the weak signals and eliminate the effects of INS errors. The design of this estimator is discussed later in this section.

4 The individual DLL/PLLs are different from those of traditional ultra-tightly coupled GPS/INS systems, which adopt ultra-tightly coupled Kalman filters for signal tracking. In this design, traditional sophisticated DLL/PLLs are still located in every individual signal tracking channels, as shown in Figure 3. Furthermore, the individual DLL/PLLs are enhanced using a data wipe off technology mentioned in item 1 to perform very-long coherent integration. The individual DLL/PLLs are combined with INS aiding loops and COOP loops to track both strong and weak signals.

5 Since INS is used in this approach to mainly provide receiver dynamics information, any other sensors such as odometers or radars can replace INS sensors to provide Doppler measurements for the enhanced receiver. This provides an effective capability to re-configure the software receiver to suit various aiding hardware.

A. Multi-Channel CO-OP Tracking Loop

Figure 4 shows the modular design of the multi-channel co-operated tracking loops, i.e., the COOP tracking loops. The basic strategy of COOP tracking is to project signals from the channel domain to the position domain and then try to track/estimate the signals in the position domain. Then the signals are projected back to the channel domain for Doppler removal (Zhodzishsky et al., 1998).

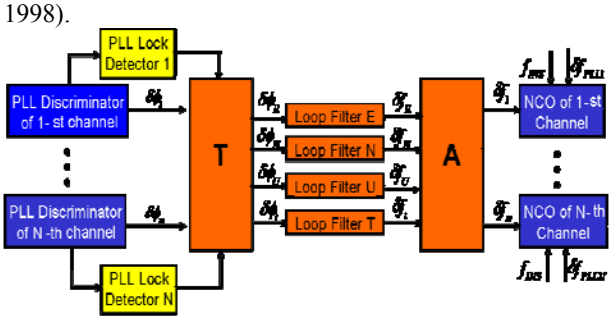

Fig. 4 Architecture of co-operated (CO-OP) Tracking Module

In Figure 4,

$$T = T^n = C_e^n \, T^e$$
$$A = A^n = A^e \, C_n^e \qquad (1)$$

where the superscripts n and e represent the Local-Level frame (LLF) and Earth-Centered Earth-Fixed (ECEF) frame, respectively. C is the rotation matrix and A is the direction cosine matrix (also called geometry matrix). T is the transfer matrix and defined as follows:

$$T^n = \left(\left(A^n\right)^T C_\phi^{-1}\left(A^n\right)\right)^{-1}\left(A^n\right)^T C_\phi^{-1} \quad (2)$$

C_ϕ^{-1} is a weighted matrix determined by FLL lock detectors in individual PLLs.

The COOP loop can be based on either a least-squares estimation method or on a Kalman filter. In this study, the least-squares method is used. Least-squares estimation is an effective optimal estimation method, especially when measurement redundancy is high. Many fault detection algorithms such as Receiver Autonomous Integrity Monitoring (RAIM) are in fact based on least-squares estimation, provided redundant measurements are available. Herein, the principle of "using an optimal estimator in GPS positioning to fully utilize measurement redundancy", is put forward from the measurement processing domain into the GPS baseband signal processing domain, which leads to the COOP tracking method. COOP loops allow a GPS receiver to do signal tracking based on the multi-channel vector tracking approach. Furthermore, based on measurement redundancy, blunder detection algorithms and adaptive estimation methods now can be realized at the signal processing stage rather than at the measurement processing stage.

B. Effect of IMU Quality

In a loosely or tightly coupled GPS/INS for use in urban canyons, the receiver frequently loses phase lock on incoming signals. Therefore the INS in this kind of integrated system must be able to accommodate GPS outages for relatively long periods, e.g., up to 30 s. In an ultra-tightly coupled GPS/INS however, the receiver can track weak GPS signals continuously because of INS aiding. Thus, the INS is constantly corrected by GPS measurements which are typically available every 1 s. Consequently, the maximum INS prediction duration may be limited to 1 s in many common scenarios. This implies that a low-cost Micro Electro-Mechanical System (MEMS) IMU might be acceptable for ultra-tightly coupled GPS/INS for certain applications.

INS velocity errors caused by sensor errors over one second can be estimated using velocity error signatures as (Scherzinger, 2004)

$$\begin{aligned} \delta V &= b_a\, t = b_a \\ \delta V &= -\frac{1}{6} g b_g t^2 = -\frac{1}{6} g b_g \end{aligned} \quad (3)$$

where δV is the velocity error, b_a is the accelerometer bias and b_g is the gyro bias, g is the gravity, and t is the time interval.

With COOP loops, the effects of an INS positioning error on the receiver Doppler prediction and clock error compensation will be limited. Therefore, the accuracy for aiding velocity need not be accurate to 1 cm/s level any more, as pointed out by Soloviev et al. (2004a). In this research, a velocity error of 0.1 m/s is used for the Doppler aiding accuracy, which makes it feasible to use a lower grade IMU quality with either an accelerometer bias of 10 mg or a gyro bias of 3.4˚/s.

In actual applications, a velocity error of 0.1 m/s in 1 s can be achieved with many MEMS IMUs available today, such as the Crista IMU from Cloud Cap Technology, as described by Godha and Cannon (2005). The in-run gyro biases and accelerometer biases of this unit are about 0.3 ˚/s and 2.5 mg, respectively.

C. Effect of Allan Oscillator Phase Noise on Carrier Phase Tracking

The measurement characteristics of a low-cost Temperature-Compensated Crystal Oscillator (TCXO) are similar to those of a low-cost gyro. Oscillator clock errors can be divided into turn-on bias, in-run drift and the remaining colored noise components, the latter being characterized by the Allan Variance. In most GPS/INS systems, the clock bias and drift are estimated and then compensated using a Kalman filter.

In theory, since the clock noise described by the Allan variance is characteristically colored, it can be modeled and thus partly estimated by GPS/INS Kalman filters. If the colored noise is modeled perfectly, the remaining part will be limited to white noise, which can be regarded as thermal noise and easily handled by receiver tracking loops. To simplify the filter design, Allan clock noise is regularly assumed to be white noise (Brown and Hwang, 1992) and thus will not be estimated. Most of the energy of the colored noise is located in the low frequency band in the frequency domain. For this reason, when low-pass loop filters in a receiver try to eliminate the thermal noise from the signal, most of the clock noise passes through these low-pass filters, since it is mixed with the GPS signal in the low frequency band of the spectrum. Therefore, Allan oscillator phase noise must be considered in receiver design, especially in weak signal environments. The effect of the correlated clock noise on signal tracking is discussed in Kaplan (1996).

D. Determination of Coherent Integration Time for Internal Individual DLLs/PLLs

The tracked GPS signal power after accumulators can be expressed as (Raquet, 2004)

$$I = \frac{AM_E}{\sqrt{2}} \frac{\sin(\pi \Delta f T)}{\pi \Delta f T} R(\tau) D \cos(\pi \Delta f T + \phi_0)$$
$$Q = \frac{AM_E}{\sqrt{2}} \frac{\sin(\pi \Delta f T)}{\pi \Delta f T} R(\tau) D \sin(\pi \Delta f T + \phi_0) \quad (4)$$

where M_E is the number of samples accumulated in one sample period, Δf is the frequency error over the integration interval, R is the self-correlation function of PRN code, and D is the navigation data bit modulated on the signal. In Equation (4), the tracked signal undergoes a power loss due to the aiding velocity errors, with the loss characterized by the function $\frac{\sin(\pi \Delta f T)}{\pi \Delta f T}$. Since the L1 carrier wavelength is about 19 cm, a velocity error of 0.1 m/s will lead to a maximum Doppler error of 0.5 Hz along a given LOS vector. The resulting signal power loss over the total integration time due to this phenomenon is shown in Figure 5. A coherent integration time of 1 s would result in a power loss of 4 dB.

The design of receiver tracking loops using continuous update approximation is discussed in details in Kaplan (1996). Stephens and Thomas (1995) have shown that, when the product of loop bandwidth (B_n) and coherent integration time (T_{coh}) is much greater than 0.1 or close to 1, the continuous update approximation does not hold. Gao (2007) and Ilir et al. (2007) give the expressions for the signal tracking errors of both FLL and PLL in a digital GPS receiver. Since the Doppler uncertainty from INS aiding is 0.5 Hz, as shown in Figure 5, coherent integration time should be shorter than 0.2 s to satisfy the condition $B_n \times T_{coh} < 0.1$ *or* $B_n \times T_{coh} << 1$.

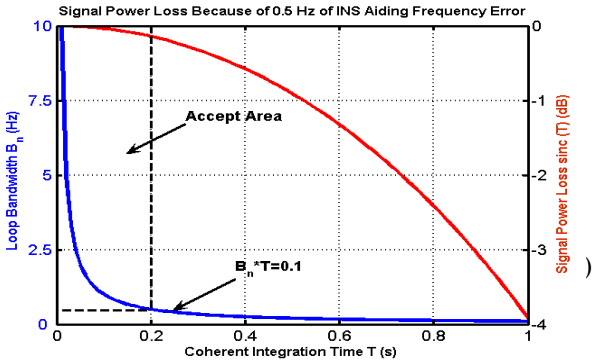

Fig. 5 Signal Power Loss over Integration Time

Another factor that limits the choice of very long coherent integration time is the navigation data bit transition. Soloviev et al. (2004a) presents an energy-based bit estimation algorithm to account for possible bit transitions during signal integration. The algorithm searches for the bit combination every 20 ms to maximize signal energy over the tracking integration interval. This approach presumes that the right bit combination is the one that maximizes the signal energy over the tracking integration interval, namely, the Signal-to-Noise Ratio (SNR) after 20 ms of integration should be above zero. This yields a limitation of -157 dBm for this approach (Gao, 2007), as given by:

$$\begin{aligned} C &= N_0 B \, SNR_{I20} \\ &= -204 + 10 \log\left(\frac{1000}{20}\right) + 0 \\ &= -157 \; dBm \end{aligned} \quad (5)$$

where SNR_{I20} is the signal-to-noise ratio after 20 ms coherent integration, N_0 is the environmental thermal noise and $N_0 = -204 \; dBm$. B is the bandwidth of the phase tracking loop using 20 ms coherent integration time and $B = 10 \log(\frac{1000}{20}) \; Hz$.

Equation (5) illustrates that an incoming signal lower than -157 dBm, which equals to 17 dB-Hz, will fail the assumption implied in this algorithm and, hence, the energy-based bit detection approach. In this case, increasing the length of coherent integration time cannot help to track this very-weak signal.

In consideration of the above factors, a maximum coherent integration time of 100 ms is used in this paper.

TEST SETUP AND TEST RESULTS

Several static and dynamic tests were conducted to assess the performance of the above INS-assisted HSGPS receiver. In the static case, a scenario where a GPS receiver is tracking strong and weak signals at the same time is simulated to examine the performance of COOP loops. In the dynamic case, a scenario where a GPS receiver is tracking all weak signals at the same time is simulated to assess the tracking sensitivity of the ultra-tightly coupled GPS/INS system.

A. Simulation Test Tools

For simulating the incoming GPS signal and IMU measurements, the GPS simulator GPS_GenTM, developed previously by Dong et al (2004), and an INS simulator INS_Sim developed by Gao (2007) were used. Figure 6 shows the architecture of the INS simulator.

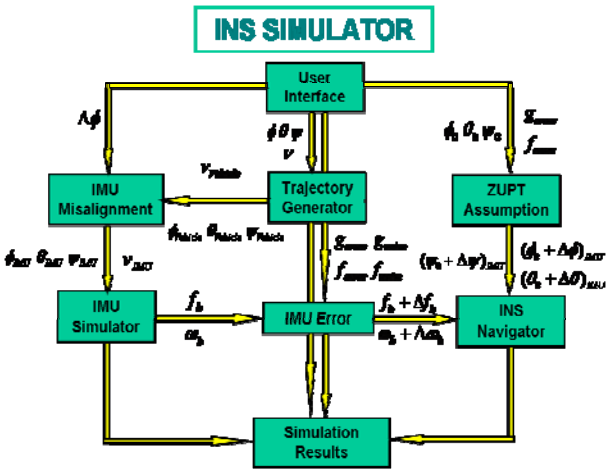

Fig. 6 The Architecture of INS Simulator

B. Static Test Scenario

The test summary is shown in Figure 7. The total duration of the data collection was 50 s. During the first 14 s, the system initialization, including GPS signal acquisition and ephemeris collection, was performed. Then, from 14 to 50 s, the INS output, i.e., the user velocity and position from the INS, was fed to the GPS receiver to assist with the GPS signal tracking. To simplify the test analysis, INS operated in stand-alone mode for this 36 s aiding period, such that it kept accumulating errors. The GPS simulator scenario was designed to output strong GPS signals (45 dB-Hz) for all satellites, except for PRN 07, as illustrated in Figure 7. For PRN 07, the signal power was 45 dB-Hz during the first 20 s period and then reduced gradually to 15 dB-Hz by the GPS simulator over the next 20 s period. During the last 10 s period, the signal power of PRN 07 was kept at 15 dB-Hz.

The error characteristics of the IMU data simulated were similar to a tactical grade HG1700 IMU, as discussed by Petovello (2003). Figure 8 shows the interface of INS_Sim and

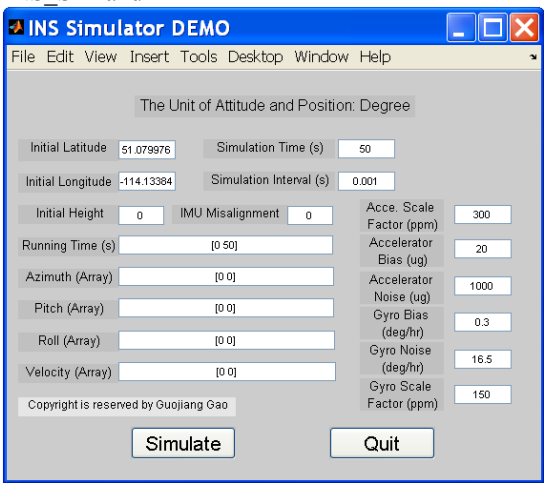

Fig. 8 The Interface of INS Simulator

Table 3 lists the parameters used for simulating the IMU data. The simulation assumes that the INS has been initialized with GPS measurements. So the accelerometer bias residual and gyro bias residual were limited to around 20 μg and 0.3 ˚/hr respectively. The IMU noise bandwidth was kept at about 10 Hz so that the gyro noise was equal to 16.5 ˚/hr. The pitch and roll alignment errors were 0.01˚ and the heading alignment error was 0.05˚.

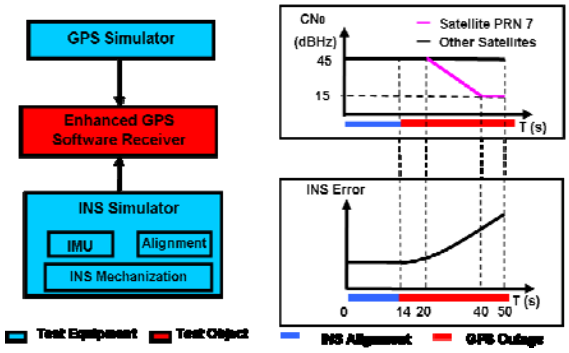

Fig. 7 Static Test Setup for INS-assisted HSGPS Receiver

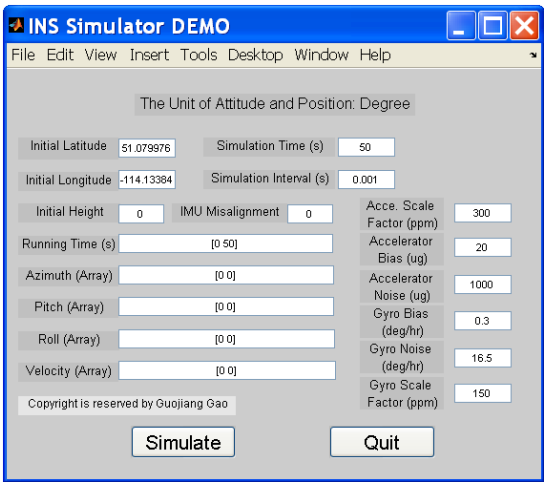

Fig. 8 The Interface of INS Simulator

Table 3 Simulated-INS Parameters

	Accelerometer	Gyro
Scale Factor	300 ppm	150 ppm
Bias Residual	20 μg	0.3 °/hr
Random Noise	1000 μg	$5.5°/\sqrt{hr \cdot Hz}$

In this test, the ephemeris at 19:27 on July 06, 2000 in UTC time was chosen for the GPS signal simulation. At that time, seven satellites were visible from the test position (Latitude 51° North, longitude -114° West), such that the GDOP and HDOP were both less than 2 during the test period. The PRNs visible above a 15° masking angle were 04, 05, 07, 09, 17, 24 and 30. The receiver parameters used in this static test were 3 Hz bandwidth and 20 ms coherent integration time for COOP, and 0.2 Hz bandwidth and 100 ms coherent integration time for individual PLL. To assess the improvement provided by INS aiding, the standard version of the software GPS receiver without INS aiding was also used in all tests to measure normal GPS tracking performance. In this standard GPS receiver, 10 Hz bandwidth and 10 ms coherent integration time were used for signal tracking. In the test, previous experience was used to empirically select these parameters for this comparison. The simulated INS velocity error is shown in Figure 9. It is noted that, at the 50-s point, INS velocity errors reach 0.1 m/s, which is similar to the velocity accuracy available from a MEMS IMU, during 1 s in the integrated GPS/INS system. As shown in Figure 9, the INS errors have low frequency content. In contrast to GPS errors, their time growth is somewhat smooth.

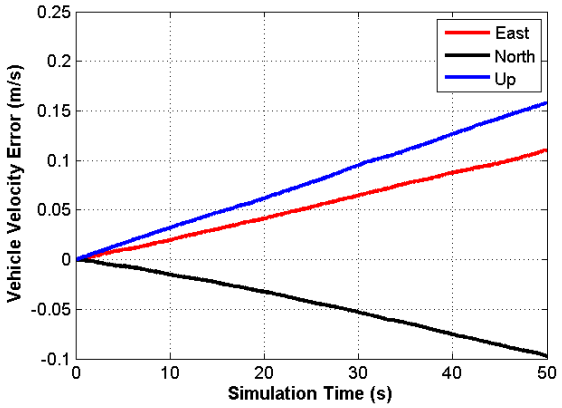

Fig. 9 Simulated INS Velocity Error

C. Static Test Results

Figure 10 shows the carrier-to-noise (C/No) density of satellite PRN 07 over the entire test period both with and without INS aiding. Figure 11 shows the carrier phase tracking error of individual PLLs and Figure 12 shows the total carrier phase tracking errors with and without INS aiding. In Figure 12, for computing the errors, the "true" reference carrier phase was determined by another test where satellite PRN 07 was kept at 45 dB-Hz and all other scenario parameters kept the same as the present test. The figures show that, although the signal power of satellite PRN 07 is attenuated from 45 to 15 dB-Hz during the last 30-s period, the INS-assisted GPS receiver can track the satellite with carrier phase locked. However, a standard GPS receiver without INS aiding cannot track satellite PRN 07 any longer when the signal power is lower than 30 dB-Hz.

As stated previously, the first 14-s period of the test is used to finish frame synchronization and receive code-time-delay (τ) from the navigation data; when the latter is received, the receiver continues to output positioning solutions from the 14 s point onward. Since INS is mainly used to aid carrier phase tracking loops in this paper, only the velocity solution is shown in Figure 13. For the standard receiver without INS aiding, because of the large tracking errors on satellite PRN 07 around epoch 30 s, there are two large faults at epoch 31 s and 32 s, which are about 4 m/s and 11.5 m/s, respectively. The reason for these two faults in the GPS solution is that an epoch-by-epoch least-squares positioning approach is used here to calculate receiver velocity. In this basic least-squares estimator, no fault testing is performed.

After the 32-s epoch, the standard receiver loses lock on satellite PRN 07 so that velocity error returns to a normal level.

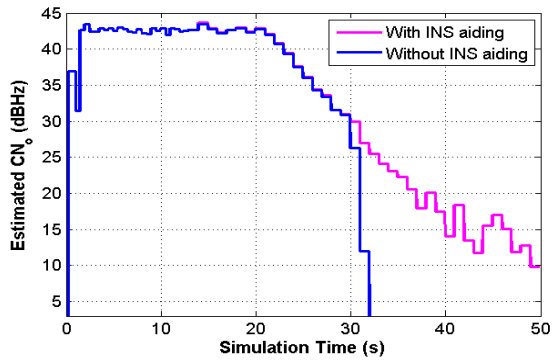

Fig. 10 CN_0 of PRN 07 Tracked in Static Test

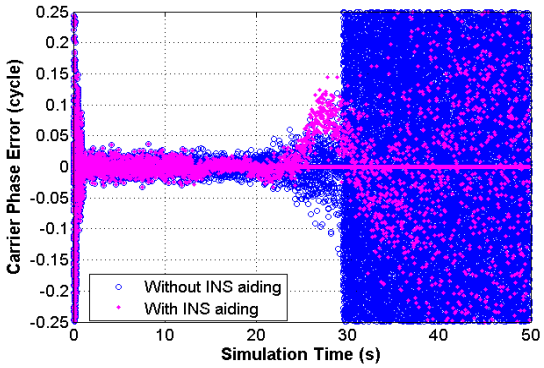

Fig. 11 PLL Carrier Phase Error in Static Test

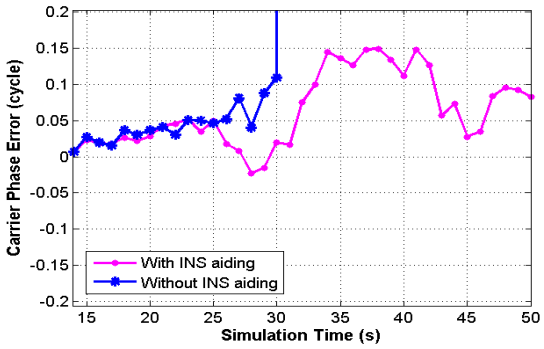

Fig. 12 Total Carrier Phase Error in Static Test

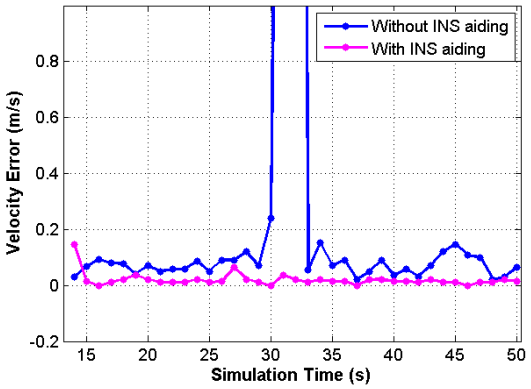

Fig. 13 Horizontal Velocity Error in Static Test

Figure 12 and Figure 13 show that the signal tracking sensitivity of the INS-assisted GPS receiver is improved for at least 15 dB as compared to that of the standard GPS receiver when there are multiple strong signals available. This improvement can be attributed to two aspects:

1. Since INS aiding removes the signal Doppler and thus decreases the receiver dynamics uncertainty, a long coherent integration (here is 100 ms) is used for weak signal tracking.

