mobile communications series

Ahmed El-Rabbany

INTRODUCTION TO GPS

THE GLOBAL POSITIONING SYSTEM

Introduction to GPS

The Global Positioning System

For a complete listing of the *Artech House Mobile Communications Series,*
turn to the back of this book.

Introduction to GPS

The Global Positioning System

Ahmed El-Rabbany

Artech House
Boston • London
www.artechhouse.com

Library of Congress Cataloging-in-Publication Data
El-Rabbany, Ahmed.
Introduction to GPS: the Global Positioning System/Ahmed El-Rabbany.
p. cm.—(Artech House mobile communications series)
Includes bibliographical references and index.
ISBN 1-58053-183-1 (alk. paper)
1. Global Postioning System. I. Title. II. Series.
G109.5E6 2002
910'.285—dc21 2001055249

British Library Cataloguing in Publication Data
El-Rabbany, Ahmed
Introduction to GPS: the global positioning system/Ahmed El-Rabbany.
—(Artech House mobile communications series)
1. Global Positioning System
I. Title
629'.045
ISBN 1-58053-183-0

Cover design by Yekatarina Ratner

685 Canton Street
Norwood, MA 02062

International Standard Book Number: 1-58053-183-0
Library of Congress Catalog Card Number: 2001055249

10 9 8 7 6 5 4 3 2 1

To the people who made significant contributions to my life—
My parents, my wife, and my children

Contents

Preface

The idea of writing an easy-to-read, yet complete, GPS book evolved during my industrial employment term during the period from 1996 to 1997. My involvement in designing and providing short GPS courses gave me the opportunity to get direct feedbacks from GPS users with a wide variety of expertise and background. One of the most difficult tasks, which I encountered, was the recommendation of an appropriate GPS reference book to the course attendees. Giving the fact that the majority of the GPS users are faced with a very tight time, it was necessary that the selected GPS book be complete and easy-to-read. Such a book did not exist.

Initially, I developed the vugraphs, which I used in the delivery of the short GPS courses. I then modified the vugraphs several times to accommodate not only the various types of GPS users but also my undergraduate students at both the University of New Brunswick and Ryerson University. The modified vugraphs were then used as the basis for this GPS book. I tried to address all aspects of GPS in a simple manner, avoiding any mathematics. The book also addresses more recent issues such as the modernization of GPS and the proposed European satellite navigation system known as Galileo. As well, the book emphasizes GPS applications, which will benefit not only the GPS users but also the GPS marketing and sales personnel.

Chapter 1 of the book introduces the GPS system and its components. Chapter 2 examines the GPS signal structure, the GPS modernization, and the key types of the GPS measurements. An in-depth discussion of the errors and biases that affect the GPS measurements, along with suggestions on how to overcome them, is presented in Chapter 3. Datums, coordinate systems, and map projections are discussed in a simple manner in Chapter 4, offering a clear understanding of this widely misunderstood area. Chapters 5 and 6 address the various modes of GPS positioning and the issue of the ambiguity resolution of the carrier-phase measurements. The various GPS services available on the market and the standard formats used for the various types of GPS data are presented in Chapters 7 and 8. Chapter 9 focuses on the integration of the GPS with other systems. The GPS applications in the various fields are given in Chapter 10. The book ends with Chapter 11, which covers the other satellite navigation systems developed or proposed in different parts of the world.

Acknowledgments

I would like to extend my appreciation to Dr. Alfred Kleusberg, Dr. Naser El-Sheimy, and Dr. David Wells for reviewing and/or commenting on the earlier version of the manuscript.

1

Introduction to GPS

The Global Positioning System (GPS) is a satellite-based navigation system that was developed by the U.S. Department of Defense (DoD) in the early 1970s. Initially, GPS was developed as a military system to fulfill U.S. military needs. However, it was later made available to civilians, and is now a dual-use system that can be accessed by both military and civilian users [1].

GPS provides continuous positioning and timing information, anywhere in the world under any weather conditions. Because it serves an unlimited number of users as well as being used for security reasons, GPS is a one-way-ranging (passive) system [2]. That is, users can only receive the satellite signals. This chapter introduces the GPS system, its components, and its basic idea.

1.1 Overview of GPS

GPS consists, nominally, of a constellation of 24 operational satellites. This constellation, known as the initial operational capability (IOC), was completed in July 1993. The official IOC announcement, however, was made on December 8, 1993 [3]. To ensure continuous worldwide coverage, GPS satellites are arranged so that four satellites are placed in each of six orbital planes (Figure 1.1). With this constellation geometry, four to ten GPS satellites will be visible anywhere in the world, if an elevation angle of 10° is considered. As discussed later, only four satellites are needed to provide the positioning, or location, information.

GPS satellite orbits are nearly circular (an elliptical shape with a maximum eccentricity is about 0.01), with an inclination of about 55° to the equator. The semimajor axis of a GPS orbit is about 26,560 km (i.e., the satellite altitude of about 20,200 km above the Earth's surface) [4]. The corresponding GPS orbital period is about 12 sidereal hours (~11 hours, 58 minutes). The GPS system was officially declared to have achieved full operational capability (FOC) on July 17, 1995, ensuring the availability of at least 24 operational, nonexperimental, GPS satellites. In fact, as shown in Section 1.4, since GPS achieved its FOC, the number of satellites in the GPS constellation has always been more than 24 operational satellites.

1.2 GPS segments

GPS consists of three segments: the space segment, the control segment, and the user segment (Figure 1.2) [5]. The space segment consists of the 24-satellite constellation introduced in the previous section. Each GPS satellite transmits a signal, which has a number of components: two sine waves (also known as carrier frequencies), two digital codes, and a navigation message. The codes and the navigation message are added to the carriers as binary biphase modulations [5]. The carriers and the codes are used mainly to determine the distance from the user's receiver to the GPS

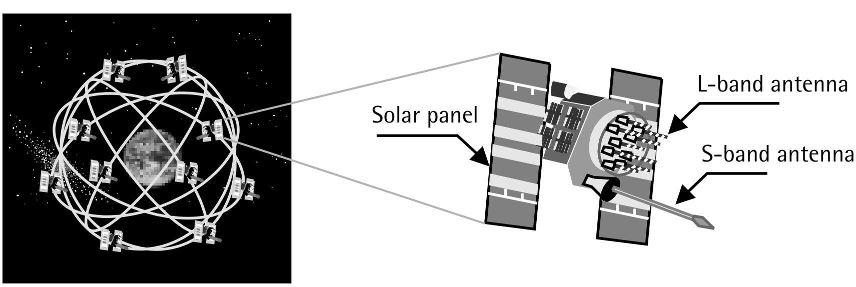

Figure 1.1 GPS constellation.

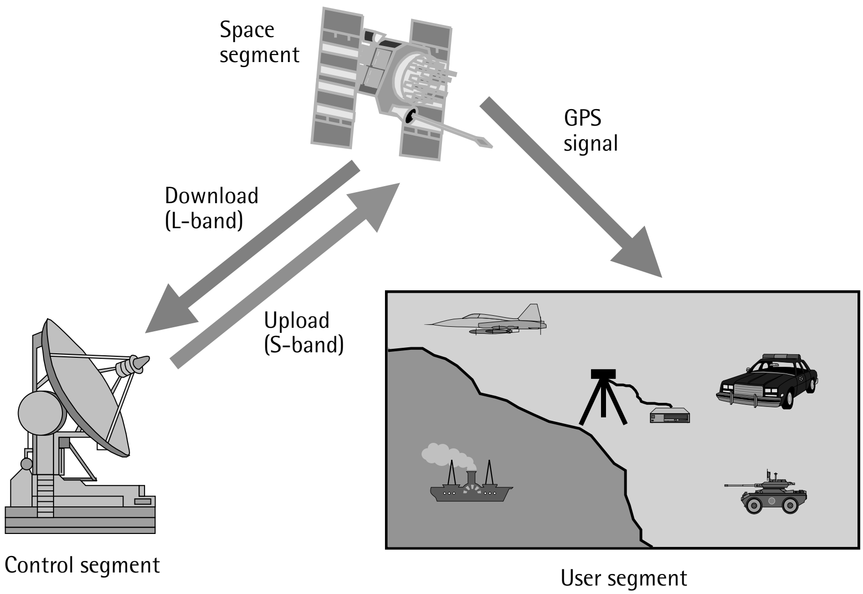

Figure 1.2 GPS segments.

satellites. The navigation message contains, along with other information, the coordinates (the location) of the satellites as a function of time. The transmitted signals are controlled by highly accurate atomic clocks onboard the satellites. More about the GPS signal is given in Chapter 2.

The control segment of the GPS system consists of a worldwide network of tracking stations, with a master control station (MCS) located in the United States at Colorado Springs, Colorado. The primary task of the operational control segment is tracking the GPS satellites in order to determine and predict satellite locations, system integrity, behavior of the satellite atomic clocks, atmospheric data, the satellite almanac, and other considerations. This information is then packed and uploaded into the GPS satellites through the S-band link.

The user segment includes all military and civilian users. With a GPS receiver connected to a GPS antenna, a user can receive the GPS signals, which can be used to determine his or her position anywhere in the world. GPS is currently available to all users worldwide at no direct charge.

1.3 GPS satellite generations

GPS satellite constellation buildup started with a series of 11 satellites known as Block I satellites (Figure 1.3). The first satellite in this series (and in the GPS system) was launched on February 22, 1978; the last was launched on October 9, 1985. Block I satellites were built mainly for experimental purposes. The inclination angle of the orbital planes of these satellites, with respect to the equator, was 63°, which was modified in the following satellite generations [6]. Although the design lifetime of Block I satellites was 4.5 years, some remained in service for more than 10 years. The last Block I satellite was taken out of service on November 18, 1995.

The second generation of the GPS satellites is known as Block II/IIA satellites (Figure 1.3). Block IIA is an advanced version of Block II, with an increase in the navigation message data storage capability from 14 days for Block II to 180 days for Block IIA. This means that Block II and Block IIA satellites can function continuously, without ground support, for periods of 14 and 180 days, respectively. A total of 28 Block II/IIA satellites were launched during the period from February 1989 to November 1997. Of these, 23 are currently in service. Unlike Block I, the orbital plane of Block II/IIA satellites are inclined by 55° with respect to the equator. The design lifetime of a Block II/IIA satellite is 7.5 years, which was exceeded by most Block II/IIA satellites. To ensure national security, some security features, known as selective availability (SA) and antispoofing, were added to Block II/IIA satellites [3, 6].

A new generation of GPS satellites, known as Block IIR, is currently being launched (Figure 1.3). These replenishment satellites will be backward compatible with Block II/IIA, which means that the changes are transparent to the users. Block IIR consists of 21 satellites with a design life of 10 years. In addition to the expected higher accuracy, Block IIR satellites have the capability of operating autonomously for at least 180 days without ground corrections or accuracy degradation. The autonomous navigation capability of this satellite generation is achieved in part through mutual satellite ranging capabilities. In addition, predicted ephemeris and clock data for a period of 210 days are uploaded by the ground control segment to support the autonomous navigation. More features will be added to the last 12 Block IIR satellites under the GPS modernization program, which will be launched at the beginning of 2003 [7]. As of July 2001, six Block IIR satellites have been successfully launched.

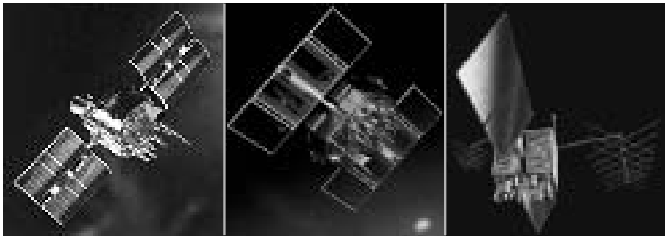

Figure 1.3 GPS satellite generations. (From http:\\www2.geod.hrcan.gc.ca/~craymer/gps.html.)

Block IIR will be followed by another system, called Block IIF (for "follow-on"), consisting of 33 satellites. The satellite life span will be 15 years. Block IIF satellites will have new capabilities under the GPS modernization program that will dramatically improve the autonomous GPS positioning accuracy (see Chapter 2 for details). The first Block IIF satellite is scheduled to be launched in 2005 or shortly after that date.

1.4 Current GPS satellite constellation

The current GPS constellation (as of July 2001) contains five Block II, 18 Block IIA, and six Block IIR satellites (see Table 1.1). This makes the total number of GPS satellites in the constellation to be 29, which exceeds the nominal 24-satellite constellation by five satellites [8]. All Block I satellites are no longer operational.

The GPS satellites are placed in six orbital planes, which are labeled A through F. Since more satellites are currently available than the nominal 24-satellite constellation, an orbital plane may contain four or five satellites. As shown in Table 1.1, all of the orbital planes have five satellites, except for orbital plane C, which has only four. The satellites can be identified by various systems. The most popular identification systems within the GPS user community are the space vehicle number (SVN) and the pseudorandom noise (PRN); the PRN number will be defined later. Block II/IIA satellites are equipped with four onboard atomic clocks: two cesium (Cs) and two rubidium (Rb). The cesium clock is used as the primary timing source to control the GPS signal. Block IIR satellites, however, use

Table 1.1 GPS Satellite Constellation as of July 2001

Sequence	SVN	PRN	Orbital Plane	Clock	Sequence	SVN	PRN	Orbital Plane	Clock
II-2	13	2	B-3	Cs	II-21	39	9	A-1	Cs
II-4	19	19	A-5	Cs	II-22	35	5	B-4	Cs
II-5	17	17	D-3	Cs	II-23	34	4	D-4	Rb
II-8	21	21	E-2	Cs	II-24	36	6	C-1	Cs
II-9	15	15	D-5	Cs	II-25	33	3	C-2	Cs
II-10	23	23	E-5	Cs	II-26	40	10	E-3	Cs
II-11	24	24	D-1	Cs	II-27	30	30	B-2	Cs
II-12	25	25	A-2	Cs	II-28	38	8	A-3	Rb
II-14	26	26	F-2	Rb	IIR-2	43	13	F-3	Rb
II-15	27	27	A-4	Cs	IIR-3	46	11	D-2	Rb
II-16	32	1	F-4	Cs	IIR-4	51	20	E-1	Rb
II-17	29	29	F-5	Rb	IIR-5	44	28	B-5	Rb
II-18	22	22	B-1	Rb	IIR-6	41	14	F-1	Rb
II-19	31	31	C-3	Cs	IIR-7	54	18	E-4	Rb
II-20	37	7	C-4	Rb					

rubidium clocks only. It should be pointed out that two satellites, PRN05 and PRN06, are equipped with corner cube reflectors to be tracked by laser ranging (Table 1.1).

1.5 Control sites

The control segment of GPS consists of a master control station (MCS), a worldwide network of monitor stations, and ground control stations (Figure 1.4). The MCS, located near Colorado Springs, Colorado, is the central processing facility of the control segment and is manned at all times [9].

There are five monitor stations, located in Colorado Springs (with the MCS), Hawaii, Kwajalein, Diego Garcia, and Ascension Island. The positions (or coordinates) of these monitor stations are known very precisely.

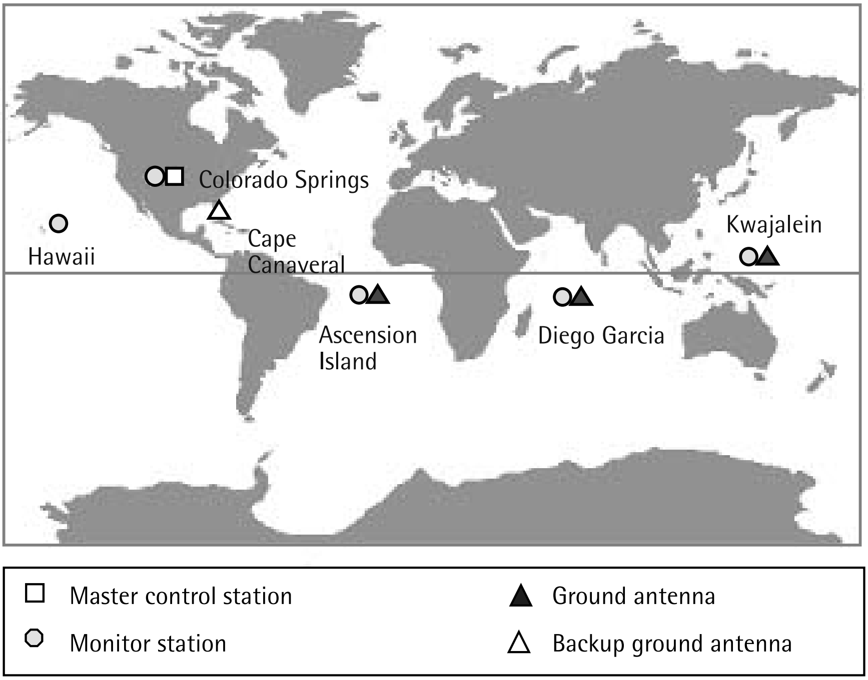

Figure 1.4 GPS control sites.

Each monitor station is equipped with high-quality GPS receivers and a cesium oscillator for the purpose of continuous tracking of all the GPS satellites in view. Three of the monitor stations (Kwajalein, Diego Garcia, and Ascension Island) are also equipped with ground antennas for uploading the information to the GPS satellites. All of the monitor stations and the ground control stations are unmanned and operated remotely from the MCS.

The GPS observations collected at the monitor stations are transmitted to the MCS for processing. The outcome of the processing is predicted satellite navigation data that includes, along with other information, the satellite positions as a function of time, the satellite clock parameters, atmospheric data, satellite almanac, and others. This fresh navigation data is sent to one of the ground control stations to upload it to the GPS satellites through the S-band link.

Monitoring the GPS system integrity is also one of the tasks of the MCS. The status of a satellite is set to unhealthy condition by the MCS during satellite maintenance or outages. This satellite health condition appears as a part of the satellite navigation message on a near real-time basis. Scheduled satellite maintenance or outage is reported in a message called Notice Advisory to Navstar Users (NANU), which is available to the public through, for example, the U.S. Coast Guard Navigation Center [8].

1.6 GPS: The basic idea

The idea behind GPS is rather simple. If the distances from a point on the Earth (a GPS receiver) to three GPS satellites are known along with the satellite locations, then the location of the point (or receiver) can be determined by simply applying the well-known concept of resection [10]. That is all! But how can we get the distances to the satellites as well as the satellite locations?

As mentioned before, each GPS satellite continuously transmits a microwave radio signal composed of two carriers, two codes, and a navigation message. When a GPS receiver is switched on, it will pick up the GPS signal through the receiver antenna. Once the receiver acquires the GPS signal, it will process it using its built-in software. The partial outcome of the signal processing consists of the distances to the GPS satellites through the digital codes (known as the pseudoranges) and the satellite coordinates through the navigation message.

Theoretically, only three distances to three simultaneously tracked satellites are needed. In this case, the receiver would be located at the intersection of three spheres; each has a radius of one receiver-satellite distance and is centered on that particular satellite (Figure 1.5). From the practical point of view, however, a fourth satellite is needed to account for the receiver clock offset [6]. More details on this are given in Chapter 5.

The accuracy obtained with the method described earlier was until recently limited to 100m for the horizontal component, 156m for the vertical component, and 340 ns for the time component, all at the 95% probability level. This low accuracy level was due to the effect of the so-called selective availability, a technique used to intentionally degrade the autonomous real-time positioning accuracy to unauthorized users [3]. With the recent presidential decision of terminating the selective availability, the obtained horizontal accuracy is expected to improve to about 22m (95%

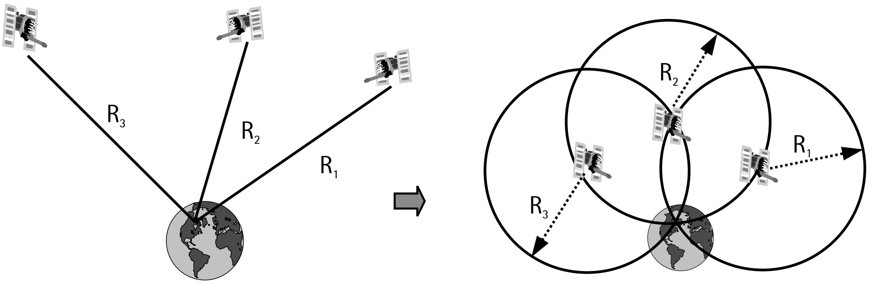

Figure 1.5 Basic idea of GPS positioning.

probability level) [7, 11]. To further improve the GPS positioning accuracy, the so-called differential method, which employs two receivers simultaneously tracking the same GPS satellites, is used. In this case, positioning accuracy level of the order of a subcentimeter to a few meters can be obtained.

Other uses of GPS include the determination of the user's velocity, which could be determined by several methods. The most widely used method is based on estimating the Doppler frequency of the received GPS signal [6]. It is known that the Doppler shift occurs as a result of the relative satellite-receiver motion. GPS may also be used in determining the attitude of a rigid body, such as an aircraft or a marine vessel. The word "attitude" means the orientation, or the direction, of the rigid body, which can be described by the three rotation angles of the three axes of the rigid body with respect to a reference system. Attitude is determined by equipping the body with a minimum of three GPS receivers (or one special receiver) connected to three antennas, which are arranged in a nonstraight line [12]. Data collected at the receivers are then processed to obtain the attitude of the rigid body.

1.7 GPS positioning service

As stated earlier, GPS was originally developed as a military system, but was later made available to civilians as well. However, to keep the military advantage, the U.S. DoD provides two levels of GPS positioning and timing services: the Precise Positioning Service (PPS) and the Standard Positioning Service (SPS) [3].

PPS is the most precise autonomous positioning and timing service. It uses one of the transmitted GPS codes, known as P(Y)-code, which is accessible by authorized users only. These users include U.S. military forces. The expected positioning accuracy provided by the PPS is 16m for the horizontal component and 23m for the vertical component (95% probability level).

SPS, however, is less precise than PPS. It uses the second transmitted GPS code, known as the C/A-code, which is available free of charge to all users worldwide, authorized and unauthorized. Originally, SPS provided positioning accuracy of the order of 100m for the horizontal component and 156m for the vertical component (95% probability level). This was achieved under the effect of selective availability. With the recent presidential decision of discontinuing the SA, the SPS autonomous positioning accuracy is presently at a comparable level to that of the PPS.

1.8 Why use GPS?

GPS has revolutionized the surveying and navigation fields since its early stages of development. Although GPS was originally designed as a military system, its civil applications have grown much faster. As for the future, it is said that the number of GPS applications will be limited only to one's imagination.

On the surveying side, GPS has replaced the conventional methods in many applications. GPS positioning has been found to be a cost-effective process, in which at least 50% cost reduction can be obtained whenever it is possible to use the so-called real-time kinematic (RTK) GPS, as compared with conventional techniques [13]. In terms of productivity and time saving, GPS could provide more than 75% timesaving whenever it is possible to use the RTK GPS method (more about RTK capabilities and limitations is given in Chapter 5) [12]. The fact that GPS does not require intervisibility between stations has also made it more attractive to surveyors over the conventional methods. For those situations in which the GPS signal is obstructed, such as in urban canyons, GPS has been successfully integrated with other conventional equipment.

GPS has numerous applications in land, marine, and air navigation. Vehicle tracking and navigation are rapidly growing applications. It is expected that the majority of GPS users will be in vehicle navigation.

Future uses of GPS will include automatic machine guidance and control, where hazardous areas can be mapped efficiently and safely using remotely controlled vehicles. The recent U.S. decision to modernize GPS and to terminate the selective availability will undoubtedly open the door for a number of other applications yet to be developed [10].

References

[1] FRP, *U.S. Federal Radionavigation Plan*, 1999.

[2] Langley, R. B., "Why Is the GPS Signal So Complex?" *GPS World*, Vol. 1, No. 3, May/June 1990, pp. 56–59.

[3] Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*, 3rd ed., New York: Springer-Verlag, 1994.

[4] Langley, R. B., "The Orbits of GPS Satellites," *GPS World*, Vol. 2, No. 3, March 1991, pp. 50–53.

[5] Wells, D. E., et al., *Guide to GPS Positioning*, Fredericton, New Brunswick: Canadian GPS Associates, 1987.

[6] Kaplan, E., *Understanding GPS: Principles and Applications,* Norwood, MA: Artech House, 1990.

[7] Shaw, M., K. Sandhoo, and D. Turner, "Modernization of the Global Positioning System," *GPS World*, Vol. 11, No. 9, September 2000, pp. 36–44.

[8] U.S. Coast Guard Navigation Center, "GPS Status," September 17, 2001, http://www.navcen.uscg.gov/gps/.

[9] Leick, A., *GPS Satellite Surveying*, 2nd ed., New York: Wiley, 1995.

[10] Langley, R. B., "The Mathematics of GPS," *GPS World*, Vol. 2, No. 7, July/August 1991, pp. 45–50.

[11] Conley, R., "Life After Selective Availability," *U.S. Institute of Navigation Newsletter*, Vol. 10, No. 1, Spring 2000, pp. 3–4.

[12] Kleusberg, A., "Mathematics of Attitude Determination with GPS," *GPS World*, Vol. 6, No. 9, September 1995, pp. 72–78.

[13] Berg, R. E., "Evaluation of Real-Time Kinematic GPS Versus Total Stations for Highway Engineering Surveys," *8th Intl. Conf. Geomatics: Geomatics in the Era of RADARSAT*, Ottawa, Canada, May 24–30, 1996, CD-ROM.

2

GPS Details

Positioning, or finding the user's location, with GPS requires some understanding of the GPS signal structure and how the measurements can be made. Likewise, as the GPS signal is received through a GPS receiver, understanding the capabilities and limitations of the various types of GPS receivers is essential. Furthermore, the GPS measurements, like all measurable quantities, contain errors and biases, which can be removed or reduced by combining the various GPS observables. This chapter discusses these issues in detail.

2.1 GPS signal structure

As mentioned in Chapter 1, each GPS satellite transmits a microwave radio signal composed of two carrier frequencies (or sine waves) modulated by two digital codes and a navigation message (see Figure 2.1). The two carrier frequencies are generated at 1,575.42 MHz (referred to as the L1 carrier) and 1,227.60 MHz (referred to as the L2 carrier). The corresponding carrier wavelengths are approximately 19 cm and 24.4 cm, respectively, which result from the relation between the carrier frequency and the speed of light in space [1, 2]. The availability of the two carrier frequencies allows for correcting a major GPS error, known as the ionospheric delay (see Chapter 3 for details). All of the GPS satellites transmit the same L1 and L2 carrier frequencies. The code modulation, however, is different for each satellite, which significantly minimizes the signal interference.

The two GPS codes are called coarse acquisition (or C/A-code) and precision (or P-code). Each code consists of a stream of binary digits, zeros and ones, known as bits or chips. The codes are commonly known as PRN codes because they look like random signals (i.e., they are noise-like signals). But in reality, the codes are generated using a mathematical algorithm. Presently, the C/A-code is modulated onto the L1 carrier only, while the P-code is modulated onto both the L1 and the L2 carriers. This modulation is called biphase modulation, because the carrier phase is shifted by 180° when the code value changes from zero to one or from one to zero [3].

The C/A-code is a stream of 1,023 binary digits (i.e., 1,023 zeros and ones) that repeats itself every millisecond. This means that the chipping rate of the C/A-code is 1.023 Mbps. In other words, the duration of one bit is approximately 1 ms, or equivalently 300m [4]. Each satellite is assigned a unique C/A-code, which enables the GPS receivers to identify which satellite is transmitting a particular code. The C/A-code range measurement is relatively less precise compared with that of the P-code. It is, however, less complex and is available to all users.

The P-code is a very long sequence of binary digits that repeats itself after 266 days [1]. It is also 10 times faster than the C/A-code (i.e., its rate is 10.23 Mbps). Multiplying the time it takes the P-code to repeat itself, 266 days, by its rate, 10.23 Mbps, tells us that the P-code is a stream of about 2.35×10^{14} chips! The 266-day-long code is divided into 38 segments; each is 1 week long. Of these, 32 segments are assigned to the various GPS satellites. That is, each satellite transmits a unique 1-week segment of the P-code, which is initialized every Saturday/Sunday midnight crossing. The remaining six segments are reserved for other uses. It is worth mentioning that a GPS satellite is usually identified by its unique 1-week segment of the P-code. For example, a GPS satellite with an ID of PRN 20 refers to a GPS satellite that is assigned the twentieth-week segment of the PRN P-code. The P-code is designed primarily for military purposes. It was available to all users until January 31, 1994 [1]. At that time, the P-code was encrypted by adding to it an unknown W-code. The resulting encrypted code is called

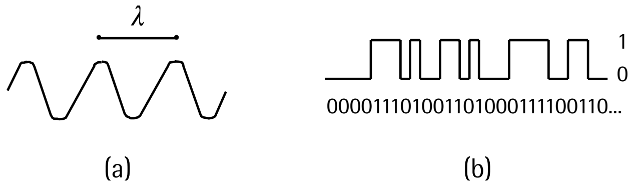

Figure 2.1 (a) A sinusoidal wave; and (b) a digital code.

the Y-code, which has the same chipping rate as the P-code. This encryption is known as the antispoofing (AS).

The GPS navigation message is a data stream added to both the L1 and the L2 carriers as binary biphase modulation at a low rate of 50 kbps. It consists of 25 frames of 1,500 bits each, or 37,500 bits in total. This means that the transmission of the complete navigation message takes 750 seconds, or 12.5 minutes. The navigation message contains, along with other information, the coordinates of the GPS satellites as a function of time, the satellite health status, the satellite clock correction, the satellite almanac, and atmospheric data. Each satellite transmits its own navigation message with information on the other satellites, such as the approximate location and health status [1].

2.2 GPS modernization

The current GPS signal structure was designed in the early 1970s, some 30 years ago [5]. In the next 30 years, GPS constellation is expected to have a combination of Block IIR satellites, currently being launched, and Block IIF and possibly Block III satellites. To meet the future requirements, the GPS decision makers have studied several options to adequately modify the signal structure and system architecture of the future GPS constellation. The modernization program aims, among other things, to provide signal redundancy and improve positioning accuracy, signal availability, and system integrity.

The modernization program will include the addition of a civil code (C/A-code) on the L2 frequency and two new military codes (M-codes) on both the L1 and the L2 frequencies [5]. These codes will be added to the last 12 Block IIR satellites, which will be launched at the beginning of 2003. The availability of two civil codes (i.e., C/A-code on both L1 and L2 frequencies) allows a user with a stand-alone GPS receiver to correct for the effect of the ionosphere (the upper layer of the atmosphere), which is a major error source (see Chapter 3 for details). With the termination of selective availability, it is expected that once a sufficient number of satellites with the new capabilities is available, the autonomous GPS horizontal accuracy will be about 8.5m (95% of the time) or better [5].

The addition of the C/A-code to L2, although it improves the autonomous GPS accuracy, was found to be insufficient for use in the civil aviation safety-of-life applications. This is mainly because of the potential interference from the ground radars that operate near the GPS L2 band. As such, to satisfy aviation user requirements, a third civil signal at 1,176.45 MHz (called L5) will be added to the first 12 Block IIF satellites along with the C/A-code on L2 and the M-code on L1 and L2, as part of the modernization program [5]. This third frequency will be robust and will have a higher power level. In addition, this new L5 signal will have wide broadcast bandwidth (a minimum of 20 MHz) and a higher chipping rate (10.23 MHz), which provide higher accuracy under noisy and multipath conditions. The new code will be longer than the current C/A-code, which reduces the system self-interference through the improvement of the auto- and cross-correlation properties. Finally, the broadcast navigation message of the new signal, although containing more or less the same data as the L1 and L2 channels, will have an entirely different, more efficient, structure. The first Block IIF satellite is scheduled to be launched in 2005 or shortly after that date. The addition of these capabilities will dramatically improve the autonomous GPS positioning accuracy. As well, the real-time kinematic (RTK) users, who require centimeter-level accuracy in real time, will be able to resolve the initial integer ambiguity parameters instantaneously. More about RTK positioning is given in Chapter 5.

The modernization of GPS will also include the studies for the next generation Block III satellites, which will carry GPS into 2030. Finally, the GPS ground control facilities will also be upgraded as a part of the GPS modernization program. With this upgrade, the expected standalone GPS horizontal accuracy will be 6m (95% of the time) or better [5].

2.3 Types of GPS receivers

In 1980, only one commercial GPS receiver was available on the market, at a price of several hundred thousand U.S. dollars [6]. This, however, has changed considerably as more than 500 different GPS receivers are available in today's market (see, for example, the January 2001 issue of *GPS World* magazine). The current receiver price varies from about $100 for the simple handheld units to about $15,000 for the sophisticated geodetic quality units. The price will continue to decline in the future as the receiver technology becomes more advanced. A GPS receiver requires an antenna attached to it, either internally or externally. The antenna receives the incoming satellite signal and then converts its energy into an electric current, which can be handled by the GPS receiver [6, 7].

Commercial GPS receivers may be divided into four types, according to their receiving capabilities. These are: single-frequency code receivers, single-frequency carrier-smoothed code receivers, single-frequency code and carrier receivers, and dual-frequency receivers. Single-frequency receivers access the L1 frequency only, while dual-frequency receivers access both the L1 and the L2 frequencies. Figure 2.2 shows examples of various types of GPS receivers. GPS receivers can also be categorized according to their number of tracking channels, which varies from 1 to 12 channels. A good GPS receiver would be multichannel, with each channel dedicated to continuously tracking a particular satellite. Presently, most GPS receivers have 9 to 12 independent (or parallel) channels. Features such as cost, ease of use, power consumption, size and weight, internal and/or external data-storage capabilities, interfacing capabilities, and multipath mitigation (i.e., type of correlator) are to be considered when selecting a GPS receiver.

The first receiver type, the single-frequency code receiver, measures the pseudoranges with the C/A-code only. No other measurements are available. It is the least expensive and the least accurate receiver type, and is mostly used for recreation purposes. The second receiver type, the single-frequency carrier-smoothed code receiver, also measures the pseudoranges with the C/A-code only. However, with this receiver type, the higher-resolution carrier frequency is used internally to improve the resolution of the code pseudorange, which results in high-precision pseudorange measurements. Single-frequency code and carrier receivers output the raw C/A-code pseudoranges, the L1 carrier-phase measurements, and the navigation message. In addition, this receiver type is capable of performing the functions of the other receiver types discussed above.

Dual-frequency receivers are the most sophisticated and most expensive receiver type. Before the activation of AS, dual-frequency receivers

Magellan handheld GPS receiver

Ashtech ZX geodetic quality GPS receiver

Figure 2.2 Examples of GPS receivers. (Courtesy of Magellan Corporation.)

were capable of outputting all of the GPS signal components (i.e., L1 and L2 carriers, C/A-code, P-code on both L1 and L2, and the navigation message). However, after the AS activation, the P-code was encrypted to Y-code. This means that the receiver cannot output either the P-code or the L2 carrier using the traditional signal-recovering technique. To overcome this problem, GPS receiver manufacturers invented a number of techniques that do not require information of the Y-code. At the present time, most receivers use two techniques known as the Z-tracking and the cross-correlation techniques. Both techniques recover the full L2 carrier, but at a degraded signal strength. The amount of signal strength degradation is higher in the cross-correlation techniques compared with the Z-tracking technique.

2.4 Time systems

Time plays a very important role in positioning with GPS. As explained in Chapter 1, the GPS signal is controlled by accurate timing devices, the atomic satellite clocks [8]. In addition, measuring the ranges (distances) from the receiver to the satellites is based on both the receiver and the satellite clocks. GPS is also a timing system, that is, it can be used for time synchronization.

A number of time systems are used worldwide for various purposes [1]. Of these, the Coordinated Universal Time (UTC) and the GPS Time are the most important to GPS users. UTC is an atomic time scale based on the International Atomic Time (TAI). TAI is a uniform time scale, which is computed based on independent time scales generated by atomic clocks located at various timing laboratories throughout the world. In surveying and navigation, however, a time system with relation to the rotation of the Earth, not the atomic time, is desired. This is achieved by occasionally adjusting the UTC time scale by 1-second increments, known as leap seconds, to keep it within 0.9 second of another time scale called the Universal Time 1 (UT1) [8, 9], where UT1 is a universal time that gives a measure of the rotation of the Earth. Leap seconds are introduced occasionally, on either June 30 or December 31. As of July 2001, the last leap second was introduced on January 1, 1999, which made the difference between TAI and UTC time scales to be exactly 32 seconds (TAI is ahead of UTC). Information about the leap seconds can be found at the U.S. Naval Observatory Web site, http://maia.usno.navy.mil.

GPS Time is the time scale used for referencing, or time tagging, the GPS signals. It is computed based on the time scales generated by the atomic clocks at the monitor stations and onboard GPS satellites. There are no leap seconds introduced into GPS Time, which means that GPS Time is a continuous time scale. GPS Time scale was set equal to that of the UTC on January 6, 1980 [8]. However, due to the leap seconds introduced into the UTC time scale, GPS Time moved ahead of the UTC by 13 seconds on January 1, 1999. The difference between GPS and UTC time scales is given in the GPS navigation message. It is worth mentioning that, as shown in Chapter 3, both GPS satellite and receiver clocks are offset from the GPS Time, as a result of satellite and receiver clock errors.