2. Since the COOP tracking method is used, the strong signals aid the weak signal tracking. In this test, the tracking of other strong satellites is of benefit to the tracking of satellite PRN 07, as will be explained later in this section.

Figure 14 shows the carrier phase tracking error of satellite PRN 07 for the INS-assisted GPS receiver, with and without the use of COOP estimators. In the case when COOP is not used, the individual PLL parameters are kept the same as for the case with COOP estimators, i.e. 0.2 Hz bandwidth and 100 ms coherent integration time. From Figure 14, it can be seen that, although the coherent integration time is 100 ms, without the COOP loop, the tracking performance is even worse than that of a standard GPS receiver (without INS aiding) where coherent integration time is 10 ms. At epoch 26 s, the PLL loses lock and the carrier phase error drifts away rapidly. The reason for the loss of phase lock and poor performance of the INS-aided receiver without COOP in Figure 14 is shown in Figure 15, which gives the carrier Doppler of satellite PRN 07 tracked by COOP and individual PLLs separately. Because of strong signals in view, COOP can perfectly track Doppler residuals caused by the INS aiding errors (shown in Figure 9). Since COOP tracking compensates for the INS aiding errors, combined COOP and INS aiding provides a nearly

perfect reference for receiver dynamics. Therefore, even though the power of PRN 07 drops to 15 dB-Hz during the last 30 s, the carrier Doppler tracked by individual PLLs is close to zero. If COOP were not used, the individual PLL would have to track the INS aiding Doppler error, as shown in Figure 14. When the coherent integration time is very long, the INS aiding Doppler error will fail the pure PLL tracking.

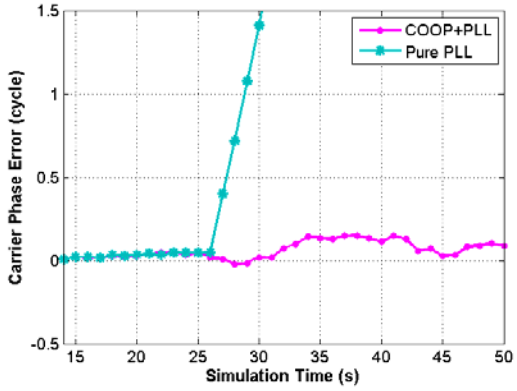

Fig. 14 Total Carrier Phase Error of PLL-Only Receiver with INS Aiding in Static Test

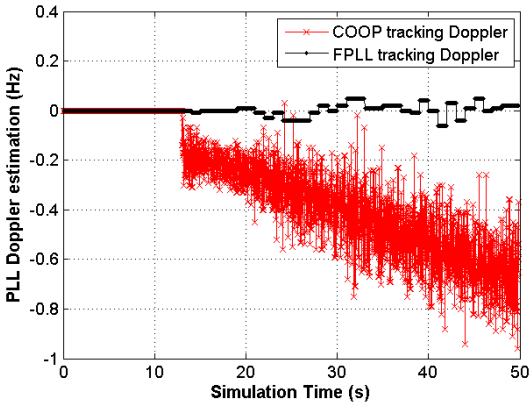

Fig. 15 Carrier Doppler Tracked by PLL and CO-OP Separately in Static Test

To investigate the signal tracking stability of the INS-assisted HSGPS receiver while the receiver parameters vary, different combinations of receiver tracking parameters were examined when all incoming GPS signals were kept at 15 dB-Hz. In this static test, the parameters adopted for the three test receivers are listed in Table 4: Receiver one used very narrow noise bandwidths for both COOP and FPLL, with a very long FPLL integration time, namely, 2 s. In receiver two, wider noise bandwidth and shorter integration time were adopted. However, compared to those used in a standard receiver, these parameters were still very stringent. Receiver three used a set of parameters that can also be used in a standard GPS-only receiver in static situations.

The test results statistics are summarized in Table 5, including the receiver PLL discriminator output and the carrier Doppler tracked by the COOP, respectively. The standard derivations of these two observations are used to assess carrier phase tracking performance of the three receivers. When the receiver parameters become more stringent from receiver three to receiver one, the standard deviation of the carrier Doppler tracked by COOP decreases, which means that COOP can track the incoming signals with increasing accuracy. One can also see that the 15 dB-Hz signal is locked in the entire test and all three receivers can output reasonable velocity solutions. Based on the test results, it is evident that, while the adopted receiver parameters vary in a large range, the INS-assisted HSGPS receiver presents very stable performance in both phase and code tracking, although there are cycle slips present in all three receivers. Finally, it should be noted that due to different receiver dynamics, different levels of signal power, etc, the range of suitable parameters for the INS-assisted HSGPS receiver may change from one case to another. In dynamic situations, a narrower range of suitable receiver parameters is expected.

Table 4 Parameters Adopted in Three Receivers

Adopted Receiver Parameters	Receiver One	Receiver Two	Receiver Three
Phase Noise Bandwidth of COOP (Hz)	0.2	1.2	3
Phase Noise Bandwidth of FPLL (Hz)	0.1	0.2	0.2
Coherent Integration Time of COOP (ms)	20	20	20
Non-Coherent Integration Times of COOP	1	1	1
Coherent Integration Time of FPLL (ms)	100	100	100
Non-Coherent Integration Times of FPLL	20	10	10

Table 5 Tracking Result Statistics of Different Receivers When the Incoming Signal is 15 dB-Hz

Observation Name	Receiver One	Receiver Two	Receiver Three
Estimated Carrier Phase Error Std on Satellite PRN 07 (cycle)	0.029	0.030	0.030
COOP Tracking Doppler Std on Satellite PRN 07 (Hz)	0.4	1.3	1.9
Estimated C/No Mean on Satellite PRN 07 (dB-Hz)	17.9	17.8	17.8
Estimated C/No Std on Satellite PRN 07 (dB-Hz)	3.1	3.2	3.1
Horizontal Velocity Error Mean (m/s)	0.21	0.38	0.40
Horizontal Velocity Error Std (m/s)	0.16	0.23	0.30
Vertical Velocity Error Mean (m/s)	-0.10	0.23	0.14
Vertical Velocity Error Std (m/s)	0.24	0.50	0.79

D. Dynamic Test Scenario

The receiver trajectory and the velocities simulated in the dynamic test are shown in Figure 16 and Figure 17. During the first 20 s, the vehicle moved east with a velocity of 100 m/s. In the next 30 s, the vehicle made an "S" shaped trajectory, with an angular rate of 6 °/s.

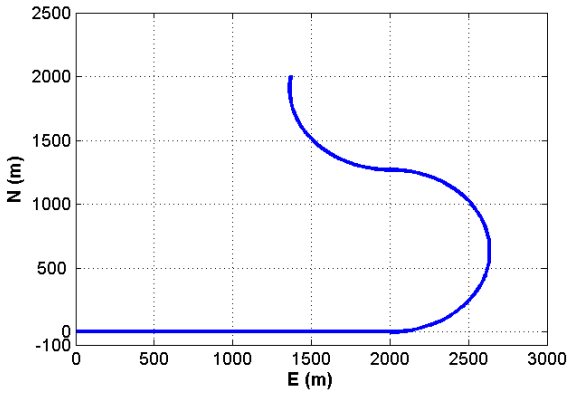

Fig. 16 Vehicle Trajectory in Dynamic Test

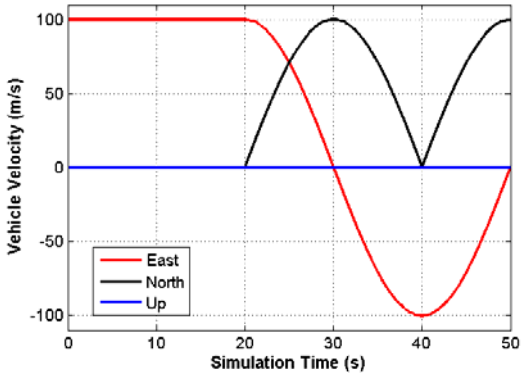

Fig. 17 Vehicle Velocity in Dynamic Test

The change of signal power in the dynamic test is shown in Figure 18. In contrast to the static test, the power of all signals was degraded simultaneously from 45 dB-Hz to 25 dB-Hz during the period of 20 s to 30 s. Then the power level for all signals was kept at 25 dB-Hz from 30 s to 40 s, and then increased back to 45 dB-Hz during the last 10 s of the test. The parameters of the INS simulator were the same as those in the static test. The INS velocity errors are shown in Figure 19. GPS parameters used in the standard GPS software receiver were the same as those in the static test, namely 10 Hz bandwidth and 10 ms coherent integration time. For INS-assisted HSGPS, the bandwidth was 0.4 Hz for individual PLL and 3 Hz for COOP. The coherent integration time of 100 ms for individual PLL and 20 ms for COOP were used. The other parameters were kept same as those of the static test.

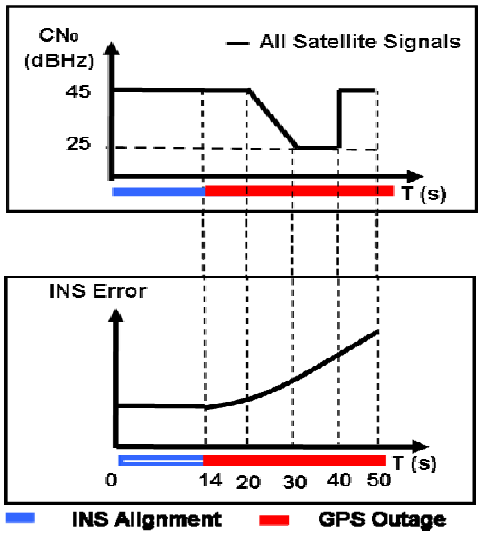

Fig. 18 Simulated Change of Signal Power

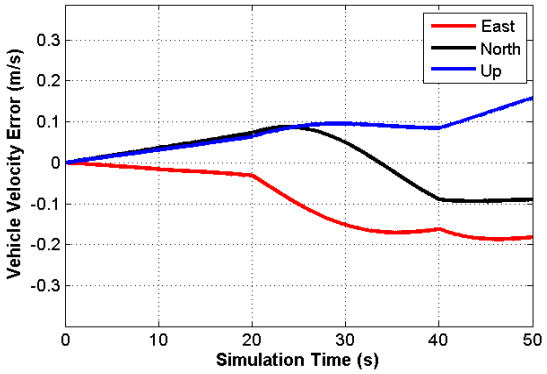

Fig. 19 INS Velocity Error in Dynamic Test

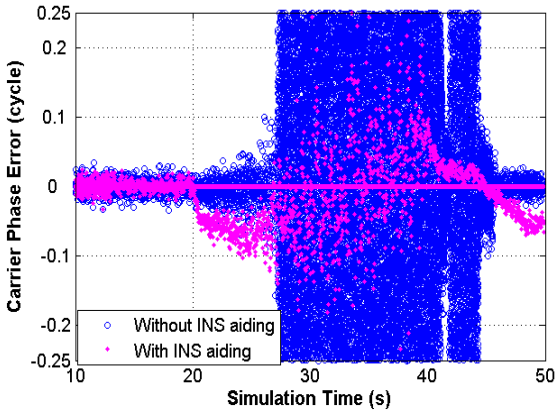

Fig. 21 PLL Carrier Phase Error in Dynamic Test

E. Dynamic Test Results

Figure 20 shows the C/No of the received PRN 07 satellite signal during the test period both with and without INS aiding. Figure 21 shows the PLL carrier phase tracking errors, Figure 22 shows the total carrier phase errors, and Figure 23 shows the horizontal velocity errors for the INS-assisted HSGPS receiver and the standard GPS receiver. It can be seen from these figures that the tracking performances of the standard receiver is very poor under dynamic conditions as compared to those of the INS-assisted HSGPS receiver. Figure 21 and Figure 22 show clearly that the standard GPS receiver cannot lock on the incoming carrier phase when the vehicle starts to make the "S" shaped trajectory. During the period between 20 s and 28 s, although the individual PLLs in the standard receiver show lock on the incoming carrier phase in Figure 21, there are cycle slips due vehicle dynamics. From epoch 28 s onward, the standard receiver stops to output the GPS solution. In contrast, the INS-assisted HSGPS receiver can track the incoming weak signals down to 25 dB-Hz with carrier phase locked during the entire test.

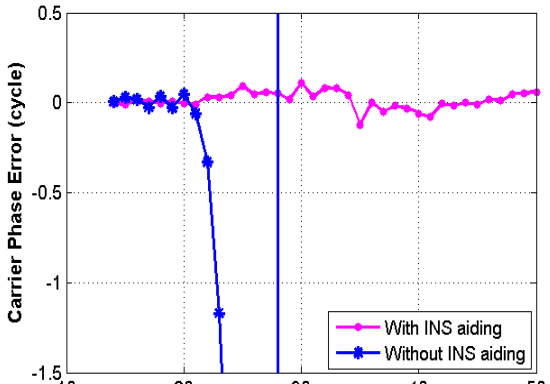

Fig. 22 Total Carrier Phase Error in Dynamic Test

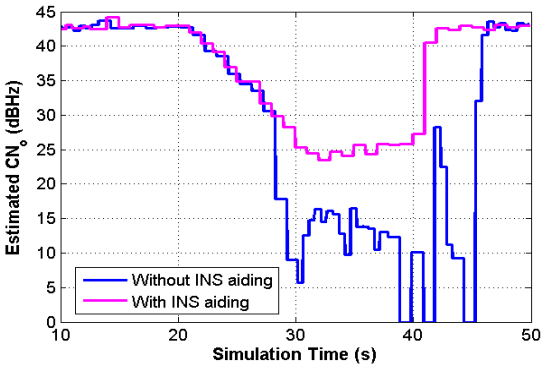

Fig. 20 CN_0 of PRN 07 in Dynamic Test

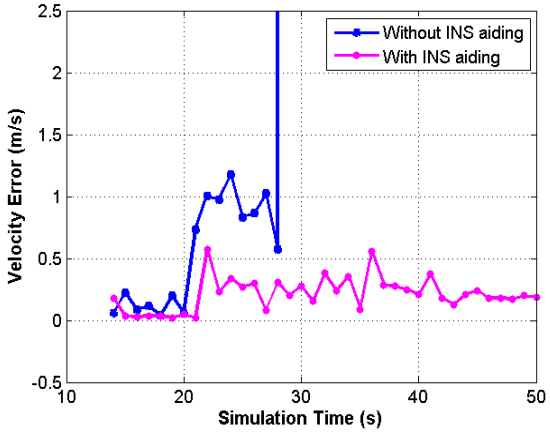

Fig. 23 Horizontal Velocity Error in Dynamic Test

CONCLUSIONS

This paper proposes a novel design for an INS-assisted GPS receiver to improve GPS tracking performance for navigation in degraded signal environments. The effects of IMU quality and receiver parameters such as coherent integration time on the designed system are analyzed. Compared to a standard GPS receiver without INS aiding, the INS-assisted GPS receiver proposed here yields much better performance under attenuated signal environments, based on the tests carried out. An analysis of the results leads to the following conclusions:

1. INS aiding can effectively reduce the receiver dynamics uncertainty and improve tracking performance of a standard GPS receiver significantly under both weak signal and high dynamic signal environments.

2. When an INS solution is available, an effective signal tracking strategy can be summarized in three steps. First, the INS solution is implemented in order to remove most of receiver dynamics uncertainty; as a consequence, the residual Doppler signal left for the COOP and FPLLs to track is close to zero. Next, a vector tracking-based COOP loop is designed to track the residual carrier Doppler effectively; since six to 10 satellites are usually in view, COOP
3. tracking yields much better performance than conventional FPLLs, especially under weak signal environments. Finally, FLL-assisted PLLs can be used to track the carrier. Since the Doppler signal is compensated by the combined INS and COOP aiding, the residual Doppler error is close to zero. This significantly decreases carrier phase tracking errors and, therefore, increases the FLL/PLL tracking sensitivity.

4. Although INS error increases rapidly with time during a GPS outage, the INS solution errors change smoothly. These errors can be easily tracked by the COOP method and thus, will not affect signal tracking significantly. Therefore, even if the INS solution error is as large as 0.1 m/s, an INS-assisted GPS receiver can track a GPS signal that is 30 dB lower than LOS signals with relatively good positioning accuracy.

5. The combined tracking of the FPLL and COOP loops presented herein have been shown to track signals as low as 15 dB-Hz. When the signal power is above 22-23 dB-Hz, this method can lock on the incoming carrier and provide accurate carrier phase measurements. When the signal is lower than 22 dB-Hz but higher than 15 dB-Hz, the method can track the incoming carrier most of the time, although cycle slips may occur. When there are several strong signals in view, the receiver can lock the other weak carrier signals as low as 15 dB-Hz due to the assistance from the strong signals.

6. Because INS aiding provides most of the Doppler measurements, high receiver dynamics do not affect signal tracking significantly in INS-assisted GPS receivers. With INS aiding and by adopting COOP tracking, long coherent integration can be implemented safely when necessary.

REFERENCES

Alban S., D. M. Akos and S. M. Rock (2003), ***Performance Analysis and Architectures for INS-Aided GPS Tracking Loops***, in Proceedings of ION National Technical Meeting (NTM), 22-24 January, Anaheim CA, pp. 611-622, U.S. Institute of Navigation, Fairfax VA.

Beser J., S. Alexander, R. Crane, S. Rounds and J. Wyman (2002), ***Trunavtm: A Low-Cost Guidance/Navigation Unit Integrating A SAASM-Based GPS And MEMS IMU In A Deeply coupled Mechanization***, in proceedings of ION GPS/GNSS, 24-27 September, Portland OR, pp. 545-555, U.S. Institute of Navigation, Fairfax VA.

Brown, R. G. and P. Y. C. Hwang (1992), ***Introduction To Random Signals and Applied Kalman Filtering***, John Wiley & Sons, Toronto ON, second edition.

Don K., S. Luis and R. B. Langley (2005), ***Compensation of the Effects of Phase Wind-up for Improving the Performance of a GPS RTK-Based Vehicle Navigation System***, in Proceedings of ION GNSS 18th International Technical Meeting of the Satellite Division, 13-16 September, Long Beach, CA, pp. 346-354, Institute of Navigation, Fairfax VA.

Dong L., C. Ma and G. Lachapelle (2004), ***Implementation and Verification of a Software-Based IF GPS Signal Simulator***, in proceedings of ION National Technical Meeting (NTM), 26-28 January, San Diego CA, pp. 378-389, Institute of Navigation, Fairfax VA.

Gao G. (2007), ***INS-Assisted High Sensitivity GPS Receivers for Degraded Signal Navigation***, PhD Thesis, Published as Reports Number 20252, Department of Geomatics Engineering, University of Calgary, Canada (Available at http:\plan.geomatics.ucalgary.ca).

Gao G. and G. Lacahpelle (2006), ***INS-Assisted High Sensitivity GPS Receivers for Degraded Signal Navigation,*** in proceedings of ION GNSS 2006, Fort Worth TX, 26-29 September 2006.

Gautier J. D. and B. W. Parkinson (2003), ***Using the GPS/INS Generalized evaluation Tool (GIGET) for the Comparison of Loosely Coupled, Tightly Coupled and Ultra-Tightly Coupled Integrated Navigation Systems***, in proceedings of ION 59th Annual Meeting/CIGTF 22nd Guidance Test Symposium, 23-25 June, Albuquerque, NM, pp. 65-76, Institute of Navigation, Fairfax VA.

Godha S. and M. E. Cannon (2005), ***Integration of DGPS with a Low Cost MEMS – Based Inertial Measurement Unit (IMU) for Land Vehicle Navigation Application***, in proceedings of ION GPS/GNSS, September13-16, Long Beach CA, pp. 333-345, Institute of Navigation, Fairfax VA.

Gustafson D., J. Dowdle and K. Flueckiger (2000), ***A High Anti-Jam GPS based Navigator***, in proceedings of ION National Technical Meeting (NTM), 26-28 January, Anaheim CA, pp. 495-503, Institute of Navigation, Fairfax VA.

Ilir F. P., W. K. Clifford, G. Gao, R. M. William, J. Wang and J. Lavrakas (2007), ***Discrete vs. Continuous Carrier Tracking Loop Theory, Implementation, and Testing with Large BnT***, in proceedings of ION GNSS 2007, 25-28 September, Fort Worth TX, Institute of Navigation, Fairfax VA.

Kaplan E. D., ed (1996), Understanding GPS: Principles and Application, Artech House, Boston MA.

Kreye C., B. Eissfeller and J. Ó. Winkel (2000), ***Improvements of GNSS Receiver Performance Using Ultra-tightly coupled INS Measurements***, in proceedings of ION GPS/GNSS, 19-22 September, Salt Lake City UT, pp. 844-854, Institute of Navigation, Fairfax VA.

Lachapelle, G. (2007), ***Pedestrian Navigation With High Sensitivity GPS Receivers and MEMS***, Journal of Personal and Ubiquitous Computing, Springer, 11, 6, 481-488.

Lachapelle G., H. Kuusniemi and D. T. H. Dao (2003), ***HSGPS Signal Analysis and Performance under Various Indoor Conditions***, in proceedings of ION GPS/GNSS, 9-12 September, Portland OR, pp. 1171-1184, Institute of Navigation, Fairfax VA.

Ma C., R. Klukas and G. Lachapelle (2007), ***A Non-Line-of-Sight Error Mitigation Method for TOA Measurements***. IEEE Transactions on Vehicular Technology, 56, 2 (March), 641-651.

Ma C., G. Lachapelle and M. E. Cannon (2004), ***Implementation of a Software GPS Receiver***, in proceedings of ION GPS/GNSS, September 21-24, Long Beach CA, pp. 956-970, Institute of Navigation, Fairfax VA.

MacGougan G. D. (2003), ***High Sensitivity GPS Performance Analysis in Degraded Signal Environments***, Master thesis, UCGE Reports Number 20176, Department of Geomatics Engineering, University of Calgary, Canada (Available at http://plan.geomatics.ucalgary.ca).

Petovello M. G. (2003), ***Real-Time Integration of a Tactical-Grade IMU and GPS for High-Accuracy Positioning and Navigation***, PhD thesis, UCGE Reports Number 20173, Department of Geomatics Engineering, University of Calgary, Canada (Available at http://plan.geomatics.ucalgary.ca).

Raquet J. F. (2004), ***GPS Receiver Design***, ENGO 699.10 Course Notes, Department of Geomatics Engineering, University of Calgary, Canada.

Scherzinger B. (2004), ***Estimation with Applications to Navigation***, ENGO 699.11 Course Notes, Department of Geomatics Engineering, University of Calgary, Canada.

Sennott J. (1997), ***Robustness of Tightly Coupled Integrations for Real-Time Centimeter GPS Positioning***, in proceedings of ION GPS, 16-19 September, Kansas City KS, pp. 655-663, Institute of Navigation, Fairfax VA.

Soloviev A., F. van Graas and S. Gunawardena (2004a), ***Implementation of Deeply Integrated GPS/Low-Cost IMU for Acquisition and Tracking of Low CNR GPS Signals***, in proceedings of ION National Technical Meeting (NTM), 26-28 January, San Diego CA, pp. 923-935, Institute of Navigation, Fairfax VA.

Soloviev A., S. Gunawardena and F. van Graas (2004b), ***Deeply Integrated GPS/Low-Cost IMU for Low CNR Signal Processing: Flight Test Results and Real Time Implementation***, in proceedings of ION GNSS 17th International Technical Meeting of the Satellite Division, 21-24 September, Long Beach CA, pp. 1598-1608, Institute of Navigation, Fairfax VA.

Stephens S. A. and J. B. Thomas(1995), ***Controlled-Root Formulation for Digital Phase-Locked Loops***, IEEE Transactions on Aerospace and Electronic Systems, Vol. 31, No. 1, pp. 78-95.

Tetewsky, A. K. and F. E. Mullen (1996), ***Carrier phase wrap-up induced by rotating GPS antennas***, in Proceedings of the Institute of Navigation ION AM-96, Cambridge, MA, pp. 21-28, , Institute of Navigation, Fairfax VA.

Van Diggelen F. and C. Abraham (2001), ***Indoor GPS Technology***, CTIA Wireless-Agenda, Dallas, May.

Watson, R., G. Lachapelle, R. Klukas, S. Turunen, S. Pietilä and I. Halivaara (2006), ***Investigating GPS Signals Indoors with Extreme High-Sensitivity Detection Techniques***, Navigation, Institute of Navigation, 52, 4, 199-213.

Zhodzishsky M., S. Yudanov, V. Veitsel and J. Ashjaee (1998), ***COOP Tracking for Carrier Phase***, in proceedings of ION GPS, 15-18 september, Nashville, Tennessee, pp. 653-664, Institute of Navigation, Fairfax VA.

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 62-71

Journal of Global Positioning Systems
2008
© CPGPS

An Evaluation of Various Ionospheric Error Mitigation Methods used in Single Frequency PPP

S. Choy, K. Zhang, and D. Silcock
Royal Melbourne Institute of Technology (RMIT) University, Melbourne, Australia.

Abstract

Precise Point Positioning (PPP) using dual frequency GPS receivers is capable of providing centimetre level point positioning accuracy anywhere around the world, without the need for a base station. However, when using single frequency GPS receivers, the accuracy of the positioning decreases, particularly in the height component. One main factor for this degradation in accuracy is the unmodeled ionospheric error.

This paper investigates the performance of three different ionospheric error mitigation methods used in single frequency PPP in the Australian Region. They are the GRAPHIC (GRoup And PHase Ionospheric Correction) algorithm, the Global Ionospheric Maps (GIMs) and the Klobuchar model. Numerical results show that the GRAPHIC and GIMs methods are able to provide point positioning accuracy better than 1m for session duration less than an hour using geodetic quality single frequency receivers. For 12 to 24 hours data sets, the positioning accuracy can be as good as <0.1m.

Keywords: Single Frequency, PPP, Ionospheric Error

1 Introduction

Global Positioning System (GPS) is currently one of the most popular satellite positioning systems due to the global availability of GPS signal and performance. Centimetre level positioning accuracy is now achievable using carrier phase-based Differential GPS (DGPS) technique, in which two or more geodetic quality GPS receivers are deployed and two or more frequencies are used to alleviate ionospheric effects (Wang et al., 2004, Wu et al., 2006). This technique is able to provide high accuracy solution because common errors, such as satellite and receiver clock errors are cancelled out in short baselines or the errors are dramatically reduced in long baselines (Witchayangkoon, 2000, Zhang et al., 2007).

The coordinates of a point can also be determined by absolute point positioning technique using a single GPS receiver (Hofmann-Wellenhof et al., 2001). However, this technique using code observations is only capable of providing positioning accuracy in the level of a few metres. Such performance is due to the nature of code observations, as well as, the lack of knowledge about the satellite clock error and the noise affecting code observations (Kwon et al., 2001). Therefore, the absolute positioning technique is often not suitable for applications that require highly precise and accurate solutions.

Since 1994, the International GNSS Service (IGS) has been providing precise GPS orbit and satellite clock corrections to the GPS community. This has pushed absolute point positioning technique to a new era, whereby users can accurately obtain their positions without the need to process their data with any base station. This new technique is known as Precise Point Positioning (PPP). PPP uses undifferenced code and carrier phase observation from a single receiver, in addition to the precise satellite orbit and clock correction products for high accuracy point positioning. Recent research has demonstrated that PPP is capable of providing centimetre level point positioning for static applications and decimetre level for kinematic applications using a dual frequency, geodetic quality receiver (Hèroux et al., 2004; Abdel-salam, 2005). As for single frequency observations, the accuracy of the estimated point positioning decreases (Yuan *et al.*, 2007), particularly in the height component. One main factor for this degradation in accuracy is the effect of unmodeled ionospheric error.

The objective of this paper is to investigate the accuracy of single frequency PPP as a function of various observation durations, using the Australian Regional GPS Network (ARGN) stations. Single frequency GPS receivers are the most widely used tools for tracking, navigation and geo-referencing. It is estimated that 75 percent of all GPS receivers used globally are single frequency receivers (Arbesser-Rastburg, 2006, Wyllie et al., 2006). Thus, any accuracy improvement on the point

positioning algorithm will clearly be of great practical importance. Different ionospheric error mitigation methods will also be assessed and their accuracy compared. Three methods were used in this investigation: the single frequency ionospheric-free code and phase delay known as GRAPHIC (GRoup And PHase Ionospheric Correction) algorithm (Yunck, 1993), the IGS rapid Global Ionospheric Maps (GIMs) and the Klobuchar model together with the broadcast ionospheric coefficients.

This paper first presents a brief overview of the basic PPP observation equations. Secondly, a description of the GRAPHIC, GIMs and Klobuchar models will be given. Absolute point positioning at eight ARGN sites with accurately known International Terrestrial Reference Frame 2000 (ITRF00) coordinates is carried out using single frequency GPS observations with different ionospheric error mitigation methods. The performance of the algorithms is evaluated based on the accuracy and precision of the derived solutions, as well as the time required for the solutions to converge. Numerical results clearly demonstrates that the GRAPHIC and GIMs methods are able to provide point positioning accuracy better than 1m for session duration less than an hour using geodetic quality single frequency receivers. For 12 to 24 hours data sets, the positioning accuracy can be as good as a few centimetres (approximately <0.1m under favourable conditions).

2 GPS Observation Equations

The basic GPS observation equations can be written and expressed as follows (Chen and Gao, 2005):

$$\Pr = p_r^s + c.(dt_r - dT^s) + dorb + dion + dtrop + drel + \varepsilon(\Pr) \quad (1)$$

$$\Phi_L = p_r^s + c.(dt_r - dT^s) + dorb - dion + dtrop + \lambda_L.N_L + drel + dpw + \varepsilon(\Phi_L) \quad (2)$$

where,

Pr	:	measured pseudorange on L1 (m)
Φ_L	:	measured carrier phase range, L1 or L2 (m)
p_r^s	:	true geometric range between the GPS receiver and the satellite (m)
c	:	speed of light (m/s)
dt_r	:	receiver clock error (s)
dT^s	:	satellite clock error (s)
$dorb$	:	satellite orbit error (m)
$dion$	:	ionospheric error (m)
$dtrop$	:	tropospheric error (m)
$drel$	:	relativistic effects (m)
dpw	:	phase windup error on the carrier phase measurements (m)
λ_L	:	wavelength of the carrier phase, L1 or L2 (m/cycle)
N_L	:	phase ambiguity including the initial phase bias on the carrier phase, L1 or L2 (cycle)
$\varepsilon(.)$	:	noise including multipath (m)

The satellite orbit error ($dorb$) and clock error (dT^s) can be eliminated by applying the precise orbit and clock correction products from the IGS. The tropospheric error ($dtrop$) can be corrected at decimetre and even centimetre level, using existing models and meteorological measurements. The relativistic effects ($drel$) and the phase windup (dpw) can be corrected to centimetre level accuracy using existing correction models (Chen and Gao, 2005).

Consequently, equations (1) and (2) can be expressed as,

$$\Pr = p_r^s + c.dt_r + dion + M.ZPD + \varepsilon(\Pr) \quad (3)$$

$$\Phi_L = p_r^s + c.dt_r - dion + M.ZPD + \lambda_L.N_L + \varepsilon(\Phi_L) \quad (4)$$

It should be noted that the original tropospheric error ($dtrop$) is now expressed in equations (3) and (4) as a function of the tropospheric Zenith Path Delay (ZPD) with Mapping function (M) relating the tropospheric error to the elevation angle of the satellite.

Equations (3) and (4) are known as the basic PPP observation equations. From Equation (3) and (4), the ionospheric error becomes a major source of error in single frequency PPP. Note that there are two observation equations for single frequency data, one code measurement and one carrier phase measurement for each satellite observed per epoch.

3 Ionospheric Error Mitigation Methods

There are a number of different mitigation methods for single frequency GPS users to correct for the ionospheric error. The simplest and most widely used method to correct for the ionospheric error is to utilise the Klobuchar model together with the eight ionospheric

coefficients broadcast as part of the navigation message. During normal operation, the parameters of the model are updated at least once every six days (ARINC Research Corporation, 2000; Øvstedal, 2002). This algorithm can be used in real-time and it was designed to provide a correction for approximately 50 percent Root Mean Square (RMS) of the ionospheric range delay (Klobuchar, 1987). Since mid July 2000, the Centre for Orbit Determination in Europe (CODE) has been providing post-fit Klobuchar ionospheric coefficients that best fit the GIMs data estimated by CODE. Øvestedal (2002) has shown that the post-fit coefficient is able to provide more consistent results than the broadcast Klobuchar model. Currently, the post-fit Klobuchar ionospheric coefficients have a latency of several days. Thus, for the purpose of this investigation, the Klobuchar model with the broadcast ionospheric coefficients was used instead. It is worth noting that the CODE has also been estimating predicted Klobuchar-style coefficients. However, the improvement was found to be not as significant as the post-fit coefficients (Chen and Gao, 2005; CODE, 2007).

Alternatively, single frequency GPS users may fully exploit the state of the art ionospheric model provided by IGS and other organisations, such as the Jet Propulsion Laboratory (JPL) and the CODE. Currently, four IGS Ionosphere Associate Analysis Centres (IAACs) provide two-dimensional GIMs in IONosphere map EXchange (IONEX) format (Schaer et al., 1998) that refer to a 450 km shell height. The four IGS IAACs are CODE, European Space Operations Centre of ESA (ESOC), JPL, and Technical University of Catalonia (UPC) (Hernández-Pajares, private communication). Each IAAC sets up a daily IONEX file that has 13 GIMs, which contains Total Electron Content (TEC) values and a set of Differential Code Biases (DCBs) values for that day. These products are contributed to the IGS Ionosphere Working Group in order to generate the combined IGS final and rapid GIMs. In April 2003, the IGS final GIMs in IONEX format became an official IGS product, which has a latency of 11 days. Meanwhile, a rapid version of the GIMs with a delay of less than 24 hours is made available to the public since December 2003. Both final and rapid GIMs have a temporal resolution of 2 hours and a spatial resolution of 5˚ in longitude (λ) and 2.5˚ in latitude (ϕ). According to the products accuracy specifications, the rapid and final GIMs have an accuracy level of 2 TEC Unit (TECU) to about 8-9 TECU (IGS, 2007), in which 1 TECU corresponds to 0.163m range error on L1 frequency (Øvstedal, 2002; Chen and Gao, 2005). The combined IGS rapid GIMs were used in this investigation due to its shorter latency. The characteristics of the GIMs are shown in Table 1.

Table 1 IGS Ionospheric Correction Products (IGS, 2007).

Product	Accuracy	Latency	Updates	Sample Interval
Rapid Ionospheric TEC Grid	2-9 TECU	<24 Hours	Daily	2 Hours; 5˚ (λ) x 2.5˚ (ϕ)
Final Ionospehric TEC Grid	2-8 TECU	~11 Days	Weekly	2 Hours; 5˚ (λ) x 2.5˚ (ϕ)

Perhaps the least appreciated technique for single frequency GPS users to correct for the ionospheric error is by using the GRAPHIC method. The ionosphere delays the code measurements and advances the carrier phase measurements, thus making it possible to eliminate this error by taking the simple average of the code and carrier phase delay observables (Yunck, 1993; Montenbruck, 2003; Simsky, 2006).

Upon adding and averaging the code and carrier phase range, the combined GRAPHIC measurement (ignoring the higher-order ionospheric error terms) can be written in a simplified manner as follows,

$$\begin{aligned} P^{CA\&L1} &= \frac{\mathrm{Pr} + \Phi_{L1}}{2} \\ &= p_r^s + c \cdot dt_r + M \cdot ZPD + \frac{\lambda_{L1} N_{L1}}{2} \\ &\quad + \frac{\varepsilon(\mathrm{Pr})}{2} + \frac{\varepsilon(\Phi_{L1})}{2} \end{aligned} \tag{5}$$

In Equation (5), one should note that the combined measurement no longer depends on the ionospheric delay and exhibits a noise, which is half the code and carrier phase noise values. However, since the bandwidth limitations keep the carrier phase observables typically 100 times more precise than the code, the noise affecting the GRAPHIC method is mainly dominated by the code measurement error (Yunck, 1993). The term GRAPHIC was first introduced by Yunck (1993) for single frequency ionospheric-free code and phase delays. Since then, various authors have addressed the potential of using GRAPHIC method in post-processing and real-time single frequency PPP (Montenbruck, 2003; Muellerschoen et al., 2004; Simsky, 2006).

4 Differential Code Biases (DCBs)

It is important for single frequency PPP users to apply the L1-L2 DCBs along with the above-mentioned ionospheric error mitigation methods. For both the broadcast and precise satellite clock corrections, the offset of the satellite clocks is always referred to the ionosphere-free linear combination of the L1 and L2 frequency. For dual frequency PPP, no such DCBs calibrations are required to be applied. However, single frequency PPP users must apply the satellite DCBs as the

satellite clock corrections are consistent with the satellite L1-L2 DCBs convention (Kouba, 2003; Le, 2004). The satellite DCBs can be found from the GIMs in IONEX format as they are constantly computed by IAACs as part of their global ionospheric delay corrections, as well as from the broadcast navigation message. In this research, the DCBs from the GIMs were used.

Due to the effects of Anti-Spoofing, some civilian GPS receivers do not output P1 code but C/A code instead, which has a different hardware delay than the P1 code. Therefore, data from these receivers must be corrected using the P1-C1 DCBs in order to achieve full consistency with the P1-P2 data and the satellite clock corrections. The RINEX conversion utility "cc2noncc.f" can be used to transform given C/A code measurements to be consistent with the P1-P2 data and satellite clock corrections. The "cc2noncc.f" conversion utility is available at the IGS Clock Products Working Group website (U.S. Naval Research Lab, 2007). Alternatively, the P1-C1 DCBs can be obtained as a separate file from the CODE website (CODE, 2007).

5 Numerical Results and Analysis

The CSRS-PPP software package was used to facilitate single frequency GPS data processing (Hèroux et al., 2004). This software package is capable of processing both dual and single frequency data, using either the broadcast or precise satellite ephemerides. After considering the accuracy and latency of these products, the IGS rapid satellite orbit and clock corrections were used in this investigation for both the GRAPHIC and GIMs methods. As for the Klobuchar model, the broadcast ephemerides were used instead. Currently, the IGS rapid orbit and clock corrections are accurate better than 5cm and 0.1ns respectively, and have a latency of 17 hours (IGS, 2007). Details of the various IGS products can be found at the IGS website (IGS, 2008).

Precise satellite orbits provided by the IGS ephemerides do not refer to the antenna phase centre, but instead, they refer to the centre of mass. As all GPS range observations are measured from the satellite transmitting antenna to the electrical phase centre of the receiving antenna, single frequency PPP users are required to apply the satellite phase centre corrections to account for the offsets. The relative GPS antenna phase centre offsets and variations contained in the *igs_01.pcv* file (IGS, 2007) were used in this research, as the GPS data were collected prior to 5 November 2006. However, one should note that the IGS has adopted the absolute antenna phase centre offsets and variations for its routine generation of the precise satellite orbits on the 5 November 2006 (Gendt, 2006). Thus, users should use the absolute antenna phase centre corrections for all GPS data processing using observations data collected after the 5 November 2006.

A static test was performed on 7th July 2006, using eight selected ARGN GPS stations. They are Hobart, (HOB2), Mount Stromlo (STR1), Alice Springs (ALIC), Yaragadee (YAR2), Ceduna (CEDU), Townsville (TOW2), Darwin (DARW) and Cocos Island (COCO) stations. The location of these stations is shown in Fig. 1. 24 hours data from these eight stations were downloaded from the Scripps Orbit and Permanent Array (SOPAC) website (SOPAC, 2008). These data were windowed into 15 minutes, 30 minutes, 1 hour, 2 hours, 4 hours and 12 hours, starting from 00:00 GPS time. These stations are equipped with dual frequency, geodetic quality GPS receivers, but only observations on L1 were used in the data processing. The time interval of the collected data is 30 seconds. A 5˚ cut-off elevation angle was applied to all datasets.

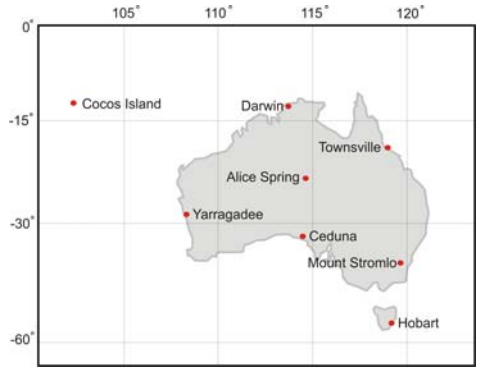

Fig. 1 The location of the eight ARGN stations.

One of the general rules of single frequency PPP algorithm is to set a realistic a priori code and carrier phase sigma values, which will adequately reflect the actual measurements noise including multipath. Since the noise on L1 code is unknown, it is a challenge to produce realistic error estimates which reflect the measurement noise. The parameter statistics are often biased, especially for a short observation period because multipath does not average out (Hèroux, private communication). Therefore, the measurement sigmas used in this investigation were based on the standard values widely used in GPS processing. In fact, the values themselves are not of great importance; only the fact the sigma values for the code measurement should be greater (approximately 4 times) than the carrier phase measurements (Simsky, 2006). In this research, a priori code and carrier phase sigma values were set to 4m and 0.03m, respectively.

These data were post-processed using the above three different ionospheric error mitigation methods and the estimated coordinates were compared to the accurately known ITRF00 coordinates for the eight ARGN stations. L1 code and carrier phase measurements were used in the GRAPHIC and IGS rapid GIMs algorithms; while only L1 code measurements were used in the Klobuchar model method. This was done to provide an indication of the

achievable point positioning accuracy using the classical single frequency code-based processing with the broadcast ionospheric coefficients. The coordinate differences between the estimated and the known values are plotted as a function of various observation periods.

The results for STR1 station point estimations are presented in Fig. 2 and 3, respectively. Fig. 2 shows error values of the east, north and height components based on the three ionospheric error mitigation processing methods, over the first 12 hours (00:00-12:00 GPS time; Greenwich Meridian Time (GMT) +10 hours). The east, north and height component errors were computed by subtracting the accurately known reference coordinates with the estimated position values. Fig. 2 presents an example of coordinates convergence of a 12 hours solution for each methods. Fig. 3 provides a more detailed outlook of the position error at specific observation period.

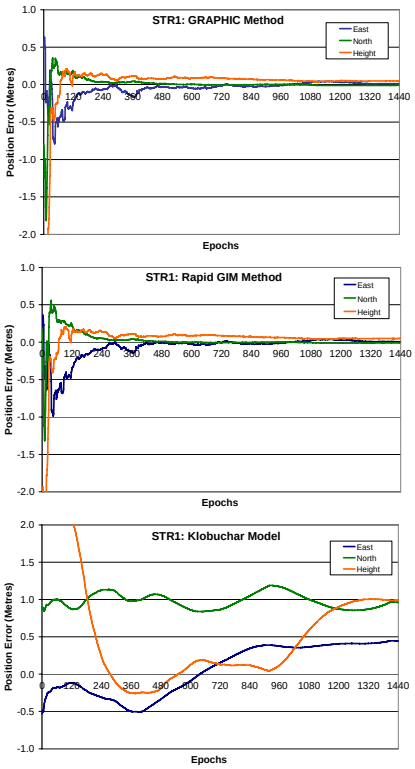

Fig. 2 Position errors with respect to the accurately known coordinates for STR1 station (sample rate: 1 epoch = 30 seconds).

It can be clearly seen from these figures that the GRAPHIC and GIMs methods follow a similar long-term pattern, whereby both methods provide comparable positioning accuracy. Both methods are able to provide horizontal accuracy better than 0.5m after an hour observation period (120 epochs). Only after 4 hours observations, the position estimates from these two methods converge to approximately 8cm of the known values. This is because the ambiguity term in L1 measurement is not known, and the multipath effect on the pseudorange measurement typically sets the accuracy limit for single frequency PPP, especially for short observation sessions. As predicted, the broadcast Klobuchar model has the least accurate positioning results.

In order to have representative data from the low or near equatorial region where ionospheric activities are considerably higher and complex than the mid latitude region, COCO station situated at latitude 12˚11'S has been selected. Fig. 4 and 5 present the position errors using the three different ionospheric error mitigation methods. Similarly, the GRAPHIC and GIMs provide similar positioning accuracy over 24 hours observation period. While, the Klobuchar model only provides comparable position estimates accuracy between 00:00 to

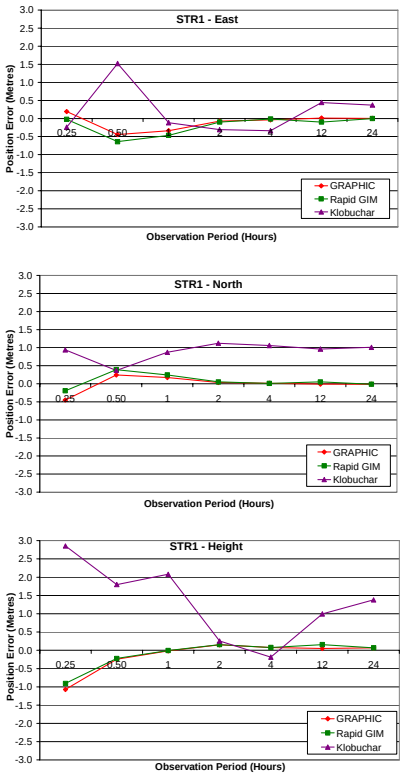

Fig. 3 A detailed outlook of the position errors for STR1 station at 0.25, 0.50, 1, 2, 4, 12 and 24 hours observation periods.

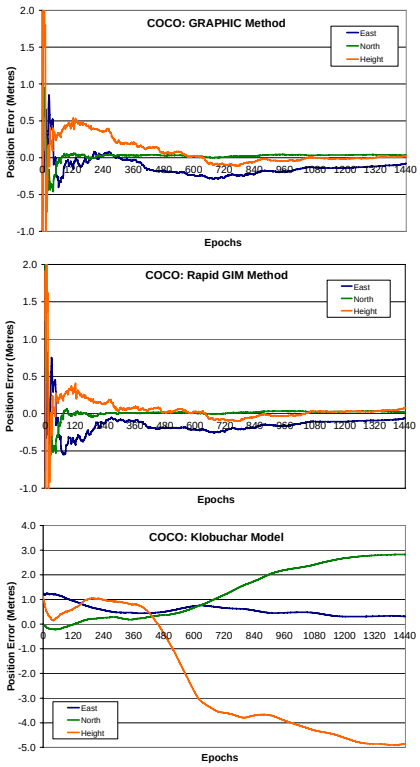

Fig. 4 Position errors with respect to the accurately known coordinates for COCO station (sample rate: 1 epoch = 30 seconds). Note the different y-axis scale on the Klobuchar model plot.