2.5 Pseudorange measurements

The pseudorange is a measure of the range, or distance, between the GPS receiver and the GPS satellite (more precisely, it is the distance between the GPS receiver's antenna and the GPS satellite's antenna). As stated before, the ranges from the receiver to the satellites are needed for the position computation. Either the P-code or the C/A-code can be used for measuring the pseudorange.

The procedure of the GPS range determination, or pseudoranging, can be described as follows. Let us assume for a moment that both the satellite and the receiver clocks, which control the signal generation, are perfectly synchronized with each other. When the PRN code is transmitted from the satellite, the receiver generates an exact replica of that code [3]. After some time, equivalent to the signal travel time in space, the transmitted code will be picked up by the receiver. By comparing the transmitted code and its replica, the receiver can compute the signal travel time. Multiplying the travel time by the speed of light (299,729,458 m/s) gives the range between the satellite and the receiver. Figure 2.3 explains the pseudorange measurements.

Unfortunately, the assumption that the receiver and satellite clocks are synchronized is not exactly true. In fact, the measured range is contaminated, along with other errors and biases, by the synchronization error between the satellite and receiver clocks. For this reason, this quantity is referred to as the pseudorange, not the range [4].

GPS was designed so that the range determined by the civilian C/A-code would be less precise than that of military P-code. This is based on the fact that the resolution of the C/A-code, 300m, is 10 times lower than the P-code. Surprisingly, due to the improvements in the receiver technology, the obtained accuracy was almost the same from both codes [4].

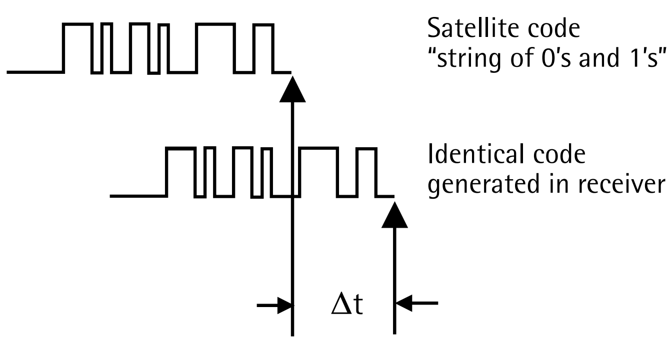

Figure 2.3 Pseudorange measurements.

2.6 Carrier-phase measurements

Another way of measuring the ranges to the satellites can be obtained through the carrier phases. The range would simply be the sum of the total number of full carrier cycles plus fractional cycles at the receiver and the satellite, multiplied by the carrier wavelength (see Figure 2.4). The ranges determined with the carriers are far more accurate than those obtained with the codes (i.e., the pseudoranges) [4]. This is due to the fact that the wavelength (or resolution) of the carrier phase, 19 cm in the case of L1 frequency, is much smaller than those of the codes.

There is, however, one problem. The carriers are just pure sinusoidal waves, which means that all cycles look the same. Therefore, a GPS receiver has no means to differentiate one cycle from another [4]. In other words, the receiver, when it is switched on, cannot determine the total number of the complete cycles between the satellite and the receiver. It can only measure a fraction of a cycle very accurately (less than 2 mm), while the initial

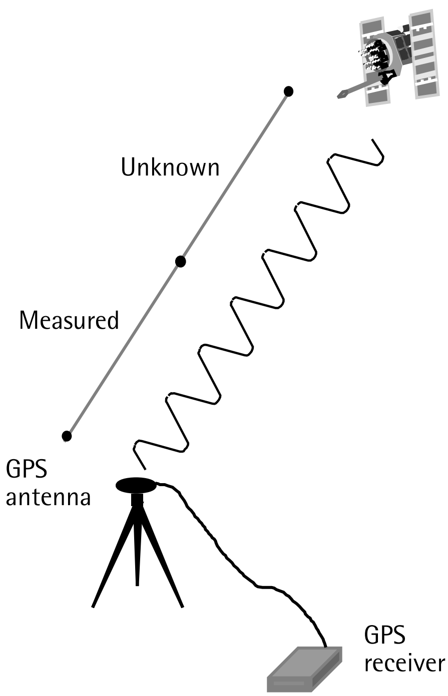

Figure 2.4 Carrier-phase measurements.

number of complete cycles remains unknown, or ambiguous. This is, therefore, commonly known as the initial cycle ambiguity, or the ambiguity bias. Fortunately, the receiver has the capability to keep track of the phase changes after being switched on. This means that the initial cycle ambiguity remains unchanged over time, as long as no signal loss (or cycle slips) occurs [3].

It is clear that if the initial cycle ambiguity parameters are resolved, accurate range measurements can be obtained, which lead to accurate position determination. This high accuracy positioning can be achieved through the so-called relative positioning techniques, either in real time or in the postprocessing mode. Unfortunately, this requires two GPS receivers simultaneously tracking the same satellites in view. More about the various positioning techniques and the ways of resolving the ambiguity parameters is given in Chapters 5 and 6, respectively.

2.7 Cycle slips

A cycle slip is defined as a discontinuity or a jump in the GPS carrier-phase measurements, by an integer number of cycles, caused by temporary signal loss [1]. Signal loss is caused by obstruction of the GPS satellite signal due to buildings, bridges, trees, and other objects (Figure 2.5). This is mainly because the GPS signal is a weak and noisy signal. Radio interference, severe ionospheric disturbance, and high receiver dynamics can also cause signal loss. Cycle slips could occur due to a receiver malfunction [1].

Cycle slips may occur briefly or may remain for several minutes or even more. Cycle slips could affect one or more satellite signals. The size of a cycle slip could be as small as one cycle or as large as millions of cycles. Cycle slips must be identified and corrected to avoid large errors in the computed coordinates. This can be done using several methods. Examining the so-called triple difference observable, which is formed by combining the GPS observables in a certain way (see Section 2.8), is the most popular in practice. A cycle slip will only affect one triple difference and therefore will appear as a spike in the triple difference data series. In some extreme cases, such as severe ionospheric activities, it might be difficult to correctly detect and repair cycle slips using triple difference observable [1, 3]. Visual inspection of the adjustment residuals might be useful to locate any remaining cycle slip.

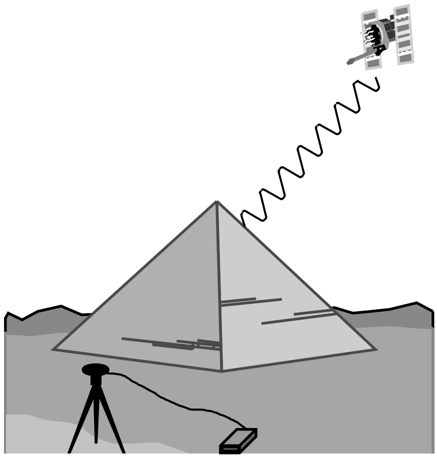

Figure 2.5 GPS cycle slips.

As shown in Chapter 3, a zero baseline test is used to detect cycle slips due to receiver malfunction. In this test, two receivers are connected to one antenna through a signal splitter. Cycle slips can be detected by examining the adjustment residuals [3].

2.8 Linear combinations of GPS observables

GPS measurements are corrupted by a number of errors and biases (discussed in detail in Chapter 3), which are difficult to model fully. The unmodeled errors and biases limit the positioning accuracy of the standalone GPS receiver. Fortunately, GPS receivers in close proximity will share to a high degree of similarity the same errors and biases. As such, for those receivers, a major part of the GPS error budget can simply be removed by combining their GPS observables.

In principle, there are three groups of GPS errors and biases: satellite-related, receiver-related, and atmospheric errors and biases [3]. The measurements of two GPS receivers simultaneously tracking a particular satellite contain more or less the same satellite-related errors and atmospheric errors. The shorter the separation between the two receivers, the more similar the errors and biases. Therefore, if we take the difference between the measurements collected at these two GPS receivers, the satellite-related errors and the atmospheric errors will be reduced significantly. In fact, as shown in Chapter 3, the satellite clock error is effectively removed with this linear combination. This linear combination is known as between-receiver single difference (Figure 2.6).

Similarly, the two measurements of a single receiver tracking two satellites contain the same receiver clock errors. Therefore, taking the difference between these two measurements removes the receiver clock errors. This difference is known as between-satellite single difference (Figure 2.6).

When two receivers track two satellites simultaneously, two between-receiver single difference observables could be formed. Subtracting these two single difference observables from each other generates the so-called double difference [3]. This linear combination removes the satellite and receiver clock errors. The other errors are greatly reduced. In addition, this observable preserves the integer nature of the ambiguity parameters. It is therefore used for precise carrier-phase-based GPS positioning.

Another important linear combination in known as the "triple difference," which results from differencing two double-difference observables over two epochs of time [3]. As explained in the previous section, the ambiguity parameters remain constant over time, as long as there are no cycle slips. As such, when forming the triple difference, the constant ambiguity parameters disappear. If, however, there is a cycle slip in the data, it will

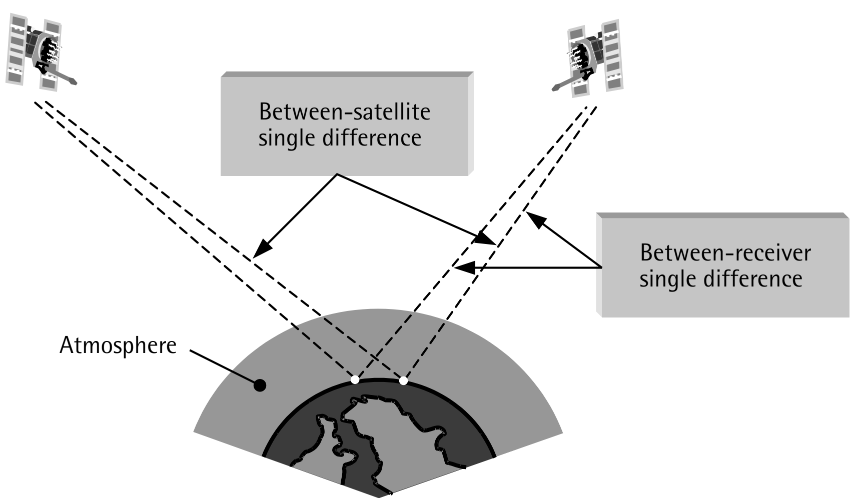

Figure 2.6 Some GPS linear combinations.

affect one triple-difference observable only, and therefore will appear as a spike in the triple-difference data series. It is for this reason that the triple-difference linear combination is used for detecting the cycle slips.

All of these linear combinations can be formed with a single frequency data, whether it is the carrier phase or the pseudorange observables. If dual-frequency data is available, other useful linear combinations could be formed. One such linear combination is known as the ionosphere-free linear combination. As shown in Chapter 3, ionospheric delay is inversely proportional to the square of the carrier frequency. Based on this characteristic, the ionosphere-free observable combines the L1 and L2 measurements to essentially eliminate the ionospheric effect. The L1 and L2 carrier-phase measurements could also be combined to form the so-called wide-lane observable, an artificial signal with an effective wavelength of about 86 cm. This long wavelength helps in resolving the integer ambiguity parameters [1].

References

[1] Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice,* 3rd ed., New York: Springer-Verlag, 1994.

[2] Langley, R. B., "Why Is the GPS Signal So Complex?" *GPS World,* Vol. 1, No. 3, May/June 1990, pp. 56–59.

[3] Wells, D. E., et al., *Guide to GPS Positioning,* Fredericton, New Brunswick: Canadian GPS Associates, 1987.

[4] Langley, R. B., "The GPS Observables," *GPS World*, Vol. 4, No. 4, April 1993, pp. 52–59.

[5] Shaw, M., K. Sandhoo, and D. Turner, "Modernization of the Global Positioning System," *GPS World*, Vol. 11, No. 9, September 2000, pp. 36–44.

[6] Langley, R. B., "The GPS Receiver: An Introduction," *GPS World*, Vol. 2, No. 1, January 1991, pp. 50–53.

[7] Langley, R. B., "Smaller and Smaller: The Evolution of the GPS Receiver," *GPS World*, Vol. 11, No. 4, April 2000, pp. 54–58.

[8] Langley, R. B., "Time, Clocks, and GPS," *GPS World*, Vol. 2, No. 10, November/December 1991, pp. 38–42.

[9] McCarthy, D. D., and W. J. Klepczynski, "GPS and Leap Seconds: Time to Change," *GPS World*, Vol. 10, No. 11, November 1999, pp. 50–57.

3

GPS Errors and Biases

GPS pseudorange and carrier-phase measurements are both affected by several types of random errors and biases (systematic errors). These errors may be classified as those originating at the satellites, those originating at the receiver, and those that are due to signal propagation (atmospheric refraction) [1]. Figure 3.1 shows the various errors and biases.

The errors originating at the satellites include ephemeris, or orbital, errors, satellite clock errors, and the effect of selective availability. The latter was intentionally implemented by the U.S. DoD to degrade the autonomous GPS accuracy for security reasons. It was, however, terminated at midnight (eastern daylight time) on May 1, 2000 [2]. The errors originating at the receiver include receiver clock errors, multipath error, receiver noise, and antenna phase center variations. The signal propagation errors include the delays of the GPS signal as it passes through the ionospheric and tropospheric layers of the atmosphere. In fact, it is only in a vacuum (free space) that the GPS signal travels, or propagates, at the speed of light.

In addition to the effect of these errors, the accuracy of the computed GPS position is also affected by the geometric locations of the GPS satellites as seen by the receiver. The more spread out the satellites are in the sky, the better the obtained accuracy (Figure 3.1).

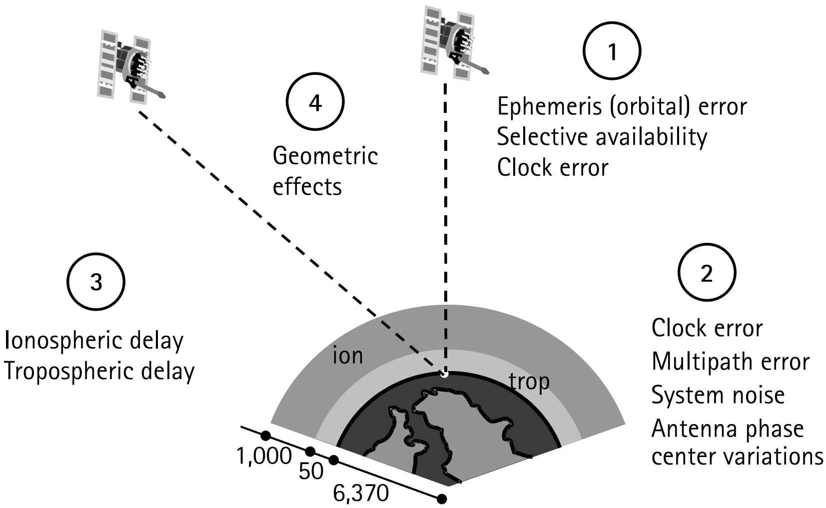

Figure 3.1 GPS errors and biases.

As shown in Chapter 2, some of these errors and biases can be eliminated or reduced through appropriate combinations of the GPS observables. For example, combining L1 and L2 observables removes, to a high degree of accuracy, the effect of the ionosphere. Mathematical modeling of these errors and biases is also possible. In this chapter, the main GPS error sources are introduced and the ways of treating them are discussed.

3.1 GPS ephemeris errors

Satellite positions as a function of time, which are included in the broadcast satellite navigation message, are predicted from previous GPS observations at the ground control stations. Typically, overlapping 4-hour GPS data spans are used by the operational control system to predict fresh satellite orbital elements for each 1-hour period. As might be expected, modeling the forces acting on the GPS satellites will not in general be perfect, which causes some errors in the estimated satellite positions, known as ephemeris errors. Nominally, an ephemeris error is usually in the order of 2m to 5m, and can reach up to 50m under selective availability [3]. According to [2], the range error due to the combined effect of the ephemeris and the satellite clock errors is of the order of 2.3m [1σ-level; σ is the standard deviation (see Appendix B)].

An ephemeris error for a particular satellite is identical to all GPS users worldwide [4]. However, as different users see the same satellite at different view angles, the effect of the ephemeris error on the range measurement, and consequently on the computed position, is different. This means that combining (differencing) the measurements of two receivers simultaneously tracking a particular satellite cannot totally remove the ephemeris error. Users of short separations, however, will have an almost identical range error due to the ephemeris error, which can essentially be removed through differencing the observations. For relative positioning (see Chapter 5), the following rule of thumb gives a rough estimate of the effect of the ephemeris error on the baseline solution: *the baseline error / the baseline length = the satellite position error / the range satellite* [5]. This means that if the satellite position error is 5m and the baseline length is 10 km, then the expected baseline line error due to ephemeris error is approximately 2.5 mm.

Some applications, such as studies of the crustal dynamics of the earth, require more precise ephemeris data than the broadcast ephemeris. To support these applications, several institutions [e.g., the International GPS Service for Geodynamics (IGS), the U.S. National Geodetic Survey (NGS), and Geomatics Canada] have developed postmission precise orbital service. Precise ephemeris data is based on GPS data collected at a global GPS network coordinated by the IGS. At the present time, precise ephemeris data is available to users with some delay, which varies from 12 hours for the IGS ultra rapid orbit to about 12 days for the most precise IGS precise orbit. The corresponding accuracies for the two precise orbits are in the order of a few decimeters to 1 decimeter, respectively. Users can download the precise ephemeris data free of charge from the IGS center, at ftp://igscb.jpl.nasa.gov/igscb/product/.

3.2 Selective availability

GPS was originally designed so that real-time autonomous positioning and navigation with the civilian C/A code receivers would be less precise than military P-code receivers. Surprisingly, the obtained accuracy was almost the same from both receivers. To ensure national security, the U.S. DoD implemented the so-called selective availability (SA) on Block II GPS satellites to deny accurate real-time autonomous positioning to unauthorized users. SA was officially activated on March 25, 1990 [3].

SA introduces two types of errors [6]. The first one, called delta error, results from dithering the satellite clock, and is common to all users worldwide. The second one, called epsilon error, is an additional slowly varying orbital error. With SA turned on, nominal horizontal and vertical errors can be up to 100m and 156m, respectively, at the 95% probability level. Figure 3.2 shows how the horizontal position of a stationary GPS receiver varies over time, mainly as a result of the effect of SA. Like the range error due to ephemeris error, the range error due to epsilon error is almost identical between users of short separations. Therefore, using differential GPS (DGPS; see Chapter 5) would overcome the effect of the epsilon error. In fact, DGPS provides better accuracy than the standalone P-code receiver due to the elimination or the reduction of the common errors, including SA [4].

Following extensive studies, the U.S. government discontinued SA on May 1, 2000, resulting in a much-improved autonomous GPS accuracy [2]. With the SA turned off, the nominal autonomous GPS horizontal and vertical accuracies would be in the order of 22m and 33m (95% of the time),

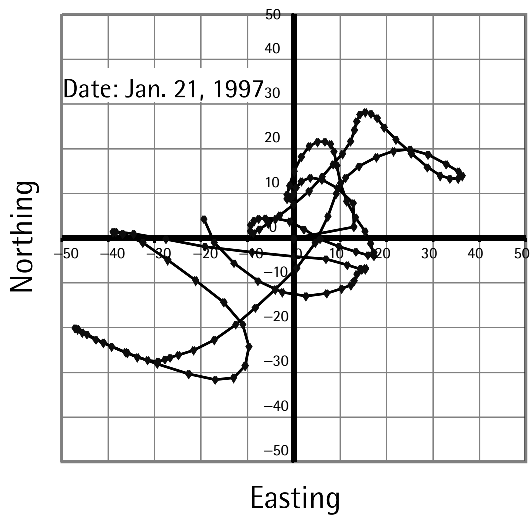

Figure 3.2 Position variation of a stationary GPS receiver due to SA.

respectively. Figure 3.3 shows the GPS errors after SA was turned off. The elimination of SA will open the door for faster growth of GPS markets (e.g., vehicle navigation and enhanced-911). Although the removal of SA would have little impact on the DGPS accuracy, it would reduce the cost of installing and operating a DGPS system. This is mainly because of the reduction in the required transmission rate.

3.3 Satellite and receiver clock errors

Each GPS Block II and Block IIA satellite contains four atomic clocks, two cesium and two rubidium [7]. The newer generation Block IIR satellites carry rubidium clocks only. One of the onboard clocks, primarily a cesium for Block II and IIA, is selected to provide the frequency and the timing requirements for generating the GPS signals. The others are backups [7].

The GPS satellite clocks, although highly accurate, are not perfect. Their stability is about 1 to 2 parts in 10^{13} over a period of one day. This means that the satellite clock error is about 8.64 to 17.28 ns per day. The corresponding range error is 2.59m to 5.18m, which can be easily calculated by multiplying the clock error by the speed of light (i.e., 299,729,458 m/s). Cesium clocks tend to behave better over a longer period of time compared with rubidium clocks. In fact, the stability of the cesium clocks

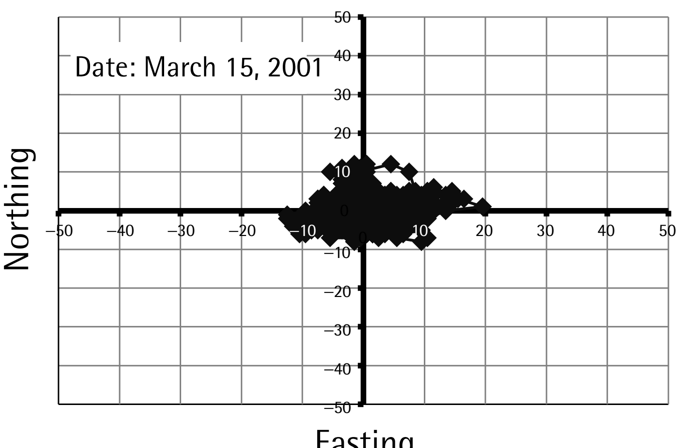

Figure 3.3 Position variation of a stationary GPS receiver after terminating SA.

over a period of 10 days or more improves to several parts in 10^{14} [7]. The performance of the satellite clocks is monitored by the ground control system. The amount of drift is calculated and transmitted as a part of the navigation message in the form of three coefficients of a second-degree polynomial [3, 8].

Satellite clock errors cause additional errors to the GPS measurements. These errors are common to all users observing the same satellite and can be removed through differencing between the receivers. Applying the satellite clock correction in the navigation message can also correct the satellite clock errors. This, however, leaves an error of the order of several nanoseconds, which translates to a range error of a few meters (one nanosecond error is equivalent to a range error of about 30 cm) [4].

GPS receivers, in contrast, use inexpensive crystal clocks, which are much less accurate than the satellite clocks [1]. As such, the receiver clock error is much larger than that of the GPS satellite clock. It can, however, be removed through differencing between the satellites or it can be treated as an additional unknown parameter in the estimation process. Precise external clocks (usually cesium or rubidium) are used in some applications instead of the internal receiver clock. Although the external atomic clocks have superior performance compared with the internal receiver clocks, they cost between a few thousand dollars for the rubidium clocks to about $20,000 for the cesium clocks.

3.4 Multipath error

Multipath is a major error source for both the carrier-phase and pseudorange measurements. Multipath error occurs when the GPS signal arrives at the receiver antenna through different paths [5]. These paths can be the direct line of sight signal and reflected signals from objects surrounding the receiver antenna (Figure 3.4).

Multipath distorts the original signal through interference with the reflected signals at the GPS antenna. It affects both the carrier-phase and pseudorange measurements; however, its size is much larger in the pseudorange measurements. The size of the carrier-phase multipath can reach a maximum value of a quarter of a cycle (about 4.8 cm for the L1 carrier phase). The pseudorange multipath can theoretically reach several tens of meters for the C/A-code measurements. However, with new advances in

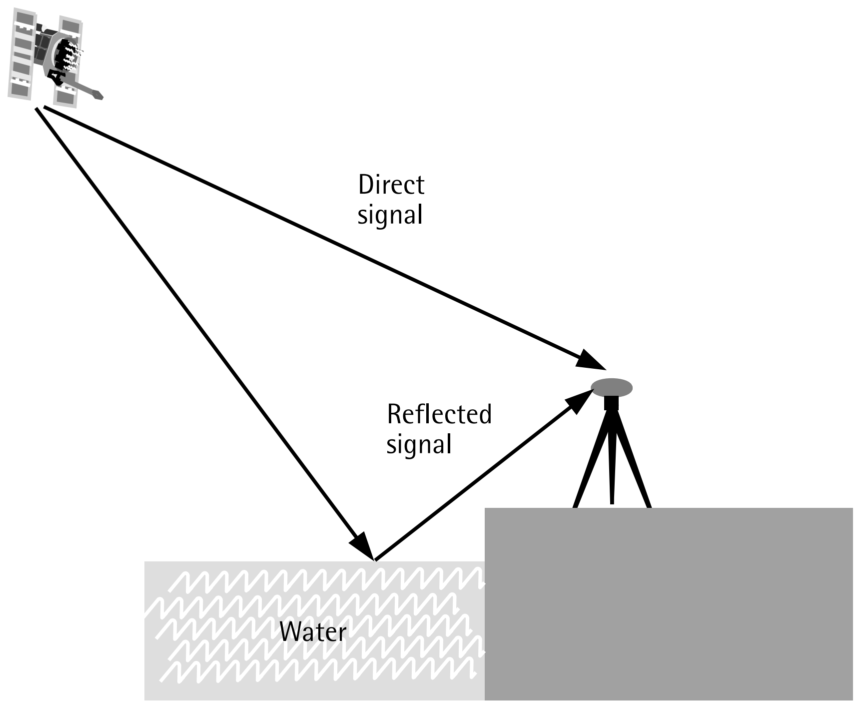

Figure 3.4 Multipath effect.

receiver technology, actual pseudorange multipath is reduced dramatically. Examples of such technologies are the Strobe correlator (Ashtech, Inc.) and the MEDLL (NovAtel, Inc.). With these multipath-mitigation techniques, the pseudorange multipath error is reduced to several meters, even in a highly reflective environment [9].

Under the same environment, the presence of multipath errors can be verified using a day-to-day correlation of the estimated residuals [3]. This is because the satellite-reflector-antenna geometry repeats every sidereal day. However, multipath errors in the undifferenced pseudorange measurements can be identified if dual-frequency observations are available. A good general multipath model is still not available, mainly because of the variant satellite-reflector-antenna geometry. There are, however, several options to reduce the effect of multipath. The straightforward option is to select an observation site with no reflecting objects in the vicinity of the receiver antenna. Another option to reduce the effect of multipath is to use a chock ring antenna (a chock ring device is a ground plane that has several concentric metal hoops, which attenuate the reflected signals). As the GPS signal is right-handed circularly polarized while the reflected signal is left-handed, reducing the effect of multipath may also be achieved by using an antenna with a matching polarization to the GPS signal (i.e., right-handed). The disadvantage of this option, however, is that the polarization of the multipath signal becomes right-handed again if it is reflected twice [9].

3.5 Antenna-phase-center variation

As stated in Chapter 2, a GPS antenna receives the incoming satellite signal and then converts its energy into an electric current, which can be handled by the GPS receiver [10]. The point at which the GPS signal is received is called the antenna phase center [3]. Generally, the antenna phase center does not coincide with the physical (geometrical) center of the antenna. It varies depending on the elevation and the azimuth of the GPS satellite as well as the intensity of the observed signal. As a result, additional range error can be expected [3].

The size of the error caused by the antenna-phase-center variation depends on the antenna type, and is typically in the order of a few centimeters. It is, however, difficult to model the antenna-phase-center variation and, therefore, care has to be taken when selecting the antenna type [1]. For short baselines with the same types of antennas at each end, the phase-center error can be canceled if the antennas are oriented in the same direction [11]. Mixing different types of antennas or using different orientations will not cancel the error. Due to its rather small size, this error is neglected in most of the practical GPS applications.

It should be pointed out that phase-center errors could be different on L1 and L2 carrier-phase observations. This can affect the accuracy of the ionosphere free linear combination, particularly when observing short baselines. As mentioned before, for short baselines, the errors are highly correlated over distance and cancel sufficiently through differencing. Therefore, using a single frequency might be more appropriate for short baselines in the static mode (see Chapter 5 for details on the static GPS positioning mode).

3.6 Receiver measurement noise

The receiver measurement noise results from the limitations of the receiver's electronics. A good GPS system should have a minimum noise level. Generally, a GPS receiver performs a self-test when the user turns it on. However, for high-cost precise GPS systems, it might be important for the user to perform the system evaluation. Two tests can be performed for evaluating a GPS receiver (system): zero baseline and short baseline tests [12].

A zero baseline test is used to evaluate the receiver performance. The test involves using one antenna/preamplifier followed by a signal splitter that feeds two or more GPS receivers (see Figure 3.5). Several receiver problems such as interchannel biases and cycle slips can be detected with this test. As one antenna is used, the baseline solution should be zero. In other words, any nonzero value is attributed to the receiver noise. Although the zero baseline test provides useful information on the receiver performance, it does not provide any information on the antenna/preamplifier noise. The contribution of the receiver measurement noise to the range error will depend very much on the quality of the GPS receiver.

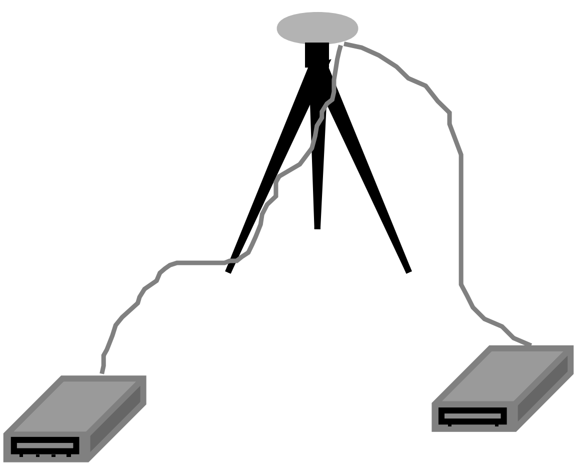

Figure 3.5 Zero baseline test for evaluating the performance of a GPS receiver.

According to [2], typical average value for range error due to the receiver measurement noise is of the order of 0.6m (1σ-level).

To evaluate the actual field performance of a GPS system, it is necessary to include the antenna/preamplifier noise component [12]. This can be done using short baselines of a few meters apart, observed on two consecutive days (see Figure 3.6). In this case, the double difference residuals of one day would contain the system noise and the multipath effect. All other errors would cancel sufficiently. As the multipath signature repeats every sidereal day, differencing the double difference residuals between the two consecutive days eliminates the effect of multipath and leaves only the system noise.

3.7 Ionospheric delay

At the uppermost part of the earth's atmosphere, ultraviolet and X-ray radiations coming from the sun interact with the gas molecules and atoms. These interactions result in *gas ionization*: a large number of free "negatively charged" electrons and "positively charged" atoms and molecules [13]. Such a region of the atmosphere where gas ionization takes place is called the ionosphere. It extends from an altitude of approximately 50 km to about 1,000 km or even more (see Figure 3.1). In fact, the upper limit of the ionospheric region is not clearly defined [14, 15].

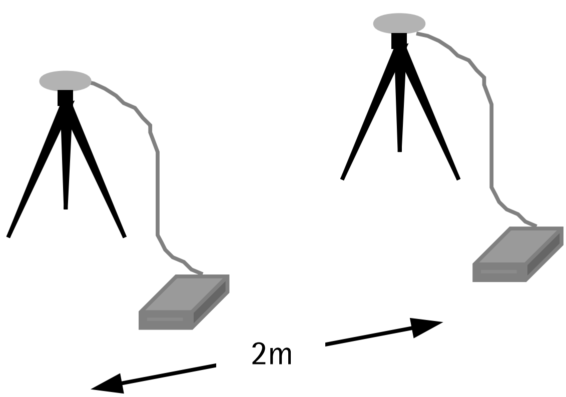

Figure 3.6 Short baseline test for evaluating the performance of a GPS system.

The electron density within the ionospheric region is not constant; it changes with altitude. As such, the ionospheric region is divided into subregions, or layers, according to the electron density. These layers are named D (50–90 km), E (90–140 km), F1 (140–210 km), and F2 (210–1,000 km), respectively, with F2 usually being the layer of maximum electron density. The altitude and thickness of those layers vary with time, as a result of the changes in the sun's radiation and the Earth's magnetic field. For example, the F1 layer disappears during the night and is more pronounced in the summer than in the winter [14].

The question that may arise is: How would the ionosphere affect the GPS measurements? The ionosphere is a dispersive medium, which means it bends the GPS radio signal and changes its speed as it passes through the various ionospheric layers to reach a GPS receiver. Bending the GPS signal path causes a negligible range error, particularly if the satellite elevation angle is greater than 5°. It is the change in the propagation speed that causes a significant range error, and therefore should be accounted for. The ionosphere speeds up the propagation of the carrier phase beyond the speed of light, while it slows down the PRN code (and the navigation message) by the same amount. That is, the receiver-satellite distance will be too short if measured by the carrier phase and too long if measured by the code, compared with the actual distance [3]. The ionospheric delay is proportional to the number of free electrons along the GPS signal path, called the total electron content (TEC). TEC, however, depends on a number of factors: (1) the time of day (electron density level reaches a daily maximum in early afternoon and a minimum around midnight at local time); (2) the time of year (electron density levels are higher in winter than in summer); (3) the 11-year solar cycle (electron density levels reach a maximum value approximately every 11 years, which corresponds to a peak in the solar flare activities known as the solar cycle peak—in 2001 we are currently around the peak of solar cycle number 23); and (4) the geographic location (electron density levels are minimum in midlatitude regions and highly irregular in polar, auroral, and equatorial regions). As the ionosphere is a dispersive medium, it causes a delay that is frequency dependent. The lower the frequency, the greater the delay; that is, the L2 ionospheric delay is greater than that of L1. Generally, ionospheric delay is of the order of 5m to 15m, but can reach over 150m under extreme solar activities, at midday, and near the horizon [5].

This discussion shows that the electron density level in the ionosphere varies with time and location. It is, however, highly correlated over relatively short distances, and therefore differencing the GPS observations between users of short separation can remove the major part of the ionospheric delay. Taking advantage of the ionosphere's dispersive nature, the ionospheric delay can be determined with a high degree of accuracy by combining the P-code pseudorange measurements on both L1 and L2. Unfortunately, however, the P-code is accessible by authorized users only. With the addition of a second C/A-code on L2 as part of the modernization program, this limitation will be removed [2]. The L1 and L2 carrier-phase measurements may be combined in a similar fashion to determine the variation in the ionospheric delay, not the absolute value. Users with dual-frequency receivers can combine the L1 and L2 carrier-phase measurements to generate the ionosphere-free linear combination to remove the ionospheric delay [5]. The disadvantages of the ionosphere-free linear combination, however, are: (1) it has a relatively higher observation noise, and (2) it does not preserve the integer nature of the ambiguity parameters. As such, the ionosphere-free linear combination is not recommended for short baselines. Single-frequency users cannot take advantage of the dispersive nature of the ionosphere. They can, however, use one of the empirical ionospheric models to correct up to 60% of the delay [13]. The most widely used model is the Klobuchar model, whose coefficients are transmitted as part of the navigation message. Another solution for users with single-frequency GPS receivers is to use corrections from regional networks [15]. Such corrections can be received in real time through communication links.

3.8 Tropospheric delay

The troposphere is the electrically neutral atmospheric region that extends up to about 50 km from the surface of the earth (see Figure 3.1). The troposphere is a nondispersive medium for radio frequencies below 15 GHz [16]. As a result, it delays the GPS carriers and codes identically. That is, the measured satellite-to-receiver range will be longer than the actual geometric range, which means that a distance between two receivers will be longer than the actual distance. Unlike the ionospheric delay, the tropospheric delay cannot be removed by combining the L1 and the L2 observations. This is mainly because the tropospheric delay is frequency independent.

The tropospheric delay depends on the temperature, pressure, and humidity along the signal path through the troposphere. Signals from satellites at low elevation angles travel a longer path through the troposphere than those at higher elevation angles. Therefore, the tropospheric delay is minimized at the user's zenith and maximized near the horizon. Tropospheric delay results in values of about 2.3m at zenith (satellite directly overhead), about 9.3m for a 15°-elevation angle, and about 20–28m for a 5°-elevation angle [17, 18].

Tropospheric delay may be broken into two components, dry and wet. The dry component represents about 90% of the delay and can be predicted to a high degree of accuracy using mathematical models [18]. The wet component of the tropospheric delay depends on the water vapor along the GPS signal path. Unlike the dry component, the wet component is not easy to predict. Several mathematical models use surface meteorological measurements (atmospheric pressure, temperature, and partial water vapor pressure) to compute the wet component. Unfortunately, however, the wet component is weakly correlated with surface meteorological data, which limits its prediction accuracy. It was found that using default meteorological data (1,010 mb for atmospheric pressure, 20°C for temperature, and 50% for relative humidity) gives satisfactory results in most cases.