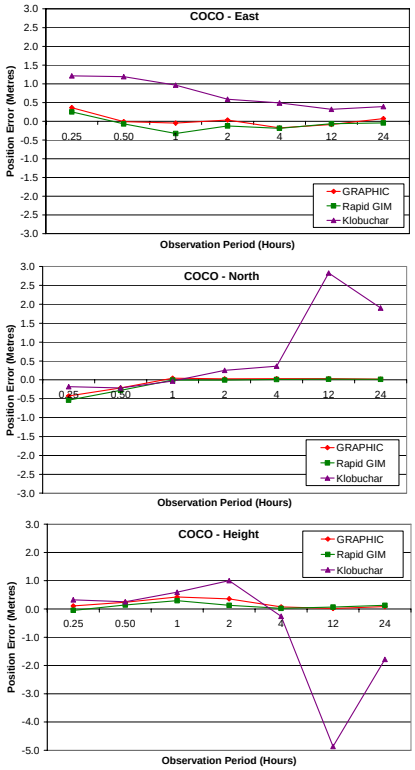

Fig. 5 A detailed outlook of the position errors for COCO station at 0.25, 0.50, 1, 2, 4, 12 and 24 hours observation periods. Note the different y-axis scale on the height component plot.

03:00 GPS time (GMT +6.5 hours), where the ionospheric activity is considered at its minimum. The Klobuchar model is unable to provide consistent and accurate solutions especially in the height estimations, over 24 hours observation period. This is mainly due to the increase of TEC in the atmosphere and the broadcast ionospheric coefficients failed to account for the sudden variation. Both the GRAPHIC and GIMs methods are not significantly affected by this phenomenon.

Another point worth noting is that the height component for this station using GRAPHIC and GIMs methods takes longer time to converge than STR1 station. This is probably caused by the software failure to account for the residual ionospheric range delay effectively, thus disabling the fixing of ambiguity over short observation period.

Fig. 6 and 7 illustrate the position estimate errors for DARW station. DARW station is situated approximately 70km away from Darwin, in the Northern Territory of Australia. Similarly, DARW station is located in the low latitude region 12˚50'S.

It can be inferred from Fig. 4, 5, 6, and 7 that the DARW position estimates based on the three algorithms are not as accurate and consistent as those of COCO station, even though both stations are located in the same latitudinal region. Two reasons that may cause this: first, the geometric factors of the satellites, and second, the multipath effects and/or the undetected cycle clips of the carrier phase measurements at DARW station. To verify this, the datasets of these stations were checked for its data quality using TEQC (Translating/Editing/Quality Control) software. It was found that the multipath effects affecting L1 frequency is more significant at DARW than COCO station. The number of cycle slips and outliers are also higher at DARW station, resulting in less accurate position solutions. Fig. 8 and 9 are the images of the COCO and DARW stations.

Tables 2 and 3 summarise the combined mean and RMS values based on the eight ARGN stations over various observation periods, i.e. 0.25 (15m), 0.50 (30m), 1, 2, 4, 12 and 24 hours. The GRAPHIC method was abbreviated in the table as "GHIC", and the Klobuchar model was abbreviated as "Klob model".

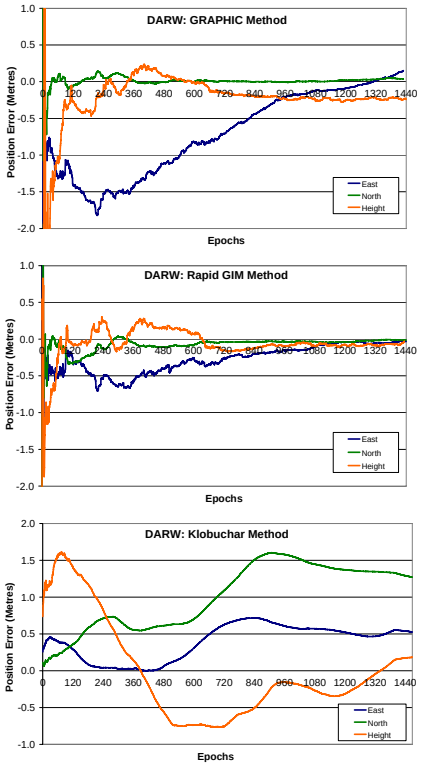

Fig. 6 Position errors with respect to the accurately known coordinates for DARW station (sample rate: 1 epoch = 30 seconds).

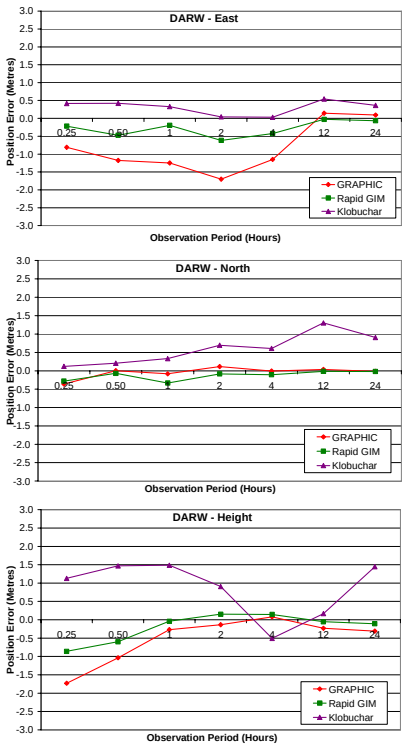

Fig. 7 A detailed outlook of the position errors for DARW station at 0.25, 0.50, 1, 2, 4, 12 and 24 hours observation periods.

Fig. 8 COCO station is equipped with an ASHTECH UZ-12 GPS receiver with an AOAD/M_T antenna type. This station is located within the perimeter of Cocos Island Airport and it offers unobstructed visibility of the sky (Geoscience Australia, 2007).

Fig. 9 DARW station is equipped with an ASHTECH UZ-12 GPS receiver with an ASH700936D_M antenna type. This station is located at an old abandoned seismic station at Manton Dam and is surrounded by trees and a building (Geoscience Australia, 2007).

Table 2 The combined mean for the three ionospheric error mitigation methods, based on GPS data collected at the eight ARGN stations.

Combined Mean (unit: m)								
		15m	30m	1hr	2hr	4hr	12hr	24hr
GHIC	E	-0.05	-0.27	-0.29	-0.38	-0.35	0.06	0.05
	N	-0.42	-0.13	-0.12	0.01	0.00	0.00	-0.01
	H	-0.69	-0.14	0.06	0.21	0.22	0.03	0.03
Rapid GIMs	E	0.02	-0.10	-0.05	-0.10	-0.07	-0.01	-0.01
	N	-0.12	0.00	-0.03	-0.02	-0.02	0.00	0.00
	H	-0.14	0.08	0.25	0.18	0.14	0.09	0.08
Klob model	E	0.27	0.48	0.20	-0.14	-0.18	0.41	0.34
	N	0.56	0.53	0.64	0.88	0.74	1.20	1.03
	H	1.71	1.63	1.51	0.67	0.04	0.01	1.03

Table 3 The combined RMS values for the three ionospheric error mitigation methods, based on GPS data collected at the eight ARGN stations.

Combined RMS value (unit: m)								
		15m	30m	1hr	2hr	4hr	12hr	24hr
GHIC	E	0.42	0.71	0.85	0.77	0.57	0.10	0.07
	N	0.44	0.31	0.23	0.09	0.05	0.04	0.03
	H	1.37	0.50	0.30	0.33	0.32	0.13	0.15
Rapid GIMs	E	0.30	0.45	0.32	0.33	0.21	0.05	0.07
	N	0.25	0.27	0.17	0.09	0.06	0.04	0.03
	H	0.91	0.36	0.37	0.23	0.18	0.14	0.12
Klob model	E	0.52	0.75	0.41	0.36	0.37	0.41	0.34
	N	0.71	0.68	0.73	0.94	0.77	1.36	1.09
	H	1.88	1.74	1.58	0.81	0.49	1.87	1.49

As expected, the Klobuchar model solution gives the least accurate positioning solutions. The bias particularly in the height component is larger using this model than the GRAPHIC and GIMs methods. The GIMs method generally provides better point positioning accuracy and precision especially for shorter observation period, as their combined mean and RMS values are the lowest among the three algorithms tested. Although the GRAPHIC provides less accurate solutions over short observation periods, this algorithm is still capable of providing high accuracy point positioning solution over longer observation periods, without the aid of any ionospheric correction products. The GIMs method and GRAPHIC method are remarkably similar especially after a few hours. This is because the use of the GIMs improves the accuracy of the code-based single frequency PPP, but this has impact only on the initial portion of the solutions. After the ambiguities are stabilized, the solutions will follow the more precise carrier phase measurements and the code measurements will only have marginal effect.

6 Discussions

The ionospheric coefficients of the Klobuchar model are transmitted as part of the satellite navigation message, and are available for all single frequency GPS users in real-time. The Klobuchar model was designed to minimise users' computational complexity, while having the potential to correct for at least 50 percent RMS of the ionospheric range error (Klobuchar, 1987). The Klobuchar model based on code solution could provide metre level positioning accuracy using good quality single frequency observations. However, the solutions, particularly in the height component are often not accurate and precise enough for many GPS applications.

Alternatively, single frequency users may fully exploit the IGS combined GIMs to correct for the ionospheric effects. The IGS GIMs come in two products, final and rapid GIMs, and their characteristics have been outlined in Table 1. Currently, the IGS ionospheric products provide accuracy of 2 to 9 TECU at grid points, in which 1 TECU corresponds to 0.16m range error on L1 measurement (Øvstedal, 2002; Chen and Gao, 2005). Users should however note that the accuracy of the map degrades for interpolated points, as the maps are constructed with a spatial resolution of 5˚ in longitude and 2.5˚ in latitude and a temporal resolution of 2 hours. Therefore, the performance of the GIMs for stations located within a grid will not be as optimal for stations located at grid points.

The IGS rapid GIMs and GRAPHIC methods provide comparable point positioning accuracy over long observation sessions (e.g. >12 hours). As for shorter observation session, the GIMs method performs better. Although the GIMs method using L1 code and phase measurement is more robust, one limitation of using this method is that the IGS GIMs are currently not available in real-time (as at February 2007). Therefore, this method is not applicable for real-time single frequency PPP.

On the other hand, the GRAPHIC algorithm is based on the single frequency ionospheric-free linear combination of L1 code and carrier phase measurements. This algorithm takes advantage of the fact that the ionosphere affects the code (delay) and carrier phase (advance) at the same magnitude but opposite in sign (Yunck, 1993; Montenbruck, 2003; Muellerschoen et al., 2004; Simsky, 2006). Limitation of using this method is that the carrier phase ambiguity is required to be estimated and the noise level of this combination is largely dominated by the code measurement noise (Chen and Gao, 2005). This means that the accuracy of the point positioning solutions using GRAPHIC is highly dependent on the L1 code measurements noise. In addition, an estimation process using cumulative measurements has to be applied and a long period of several hours is also required for the float ambiguity parameters to converge (Hèroux et al., 2004; Chen and Gao, 2005). This slow ambiguity convergence issue continues to limit the applicability of single frequency PPP for short occupancy, real-time use, and thus requires further research (Hèroux et al., 2004). Nevertheless, the uniqueness of GRAPHIC algorithm is that it has the obvious advantage of eliminating the first-order ionospheric error in real-time, while being able to provide high accuracy point positioning solutions after long observation sessions, typically more than 12 hours.

7 Conclusions

Different ionospheric error mitigation methods for single frequency PPP have been described, compared and evaluated. It has been shown that both GRAPHIC and GIMs methods are capable of providing high accuracy point positioning solutions to single frequency PPP users. The accuracy of the improved absolute positioning can be confirmed at a sub-metre to metre level for less than 1-hour observation sessions, using high quality single frequency GPS receivers. After 12 to 24 hours, the accuracy of the solution can be as good as a few centimetres (under favourable condition, e.g. low multipath). For short observation sessions, the GIMs method performs better than the GRAPHIC and Klobuchar model. It is expected that the effects of ionosphere will be more pronounced at stations located in the low latitude regions, and that the results obtained from the mid to high latitude stations tend to be more accurate and consistent. However, the effects of the station location on positional accuracy using both GRAPHIC and GIMs methods seem unclear as a result of the test compiled for this paper. Further research will be

undertaken on both ionospheric quiet and disturbed days and their results compared.

One of the limitations using single frequency PPP approach is that the phase ambiguities on L1 are not of integer values, as in double-difference. This is because they are corrupted by the satellite and receiver initial phase biases. PPP technique also requires a long initialization time, typically more than 30 minutes to 2 hours for the float solution to converge (Gao and Garin, 2006). The PPP convergence time varies based on the number and geometry of visible satellites, observation sampling rate and quality, as well as users' defined environment. Therefore, further investigation on the ambiguity convergence time and cycle slips detection algorithm is recommended to improve the robustness of PPP for real-time applications.

Acknowledgements

The authors would like to thank the Geodetic Survey Division of Natural Resources Canada (NRCan) for kindly providing the CSRS-PPP software package and Mr. François Lahaye from NRCan for answering questions concerning the usage of the software. The first author would also like to thank Mr. Pierre Héroux from NRCan for his useful information and valuable technical discussion regarding PPP. The helpful comments and suggestions from the two anonymous reviewers are gratefully acknowledged.

References

Abdel-salam M. (2005) ***Precise Point Positioning Using Un-Differenced Code and Carrier Phase Observations***. PhD Dissertation, The University of Calgary, Canada.

Arbesser-Rastburg B. (2006) ***The Galileo Single Frequency Ionospheric Correction Algorithm***, <http://sidc.oma.be/esww3/presentations/Session4/Arbesser.pdf> (January 2007).

ARINC Research Corporation (2000) ***GPS Interface Control Document ICD-GPS-200 (IRN-200C-004): Navstar GPS Space Segment and Aviation User Interfaces***. CA, USA.

Chen K. and Gao Y. (2005) ***Real-Time Precise Point Positioning Using Single Frequency Data***. Proceedings of the ION GNSS 18th International Technical Meeting of the Satellite Division, Long Beach, CA, 13-16 September, 1514-1523.

CODE (2007) ***Global Ionosphere Maps Produced by CODE***. <http://www.aiub.unibe.ch/content/research/gnss/code___research/igs/global_ionosphere_maps_produced_by_code/index_eng.html> (January 2007).

Gao Y. and Garin L. (2006) ***GNSS Solution: Precise Point Positioning and Its Challenges, Aided-GNSS and Signal Tracking***. <http://www.insidegnss.com/pdf/12-06-GNSSSol.pdf> (December 2006).

Gendt G. (2006) ***[IGSMAIL-5438]: IGS Switch to Absolute Antenna Model and ITRF 2005***. <http://igscb.jpl.nasa.gov/mail/igsmail/2006/msg00161.html> (November 2006).

Geoscience Australia (2007) ***Australian Regional GPS Network***. <http://www.ga.gov.au/geodesy/argn/> (January 2007).

Hernández-Pajares, M. (2007) ***Private Communication***. Group Astronomy and Geomatics, Technical University of Catalonia.

Hèroux P., Gao Y., Kouba J., Lahaye F., Mireault Y., Collins P., Macleod K., Tétreault P. and Chen K. (2004) ***Products and Applications for Precise Point Positioning - Moving Towards Real-Time***. Proceedings of the ION GNSS 17th International Technical Meeting of the Satellite Division, Long Beach, CA, 21-24 September, 1832-1843.

Hèroux P. (2006) ***Private Communication***. Geodetic Survey Division of Natural Resources Canada.

Hofmann-Wellenhof B. Lichtengegger H. and Collins J. (2001) ***GPS Theory and Practice***. (5th edition) Springer-Verlag Wien New York.

IGS (2007) ***International GNSS Service***. <http://igscb.jpl.nasa.gov/> (January 2007).

Klobuchar J.A. (1987) ***Ionospheric Time-Delay Algorithm for Single-Frequency GPS Users***. IEEE Transactions on Aerospace and Electronic Systems, AES-23(3), 325-331.

Kouba J. (2003) ***A Guide to using International GPS Service (IGS) Products***. <ftp://igscb.jpl.nasa.gov/igscb/resource/pubs/GuidetoUsingIGSProducts.pdf> (April 2006).

Kwon J.H., Kim J.W. and Lee D.C. (2001) ***Absolute Kinematic GPS Positioning for Remote Area***. Proceedings of the IEEE Geoscience and Remote Sensing Symposium (IGARSS) 2001, Sydney, Australia, 9-13 July, 2067-2069.

Le A.Q. (2004) ***Achieving Decimetre Accuracy with Single Frequency Standalone GPS Positioning***. Proceedings of the ION GNSS 17th International Technical Meeting of the Satellite Division, Long Beach, CA, 1881-1891.

Montenbruck O. (2003) ***Kinematic GPS Positioning of LEO Satellites Using Ionosphere-Free Single Frequency Measurements***. Aerospace Science and Technology, 7, 396-405.

Muellerschoen R., Iijima B., Meyer R., Bar-Sever Y. and Accad E. (2004) ***Real-Time Point-Positioning Performance Evaluation of Single-Frequency Receivers Using NASA's Global Differential GPS System***. Proceedings of the ION GNSS 17th International Technical Meeting of the Satellite Division, Long Beach, CA, 21-24 September, 1872-1880.

Øvstedal O. (2002) ***Absolute Positioning with Single-Frequency GPS Receivers, GPS Solutions***. 5(4), 33-44.

Schaer S., Gurtner W. and Feltens J. (1998) ***IONEX: The IONosphere Map EXchange Format Version 1***. Proceedings of the IGS AC Workshop, Darmstadt, Germany, 9-11 February.

Simsky A. (2006) ***Standalone Real-Time Navigation Algorithm for Single-Frequency Ionosphere-Free Positioning Based on Dynamic Ambiguities (DARTS-SF)***. Proceedings of the ION GNSS 18th International Technical Meeting of the Satellite Division, Fort Worth, Texas, 301-308.

SOPAC (2007) ***Scripps Orbit and Permanent Array Center***. <http://sopac.ucsd.edu/> (January 2007).

U.S. Naval Research Lab (2007) ***IGS Clock Products Working Group***. <https://goby.nrl.navy.mil/IGStime/ cc2noncc/> (August 2007).

Wang Z., Wu Y., Zhang K. and Meng Y. (2004) ***Triple Frequency Method for High-order Ionospheric Refractive Error Modelling in GPS Modernization,*** J of Global Positioning Systems Vol. 4, No.1+2, 291-295.

Witchayangkoon B. (2000) ***Elements of GPS Precise Point Positioning***, PhD Dissertation, The University of Maine, USA.

Wu S., Yuan Y., Zhang K. and Grenfell R. (2006) ***Temporal and Spatial Variations of the Ionospheric TEC over Victoria for GPSnet-based Real-time Positioning***, J of Global Positioning Systems, Vol.5, No.1-2, pp.52-57.

Wyllie S., Zhang K. and Talbot N. (2006) ***An Analysis of the Temporal Correlation of the Ionospheric Bias Affecting GPS Carrier Phase Observations,*** Proceedings of IGNSS Symposium. Holiday Inn Surfers Paradise, Australia, 17–21 July (79).

Yuan Y., Huo X. and Ou J. (2007) ***Models and Methods for Precise Determination of Ionospheric Delay using GPS***. Progress in Natural Science, 2(17), 187-196.

Yunck T. (1993) ***Coping With the Atmosphere and Ionosphere in Precise Satellite and Ground Positioning***. Geophysical Monograph 73, 13.

Zhang K.; Wu S. and Wu F. (2007) ***The Latest Development of a Network-based RTK System in Australia***, International Journal of Science and Research, Vol. 2(1), pp.87-94.

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 72-79

Journal of Global Positioning Systems
2008
© CPGPS

Statistical Comparison of Various Interpolation Algorithms for Grid-Based Single Shell Ionospheric Model over Indian Region

Ashish K Shukla, Neha Nagori, Saurabh Das, Nishkam Jain, M R Sivaraman, K Bandyopadhyay
SATCOM & IT Applications Area, Space Applications Centre,
Indian Space Research Organisation, Ahmedabad, 380015, INDIA

Abstract

Future Satellite Based Navigation Systems based on GPS, like US Wide Area Augmentation Sysytem (WAAS), are of global interest to the scientific community. Precise estimation of ionospheric delay is most crucial for successful implementation of the systems. Due to the complex ionospheric structure and large variation of Total Electron Content (TEC) in low latitudes, it is necessary to compare and validate the efficiency of the existing algorithms in this region. In this study, performances of the various interpolation algorithms for TEC calculation at ionospheric grid point (IGP) and user position for grid-based Single Shell Model have been tested for 72 test days of 2005. Based on the results obtained from this analysis, it has been found that, for the Indian region, it would be more suitable to use Ordinary Kriging in place of Planar Fit to estimate delay at an IGP (as used by US WAAS). It has also been found that Ordinary Kriging performs better than the Bilinear Interpolation technique at the user end.

Keywords: GAGAN, Ionosphere, Interpolation algorithm, Single Shell Model

1. Introduction

Satellite Based Augmentation Systems (SBAS) for airline navigation are currently under development worldwide. United States was the first country to develop and demonstrate the Wide Area Augmentation System (WAAS, 2008, Walter T, 2002 and http://waas.stanford.edu.). Later on, Europe started European Geostationary Navigation Overlay Service (EGNOS) program and similar programs are at the various stages of development in the countries like Japan, China, Brazil and India. Indian Space Research Organization (ISRO) is involved in the development and establishment of a full complement of Wide Area Augmentation System (WAAS) over Indian Airspace, called GAGAN (GPS Aided Geo Augmented Navigation), in collaboration with Airports Authority of India (AAI).

The primary goal of SBAS is to enhance the accuracy and integrity of user position estimates based upon Global Positioning System (GPS) measurements. In the absence of the selective availability, the ionosphere is the largest source of errors for single-frequency users of GPS. Currently, the ionospheric models used in SBAS rely on the Single-Thin- Shell approximation (Klobuchar,1987; Mannucci et al.,1998). WAAS derives slant ionospheric delay error and confidence bounds from estimates of vertical delay modelled on a grid at regularly spaced intervals of latitude and longitude (5x5 degree). The vertical delay estimates at each IGP is calculated from a Planar Fit technique using measured delay values surrounding that IGP. User estimates delay at its position using Bilinear Interpolation from the delay values at four corners of the grid in which it lies.