3.9 Satellite geometry measures

The various types of errors and biases discussed earlier directly affect the accuracy of the computed GPS position. Proper modeling of those errors and biases and/or appropriate combinations of the GPS observables will improve the positioning accuracy. However, these are not the only factors that affect the resulting GPS accuracy. The satellite geometry, which represents the geometric locations of the GPS satellites as seen by the receiver(s), plays a very important role in the total positioning accuracy [5]. The better the satellite geometry strength, the better the obtained positioning accuracy. As such, the overall positioning accuracy of GPS is measured by the combined effect of the unmodeled measurement errors and the effect of the satellite geometry.

Good satellite geometry is obtained when the satellites are spread out in the sky [19]. In general, the more spread out the satellites are in the sky, the better the satellite geometry, and vice versa. Figure 3.7 shows a simple graphical explanation of the satellite geometry effect using two satellites [assuming a two-dimensional (2-D) case]. In such a case, the receiver will be located at the intersection of two arcs of circles; each has a radius equal to the receiver-satellite distance and a center at the satellite itself. Because of the measurement errors, the measured receiver-satellite distance will not be exact and an uncertainty region on both sides of the estimated distance will be present. Combining the measurements from the two satellites, it can be seen that the receiver will in fact be located somewhere within the uncertainty area, the hatched area. It is known from statistics that, for a certain probability level, if the size of the uncertainty area is small, the computed receiver's position will be precise. As shown in Figure 3.7(a), if the two satellites are far apart (i.e., spread out), the size of the uncertainty area will be small, resulting in good satellite geometry. Similarly, if the two satellites are close to each other [Figure 3.7(b)], the size of the uncertainty area will be large, resulting in poor satellite geometry.

The satellite geometry effect can be measured by a single dimensionless number called the dilution of precision (DOP). The lower the value of the DOP number, the better the geometric strength, and vice versa [3, 8]. The DOP number is computed based on the relative receiver-satellite geometry at any instance, that is, it requires the availability of both the receiver and the satellite coordinates. Approximate values for the coordinates are

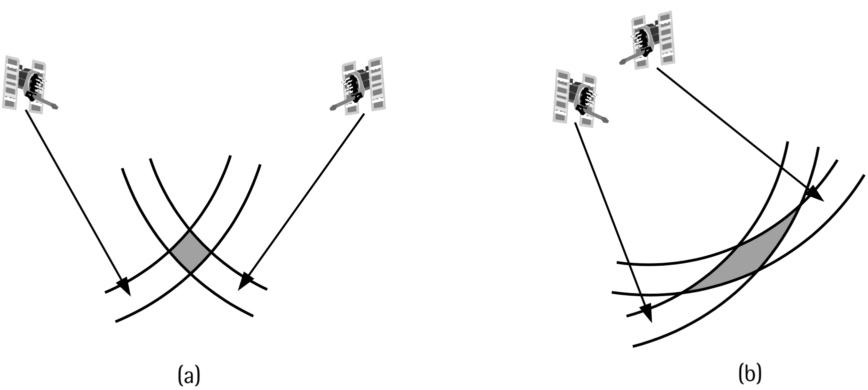

Figure 3.7 (a) Good satellite geometry; and (b) bad satellite geometry.

generally sufficient though, which means that the DOP value can be determined without making any measurements. As a result of the relative motion of the satellites and the receiver(s), the value of the DOP will change over time. The changes in the DOP value, however, will generally be slow except in the following two cases: (1) a satellite is rising or falling as seen by the user's receiver, and (2) there is an obstruction between the receiver and the satellite (e.g., when passing under a bridge).

In practice, various DOP forms are used, depending on the user's need [19]. For example, for the general GPS positioning purposes, a user may be interested in examining the effect of the satellite geometry on the quality of the resulting three-dimensional (3-D) position (latitude, longitude, and height). This could be done by examining the value of the position dilution of precision (PDOP). In other words, PDOP represents the contribution of the satellite geometry to the 3-D positioning accuracy. PDOP can be broken into two components: horizontal dilution of precision (HDOP) and vertical dilution of precision (VDOP). The former represents the satellite geometry effect on the horizontal component of the positioning accuracy, while the latter represents the satellite geometry effect on the vertical component of the positioning accuracy. Because a GPS user can track only those satellites above the horizon, VDOP will *always* be larger than HDOP. As a result, the GPS height solution is expected to be less precise than the horizontal solution. The VDOP value could be improved by supplementing GPS with other sensors, for example, the pseudolites (see Chapter 9 for details). Other commonly used DOP forms include the time dilution of precision (TDOP) and the geometric dilution of precision (GDOP). GDOP represents the combined effect of the PDOP and the TDOP.

To ensure high-precision GPS positioning, it is recommended that a suitable observation time be selected to obtain the highest possible accuracy. A PDOP of five or less is usually recommended. In fact, the actual PDOP value is usually much less than five, with a typical average value in the neighborhood of two. Most GPS software packages have the ability to predict the satellite geometry based on the user's approximate location and the approximate satellite locations obtained from a recent almanac file for the GPS constellation. The almanac file is obtained as part of the navigation message, and can be downloaded free of charge over the Internet (e.g., from the U.S. Coast Guard Navigation Center [20]).

3.10 GPS mission planning

Even under the full constellation of 24 GPS satellites, there exist some periods of time where only four satellites are visible above a particular elevation angle, which may not be enough for some GPS works. Such a satellite visibility problem is expected more at high latitudes (higher than about 55°) because of the nature of the GPS constellation [4]. This problem may also occur in some low- or midlatitude areas for a particular period of time. For example, in urban and forested areas, the receiver's sky window is reduced as a result of the obstruction caused by the high-rise buildings and the trees. Because the satellite geometry changes over time, the satellite visibility problem may be overcome by selecting a suitable observation time, which ensures a minimum number of visible satellites and/or a particular maximum DOP value. To help users in identifying the best observation periods, GPS manufacturers have developed mission-planning software packages, which predict the satellite visibility and geometry at any given location.

Mission-planning software provides a number of plots to help in planning the GPS survey or mission. The first plot is known as the sky plot, which represents the user's sky window by a series of concentric circles. Figure 3.8 shows the sky plot for Toronto on April 13, 2001, which was

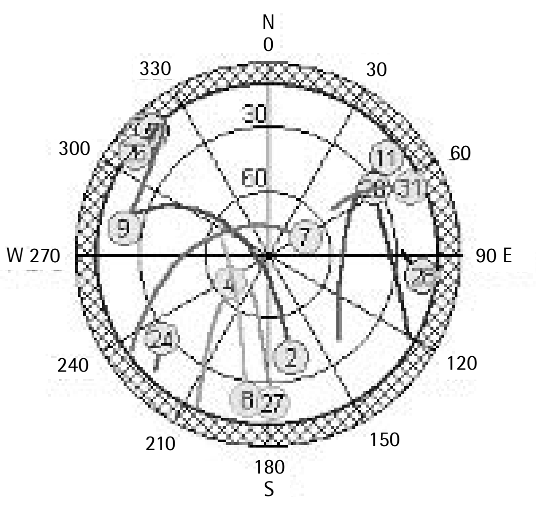

Figure 3.8 GPS sky plot for Toronto on April 13, 2001.

produced by the Ashtech Locus processor software. The center point represents the user's zenith, while the outer circle represents his or her horizon. Intermediate circles represent different elevation angles. The outer circle is also graduated from 0° to 360° to represent the satellite azimuth (direction) at any time. Once the user inputs his or her approximate location and the desired observation period, the path of each satellite in his or her view will be shown on the sky plot. This means that relative satellite locations, the satellite azimuth and elevation, can be obtained. The user may also specify a certain elevation angle, normally 10° or 15°, to be used as a mask or cutoff angle. A mask angle is the angle below which the receiver will not track any satellite, even if the satellite is above the receiver's horizon.

Other important plots include the satellite availability plot, which shows the total number of the visible satellites above the user-specified mask angle, and the satellite geometry plot. Figure 3.9 shows the satellite availability and geometry for Toronto on April 13, 2001. The satellite geometry plot is normally represented by the PDOP, HDOP, and VDOP.

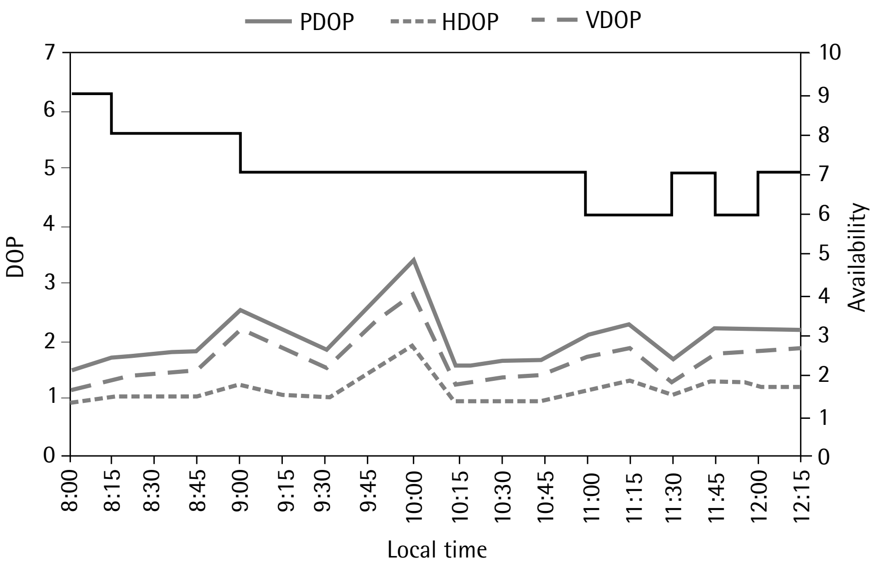

Figure 3.9 Satellite availability and geometry for Toronto on April 13, 2001.

3.11 User equivalent range error

It has been shown that the GPS positioning accuracy is measured by the combined effect of the unmodeled measurement errors and the effect of the satellite geometry. The unmodeled measurement errors will certainly be different from one satellite to another, mainly because of the various view angles. In addition, the ranging errors for the various satellites will have a certain degree of similarity (i.e., correlated). To rigorously determine the expected GPS positioning accuracy, we may apply an estimation technique such as the least-squares method [8, 18]. The least-squares method estimates the user's position (location) as well as its covariance matrix. The latter tells us how well the user's position is determined. In fact, the covariance matrix reflects the combined effect of the measurement errors and the satellite geometry.

A more simplified way of examining the GPS positioning accuracy may be achieved through the introduction of the user equivalent range error (UERE). Assuming that the measurement errors for all the satellites are identical and independent, then a quantity known as the UERE may be defined as the root-sum-square of the various errors and biases discussed earlier [3]. Multiplying the UERE by the appropriate DOP value produces the expected precision of the GPS positioning at the one-sigma (1-σ) level. To obtain the precision at the 2-σ level, sometimes referred to as approximately 95% of the time, we multiply the results by a factor of two. For example, assuming that the UERE is 8m for the standalone GPS receiver, and taking a typical value of HDOP as 1.5, then the 95% positional accuracy will be $8 \times 1.5 \times 2 = 24$m.

References

[1] Kleusberg, A., and R. B. Langley, "The Limitations of GPS," *GPS World*, Vol. 1, No. 2, March/April 1990, pp. 50–52.

[2] Shaw, M., K. Sandhoo, and D. Turner, "Modernization of the Global Positioning System," *GPS World*, Vol. 11, No. 9, September 2000, pp. 36–44.

[3] Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice,* 3rd ed., New York: Springer-Verlag, 1994.

[4] El-Rabbany, A., *The Effect of Physical Correlations on the Ambiguity Resolution and Accuracy Estimation in GPS Differential Positioning*,

Technical Report No. 170, Department of Geodesy and Geomatics Engineering, Fredericton, New Brunswick, Canada: University of New Brunswick, 1994.

[5] Wells, D. E., et al., *Guide to GPS Positioning,* Fredericton, New Brunswick: Canadian GPS Associates, 1987.

[6] Georgiadou, Y., and K. D. Doucet, "The Issue of Selective Availability," *GPS World*, Vol. 1, No. 5, September/October 1990, pp. 53–56.

[7] Langley, R. B., "Time, Clocks, and GPS," *GPS World*, Vol. 2, No. 10, November/December 1991, pp. 38–42.

[8] Kaplan, E., *Understanding GPS: Principles and Applications,* Norwood, MA: Artech House, 1996.

[9] Weill, L. R., "Conquering Multipath: The GPS Accuracy Battle," *GPS World*, Vol. 8, No. 4, April 1997, pp. 59–66.

[10] Langley, R. B., "The GPS Receiver: An Introduction," *GPS World*, Vol. 2, No. 1, January 1991, pp. 50–53.

[11] Schupler, B. R., and T. A. Clark, "How Different Antennas Effect the GPS Observable," *GPS World*, Vol. 2, No. 10, November/December 1991, pp. 32–36.

[12] Nolan, J., S. Gourevitch, and J. Ladd, "Geodetic Processing Using Full Dual Band Observables," *Proc. ION GPS-92, 5th Intl. Technical Meeting*, Satellite Div., Institute of Navigation, Albuquerque, NM, September 16–18, 1992, pp. 1033–1041.

[13] Klobuchar, J. A., "Ionospheric Effects on GPS," *GPS World*, Vol. 2, No. 4, April 1991, pp. 48–51.

[14] Komjathy, A., *Global Ionospheric Total Electron Content Mapping Using the Global Positioning System,* Ph.D. dissertation, Department of Geodesy and Geomatics Engineering, Technical Report No. 188, University of New Brunswick, Fredericton, New Brunswick, Canada, 1997.

[15] Langley, R. B., "GPS, the Ionosphere, and the Solar Maximum," *GPS World*, Vol. 11, No. 7, July 2000, pp. 44–49.

[16] Hay, C., and J. Wong, "Enhancing GPS: Tropospheric Delay Prediction at the Master Control Station," *GPS World*, Vol. 11, No. 1, January 2000, pp. 56–62.

[17] Brunner, F. K., and W. M. Welsch, "Effect of the Troposphere on GPS Measurements," *GPS World*, Vol. 4, No. 1, January 1993, pp. 42–51.

[18] Leick, A., *GPS Satellite Surveying,* 2nd ed., New York: Wiley, 1995.

[19] Langley, R. B., "Dilution of Precision," *GPS World*, Vol. 10, No. 5, May 1999, pp. 52–59.

[20] U.S. Coast Guard Navigation Center, accessed 2001, http://www.navcen.uscg.gov/GPS/default.htm#almanacs.

4

Datums, Coordinate Systems, and Map Projections

The ability of GPS to determine the precise location of a user anywhere, under any weather conditions, attracted millions of users worldwide from various fields and backgrounds. With advances in GPS and computer technologies, GPS manufacturers were able to come up with very user-friendly systems. However, one common problem that many newcomers to the GPS face is the issue of datums and coordinate systems, which require some geodetic background. This chapter tackles the problem of datums and coordinate systems in detail. As in the previous chapters, complex mathematical formulas are avoided. As many users are interested in the horizontal component of the GPS position, the issue of map projections is also introduced. For the sake of completeness, the height systems are introduced as well, at the end of this chapter.

4.1 What is a datum?

The fact that the topographic surface of the Earth is highly irregular makes it difficult for the geodetic calculations—for example, the determination of the user's location—to be performed. To overcome this problem, geodesists adopted a smooth mathematical surface, called the reference surface, to approximate the irregular shape of the earth (more precisely to approximate the global mean sea level, the geoid) [1, 2]. One such mathematical surface is the sphere, which has been widely used for low-accuracy positioning. For high-accuracy positioning such as GPS positioning, however, the best mathematical surface to approximate the Earth and at the same time keep the calculations as simple as possible was found to be the biaxial ellipsoid (see Figure 4.1). The biaxial reference ellipsoid, or simply the reference ellipsoid, is obtained by rotating an ellipse around its minor axis, b [2]. Similar to the ellipse, the biaxial reference ellipsoid can be defined by the semiminor and semimajor axes (a, b) or the semimajor axis and the flattening (a, f), where $f = 1 - (b / a)$.

An appropriately positioned reference ellipsoid is known as the geodetic datum [2]. In other words, a geodetic datum is a mathematical surface, or a reference ellipsoid, with a well-defined origin (center) and orientation. For example, a geocentric geodetic datum is a geodetic datum with its origin coinciding with the center of the Earth. It is clear that there are an infinite number of geocentric geodetic datums with different orientations. Therefore, a geodetic datum is uniquely determined by specifying eight parameters: two parameters to define the dimension of the reference ellipsoid; three parameters to define the position of the origin; and three parameters to define the orientation of the three axes with respect to the earth. Table 4.1 shows some examples of three common reference systems and their associated ellipsoids [3].

In addition to the geodetic datum, the so-called vertical datum is used in practice as a reference surface to which the heights (elevations) of points are referred [2]. Because the height of a point directly located on the vertical datum is zero, such a vertical reference surface is commonly known as the surface of zero height. The vertical datum is often selected to be the geoid; the surface that best approximates the mean sea level on a global basis [see Figure 4.1(a)].

In the past, positions with respect to horizontal and vertical datums have been determined independent of each other [2]. However, with the

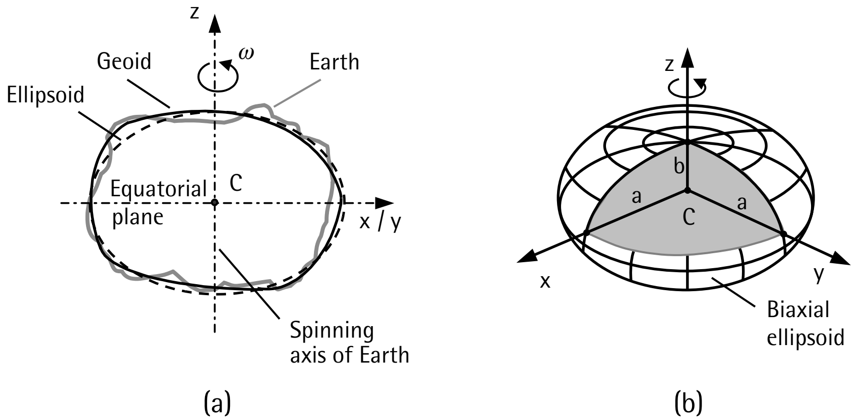

Figure 4.1 (a) Relationship between the physical surface of the Earth, the geoid, and the ellipsoid; and (b) ellipsoidal parameters.

advent of space geodetic positioning systems such as GPS, it is possible to determine the 3-D positions with respect to a 3-D reference system.

4.2 Geodetic coordinate system

A coordinate system is defined as a set of rules for specifying the locations (also called coordinates) of points [4]. This usually involves specifying an origin of the coordinates as well as a set of reference lines (called axes) with known orientation. Figure 4.2 shows the case of a 3-D coordinate system that uses three reference axes (x, y, and z) that intersect at the origin (C) of the coordinate system.

Table 4.1 Examples of Reference Systems and Associated Ellipsoids

Reference Systems	Ellipsoid	a(m)	$1/f$
WGS 84	WGS 84	6378137.0	298.257223563
NAD 83	GRS 80	6378137.0	298.257222101
NAD 27	Clarke 1866	6378206.4	294.9786982

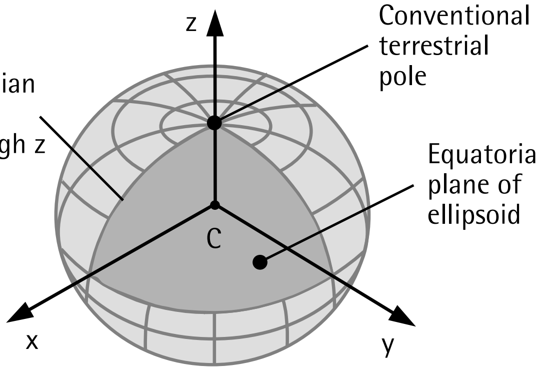

Figure 4.2 3-D coordinate system.

Coordinate systems may be classified as one-dimensional (1-D), 2-D, or 3-D coordinate systems, according to the number of coordinates required to identify the location of a point. For example, a 1-D coordinate system is needed to identify the height of a point above the sea surface.

Coordinate systems may also be classified according to the reference surface, the orientation of the axes, and the origin. In the case of a 3-D geodetic (also known as geographic) coordinate system, the reference surface is selected to be the ellipsoid. The orientation of the axes and the origin are specified by two planes: the meridian plane through the polar or z-axis (a meridian is a plane that passes through the north and south poles) and the equatorial plane of the ellipsoid (see Figure 4.2 for details).

Of particular importance to GPS users is the 3-D geodetic coordinate system. In this system, the coordinates of a point are identified by the geodetic latitude (ϕ), the geodetic longitude (λ), and the height above the reference surface (h). Figure 4.3 shows these parameters. Geodetic coordinates (ϕ, λ, and h) can be easily transformed to Cartesian coordinates (x, y, and z) as shown in Figure 4.3(b) [2]. To do this, the ellipsoidal parameters (a and f) must be known. It is also possible to transform the geodetic coordinates (ϕ and λ) into a rectangular grid coordinate (e.g., Northing and Easting) for mapping purposes [5].

4.2.1 Conventional Terrestrial Reference System

The Conventional Terrestrial Reference System (CTRS) is a 3-D geocentric coordinate system, that is, its origin coincides with the center of the Earth

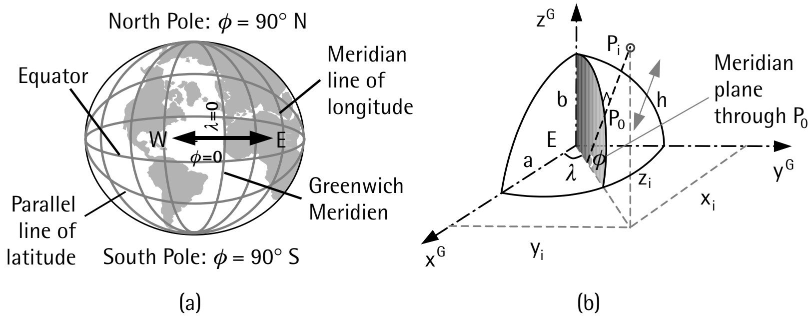

Figure 4.3 (a) Concept of geodetic coordinates; and (b) geodetic and Cartesian coordinates.

(Figure 4.2). The CTRS is rigidly tied to the Earth, that is, it rotates with the Earth [5]. It is therefore also known as the Earth-centered, Earth-fixed (ECEF) coordinate system.

The orientation of the axes of the CTRS is defined as follows: The z-axis points toward the conventional terrestrial pole (CTP), which is defined as the average location of the pole during the period 1900–1905 [3]. The x-axis is defined by the intersection of the terrestrial equatorial plane and the meridional plane that contains the mean location of the Greenwich observatory (known as the mean Greenwich meridian). It is clear from the definition of the x and z axes that the xz-plane contains the mean Greenwich meridian. The y-axis is selected to make the coordinate system right-handed (i.e., 90° east of the x-axis, measured in the equatorial plane). The three axes intersect at the center of the Earth, as shown in Figure 4.2.

The CTRS must be positioned with respect to the Earth (known as realization) to be of practical use in positioning [2]. This is done by assigning coordinate values to a selected number of well-distributed reference stations. One of the most important CTRSs is the International Terrestrial Reference System (ITRS), which is realized as the International Terrestrial Reference Frame (ITRF). The ITRF solution is based on the measurements from globally distributed reference stations using GPS and other space geodetic systems. It is therefore considered to be the most accurate coordinate system [6]. The ITRF is updated every 1 to 3 years to achieve the highest possible accuracy. The most recent version at the time of this writing is the ITRF2000.

4.2.2 The WGS 84 and NAD 83 systems

The World Geodetic System of 1984 (WGS 84) is a 3-D, Earth-centered reference system developed by the former U.S. Defense Mapping Agency now incorporated into a new agency, National Imagery and Mapping Agency (NIMA). It is the official GPS reference system. In other words, a GPS user who employs the broadcast ephemeris in the solution process will obtain his or her coordinates in the WGS 84 system. The WGS 84 utilizes the CTRS combined with a reference ellipsoid that is identical, up to a very slight difference in flattening, with the ellipsoid of the Geodetic Reference System of 1980 (GRS 80); see Table 4.1. The latter was recommended by the International Association of Geodesy for use in geodetic applications [5]. WGS 84 was originally established (realized) using a number of Doppler stations. It was then updated several times to bring it as close as possible to the ITRF reference system. With the most recent update, WGS 84 is coincident with the ITRF at the subdecimeter accuracy level [7].

In North America, another nominally geocentric datum, the North American Datum of 1983 (NAD 83), is used as the legal datum for spatial positioning. NAD 83 utilizes the ellipsoid of the GRS 80, which means that the size and shape of both WGS 84 and NAD 83 are almost identical. The original realization of NAD 83 was done in 1986, by adjusting "primarily" classical geodetic observations that connected a network of horizontal control stations spanning North America, and several hundred observed Doppler positions. Initially, NAD 83 was designed as an Earth-centered reference system [8]. However, with the development of more accurate techniques, it was found that the origin of NAD 83 is shifted by about 2m from the true Earth's center. In addition, access to NAD 83 was provided mainly through a horizontal control network, which has a limited accuracy due to the accumulation of errors. To overcome these limitations, NAD 83 was tied to ITRF using 12 common, very long baseline interferometry (VLBI) stations located in both Canada and the United States (VLBI is a highly accurate, yet complex, space positioning system). This resulted in an improved realization of the NAD 83, which is referred to as NAD 83 (CSRS) and NAD 83 (NSRS) in both Canada and the United States, respectively [8]. The acronyms CSRS and NSRS refer to the Canadian Spatial Reference System and National Spatial Reference System, respectively. It should be pointed out that, due to the different versions of the ITRF, it is important to define to which epoch the ITRF coordinates refer.

4.3 What coordinates are obtained with GPS?

The satellite coordinates as given in the broadcast ephemeris will refer to the WGS 84 reference system. Therefore, a GPS user who employs the broadcast ephemeris in the adjustment process will obtain his or her coordinates in the WGS 84 system as well. However, if a user employs the precise ephemeris obtained from the IGS service (Chapter 7), his or her solution will be referred to the ITRF reference system. Some agencies provide the precise ephemeris in various formats. For example, Geomatics Canada provides its precise ephemeris data in both the ITRF and the NAD 83 (CSRS) formats.

The question that may arise is what happens if the available reference (base) station coordinates are in NAD 83 rather than in WGS 84? The answer to this question varies, depending on whether the old or the improved NAD 83 system is used. Although the sizes and shapes of the reference ellipsoids of the WGS 84 and the old NAD 83 are almost identical; their origins are shifted by more than 2m with respect to each other [3]. This shift causes a discrepancy in the absolute coordinates of points when expressed in both reference systems. In other words, a point on the Earth's surface will have WGS 84 coordinates that are different from its coordinates in the old NAD 83. The largest coordinate difference is in the height component (about 0.5m). However, the effect of this shift on the relative GPS positioning is negligible. For example, if a user applies the NAD 83 coordinates for the reference station instead of its WGS 84, his or her solution will be in the NAD 83 reference system with a negligible error (typically at the millimeter level). The improved WGS 84 and the NAD 83 systems are compatible.

4.4 Datum transformations

As stated in Section 4.1, in the past, positions with respect to horizontal and vertical datums have been determined independent of each other [2]. In addition, horizontal datums were nongeocentric and were selected to best fit certain regions of the world (Figure 4.4). As such, those datums were commonly called local datums. More than 150 local datums have been used by different countries of the world. An example of the local datums is the North American datum of 1927 (NAD 27). With the advent of space

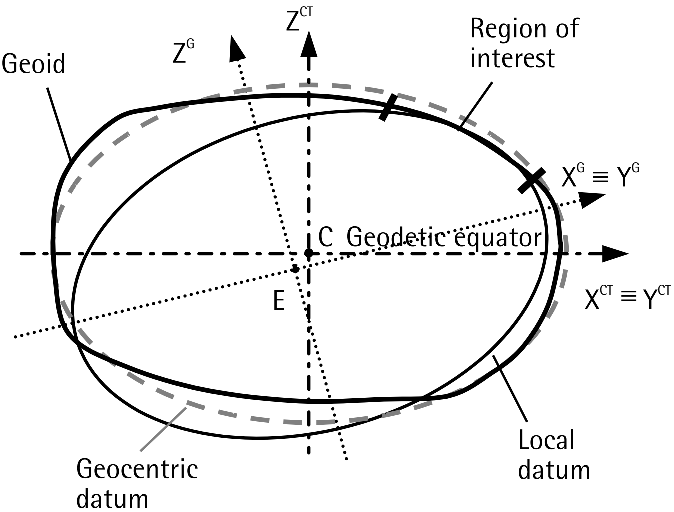

Figure 4.4 Geocentric and local datums.

geodetic positioning systems such as GPS, it is now possible to determine global 3-D geocentric datums.

Old maps were produced with the local datums, while new maps are mostly produced with the geocentric datums. Therefore, to ensure consistency, it is necessary to establish the relationships between the local datums and the geocentric datums, such as WGS 84. Such a relationship is known as the datum transformation (see Figure 4.5). NIMA has published the transformation parameters between WGS 84 and the various local datums used in many countries. Many GPS manufacturers currently use these parameters within their processing software packages. It should be clear, however, that these transformation parameters are only approximate and should not be used for precise GPS applications. In Toronto, for example, a difference as large as several meters in the horizontal coordinates is obtained when applying NIMA's parameters (WGS 84 to NAD 27) as compared with the more precise National Transformation software (NTv2) produced by Geomatics Canada. Such a difference could be even larger in other regions. The best way to obtain the transformation parameters is by comparing the coordinates of well-distributed common points in both datums.

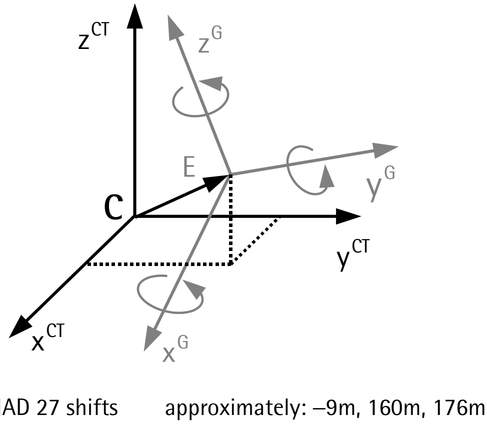

Figure 4.5 Datum transformations.

4.5 Map projections

Map projection is defined, from the geometrical point of view, as the transformation of the physical features on the curved Earth's surface onto a flat surface called a map (see Figure 4.6). However, it is defined, from the mathematical point of view, as the transformation of geodetic coordinates (ϕ, λ) obtained from, for example, GPS, into rectangular grid coordinates often called easting and northing. This is known as the direct map

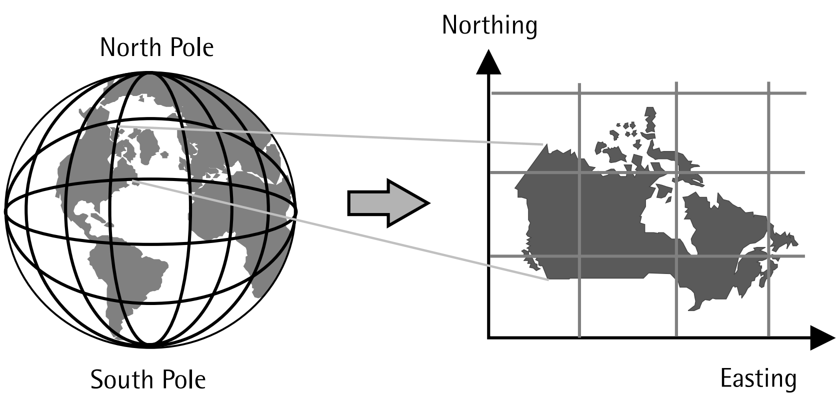

Figure 4.6 Concept of map projection.

projection [2, 4]. The inverse map projection involves the transformation of the grid coordinates into geodetic coordinates. Rectangular grid coordinates are widely used in practice, especially the Geomatics-related works. This is mainly because mathematical computations are performed easier on the mapping plane as compared with the reference surface (i.e., the ellipsoid).

Unfortunately, because of the difference between the ellipsoidal shape of the Earth and the flat projection surface, the projected features suffer from distortion [3]. In fact, this is similar to trying to flatten the peel of one-half of an orange; we will have to stretch portions and shrink others, which results in distorting the original shape of the peel. A number of projection types have been developed to minimize map distortions. In most of the GPS applications, the so-called conformal map projection is used [2]. With conformal map projection, the angles on the surface of the ellipsoid are preserved after being projected on the flat projection surface (i.e., the map). However, both the areas and the scales are distorted; remember that areas are either squeezed or stretched [9]. The most popular conformal map projections are transverse Mercator, universal transverse Mercator (UTM), and Lambert conformal conic projections.

It should be pointed out that not only the projection type should accompany the grid coordinates of a point, but also the reference system. This is because the geodetic coordinates of a particular point will vary from one reference system to another. For example, a particular point will have different pairs of UTM coordinates if the reference systems are different (e.g., NAD 27 and NAD 83).

4.5.1 Transverse Mercator projection

Transverse Mercator projection (also known as Gauss-Krüger projection) is a conformal map projection invented by Johann Lambert (Germany) in 1772 [9]. It is based on projecting the points on the ellipsoidal surface mathematically onto an imaginary transverse cylinder (i.e., its axis lies in the equatorial plane). The cylinder can be either a tangent to the ellipsoid along a meridian called the central meridian, or a secant cylinder (see Figure 4.7 for the case of tangent cylinder). In the latter case, two small complex curves at equal distance from the central meridian are produced.

Upon cutting and unfolding the imaginary cylinder, the required flat map (i.e., transverse Mercator projection) is produced. Again, it should be understood that the transverse cylinder is only an imaginary surface. As

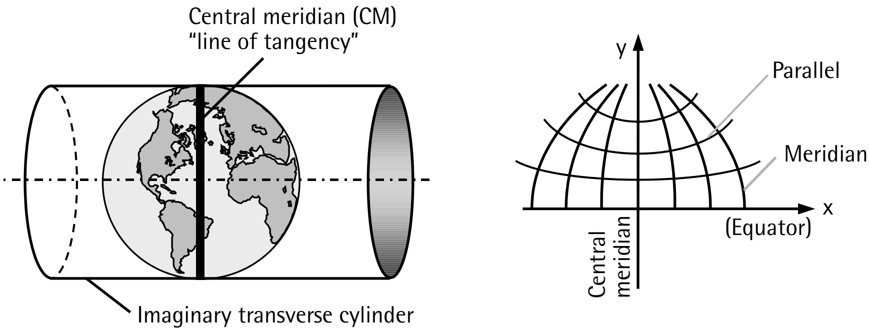

Figure 4.7 Transverse Mercator map projection.

explained earlier, the projection is made mathematically through the transformation of the geodetic coordinates into the grid coordinates.

In the case of a tangent cylinder, all features along the line of tangency, the central meridian, are mapped without distortion. This means that the scale, which is a measure of the amount of distortion, is true (equals one) along the central meridian. As we move away from the central meridian, the projected features will suffer from distortion. The farther we are from the central meridian, the greater the distortion. In fact, the scale factor increases symmetrically as we move away from the central meridian. This is why this projection is more suitable for areas that are long in the north-south direction.

In the case of a secant cylinder, all features along the two small complex curves will be mapped without distortion (see Figure 4.8). The scale is true along the two small complex curves, not the central meridian. Similar to the tangent cylinder case, the distortion increases as we move away from the two small complex curves.

4.5.2 Universal transverse Mercator projection

The universal transverse Mercator (UTM) is a map projection that is based completely on the original transverse Mercator, with a secant cylinder (Figure 4.8). With UTM, however, the Earth (i.e., the ellipsoid) is divided into 60 zones of the same size; each zone has its own central meridian that is located at exactly the middle of the zone [9]. This means that each zone covers 6° of longitude, 3° on each side of the zone's central meridian. Each zone is projected separately (i.e., the imaginary cylinder will be rotated

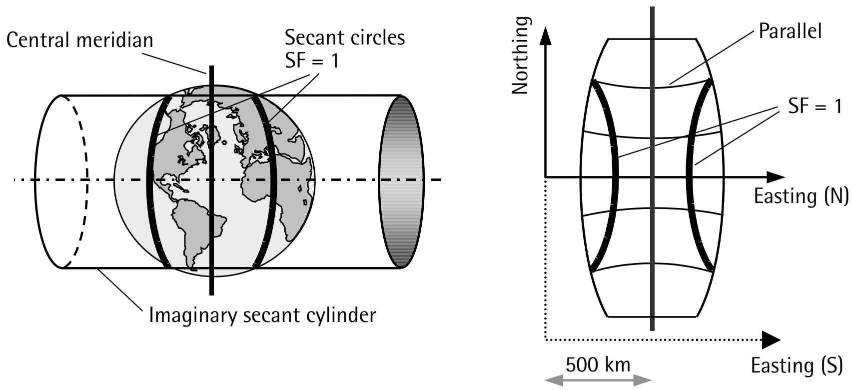

Figure 4.8 UTM projection.

around the Earth), which leads to a much smaller distortion compared with the original transverse Mercator projection. Each zone is assigned a number ranging from 1 to 60, starting from $\lambda = 180°$ W, and increases eastward (i.e., zone 1 starts at 180° W and ends at 174° W with its central meridian at 177° W); see Figure 4.9.