Due to presence of complex ionospheric structure and the Equatorial Ionospheric Anomaly over the Indian region (Rama Rao et al, 2006), it is required to test the performance of grid-based thin-shell model with different user Interpolation techniques. In order to find a grid-based model suitable for the Indian region, performance study of the algorithms over large time period is essential. Present analysis has been done for all quiet days (A_P index <50) of the year 2005, and an attempt has been made to find out an algorithm which performs well for the Indian region under these conditions. Choice of testing under quiet conditions was made only to exclude the effects due to various abrupt variations derived from enhanced geomagnetic activities. Issues related to the enhanced geomagnetic activities are beyond the scope of this paper.

In the present study, grid-based single-shell model using Planar Fit (as used by WAAS), Ordinary Kriging and Universal Kriging have been attempted to obtain the vertical ionospheric delay values at the IGPs over the Indian region. In this study, besides Bilinear Interpolation Kriging (Ordinary and Universal) interpolation and Planar Fit has also been used to determine user ionospheric delays. A statistical comparison and

validation of these different algorithms has been done by obtaining the Root Mean Square Error (RMSE) between observed (truth data) and calculated slant ionospheric delay values.

In the following sections, an overview of the grid based estimation of the ionospheric delay by Single Shell Model with a brief description of the different techniques used at an IGP and Ionospheric Pierce Point (IPP) is given.

2. Grid Based Single Shell Model:

GPS signal delay caused by the ionosphere is directly proportional to the number of the free electrons along the signal ray path. The total electron content (TEC) along the ray path of the signal from satellite to receiver may be written as the path integral of the electron density along with the line of sight.

$$TEC = \int_S^R Ne(l)dl \tag{1}$$

where the subscripts S and R identify, respectively, the satellite and receiver in question, and Ne is the electron density.

Thus, solving for slant delay is an inherently three dimensional problem. The Single-Shell Model reduces this three dimensional problem to two-dimensions by introducing the simplifying assumption that the whole ionosphere is compressed only in a neighbourhood of a fixed altitude, taken as 350 km (RTCA special committee 159, 2001, Gao Y. and Liu Z.,2002).

The relation given below allows to estimate the vertical delay at the IPP to be inferred from a measurement of slant delay. Mathematically, it is expressed as:

$$TEC_{slant} = M(E,h).TEC_{Vertical} \tag{2}$$

where

$$M(E,h) = [1-(\mathrm{Re}.\cos E/(\mathrm{Re}+h))^2]^{-1/2} \tag{3}$$

is a mapping function. Where E , h and R_e denote elevation angle, maximum electron density altitude and radius of the Earth respectively. After determining the vertical delays at IGP, user interpolates the vertical delay at its position using these delay values. Using mapping function, these vertical delays are converted to slant delays (RTCA special committee 159, 2001; Mishra and Enge, 2001).

2.1. Planar Fit Model

With Planar Fit method, a smoothened surface of delay variations is obtained by averaging highly variable spatial data (IPP delays). Optimal weights for IPPs are calculated based on the residual errors. The residual error is obtained by using the correlation structure, which includes measurement variances of IPP, de-correlation factor and biases forming weight matrix (Walter et al, 2000).

The ionospheric delay around an IGP is estimated using Planar Fit linear function given as:

$$I_{y,IGP}(x,y) = a_0 + a_1 x + a_2 y \tag{4}$$

Here, a_0 represents exact vertical delay at the IGP, a_1 and a_2 represents delay variation along X and Y axes respectively.

The solution for Planar coefficients (a_0 , a_1 , a_2) are obtained by the weighted least square method and is given as,

$$[a_0 \quad a_1 \quad a_2] = [(G^T.W.G)^{-1}.G^T.W.I_{y,IPP}] \tag{5}$$

where $I_{y,IPP}$ is the Matrix of measured vertical delay at IPPs surrounding an IGP, G is the observation matrix and W is the weight matrix.

As for the calculation of W , the correlation coefficient value for the Indian region is taken as four (Acharya etal,2006). Since IGP is located at the origin in the local coordinate system, vertical delay at IGP can be obtained by determining Planar coefficient a_0 , making other coefficients zero. Finally, the delay estimate at the IGP is given by (Walter et al, 2000)

$$I_{y,IPP} = [1 \quad 0 \quad 0].[(G^T.W.G)^{-1}.G^T.W.I_{y,IPP}] \tag{6}$$

2.2. Ordinary Kriging

It has been observed that the nominal ionosphere can be well characterized by a Planar trend with a covariance depending on the distance, but because of irregular behaviour of the ionosphere, it is inappropriate to estimate delay values using Planar Fit. Therefore another technique called Kriging, has been implemented (Wielgosz P. et al, 2003). This uses a slack variable, called Lagrange multiplier λ to obtain the delay in an optimum and unbiased manner. Kriging is a geo-statistical interpolation technique that considers both the distance and the degree of variation between known data

points when estimating values in unknown areas. A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated (Davis J. C., 1986). Applied properly, Kriging allows the user to derive weights that result in optimal and unbiased estimates. In this technique, the ionospheric estimation is assumed to be a statistical process and the relation between the statistical variables is defined by:

$$\mathrm{var}(Z(x+h) - Z(x)) = 2\gamma h \tag{7}$$

where, $Z(x+h)$ and $Z(x)$ are two spatially separated random outcomes, γ is called the semivariogram. The variogram of measurements is only a function of the distance between measurements. Thus, before the actual interpolation can begin, Kriging must calculate every possible distance weighing function. This is done by generating the experimental semivariogram of the data set and choosing a mathematical model which best approximates the shape of the semivariogram. Spherical model of the semivariogram has been used here and its mathematical form is given as follows:

$$\gamma(h) = c\left(\left(\frac{3d}{2a}\right) - \left(\frac{d}{2a}\right)^3\right), \text{ when } d \le a \tag{8a}$$

$$= c \quad , \text{ when } d > a \tag{8b}$$

where, d is any pre-defined distance, h is the distance between estimated location and observed location. The semivariances have been calculated for different values of h and have been plotted in the form of a semivariogram. If the difference of different values of h is increased the points being compared are less and less closely related to each other and their difference becomes larger, resulting in larger values of $\gamma(h)$. At some distance, the points being compared are so far apart that they are not related to each other. And their squared differences become equal in magnitude to the variance around the average value. The semivariance no longer increases and the semivariogram developed a flat region, called a sill c . The distance at which the semivariance approaches the variance is referred to as range (a) of the rationalised variable (Davis J. C., 1986).

Finally, the delay estimate at IGP, $I_{est}(x)$, is given by

$$I_{est}(x) = \sum_{k=1}^{n} W_k . I_{mes}(x_k) \tag{9}$$

where W_k is the weight factor depending on the distance and $I_{mes}(x_k)$ is the measured ionospheric delay at the k[th] IPP (Cressie, 1993).

2.3. Universal Kriging

In the presence of a trend or slow change in the average value, as in case of low latitude ionosphere, a linear estimator is no longer unbiased. Universal Kriging can then be used and can be regarded as consisting of three operations: first the drift must be estimated and removed. Then the stationary residuals are Kriged to obtain the needed estimates. Finally, the estimated residuals are combined with the drift to obtain the estimates of the actual surface. Therefore, Kriging with a trend or Universal Kriging is the technique, which minimizes the mean-square estimation error given that the spatial structure is given by (Blanch J, 2003)

$$I_y(x, y) = a_0 + a_1 x + a_2 y + R(x, y) \tag{10}$$

where the first three terms define the Planar trend and $R(x, y)$ is a random function with zero mean and a given variogram. The expected value of delay estimate is given by:

$$E(I_{est}(x)) = E\left(\sum_{k=1}^{n} \lambda_k . I_{mes}(x_k)\right) \tag{11}$$

Now, the following equation can be simplified

$$E(a_0 + a_1 x^{est} + a_2 x^{north}) = E(\lambda_1 . \mathrm{Im}ea(x_1) + \lambda_2 . \mathrm{Im}ea(x_2) + \ldots + \lambda_n . \mathrm{Im}ea(x_n)) \tag{12}$$

To get the following form

$$X = \lambda . G^T \tag{13}$$

where G is the geometry matrix, X is the position geometry of the IGP along the east and the north directions and is given as:

$$X = [1 \quad x^{est} \quad x^{north}] \tag{14}$$

The final aim is to find λ such that it minimises the difference between the estimated and the true values at the IGP. λ is obtained by applying the optimization using Lagrange Multiplier and is given as:

$$\lambda = (W - W.G(G^T.W.G)^{-1}.G^T.W)C(x, x_k) + W.G(G^T.W.G)^{-1}.X \tag{15}$$

where, $W = (C(x_i, x_j))^{-1}$

$C(x_i, x_j)$ is an nxn matrix whose elements are computed using the assumed covariance for the process noise. $C(x, x_k)$ is a vector whose elements are the covariance between the residual at the IGP location and each of the residuals at the measurement location.

Finally, vertical delay at the IGP is obtained by (Blanch, 2003),

$$I_{est}(x) = \sum_{k=1}^{n} \lambda_k . I_{mes}(x_k) \qquad (16)$$

Here, an exponential model of the semivariogram has been used to calculate c (Blanch J, 2003).

2.4. Bilinear Interpolation

For four-point interpolation, the mathematical formulation for interpolated vertical IPP delay as a function of IPP latitude and longitude is provided in RTCA document using Bilinear Interpolation (RTCA special committee 159, 2001).

3. Data Collection, Pre-processing and Analysis

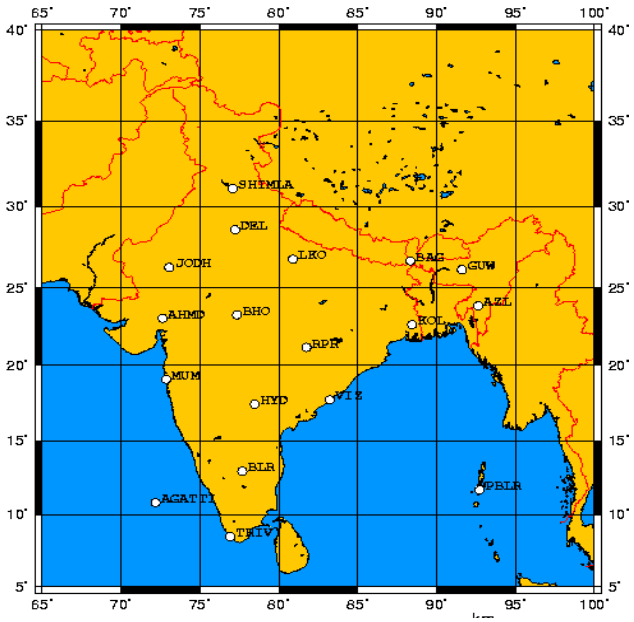

Fig. 1 Map of the TEC stations collecting data for GAGAN

To carry out the comparison and validation of different ionospheric models for GAGAN, TEC data has been collected from 18 TEC stations having dual frequency receivers, spread all over the Indian region (5ºN-35ºN degrees latitude and 70ºE -95ºE degrees longitude). Locations of the TEC stations are shown in Figure 1. These stations are selected nearly at the centres of the eighteen 5ºx5º grids over India. GPS data has been collected with an elevation angle greater than 15 degrees from each of the 18 stations and has been used to calculate TEC values for the present study. These receivers receive pseudorange and carrier phase data from all the visible satellites above the said cut off elevation angle. TEC has been calculated from the data collected over the one minute interval and recorded on a PC.

Before using the TEC data, receiver bias has been removed by using Kalman Filter. Satellite bias correction has also been applied. A brief description of pre-processing and analysis is given as follows:

3.1. TEC estimation

Raw pseudo range and carrier phase data is recorded in .GPS file at 5 Hz interval in dual frequency GPS receiver (Novatel OEM4). Then the .GPS file is converted into RINEX format using a software to get Observation and Navigation files. Thus, we get C_1 and P_2 (pseudo range in meters) and $f1$ and $f2$ (carrier phase in cycles) from L1 and L2 of all visible satellites from the Observation file.

TEC can be estimated from relative code delay and also from relative carrier phase measurements at L_1 and L_2 using dual frequency GPS receiver.
In our analysis estimation of TEC has been done using pseudorange measurement and the relation can be summarized as follows,

$$TEC_\rho = 9.5196(P_2 - C_1) \qquad (17)$$

where C_1 and P_2 are the pseudoranges in meters using C/A code measurement and P code measurement respectively.

Similar expression with carrier phase measurements for estimating TEC may be given as:

$$TEC_\Phi = \frac{(\Phi_2\lambda_2 - \Phi_1\lambda_1) + (N_{f2}\lambda_2 - N_{f1}\lambda_1)}{1.34.10^{-7}(\lambda_2 / f_2 - \lambda_1 / f_1)}$$

$$TEC_\Phi = 9.5499\{\ (\Phi_2\lambda_2 - \Phi_1\lambda_1) + (N_{f2}\lambda_2 - N_{f1}\lambda_1)\ \} \qquad (18)$$

Here, $(\Phi_2\lambda_2 - \Phi_1\lambda_1)$ is in meters, λ_1 and λ_2 are wavelengths at $f1$ and $f2$, and N_{f1} and N_{f2} are integer ambiguity terms at $f1$ and $f2$ respectively.

The above expression (18) is approximately equal to the expression obtained for TEC from the pseudorange (code) difference measurements given in equation (17) except for the integer ambiguity term.

It is well known fact that carrier phase can be measured more precisely than code measurement and so the TEC (Misra P. and Enge P. , 2001). But, carrier phase contains integer ambiguity and code based measurements are noisy. So, code smoothening has been done using the carrier phase measurements, while determining TEC, and this does not involve solving for the integer ambiguity in carrier phase measurements.

To get noise free and absolute TEC, carrier smoothing is done by moving average of the difference between TEC from code and carrier phase calculated at each epoch and added to the carrier phase TEC and is expressed as:

$$TEC = TEC_{\Phi} + (1/K)\sum_{k=0}^{K} (TEC_{\Phi}^{k} - TEC_{\rho}^{k}) \quad (19)$$

where k is the epoch time. The smoothed carrier phase TEC fitted to the level of code TEC to get smoothed code TEC.

Satelite and receiver Inter Frequency Biases (IFB) have also been estimated and removed while calculating TEC. Satellite IFB includes differential P_1 - P_2 code bias and the differential P_1 - C_1 code bias, which are estimated and measured separately. Differential P_1 - P_2 code bias can be determined from the GPS broadcast differential group delay (τ_{GD}) values by calculating GPS offset time and mean τ_{GD} values. Method of estimating P_1 - P_2 bias has been studied and working algorithm is developed. The P_1 - C_1 bias values are available on CODE website (//www.aiub.unibe.ch/ionosphere.html). These values are used and after converting into TEC units, it is also applied to the above equation. Receiver IFB is caused by the difference between the frequency response of filtering the GPS L_1 and L_2 signals, resulting in a time-misalignment of the two pseudoranges. Receiver IFB is not constant and varies with time (Sardon E. and Zarraoa N, 1997). In our approach, these biases have been estimated by Kalman Filter technique (Sardon et al, 1994). For more details on receiver bias correction one may refer to Rajat et al (2007).

Combining all the biases one can write the true TEC by (Wilson B.D. and Manuncci A. J, 1993, Sardon E. and Zarraoa N., 1997),

$$TEC_{true} = TEC + K_r + K^s \quad (20)$$

where, $K_r = R_{bias}$ is the receiver inter frequency bias and $K^s = P_1P_{2bias} - P_1C_{1bias}$ is the satellite bias. It can be expressed in more general form as follows,

$$TEC_{true} = TEC + P_1P_{2bias} - P_1C_{1bias} + R_{bias} \quad (21)$$

3.2. Analysis

For the present analysis, data from all the 18 TEC stations has been used. Analysis has been done for all quiet days of year 2005 with A_p index <50. A statistical validation of the different techniques, described below, has been done.

Bilinear Interpolation is implemented by WAAS to calculate the User Ionospheric Vertical Delay (UIVD) from the four grid values of grid ionosphere vertical delay (GIVD). But, in the present study, authors have also used Kriging (Ordinary and Universal) and Planar Fit techniques to estimate slant delay at user's position. In total, there are three different techniques implemented at the IGP which are Planar Fit, Ordinary Kriging and Universal Kriging and four different techniques at the IPP i.e. Planar Fit, Ordinary Kriging, Universal Kriging and Bilinear Interpolation. From these total six combinations: (a) Ordinary Kriging at IGP with Ordinary Kriging at IPP (b) Ordinary Kriging at IGP with Bilinear Interpolation at IPP (c) Universal Kriging at IGP with Universal Kriging at IPP (d) Universal Kriging at IGP with Bilinear Interpolation at IPP (e) Planar Fit at IGP with Planar Fit at IPP and (f) Planar Fit at IGP with Bilinear Interpolation at IPP, have been tested to find the combination which gives least RMSE in slant TEC (STEC) and highest percentage counts less than three TEC over the Indian region.

3.3. Validation Methodology

In the present study for validation purpose the methodology adopted is to remove each ray and reconstruct it separately with the help of remaining rays. Reconstruction is done in following two steps. First, ionospheric delay values at IGPs have been determined by using Single-Shell model with different algorithms without using that particular ray. In second step, using vertical delay values at the four corners of IGP in which the particular IPP lies, slant ionospheric delays (STEC) at IPP has been reconstructed using all the algorithms considered. In total six different combinations as described above has been analyzed. Yearly averaged Root Mean Square Error (RMSE) in STEC has been estimated for all quiet days of the year 2005. Following formula has been used for the calculation of the RMSE in Slant TEC values:

$$RMSE_{STEC} = \sqrt{\frac{\sum_{i=1}^{N} (STEC_m(i) - STEC_r(i))^2}{N}} \quad (22)$$

where $STEC_m$ is the measured (truth data) STEC, $STEC_r$ is the reconstructed STEC and N is the total number of IPPs.

4. Results and Discussion

The results are shown in the Figures 2 and 3. In Figure 2, middle curve represents the yearly average of the RMSE in STEC for different hours (00UT to 23 UT) with the maximum and minimum values shown by vertical lines. For having a better insight of the error variation, one-sigma value above and below the average has also been plotted along with the average curve. As one-sigma value around the average contains the 68 % of the total points, it can be concluded that the maximum errors occurs only for limited cases and average curve represents the true nature of the errors. All the combinations show significant variation of the errors from the mean line. This is shown by the error bars in Figure 2. Difference in maximum and minimum error values is particularly very large for the Universal Kriging and Planar Fit techniques in comparison with the Ordinary Kriging. This shows the consistency and reliability of the Ordinary Kriging over the other considered algorithms. These results are expected as the equatorial ionosphere varies vary rapidly.

Further, from Figure 2, it can be seen that most of the time maximum error in STEC occurs in the peak hours i.e., from 6-10 UT. This is the time period in which ionospheric anomaly is prominent over the Indian region. Also it is clearly visible that 'Ordinary Kriging at IGP with Ordinary Kriging at IPP' and 'Ordinary Kriging at IGP with Bilinear Interpolation at IPP' performs better than all other combinations considered. Further, Ordinary Kriging estimates the delay more reliably than Bilinear Interpolation technique at IPPs for all the time. This clearly indicates that the ionospheric delay over Indian region can well be modelled with Ordinary Kriging both at IGP and IPP in comparison to other techniques.

Although Universal Kriging technique is theoretically more sound than the Ordinary Kriging, it does not show the improvement over the Ordinary Kriging. Authors feel that the reason for this is the use of a constant correlation coefficient in Universal Kriging, which has been taken as four (Acharya et al, 2006), may not be appropriate for the Indian region and should have been taken as variable.

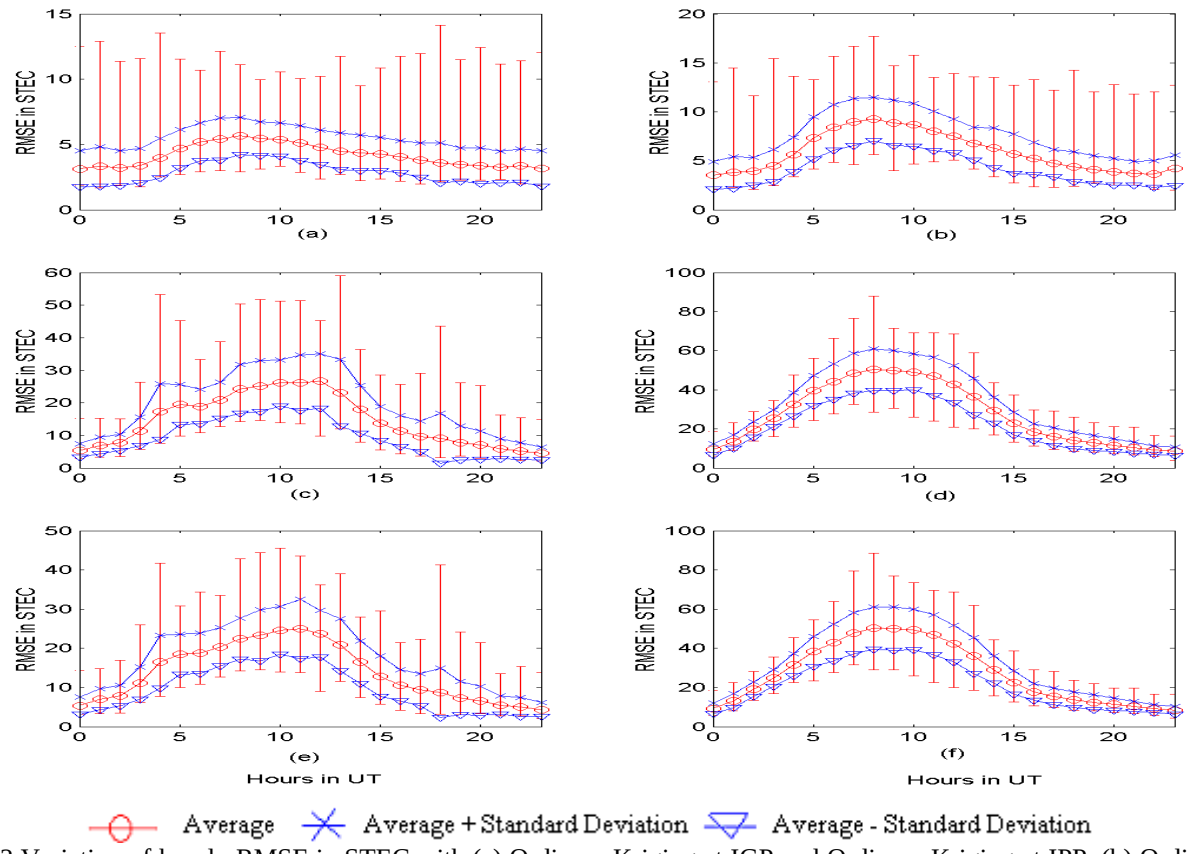

Fig. 2 Variation of hourly RMSE in STEC with (a) Ordinary Kriging at IGP and Ordinary Kriging at IPP, (b) Ordinary Kriging at IGP and Bilinear Interpolation at IPP, (c) Universal Kriging at IGP and Universal Kriging at IPP, (d) Universal Kriging at IGP and Bilinear Interpolation at IPP, (e) Planar Fit at IGP and Planar Fit at IPP, (f) Planar Fit at IGP and Bilinear Interpolation at IPP. All errors shown are in TEC units.

As an stringent accuracy criteria for the Indian region, it was required that the error should not exceed three TEC values (0.48 meters), therefore a comparison plot for percentage count between two best performing combinations have also been given in Figure 3.