UTM utilizes a scale factor of 0.9996 along the zone's central meridian (Figure 4.8). The reason for selecting this scale factor is to have a more uniformly distributed scale, with a minimum deviation from one, over the entire zone. For example, at the equator, the scale factor changes from 0.9996 at the central meridian to 1.00097 at the edge of the zone, while at

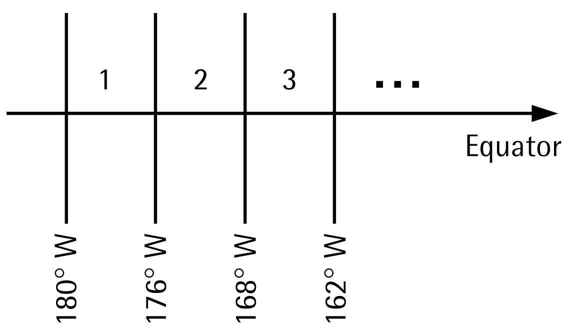

Figure 4.9 UTM zoning.

midlatitude (ϕ = 45° N), the scale changes from 0.9996 at the central meridian to 1.00029 at the edge of the zone. This shows how the distortion is kept at a minimal level with UTM.

To avoid negative coordinates, the true origin of the grid coordinates (i.e., where the equator meets the central meridian of the zone) is shifted by introducing the so-called false northing and false easting (Figure 4.8). The false northing and false easting take different values, depending on whether we are in the northern or the southern hemisphere. For the northern hemisphere, the false northing and false easting are 0.0 km and 500 km, respectively, while for the southern hemisphere, they are 10,000 km, and 500 km, respectively.

A final point to be made here is that UTM is not suitable for projecting the polar regions. This is mainly due to the many zones to be involved when projecting a small polar area. Other projection types, such as the stereographic double projection, may be used (see Section 4.5.5).

4.5.3 Modified transverse Mercator projection

The modified transverse Mercator (MTM) projection is another projection that, similar to the UTM, is based completely on the original transverse Mercator, with a secant cylinder [9]. MTM is used in some Canadian provinces such as the province of Ontario. With MTM, a region is divided into zones of 3° of longitude each (i.e., 1.5° on each side of the zone's central meridian). Similar to UTM, each zone is projected separately, which leads to a small distortion. In Canada, the first zone starts at some point just east of Newfoundland (λ = 51° 30′ W), and increases westward. Canada is covered by a total of 32 zones, while the province of Ontario is covered by 10 zones (zones 8 through 17). Figure 4.10 shows zone 10, where the city of Toronto is located.

MTM utilizes a scale factor of 0.9999 along the zone's central meridian (Figure 4.10). This leads to even less distortion throughout the zone, as compared with the UTM. For example, at a latitude of ϕ = 43.5° N, the scale factor changes from 0.9999 at the central meridian to 1.0000803 at the boundary of the zone. This shows how the scale variation and, consequently, the distortion are minimized with MTM [9]. This, however, has the disadvantage that the number of zones is doubled.

Similar to UTM, to avoid negative coordinates, the true origin of the grid coordinates is shifted by introducing the false northing and false easting. As Canada is completely located in the northern hemisphere, there is

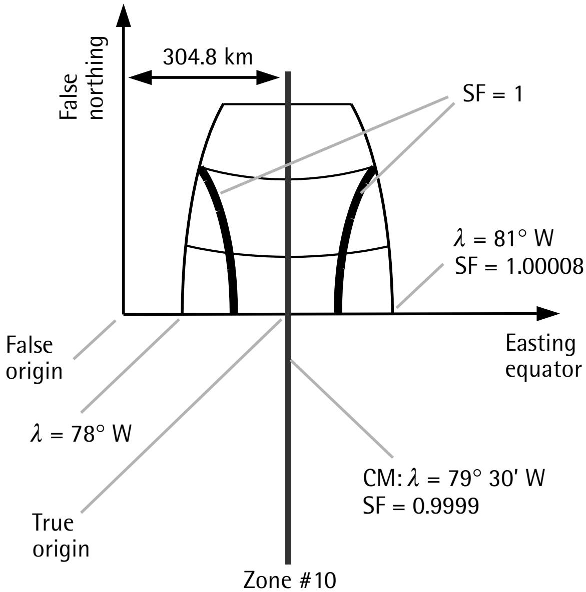

Figure 4.10 MTM projection.

only one false northing and one false easting of 0.0m and 304,800m, respectively (see Figure 4.10).

4.5.4 Lambert conical projection

Lambert conical projection is a conformal map projection developed by Johann Lambert (Germany) in 1772 (the same year in which he developed the transverse Mercator projection). It is based on projecting the points on the ellipsoidal surface mathematically onto an imaginary cone [9]. The cone may either touch the ellipsoid along one of the parallels or intersect the ellipsoid along two parallels. The resulting parallels are called the standard parallels (i.e., one standard parallel is produced in the first case while two standard parallels are produced in the second case). Upon cutting and unfolding the imaginary cone, the required flat map is produced (Figure 4.11).

As with the case of the transverse Mercator projection, all features along the standard parallels are mapped without distortion. As we move away from the standard parallels, the projected features will suffer from

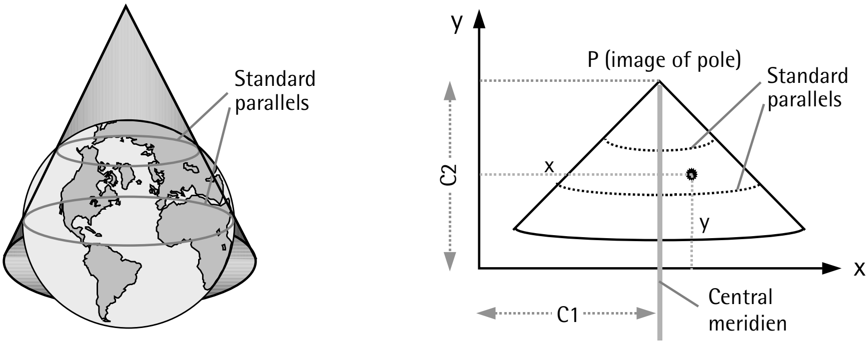

Figure 4.11 Lambert conical projection.

distortion. This means that this projection is more suitable for areas that extend in the east-west direction.

This projection is designed so that all the parallels are projected as parts of concentric circles with the center at the apex of the cone, while all the meridians are projected as straight lines converging at the apex of the cone. In other words, the meridians will be the radii of concentric centers. A central meridian that nearly passes through the middle of the area to be mapped is selected to establish the direction of the grid north (i.e., the y-axis). To avoid negative coordinates, the origin of the grid coordinates is shifted by introducing two constants, C1 and C2 (see Figure 4.11). The values of C1, C2, and the latitude of the standard parallels are determined by the mapping authorities.

4.5.5 Stereographic double projection

The stereographic double projection is a map projection used in some parts of the world, including the Canadian province of New Brunswick. With this mapping projection, points on the reference ellipsoid are projected onto the projection plane through an intermediate surface: an imaginary sphere [9]. In other words, the projection is done in two steps, hence the name "double projection." First, features on the reference ellipsoid are conformally projected onto an imaginary sphere. Second, features on the sphere are conformally projected onto a tangent or a secant plane to produce the required map (Figure 4.12). The latter projection is known as stereographic projection.

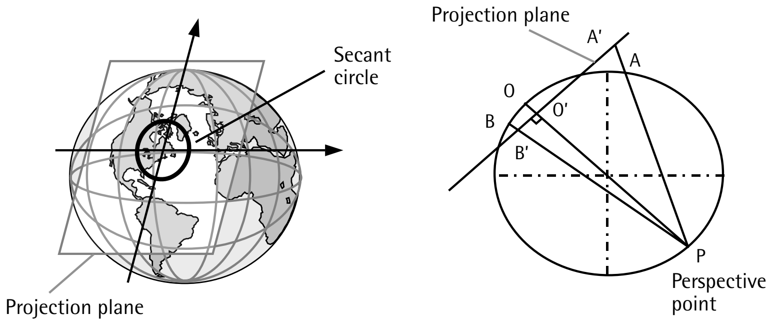

Figure 4.12 Stereographic double projection.

There are three cases of stereographic projection to be obtained depending on the position of the projection plane relative to the sphere (i.e., the origin O). If the origin is selected at one of the poles of the sphere, the projection is called polar stereographic. However, if it is selected at some point on the equator of the sphere, the projection is called transverse or equatorial stereographic. The general case in which the origin is selected at an arbitrary point is called oblique stereographic. In the last case, the meridian passing through the map origin is projected as a straight line. All other meridians and parallels are projected as circles. In New Brunswick, a secant projection plane is used with an origin selected at $\phi = 46° 30'$ N and $\lambda = 66° 30'$ W.

In the stereographic projection, a perspective point (P) is first selected to be diametrically opposite to the origin (O). If a secant projection plane is used, a point (A) on the surface of the sphere is projected by drawing a line (PA) and extending it outward to A′ on the projection plane (see Figure 4.12). Points inside the secant circle, such as point B, are projected inward. As discussed before, features along the secant circle are projected without distortion, while other features suffer from distortion. In New Brunswick, a scale factor of 0.999912 is selected at the origin. Similar to the previous three map projections, a false northing and a false easting are introduced to avoid negative coordinates.

4.6 Marine nautical charts

Marine nautical charts are maps used by mariners for navigation purposes. They contain information such as aids to navigation and hazards. Until recently, paper charts were the only source of information available to mariners. However, over a decade ago, the Electronic Chart Display and Information System (ECDIS) was introduced, revolutionizing the field of marine navigation.

ECDIS is a computerized navigation system that integrates geographic information with navigation instrumentation [10]. It consists mainly of a computer processor and display, a digital database, and navigation sensors (see Figure 4.13). ECDIS is not only capable of displaying the navigation-related information in real time but also supporting other advanced functions [11]. For example, rout planning, rout monitoring, and automatic alarms, to name a few, are all supported by ECDIS. ECDIS and Radar/Automatic Radar Plotting Aid (ARPA) may be superimposed on a single display to provide a system that can be used for collision avoidance as well. The International Maritime Organization (IMO) adopted the performance standards for ECDIS in November 1995. The standards specify, among other things, that two independent positioning sources are required for ECDIS [10].

A number of hydrographic offices are currently involved in producing ECDIS databases by digitizing existing paper charts (i.e., converting paper charts into digital computer files). However, this has the disadvantage that the paper charts are generally based on local datums. This means that correct datum shifts must be considered to ensure consistency [11]. In addition, the paper charts in some areas were based on old survey methods, which are far less accurate than the required standards. A complete resurvey of those areas might be required to overcome this problem.

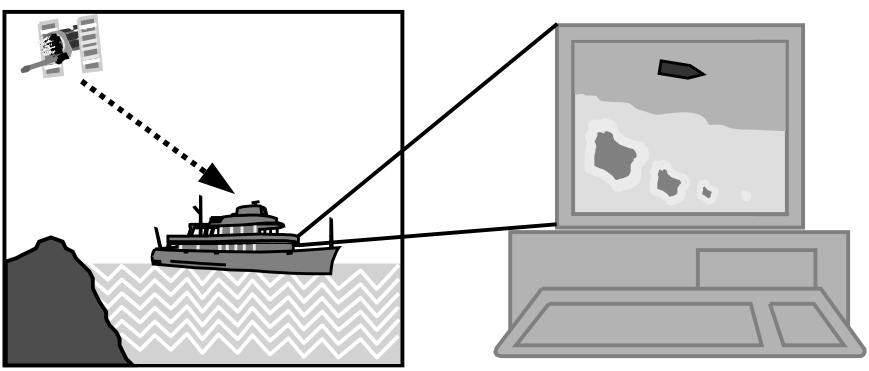

Figure 4.13 Marine nautical chart system.

4.7 Local arbitrary mapping systems

When surveying small areas, it is often more appropriate to employ a user-defined local plane coordinate system. In this case, the curved Earth's surface may be considered as a plane surface with a negligible amount of distortion. To establish a local coordinate system with GPS, a set of points with known coordinate values in both the WGS 84 and the local system must be available [5].

By comparing the coordinates of the common points (i.e., points with known coordinates in both the local system and the WGS 84 system), the transformation parameters may be obtained using the least squares technique. These transformation parameters will be used to transform all the new GPS-derived coordinates into the local coordinate system. It should be noted that the better the distribution of the common points, the better the solution (see Figure 4.14). The number of common points also plays an important role. The greater the number of common points, the better the solution [2].

Establishing a local coordinate system is usually done in either of two ways. One way is to supply the transformation parameters software

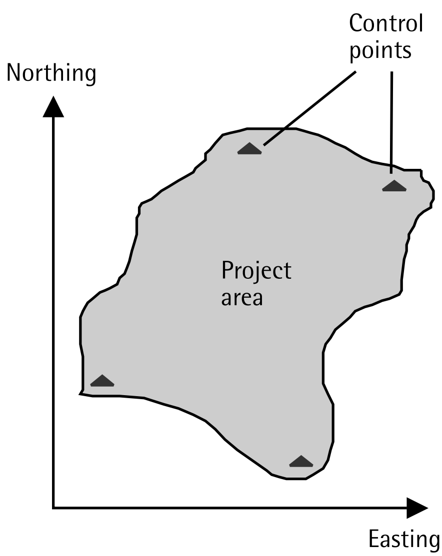

Figure 4.14 Local arbitrary mapping system.

(usually provided by the manufacturers of the GPS receivers) with the coordinates of the common points in both systems, if they are available. The software will then compute the transformation parameters, which once downloaded into the GPS data collector will be used to automatically transform all the new coordinates into the local coordinate system. Alternatively, if the coordinates of a set of points are known only in the local coordinate system, the user may occupy those points with the rover receiver to obtain their coordinates in the WGS 84 system. Real-time kinematic (RTK) GPS surveying (see Chapter 5) is normally used for this purpose. This allows the determination of the transformation parameters while in the field.

4.8 Height systems

The height (or elevation) of a point is defined as the vertical distance from the vertical datum to the point (Figure 4.15). As stated in Section 4.1, the geoid is often selected to be the vertical datum [2]. The height of a point above the geoid is known as the orthometric height. It can be positive or negative depending on whether the point is located above or below the geoid. Because they are physically meaningful, orthometric heights are often needed in practice and are usually found plotted on topographic maps [2].

In some cases, such as the case of GPS, the obtained heights are referred to the reference ellipsoid, not the geoid (Figure 4.15). Therefore, these heights are known as the ellipsoidal heights. An ellipsoidal height can also be positive or negative depending on whether the point is located above or below the surface of the reference ellipsoid. Unfortunately, ellipsoidal heights are purely geometrical and do not have any physical meaning. As such, the various Geomatics instruments (e.g., the total stations) cannot directly sense them.

The geoid-ellipsoid separation is known as the geoidal height or undulation (Figure 4.15). This distance can reach up to about 100m, and it can be positive or negative depending on whether the geoid is above or below the reference ellipsoid at a particular point [12]. Accurate information about the geoidal height leads to the determination of the orthometric height through the ellipsoidal height, and vice versa. Geoid models that describe the geoidal heights for the whole world have been developed.

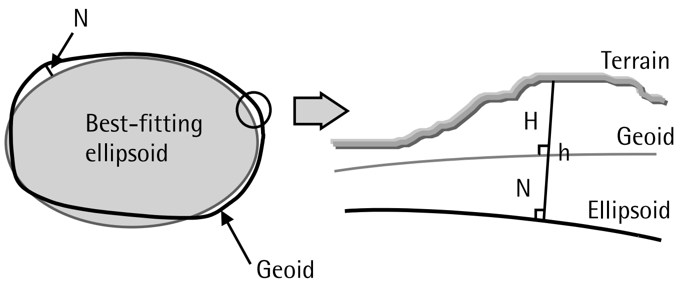

Figure 4.15 Height systems.

Unfortunately, these models do not have consistent accuracy levels everywhere, mainly because of the lack of local gravity data and the associated height information in some parts of the world [12]. Many GPS receivers and software packages have built-in geoid models for automatic conversion between orthometric and ellipsoidal heights. However, care must be taken when applying them, as they are usually low-accuracy models.

References

[1] Torge, W., *Geodesy,* New York: Walter de Gruyter, 1991.

[2] Vanicek, P., and E. J. Krakiwsky, *Geodesy: The Concepts*, 2nd ed., New York: North Holland, 1986.

[3] Leick, A., *GPS Satellite Surveying,* 2nd ed., New York: Wiley, 1995.

[4] National Geodetic Survey, "Geodetic Glossary," U.S. Department of Commerce, NOAA, Rockville, MD: U.S. Department of Commerce, NOAA, 1986.

[5] Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice*, 3rd ed., New York: Springer-Verlag, 1994.

[6] Boucher, C., and Z. Altamimi, "International Terrestrial Reference Frame," *GPS World,* Vol. 7, No. 9, September 1996, pp. 71–74.

[7] Malys, S., et al., "Refinements to the World Geodetic System 1984," *Proc. ION GPS-97, 10th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Kansas City, MO, September 16–19, 1997, pp. 841–850.

[8] Craymer, M., R. Ferland, and R. Snay, "Realization and Unification of NAD83 in Canada and the U.S. Via the ITRF," *Proc. Intl. Symp. Intl. Assoc.*

of Geodesy, Sec. 2, "Towards an Integrated Geodetic Observing System (IGGOS)," Munich, Germany, October 5–9, 1998.

[9] Krakiwsky, E. J., "Conformal Map Projections in Geodesy," Department of Geodesy and Geomatics Engineering, L.N. No. 37, University of New Brunswick, Fredericton, New Brunswick, Canada, 1973.

[10] Alexander, L., "What Is an ENC?" *Hydro International*, Vol. 2, No. 5, July/August 1998, pp. 60–63.

[11] Casey, M. J., and P. Kielland, "Electronic Charts and GPS," *GPS World*, Vol. 1, No. 4, July/August 1990, pp. 56–59.

[12] Schwarz, K. P., and M. G. Sideris, "Heights and GPS," *GPS World*, Vol. 4, No. 2, February 1993, pp. 50–56.

5

GPS Positioning Modes

Positioning with GPS can be performed by either of two ways: point positioning or relative positioning. GPS point positioning employs one GPS receiver that measures the code pseudoranges to determine the user's position instantaneously, as long as four or more satellites are visible at the receiver. The expected horizontal positioning accuracy from the civilian C/A-code receivers has gone down from about 100m (2 drms) when selective availability was on, to about 22m (2 drms) in the absence of selective availability [1]. GPS point positioning is used mainly when a relatively low accuracy is required. This includes recreation applications and low-accuracy navigation.

GPS relative positioning, however, employs two GPS receivers simultaneously tracking the same satellites. If both receivers track at least four common satellites, a positioning accuracy level of the order of a subcentimeter to a few meters can be obtained [2]. Carrier-phase or/and pseudorange measurements can be used in GPS relative positioning, depending on the accuracy requirements. The former provides the highest possible accuracy. GPS relative positioning can be made in either real-time or post-mission modes. GPS relative positioning is used for high-accuracy applications such as surveying and mapping, GIS, and precise navigation.

5.1 GPS point positioning

GPS point positioning, also known as standalone or autonomous positioning, involves only one GPS receiver. That is, one GPS receiver simultaneously tracks four or more GPS satellites to determine its own coordinates with respect to the center of the Earth (Figure 5.1). Almost all of the GPS receivers currently available on the market are capable of displaying their point-positioning coordinates.

To determine the receiver's point position at any time, the satellite coordinates as well as a minimum of four ranges to four satellites are required [2]. The receiver gets the satellite coordinates through the navigation message, while the ranges are obtained from either the C/A-code or the P(Y)-code, depending on the receiver type (civilian or military). As mentioned before, the measured pseudoranges are contaminated by both the satellite and receiver clock synchronization errors. Correcting the satellite clock errors may be done by applying the satellite clock correction in the navigation message; the receiver clock error is treated as an additional unknown parameter in the estimation process [2]. This brings the total number of unknown parameters to four: three for the receiver coordinates and one for the receiver clock error. This is the reason why at least four satellites are needed. It should be pointed out that if more than four satellites are tracked, the so-called least-squares estimation or Kalman filtering technique is applied [2–4]. As the satellite coordinates are given in the WGS 84 system, the obtained receiver coordinates will be in the WGS 84 system as

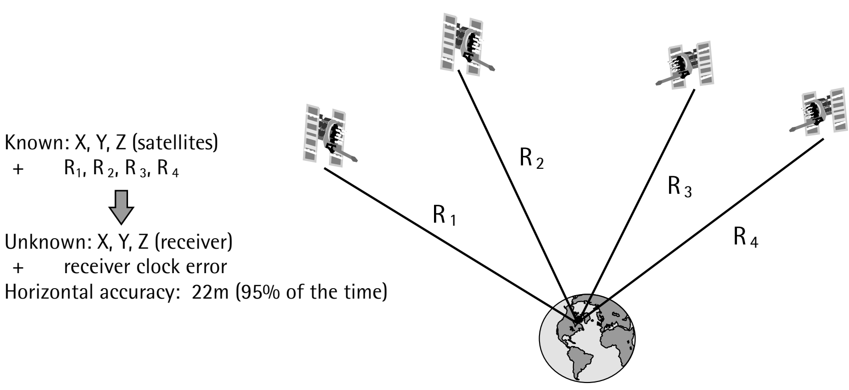

Figure 5.1 Principle of GPS point positioning.

well, as explained in Chapter 4. However, most GPS receivers provide the transformation parameters between WGS 84 and many local datums used around the world.

5.2 GPS relative positioning

GPS relative positioning, also called differential positioning, employs two GPS receivers simultaneously tracking the same satellites to determine their relative coordinates (Figure 5.2). Of the two receivers, one is selected as a reference, or base, which remains stationary at a site with precisely known coordinates. The other receiver, known as the rover or remote receiver, has its coordinates unknown. The rover receiver may or may not be stationary, depending on the type of the GPS operation.

A minimum of four common satellites is required for relative positioning. However, tracking more than four common satellites simultaneously would improve the precision of the GPS position solution [2]. Carrier-phase and/or pseudorange measurements can be used in relative positioning. A variety of positioning techniques are used to provide a

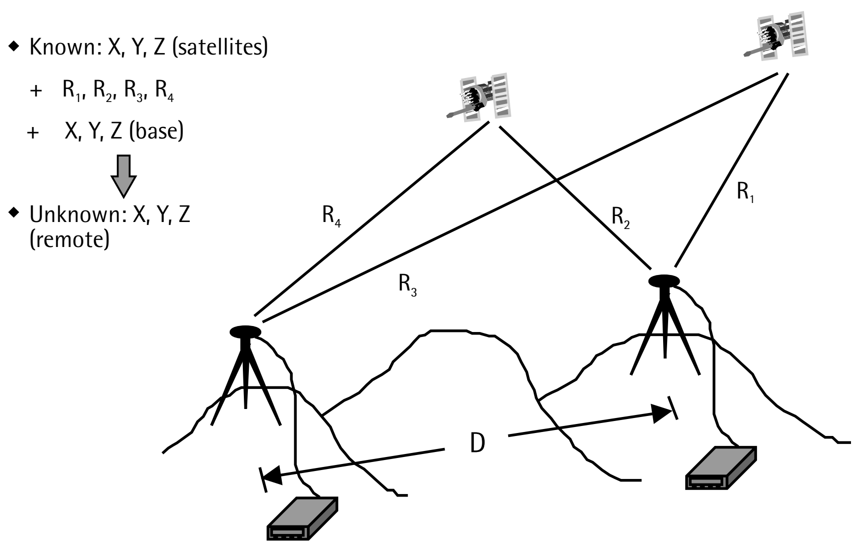

Figure 5.2 Principle of GPS relative positioning.

postprocessing (postmission) or real-time solution. Details of the commonly used relative positioning techniques are given in Sections 5.3 to 5.7. GPS relative positioning provides a higher accuracy than that of autonomous positioning. Depending on whether the carrier-phase or the pseudorange measurements are used in relative positioning, an accuracy level of a subcentimeter to a few meters can be obtained. This is mainly because the measurements of two (or more) receivers simultaneously tracking a particular satellite contain more or less the same errors and biases [5]. The shorter the distance between the two receivers, the more similar the errors. Therefore, if we take the difference between the measurements of the two receivers (hence the name "differential positioning"), the similar errors will be removed or reduced.

5.3 Static GPS surveying

Static GPS surveying is a relative positioning technique that depends on the carrier-phase measurements [2]. It employs two (or more) stationary receivers simultaneously tracking the same satellites (see Figure 5.3). One receiver, the base receiver, is set up over a point with precisely known coordinates such as a survey monument (sometimes referred to as the known point). The other receiver, the remote receiver, is set up over a point whose coordinates are sought (sometimes referred to as the unknown point). The base receiver can support any number of remote receivers, as long as a minimum of four common satellites is visible at both the base and the remote sites.

In principle, this method is based on collecting simultaneous measurements at both the base and remote receivers for a certain period of time, which, after processing, yield the coordinates of the unknown point. The observation, or occupation, time varies from about 20 minutes to a few hours, depending on the distance between the base and the remote receivers (i.e., the baseline length), the number of visible satellites, and the satellite geometry. The measurements are usually taken at a recording interval of 15 or 20 seconds, or one sample measurement every 15 or 20 seconds.

After completing the field measurements, the collected data is downloaded from the receivers into the PC for processing. Different processing options may be selected depending on the user requirements, the baseline length, and other factors. For example, if the baseline is relatively short, say,

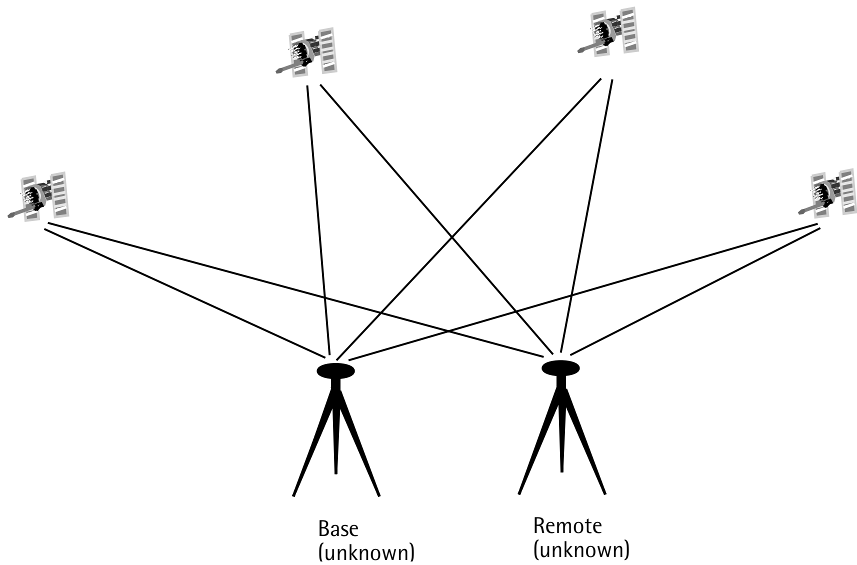

Figure 5.3 Static GPS surveying.

15 or 20 km, resolving the ambiguity parameters would be a key issue to ensure high-precision positioning. As such, in this case the option of fixing the ambiguity parameters should be selected. In contrast, if the baseline is relatively long, a user may select the ionosphere-free linear combination option to remove the majority of the ionospheric error (see Chapter 2 for details on the various linear combinations of the GPS observables). This is because the ambiguity parameters may not be fixed reliably at the correct integer values. For very long baselines, for example, over 1,000 km, it is recommended that the user processes the data with one of the scientific software packages available, such as the BERENSE software developed by the University of Bern, rather than a commercial software package. The precise ephemeris should also be used in this case, as the effect of the orbital errors will be considerably different at the two ends of the baseline.

Static GPS surveying with the carrier-phase measurements is the most accurate positioning technique. This is mainly due to the significant change in satellite geometry over the long observation time span. Although both the single- and dual-frequency receivers can be used for static positioning, the latter is often used, especially for baselines exceeding 20 km.

The expected accuracy from a geodetic quality receiver is typically 5 mm + 1 ppm (rms), ppm for parts per million and rms for root-mean-square. That is, for a 10-km baseline, for example, the expected accuracy of the static GPS surveying is 1.5 cm (rms). Higher accuracy may be obtained by, for example, applying the precise ephemeris.

5.4 Fast (rapid) static

Fast, or rapid, static surveying is a carrier-phase–based relative positioning technique similar to static GPS surveying. That is, it employs two or more receivers simultaneously tracking the same satellites. However, with rapid static surveying, only the base receiver remains stationary over the known point during the entire observation session (see Figure 5.4). The rover receiver remains stationary over the unknown point for a short period of time only, and then moves to another point whose coordinates are sought [2]. Similar to the static GPS surveying, the base receiver can support any number of rovers.

This method is suitable when the survey involves a number of unknown points located in the vicinity (i.e., within up to about 15 km) of a known point. The survey starts by setting up the base receiver over the

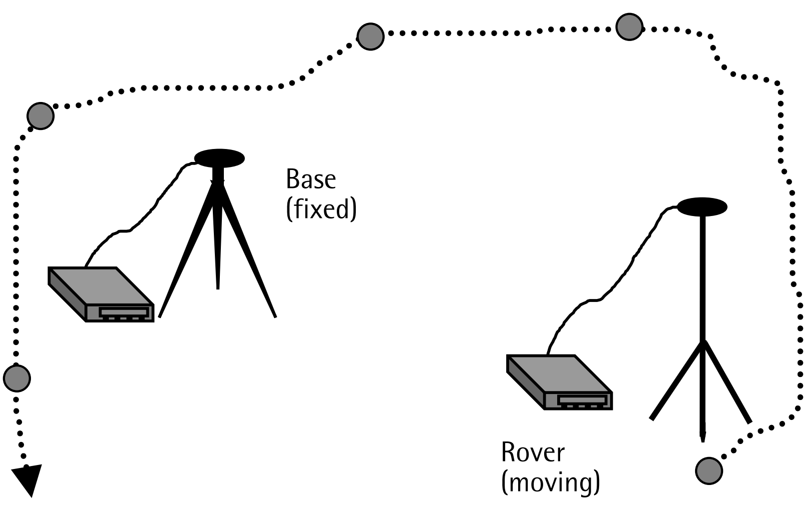

Figure 5.4 Fast (rapid) static GPS surveying.

known point, while setting up the rover receiver over the first unknown point (Figure 5.4). The base receiver remains stationary and collects data continuously. The rover receiver collects data for a period of about 2 to 10 minutes, depending on the distance to the base as well as the satellite geometry [2]. Once the rover receiver has collected the data, the user moves to the following point with unknown coordinates and repeats the procedures. It should be pointed out that, while moving, the rover receiver may be turned off. Due to the relatively short occupation time for the rover receiver, the recording interval is reduced to 5 seconds.

After collecting and downloading the field data from both receivers, the PC software is used for data processing. Depending on whether enough common data was collected, the software may output a fixed solution, which indicates that the ambiguity parameters were fixed at integer values (see Chapter 6 for details). Otherwise, a float solution is obtained, which means that the software was unable to fix ambiguity parameters at integer values (i.e., only real-valued ambiguity parameters were obtained). This problem occurs mainly when the collected GPS data is insufficient. A fixed solution means that the positioning accuracy is at the centimeter level, while the float solution means that the positioning accuracy is at the decimeter or submeter level. Although both the single- and dual-frequency receivers can be used for fast static surveying, the probability of getting a fixed solution is higher with the latter.

5.5 Stop-and-go GPS surveying

Stop-and-go surveying is another carrier-phase-based relative positioning technique. It also employs two or more GPS receivers simultaneously tracking the same satellites (Figure 5.5): a base receiver that remains stationary over the known point and one or more rover receivers [2]. The rover receiver travels between the unknown points, and makes a brief stop at each point to collect the GPS data. The data is usually collected at a 1- to 2-second recording rate for a period of about 30 seconds per each stop. Similar to the previous methods, the base receiver can support any number of rovers. This method is suitable when the survey involves a large number of unknown points located in the vicinity (i.e., within up to 10–15 km) of a known point.

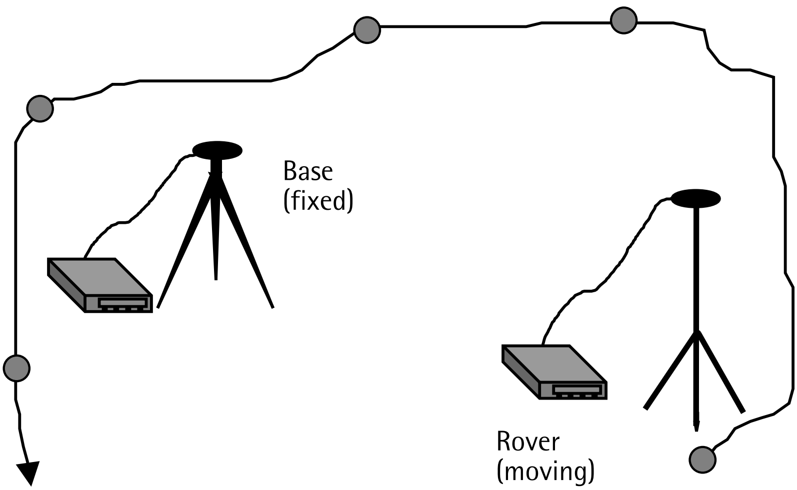

Figure 5.5 Stop-and-go GPS surveying.

The survey starts by first determining the initial integer ambiguity parameters, a process known as receiver initialization. This could be done by various methods, discussed in the next chapter. Once the initialization is performed successfully, centimeter-level positioning accuracy can be obtained instantaneously. This is true as long as there is a minimum of four common satellites simultaneously tracked by both the base and the rover receivers at all times. If this condition is not fulfilled at any moment during the survey, the initialization process must be repeated to ensure centimeter-level accuracy.

Following the initialization, the rover moves to the first unknown point. After collecting about 30 seconds of data, the rover moves, without being switched off, to the second point and the procedures are repeated. It is of utmost importance that at least four satellites are tracked, even during the move; otherwise the initialization process must be repeated again by, for example, reoccupying the previous point. Some manufacturers, for example, Ashtech Inc., recommend the reoccupation of the first point at the end of the survey. This turned out to be very useful in obtaining a fixed solution provided that the processing software has the forward and backward processing functions. Once the data is collected and downloaded, PC software is used to process it. Some software packages have the forward and backward processing functions, which help in obtaining a fixed solution, or centimeter-level accuracy. Both single- and dual-frequency receivers may use the stop-and-go surveying method.

A special case of stop-and-go surveying is known as kinematic GPS surveying. Both methods are the same in principle; however, the latter requires no stops at the unknown points. The positional accuracy is expected to be higher with the stop-and-go surveying, as the errors are averaged out when the receiver stops at the unknown points.

5.6 RTK GPS

RTK surveying is a carrier phase–based relative positioning technique that, like the previous methods, employs two (or more) receivers simultaneously tracking the same satellites (Figure 5.6). This method is suitable when: (1) the survey involves a large number of unknown points located in the vicinity (i.e., within up to about 10–15 km) of a known point; (2) the coordinates of the unknown points are required in real time; and (3) the line of sight, the propagation path, is relatively unobstructed [6]. Because

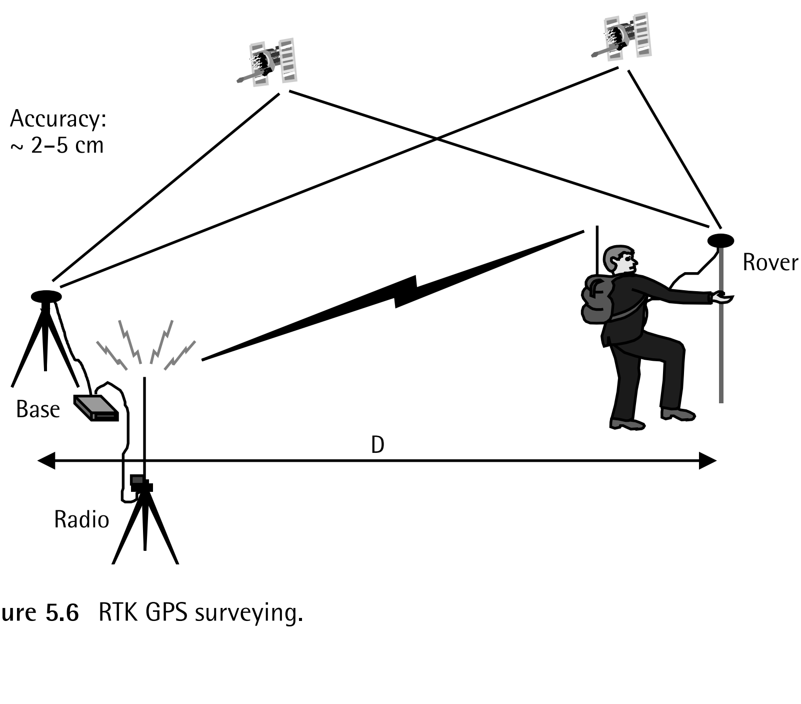

Figure 5.6 RTK GPS surveying.

of its ease of use as well as its capability to determine the coordinates in real time, this method is the preferred method by many users.

In this method, the base receiver remains stationary over the known point and is attached to a radio transmitter (Figure 5.6). The rover receiver is normally carried in a backpack and is attached to a radio receiver. Similar to the conventional kinematic GPS method, a data rate as high as 1 Hz (one sample per second) is required. The base receiver measurements and coordinates are transmitted to the rover receiver through the communication (radio) link [7, 8]. The built-in software in a rover receiver combines and processes the GPS measurements collected at both the base and the rover receivers to obtain the rover coordinates.

The initial ambiguity parameters are determined almost instantaneously using a technique called on-the-fly (OTF) ambiguity resolution, to be discussed in the next chapter. Once the ambiguity parameters are fixed to integer values, the receiver (or its handheld computer controller) will display the rover coordinates right in the field. That is, no postprocessing is required. The expected positioning accuracy is of the order of 2 to 5 cm (rms). This can be improved by staying over the point for a short period of time, for example, about 30 seconds, to allow for averaging the position. The computed rover coordinates for the entire survey may be stored and downloaded at a later time into CAD software for further analysis. This method is used mainly, but not exclusively, with dual-frequency receivers.

Under the same conditions, the positioning accuracy of the RTK method is slightly degraded compared with that of the conventional kinematic GPS method. This is mainly because the time tags (or time stamps) of the conventional kinematic data from both the base and the rover match perfectly in the processing. With RTK, however, the base receiver data reaches the rover after some delay (or latency). Data latency occurs as a result of formatting, packetizing, transmitting, and decoding the base data [7]. To match the time tag of the rover data, the base data must be extrapolated, which degrades the positioning accuracy.

5.7 Real-time differential GPS

Real-time differential GPS (DGPS) is a code-based relative positioning technique that employs two or more receivers simultaneously tracking the same satellites (Figure 5.7). It is used when a real-time meter-level accuracy

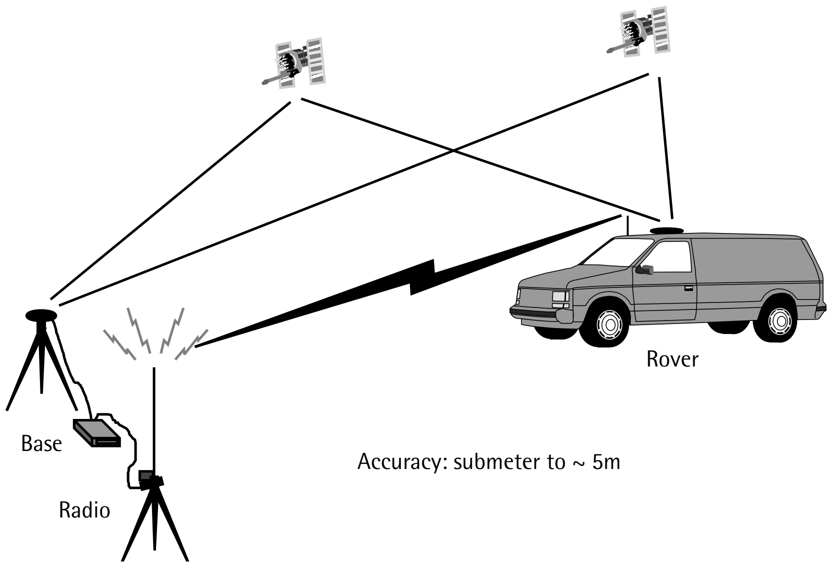

Figure 5.7 Real-time differential GPS operation.

is enough. The method is based on the fact that the GPS errors in the measured pseudoranges are essentially the same at both the base and the rover, as long as the baseline length is within a few hundred kilometers.

As before, the base receiver remains stationary over the known point. The built-in software in the base receiver uses the precisely known base coordinates as well as the satellite coordinates, derived from the navigation message, to compute the ranges to each satellite in view. The software further takes the difference between the computed ranges and the measured code pseudoranges to obtain the pseudorange errors (or DGPS corrections). These corrections are transmitted in a standard format called Radio Technical Commission for Maritime Service (RTCM) to the rover through a communication link (see Chapter 8 for more about RTCM). The rover then applies the DGPS corrections to correct the measured pseudoranges at the rover. Finally, the corrected pseudoranges are used to compute the rover coordinates.

The accuracy obtained with this method varies between a submeter and about 5m, depending on the base-rover distance, the transmission rate of the RTCM DGPS corrections, and the performance of the C/A-code receivers [2]. Higher accuracy is obtained with short base-rover separation, high transmission rate, and carrier-smoothed C/A-code ranges. With the termination of selective availability, the data rate could be reduced to 10 seconds or lower without noticeable accuracy degradation. Further accuracy improvement could be achieved if the receivers are capable of storing the raw pseudorange measurements, which could be used at a later time in the postprocessing mode. As the real-time DGPS is widely used, some governmental agencies as well as private firms are providing the RTCM DGPS corrections either at no cost or at certain fees. More about these services will be given in Chapter 7.

5.8 Real time versus postprocessing

The term *real time* means that the results are obtained almost instantaneously, while the term *postprocessing* means that the measurements are collected in the field and processed at a later time to obtain the results. Each of these modes has some advantages and some disadvantages.

The first advantage of the real-time mode is that the results as well as the accuracy measures (or quality control) are obtained while in the field. This is especially important for RTK surveying, as the user would not store the displayed coordinates unless the ambiguity parameters are shown to be fixed at integer values and centimeter-level accuracy is achieved. This leads to a higher productivity compared with the postprocessing mode, as only enough GPS data to obtain a fixed solution is collected. In addition, processing the GPS data is done automatically in the field by the built-in software. This means that no postprocessing software training is required. The user also saves the time spent in data processing.

There are, however, some advantages in the postprocessing mode as well. The first of these is that more accurate results are generally obtained with the postprocessing mode. One reason for this is more flexibility in editing and cleaning of the collected GPS data. As well, there is no accuracy degradation due to data latency, as explained in Section 5.7. Another important advantage is that the communication link problems, such as the relatively unobstructed line-of-sight requirement, are avoided. In some cases, the input parameters, such as the base station coordinates or the antenna height, may contain some errors, which lead to errors in the computed rover coordinates. These errors can be corrected in the postprocessing mode, while they cannot be completely corrected in the real-time mode.

5.9 Communication (radio) link

RTK and real-time DGPS operations require a communication, or radio, link to transmit the information from the base receiver to the rover receiver (Figures 5.6 and 5.7). RTK data are typically transmitted at a baud rate of 9,600, while the DGPS corrections are typically transmitted at 200 Kbps. A variety of radio links that use different parts of the electromagnetic spectrum are available to support such operations. The spectrum parts mostly used in practice are the low/medium frequency (LF/MF) bands (i.e., 30 kHz to 3 MHz) and the very high and ultrahigh frequency (VHF/UHF) bands (i.e., 30 MHz to 3 GHz) [7, 8]. Often, GPS users utilize their own dedicated radio links to transmit base station information.

Dedicated ground-based GPS radio links are mostly established using the VHF/UHF band. Radio links in this band provide line-of-sight coverage, with the ability to penetrate into buildings and other obstructions. One example of such a radio link is the widely used RFM96W from Pacific Crest Corporation, which is available in different models based on the supported frequencies in the VHF/UHF band. This type of radio link requires a license to operate. A new radio link that was recently produced by the same company is called the Position Data Link (PDL) (see Figure 5.8). PDL allows for a baud rate of 19,200, and is characterized by low power consumption and enhanced user interface. Another example is the license-free spread-spectrum radio transceiver, which operates in the 902–928 MHz portion of the UHF band (Figure 5.8). This radio link has coverage of 1–5 km and 3–15 km in urban and rural areas, respectively. More recently, some GPS manufacturers adopted cellular technology, the digital Personal Communication Services (PCS), as an alternative communication link. In the near future, it is expected that the third-generation (3G) wideband digital networks will be used extensively as the GPS communication link. The 3G technology uses common global standards, which reduces the service cost. In addition, this technology allows the devices to be kept in the "on" position all the time for data transmission or reception, while the subscribers pay for the packets of data they transmit/receive.

Figure 5.8 Examples of radio modems. (Courtesy of Magellan Corporation.)

It should be pointed out that obstructions along the propagation path, such as buildings and terrain, attenuate the transmitted signal, which leads to limited signal coverage. The transmitted signal attenuation may also be caused by ground reflection (multipath), the transmitting antenna, and other factors [7]. To increase the coverage of a radio link, a user may

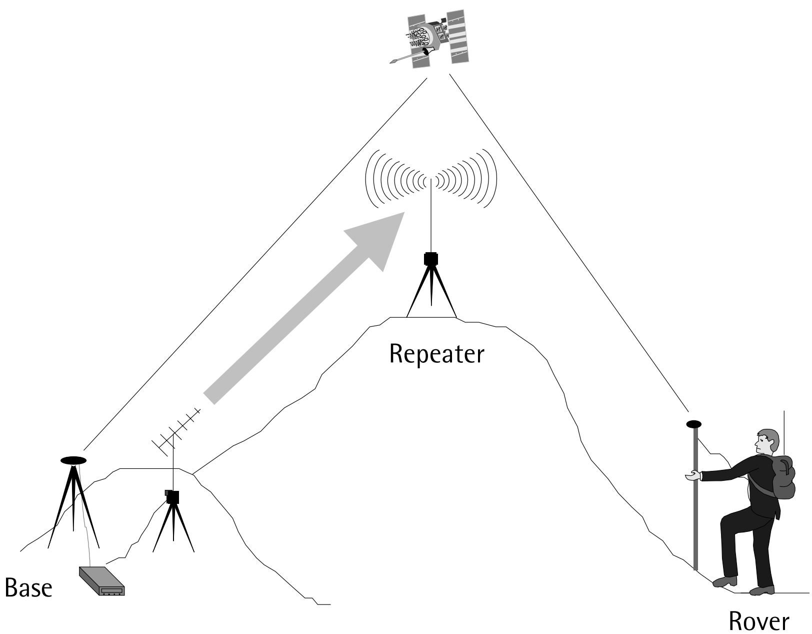

Figure 5.9 Use of repeaters to increase radio coverage.

employ a power amplifier or high-quality coaxial cables, or he or she may increase the height of the transmitting/receiving radio antenna. If a user employs a power amplifier, however, he or she should be cautioned against signal overload, which usually occurs when the transmitting and the receiving radios are very close to each other [7].

A user may also increase the signal coverage by using a repeater station. In this case, it might be better to use a unidirectional antenna, such as a Yagi, at the base station and an omnidirectional antenna at the repeater station (see Figure 5.9) [8].

References

[1] Shaw, M., K. Sandhoo, and D. Turner, "Modernization of the Global Positioning System," *GPS World*, Vol. 11, No. 9, September 2000, pp. 36–44.

[2] Hoffmann-Wellenhof, B., H. Lichtenegger, and J. Collins, *Global Positioning System: Theory and Practice,* 3rd ed., New York: Springer-Verlag, 1994.

[3] Leick, A., *GPS Satellite Surveying,* 2nd ed., New York: Wiley, 1995.

[4] Levy, L. J., "The Kalman Filter: Navigation's Integration Workhorse," *GPS World*, Vol. 8, No. 9, September 1997, pp. 65–71.

[5] Langley, R. B., "The GPS Observables," *GPS World*, Vol. 4, No. 4, April 1993, pp. 52–59.

[6] Langley, R. B., "RTK GPS," *GPS World,* Vol. 9, No. 9, September 1998, pp. 70–76.

[7] Langley, R. B., "Communication Links for DGPS," *GPS World*, Vol. 4, No. 5, May 1993, pp. 47–51.

[8] Pacific Crest Corporation, "The Guide to Wireless GPS Data Links," 2000.

6

Ambiguity-Resolution Techniques

The previous chapter showed that centimeter-level positioning accuracy could be achieved with the carrier-phase observables in the relative positioning mode. A prerequisite to this, however, is the successful determination of the initial integer ambiguity parameters (in fact, the integer double-difference ambiguity parameters). This process is commonly known as ambiguity resolution. Resolving the ambiguity parameters correctly is equivalent to having very precise ranges to the satellites, which leads to high-accuracy positioning [1].

The ambiguity parameters are initially determined as part of the least-squares, or Kalman filtering, solution [2, 3]. Unfortunately, however, neither method can directly determine the integer numbers of the ambiguity parameters. What can be obtained are the real-valued numbers along with their uncertainty parameters (so-called covariance matrix) only. These real-valued numbers are in fact difficult to separate from the baseline solution [4]. As such, since we know in advance that the ambiguity parameters are integer numbers, it becomes clear that further analysis is required.

Traditionally, high-precision GPS relative positioning with carrier-phase observables was carried out using long observational time spans (typically a few hours). This allows for the receiver-satellite geometry to change considerably, which helps in separating the ambiguity parameters from the baseline solution. As such, even though the least-squares solution would contain real-valued numbers for the ambiguity parameters, they were very close to integer values. Consequently, the correct integer values were simply obtained by rounding off the real-valued numbers to the nearest integers [4]. Another least-squares adjustment was then to be carried out, considering the integer-valued ambiguity parameters as known values while the baseline components are unknowns. It is clear that, although this method is capable of determining the correct integer values of the ambiguity parameters, it is time-consuming. As such, the use of this method is currently limited to long baselines in the static mode.

Various methods have been developed to overcome the limitation of the previous method (i.e., the use of long observational time spans). One such method is to use a known baseline (i.e., the coordinates of its end points are accurately known), which might be available within the project area. The ambiguity parameters are determined by simply occupying the two end points of the known baseline with the base and the rover receivers for a short period of time. This process is commonly known as receiver initialization. Following receiver initialization, the rover receiver can move to the points to be surveyed. With this method, the receiver uses the ambiguity parameters determined during the initialization to solve for the coordinates of the new points. As mentioned in Chapter 2, the initial integer number of cycles (the ambiguity parameter) remains constant over time, even if the receiver is in motion, as long as no cycle slips have occurred. In other words, it is necessary that the receivers be kept "on" all the time and that at least four common satellites are tracked at any moment. An alternative initialization method is known as the antenna swap method, which can be used when no known baseline is available within the project area. This method, which was introduced by Dr. Ben Remondi in 1986, is based on exchanging the antennas between the base and the rover while tracking at least four satellites. More details on this method are given later. Both the known baseline and the antenna swap methods are more suitable for kinematic positioning in the postprocessing mode.

These three methods are suitable for non-real-time applications, with which the data are collected in the field and then postprocessed at later times. RTK positioning, however, requires that the integer ambiguity parameters be determined while the receiver is in motion, or on the fly [5]. Resolving the ambiguities on the fly, often called on-the-fly ambiguity resolution, is different from the ambiguity-resolution techniques mentioned earlier in the sense that the initialization is performed in the field using very short observational time spans. Due to the high altitudes of the GPS satellites, the receiver-satellite geometry changes very slowly over time. As such, a short time span of data causes some difficulties in resolving the ambiguities. Fortunately, a more advanced technique has been developed to overcome this limitation; this technique is discussed later.

6.1 Antenna swap method

The antenna swap is a method used for a fast and reliable determination of the initial ambiguity parameters (i.e., initialization) in the postprocessing mode [6]. This method is used mainly when single-frequency receivers are used for kinematic surveying, although it can be used with dual frequency as well.

The initialization procedures with the antenna swap method start by setting up the reference (base) receiver over the known point while setting up the rover within a few meters from it (Figure 6.1). About 1-minute simultaneous GPS data is then to be collected at both receivers. Usually, a data rate as high as 1 or 2 seconds is used. Once the data is collected, the two antennas (with the two GPS receivers connected to them) are exchanged (see Figure 6.1). This is done without changing the original antenna heights. Care must be taken to keep tracking to a minimum of four, preferably five, common satellites. With this new setup, another simultaneous 1-minute GPS data, at the previous rate, is collected by both receivers. After this step, the receivers are returned to their original setup, which ends the initialization procedures.

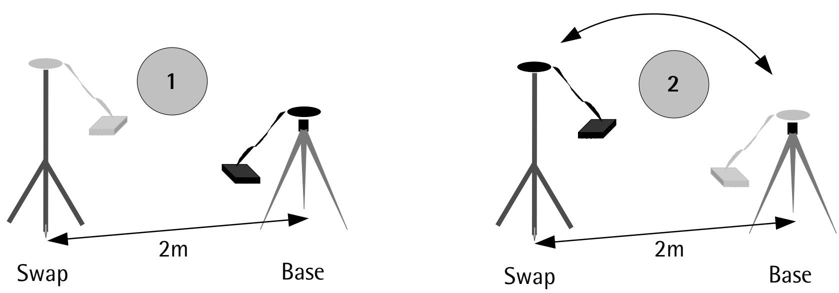

Figure 6.1 Antenna swap method.

Once the initialization is performed, the base receiver must be kept stationary over the known point while the rover moves between the points to be surveyed, as discussed before in the kinematic method. After finishing the fieldwork, the data is downloaded into the PC processing software, which will first use the initialization data to determine the initial ambiguity parameters. Once determined, the software will use these parameters to determine the coordinates of the survey points at centimeter-level accuracy. It should be pointed out that a shorter observational time span would be enough for the receiver initialization.

6.2 On-the-fly ambiguity resolution

On-the-fly (OTF) ambiguity resolution is an advanced technique developed recently to determine the initial integer ambiguity parameters without static initialization (i.e., while the rover receiver is moving). This technique may be applied with either single- or dual-frequency data. However, resolving the ambiguities is faster and more reliable with dual-frequency data. It is used mainly for, but not restricted to, real-time kinematic operations.

Several OTF techniques have been developed over the past several years. Only one method is summarized here [4]. The base and rover measurements are combined in the double differenced mode and an initial adjustment by, for example, the least squares or Kalman filtering technique, is then performed. The outcome of this initial adjustment is an initial rover position along with estimates (real values) for the ambiguity parameters and their uncertainty values, or the covariance matrix.

The covariance matrix can be represented geometrically to form a region, known as the confidence region, around the estimated real-value ambiguity parameters [4]. The size of such a confidence region depends on the size of the uncertainty parameters of the ambiguities as well as the used probability level. The larger the uncertainty values and/or the probability level, the larger the size of the confidence region. The confidence region takes the shape of an ellipse if the number of the estimated parameters is two, and an ellipsoid if it is three. If the number of estimated parameters is more than three, which is the case if the number of satellites is more than four, a confidence region of a hyperellipsoid is obtained.

Generally, a confidence region of a hyperellipsoid is formed around the estimated real-valued ambiguity parameters. Such a hyperellipsoid contains the likely integer ambiguity parameters at a certain probability level. For example, if a probability value of 99% is used to scale the hyperellipsoid, it means that there is a 99% chance that the true integer ambiguity parameters are located inside that hyperellipsoid. Since we know in advance that the ambiguity parameters must be integer numbers, we may draw (mathematically) gridlines that intersect at integer values inside the hyperellipsoid. If the grid spacing is selected to be equal to one carrier cycle, then the likely integer ambiguity parameters would be represented by one of the points of intersection inside the hyperellipsoid. Figure 6.2 simplifies this, using a 2-D case as an example. The hyperellipsoid is then used for searching the likely integer values for ambiguity parameters (i.e., all the points inside the confidence region with integer values). Based on statistical evaluation, only one point is selected as the most likely candidate for the integer ambiguity parameters. Once the ambiguities are correctly resolved, a final adjustment is performed to obtain the rover coordinates at centimeter-level accuracy. It should be pointed out that the OTF technique, although designed mainly for resolving the ambiguity parameters in real time, could also be used in the non-real-time mode.

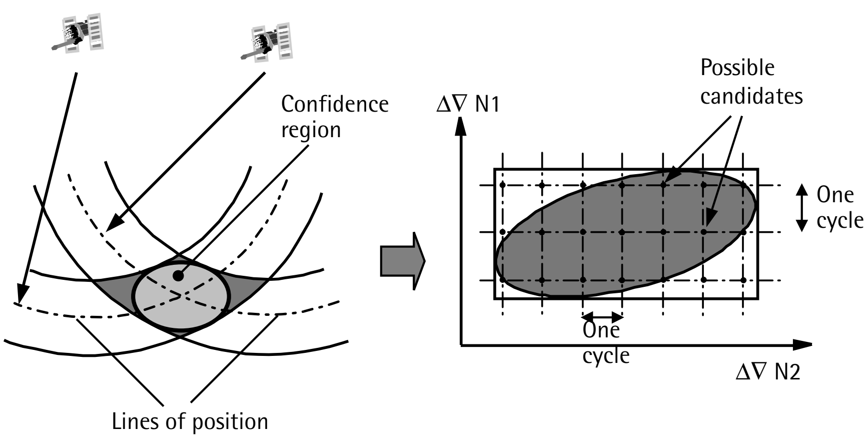

Figure 6.2 OTF ambiguity resolution.

References

[1] Joosten, P., and C. Tiberius, "Fixing the Ambiguities: Are You Sure They're Right?" *GPS World*, Vol. 11, No. 5, May 2000, pp. 46–51.

[2] Leick, A., *GPS Satellite Surveying*, 2nd ed., New York: Wiley, 1995.

[3] Levy, L. J., "The Kalman Filter: Navigation's Integration Workhorse," *GPS World,* Vol. 8, No. 9, September 1997, pp. 65–71.

[4] Teunissen, P. J. G., P. J. de Jonge, and C. C. J. M. Tiberius, "A New Way to Fix Carrier-Phase Ambiguities," *GPS World*, Vol. 6, No. 4, April 1995, pp. 58–61.

[5] DeLoach, S. R., D. Wells, and D. Dodd, "Why On-the-Fly?" *GPS World*, Vol. 6, No. 5, May 1995, pp. 53–58.

[6] Remondi, B., "Performing Centimeter-Level Surveys in Seconds with GPS Carrier Phase: Initial Results," *Proc. 4th Intl. Geodetic Symposium on Satellite Positioning*, Austin, TX, April 28–May 2, 1986, Vol. 2, pp. 1229–1249.

7

GPS Data and Correction Services

GPS users employ differential techniques to achieve the highest possible positional accuracy. A straightforward manner of doing this is to use two GPS receivers, a base and a rover, as discussed in Chapter 5. However, this may not be cost effective in many instances. An alternative, which could significantly reduce the survey cost without degrading the positional accuracy, is to use one of the available GPS data and correction services. If such a service is available within the project area, a GPS user would only need one receiver to be used as a rover; no base receiver is required. A number of various GPS services are currently available with various levels of accuracy and cost. Some services are even provided with no user fees required.

For high-accuracy static positioning, primarily for establishing new control points, GPS users may use one of the highly precise permanent GPS reference station networks established by several organizations around the world. These services are currently available free of charge at the global level [e.g., the International GPS Service for Geodynamics (IGS) network] as well as at the regional level [e.g., the Continuously Operating Reference Station (CORS) network in the United States] [1–3]. The Canadian Active Control System (CACS) is another regional GPS service, which is available to users at nominal fees [4]. The reference stations within these systems are operating on a continuous basis, and provide access to the modern reference frames, such as the ITRF and the improved NAD 83.

Some countries around the world have established networks of reference stations around their coastal areas, which continuously broadcast real-time DGPS corrections in a special format known as the RTCM format (see Chapter 8). This service is primarily designed to enhance the safety of marine navigation, but is available at no cost to all users within the coverage area. Although this service is available primarily free of charge, it requires a beacon receiver connected to a GPS receiver that accepts the RTCM corrections [5]. GPS receivers that accept the RTCM corrections are commonly known as differential-ready GPS receivers. The accuracy obtained from such a service is usually in the range of a submeter to a few meters.

At the commercial level, two real-time DGPS correction services are widely used. One broadcasts the DGPS corrections through FM broadcast stations, while the other through communication satellites [6–8]. The DGPS corrections from the two systems are based on the DGPS corrections from a network of ground reference stations that cover a wide area, for example, a continent. The system is therefore known as the wide-area differential GPS (WADGPS). Both systems require a special receiver to decode the DGPS correction information, which would be interfaced to the GPS rover receiver to output positional information at the meter-level accuracy. WADGPS systems have several advantages over conventional single-station DGPS systems, including coverage of large, inaccessible regions using fewer reference stations [9].

Multisite, real-time, carrier-phase–based RTK positioning at subdecimeter-level accuracy is a new service that is currently being developed [10]. With this service, as little as four GPS reference receivers could cover an entire city or even a number of adjacent small cities. Advancements in wireless communication and the Internet are expected to make this service very promising [11]. This chapter summarizes each of the GPS services.

7.1 Data service

Several organizations around the world have established highly precise permanent GPS reference station networks, which are used for various geodetic purposes. The GPS data collected at these reference stations is made available to users, and could be used for high-accuracy static-positioning operations, such as establishing new control points. One such organization is the IGS, which is a service with international multiagency membership to support global geodetic and geophysical activities [1]. The International Association of Geodesy (IAG) has formally recognized the IGS since 1993. The IGS service is accomplished through a global network of 250 tracking stations (as of April 2001) equipped with continuously operating dual-frequency receivers and a number of data and analysis centers. Figure 7.1 shows the current IGS tracking stations.

The raw GPS data collected at each tracking station is formatted in a standard format, called the Receiver Independent Exchange (RINEX) (see details in Chapter 8), by the operational data centers. The formatted data is then collected by the global data centers for archiving and providing on-line accessibility. At this stage, the analysis centers use the on-line data to create a number of products. These products include GPS precise ephemeris, satellite and tracking station clock information, tracking station coordinates and velocities, and Earth rotation parameters. There are different quality levels for the precise ephemeris, depending on the time of availability (see Chapter 8).

Figure 7.1 IGS tracking stations. (From http://igscb.jpl.nasa.gov/.)

IGS GPS data and products are currently available, at no cost, to users worldwide through the Internet (the IGS URL address is given in Appendix B). The data is available in the standard RINEX format. It should be pointed out that both the IGS precise ephemeris and the tracking station coordinates are referred to the ITRF reference system. That is, if a user employs the IGS precise ephemeris, his or her solution coordinates will be referred to the ITRF reference system. In North America, users could also access GPS data through the CORS and CACS networks, which are operated by the U.S. NGS and Geomatics Canada, respectively [3, 4]. One advantage of using CORS and CACS data is that the reference stations are relatively closer to each other. Additionally, the precise ephemeris and the coordinates of the reference stations may be obtained in either of the ITRF or the improved NAD 83 systems.

7.2 DGPS radio beacon systems

Marine radio beacons are electronic aids to navigation that operate in the low-to-medium frequency band of 283.5–325 kHz [5]. They are installed at lighthouses and other coastal locations. To enhance maritime safety, a number of marine radio beacons throughout the world have been modified to broadcast real-time DGPS corrections in the RTCM format (see Figure 7.2). This service is available at no charge in most cases.

In the DGPS beacon system, a reference station (RS) creates the real-time DGPS corrections in the RTCM format as discussed in Chapter 5. These corrections are digitally modulated using a special form of frequency modulation known as minimum shift keying (MSK). The modulated correction data is then transmitted from the radio beacon at rates between 25 and 200 bps [5]. Typical rates, however, are 100 and 200 bps. In most of the cases, an integrity monitoring (IM) unit is colocated with the RS to monitor its performance (see Figure 7.2).

A user equipped with an MSK beacon receiver can receive the transmitted DGPS corrections as long as he or she is within the coverage area of a particular beacon station. The coverage depends, among other factors, on the transmitter power output, the atmospheric noise, and the receiver sensitivity. The coverage also depends on the characteristics of the propagation path or conductivity; it is greater over water than inland. Beacon locations are usually selected to provide overlapping coverage to increase

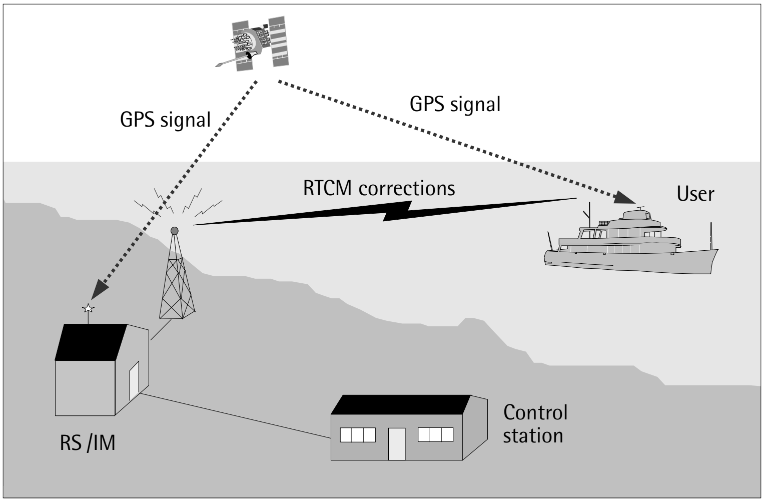

Figure 7.2 DGPS beacon service.

the overall signal availability. The service provider (e.g., the coast guard of a particular country) usually publishes the expected coverage area of a DGPS beacon system. Some manufacturers of the radio beacon receivers publish detailed information about the availability of the beacon service worldwide.

It should be pointed out that MSK beacon receivers are available as single- or dual-channel receivers. The latter are more expensive, but more reliable. The second channel is useful when searching for an adjacent beacon station, if available, with a better signal quality. The receiver will automatically switch to the adjacent beacon station once it is found. Evidently, if an area is known to be covered by one beacon station only, a single-channel MSK receiver would be enough and more cost-effective. To be useful, the MSK receiver should be interfaced to a differential-ready GPS receiver, which would then output the corrected station coordinates.

7.3 Wide-area DGPS systems

Real-time DGPS with a single reference station has the disadvantage that the positioning accuracy tends to deteriorate as the user moves away from the reference station. That is, the highest positioning accuracy is limited to the relatively small area surrounding the reference station. To overcome this problem, a system based on a number of widely separated reference stations, known as WADGPS, has been developed [9]. With this system, the reference stations collect and preprocess the GPS data to compute the DGPS corrections, which are then forwarded to a master station via terrestrial links such as fiber optic cables. The master station analyzes and combines the received data to determine a number of correction parameters for each GPS satellite, which would be valid within the system coverage area (see Figure 7.3). These parameters are packed and uploaded into a geostationary satellite, which rebroadcasts them back to the Earth to ensure a wide coverage. A user within the system coverage area will receive only one set of DGPS corrections, which is valid for his or her location.

A number of commercial WADGPS systems are currently available, including OMNISTAR and RACAL LandStar [7, 8]. Both of these systems use satellite data link and cover various regions of the world (see Figure 7.3). The OMNISTAR service operates in the C-band of the frequency spectrum, while the LandStar service operates in the L-band. To access either service, a subscriber needs the system data receiver to receive and decode the DGPS corrections. The data receiver must be interfaced to a differential-ready GPS receiver to obtain the corrected position. Accuracy of the order of a submeter to a few meters can be obtained, depending mainly on the GPS receiver type.

To reduce the cost of the WADGPS service, some service providers have developed an alternative way of broadcasting the real-time DGPS corrections of their WADGPS system [6]. The system is based on using the already existing FM radio broadcasts to deliver the DGPS corrections to local users (Figure 7.4). That is, the correction parameters transmitted by the geostationary satellite will be received by a number of FM broadcast stations, which will then compute and broadcast the DGPS corrections to the local users. A technology known as radio data system (RDS) allows the FM radio broadcasts to carry the digital DGPS corrections information. To access a service of this type, a subscriber within the coverage area of the FM radio broadcast needs an RDS FM receiver to construct the DGPS corrections. This pager-sized FM receiver must be interfaced to a differential-ready GPS receiver to obtain the corrected position. Accuracy of the order of 1m to 10m can be obtained, depending mainly on the subscription option and the GPS receiver type. This system has the advantage that the

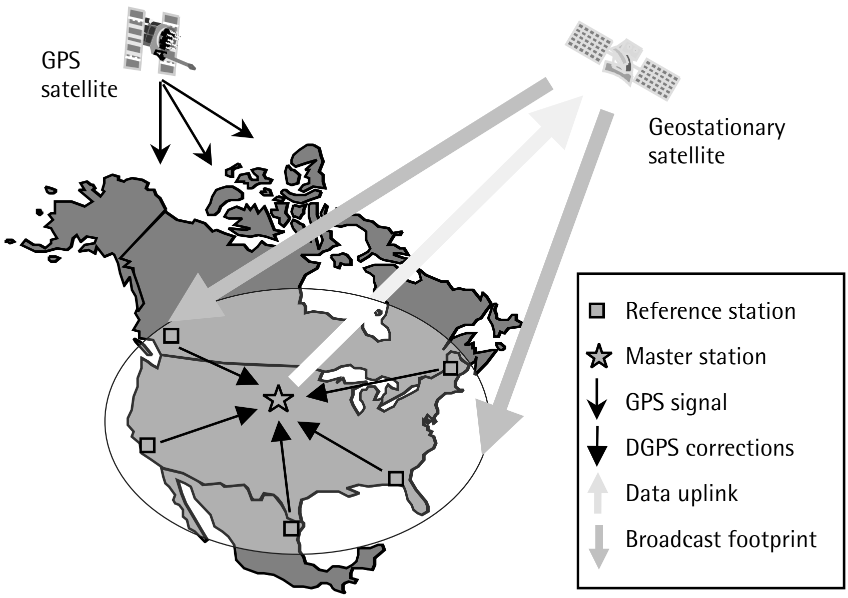

Figure 7.3 Principle of WADGPS system.

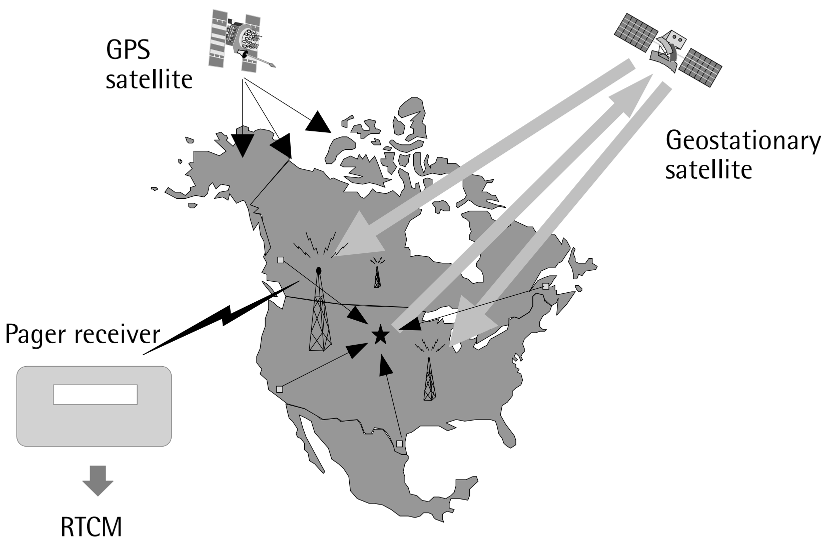

Figure 7.4 Distribution of WADGPS corrections through FM broadcast stations.

subscription fee as well as the FM receiver cost are low compared with the satellite-based broadcast service. It suffers, however, from the limited service coverage of the FM broadcast stations. Reported North American coverage is about 90% of the populated areas.

7.4 Multisite RTK system

As mentioned in Chapter 5, RTK positioning with a single reference station is limited to a distance of about 10 to 15 km. Beyond this distance limit, the errors at the reference and the rover receivers become less correlated (i.e., dissimilar) and would not cancel out sufficiently through the double differencing [10]. This leads to unsuccessful fixing for the ambiguity parameters, which in turn deteriorates the positioning accuracy. To overcome this limitation, research groups have developed multisite real-time, carrier-phase–based RTK positioning [10].

The idea behind multisite RTK positioning is based on using a network of reference stations to create raw GPS measurements for a virtual reference station, which is located very close to the mobile, or the rover, receiver. Once created, the virtual reference station measurements are transmitted to the mobile receiver, where the normal single reference station RTK positioning can be performed. The way the virtual reference station measurements are created can be summarized as follows. First, the differential errors between the reference stations within the network are determined, based on their known precise coordinates. The differential errors at any point within the network (e.g., a mobile receiver's location) can then be determined by interpolation. Once the mobile user provides his or her approximate position to the control station, the differential errors at that location are determined. The raw measurements are then created based on the differential errors and the approximate position of the mobile user [10].

Other forms of multisite RTK positioning have also been developed. The principle, however, remains similar to the virtual reference station technique. It should be pointed out that multisite RTK positioning is growing fast. The recent developments in Internet and wireless communication technologies make this service very promising [11].

References

[1] Beutler, G., et al., "International GPS Service (IGS): An Interdisciplinary Service in Support of Earth Sciences," *32nd COSPAR Scientific Assembly*, Nagoya, Japan, July 12–19, 1998.

[2] Cheves, M., "No-Cost GPS Observations Available Nationally," *Professional Surveyor*, Vol. 16, No. 2, 1996, pp. 6, 8, 10.

[3] Cheves, M., "CORS: A Case Study of a One-Man GPS Crew," *Professional Surveyor*, Vol. 16, No. 2, 1996, pp. 10, 12, 13.

[4] Duval, R., P. Héroux, and N. Beck, "Canadian Active Control System Delivering the Canadian Spatial Reference System," *GIS '96*, Vancouver, Canada, March 1996.

[5] Radio Technical Commission for Maritime Services, "RTCM Recommended Standards for Differential GNSS Service," Version 2.2, Alexandria, VA, January 15, 1998.