'Ordinary Kriging at IGP with Ordinary Kriging at IPP' has shown highest percentage of counts for less than three TEC for all the 24 hours. For this combination, at 00 Hrs UT, percentage count less than three TEC is around 49, it decreases to one around 10 Hrs UT (around peak anomaly time) and again increases to 50 at 23 Hrs UT. During peak anomaly hours of the day percentage count (< 3 TEC) is very low.

From Figure 3, it can also be observed that none of the algorithms used in the single shell model satisfy the criteria required for GAGAN for all hours of the day. Grid spacing of 5x5 degree may be a possible reason for getting low percentage counts less than 3 TEC, which may improve by reducing the grid size for Indian region. Increasing the number of stations may also improve the percentage counts.

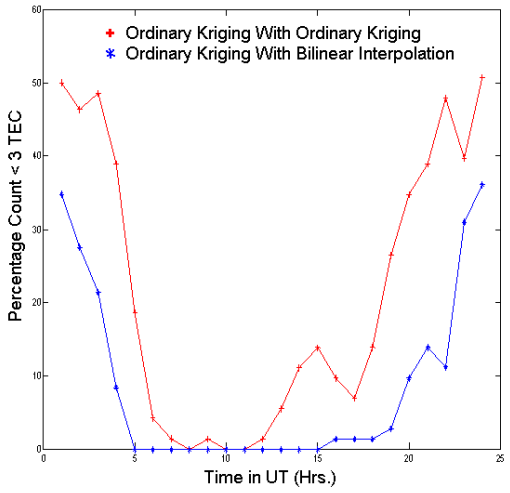

Fig. 3 Comparison of percentage count less than 3 TEC for Ordinary Kriging at IGP with Ordinary Kriging at IPP and Ordinary Kriging at IGP with Bilinear Interpolation at IPP at every hour (00-23 UT) for all the quiet days of 2005.

5. Conclusions

Preliminary study has been performed using different interpolation techniques for grid based Single Shell Model over Indian region for 72 quiet days of 2005 using GPS TEC measurement. It can be concluded from the present study that Planar Fit (as used by US WAAS) technique to calculate VTEC at IGPs may not be suitable for Indian region. On the basis of present study, it has been observed that Ordinary Kriging seems to be a better choice than Planar Fit technique for estimation of VTEC at IGP. Also, instead of Bilinear Interpolation technique at user end, Kriging Interpolation technique seems to be more appropriate for this region.

A more detailed study including disturbed days with more years data is undergoing to meet the GAGAN requirement.

6. Acknowledgements

Authors express their sincere thanks to many Scientists/Engineers from AAI and ISRO, who helped us in the establishment of 18 GPS Stations and TEC data collection, for carrying out this study. Authors are also thankful to Dr K S Dasgupta, Deputy Director, SATCOM & IT Applications Area and Deval Mehta for the internal review of the manuscript. Authors also express their sincere gratitude to the anonymous reviewers for their valuable suggestions.

7. References

Acharya R., Nagori N., Jain N., Sunda S., Sivaraman M. R., Bandyopadhyay K. (2006) ***Ionospheric Correlation Analysis for GAGAN***, Proceedings of CODEC-2006,Kolkata,India.

Acharya R., Nagori N., Jain N., Sunda S., Sivaraman M. R., Bandyopadhyay K. (2007) ***Ionospheric studies for the implementation of GAGAN***, Indian Journal of Radio and Space Physics.

Blanch J (September 2002). ***An Ionospheric Estimation Algorithm for WAAS based on Kriging,***
ION's GPS Meeting, Portland,

Blanch J. (2003) ***Using Kriging bound satellite ranging errors due to the ionosphere***, PhD. Thesis, Stanford University.

Blanch J., Walter T., Enge P.(January 2003), ***Adapting Kriging to the WAAS MOPS Ionospheric Grid,*** ION's National Technical Meeting, Anaheim, CA

CODE website, http://www.aiub.unibe.ch/***ionosphere.html***.

Cressie N.A.C, (1993) ***Statistics for spatial data.*** Revised edition. John Wiley and Sons, New York.

Davis J. C.(1986), ***Statistics and data analysis in geology***, 2nd edition, John Wiley and sons, NewYork

Gao Y. and Liu Z.(2002): ***Precise Ionosphere Modeling Using Regional GPS Network Data,*** Journal of Global Positioning Systems vol 1(1): 18-24

Klobuchar J A (1987), ***Ionospheric Time Delay Algorithm for Single Frequency GPS Users***, IEEE Transactions on Aerospace and Electronics Systems,Vol. AES-23 (3): 325-31.

Mannucci A.J, Wilson B.D, Yuan D.N, Ho C.H, Lindqwister U.J, Runge T.F. (1998), ***A Global mapping technique for GPS-derived ionospheric total electron content measurements***, Radio Science, 33(3): 565-582.

Misra P. and Enge P. (2001), ***Global Positioning System Signals, Measurements and Performance,*** G-J Press, Massachusetts.

Parkinson B. W., Spilker J. J., Axelrad P., and Enge P., Ed (1996), ***Global Positioning System: Theory and Applications (Progress in Astronautics, vols. 62 and 63)*** Washington, DC: AIAA, vol. 1, Chap. 12: 492.

Rama Rao P.V.S, Niranjan K., Prasad D.S.V.V.D, Gopi Krishna S, Uma G (2006), ***On the validity of the ionospheric pierce point (IPP) altitude of 350 km in the Indian equatorial and low-latitude sector.*** Annales Geophysics, Vol 24: 2159-2168.

RTCA special committee 159 (November 2001), ***Minimum Operational Performance Standards for Airborne equipment using Global Positioning System/Wide Area Augmented System,*** RTCA/DO – 229 C

Sardon E., Rus A. and Zarroa N. (1994) ***Estimation of the transmitter and receiver differential biases and the ionospheric Total Electron Content from GPS observations***, Radio Science, vol. 29: 577.

Sardon E., and Zarrao N. (1997), ***Estimation of total electron content using GPS data: How stable are the differential satellite and receiver instrumental biases?,*** Radio Science, vol. 32(5):1899-1910.

Walter T, Hansen A, Blanch J, and Enge P (September 2000) ***Robust detection of ionospheric irregularities,*** in proceedings of ION GPS, Salt Lake City, UT.

Walter T(2002), ***Introduction to the Wide Area Augmentation System,*** Journal of Global Positioning Systems, vol 1(2): 151-153

WAAS (2008) ***Wide area augmentation system***, http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/techops/navservices/gnss/waas/.

Wielgosz P, Brzezinska D and Kashani I(2003), ***Regional Ionosphere Mapping with Kriging and Multiquadric Methods***, Journal of Global Positioning Systems, vol 2(1):48-55

Wilson, B. D. and Mannucci A. J., ***Instrumental Biases in Ionospheric Measurements derived from GPS data***, Proceedings of ION GPS'93, Salt Lake City, 1993.

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 80-80

Journal of Global Positioning Systems
2008
© CPGPS

Technical Notes

"Technical Notes" is a new column in this Journal, featuring reviews of technical or theoretical tools for topics of positioning systems and their applications. Specialists in various fields are welcome to contribute a normal article to outline the issues of interest as systematic possible. In general, the manuscripts may aim to fill the gaps between the textbooks and scientific papers published for the specific topic. In this issue, Dr Jianguo Wang, York University, will review testing statistics in the context of Kalman filter, providing useful tools for design and use of a Kalman filter in navigation and positioning applications.

The column of this issue is coordinated by Dr Jianguo Wang, who appreciates your contribution to this column, along with your comments or ideas for topics for future issues (jgwang@yorku.ca).

Journal of Global Positioning Systems (2008)
Vol. 7, No. 1 : 81-90

Journal of Global Positioning Systems
2008
© CPGPS

Test Statistics in Kalman Filtering

Jian-Guo Wang
Faculty of Science and Engineering, York University

Abstract

Many estimation problems can be modeled using a Kalman filter. One of the key requirements for Kalman filtering is to characterize various error sources, essentially for the quality assurance and quality control of a system. This characterization can be evaluated by applying the principle of multivariate statistics to the system innovations and the measurement residuals. This manuscript will systematically examine the test statistics in Kalman filter on the ground of the normal, χ^2-, ***t***- and ***F***- distributions, and the strategies for global, regional and local statistical tests as well. It is hoped that these test statistics can generally help better understand and perform the statistical analysis in specific applications using a Kalman filter.

Key words: Kalman filter, test statistics, normal distribution, χ^2 distribution, ***t***-distribution, ***F***-distribution

---

1 Introduction

Since 1980s, Geomatics professionals both in research and industry have increasingly shown their profound interest in applying the Kalman filter to various applications, such as kinematic positioning or navigation systems, image processing, and data processing of deformation monitoring etc. Undoubtedly, knowledge of Kalman filtering has become essential to the Geomatics researchers and professionals.

A Kalman filter is simply an optimal recursive data processing algorithm that blends all available information, including measurement outputs, prior knowledge about the system and measuring sensors, to estimate the state variables in such a manner that the error is statistically minimized [Maybeck, 1979]. In practice, linear equation system with white Gaussian noises is commonly taken as the standard model of a Kalman filter. However, one must generally face the following facts [Maybeck, 1979]: (1) no mathematical model is perfect, (2) dynamic systems are driven not only by own control inputs, but also by disturbances which can neither be controlled nor modelled deterministically, and (3) sensors do not provide perfect and complete data about a system.

Hence, a Kalman filter can function properly only if the assumptions about its model structures, dynamical process and measurement noise are correct or realistic. It can become divergent if any of the following situations occurs [Schlee et al, 1967; Tarn et al, 1970; Gelb, 1974; Stöhr, 1986; Loffeld, 1990]:

- Improper system model;
- False modeled process noise;
- False modeled measurement noise or
- Unexpected sudden changes of the state vectors

Correspondingly, one needs to study the behaviors of the errors associated with the system model. This may be called as *system identification* or *system diagnostics*, one of the advanced topics in Kalman filter.

There are different ways to perform system identification. Statistic tests belong to the essential methods of system identification. Herewith, the system model under the Null hypothesis is tested against one or multiple alternative hypotheses. The statistic algorithms can be divided into two categories. The first one is to make multiple hypotheses about the stochastic characteristics of a system (Multiple Hypothesis Filter Detectors) [Willsky, Deyst, Crawford, 1974, 1975; Willsky, 1976]. In order to reach a statistic decision, the posteriori probabilities of the state vectors will be calculated, for instance, through the sequential probability ratio test- SPRT [Willsky, 1976; Yoshimura, et al, 1979]. The second one is to perform the system identification with the help of series of system innovations ("Innovation-based detection") [Mehra, Peschon, 1971; Stöhr, 1986; Salzmann, Teunissen, 1989; etc.]. With this method, the signal is filtered using a normal model, until a failure is found through the statistic tests, for example, through GLR (Generalized likelihood ratio) method [Willsky, 1976; Huep, 1986] or more often through the specific test statistics based on normal, χ^2, ***t*** - or ***F*** - distribution.

Since the Kalman filter was introduced, how to characterize the error sources of a system, especially the

series of the system innovation, has caught certain research attention. [Stöhr, 1986] studied the statistic tests on the ground of Normal Distribution and χ^2 - Distribution using system innovation. [Salzmann, 1993] summarized a three-part test procedure as Detection, Identification and Adaptation (DIA) using system innovation for Kalman filter, in which the construction of test statistics is essential. [Wang, 1997] further discussed the test statistics not only on the basis of normal and χ^2 - distributions, but also on the basis of t - and F - distributions using both of the system innovation and measurement residuals in Kalman filter.

A statistical test is no thing else, but a method of making statistical decisions based on the existing system model using experimental data. One needs statistic tests, for example, to identify abnormal dynamic changes of the system states, or to statistically verify the significance of the additional parameters, such as different sensor biases, in an integrated navigation system. Statistic tests can also help with studying the whiteness of system innovation. The detection of measurement outliers definitely needs statistic tests. A lot more examples exist in practice. They show how essential statistic tests are in Kalman filter so that a developer has to be capable of constructing the proper test statistics and applying them to practice.
However, there is still a lack of systematic description of fundamentals of test statistics for Kalman filter in textbooks. Most of the available works have missed out this topic or, if not the case, the test statistics is mostly based on the Normal Distribution and the χ^2 - Distribution only using system innovation. Applying of the t - and F - tests is not common.

From the perspective of both research and industry, it could be helpful to have a systematical understanding to the fundamentals of test statistics for Kalman filter, in order to use a Kalman filter well or develop new algorithms. However, the existing textbooks about Kalman filtering and applications do not normally talk about testing statistics, although they may be found miscellaneously in scientific publications of different fields. This manuscript aims to fill the gaps between the textbooks and scientific papers in the context of Kalman filter theory and applications.

This manuscript is organized as follows. The algorithm of Kalman filter is summarized in Section 2. Section 3 gives the estimation of the variance factor, or more precisely, the variance of unit weight. The statistic characteristics of filtering solutions are described in Section 4. Sections 5 and 6 are dedicated to building up various test statistics for system innovation and measurement residuals. The concluding remarks are given in the last section.

2. ALGORITHM OF KALMAN FILTERING

The Kalman filter is a set of mathematical equations that provide an efficient recursive means to estimate the state of a process through minimizing its mean squared errors. This section is to provide a brief introduction to the discrete Kalman filter, which includes its description and some discussion of the algorithm.

2.1. The Model

We consider a linear or linearized system with the state-space notation and assume that the data are available over a discrete time series $\{\boldsymbol{t}_0, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_N\}$, which will often be simplified to $\{0,1,\ldots,\boldsymbol{N}\}$. Without loss of generality, a deterministic system input vector will be droped in all of the expressions in this paper. Hence, at any time instant $\boldsymbol{k}$ ($1 \le \boldsymbol{k} \le \boldsymbol{N}$) the system can be written as follows:

$$\boldsymbol{x}(\boldsymbol{k}+1) = \boldsymbol{A}(\boldsymbol{k}+1,\boldsymbol{k})\boldsymbol{x}(\boldsymbol{k}) + \boldsymbol{B}(\boldsymbol{k})\boldsymbol{w}(\boldsymbol{k}) \quad (1)$$

$$\boldsymbol{z}(\boldsymbol{k}+1) = \boldsymbol{C}(\boldsymbol{k}+1)\boldsymbol{x}(\boldsymbol{k}+1) + \Delta(\boldsymbol{k}+1) \quad (2)$$

where $\boldsymbol{x}(\boldsymbol{k})$ is the *n*-dimensional state-vector, $\boldsymbol{z}(\boldsymbol{k})$ is the *p*-dimensional observation vector, $\boldsymbol{w}(\boldsymbol{k})$ is the *m*-dimensional process noise vector, $\Delta(\boldsymbol{k})$ is the *p*-dimensional measurement noise vector, $\boldsymbol{A}(\boldsymbol{k}+1,\boldsymbol{k})$ is the $\boldsymbol{n}\times\boldsymbol{n}$ coefficient matrix of $\boldsymbol{x}(\boldsymbol{k})$, $\boldsymbol{B}(\boldsymbol{k})$ is the $\boldsymbol{n}\times\boldsymbol{m}$ coefficient matrix of $\boldsymbol{w}(\boldsymbol{k})$, $\boldsymbol{C}(\boldsymbol{k})$ is the $p\times n$ coefficient matrix of $\boldsymbol{z}(\boldsymbol{k})$. The random vectors $\boldsymbol{w}(\boldsymbol{k})$ and $\Delta(\boldsymbol{k})$ are generally assumed to be Gaussian with zero-mean:

$$\boldsymbol{w}(\boldsymbol{k}) \sim \boldsymbol{N}(\boldsymbol{o}, \boldsymbol{Q}(\boldsymbol{k})) \quad (3)$$

$$\Delta(\boldsymbol{k}) \sim \boldsymbol{N}(\boldsymbol{o}, \boldsymbol{R}(\boldsymbol{k})) \quad (4)$$

where $\boldsymbol{Q}(\boldsymbol{k})$ and $\boldsymbol{R}(\boldsymbol{k})$ are positive definite variance matrices, respectively. Further assumptions about the random noise are made and specified as follows ($\boldsymbol{i} \ne \boldsymbol{j}$):

$$\boldsymbol{Cov}(\boldsymbol{w}(\boldsymbol{i}), \boldsymbol{w}(\boldsymbol{j})) = \boldsymbol{O} \quad (5)$$

$$\boldsymbol{Cov}(\Delta(\boldsymbol{i}), \Delta(\boldsymbol{j})) = \boldsymbol{O} \quad (6)$$

$$\boldsymbol{Cov}(\boldsymbol{w}(\boldsymbol{i}), \Delta(\boldsymbol{j})) = \boldsymbol{O} \quad (7)$$

Very often, we also have to assume the initial mean and variance-covariance matrix $\boldsymbol{x}(0)$ and $\boldsymbol{D}_{\boldsymbol{xx}}(0)$ for the system state at the time epoch 0. In addition, the initial state $\boldsymbol{x}(0)$ is also assumed to be independent of $\boldsymbol{w}(\boldsymbol{k})$ and $\Delta(\boldsymbol{k})$ for all $\boldsymbol{k}$.

2.2. Kalman Filtering Equations

To derive the optimal estimate $\hat{\boldsymbol{x}}(\boldsymbol{k})$ of $\boldsymbol{x}(\boldsymbol{k})$, one may use one of several optimality criterions to construct the optimal filter. For example, if the least-squares method is used, the optimality is defined in the sense of linear unbiased minimum variance, namely,

$$\left.\begin{aligned} &\boldsymbol{E}\{\hat{\boldsymbol{x}}\} = \boldsymbol{x} \\ &\boldsymbol{E}\{(\boldsymbol{x}-\hat{\boldsymbol{x}})(\boldsymbol{x}-\hat{\boldsymbol{x}})^{\boldsymbol{T}}\} = \min \end{aligned}\right\} \tag{8}$$

where $\hat{\boldsymbol{x}}$ is the unbiased minimum variance estimate of $\boldsymbol{x}$.

Under the given stochastic conditions in Section 2.1, one can derive the Kalman filtering for the state vector at $\boldsymbol{k}+1$:

$$\hat{\boldsymbol{x}}(\boldsymbol{k}+1) = \hat{\boldsymbol{x}}(\boldsymbol{k}+1/\boldsymbol{k}) + \boldsymbol{G}(\boldsymbol{k}+1)\boldsymbol{d}(\boldsymbol{k}+1/\boldsymbol{k}) \tag{9}$$

and its variance-covariance matrix

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k}+1) &= \{\boldsymbol{E}-\boldsymbol{G}(\boldsymbol{k}+1)\boldsymbol{C}(\boldsymbol{k}+1)\}\boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k}+1/\boldsymbol{k}) \\ &\{\boldsymbol{E}-\boldsymbol{G}(\boldsymbol{k}+1)\boldsymbol{C}(\boldsymbol{k}+1)+\boldsymbol{G}(\boldsymbol{k}+1)\boldsymbol{R}(\boldsymbol{k}+1)\boldsymbol{G}^{\boldsymbol{T}}(\boldsymbol{k}+1) \end{aligned} \tag{10}$$

where

$$\hat{\boldsymbol{x}}(\boldsymbol{k}+1/\boldsymbol{k}) = \boldsymbol{A}(\boldsymbol{k}+1,\boldsymbol{k})\hat{\boldsymbol{x}}(\boldsymbol{k}) \tag{11}$$

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k}+1/\boldsymbol{k}) &= \boldsymbol{A}(\boldsymbol{k}+1,\boldsymbol{k})\boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k})\boldsymbol{A}^{\boldsymbol{T}}(\boldsymbol{k}+1,\boldsymbol{k}) \\ &+ \boldsymbol{B}(\boldsymbol{k})\boldsymbol{Q}(\boldsymbol{k})\boldsymbol{B}^{\boldsymbol{T}}(\boldsymbol{k}) \end{aligned} \tag{12}$$

$$\boldsymbol{d}(\boldsymbol{k}+1/\boldsymbol{k}) = \boldsymbol{z}(\boldsymbol{k}+1) - \boldsymbol{C}(\boldsymbol{k}+1)\hat{\boldsymbol{x}}(\boldsymbol{k}+1/\boldsymbol{k}) \tag{13}$$

$$\begin{aligned} \boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}+1/\boldsymbol{k}) &= \boldsymbol{C}(\boldsymbol{k}+1)\boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k}+1/\boldsymbol{k})\boldsymbol{C}^{\boldsymbol{T}}(\boldsymbol{k}+1) \\ &+ \boldsymbol{R}(\boldsymbol{k}+1) \end{aligned} \tag{14}$$

$$\boldsymbol{G}(\boldsymbol{k}+1) = \boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k}+1/\boldsymbol{k})\boldsymbol{C}^{\boldsymbol{T}}(\boldsymbol{k}+1)\boldsymbol{D}_{\boldsymbol{dd}}^{-1}(\boldsymbol{k}+1/\boldsymbol{k}) \tag{15}$$

Here $\hat{\boldsymbol{x}}(\boldsymbol{k}+1/\boldsymbol{k})$ is the one-step prediction of the state vector from the past epoch k with its variance matrix $\boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k}+1/\boldsymbol{k})$, $\boldsymbol{d}(\boldsymbol{k}+1/\boldsymbol{k})$ is the system innovation vector with its variance matrix $\boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}+1/\boldsymbol{k})$, and $\boldsymbol{G}(\boldsymbol{k}+1)$ is the Kalman gain matrix.

An essential characteristic of the sequence $\boldsymbol{d}(1/0)$, …, $\boldsymbol{d}(\boldsymbol{i}+1/\boldsymbol{i})$, …, $\boldsymbol{d}(\boldsymbol{k}+1/\boldsymbol{k})$ is that they are independent from each other epochwise [Stöhr, 1986; Chui, Chen, 1987]:

$$\boldsymbol{Cov}\{\boldsymbol{d}(\boldsymbol{i}+1/\boldsymbol{i}),\boldsymbol{d}(\boldsymbol{j}+1/\boldsymbol{j})\} = \boldsymbol{O} \quad \text{for } (\boldsymbol{i}\neq\boldsymbol{j}) \tag{16}$$

The stochastic characteristics of $\boldsymbol{d}(\boldsymbol{k}+1/\boldsymbol{k})$ are obviously the mixture of the stochastic information from the real observation noise $\{\Delta(1),\Delta(2),\ldots\}$ and the system noise $\{\boldsymbol{w}(0),\boldsymbol{w}(1),\ldots\}$. Traditionally, the system innovation sequences are analyzed and used to build up the test statistics.

2.3. An alternate Derivation of Kalman Filtering

Let us analyze the error sources in Kalman filter in a different way. The optimal estimate $\hat{\boldsymbol{x}}(\boldsymbol{k}+1)$ of $\boldsymbol{x}(\boldsymbol{k}+1)$ at the instant $\boldsymbol{k}$ is always associated with the stochastic information, which may be divided into three independent groups:

a. The real observation noise $\Delta(\boldsymbol{k}+1)$,
b. The system noise $\boldsymbol{w}(\boldsymbol{k})$,
c. The noise from the predicted $\hat{\boldsymbol{x}}(\boldsymbol{k}+1/\boldsymbol{k})$ through $\hat{\boldsymbol{x}}(\boldsymbol{k})$, on which the stochastic characteristics of $\{\Delta(1),\Delta(2),\ldots,\Delta(\boldsymbol{k})\}$, $\{\boldsymbol{w}(0),\boldsymbol{w}(1),\ldots,\boldsymbol{w}(\boldsymbol{k}-1)\}$ are propagated through the system state model.

If these different error resources could be studied separately, it could be very helpful to evaluate the performance of a system in Kalman filter. Along with this line of thinking, the system model as in 2.1 can be reformulated through the three groups of the observation or residual equations as follows:

$$\boldsymbol{v}_{\boldsymbol{l}_x}(\boldsymbol{k}+1) = \hat{\boldsymbol{x}}(\boldsymbol{k}+1) - \boldsymbol{B}(\boldsymbol{k})\,\hat{\boldsymbol{w}}(\boldsymbol{k}) - \boldsymbol{l}_x(\boldsymbol{k}+1) \tag{17}$$

$$\boldsymbol{v}_{\boldsymbol{l}_w}(\boldsymbol{k}+1) = \hat{\boldsymbol{w}}(\boldsymbol{k}) - \boldsymbol{l}_w(\boldsymbol{k}+1) \tag{18}$$

$$\boldsymbol{v}_{\boldsymbol{l}_z}(\boldsymbol{k}+1) = \boldsymbol{C}(\boldsymbol{k}+1)\,\hat{\boldsymbol{x}}(\boldsymbol{k}+1) - \boldsymbol{l}_z(\boldsymbol{k}+1) \tag{19}$$

where the independent (pseudo-)observation groups are simply listed by

$$\boldsymbol{l}_x(\boldsymbol{k}+1) = \boldsymbol{A}(\boldsymbol{k}+1,\boldsymbol{k})\hat{\boldsymbol{x}}(\boldsymbol{k}) \tag{20}$$

$$\boldsymbol{l}_w(\boldsymbol{k}+1) = \boldsymbol{w}_0(\boldsymbol{k}) \tag{21}$$

$$\boldsymbol{l}_z(\boldsymbol{k}+1) = \boldsymbol{z}(\boldsymbol{k}+1) \tag{22}$$

with their variance-covariance matrices by

$$\boldsymbol{D}_{\boldsymbol{l}_x\boldsymbol{l}_x}(\boldsymbol{k}+1) = \boldsymbol{A}(\boldsymbol{k}+1,\boldsymbol{k})\boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k})\boldsymbol{A}^{\boldsymbol{T}}(\boldsymbol{k}+1,\boldsymbol{k}) \tag{23}$$

$$\boldsymbol{D}_{\boldsymbol{l}_w\boldsymbol{l}_w}(\boldsymbol{k}+1) = \boldsymbol{Q}(\boldsymbol{k}) \tag{24}$$

$$\boldsymbol{D}_{\boldsymbol{l}_z\boldsymbol{l}_z}(\boldsymbol{k}+1) = \boldsymbol{R}(\boldsymbol{k}+1) \tag{25}$$

$\boldsymbol{l}_x(\boldsymbol{k}+1)$, $\boldsymbol{l}_w(\boldsymbol{k}+1)$ and $\boldsymbol{l}_z(\boldsymbol{k}+1) = \boldsymbol{z}(\boldsymbol{k}+1)$ are the $\boldsymbol{n}$-, $\boldsymbol{m}$- and $\boldsymbol{p}$-dimentional measurement or pseudo-measurement vectors, respectively. Usually $\boldsymbol{w}_0(\boldsymbol{k}) = \boldsymbol{o}$.

Again, by applying the least squares method, the identical estimate $\hat{\boldsymbol{x}}(\boldsymbol{k}+1)$ of $\boldsymbol{x}(\boldsymbol{k}+1)$ as in the section 2.2 can be obtained. For more details on this alternate derivation of Kalman filter and its advantages, the reader is referred to [Wang, 1997; Caspary and Wang, 1998].

This alternate derivation of Kalman filtering will directly make the measurement residual vectors available for error analysis and possibly to build up the test statistics in Kalman filter, since it is based on the measurement residual vectors. One can now analyse any of three measurement vectors through their own residual vectors. The measurement residual vectors are the functions of the system innovation vector epochwise

$$\boldsymbol{v}_{l_x l_x}(\boldsymbol{k}) = \boldsymbol{D}_{l_x l_x}(\boldsymbol{k})\boldsymbol{D}_{\boldsymbol{xx}}^{-1}(\boldsymbol{k}/\boldsymbol{k}-1)\boldsymbol{K}(\boldsymbol{k})\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1) \quad (26)$$

$$\boldsymbol{v}_{l_w l_w}(\boldsymbol{k}) = \boldsymbol{Q}(\boldsymbol{k}-1)\boldsymbol{B}^{\boldsymbol{T}}(\boldsymbol{k}-1)\boldsymbol{D}_{\boldsymbol{xx}}^{-1}(\boldsymbol{k}/\boldsymbol{k}-1)\boldsymbol{K}(\boldsymbol{k})\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1) \quad .....(27)$$

$$\boldsymbol{v}_{l_z l_z}(\boldsymbol{k}) = \{\boldsymbol{C}(\boldsymbol{k})\boldsymbol{K}(\boldsymbol{k}) - \boldsymbol{E}\}\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1) \quad (28)$$

Similar to (16), we can readily prove the following results of independence:

$$\boldsymbol{Cov}\{\boldsymbol{v}(\boldsymbol{i}),\boldsymbol{v}(\boldsymbol{j})\} = \boldsymbol{O} \quad \text{for } (\boldsymbol{i} \neq \boldsymbol{j}) \quad (29)$$

3. VARIANCE OF WEIGHT UNIT

The posteriori estimation of the variance of weight unit σ_0^2 is essential in Geomatics. Some confusion has been out there in applications of Kalman filer, because the variance-covariance matrices are directly used in Kalman filter. Surely the variance of weight unit, also called as *variance factor*, should be close to unity for a perfect model of system. However, this barely happens in practice.

An algorithm for the estimation of unknown variance factor σ_0^2 was constructed on the ground of the normal-Gamma distribution in [Koch, 1990]. It allows estimating the variance factor together with the state vector in Kalman filter. Alternatively, σ_0^2 can also be estimated by taking advantages of the sequences of the system innovation or the measurement residual vectors in [Wang, 1997 etc]. The single epoch estimate of σ_0^2, also called as the local variance of unit weight, is given by

$$\hat{\sigma}_{l0}^2(\boldsymbol{k}) = \frac{\boldsymbol{v}^{\boldsymbol{T}}(\boldsymbol{k})\boldsymbol{D}_{\boldsymbol{ll}}^{-1}(\boldsymbol{k})\boldsymbol{v}(\boldsymbol{k})}{\boldsymbol{r}(\boldsymbol{k})} \quad (30)$$

where

$$\boldsymbol{v}(\boldsymbol{k}) = [\boldsymbol{v}_{l_x}^{\boldsymbol{T}}(\boldsymbol{k}), \boldsymbol{v}_{l_w}^{\boldsymbol{T}}(\boldsymbol{k}), \boldsymbol{v}_{l_z}^{\boldsymbol{T}}(\boldsymbol{k})]^{\mathrm{T}} \quad (31)$$

$$\boldsymbol{D}_{\boldsymbol{ll}}(\boldsymbol{k}) = \mathrm{diag}\{\boldsymbol{D}_{l_x l_x}(\boldsymbol{k}), \boldsymbol{D}_{l_w l_w}(\boldsymbol{k}), \boldsymbol{D}_{l_z l_z}(\boldsymbol{k})\} \quad (32)$$

and $\boldsymbol{r}(\boldsymbol{k})$ is the number of the redundant measurements at epoch k ($t_0 < t_k \leq t_N$). An alternate expression exists:

$$\hat{\sigma}_{l0}^2(\boldsymbol{k}) = \frac{\boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}/\boldsymbol{k}-1)\boldsymbol{D}_{\boldsymbol{dd}}^{-1}(\boldsymbol{k}/\boldsymbol{k}-1)\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)}{\boldsymbol{r}(\boldsymbol{k})} \quad (33)$$

The proof of equivalence between (30) and (33) can be referred to [Pelzer, 1987; Tao, 1992; Wang, 1997]. One can also estimate the variance factor σ_0^2 over a specific time interval as the regional estimate of variance of unit weight. For example, over a certain specified time interval from epoch ($\boldsymbol{k}-\boldsymbol{s}+1$) to epoch $\boldsymbol{k}$, one can order the system innovation for these $\boldsymbol{s}$ epochs as follows

$$\begin{aligned}\boldsymbol{d}_{\boldsymbol{r}}(\mathrm{s}) = [&\boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}+\boldsymbol{s}-1/\boldsymbol{k}+\boldsymbol{s}-2),\\ &\boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}+\boldsymbol{s}-2/\boldsymbol{k}+\boldsymbol{s}-3),...,\boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}/\boldsymbol{k}-1)]^{\mathrm{T}}\end{aligned} \quad (34)$$

with its variance matrix

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{d}_r(\mathrm{s})\boldsymbol{d}_r(\mathrm{s})} = \mathrm{diag}\{&\boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}+\boldsymbol{s}-1/\boldsymbol{k}+\boldsymbol{s}-2),\\ &\boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}+\boldsymbol{s}-2/\boldsymbol{k}+\boldsymbol{s}-3),...,\boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}/\boldsymbol{k}-1)\}\end{aligned} \quad (35)$$

The regional estimate of σ_0^2 is then equal to

$$\hat{\sigma}_{r0}^2(\boldsymbol{k}) = \frac{\boldsymbol{d}^{\boldsymbol{T}}(\mathrm{s})\boldsymbol{D}_{\boldsymbol{d}_r(\mathrm{s})\boldsymbol{d}_r(\mathrm{s})}^{-1}\boldsymbol{d}(\mathrm{s})}{\boldsymbol{f}_{\boldsymbol{r}}(\mathrm{s})} \quad (36)$$

where $\boldsymbol{f}_{\boldsymbol{r}}(\mathrm{s})$ is the total number of the redundant measurements of $\boldsymbol{s}$ epochs:

$$\boldsymbol{f}_{\boldsymbol{r}}(\mathrm{s}) = \sum_{\boldsymbol{j}=1}^{\boldsymbol{s}} \boldsymbol{r}(\boldsymbol{k}-\boldsymbol{s}+\boldsymbol{j}) \quad (37)$$

Furthermore, the global estimate of σ_0^2 for all of the past $\boldsymbol{k}$ epochs can be calculated by

$$\hat{\sigma}_{g0}^2(\boldsymbol{k}) = \frac{\boldsymbol{d}_{\boldsymbol{g}}^{\boldsymbol{T}}(\boldsymbol{k})\boldsymbol{D}_{\boldsymbol{d}_g(\boldsymbol{k})\boldsymbol{d}_g(\boldsymbol{k})}^{-1}\boldsymbol{d}(\boldsymbol{k})}{\boldsymbol{f}_{\boldsymbol{g}}(\boldsymbol{k})} \quad (38)$$

where

$$\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k}) = [\boldsymbol{d}^{\boldsymbol{T}}(1/0), \boldsymbol{d}^{\boldsymbol{T}}(2/1), \ldots, \boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}/\boldsymbol{k}-1)]^{\mathrm{T}} \quad (39)$$

$$\boldsymbol{D}_{\boldsymbol{d}_g(\boldsymbol{k})\boldsymbol{d}_g(\boldsymbol{k})} = \mathrm{diag}\{\boldsymbol{D}_{\boldsymbol{dd}}(1/0), \boldsymbol{D}_{\boldsymbol{dd}}(2/1),..., \boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}/\boldsymbol{k}-1)\} \quad \ldots\ldots(40)$$

$$\boldsymbol{f}_{\boldsymbol{g}}(\boldsymbol{k}) = \sum_{\boldsymbol{j}=1}^{\boldsymbol{k}} \boldsymbol{r}(\boldsymbol{j}) \quad (41)$$

4. STATISTIC CHARACTERISTICS OF FILTERING SOLUTIONS

In order to evaluate the quality of the solutions and construct different test statistics, the statistic distributions of various random vectors are used in Kalman filter on the ground of the hypothesis: the normal distributed process and measurement noise as given in 3.1 will be discussed.
At an arbitrary epoch k, all of the derived random variables or vectors are the functions of the measurement vector $\boldsymbol{l}(\boldsymbol{k}) = [\boldsymbol{l}_{\boldsymbol{x}}^{\boldsymbol{T}}(\boldsymbol{k}), \boldsymbol{l}_{\boldsymbol{w}}^{\boldsymbol{T}}(\boldsymbol{k}), \boldsymbol{l}_{\boldsymbol{z}}^{\boldsymbol{T}}(\boldsymbol{k})]^{\boldsymbol{T}}$ (see (20) ~ (22)). Among them, $\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)$, $\hat{\boldsymbol{x}}(\boldsymbol{k})$, $\boldsymbol{v}(\boldsymbol{k})$, $\hat{\sigma}_0^2(\boldsymbol{k})$ are essential for quality control of a system. By applying the law of error propagation, their distributions are easily known as:

$$\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{dd}}(\boldsymbol{k}/\boldsymbol{k}-1)) \tag{42}$$

$$\hat{\boldsymbol{x}}(\boldsymbol{k}) \sim \boldsymbol{N}(\mu_{\boldsymbol{x}}, \boldsymbol{D}_{\boldsymbol{xx}}(\boldsymbol{k})) \tag{43}$$

$$\boldsymbol{v}(\boldsymbol{k}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{vv}}(\boldsymbol{k})) \tag{44}$$

$$\begin{aligned}\boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}/\boldsymbol{k}-1)\boldsymbol{D}_{\boldsymbol{dd}}^{-1}(\boldsymbol{k}/\boldsymbol{k}-1)\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)\\ = \boldsymbol{v}^{\boldsymbol{T}}(\boldsymbol{k})\boldsymbol{D}_{\boldsymbol{ll}}^{-1}(\boldsymbol{k})\boldsymbol{v}(\boldsymbol{k}) \sim \chi^2(\boldsymbol{r}(\boldsymbol{k}))\end{aligned} \tag{45}$$

where $\boldsymbol{N}(\boldsymbol{a}, \boldsymbol{b})$ represents a normal distribution with $\boldsymbol{a}$ and $\boldsymbol{b}$ as its expectation and variance, respectively.

The i-th component $\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1)$ in $\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)$ is normally distributed as follows:

$$\begin{aligned}\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1) \sim \boldsymbol{N}(0, \sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1)})\\ (i = 1, 2, \ldots, p;\ k = 1, 2, \ldots N)\end{aligned} \tag{46}$$

Any arbitrary subvector of $\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)$ is also normally distributed. Based on the independency of the innovation vectors between two arbitrary epochs shown in (16), the vectors $\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s})$ and $\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})$ as in (34) and (39) belong to the following normal distributions:

$$\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s})\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s})}) \tag{47}$$

$$\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})}) \tag{48}$$

wherein $\boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s})\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s})}$ and $\boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})}$ are as in (35) and (40). Analog to (46), any arbitrary components $\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})$ and $\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s})$ for the $\boldsymbol{i}$-th type of observations also belong to the normal distribution:

$$\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})}) \quad (i = 1, 2, \ldots, p) \tag{49}$$

$$\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s})\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s})}) \quad (i = 1, 2, \ldots, p) \tag{50}$$

where, for $i = 1, 2, \ldots, n+m+p$,

$$\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k}) = \left(\boldsymbol{d}_{\boldsymbol{i}}(1/0),\ \boldsymbol{d}_{\boldsymbol{i}}(2/1),\ \cdots,\ \boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1)\right)^{\boldsymbol{T}} \tag{51}$$

$$\begin{aligned}\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s}) = \big(\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+1/\boldsymbol{k}-\boldsymbol{s}),\ \boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+2/\boldsymbol{k}-\boldsymbol{s}+1),\\ \cdots,\ \boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1)\big)^{\boldsymbol{T}}\end{aligned} \tag{52}$$

with their corresponding variance-covariance matrices

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})} =\\ \boldsymbol{diag}\left(\sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(1/0)},\ \sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(2/1)},\ \cdots,\ \sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1)}\right)\end{aligned} \tag{53}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s})\boldsymbol{d}_{\boldsymbol{ri}}(\boldsymbol{s})} = \boldsymbol{diag}\big(\sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+1/\boldsymbol{k}-\boldsymbol{s})},\ \sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+2/\boldsymbol{k}-\boldsymbol{s}+1)},\\ \cdots,\ \sigma^2_{\boldsymbol{d}_{\boldsymbol{i}}(\boldsymbol{k}/\boldsymbol{k}-1)}\big)\end{aligned} \tag{54}$$

The i-th component $\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k})$ of $\boldsymbol{v}(\boldsymbol{k})$ is of the normal distribution, too:

$$\begin{aligned}\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k}) \sim \boldsymbol{N}(0, \sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k})})\\ (i = 1, 2, \ldots, n+m+p;\ k = 1, 2, \ldots, N)\end{aligned} \tag{55}$$

For any arbitrary subvector of $\boldsymbol{v}(\boldsymbol{k})$, e.g. $\boldsymbol{v}_{\boldsymbol{l}_{\boldsymbol{x}}}(\boldsymbol{k})$, $\boldsymbol{v}_{\boldsymbol{l}_{\boldsymbol{w}}}(\boldsymbol{k})$ or $\boldsymbol{v}_{\boldsymbol{l}_{\boldsymbol{z}}}(\boldsymbol{k})$, the normal distributions apply. The global cumulative measurement residual vector for all of the past epochs can be defined as

$$\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k}) = \left(\boldsymbol{v}^{\boldsymbol{T}}(1),\ \boldsymbol{v}^{\boldsymbol{T}}(2),\ \cdots,\ \boldsymbol{v}^{\boldsymbol{T}}(\boldsymbol{k})\right)^{\boldsymbol{T}} \tag{56}$$

and the regional cumulative from the past s epochs as

$$\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s}) = \left(\boldsymbol{v}^{\boldsymbol{T}}(\boldsymbol{k}-\boldsymbol{s}+1),\ \boldsymbol{v}^{\boldsymbol{T}}(\boldsymbol{k}-\boldsymbol{s}+2),\ \cdots,\ \boldsymbol{v}^{\boldsymbol{T}}(\boldsymbol{k})\right)^{\boldsymbol{T}} \tag{57}$$

with the following corresponding variance-covariance matrices

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})} =\\ \boldsymbol{diag}\left(\boldsymbol{D}_{\boldsymbol{v}(1)\boldsymbol{v}(1)},\ \boldsymbol{D}_{\boldsymbol{v}(2)\boldsymbol{v}(2)},\ \cdots,\ \boldsymbol{D}_{\boldsymbol{v}(\boldsymbol{k})\boldsymbol{v}(\boldsymbol{k})}\right)\end{aligned} \tag{58}$$

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s})\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s})} = \boldsymbol{diag}\big(\boldsymbol{D}_{\boldsymbol{v}(\boldsymbol{k}-\boldsymbol{s}+1)\boldsymbol{v}(\boldsymbol{k}-\boldsymbol{s}+1)},\ \boldsymbol{D}_{\boldsymbol{v}(\boldsymbol{k}-\boldsymbol{s}+2)\boldsymbol{v}(\boldsymbol{k}-\boldsymbol{s}+2)},\\ \cdots,\ \boldsymbol{D}_{\boldsymbol{v}(\boldsymbol{k})\boldsymbol{v}(\boldsymbol{k})}\big)\end{aligned} \tag{59}$$

So $\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})$ and $\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{k})$ are obviously normally distributed:

$$\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})}) \tag{60}$$

$$\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s})\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s})}) \tag{61}$$

Similar to (49) and (50), the individual components of $\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})$ and $\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{k})$:

$$\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k}) = \left(\boldsymbol{v}_{\boldsymbol{i}}(1),\ \ \boldsymbol{v}_{\boldsymbol{i}}(2),\ \ \cdots,\ \ \boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k})\right)^{\boldsymbol{T}} \tag{62}$$

$$\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s}) = \left(\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+1),\ \ \boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+2),\ \ \cdots,\ \ \boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k})\right)^{\boldsymbol{T}} \tag{63}$$

belong to the following normal distributions

$$\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})}) \tag{64}$$

$$\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s}) \sim \boldsymbol{N}(0, \boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})}) \tag{65}$$

with i = 1, 2, …, n+m+p, where

$$\boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})} = \boldsymbol{diag}\left(\sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(1)},\ \ \sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(2)},\ \ \cdots,\ \ \sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k})}\right) \tag{66}$$

$$\boldsymbol{D}_{\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})} = \boldsymbol{diag}\left(\sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+1)},\ \ \sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k}-\boldsymbol{s}+2)},\ \ \cdots,\ \ \sigma^2_{\boldsymbol{v}_{\boldsymbol{i}}(\boldsymbol{k})}\right) \tag{67}$$

are the variance matrices of $\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})$ and $\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})$.

For multiple components in $\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})$, $\boldsymbol{d}_{\boldsymbol{r}}(\boldsymbol{s})$, $\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})$ or $\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{k})$, the same rule applies.

5. TEST STATISTICS FOR SYSTEM INNOVATION

This section will construct test statistics using system innovation. Under the assumption that no outliers exist in measurements, one could diagnose the possible failure caused by inappropriate state equations. Contrarily, one can identify the possible outliers under the assumption if the system model is assumed to be correct. The cause of a failure may be ambiguous and need to be analyzed in more details.

In this and next sections, we turn to perform statistic tests the epoch $\boldsymbol{k}$ = 1, 2, … from the very beginning to an arbitrary epoch. The statistic tests will be introduced in three different levels, namely, global for all of the past k epochs, regional for an arbitrary continuous epoch group, e.g., the $\boldsymbol{s}$ epochs in the past, and local for a single epoch (often the current epoch). The first two tests are very meaningful for the identification of systematic errors and the local one aims directly at the potential outliers or the unexpected sudden state changes.

5.1. Global Test Statistics

Global tests can be introduced in two different ways to investigate the system behaviors. Right after the first k epochs are completed, one can perform the statistic tests with all of the system innovation information from the past and with their individual components by constructing the corresponding χ^2 – test statistics.