[6] Differential Correction Inc., http://www.dgps.com/.

[7] OMNISTAR, http://www.omnistar.com/.

[8] RACAL LandStar, http://www.racal-landstar.com/.

[9] Mueller, T., "Wide Area Differential GPS," *GPS World*, Vol. 5, No. 6, June 1994, pp. 36–44.

[10] Raquet, J., and G. Lachapelle, "Efficient Precision Positioning: RTK Positioning with Multiple Reference Stations," *GPS World*, Vol. 12, No. 4, April 2001, pp. 48–53.

[11] Hada, H., et al., "The Internet, Cars, and DGPS: Bringing Mobile Sensors and Global Correction Services On Line," *GPS World*, Vol. 11, No. 5, May 2000, pp. 38–43.

8

GPS Standard Formats

Since individual GPS manufacturers have their own proprietary formats for storing GPS measurements, it can be difficult to combine data from different receivers. A similar problem is encountered when interfacing various devices, including the GPS system. To overcome these limitations, a number of research groups have developed standard formats for various user needs. This chapter discusses the most widely used standard formats, namely, RINEX, NGS-SP3, RTCM SC-104, and NMEA 0183.

8.1 RINEX format

To save storage space, proprietary formats developed by GPS receiver manufacturers are mostly binary, which means that they are not directly readable when displayed [1]. This creates a problem when combining data (in the postprocessing mode) from different GPS receivers. To overcome this problem, a group of researchers have developed an internationally accepted data exchange format [1]. This format, known as the RINEX format, is in the standard ASCII format (i.e., readable text). Although a file in the ASCII format is known to take more storage space than a file in the binary format, it provides more distribution flexibility.

A RINEX file is a translation of the receiver's own compressed binary files. A draft version of the RINEX format was introduced in 1989 followed by a number of updates to accommodate more data types (e.g., GLONASS data) and other purposes [1]. The current RINEX version 2.10 defines six different RINEX files; each contains a header and data sections: (1) observation data file, (2) navigation message file, (3) meteorological file, (4) GLONASS navigation message file, (5) geostationary satellites (GPS signal payloads) data file, and (6) satellite and receiver clock data file. A new version 2.20 is currently proposed to accommodate data from low Earth orbit (LEO) satellites equipped with GPS or GPS/GLONASS receivers [2]. For the majority of GPS users, the first three files are the most important, and therefore will be the only ones discussed here. The record, or line, length of all RINEX files is restricted to a maximum of 80 characters.

The recommended naming convention for RINEX files is "ssssdddf.yyt." The first four characters, "ssss," represent the station name; the following three characters, "ddd," represent the day of the year of first record; the eighth character, "f," represents the file sequence number within the day. The file extension characters "yy" and "t" represent the last two digits of the current year and the file type, respectively. The file type takes the following symbols: "O" for observation file, "N" for navigation file, "M" for meteorological data file, "G" for GLONASS navigation file, and "H" for geostationary GPS payload navigation message file. For example, a file with the name "abcd032.01o" is an observation file for a station "abcd," which was observed on February 1, 2001.

The observation file contains in its header information that describes the file's contents such as the station name, antenna information, the approximate station coordinates, number and types of observation, observation interval in seconds, time of first observation record, and other information. The observation types are defined as L1 and L2, and represent the phase measurements on L1 and L2 (cycles); C1 represents the pseudorange using C/A-code on L1 (meters); P1 and P2 represent the pseudorange using P-code on L1 and L2 (meters); D1 and D2 represent the Doppler frequency on L1 and L2 (Hertz). The GPS time frame is used for the GPS files, while the UTC time frame is used for GLONASS files. The header section may contain some optional records such as the leap seconds. The last 20 characters of each record (i.e., columns 61 to 80) contain textual descriptions of that record. The last record in the header section must be "END OF HEADER." Figure 8.1 shows an example of a RINEX observation file for single-frequency data, which was created using the Ashtech Locus processor software.

The data section is divided into epochs; each contains the time tag of the observation (the received-signal receiver time, in the GPS time frame for GPS files), the number and list of satellites, the various types of measurements in the same sequence as given in the header, and the signal strength. Other information, such as the loss of lock indicator, is also included in the data section. The data section may optionally contain the receiver clock offset in seconds (see Figure 8.1).

The navigation message file contains the satellite information as described in Chapter 2. In its header, the navigation message contains information such as the date of file creation, the agency name, and other relevant information. Similar to the observation file, the last record in the

```
     2              OBSERVATION DATA    G (GPS)             RINEX VERSION / TYPE
ASHTORIN                                09 - APR - 01 17:27 PGM / RUN BY / DATE
                                                            COMMENT
TEST                                                        MARKER NAME
                                                            MARKER NUMBER
                                                            OBSERVER / AGENCY
                    LOCUS               L_42      UNKNOWN   REC # / TYPE / VERS
                                                            ANT # / TYPE
 -2687840.8300 -4301491.3200  3853858.0200                  APPROX POSITION XYZ
        0.0000        0.0000        0.0000                  ANTENNA: DELTA H/E/N
     1     0                                                WAVELENGTH FACT L1/2
     3    L1     C1     D1                                  # / TYPES OF OBSERV
     10.0000                                                 INTERVAL
                                                            LEAP SECONDS
  1998     9     23     18     27   10.000000      GPS      TIME OF FIRST OBS
  1998     9     23     19      1   59.997000      GPS      TIME OF LAST OBS
                                                            END OF HEADER
 98  9 23 18 27 10.0000000  0  5G03G31G01G23G08                      0.000060824
   7877626.975 6  21949801.811          -48.022
   7858214.382 6  22175367.525         1996.393
   7842888.958 6  20376440.935         2817.693
   7874476.800 6  22485604.397          233.618
   7843609.590 6  22959447.916         3287.071
 98  9 23 18 27 20.0000000  0  6G03G31G01G23G08G09                   0.000047432
   7878091.833 6  21949887.017          -45.258
   7838246.804 6  22171573.369         1997.588
   7814702.570 6  20371080.421         2819.669
   7872108.827 6  22485156.722          239.992
   7810730.579 6  22953202.951         3289.061
     -1195.47216  24085463.326          937.326
                                        ,
                                        ,
                                        ,
```

Figure 8.1 Example of a RINEX observation file for single-frequency data.

header section of the navigation file must be "END OF HEADER." Optionally, the header section may contain additional information such as the parameters of the ionospheric model for single-frequency users (Chapter 3). As well, almanac parameters relating GPS time and UTC and the leap seconds may optionally be included in the header section of the navigation message. The first record in the data section contains the satellite PRN number, the time tag, and the satellite clock parameters (bias, drift, and drift rate). The subsequent records contain information about the broadcast orbit of the satellite, the satellite health, the GPS week, and other relevant information (see Figure 8.2).

The meteorological file contains time-tagged information such as the temperature (in degrees Celsius), the barometric pressure (in millibars), and the humidity (in percent) at the observation site. The meteorological file starts with a header section containing the observation types (e.g., pressure), the sensors-related information, the approximate position of the meteorological sensor, and other related information. As with the other files, the last record in the header section must be "END OF HEADER." The data section contains the time tags (in GPS time) followed by the

```
     2.10           N: GPS NAV DATA                         RINEX VERSION / TYPE
XXRINEXN V2.10      AIUB                3-SEP-99 15:22      PGM / RUN BY / DATE
EXAMPLE OF VERSION 2.10 FORMAT                              COMMENT
     .1676D-07   .2235D-07  -.1192D-06  -.1192D-06          ION ALPHA
     .1208D+06   .1310D+06  -.1310D+06  -.1966D+06          ION BETA
     .133179128170D-06  .107469588780D-12   552960     1025 DELTA-UTC: A0,A1,T,W
    13                                                      LEAP SECONDS
                                                            END OF HEADER
 6 99  9  2 17 51 44.0 -.839701388031D-03 -.165982783074D-10  .000000000000D+00
     .910000000000D+02  .934062500000D+02  .116040547840D-08  .162092304801D+00
     .484101474285D-05  .626740418375D-02  .652112066746D-05  .515365489006D+04
     .409904000000D+06 -.242143869400D-07  .329237003460D+00 -.596046447754D-07
     .111541663136D+01  .326593750000D+03  .206958726335D+01 -.638312302555D-08
     .307155651409D-09  .000000000000D+00  .102500000000D+04  .000000000000D+00
     .000000000000D+00  .000000000000D+00  .000000000000D+00  .910000000000D+02
     .406800000000D+06  .000000000000D+00
13 99  9  2 19  0  0.0  .490025617182D-03  .204636307899D-11  .000000000000D+00
     .133000000000D+03 -.963125000000D+02  .146970407622D-08  .292961152146D+01
    -.498816370964D-05  .200239347760D-02  .928156077862D-05  .515328476143D+04
     .414000000000D+06 -.279396772385D-07  .243031939942D+01 -.558793544769D-07
     .110192796930D+01  .271187500000D+03 -.232757915425D+01 -.619632953057D-08
    -.785747015231D-11  .000000000000D+00  .102500000000D+04  .000000000000D+00
     .000000000000D+00  .000000000000D+00  .000000000000D+00  .389000000000D+03
     .410400000000D+06  .000000000000D+00
                                          ,
                                          ,
                                          ,
```

* obtained from: ftp://ftp.unibe.ch/aiub/rinex/rinex210.txt

Figure 8.2 Example of a RINEX navigation file.

meteorological data arranged in the same sequence as specified in the header (see Figure 8.3).

Most GPS receiver manufacturers have developed postprocessing software packages that accept GPS data in the RINEX format. Most of these ackages are also capable of translating the GPS data in the manufacturer's proprietary format to the RINEX format. The users should, however, be aware that some software packages change the original raw observations in the translation process (e.g., smoothing the raw pseudorange measurements).

8.2 NGS-SP3 format

As discussed in Chapter 3, several institutions are producing precise orbital (ephemeris) data to support applications requiring high-accuracy positioning. To facilitate exchanging such precise orbital data, the U.S. NGS developed the SP3 format, which later became the international standard [3]. The SP3 is an acronym for Standard Product #3, which was originally introduced as SP1 in 1985. The SP3 file is an ASCII file that contains information about the precise orbital data (in the ITRF reference frame) and the associated satellite clock corrections. The line length of the SP3 files is restricted to 60 columns (characters). All times are referred to the GPS time system in the SP3 data standards.

```
     2.10           METEOROLOGICAL DATA                        RINEX VERSION / TYPE
XXRINEXM V9.9       AIUB                   3-APR-96 00:10      PGM / RUN BY / DATE
EXAMPLE OF A MET DATA FILE                                     COMMENT
A 9080                                                         MARKER NAME
     3    PR    TD    HR                                       # / TYPES OF OBSERV
PAROSCIENTIFIC      740-16B                          0.2    PR SENSOR MOD/TYPE/ACC
HAENNI                                               0.1    TD SENSOR MOD/TYPE/ACC
ROTRONIC            I-240W                           5.0    HR SENSOR MOD/TYPE/ACC
        0.0           0.0           0.0        1234.5678 PR SENSOR POS XYZ/H
```

Figure 8.3 Example of a RINEX meteorological file.

A precise ephemeris file in the SP3 format consists of two sections: a header and data. The header section is a 22-line section (see Figure 8.4). The first line starts with the version symbols (#a) and contains information such as the Gregorian date and time of day of the first epoch of the orbit, and the number of epochs in the ephemeris file. Line 2 starts with the symbols (##) and shows the GPS week number, the seconds of the week, the epoch interval, and the modified Julian day. Lines 3–7 start with the symbol (+) and show the total number of satellites (on line 3) as well as list the satellites by their respective identifiers (PRN number). Lines 8–12 start with the symbols (++) and show the accuracy exponents for the satellites shown on lines 3–7. The meaning of the accuracy exponent (ae) is explained as follows: the standard deviation of the orbital error for a particular satellite = 2^{ae} mm. For example, as shown in Figure 8.4, satellite PRN 1 has an accuracy exponent of 6, which means that the standard deviation of its orbital error is 2^6 = 64 mm or 6.4 cm. Lines 13–19 of the SP3 header are reserved for future modification, while lines 19–22 are used freely for comments.

```
#aP2001  3 30  0  0  0.00000000     192 ORBIT IGS97 HLM  IGS
## 1107 432000.00000000   900.00000000 51998 0.0000000000000
+   26      1  2  3  4  5  6  7  8  9 10 11 13 14 18 19 20 22
+          23 24 25 26 27 28 29 30 31  0  0  0  0  0  0  0  0
+           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
+           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
+           0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
++          6  6  7  7  5  7  6  6  7  6  9  6  7  7  8  7  8
++          8  6  7  7  6  7  7  7  6  0  0  0  0  0  0  0  0
++          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
++          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
++          0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
%c cc cc ccc ccc cccc cccc cccc cccc ccccc ccccc ccccc ccccc
%c cc cc ccc ccc cccc cccc cccc cccc ccccc ccccc ccccc ccccc
%f  0.0000000  0.000000000  0.00000000000  0.000000000000000
%f  0.0000000  0.000000000  0.00000000000  0.000000000000000
%i    0    0    0    0      0      0      0      0         0
%i    0    0    0    0      0      0      0      0         0
/* ULTRA ORBIT COMBINATION FROM WEIGHTED AVERAGE OF:
/* cou emu esu gfu jpu siu usu
/* REFERENCED TO cou CLOCK AND TO WEIGHTED MEAN POLE:
/* CLK ANT Z-OFFSET (M): II/IIA 1.023; IIR 0.000
```

Figure 8.4 Example of header section of an SP3 file.

The data section of the precise ephemeris in the SP3 format starts at line 23, which contains the data and time of the first record (epoch). In fact, this is the same time shown in the first line of the header section. Subsequent lines contain the satellite coordinates and the satellite clock data for the current epoch. Each line is assigned for a particular satellite and starts with the character "P," which means a position line. The character "P" is followed by the satellite PRN number, the x, y, and z coordinates of the satellite in kilometers, and the satellite clock correction in microseconds (see Figure 8.5). In some cases, satellite velocity values and the rate of clock corrections are mixed with this information. To handle this, the position and clock correction record will be on one line, followed by a line containing the velocity and the rate of clock correction record for the same satellite. The line containing the velocity record starts with the letter "V." Subsequent epochs will have the same structure, and the file ends with the symbol "EOF."

```
/* CLK ANT Z-OFFSET (M): II/IIA 1.023; IIR 0.000
*  2001  3 30  0  0  0.00000000
P  1   -116.031103  26515.622573   1331.872298    170.652861
P  2  24757.390995   9275.128350  -3848.577237   -359.708080
P  3 -13117.929564  13968.983112  18315.041573     15.998805
P  4  23740.479526  -3537.874866 -11560.053546    700.699352
P  5  -3512.827227 -17951.461871 -19334.408201    292.906571
P  6  -5935.494799 -24254.527474   8889.371588     -0.341952
P  7  14798.294349   7536.247891 -20440.059001    583.158450
P  8  18610.888633   4767.865045  18173.364660     54.770770
P  9   9426.770116 -18913.806117 -16067.347963    -37.796993
P 10  13891.509528  -8251.910439  21127.566769      1.704797
P 11  -8941.716559  19453.856287 -15733.870061      1.569531
P 13   7038.374572  23279.495964  10806.821255     -0.822725
P 14 -14521.250452  -7158.053525 -20986.923406    -97.891864
P 18 -19581.538963 -17313.825718   4877.563537    -46.041799
P 19    429.793263  17637.998255  19905.627091    480.354763
P 20   5114.035928  18254.558669 -18635.373625    -62.353966
P 22 -20658.494478   2973.530545  16434.436461    570.014553
P 23 -17496.621848 -18488.261324   8392.593309     10.342370
P 24  23277.350666 -12714.473270   1361.486226     39.344463
P 25 -23661.057165   6947.104246  -9357.073325     12.450119
P 26   8280.485341 -22212.244294  11256.070469    405.224785
P 27  11045.683417  11584.034891  21496.856169     15.920673
P 28  -9386.791992  12141.516807  21783.012543     13.119881
P 29 -14938.572048  -3401.344352 -21449.873237    495.675550
P 30 -13949.689805 -18738.175431 -12872.779392    -13.581379
P 31  -6009.989874  24108.310257   8665.943843     37.012988
*  2001  3 30  0 15  0.00000000
P  1   -366.735215  26484.551355  -1531.200191    170.653668
                               ,
                               ,
                               ,
```

Figure 8.5 Example of data section of an SP3 file.

8.3 RTCM SC-104 standards for DGPS services

Real-time DGPS operations require the estimation of the pseudorange corrections at the reference receiver, which is then transmitted to the rover receiver through a communication link. To ensure efficiency of operations, the pseudorange corrections are sent in an industry standard format known as the RTCM SC-104 [4]. This format was proposed by the Radio Technical Commission for Maritime Services (RTCM), an advisory organization established in 1947 to investigate issues related to maritime telecommunications. Special Committee No. 104 (SC-104) was established in 1983 to develop recommendations for transmitting differential corrections to GPS users. A draft version of the recommendations was published in November 1985, followed by other updated versions. The most recent version as of this writing, Version 2.2, was published in January 1998 [4]. Originally, the RTCM SC-104 format was designed to support the public marine radio beacon broadcasts of DGPS corrections. However, it has become the industry standard format for transmitting real-time DGPS corrections.

The RTCM SC-104 standards consist of 64 message types [4]. These messages contain information such as the pseudorange correction (PRC) for each satellite in view of the reference receiver, the rate of change of the pseudorange corrections (RRC), and the reference station coordinates. Of interest to the majority of real-time DGPS users are message types 1 and 9. Both contain the PRC and the RRC corrections. However, message type 1 contains the corrections for all the satellites in view of the reference station, while in message type 9 the corrections are packed in groups of three. This leads to a lower latency for message type 9 compared with message type 1, which is useful in the presence of selective availability. The disadvantage of using message type 9, however, is that the reference station requires a more stable clock. Some tentative messages were added in Version 2.2 to support the RTK and differential GLONASS operations. Table 8.1 shows a list of the current message types.

The RTCM SC-104 messages are not directly readable; they are streams of binary digits, zeros and ones. Each RTCM SC-104 message or frame consists of an "N + 2" 30-bit words long; where N represents the number of words containing the actual data within the message and the remaining two words represent a two-word header at the beginning of each message. The size of N varies, depending on the message type and the message content (e.g., the varying number of satellites in view of the reference station).

Table 8.1 Current RTCM Message Types

Message Type Number	Current Status	Title	Message Type Number	Current Status	Title
1	Fixed	DGPS corrections	18	Fixed	RTK uncorrected carrier phases
2	Fixed	Delta DGPS corrections	19	Fixed	RTK uncorrected pseudoranges
3	Fixed	GPS reference station parameters	20	Tentative	RTK carrier-phase corrections
4	Tentative	Reference station datum	21	Tentative	RTK/high PRC account
5	Fixed	GPS constellation health	22	Tentative	Extended reference station parameters
6	Fixed	GPS null frame	23–30	—	Undefined
7	Fixed	DGPS radio beacon almanac	31	Tentative	Differential GLONASS corrections
8	Tentative	Pseudolite almanac	32	Tentative	Differential GLONASS reference standard parameters
9	Fixed	GPS partial correction set	33	Tentative	GLONASS constellation health
10	Reserved	P-code differential correction	34	Tentative	GLONASS partial differential correction set
11	Reserved	C/A-code L1, L2 delta corrections	35	Tentative	GLONASS beacon almanac

Table 8.1 (continued)

Message Type Number	Current Status	Title	Message Type Number	Current Status	Title
12	Reserved	Pseudolite station parameters	36	Tentative	GLONASS special message
13	Tentative	Ground transmitter parameters	37	Tentative	GNSS system time offset
14	Tentative	GPS time of week	38–58	—	Undefined
15	Tentative	Ionospheric delay message	59	Fixed	Proprietary message
16	Fixed	GPS special message	60–63	Reserved	Multipurpose usage
17	Tentative	GPS ephemeris	64	—	Not reported

The word size and the parity check algorithm are the same as those of the GPS navigation message. The remaining part of this section discusses the structure of message type 1, which is commonly used in real-time DGPS operations.

Figure 8.6 shows the structure of a message type 1, where five satellites were visible at the reference station. The first word of the header starts with an 8-bit preamble, which is a fixed sequence 01100110. Following the preamble are 6-bit message type identifier and a 10-bit reference station ID. The last 6 bits of this word and of all other words are assigned for parity, which checks for any error. The second word starts with a 13-bit modified z-count, a time reference for the transmitted message, followed by a 3-bit sequence number for verifying the frame synchronization. The length of frame is assigned bits 17–21 and is used to identify the start of the next frame. Bits 22–24 define the reference station health status; for example, a code of "111" means that the reference station is not working properly. The actual data set for all the satellites is contained in the remaining words. Each satellite requires a total of 40 bits for the correction, distributed in the following sequence: (1) scale factor, S (1 bit); (2) user differential range error, UDRE (2 bits); (3) satellite ID (5 bits); (4) pseudorange correction,

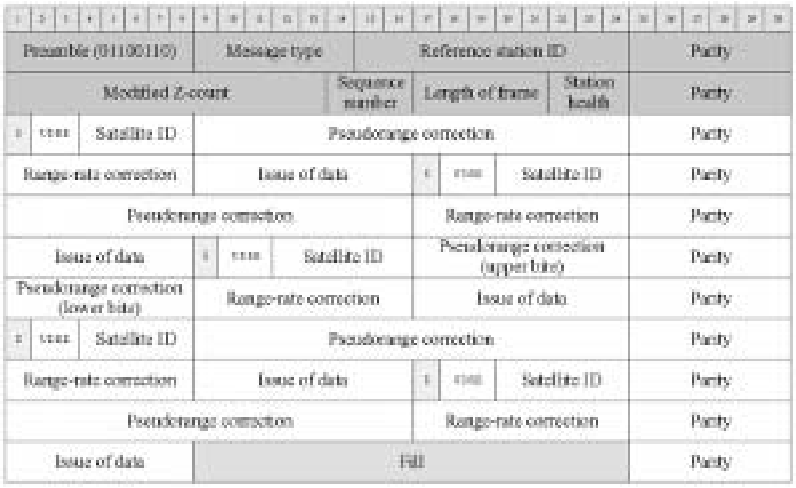

Figure 8.6 Structure of a five-satellite message type 1.

PRC (16 bits); (5) range-rate correction, RRC (8 bits); and (6) issue of data (8 bits). The scale factor is used to scale the PRC/RRC, while the UDRE is a measure of the uncertainty in the PRC. However, the issue of data identifies the GPS navigation message that the reference station used to calculate the satellite position and clock offset. Evidently, there will be cases where the required number of words is not an exact integer number. For example, as shown in Figure 8.6, if the number of satellites in the message is five, 16 bits are required to fill the frame. Similarly, if the number of satellites is four, 8 bits are required to fill the frame. To avoid confusion with the preamble, the fill uses alternating ones and zeros (i.e., 1010101010101010 or 10101010).

To obtain the DGPS corrections, the transmitted messages must be decoded and converted to 30-bit long words (strings of zeros and ones). Once this is done, party checks should be performed and then the DGPS corrections information can be extracted according to Figure 8.6. Figure 8.7 shows a "real" example showing four type 1 messages; each represents the DGPS corrections for all the satellites in view of the reference station at a particular epoch. Figure 8.8 shows how the first word of the header information is decoded. Most GPS receivers support the RTCM SC-104 standards, which allows the use of different receivers in the real-time mode. It should be noted, however, that not all the differential-ready GPS receivers could output the RTCM standards.

```
Y~}o}  _X~Cp|_TSVA@VL_OJ]K|C~_^@gJuDAA@UVwIAjI`BAoOxc~WZSc^jTQB`TAPMRUq/
Y~}Oj  _^~GX}}FXtrOL`@T\Js    A`deZuJo~~  bxHv~Ez`zyEHiFI@GbCuHeW}m]zc^`jtH
fABhH@`oA`n}_k~or}i|`dv\gC|AP_YfDuIp|  tri|  Ef`bwEBiFD@Jt\gD\hBJi_|b_Uoå
Y~}wE@`kA^X}m__tqOLZ  Sjmw  _AhWe`QjF`  GjiVorU_`\JGFxGI@y]~Jwv~~  jiHvnjqĐ
```


Figure 8.7 Example of raw RTCM SC-104 corrections (message type 1).

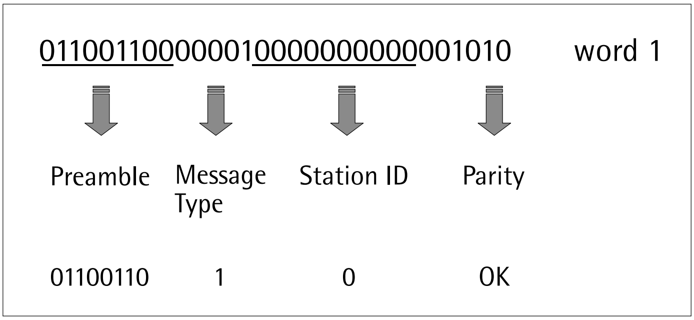

Figure 8.8 First-word decoding example.

8.4 NMEA 0183 format

NMEA is an acronym for the National Marine Electronics Association, which was founded in 1957 by a group of electronics dealers to strengthen their relationships with electronic manufacturers [5]. In 1983, with input from the manufacturers and private and governmental organizations, the association adopted the NMEA 0183 as a format for interfacing marine electronic devices. It has been updated several times; the latest release, as of this writing, Version 3.0, appeared in July 2000. The NMEA 0183 standards are data streams in the ASCII format, transmitted at a rate of 4,800 bps, from a talker to a listener, where a talker is a device that sends data to other devices (e.g., a GPS receiver) and a listener is a device that receives data from another device (e.g., a laptop computer interfaced with the GPS receiver) [5].

The NMEA 0183 data streams may include information on position, datum, water depth, and other variables. The data is sent in the form of sentences; each starts with a dollar sign "$" and terminates with a carriage return-line feed <CR><LF>. The dollar sign "$" is followed by a five-character address field, which identifies the talker (the first two characters), the data type, and the string format of the successive fields (the last three characters). The last field in any sentence is a checksum field, which follows a checksum delimiter character "*". The maximum total number of characters in any sentence is 82; that is, a maximum of 79 characters between the starting delimiter "$" and the terminating <CR><LF>. A number of these sentences are dedicated to GPS and GLONASS systems, while the remaining sentences support other devices such as echo sounders, gyros, and others [5].

Our discussion will be restricted to one sentence only, the GGA: Global Positioning System fix data. This sentence represents the time and position, and solution-related information. Figure 8.9 shows the general structure of the GGA sentence, while Table 8.2 explains the terms of the sentence.

```
$GPGGA,hhmmss.ss,llll.ll,a,yyyyy.yy,a,x,xx,x.x,x.x,M,x.x,M,x.x,xxxx*hh<CR><LF>
$GPGGA,115417.00,4338.123456,N,07938.123456,W,1,10,01.1,095.095,M,,M,999,0000
```

Figure 8.9 General structure of a GGA sentence.

Table 8.2 Explanation of GGA Sentence Terms

$	Start of sentence delimiter
GP	Talker identifier (GPS in this case)
GGA	Data identifier (GPS fix data in this case)
,	Data field delimiter
hhmmss.ss	Time of position in UTC system (hoursminutesseconds.decimal)
llll.ll	Latitude (degreesminutes.decimal)
a	N/S (North or South)

TABLE 8.2 (CONTINUED)

yyyyy.yy	Longitude (degreesminutes.decimal)
a	E/W (East or West)
x	GPS quality indicator (1 = point positioning with C/A-code)
	(2 = DGPS with C/A-code)
	(3 = point positioning with P-code)
	(4 = RTK with ambiguity parameters fixed to integer values)
	(5 = RTK with float ambiguity parameters)
xx	Number of satellites used in producing the solution
x.x	HDOP
x.x	Orthometric Height
M	Meters (units of Orthometric Height)
x.x	Geoidal Height above the WGS 84 ellipsoid
M	Meters (units of Geoidal Height)
x.x	Age of DGPS data in seconds (time since last RTCM message type 1 or 9 was received; null field when DGPS mode is not used)
xxxx	Reference station ID (in case of DGPS; use the range 0000–1,023)
*	Checksum delimiter character
hh	Checksum field (last field in the sentence)
<CR><LF>	Sentence terminator

Most GPS receivers available on the market support the NMEA 0183 standards. However, not all receivers with the NMEA 0183 port output all the GPS-specific messages. In addition, some GPS receiver manufacturers may slightly change the standard format. However, they typically provide software to interpret the data sentence.

References

[1] Gurtner, W., "RINEX: The Receiver-Independent Exchange Format," *GPS World*, Vol. 5, No. 7, July 1994, pp. 49–52.

[2] Gurtner, W., and L. Estey, "RINEX Version 2.20: Modifications to Accommodate Low Earth Orbiter Data," accessed 2001, ftp://ftp.unibe.ch/aiub/rinex/rnx_leo.txt.

[3] Remondi, B., *Extending the National Geodetic Survey Standard Orbit Formats*, NOAA Technical Report NOS 133 NGS 46, 1989.

[4] Radio Technical Commission for Maritime Services, "RTCM Recommended Standards for Differential GNSS Service," Version 2.2, Alexandria, VA, January 15, 1998.

[5] National Marine Electronic Association, "NMEA 0183 Standards for Interfacing Marine Electronics," Version 3.0, New Bern, NC, July 2000.

9
GPS Integration

GPS has found its way into many applications, mainly as a result of its accuracy, global availability, and cost-effectiveness. Unfortunately, however, there exist some situations in which part of the GPS signal may be obstructed to the extent that the GPS receiver may not "see" enough satellites for positioning. Examples of those situations are positioning in urban canyons and deep open-pit mining. This signal-obstruction problem, however, was successfully overcome by integrating GPS with other positioning systems. In fact, reported results showed that the performance of the integrated system is better than either system alone. Augmenting GPS is not limited to sensor integration. As shown below, GPS can be augmented with computer-based tools, such as GIS, for efficient data collection and analysis.

9.1 GPS/GIS integration

A geographic information system (GIS) is a computer-based tool capable of acquiring, storing, manipulating, analyzing, and displaying spatially referenced data [1]. Spatially referenced data is data that is identified according to its geographic location (e.g., features such as streets, light poles, and fire hydrants are linked by geography).

Spatial, or geographic, data can be obtained from a variety of sources such as existing maps, satellite imagery, and GPS. Once the information is collected, a GIS stores it as a collection of layers in the GIS database (see Figure 9.1). The GIS can then be used to analyze the information and decisions can be made efficiently. (For example, the decision to build a new road can be made by studying the effect of one feature, such as traffic volume.)

GPS is used to collect the GIS field data efficiently and accurately [2]. With GPS, the data is collected in a digital format in either real-time or postprocessed mode. A number of GPS/GIS systems that provide centimeter- to meter-level accuracy are now available on the market. Most of these systems allow the user to enter user-defined attributes for each feature. Built-in navigation functions to relocate field assets are also available. Pen computer-based systems are used by some GPS receiver manufacturers to allow the data to be edited and displayed as it is collected [2].

Many industries, including utilities management, forestry, agriculture, public safety, and fleet management, can benefit from integrated GPS/GIS systems.

9.2 GPS/LRF integration

In areas with heavy tree canopy, GPS receivers will normally lose lock to the GPS satellites. In addition, real-time differential GPS corrections may not be received as well. To overcome these problems, integrated GPS/handheld laser units, or laser range finders (LRFs), were developed [3]. The way the integrated system operates is to set up the GPS antenna in a nearby open area, which allows the GPS system to operate normally without losing lock to the GPS satellites. With the help of a digital compass, a reflectorless handheld laser, colocated with the GPS receiver, can be used to determine the distance and azimuth to the inaccessible points (see Figure 9.2). This operation is commonly known as the offset function. Software residing in the handheld computer helps in collecting both the offset data and the GPS data. At a later time, all the available information is processed using PC software to determine the coordinates of the inaccessible points. Collecting and processing the data may also be done in real time, while in the field,

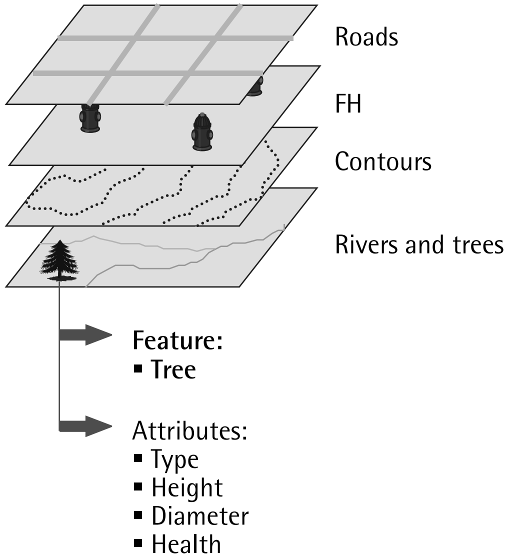

Figure 9.1 GPS/GIS integration.

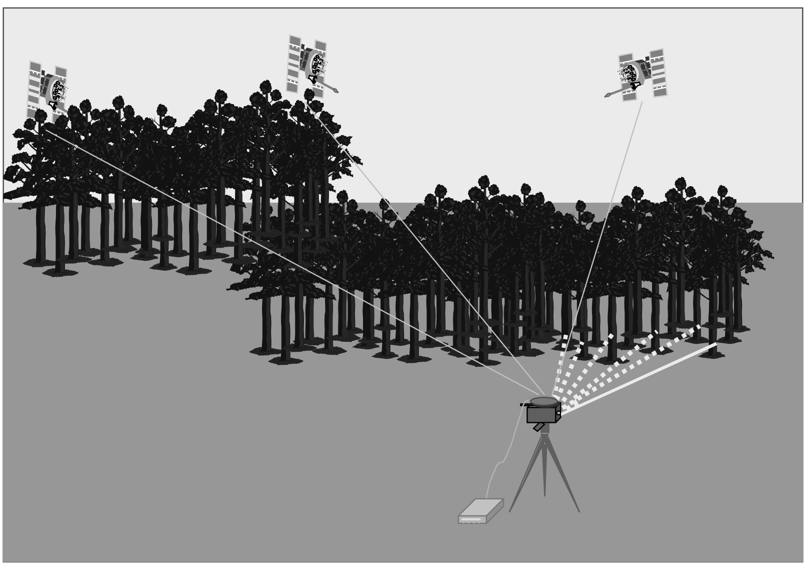

Figure 9.2 GPS/LRF integration.

provided that the real-time DGPS corrections can be received. Once the processing is done, the user can export the output to the required GIS or CAD software. This eliminates the need to place the GPS antenna directly on the features to be mapped [2].

GPS/laser integration is an attractive tool, especially for the forestry industry. Tree offsets, heights, and diameters can be measured easily with the laser unit. From a single location, a stationary user in a relatively open area can offset any number of points or features. In this case, the user location will be determined precisely by averaging all the GPS data collected while taking the offset measurements. Other applications of the GPS/laser integration include mapping points under bridges, mapping points on a busy roadway, mapping highway signs, and mapping shore lines, to name a few. GPS/laser integration can be used to map point features, line features, or area features.

9.3 GPS/dead reckoning integration

Another system that has been used to supplement GPS under poor signal reception is the dead reckoning (DR) system. Dead reckoning is a low-cost system, commonly comprising an odometer sensor and a vibration gyroscope. The integrated GPS/DR system is widely used in automatic vehicle location (AVL) applications [4].

DR navigation requires that the vehicle travel-distance and direction (heading) be available on a continuous basis. The travel-distance information is obtained from the odometer sensor, while the direction information is obtained from the gyroscope. If the vehicle starts the trip from a known location, the distance and direction information can be used to determine the vehicle location at any time. In other words, assuming that the vehicle is traveling in a horizontal plane, the travel and direction information can be integrated over time to compute the vehicle location (position).

Odometer sensors are already installed in all vehicles, mainly to evaluate their age and whether a service is required. An odometer sensor counts the number of revolutions of the vehicle's wheels, which can be converted to a travel distance through an initial calibration. This conversion is known as the odometer scale-factor determination. One way of determining the scale factor is by driving the vehicle over a known distance. Unfortunately, however, the odometer scale factor changes over time due mainly to wheel slipping and skidding, tire pressure variation, tire wear, and vehicle speed. If left uncompensated, the scale-factor error will accumulate rapidly, causing significant positional error [5].

Vibration gyroscopes, however, are low-cost sensors that measure the angular rate (heading rate) based on the so-called Coriolis acceleration. A vibration gyro outputs a voltage that is proportional to the angular velocity of the vehicle. The vehicle's heading rate is obtained by multiplying the output voltage by a scale factor. Similar to the odometer sensors, gyroscopes suffer from error accumulation due to gyro bias and scale-factor instability. A gyro bias is a temperature-sensitive variable error that affects the gyro measurements at all times. As such, a gyro will read a nonzero value even if the angular velocity is zero. It is observable when the vehicle is stationary or when it is moving in a straight line. Gyro scale-factor error, however, affects the gyro measurements only when the vehicle is taking a turn. This error could be greatly reduced by taking equal clockwise and counterclockwise rotations [4].