With all of the past k epochs together (k = 1, 2, …, N), the null hypothesis about $d_g(k)$

$$\boldsymbol{H}_0 : \boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k}) = 0 \ \text{ or } \ \boldsymbol{H}_0 : \boldsymbol{E}(\chi^2_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})}) = \tilde{\sigma}^2_0 = 1.0 \tag{68}$$

and its alternative

$$\boldsymbol{H}_1 : \boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k}) \neq 0 \ \text{ or } \ \boldsymbol{H}_1 : \boldsymbol{E}(\chi^2_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})}) \neq \tilde{\sigma}^2_0 = 1.0 \tag{69}$$

can be performed according to (48) by using the test statistic [Salzmann, Teunissen, 1989]

$$\chi^2_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})} = \boldsymbol{d}^{\boldsymbol{T}}_{\boldsymbol{g}}(\boldsymbol{k})\boldsymbol{D}^{-1}_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})}\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k}) \sim \chi^2(\alpha_{\boldsymbol{g}}, \boldsymbol{f}_{\boldsymbol{g}}(\boldsymbol{k})) \tag{70}$$

at a significance level with the Type I error α_g. $\chi^2(\alpha, \boldsymbol{f})$ is the $(1-\alpha)$ - critical value from χ^2 – Distribution with the degrees of freedom of f after (41). The null hypothesis (68) will be rejected if

$$\chi^2_{\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})} > \chi^2(\alpha_{\boldsymbol{g}}, \boldsymbol{f}_{\boldsymbol{g}}(\boldsymbol{k})) \tag{71}$$

The test can easily be extended to the i-th component $\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})$ in $\boldsymbol{d}_{\boldsymbol{g}}(\boldsymbol{k})$ for i = 1, 2, …, p and k = 1, 2, …, N. Under the null hypothesis

$$\boldsymbol{H}_0 : \boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k}) = 0 \tag{72}$$

against its alternative

$$\boldsymbol{H}_1 : \boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k}) \neq 0 \tag{73}$$

Based on (49), the test statistic can be given by

$$\chi^2_{\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})} = \boldsymbol{d}^{\boldsymbol{T}}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{D}^{-1}_{\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})}\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k}) \sim \chi^2(\boldsymbol{k}) \tag{74}$$

If

$$\chi^2_{\boldsymbol{d}_{\boldsymbol{gi}}(\boldsymbol{k})} > \chi^2(\alpha_{\boldsymbol{gi}}, \boldsymbol{k}) \qquad (i = 1, \ldots, p) \tag{75}$$

under the given significance level $\alpha_{\boldsymbol{gi}}$, the null hypothesis will be rejected.

5.2. Regional Tests

For the regional system diagnose, the processed k epochs can be grouped at the user's wish. Without loss of the generality, the discussion here will be limited to two groups. We consider having the first group for the first k – s epochs (from 1 to epoch k – s) and the second group for

the rest of the epochs (from epoch $k - s + 1$ to epoch k) as $\boldsymbol{d}_r(\boldsymbol{k-s})$ (equivalent to $\boldsymbol{d}_g(\boldsymbol{k-s})$) and $\boldsymbol{d}_r(\boldsymbol{s})$.

The null hypothesis about $\boldsymbol{d}_r(\boldsymbol{s})$ is

$$\boldsymbol{H}_0 : \boldsymbol{d}_r(\boldsymbol{s}) = 1.0 \tag{76}$$

against the alternative

$$\boldsymbol{H}_1 : \boldsymbol{d}_r(\boldsymbol{s}) \neq 0 \tag{77}$$

On the ground of the test statistic [Willsky, 1976; Stöhr, 1986; Salzmann, Teunissen, 1989], the following test is performed

$$\chi^2_{\boldsymbol{d}_r(\boldsymbol{s})} = \boldsymbol{d}_r^{\boldsymbol{T}}(\boldsymbol{s})\boldsymbol{D}^{-1}_{\boldsymbol{d}_r(\boldsymbol{s})\boldsymbol{d}_r(\boldsymbol{s})}\boldsymbol{d}_r(\boldsymbol{s}) \sim \chi^2(\boldsymbol{f}_r(\boldsymbol{s})) \tag{78}$$

at the significance level of α_r, where the number $\boldsymbol{f}_r(\boldsymbol{s})$ is the degrees of freedom as in (37). For the second group $\boldsymbol{d}_r(\boldsymbol{k-s})$ also has the χ^2 - distribution as

$$\chi^2_{\boldsymbol{d}_g(\boldsymbol{k-s})} = \boldsymbol{d}_g^{\boldsymbol{T}}(\boldsymbol{k-s})\boldsymbol{D}^{-1}_{\boldsymbol{d}_g(\boldsymbol{k-s})\boldsymbol{d}_g(\boldsymbol{k-s})}\boldsymbol{d}_g(\boldsymbol{k-s}) \sim \chi^2(\boldsymbol{f}_g(\boldsymbol{k-s})) \tag{79}$$

An additional *F*–test statistic can be constructed to test the variance homogeneity between (78) and (79) or (70) because $\boldsymbol{d}_r(\boldsymbol{s})$ is independent from $\boldsymbol{d}_r(\boldsymbol{k-s})$. This *F*–Test is given by

$$\boldsymbol{F}_{\boldsymbol{d}_r(\boldsymbol{s})} = \frac{\hat{\sigma}^2_{r0}(\boldsymbol{s})}{\hat{\sigma}^2_{g0}(\boldsymbol{k-s})} \sim \boldsymbol{F}(\boldsymbol{f}_r(\boldsymbol{s}), \boldsymbol{f}_g(\boldsymbol{k-s})) \tag{80}$$

with

$$\hat{\sigma}^2_{r0}(\boldsymbol{s}) = \frac{\boldsymbol{d}_r^{\boldsymbol{T}}(\boldsymbol{s})\boldsymbol{D}^{-1}_{\boldsymbol{d}_r(\boldsymbol{s})\boldsymbol{d}_r(\boldsymbol{s})}\boldsymbol{d}_r(\boldsymbol{s})}{\boldsymbol{f}_r(\boldsymbol{s})} \tag{81}$$

$$\hat{\sigma}^2_{g0}(\boldsymbol{k-s}) = \frac{\boldsymbol{d}_g^{\boldsymbol{T}}(\boldsymbol{k-s})\boldsymbol{D}^{-1}_{\boldsymbol{d}_g(\boldsymbol{k-s})\boldsymbol{d}_r(\boldsymbol{k-s})}\boldsymbol{d}_g(\boldsymbol{k-s})}{\boldsymbol{f}_g(\boldsymbol{k-s})} \tag{82}$$

$\boldsymbol{F}(\boldsymbol{a},\boldsymbol{b})$ in (80) is the critical value of the Fisher distribution with the 1st degrees of freedom $\boldsymbol{a}$ for the numerator and the 2nd one $\boldsymbol{b}$ for the denominator. This test is always one-sided under a user- specified Type I error α as the significance level. An exchange between the numerator and the denominator may need in case $\hat{\sigma}^2_{g0}(\boldsymbol{k-s})$ greater than $\hat{\sigma}^2_{r0}(\boldsymbol{s})$. This test is commonly employed to diagnose the significant difference between the first $k - s$ epochs and the rest of s epochs.

For the i-th component $\boldsymbol{d}_{ri}(\boldsymbol{s})$ in $\boldsymbol{d}_r(\boldsymbol{s})$, one can also construct a χ^2 test based on (50) and another $\boldsymbol{F}$–test analogue to (80). It runs

$$\boldsymbol{H}_0 : \boldsymbol{d}_{ri}(\boldsymbol{s}) = 0 \tag{83}$$

against the alternative

$$\boldsymbol{H}_1 : \boldsymbol{d}_{ri}(\boldsymbol{s}) \neq 0 \tag{84}$$

by using the χ^2 test statistics

$$\chi^2_{\boldsymbol{d}_{ri}(\boldsymbol{s})} = \boldsymbol{d}_{ri}^{\boldsymbol{T}}(\boldsymbol{s})\boldsymbol{D}^{-1}_{\boldsymbol{d}_{ri}(\boldsymbol{s})\boldsymbol{d}_{ri}(\boldsymbol{s})}\boldsymbol{d}_{ri}(\boldsymbol{s}) \sim \chi^2(\boldsymbol{s}) \tag{85}$$

An $\boldsymbol{F}$-test runs for their variance homogeneity between (85) and (79) or (70) as follows

$$\boldsymbol{F}_{\boldsymbol{d}_{gi}(\boldsymbol{s})} = \frac{\boldsymbol{d}_{ri}^{\boldsymbol{T}}(\boldsymbol{s})\boldsymbol{D}^{-1}_{\boldsymbol{d}_{ri}(\boldsymbol{s})\boldsymbol{d}_{ri}(\boldsymbol{s})}\boldsymbol{d}_{ri}(\boldsymbol{s})/\boldsymbol{s}}{\hat{\sigma}^2_{g0}(\boldsymbol{k-s})} \sim \boldsymbol{F}(\boldsymbol{s}, \boldsymbol{f}_g(\boldsymbol{k-s})) \tag{86}$$

for $i = 1, \ldots, p$ at the significant level of α_{ri}.

5.3. Local Tests

Through the local system diagnose, the tests can be introduced for the innovation vector as a whole and for its components, respectively.

At an arbitrary epoch k, the null hypothesis for $\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1})$

$$\boldsymbol{H}_0 : \boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1}) = 0 \tag{86}$$

against the alternative

$$\boldsymbol{H}_0 : \boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1}) \neq 0 \tag{87}$$

can be given. Its test statistic runs

$$\chi^2_{\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1})} = \boldsymbol{d}^{\boldsymbol{T}}(\boldsymbol{k}/\boldsymbol{k-1})\boldsymbol{D}^{-1}_{\boldsymbol{dd}}(\boldsymbol{k}/\boldsymbol{k-1})\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1}) \sim \chi^2(\boldsymbol{r}(\boldsymbol{k})) \tag{88}$$

at the significance level of α_l with the degrees of freedom $r(k)$.

A further $\boldsymbol{F}$–test statistic can be introduced as

$$\boldsymbol{F}_{\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1})} = \frac{\hat{\sigma}^2_0(\boldsymbol{k})}{\hat{\sigma}^2_{r0}(\boldsymbol{s})} \sim \boldsymbol{F}(\boldsymbol{r}(\boldsymbol{k}), \boldsymbol{f}_r(\boldsymbol{s})) \; (k = 2, 3, \ldots, N) \tag{89}$$

for the variance homogeneity between (81) and (33) or (88). $\boldsymbol{s}$ means arbitrary specific epochs between epoch 1 and epoch $k - 1$.

The quadratic form in (45) contains the entire information from the system innovation for an arbitrary epoch. Therefore, the causes of a system failure must be localized after the rejection of a $\chi^2_{\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1})}$ or $\boldsymbol{F}_{\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k-1})}$ test. It should orient to the individual error sources, e.g.

the individual measurements or the individual process noise factors etc. in kinematic positioning or navigation. One should perform the further statistic tests for the individual measurements.

In order to perform the statistic tests for multiple components in $\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)$, the method for detection of position displacements in deformation analysis can be employed. More on this can be found in [Chrzanowski, Chen, 1986].

The test statistic for single component of $\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)$ can directly be constructed on the ground of the normal distribution or the *t*–distribution. The null hypothesis is

$$\boldsymbol{H}_0 : \boldsymbol{E}(\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)) = 0 \tag{90}$$

with its alternative

$$\boldsymbol{H}_1 : \boldsymbol{E}(\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)) \neq 0 \tag{91}$$

for i = 1,2, …, p and k = 1, 2, …, N. According to (46) the test is performed

$$\boldsymbol{N}_{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)} = \frac{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)}{\sigma_{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)}} \sim \boldsymbol{N}(0,1) \tag{92}$$

at the significant level of $\alpha_{\boldsymbol{li}}$. The null hypothesis will be accepted if the two-sided test satisfies

$$-\boldsymbol{u}_{1-\frac{\alpha_{li}}{2}} \leq \frac{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)}{\sigma_{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)}} \leq \boldsymbol{u}_{1-\frac{\alpha_{li}}{2}} \qquad (i = 1, 2, \ldots, p;\ k = 1, 2, \ldots, N) \tag{93}$$

where $\boldsymbol{u}_{1-\frac{\alpha_{li}}{2}}$ is a $(1-\frac{\alpha_{li}}{2})$ -critical value from the standard normal distribution. Furthermore, based on the past system information, (93) can be extended to the following *t*–test

$$\boldsymbol{T}_{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)} = \frac{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)/\sigma_{\boldsymbol{d}_i(\boldsymbol{k}/\boldsymbol{k}-1)}}{\hat{\sigma}_{\boldsymbol{r}0}(\boldsymbol{s})} \sim \boldsymbol{t}(\boldsymbol{f}_{\boldsymbol{r}}(\boldsymbol{s})) \qquad (i = 1, 2, \ldots, p;\ k = 2, 3, \ldots, N) \tag{94}$$

The most common case is to test the current epoch $\boldsymbol{k}$ vs. the past $k-1$ epochs.

The differences between (93)-(94) and (88)-(89) are obvious. However, which one is preferable will absolutely depend on the user. A ***t***-test or an ***F***-test may deliver the more reliable results of fit to the real data, while a normal or a χ^2 test is introduced with respect to the a-priori assumption.

6. TEST STATISTICS FOR MEASUREMENT RESIDUALS

As it can be seen, the system innovation mixes up different types of information. But it is transferred to the individual measurement residuals epoch by epoch through (26) ~ (28), i.e., the residual vector $\boldsymbol{v}_{\boldsymbol{l}_x}(\boldsymbol{k})$ for the predicted state vector, the residual vector $\boldsymbol{v}_{\boldsymbol{l}_w}(\boldsymbol{k})$ for the process noise and the residual vector $\boldsymbol{v}_{\boldsymbol{l}_z}(\boldsymbol{k})$ for the direct measurements. In this way, these different types of random information can separately be studied. On the basis of the fact that (30) and (33) are equivalent, the test statistics $\chi^2_{\boldsymbol{d}_g(\boldsymbol{k})}$, $\chi^2_{\boldsymbol{d}_r(\boldsymbol{s})}$ and $\chi^2_{\boldsymbol{d}(\boldsymbol{k}/\boldsymbol{k}-1)}$ in (70), (78) and (88) can also be derived using the measurement residual vector $\boldsymbol{v}(\boldsymbol{k})$. But it is not necessary to be repeated here. Therefore, only the test statistics for the individual components will be discussed in this section.

6.1. Global Tests

For the i-th component $\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})$ in $\boldsymbol{v}_{\boldsymbol{g}}(\boldsymbol{k})$, the null hypothesis is

$$\boldsymbol{H}_0 : \boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k}) = 0 \qquad (i = 1, 2, \ldots, n+m+p) \tag{95}$$

against the alternative

$$\boldsymbol{H}_1 : \boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k}) \neq 0 \tag{96}$$

The test statistic is given by

$$\chi^2_{\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})} = \boldsymbol{v}^{\boldsymbol{T}}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{D}^{-1}_{\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k})}\boldsymbol{v}_{\boldsymbol{gi}}(\boldsymbol{k}) \sim \chi^2(\boldsymbol{k}) \tag{97}$$

with the degrees of freedom of k. The null hypothesis will be rejected if

$$\chi^2_{V_{gi}(k)} > \chi^2(\boldsymbol{k}) \qquad (i = 1, 2, \ldots, n+m+p) \tag{98}$$

at the significant level of $\overline{\alpha}_{\boldsymbol{gi}}$.

6.2. Regional Tests

For the i-th component $\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})$ from $\boldsymbol{v}_{\boldsymbol{r}}(\boldsymbol{s})$, a χ^2–test and a *F*–test can be constructed. The null hypothesis is

$$\boldsymbol{H}_0 : \boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s}) = 0 \tag{99}$$

with the alternative

$$\boldsymbol{H}_1 : \boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s}) \neq 0 \tag{100}$$

The corresponding test statistic is given by

$$\chi^2_{\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})} = \boldsymbol{v}^{\boldsymbol{T}}_{\boldsymbol{ri}}(\boldsymbol{s})\boldsymbol{D}^{-1}_{\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s})}\boldsymbol{v}_{\boldsymbol{ri}}(\boldsymbol{s}) \sim \chi^2(\boldsymbol{s}) \qquad (i = 1, 2, \ldots, n+m+p) \tag{101}$$

with the degrees of freedom of *s*. For the variance homogeneity between the independent $\chi^2_{v_{ri}(s)}$ and $\chi^2_{d_g(k-s)}$ similar to (80), the *F*–test statistic can be introduced as

$$F_{v_{gi}(s)} = \frac{v_{ri}^T(s) D^{-1}_{v_{ri}(s)v_{ri}(s)} v_{ri}(s)/s}{\hat{\sigma}^2_{g0}(k-s)} \sim F(s, f_g(k-s)) \quad \ldots\ldots (102)$$

6.3. Local Tests

The single outlier detection at a single epoch can be modeled through the null hypothesis

$$H_0 : v_i(k) = 0 \quad (103)$$

against the alternative

$$H_1 : v_i(k) \neq 0 \quad (104)$$

for $i = 1, 2, \ldots, m+n+p$ and $k = 1, 2, \ldots$ after the test of its standardized residual

$$N_{v_i(k)} = \frac{v_i(k)}{\sigma_{v_i(k)}} \sim N(0,1) \quad (105)$$

at the significant level of $\bar{\alpha}_{li}$. The null hypothesis (102) will be accepted if

$$-u_{1-\frac{\bar{\alpha}_{li}}{2}} \leq \frac{v_i(k)}{\sigma_{v_i(k)}} \leq u_{1-\frac{\bar{\alpha}_{li}}{2}} \quad (i = 1, 2, \ldots, m+n+p;\ k = 1, 2, \ldots, N) \quad (106)$$

Besides a *t*-test between (105) and (78) or (70) can be introduced as follows

$$T_{v_i(k)} = \frac{v_i(k)/\hat{\sigma}_{v_i(k)}}{\sigma_{r0}(s)} \sim t(f_r(s)) \quad (i = 1, 2, \ldots, n+m+p;\ k = 2, \ldots, N) \quad (107)$$

For any epoch *k*, one can use $v_{l_x}(k)$, $v_{l_w}(k)$ and $v_{l_z}(k)$, along with $v(k)$ to investigate the statistic characteristics of measurement vectors $l_x(k)$, $l_w(k)$ and $l_z(k)$, especially two latter ones. Multiple components are possibly diagnosed together in the same way as mentioned in 5.3.

7. CONCLUDING REMARKS

Based on the standard model of Kalman filter, different test statistics have been elaborated on the basis of the normal, χ^2-, ***t***- and ***F***-distributions in this manuscript. This work can be conducive to better understanding of the statistic fundamentals in Kalman filter, provides some insights into statistic testing methods and applications.

In particular, the system innovation vector is transformed to the residual vectors of three measurement and pseudo-measurement groups by the aid of an alternative derivation of Kalman filter algorithm. This makes possible to construct test statistics directly using the measurement residual vectors so that the system diagnosis can directly aim at different error sources of interests to users. The given posteriori estimate of variance of weight unit in Section 3 can be used either to scale the variance and covariance matrices, or to reveal the difference between the model and the processed data set. On the ground of statistic characteristics of filter solutions summarized in the section 4, the test statistics using the series of system innovation are constructed in the section 5 globally with χ^2- test, regionally either with χ^2- test or ***F*** - test, and locally either with ***t*** - test or the normal test according to the normal distribution. Analogous to the section 5, the section 6 constructs the corresponding test statistics using the measurement residuals. Fortunately, χ^2- test, ***F*** - test, ***t*** - test and the normal test are four most commonly used statistic tests. The choice between a χ^2 - test and a ***F*** - test, or between a ***t*** - test and a normal test, wherever two parallel tests are available, is left to the user.

How to construct a test statistics is more or less a theoretical task. But how to efficiently design the procedures to introduce the statistic tests in practice mostly depends on the understanding about the theory and the application. Practical experience plays an essential role in helping deliver a realistic and reliable test scheme. This manuscript however has limited to the testing statistics for general purposes instead of a specific application.

References

Caspary, Wilhelm; Wang, Jianguo (1998). ***Redundanzanteile und Varianzkomponenten im Kalman Filter***. Vol.123, No.4, 1998, pp.121-128.

Chrzanowski, A.; Chen, Y. Q. (1986). ***Report on the Ad Hoc Committee on the Analysis of Deformation Surveys***. XVIII International Congress FIG, Toronto, Canada, 1~11 June 1986, pp. 166–185.

Gelb, Arthur et al. (1974). ***Applied Optimal Estimation***. The Analytic Sciences Corporation & The M.I.T. Press, Cambridge Massachusetts London, 1974.

Huep, Wolfgang (1986). ***Zur Positionsschätzung im gestörten Kalman–Filter am Beispiel eines manövrierenden Wasserfahrzeuges***. Dissertation, Wissenschaftliche Arbeiten der Fachrichtung Vermessungswesen der Universität Hannover, Heft 143, Hannover, 1986.

Koch, K. Rudolf (1990). ***Bayesian Inference with Geodetic Applications***, 2nd edition, Springer-Verlag, Berlin, 1990.

Loffeld, Otmar. (1990). ***Estimationstheorie II, Anwendung–Kalman–Filter***. R. Oldenbourg Verlag, München Wien, 1990.

Maybeck, Peter S. (1979). ***Stochastic models, estimation, and control***. Volume 1, Academic Press, New York, 1979.

Mehra, R. K.; Peschon, J. (1971). ***An Innovations Approach to Fault Detection and Diagnosis in Dynamic Systems***. Automatica, Vol. 16, 1971, S. 637–640.

Pelzer, Hans (1987). ***Deformationsuntersuchungen auf der Basis kinematischer Bewegungsmodelle***. AVN, Year 94, No. 2, February 1987, pp. 49~62.

Salzmann, Martin; Teunissen, P.J.G. (1989). ***Quality Control in Kinematic Data Processing***. Land Vehicle Navigation 1989, DGON, Verlag TÜV Rheinland, Köln, 1989, pp. 355–366.

Salzmann, Martin (1993). ***Least Squares Filtering and Testing for Geodetic Navigation Applications***. Number 37, Publication on Geodesy, Netherlands Geodetic Commission, Delft, 1993.

Schlee, F. H.; Standish, C. J.; et al (1967). ***Divergence in the Kalman Filter***. AIAA Journal, Vol. 5, No. 6, June 1967, pp.1114–1120.

Stöhr, M. (1986). ***Der Kalman–Filter und seine Fehlerprozeße unter besonderer Berücksichtigung der Auswirkung von Modellfehlern***. Forschung–Ausbildung–Weiterbildung Bericht Heft 19, Universität Kaiserlautern, Fachbereich Mathematik, August 1986.

Tao, Benzao (1992). ***Statistic Analysis of Measurements***. Publishing house of Surveying and Mapping, Beijing, June 1992.

Tarn, T. J.; Zaborszky, J. (1970). ***A Practical, Nondiverging Filter***. AIAA Journal, Vol. 8, No. 6, June 1970, pp.1127–1133.

Wang, Jianguo (1997). ***Filtermethoden zur fehlertoleranten kinematischen Positions-bestimmung***. Schritenreihe Studiengang Vermessungswesen, Federal Arm-Forced University Munich, Germany, No. 52, Neubiberg, 1997.

Willsky, Alan S., Deyst, J. J.; Crawford, B. S. (1974). ***Adaptive Filtering and Self–Test Methods for Failure Detection and compensation***. Proceedings of the 1974 JACC, Austin, Texas, June 19–21, 1974.

Willsky, Alan S., Deyst, J. J.; Crawford, B. S. (1975). ***Two Self–Test Methods Applied to an Inertial System Problem***. Journal of Spacecraft, Vol. 12, No. 7, July 1975, S. 434–437.

Willsky, Alan S. (1976). ***A Survey of Design Methods for Failure Detection in Dynamic Systems***. Automatica, Vol. 12, 1976, pp. 601–611.

Yoshimura, Toshio; et al. (1979). ***A Sequential Failure Detection Approach and the Identification of Failure Parameters***. International Journal of System Science, Vol. 10, No. 7, 1979, S. 827 –836.