It can be seen that each of the GPS and DR systems suffers from limitations. While the GPS signal may not be available in obstructed areas, the DR system drifts over time causing large positional error. This suggests that an optimal positioning solution may be developed, based on the two positioning systems. Kalman filtering technique is commonly used for system integration [5]. With the integrated system, GPS helps in controlling the drift of the DR components through frequent calibration, while DR becomes the main positioning system during the GPS outages. As such, the performance of the integrated system will be better than either system alone.

Currently, a promising new inertial navigation technology, microelectro mechanical system (MEMS) technology, is under development. MEMS technology will be used to provide the heading and the traveled distance of the vehicle, replacing the traditional DR system. MEMS-based gyroscopes and accelerometers are expected to overcome the size and the cost of the current technology [6].

9.4 GPS/INS integration

There exist a number of applications that require high-accuracy positioning in obstructed areas and/or under high dynamic conditions. Examples of these applications are deep open-pit mining and airborne mapping (see Chapter 10 for details about these applications). As discussed earlier, a major problem with GPS is its limitation when used in obstructed areas. In addition, a GPS receiver has limited dynamic capabilities. As mentioned in Section 2.7, GPS signal obstruction and high receiver dynamics can cause temporary signal losses, or cycle slips. To overcome these limitations, GPS can be integrated with a relatively environment-independent system, the inertial navigation system (INS).

An INS is a system that, once initialized (by acquiring the initial position, velocity, and orientation information), becomes an autonomous navigation system providing 3-D position, velocity, and attitude information [7]. An inertial sensor, also known as the inertial measurement unit (IMU), is a device consisting of accelerometers, gyroscopes, other electronics components, and a computer. When mounted on a moving object, the accelerometers measure the object's acceleration plus the gravitational force, while the gyroscopes provide information on the orientation of the inertial platform. These sets of information are accumulated by the sensor's computer to produce the velocity and position information. In addition to being a relatively environment-independent system, an inertial system provides accuracy as high as that of GPS for the short period of time following the initialization [7]. Moreover, inertial systems provide very high update rates compared with GPS. A major drawback of the inertial system, however, is that it suffers from drift if left unaided for a long period of time. In particular, the performance of the gyroscopes limits the overall performance of the inertial system.

Integrating GPS and INS overcomes the limitations of both systems [7]. In fact, GPS and INS complement each other. While GPS provides the initialization and the calibration to the inertial system, the latter bridges the GPS gaps when the satellite signal is blocked or temporarily lost. GPS/INS integration is commonly done in either of two modes, namely, loose coupling or tight coupling mechanisms. Loosely coupled integration is carried out in the solution domain, while tightly coupled integration is carried out in the raw measurements domain. In addition, tightly coupled integration requires extensive computations as compared with loosely coupled integration. It results, however, in a nearly optimal integration solution. Similar to the GPS/DR, the Kalman filtering technique is commonly used for GPS/INS integration [5].

9.5 GPS/pseudolite integration

One of the fastest growing applications of GPS is open-pit mining. The use of GPS in open-pit mining can remarkably reduce the cost of various mining operations. The availability of real-time GPS positioning at centimeter-level accuracy has attracted the attention of the mining industry. This is mainly because accurate real-time positioning is a key component that leads to automating the heavy and expensive mining machines. As such, smart mining systems can be developed that not only increase mining safety but also reduce costly labor [8].

Unfortunately, similar to the earlier cases, the satellite signal will be partially blocked as the pit deepens (see Figure 9.3). As such, in deep open-pit mining, GPS alone cannot be used reliably for mining positioning. One promising system that can augment GPS to ensure high-accuracy positioning at all times is the pseudolite (short for pseudosatellite) system. A pseudolite is a ground-based electronic device that transmits a GPS-like signal (code, carrier frequency, and data message), which can be acquired by a GPS receiver. Unlike GPS, which uses atomic clocks onboard the satellites, pseudolites typically use low-cost crystal clocks to generate the signal [9].

The addition of pseudolite signals improves both system availability and geometry. The number and locations of the pseudolites can be

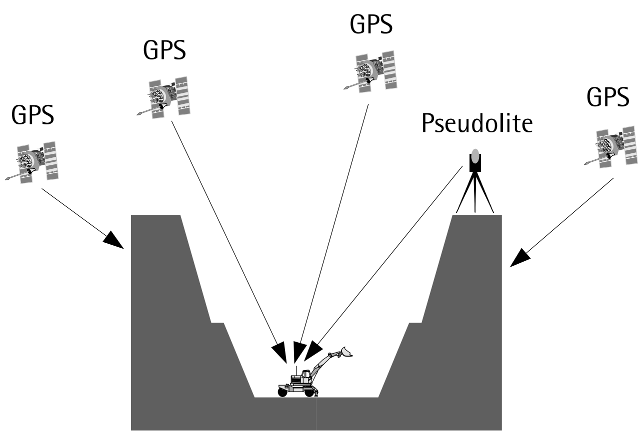

Figure 9.3 GPS/pseudolite integration.

optimized to ensure the best performance of the system. The vertical dilution of precision, in particular, can be improved dramatically, which leads to improved accuracy for the height component. Another advantage of using the pseudolites is that, being ground-based transmitters, their signals are not affected by the ionosphere. Pseudolites, however, suffer from a number of drawbacks that must be overcome to ensure high-accuracy positioning. The first is known as the near-far problem, which results from the variation in the received pseudolite signal power as the receiver-pseudolite distance changes. The closer the receiver to the pseudolite transmitter, the higher the signal power, and vice versa. This problem does not exist with GPS-only positioning, as the received GPS signal power remains almost constant, because the satellite-receiver distance does not change significantly. Consequently, in GPS/pseudolite integration, if the pseudolite signal is much stronger than the other pseudolite and GPS signals, it may overwhelm the other signals and jam the receiver. This is what is known as the near problem. However, if the pseudolite signal is much weaker, the receiver may not be able to track it, which is known as the far problem. Transmitting the pseudolite signal in short pulses with a low duty cycle may, however, minimize the effect of the near-far problem [9].

The use of inaccurate clocks to generate the pseudolite signal causes synchronization error in the sampling time. This error will cause a range error, even if double differences are formed. A possible solution to this problem is through the use of a content-free data message of a master pseudolite. Another problem that requires the pseudolite user's attention is the multipath error. Pseudolite multipath error occurs as a result of reflected signals from objects surrounding the antennas of both the receiver and the transmitter. Some researchers have suggested the use of patterned antennas as a feasible way of reducing the multipath effect. Unlike GPS-only positioning where ephemeris errors do not affect the position solution significantly, errors in the pseudolite coordinates will be propagated into the solution, causing large positioning errors. This is caused by the relatively short receiver-pseudolite separation [10]. Careful calibration of the pseudolite antenna location solves this problem.

It should be pointed out that the application of the integrated GPS/pseudolite system in not limited to deep open-pit mining. Such an integrated system has been successfully used in precise aircraft landing, deformation monitoring, and other applications. Being similar in principle to GPS, pseudolite-only positioning has the potential of being the system of the future for indoor applications, such as underground mining (see Figure 9.3). A challenging problem to overcome, however, is the pseudolite location problem.

9.6 GPS/cellular integration

Cellular communication technology is becoming widely accepted throughout the world. Both the number of subscribers and the cellular coverage areas are increasing continuously. In addition, more advanced digital cellular coverage is on the rise, allowing voice and data to be mixed seamlessly. This makes the cellular system very attractive to a number of markets, including emergency 911, AVL, and RTK GPS.

A major limitation with the current cellular system, however, is its ability to precisely determine where a call was originated [11]. Although this limitation is not critical for applications like RTK GPS, it is of utmost importance for other applications such as emergency 911 and AVL. In the United States, for example, about one-third of all emergency 911 calls come from cellular phones. Of these, nearly one-fourth cannot describe their location precisely, which makes it very difficult for an operator to effectively send out assistance. As such, the U.S. Federal Communications Commission (FCC) has made it mandatory that, as of October 2001, wireless emergency 911 callers must be located with an accuracy of 125m (67% probability level) or better [11].

To meet the FCC location requirement, wireless network operators can either use the network-based location or the handset-based location. Most network-based caller location systems employ either the time-difference of arrival (TDOA) approach or the angle of arrival (AOA) approach to determine the caller's location. The former measures the differences in the arrival times of an emergency 911 signal at the cell sites or base stations. The caller's location can be determined if the signal is received at a minimum of three base stations. Obviously, time synchronization is essential with this technique, which can be ensured by equipping each cell site with a GPS timing receiver. The second technique, the AOA, uses phased-array antennas to compute the angles at which the signal arrives at the base stations. A minimum of two sites is required to compute the caller's location with this method. As both the TDOA and AOA methods have advantages and drawbacks, some network operators combine the two methods [11].

Handset-based location technology integrates GPS with cellular communication through the installation of a GPS chipset in the handset of the wireless phone. With selective availability being turned off permanently, this technology would locate the wireless emergency 911 callers with an accuracy that exceeds the FCC requirement by a factor of ten. Unlike network-based technology, handset-based location technology is very simple to implement and does not require the installation of additional equipment at the base stations (e.g., GPS timing receivers). One of the drawbacks of the handset-based location technology, however, is that only new cellular phones can be equipped with GPS. In addition, the GPS signal is very weak to be received inside buildings. This limitation, however, could be efficiently overcome in the near future with the development of integrated GPS/MEMS technology, described in Section 9.3.

In the near future, the development of a new generation of cellular technology, the 3G wideband digital networks, will be completed. The 3G cellular technology supports voice, high-speed data, and multimedia applications. In addition, this technology uses common global standards, which not only reduces the operational cost but also makes the system useable worldwide. Moreover, with this new technology, devices can be turned on all the time for data transmission, as subscribers pay for the packets of data they receive/transmit.

The advances in the wireless communication and caller's location technologies discussed earlier will greatly impact a number of industries. The vehicle navigation market, for example, is expected to greatly benefit from the advances in wireless communication, location, and Internet technologies (see Section 10.11 for details about vehicle navigation). Currently, vehicles use complex systems that integrate location technology with in-car computer navigation systems containing electronic digital road maps and other related information. Clearly, the in-car system will not be aware of any real-world changes in the navigation system's database (e.g., a change in the traffic direction). With the availability of wireless Internet service, however, an up-to-date database residing at a central location could be accessed by drivers, eliminating the need for a complex in-car computer navigation system. Furthermore, with the availability of a precise location system, drivers could customize the information they need according to their locations, such as turn-by-turn navigation, traffic information, and local weather conditions. This method is simple, cost-effective, and flexible, and has the potential of being the way of the future.

References

[1] Elfick, M., et al., *Elementary Surveying*, 8th ed., New York: HarperCollins, 1994.

[2] Ashtech Inc., "Reliance Field Asset Management Tools," Magellan Corporation, Santa Clara, CA, 2001.

[3] Laser Technology Inc., "Survey Laser For Forestry." PowerPoint Presentation, accessed July 18, 2001, http://www.losertech.com/downloads.html.

[4] Madhukar, B. R., et al., "GPS-DR Integration Using Low Cost Sensors," *Proc. ION GPS-99*, *12th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Nashville, TN, September 14–17, 1999, pp. 537–544.

[5] Kaplan, E., *Understanding GPS: Principles and Applications,* Norwood, MA: Artech House, 1996.

[6] Schwarz, K. P., and N. El-Sheimy, "Future Positioning and Navigation (POS/NAV) Technologies Technical Study," Study performed under Scientific Services Agreement with U.S. Topographic Engineering Center, Fort Belvoir, VA, March 1999.

[7] May, M. B., "Inertial Navigation and GPS," *GPS World*, Vol. 4, No. 9, September 1993, pp. 56–66.

[8] El-Rabbany, A., "Mining Positioning," *Proc. Smart Systems for Mineral Resources Workshop,* Toronto, Ontario, February 14, 2001.

[9] Cobb, S., and M. O'Connor, "Pseudolites: Enhancing GPS with Ground-Based Transmitters," *GPS World*, Vol. 9, No. 3, March 1998, pp. 55–60.

[10] Wang, J., et al., "Integrating GPS and Pseudolite Signals for Position and Attitude Determination: Theoretical Analysis and Experiment Results," *Proc. ION GPS 2000*, *13th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Salt Lake City, UT, September 19–22, 2000, pp. 2252–2262.

[11] Driscoll, C., "Wireless Caller Location Systems," *GPS World*, Vol. 9, No. 10, October 1998, pp. 44–49.

10

GPS Applications

GPS has been available for civil and military use for more than two decades. That period of time has witnessed the creation of numerous new GPS applications. Because it provides high-accuracy positioning in a cost-effective manner, GPS has found its way into many industrial applications, replacing conventional methods in most cases. For example, with GPS, machineries can be automatically guided and controlled. This is especially useful in hazardous areas, where human lives are endangered. Even some species of birds are benefiting from GPS technology, as they are being monitored with GPS during their immigration season. This way, help can be presented as needed. This chapter describes how GPS is being used in land, marine, and airborne applications.

10.1 GPS for the utilities industry

Accurate and up-to-date maps of utilities are essential for utility companies. The availability of such maps helps electric, gas, and water utility companies to plan, build, and maintain their assets.

The GPS/GIS system provides a cost-effective, efficient, and accurate tool for creating utility maps. With the help of GPS, locations of features such as gas lines can be accurately collected, along with their attributes (such as their conditions and whether or not a repair is needed). The collected information can then be used by a GIS system to create updated utility maps.

In situations of poor GPS reception, such as in urban canyons, it might be useful to use integrated GPS and LRF systems [1]. This integrated system is an efficient tool for rapid utility mapping. A GPS receiver remains in the open for the best signal reception, while the LRF measures the offset information (range and azimuth) to the utility assets such as light poles (see Figure 10.1). The processing software should be able to combine both the GPS and the LRF information.

Buried utilities such as electric cables or water pipes can also be mapped efficiently using GPS (Figure 10.1). With the help of a pipe/cable locator attached to the second port of the GPS handheld controller, accurate information on the location and the depth of the buried utility can be collected. This is a very cost-effective and efficient tool, as no ground marking is required.

Figure 10.1 GPS for utility mapping.

10.2 GPS for forestry and natural resources

GPS has been applied successfully in many areas of the forest industry. Typical applications include fire prevention and control, harvesting operations, insect infestation, boundary determination, and aerial spraying [2].

With thousands of fires facing the forest services every year, an efficient resource-management system is essential. GPS is a key technology that enables the system operator to identify and monitor the exact location of the resources (Figure 10.2). With the help of GIS and a good communication system, appropriate decisions can be made.

In the past, aerial photography was the only means of providing information on the shape and location of cut blocks before completing harvesting operations. Such information was often lacking accuracy. With the use of differential GPS, however, this information can be accurately determined in real time.

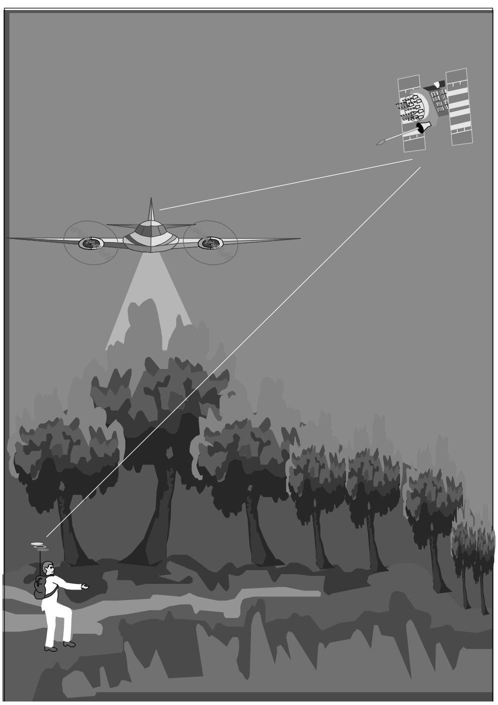

Figure 10.2 GPS applications in forestry.

GPS has also been a very useful tool for wildlife management and insect infestation. Using its precise positioning capability, GPS can determine the locations of activity centers. These locations can be easily accessed using GPS waypoint navigation (see Section 10.15).

GPS surveying is becoming the preferred method for forest boundaries determination. With real-time GPS, up to 75% time and cost reductions can be obtained. As discussed in Chapter 9, in case of poor GPS reception under heavy tree canopy, it might be useful to use integrated GPS and LRF systems. Other integrated systems, including GPS/digital barometers and GPS/laser digital videography, have been applied successfully in the forest industry as well.

10.3 GPS for precision farming

The ability of DGPS to provide real-time submeter- or even decimeter-level accuracy has revolutionized the agricultural industry [3, 4]. GPS applications in precision farming include soil sample collection, chemical applications control, and harvest yield monitors (Figure 10.3).

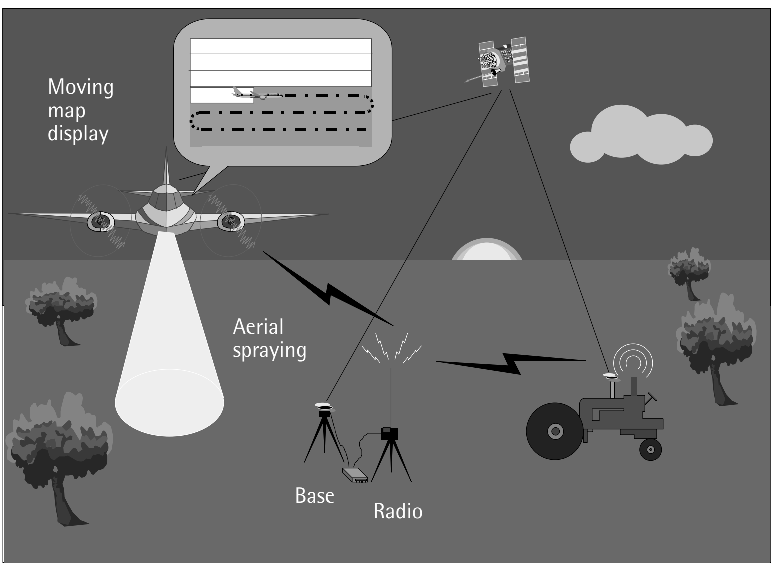

Figure 10.3 GPS for precision farming.

When collecting soil samples, GPS is used to precisely locate the sample points from a predefined grid (Figure 10.3). After testing the soil samples, information such as nitrogen and organic material contents can be obtained. This type of information is mapped and used as a reference to guide farmers in efficiently and economically treating soil problems.

When GPS is integrated with an aerial guidance system, the field sprayer can be guided through a moving map display. Based on the sprayer's location, the system will apply the chemicals at the right spots, with minimal overlap, and automatically adjust their rate. This, in addition to increasing productivity, ensures that chemicals and fuel are used efficiently.

GPS is also used to map crop yields. As the DGPS-equipped harvester moves across the field, yield rates are recorded along with DGPS-derived positions. This information is then mapped to show the yield rate.

Easy-to-use integrated systems with only a few buttons are now available on the market. DGPS corrections are available from the government-operated DGPS/beacon service free of charge, as well as from a number of commercial services. The user's own base station may be built as well (Figure 10.3).

10.4 GPS for civil engineering applications

Civil engineering works are often done in a complex and unfriendly environment, making it difficult for personnel to operate efficiently. The ability of GPS to provide real-time submeter- and centimeter-level accuracy in a cost-effective manner has significantly changed the civil engineering industry. Construction firms are using GPS in many applications such as road construction, Earth moving, and fleet management.

In road construction and Earth moving, GPS, combined with wireless communication and computer systems, is installed onboard the Earth-moving machine [5]. Designed surface information, in a digital format, is uploaded into the system. With the help of the computer display and the real-time GPS position information, the operator can view whether the correct grade has been reached (see Figure 10.4). In situations in which millimeter-level elevation is needed, GPS can be integrated with rotated beam lasers [6].

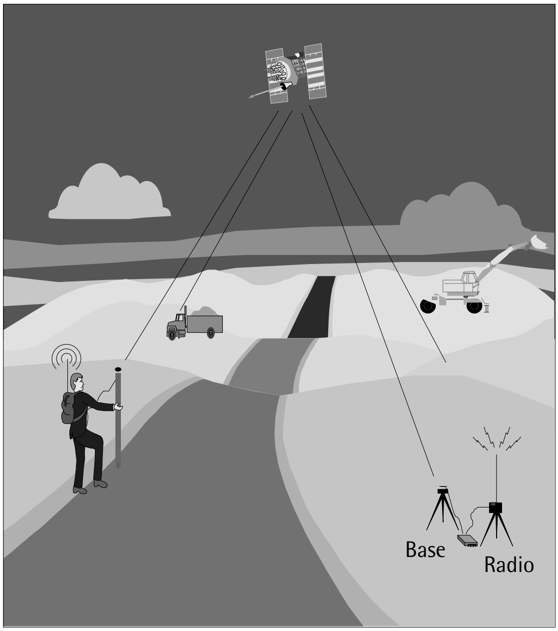

Figure 10.4 GPS for construction applications.

The same technology (i.e., combined GPS, wireless communications, and computers) is also used for foundation works (e.g., pile positioning) and precise structural placement (e.g., prefabricated bridge sections and coastal structures). In these applications, the operators are guided through the onboard computer displays, eliminating the need for conventional methods [7].

GPS is also used to track the location and usage of equipment at different sites. By sending this information to a central location, GPS enables contractors to deploy their equipment more efficiently. Moreover, vehicle operators can be efficiently guided to their destinations.

10.5 GPS for monitoring structural deformations

Since its early development, GPS has been used successfully in monitoring the stability of structures, an application that requires the highest possible accuracy. Typical examples include monitoring the deformation of dams, bridges, and TV towers. Monitoring ground subsidence of oil fields and mining areas are other examples where GPS has been used successfully. In some cases, GPS may be supplemented by other systems such as INS or total stations to work more efficiently. Deformation monitoring is done by taking GPS measurements over the same area at different time intervals [7].

Slow-deforming structures such as dams require submillimeter- to millimeter-level accuracy to monitor their displacement. Although this accuracy level may be achieved with GPS alone under certain conditions, it is not a cost-effective method [7]. To effectively monitor such structures, GPS should be supplemented with geotechnical sensors and special types of total stations.

Bridges, in contrast, are subjected to vibrations caused by dynamic traffic loads. To effectively monitor such cyclic deforming structures, dual GPS receivers should be located at several points with maximum amplitude of cyclic deformation [7]. For example, in monitoring the world's longest suspension bridge (Akashi Bridge, Japan), a GPS receiver is installed at the midpoint of the bridge while two others are installed at the main towers. Figure 10.5 shows another example in which the Ashtech Z12 dual-frequency receiver is used for monitoring bridge deformation. As the GPS data collection rate is currently limited to 10 Hz, an INS system may supplement the GPS system, in some cases, to monitor the high-frequency portion of the structure vibration.

10.6 GPS for open-pit mining

Until recently, conventional surveying was the only method available for staking drill patterns and other mining surveying. As a result of the harsh mining environment, however, stakes were often buried or displaced. In addition, drill operators had no precise way of determining the actual bit depth. Likewise, there was no way of monitoring the drill performance in the various geological layers or monitoring the haul trucks in an efficient way. More recently, however, the development of modern positioning systems and techniques, particularly RTK GPS, has dramatically improved various mining operations [8, 9]. In open-pit mines, for example, the use of RTK GPS has significantly improved several mining operations such as drilling, shoveling, vehicle tracking, and surveying. RTK GPS provides

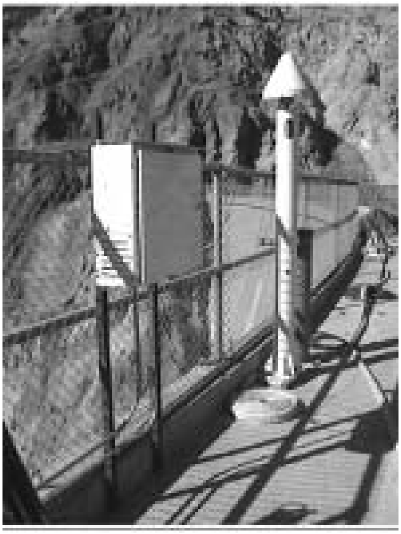

Figure 10.5 GPS for monitoring bridge deformation. (Courtesy of Magellan Corporation.)

centimeter-level positioning accuracy, and requires only one base receiver to support any number of rovers. As the pit deepens, part of the GPS signal may be blocked by the steep walls of the mine, causing a positioning problem. However, this problem, has been successfully overcome by integrating GPS with other positioning systems, mainly the pseudolite system (see Section 9.5) [10].

The mining cycle includes several phases, with ore excavation being one of the most important [11]. Excavating the ore is done by drilling a predefined pattern of blast holes, which are then loaded with explosive charges. The pattern of blast holes is designed in such a way that the size of the rock fragmentation is optimized. As such, it is important that the drills be precisely positioned over the blast holes, or otherwise redrilling may be required. An efficient way of guiding the drills is through integrating GPS with a drill navigation and monitoring system consisting of an onboard computer and drilling software. Some systems utilize two GPS receivers, mounted on the top of the drill mast, for precise real-time position and orientation of the drill. The designed drill patterns are sent to the onboard computer via radio link, and are then used by the integrated system to guide the drill operator to precisely position the drill over blast holes (see Figure 10.6). This is done automatically without staking out. In addition, the onboard computer displays other information such as the location and depth of each drill hole. This is very important as it allows the operator to view whether or not the target depth has been reached. As well, the system accumulates information on the rock hardness and the drill productivity, which can be sent to the engineering office in near real time via radio link. Such information can be used not only in monitoring the drill productivity from the engineering office, but also in understanding the rock properties, which enables better future planning [11].

GPS is also used for centimeter-level-accuracy guidance of shoveling operations (Figure 10.6). Shovels are used in loading the ore into the haul trucks, which then transport it and unload it in stockpiles. With an integrated GPS and shovel guidance and monitoring system, elevation control can be automated. With the help of the system display, shovel operators are able to keep the correct grade. This is done automatically without the need for grade control by conventional surveying methods. Similar to the drilling, shoveling productivity can be sent to the engineering office in near real time via radio link for monitoring and analysis.

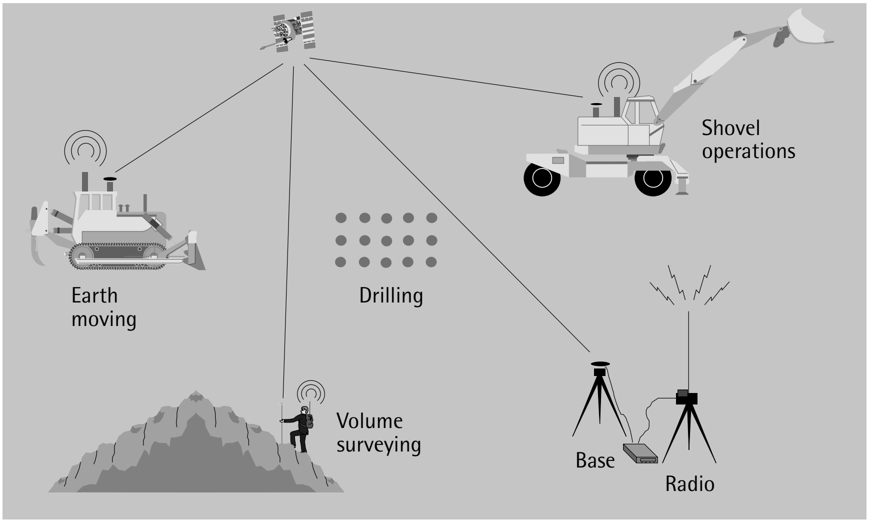

Figure 10.6 GPS use in open-pit mining.

In transporting the ore, haul trucks use continuously changing mining roads and ramps. Unless efficiently routed, safety and traffic problems would be expected, which cause an increase in the truck cycle time. The use of GPS, wireless communication, and a computer system onboard the haul trucks solve this problem efficiently. With the help of a computerized dispatch system, haul trucks can be guided to their destination using the best routes. In addition, the dispatch center can collect information on the status of each haul truck as well as the traffic conditions. Analyzing the traffic conditions is particularly important in devising a more appropriate road design [11].

GPS is also used in other phases of the mining cycle, for example, in checking the coordinates of the individual points and volume surveying. Either the RTK or the non-RTK GPS could be used for these functions (Figure 10.6).

10.7 GPS for land seismic surveying

Oil and gas exploration requires mapping of the subsurface geology through seismic surveying. In land seismic surveys, low-frequency acoustic energy is sent down into the underground rock layers (Figure 10.7). The source of the acoustic energy is often selected to be a mechanical vibrator consisting of a metal plate mounted on a truck. The plate is pressed against the ground and vibrated to produce the acoustic energy. In rough areas, dynamite is still being used as the energy source.

As the acoustic energy (signal) crosses the various underground rock layers, it is affected by the physical properties of the rocks. Portions of the signal are reflected back to the surface by the various layers. The reflected energy can be detected by special seismic devices called geophones, which are laid out at known distances from the energy source along the survey line (Figure 10.7). Upon detecting seismic energy, geophones output electrical signals that are proportional to the intensity of the reflected energy [12]. The electrical signals are then recorded on magnetic tapes for geophysical analysis and interpretation.

It is clear that unless the positions of the energy source and the geophones are known with sufficient accuracy, the very expensive seismic data becomes useless. GPS is used to provide the positioning information in a standard or a user-defined coordinate system. Integrated GPS/GLONASS

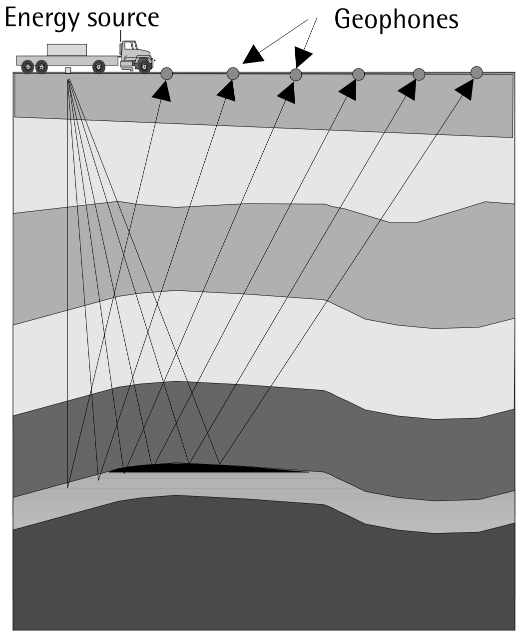

Figure 10.7 GPS for land seismic surveying.

and GPS/digital barometer systems have been used successfully in situations of poor GPS signal reception [13]. With the help of GPS, the environmental impacts (e.g., the need to cut trees) as well as the operating cost of seismic surveys have been reduced significantly.

10.8 GPS for marine seismic surveying

Marine seismic surveying is similar in principle to land seismic surveying. That is, a low-frequency acoustic energy is sent down into the subsurface rock layers, and is reflected back to the surface to reveal information about the composition of subsurface rocks (Figure 10.8).

Different methods are used in marine seismic surveys depending on the water depth. In deep waters, seismic vessels tow seismic cables, known as streamers, which contain devices called hydrophones used for detecting reflected energy. A single vessel will normally tow four to eight parallel streamers; each has a length of several kilometers [12]. The low-frequency

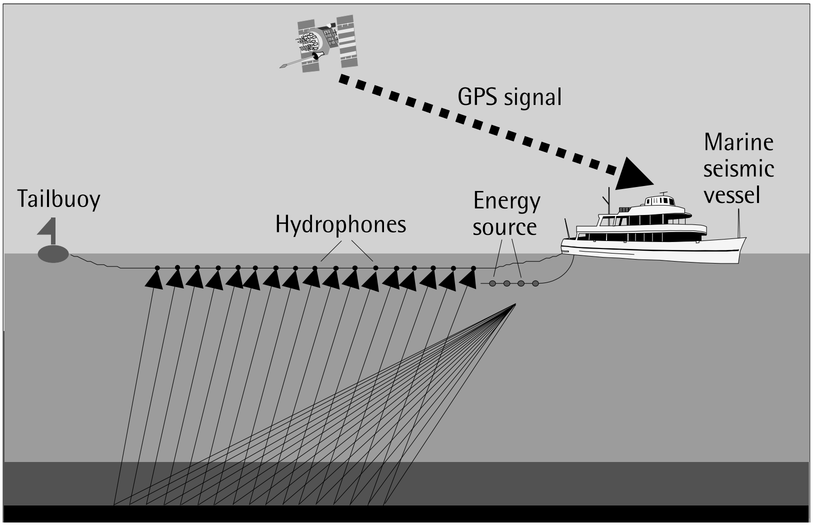

Figure 10.8 GPS for marine seismic surveying.

acoustic energy is generated using a number of air guns towed behind the vessel at about 6m below the surface. In shallow waters, both the land and the marine methods are used. Ocean bottom cable (OBC) survey is a relatively new technology that has been used recently for water depth of up to about 200m. In this method, hydrophones and geophones are combined in a single receiver to avoid water column reverberation (Figure 10.8).

To obtain meaningful results, the positions of the energy source and the hydrophones must be known with sufficient accuracy. This can be easily achieved, at lower cost, with GPS. Moreover, it is possible to revisit the points precisely with the GPS waypoint feature (see Section 10.15).

As the operation of marine seismic surveys is very expensive, the issue of quality control (QC) is essential. To maintain QC, the seismic industry has suggested the use of two independent positioning systems, with GPS being the primary one [12].

10.9 GPS for airborne mapping

GPS alone has been successfully used for topographic mapping of small-size areas. Using either conventional GPS kinematic surveying or GPS

RTK, a user takes positions of the points on the ground where the topography changes, which can be used at a later time to produce the topographic map of that area. Even in rough areas, GPS can be mounted on all-terrain vehicles (ATVs) to precisely map those areas. However, there exist situations in which the use of GPS alone becomes time-consuming and/or cost-ineffective [14]. Examples include mapping large areas, coastal areas, forests, and inaccessible areas.

Traditionally, mapping large and inaccessible areas was done using classical airborne photogrammetry. With this method, an aircraft-mounted camera is used to capture a sequence of images for the area to be mapped, which after processing construct the map. To be of practical use, the captured images must first be related to the geodetic reference system (e.g., WGS 84), a process known as georeferencing the images. In classical airborne photogrammetry, the georeferencing is done indirectly with the help of a number of ground control stations with known geodetic coordinates and their corresponding image coordinates. In recent years, GPS has been used onboard the aircraft to provide the precise position of the aerial camera as well as the precise time of each aerial exposure (Figure 10.9) [15].

The use of GPS in airborne photogrammetry has significantly reduced the required number of ground control points. It has not, however,

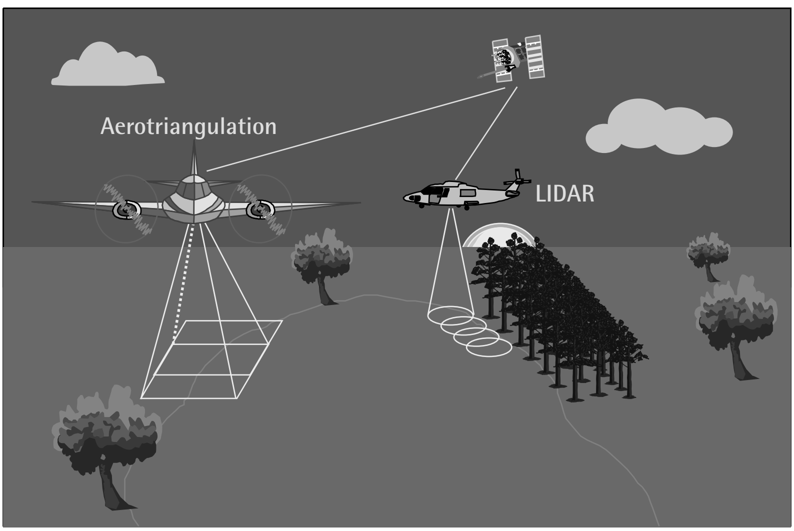

Figure 10.9 GPS for airborne mapping.

eliminated the necessity for aerial triangulation or the ground control points. This limitation, however, was overcome by augmenting GPS with a high-quality IMU; see Section 9.4 for details about this integration. The integrated GPS/inertial system provides not only the precise position of the imaging sensor but also its orientation (attitude). This allows for the captured images to be directly related to the geodetic reference system without using ground control points. In other words, direct georeferencing of the captured images can be achieved when using an integrated GPS/inertial system onboard the aircraft. In practice, however, a minimum number of ground control points may be needed for assessing the resulting accuracy [15].

Direct georeferencing using the integrated GPS/inertial system is gaining wide acceptance, and is expected to become the standard tool for rapid determination of the sensor position and orientation. With recent advances in digital airborne imaging sensors and digital photogrammetric workstations, the integrated GPS/inertial system will allow for fully digital photogrammetric workflow to be developed. With digital imaging sensors, no film development and scanning are required, which further reduces the time and the cost of photogrammetric work. Other applications such as airborne remote sensing and light detection and ranging (LIDAR) will greatly benefit from the direct georeferencing using the integrated GPS/inertial system. The latter system, LIDAR, uses an airborne laser scanner to measure the altitude of the points above the ground level [16]. Combining the GPS/inertial-based position and orientation of the laser with the measured altitude of the points leads to direct acquisition of accurate digital elevation models (DEM). Another advantage of the LIDAR system is that the data can be collected at night as well as under cloudy and high wind conditions. In addition, the ability of the laser to penetrate to the ground, even in forest areas, makes the airborne laser system attractive to the forest industry. Moreover, the LIDAR system can be used in mapping featureless areas such as deserts and areas covered by snow and ice [17].

10.10 GPS for seafloor mapping

Safe and efficient marine navigation requires, among other factors, accurate information about the water depth and the sea bottom [18]. In addition, the availability of accurate water depth is vital for making use of maximum cargo capabilities. This is especially important for areas with shallow water depth. The traditional way of obtaining the water depth was done using a single-beam echo sounder installed on a survey vessel. With this method, the single-beam echo sounder generates a sounding wave (pulse), which is transmitted to the sea bottom and then reflected back to the echo sounder (see Figure 10.10). The water depth is then computed based on the recorded travel time of the sounding pulse and the velocity of the sound in the water [18]. It should be pointed out that the echo sounder uses a hull-mounted device called the transducer to convert the electrical energy into sound energy and vice versa.

To map an area with a single-beam echo sounder, a survey vessel follows preplanned track lines while the echo sounder generates soundings along the track. Line spacing (the distance between tracks) is selected to provide the best coverage of the area. The accuracy and the reliability of the surveyed depths and locations are verified by supplementing the primary sounding lines by a series of cross lines [18]. This method is characterized by its simplicity. In addition, the echo sounder orientation is not required. A major drawback, however, is that it is time-consuming and does not provide complete coverage of the seafloor.

In recent years, a new technology for seafloor mapping has evolved that combines multibeam echo sounders, GPS, and INS. Multibeam echo

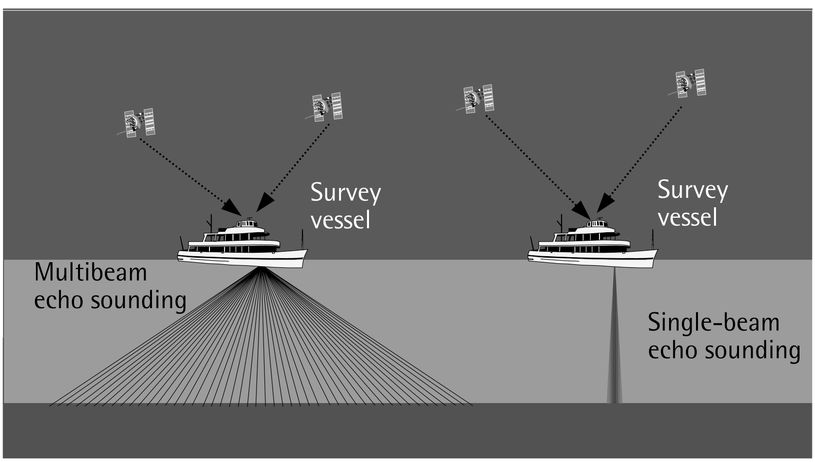

Figure 10.10 GPS for seafloor mapping.

sounders employ multiple sounding waves propagating at varying angles, which allow whole swaths of acoustic information to be collected on both sides of the track lines (see Figure 10.10). Unlike single-beam echo sounders, this multibeam technology offers complete coverage of the seafloor with high resolution, provided that the track lines are optimally designed [19]. Optimal line spacing is determined based on the approximate water depth, the footprint of sound, and the bottom profile. GPS waypoint navigation can be employed in the field to ensure that the vessel follows the designed track lines.

Because of their wide swath (usually 150°), multibeam echo sounders require accurate positioning and attitude of the vessel. This is especially important for the outer beams. Integrated GPS/INS is used for this purpose. Some manufacturers have developed an integrated GPS/INS system that utilizes two GPS receivers and antennas. Besides offering accurate positioning and attitude of the vessel, this integrated system estimates the heading of the vessel at high accuracy regardless of the vessel's dynamics and latitude.

Another state-of-the-art technology that has found wide acceptance within the hydrographic community is the airborne laser bathymetry system (LBS). LBS operates on the same principle as the land-based airborne laser system, that is, an aircraft-mounted laser sensor transmits a laser beam, which is partially reflected from the sea surface and from the seafloor. The water depth can then be computed by measuring the time difference between the returns of the two reflected pulses. An accurate 3-D seafloor map can be derived from the depth information and the GPS/inertial-based position and orientation of the laser. The major advantages of this method are high productivity and efficiency in mapping difficult areas such as narrow passages. It is, however, limited to shallow water areas (maximum depth about 50m). In addition, it is very sensitive to the water clarity.

10.11 GPS for vehicle navigation

When traveling through unfamiliar areas, vehicle drivers often use paper road maps for route guidance. However, besides being inefficient, searching for a destination using a paper map is unsafe, especially in busy areas. A new technology, incorporating GPS with digital road maps and a computer system, has been developed so that route guidance can be obtained electronically with a touch of a button [20]. Figure 10.11 illustrates this concept.

The role of GPS in this technology is to continuously determine the vehicle's location. In obstructed areas, such as urban canyons and tunnels, GPS is supplemented by a terrestrial system such as the DR system to overcome the GPS signal blockage. As discussed in Section 9.4, DR is a system that uses the vehicle's odometer and a selection from accelerometers, compasses, and gyros to determine the vehicle's direction and traveled distance. This system is accurate only over a short period of time.

The GPS-determined vehicle location is superimposed on an electronic digital road map, containing in its database digital information such as street names and directions, business listings, airports, attractions, and other related information. Once the driver inputs a destination, the built-in computer finds the best route to reach that destination. Factors such as shortest distance and time to destination, one-way roads, illegal turns, and rush-hour restrictions, are all considered in the path finding. Some systems allow the drivers to input other factors such as accident avoidance. The driver usually gets turn-by-turn instructions, with audio and/or visual indications, to the destination. If the driver misses a turn, the system displays a warning message and finds an alternative best route based

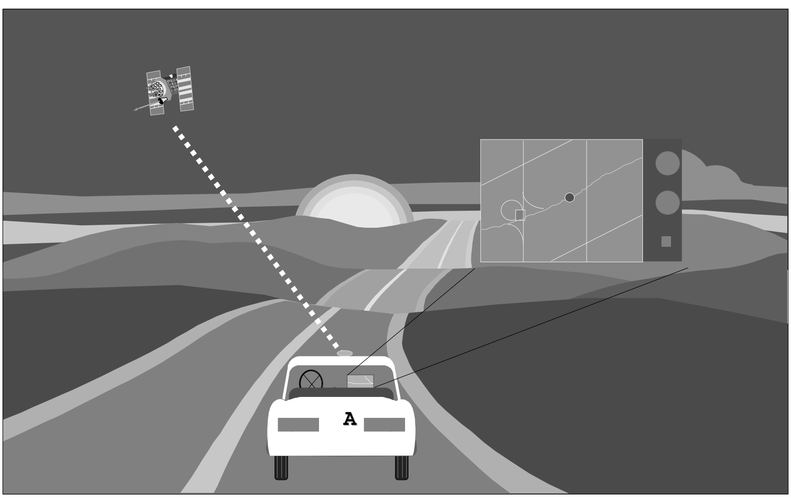

Figure 10.11 GPS for vehicle navigation.

on the current location of the vehicle. Some manufacturers add cellular systems to provide weather and traffic information and to locate the vehicles in case of emergency. Recent advances in wireless communication technology even make it possible for drivers to remotely access the Internet from their vehicles [21].

10.12 GPS for transit systems

Transit system authorities in many countries are faced with a challenging trend of fiscal constraints, which limits their capabilities to expand existing services and to increase ridership. Until recently, transit systems used old technologies such as odometer/compass sensors and signposts for position determination [22]. Odometers are sensors that measure the number of rotation counts generated by the vehicle's wheels, which are then used to estimate the distance traveled by the vehicle. With the help of a compass, the vehicle's direction of travel can be determined at any time. Combining the measurements from the odometer and the compass, the vehicle's position can be determined with respect to an initial (known) position. Unfortunately, both the odometer and the compass drift over time, which causes significant error in the estimated position. Signposts, in contrast, are radio beacon transmitters that are placed at known locations along the bus routes [23]. Each beacon transmits a low-power microwave signal, which is detected by a receiver on the bus, to account for the odometer's drift error. Unfortunately, this system has a number of limitations, including its incapability of knowing the exact location of a vehicle in between two signposts. In addition, it is not possible to track a vehicle that goes off-route as a result of, for example, a road closure [22].

To overcome the limitations of these systems, transit authorities are integrating a low-cost autonomous GPS system with one or more of these conventional systems. GPS helps in controlling the drift of the conventional systems through frequent calibration. In addition, the vehicle's position can be obtained reliably with GPS if the vehicle goes off-route. However, since some of the GPS signals will be obstructed in areas with high-rise buildings, such as downtown areas, the vehicle's position may be obtained with the help of conventional systems. As such, the performance of the integrated system is indeed better than either positioning system alone.

The integrated positioning system not only helps the transit authorities to locate their fleet of buses on a digital base map in real time, but also helps in performing other advanced functions (see Figure 10.12). For example, if the bus locations are available in real time, the bus arrival times at the bus stops can be computed reliably, thus minimizing the waiting time at the bus stops. This is a very important feature, especially under severe weather conditions. In addition, the availability of the real-time bus location information enables the transit authorities to dynamically design more efficient bus scheduling, thus improving bus efficiency and customer service. This information can be accessed through the Internet, greatly enhancing customer satisfaction [22].

10.13 GPS for the retail industry

In today's competitive market, efficiency and cost reduction play a significant role in keeping a retailer in business. GPS integrated with a GIS can help in achieving that goal. Other technologies such as wireless data communication and speech recognition are also becoming key components.

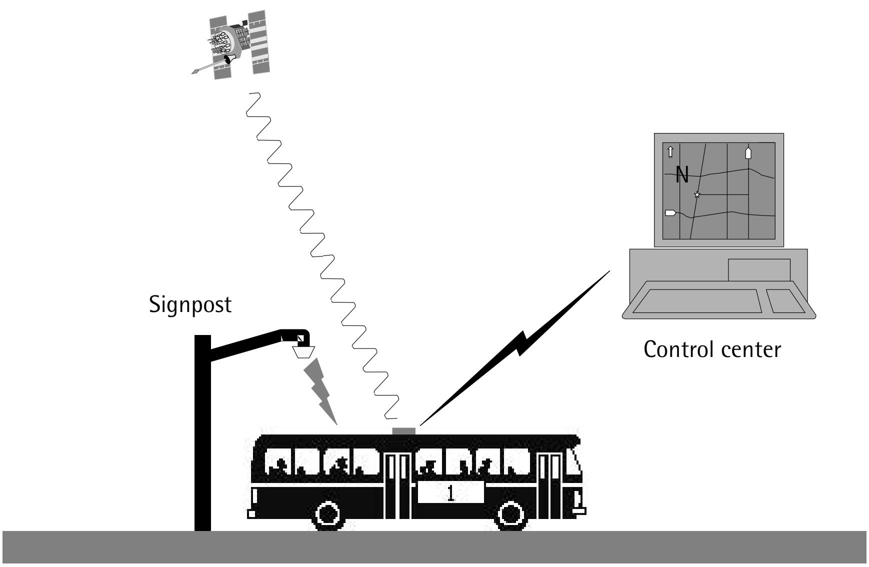

Figure 10.12 GPS for transit systems.

This section describes how the different technologies are integrated to ensure efficiency and cost reduction in two components of the retail business: delivery and real-time inventory monitoring.

The integrated system consists of two main components: an efficient route analysis for the delivery area and a GPS-based real-time fleet-monitoring system. In designing an efficient route analysis, it is necessary that an up-to-date digital map (street map database) for the delivery area be available. In addition, retail stores' location database should also be available. Based on these two sets of databases as well as the traffic conditions, it is possible to optimize the vehicles' routing and scheduling. A software package such as ESRI's Arclogistics may be used for that optimization purpose [24]. An optimized real-time fleet-monitoring system requires the availability of both GPS and a suitable wireless data communication. Speech recognition technology may also be used to speed up the delivery service.

To ensure that the GPS-derived position and the map database for the proposed delivery area are compatible, the digital map should be based on WGS 84, the parent datum for GPS. In addition, vector maps rather than raster maps should be used to increase flexibility. Information about driving restrictions, such as one-way streets or vehicle-type restrictions, should accompany the street map database. Retail stores' location is another database that should also be available. Similar to the street map database, the stores' location database should be based on WGS 84. Ultimately, the actual network drive times as a function of the time of day should be considered to optimize the route analysis. Other factors that should be considered in the route analysis include order characteristics, delivery time windows, and the varying vehicles' volume, weight, and operating costs. A good software package should have the capability of using this information to find the best route and stop sequence of each vehicle in the fleet. In addition, the software should have the capability of rerouting a vehicle in cases such as road closure or unexpected heavy traffic due to, for example, accidents. Moreover, the software should have the capability of producing detailed maps with directions to the drivers and a summary report to the dispatching center.

With this optimization process, the routing time would be reduced, which means that the number of vehicles needed for the same service is reduced. This ensures that the delivery cost is reduced. Moreover, as a result of the optimization process, it is expected that the delivery time windows will be reduced as well. This means that goods are delivered more efficiently.

The second main component of the integrated system is a GPS-based real-time fleet-monitoring system. Low-cost standalone GPS receivers may be used, as only low-accuracy positioning information is needed. To avoid GPS signal obstruction, it may be necessary for some vehicles in the fleet that operate in urban areas to be equipped with another, complementary, navigation system such as the DR system [22, 25]. In order for the dispatcher to locate a vehicle in the fleet, it is necessary that the GPS position be sent to the dispatching center via communication link [26]. This could be done in either real time or near real time, depending on the need. In fact, a combination of both might be an ideal solution, if the system is used in-state and out-of-state. At each delivery site, the driver may send some information about the products being delivered. Speech-recognition technology could be used to transfer the driver's voice into digital data that would be sent along with the GPS positioning information. Alternatively, bar codes may be used for the same purpose. The vehicle position should be sent automatically at each delivery site. This may be done by comparing the vehicle's location and the stored location of the delivery site.

The received vehicles' position along with other information would be overlaid on a base map, which helps the dispatcher to locate each vehicle in the fleet and to know of its contents. This also helps in making efficient decisions, for example, in case of an emergency. Moreover, better estimates for arrival times, based on the actual vehicles' location, can be made. If necessary, dynamic rerouting may be done upon receiving new information on the fleet.

Monitoring the fleet as indicated ensures that each driver follows the preassigned route (i.e., it gives the authorities a means for driver accountability). Although this might be a good feature in the system, the driver's privacy should be considered by, for example, restricting the accessibility of certain information to authorized users.

10.14 GPS for cadastral surveying

Cadastral surveys establish property corners, boundaries, and areas of land parcels [6]. Conventional surveying methods have been used, and are still being used, for that purpose. Conventional methods, however, have the drawback that extensive traversing is required. Moreover, extensive clear-cutting and intervening private properties might be required as well. GPS overcomes these conventional-method drawbacks.

Any of the GPS surveying methods, such as kinematic GPS or RTK GPS, can be used depending on the project requirements, location, and other factors. The RTK surveying, however, seems to be the most suitable method, especially in unobstructed areas. This is mainly because of its ease of use and the availability of the results while in the field. Inaccessible locations or obstructed areas can be surveyed with integrated systems such as GPS/LRF or GPS/total station.

There are several advantages of using GPS for cadastral surveying. The most important one is that intervisibility between the points is not required with GPS. This means that extensive traversing is eliminated, clear-cutting is not required, and intervening private properties is avoided. Other advantages include the fact that GPS provides user-defined coordinates in a digital format, which can be easily exported to any GIS system for further analysis. The accuracy obtained with GPS is consistent over the entire network; such accuracy is lacked by conventional surveying methods. Also, with GPS, one reference station can support an unlimited number of rover receivers. A number of governmental and private organizations have reported that the use of GPS in cadastral surveying is cost-effective.

10.15 GPS stakeout (waypoint navigation)

Waypoint navigation, or stakeout as it is called by surveyors, provides guidance to a GPS user in reaching his or her destination in the best way (shortest time and/or distance). By feeding the GPS receiver (or the GPS receiver controller) with the coordinates of his or her destination, a GPS user receives on-screen guidance instantaneously (see Figure 10.13 for details). Surveyors use this principle to lay out points and lines.

The idea behind GPS waypoint navigation is simple. As a first step, the user must feed the GPS receiver (or the GPS controller) with the coordinates of his or her destination. Most GPS receivers are capable of storing a number of destination points (waypoints) in their internal memory. The second step is to let the GPS receiver compute its own position, that is, the user's positions). Based on the receiver and the destination positions, the built-in receiver computer calculates the distance and the azimuth of

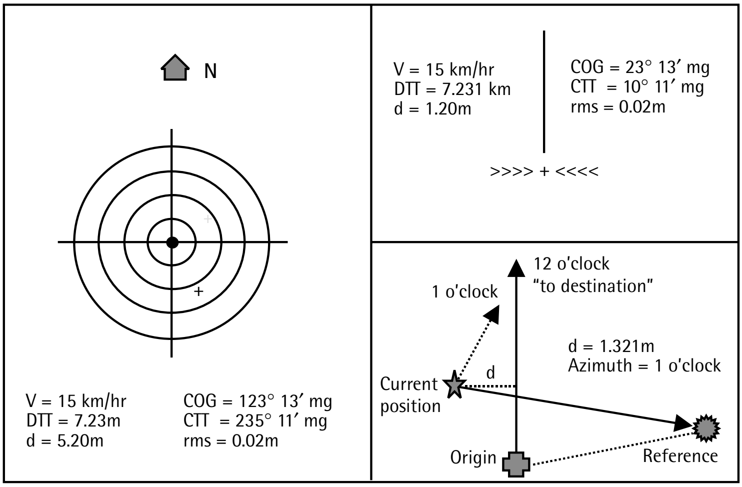

Figure 10.13 GPS waypoint navigation.

the line connecting the receiver's position and the destination points. The built-in computer uses the position information to calculate other parameters such as the expected arrival time to the user's destination based on the user's speed. In addition, the offset distance from the receiver position to the original line between the starting point and the destination can be calculated. All of this information and other data are displayed on a continuous basis to guide the GPS user.

This guidance information can be displayed in different ways [25]. One of these displays is the bull's-eye, where the destination point is located at the center of the displayed concentric circles while the user's location is displayed as a moving cursor. The top point of the bull's-eye is normally selected to represent the north. The user will reach his or her destination point when the moving cursor stays at the center of the concentric circles. In addition to this, a number of navigation parameters are displayed to help the user as well.

References

[1] Laser Technology Inc., "Survey Laser for Forestry," PowerPoint Presentation, accessed July 18, 2001, http://www.lasertech.com/downloads.html.

[2] Phillips, B., "GPS Field Applications in Forestry Consulting," *Global Positioning System in Forestry Workshop*, Kelowna, British Columbia, Canada, November 25–28, 1996.

[3] Bauer, W. D., and M. Schefcik, "Using Differential GPS to Improve Crop Yields," *GPS World*, Vol. 5, No. 2, February 1994, pp. 38–41.

[4] Petersen, C., "Precision GPS Navigation for Improving Agriculture Productivity," *GPS World*, Vol. 2, No. 1, January 1991, pp. 38–44.

[5] Smith, B. S., "GPS Grade Control for Construction," *Proc. ION GPS 2000*, *13th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Salt Lake City, UT, September 19–22, 2000, pp. 1034–1037.

[6] Elfick, M., et al., *Elementary Surveying*, 8th ed., New York: HarperCollins, 1994.

[7] El-Rabbany, A., A. Chrzanowski, and M. Santos, "GPS Applications in Civil Engineering," *Ontario Land Surveyor Quarterly*, Summer 2001, pp. 6–8.

[8] Flinn, J. A., C. Waddell, and M. A. Lowery, "Practical Aspects of GPS Implementation at the Morenci Copper Mine," *Proc. ION GPS-99*, *12th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Nashville, TN, September 14–17, 1999, pp. 915–919.

[9] Flinn, J. A., and S. M. Shields, "Optimization of GPS on Track-Dozers at a Large Mining Operation," *Proc. ION GPS-99*, *12th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Nashville, TN, September 14–17, 1999, pp. 927–931.

[10] Dai, L., et al., "GPS and Pseudolite Integration for Deformation Monitoring Applications," *Proc. ION GPS 2000*, *13th Intl. Technical Meeting*, Satellite Division, Institute of Navigation, Salt Lake City, UT, September 19–22, 2000, pp. 1–8.

[11] Shields, S., J. Flinn, and A. Obregon, "GPS in the Pits: Differential GPS Application at the Morenci Copper Mine," *GPS World*, Vol. 11, No. 10, October 2000, pp. 34–39.

[12] Jensen, M. H., "Quality Control for Differential GPS in Offshore Oil and Gas Exploration," *GPS World*, Vol. 3, No. 8, September 1992, pp. 36–48.

[13] McLintock, D., G. Deren, and E. Krakiwsky, "Environment Sensitive: DGPS and Barometry for Seismic Surveys," *GPS World*, Vol. 5, No. 2, February 1994, pp. 20–26.

[14] Flood, M., B. Gutelius, and M. Orr, "Airborne Terrain Mapping," *Earth Observation Magazine (EOM),* Vol. 6, No. 2, February 1997, pp. 40–42.

[15] Cramer, M., "On the Use of Direct Georeferencing in Airborne Photogrammetry," *Proc. 3rd Intl. Symp. Mobile Mapping Technology,* Cairo, Egypt, January 3–5, 2001, CD-ROM.

[16] Abwerzger, G., "Georeferencing of Laser Scanner Data Using GPS Attitude and Position Determination," *Proc. 3rd Intl. Symp. Mobile Mapping Technology,* Cairo, Egypt, January 3–5, 2001, CD-ROM.

[17] Favey, E., et al., "3D-Laser Mapping and Its Application in Volume Change Detection of Glaciers," *Proc. 3rd Intl. Symp. Mobile Mapping Technology,* Cairo, Egypt, January 3–5, 2001, CD-ROM.

[18] Bowditch, N., *American Practical Navigator*, Bethesda, MD: Defense Mapping Agency Hydrographic/Topographic Center, 1995.

[19] Maxfield, H. E., "Recent Developments in Seafloor Mapping Capabilities," *Hydro International*, Vol. 2, No. 1, January/February 1998, pp. 45–47.

[20] Zhao, Y., *Vehicle Location and Navigation Systems*, Norwood, MA: Artech House, 1997.

[21] Hada, H., et al., "The Internet, Cars, and DGPS: Bringing Mobile Sensors and Global Correction Services On Line," *GPS World*, Vol. 11, No. 5, May 2000, pp. 38–43.

[22] El-Rabbany, A., A. Shalaby, and S. Zolfaghari, "Real-Time Bus Location, Passenger Information and Scheduling for Public Transportation," Presented at GPS meeting, *GOEIDE 2000 Conference*, Calgary, Alberta, Canada, May 24–26, 2000.

[23] Drane, C., and C. Rizos, *Positioning Systems in Intelligent Transportation Systems*, Norwood, MA: Artech House, 1998.

[24] ESRI, "ArcLogistics Route," *Business Solution News*, accessed September 17, 2001, http://www.esri.com/software/arclogistics/index.htm.

[25] SOKKIA Corporation, *GSR2300 Operation Manual*, 1996.

[26] Beerens, J. A. J., "Fleet Monitoring with GPS and Satellite Communications," *GPS World*, Vol. 4, No. 4, April 1993, pp. 42–46.

11
Other Satellite Navigation Systems

11.1 GLONASS satellite system

GLONASS is an all-weather global navigation satellite system developed by Russia. The GLONASS satellite system has much in common with the GPS system. The nominal constellation of the GLONASS system consists of 21 operational satellites plus three spares at a nominal altitude of 19,100 km. Eight GLONASS satellites are arranged in each of three orbital planes (see Figure 11.1). GLONASS orbits are approximately circular, with an orbital period of 11 hours and 15 minutes and an inclination of 64.8° [1, 2].

Similar to GPS, each GLONASS satellite transmits a signal that has a number of components: two L-band carriers, C/A-code on L1, P-code on both L1 and L2, and a navigation message. However, unlike GPS, each GLONASS satellite transmits its own carrier frequencies in the bands 1,602–1,615.5 MHz for L1 and 1,246–1,256.5 MHz for L2, depending on the channel number. These two bands are on their way to being shifted to 1,598.0625–1,604.25 MHz and 1,242.9375–1,247.75 MHz, respectively, to avoid interference with radio astronomers and operators of low-Earth-orbiting satellites. With this shift, each pair of GLONASS satellites will be assigned the same L1 and L2 frequencies. The satellite pairs, however, will

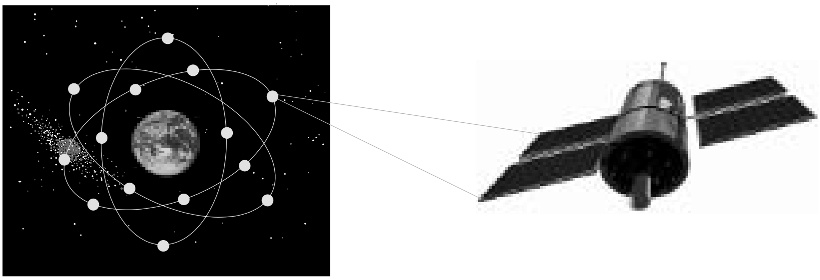

Figure 11.1 GLONASS system. (Satellite image courtesy of Magellan Corporation.)

be placed on the opposite sides of the Earth (antipodal), which means that a user cannot see them simultaneously. GLONASS codes are the same for all the satellites. As such, GLONASS receivers use the frequency channel rather than the code to distinguish the satellites. The chipping rates for the P-code and the C/A-code are 5.11 and 0.511 Mbps, respectively. The GLONASS navigation message is a 50-bps data stream, which provides, among other things, the satellite ephemeris and the channel allocation [2]. The signal of GLONASS system is not affected by either SA or antispoofing. The GLONASS system completed 24 working satellites in January 1996. Unfortunately, however, the number of GLONASS satellites had dropped to only seven satellites by May 2001 [3]. It is expected that a new genera-m tion GLONASS satellites, GLONASS-M, will be launched in the near future. GLONASS-M has a lifetime of 5 years, improved onboard atomic clocks, and the facility to transmit the C/A code on both L1 and L2 carrier frequencies [4].

GPS and GLONASS systems may be integrated to improve geometry and positioning accuracy, particularly under poor satellite visibility, such as in urban areas. There are, however, two problems with GPS/GLONASS integration. The first one is that both GPS and GLONASS systems use different coordinate frames to express the position of their satellites. GPS uses the WGS 84 system, while GLONASS uses the Earth Parameter System 1990 (PZ-90) system. The two systems differ by as much as 20m on the Earth's surface. The transformation parameters between the two systems may be obtained by simultaneously observing reference points in both systems. Various research groups have developed various sets of transformation parameters [2, 5]. However, accurate determination of the transformation parameters is still unavailable. The second problem with the GPS/GLONASS integration is that both systems use different reference times. The offset between the two time systems changes slowly and reaches several tens of microseconds. One way of determining the time offset is by treating it as an additional variable in the receiver solution.

11.2 Chinese regional satellite navigation system (Beidou system)

China has recently launched two domestically built navigation satellites, which form the first generation of a satellite-based navigation system [6–8]. It is an all-weather regional navigation system, which is known as the Beidou Navigation System. The satellites are placed in geostationary orbits at an altitude of approximately 36,000 km above the Earth's surface. The primary use of the system is in land and marine transportation. China is also planning to build its second-generation satellite positioning and navigation system, which will have more satellites and more coverage area.

11.3 Regional augmentations

The current satellite-based global navigation systems, GPS and GLONASS, do not meet all of the civil aviation requirements. To overcome these limitations, regional augmentation systems are currently being developed. A regional augmentation system typically combines one or more satellite constellations such as GPS and GLONASS, geostationary satellites equipped with navigation transponders and a number of ground reference stations [9]. Merging the various interoperable regional systems leads to a Global Navigation Satellite System (commonly known as GNSS-1) that meets the civil aviation requirements. In fact, the International Civil Aviation Organization (ICAO) has endorsed the GNSS as the core system for international aviation use [9].

Various regional augmentation systems are currently being developed as part of the worldwide GNSS. The United States is developing a GPS-based regional system called the Wide Area Augmentation System (WAAS), which covers North America with the possibility of extending to include South America. Europe is developing a similar regional system called European Geostationary Navigation Overlay System (EGNOS), which is based on both GPS and GLONASS. It covers Europe and North Africa with the possibility of extending to include all of Africa and the Middle East. A third regional GPS-based system, called Multi-function Transport Satellite (MTSAT), is being developed in Japan, and covers parts of Asia and the Pacific region. Australia is also in the process of developing its own regional system. The regional systems are expected to merge and be interoperable [9].

11.4 Future European global satellite navigation system (Galileo system)

Galileo is a satellite-based global-navigation system proposed by Europe. Galileo is a civil-controlled satellite system to be delivered through a public-private partnership [10]. Three different constellation types were investigated to ensure the optimum selection of the Galileo architecture, namely low Earth orbits (LEO), medium Earth orbits (MEO), and inclined geosynchronous orbits (IGSO). Combinations of various constellation types were also studied. Following this study, the Galileo decision makers adopted a constellation of 30 MEO satellites. The satellites will be evenly distributed over three orbital planes at an altitude of about 23,000 km. This selection ensures that more uniform performance is obtained for all regions (i.e., independent of the region's latitude). The signal characteristics of the Galileo system were to be determined sometime in 2001 [10].

Galileo will be compatible at the user level with the existing GPS and GLONASS systems. However, unlike GPS and GLONASS, Galileo will provide two levels of services: a basic, free-of-direct-charge service and a chargeable service that offers additional features. Some security measures, such as withholding of the service, have been studied to ensure that the system is properly used. A European political body, independent of Galileo management, will have the authority to take the proper measures in the event of a crisis.

The Galileo development plan will be divided into three different phases.

1. The definition phase was concluded at the end of 2000.
2. The development and validation phase began in 2001 and has been extended for a period of 4 years. This phase comprises a more

detailed definition of the Galileo system (e.g., frequency allocation). As well, it includes the construction of the various segments of the system (space, ground, and receiver). Some prototype satellites will be launched in 2004, along with the establishment of a minimal ground infrastructure, to validate the system.

3. The constellation deployment phase is scheduled to begin in 2006 and extend until 2007. With the experience gained during the system validation phase, operational satellites will be gradually launched during this phase. In addition, ground infrastructure will be completed.The target date for the gradual introduction of Galileo operational service is 2008 or shortly thereafter. At that time, EGNOS service will be provided in parallel until it is phased out in 2015 [10].

References

[1] Kleusberg, A., "Comparing GPS and GLONASS," *GPS World*, Vol. 1, No. 6, November/December 1990, pp. 52–54.

[2] Langley, R. B., "GLONASS: Review and Update," *GPS World*, Vol. 8, No. 7, July 1997, pp. 46–51.

[3] *Navtech Seminars and GPS Supply*, GPS/GNSS newsletter, May 17, 2001.

[4] Johnson, N. L., "GLONASS Spacecraft," *GPS World*, Vol. 5, No. 11, November 1994, pp. 51–58.

[5] Bazlof, Y. A., et al., "GLONASS to GPS: A New Coordinate Transformation," *GPS World*, Vol. 10, No. 1, January 1999, pp. 54–58.

[6] CANSPACE, "China Puts Second Navigation Positioning Satellite into Orbit," *Canadian Space Geodesy Forum*, December 21, 2000.

[7] CANSPACE, "Chinese Launch Navigation Satellite," *Canadian Space Geodesy Forum*, October 31, 2000.

[8] CANSPACE, "Chinese Satellite Navigation System Update," *Canadian Space Geodesy Forum*, January 9, 2001.

[9] Kayton, M., and W. R. Fried, *Avionics Navigation Systems,* 2nd ed., New York: Wiley, 1997.

[10] Commission of the European Communities, "Commission Communication to the European Parliament and the Council on GALILEO," Brussels, Belgium, November 22, 2000.

Appendix A
GPS Accuracy and Precision Measures

The term *accuracy* is used to express the degree of closeness of a measurement, or the obtained solution, to the true value. The term *precision*, however, is used to describe the degree of closeness of repeated measurements of the same quantity to each other. In the absence of systematic errors, accuracy and precision would be equivalent [1]. For this reason, the two terms are used indiscriminately in many practical purposes. Accuracy can be measured by a statistical quantity called the standard deviation, assuming that the GPS measurements contain no systematic errors or blunders. The lower the standard deviation, the higher the accuracy.

For the 1-D case, for example, measuring the length of a line between two points, the accuracy is expressed by the so-called root mean square (rms). The rms is associated with a probability level of 68.3%. For example, the accuracy of the static GPS surveying could be expressed as "5 mm + 1 ppm" (rms). This means that there is a 68.3% chance (or probability) that we get an error of less than or equal to "5 mm + 1 mm for every kilometer." In other words, if we measure a 10-km baseline, then there is a 68.3%

chance that we get an error of less than or equal to 15 mm in the measured line.

Horizontal component (e.g., easting and northing) accuracy, a 2-D case, is expressed by either the circular error probable (CEP) or twice distance rms (2drms). CEP means that there is a 50% chance that the true horizontal position is located inside a circle of radius equal to the value of CEP [1]. The corresponding probability level of the 2drms varies from 95.4% to 98.2% depending on the relative values of the errors in the easting and northing components. The ratio of the 2drms to the CEP varies from 2.4 to 3. This means that an accuracy of 40m (CEP) is equivalent to 100m (2drms) for a ratio of 2.5.

The spherical error probable (SEP) is used to express the accuracy of the 3-D case. SEP means that there is a 50% chance that the true 3-D position is located inside a sphere of a radius equal to the value of SEP [1].

Reference

[1] Mikhail, E., *Observations and Least Squares,* New York: University Press of America, 1976.

Appendix B
Useful Web Sites

B.1 GPS/GLONASS/Galileo information and data

Canadian Active Control System (CACS data and service):
http://www.geod.nrcan.gc.ca/htmlpublic/GSDproducts-Guide/CACS/English/cacstest.html

Geodetic Survey Division of Geomatics Canada:
http://www.geod.nrcan.gc.ca/index_English_text_based.html

Galileo EC page:
http://www.Galileo-pgm.org/

Galileo World magazine:
http://www.galileosworld.com/

GPS World magazine:
http://www.gpsworld.com/

International GPS Service for Geodynamics (RINEX and precise ephemeris data):
http://igscb.jpl.nasa.gov/

International Terrestrial Reference Frame (ITRF):
http://large.ensg.ign.fr/ITRF/index-old.html

Ministry of Defence of the Russian Federation (GLONASS Web page):
http://www.rssi.ru/SFCSIC/English.html

National Imagery and Mapping Agency (NIMA):
http://164.214.2.59/nimahome.html

Navtech Seminars and GPS Supply:
http://www.navtechgps.com/

UNB Internet resources:
http://gauss.gge.unb.ca/GPS.INTERNET.SERVICES.HTML

University Navstar Consortium (UNAVCO data and service):
http://www.unavco.ucar.edu/

U.S. Coast Guard Navigation Center (GPS NANU, GPS Almanac, FRP, and others):
http://www.navcen.uscg.gov/

U.S. Continuously Operating Reference Station (CORS data):
http://www.ngs.noaa.gov/CORS

U.S. National Geodetic Survey:
http://www.ngs.noaa.gov/index.shtml

U.S. National Geodetic Survey GEOID page:
http://www.ngs.noaa.gov/GEOID

U.S. National Geodetic Survey Orbit data:
http://www.ngs.noaa.gov/GPS/GPS.html

U.S. Naval Observatory (GPS timing data and information):
http://tycho.usno.navy.mil/gps_datafiles.html

B.2 GPS manufacturers

Applanix Corporation (integrated systems):
http://www.applanix.com/

Integrinautics (pseudolites):
http://www.integrinautics.com/

Leica:
http://www.leica.com/

Magellan Corporation (Ashtech precision products):
http://www.ashtech.com/

NovAtel:
http://www.novatel.ca

Pacific Crest Corporation (radio link systems):
http://www.paccrst.com/

SOKKIA Corporation:
http://www.sokkia.com/

Trimble Navigation:
http://www.trimble.com/

About the Author

Dr. Ahmed El-Rabbany is an assistant professor in the Department of Civil Engineering, Ryerson University, in Toronto, Canada. He received a Ph.D. in GPS from the Department of Geodesy and Geomatics Engineering at the University of New Brunswick. He also worked in the same department as a postdoctoral fellow and as an assistant professor. Dr. El-Rabbany has more than 17 years of research, instructional, and industrial experience in the general discipline of geomatics engineering, with specializations in GPS, geodesy, data modeling and estimation, and hydrographic surveying. He leads a number of research activities in the areas of GPS, integrated navigational chart systems, and integrated navigation systems for land navigation and hydrographic surveying. Dr. El-Rabbany currently holds leading positions with a number of local, national, and international professional organizations that directly influence the geomatics profession. He was recently appointed an honorary research associate and an adjunct professor at the University of New Brunswick and York University, respectively.

Index