ISSN 0280-5731, LMV-rapport 1997:16

Reports in Geodesy and Geographical Information Systems

GEODETIC APPLICATIONS of GPS

LECTURE NOTES for
NORDIC AUTUMN SCHOOL

organized by
NORDIC GEODETIC COMMISSION

BÅSTAD, SWEDEN
26 - 31 August 1996

edited by BO JONSSON

NATIONAL LAND SURVEY

Gävle, Sweden
1997

ISSN 0280-5731, LMV-rapport 1997:16

Reports in Geodesy and Geographical Information Systems

GEODETIC APPLICATIONS of GPS

LECTURE NOTES for
NORDIC AUTUMN SCHOOL

organized by
NORDIC GEODETIC COMMISSION

BÅSTAD, SWEDEN
26 - 31 August 1996

edited by BO JONSSON

NATIONAL LAND SURVEY

Gävle, Sweden
1997

PREFACE

The Fourth NKG Autumn School in Geodesy organised by the Nordiska Kommissionen för Geodesi (Nordic Geodetic Commission; NKG) was held in Båstad, Sweden during August 26 - 31, 1996. The NKG is mainly concerned with the establishment of national/global reference frames and reference networks and studies of crustal movements that affect the stability of these reference networks. Geodetic Applications of the GPS was chosen as the main topic for the Fourth NKG Autumn School, since the GPS technique has created new and very powerful tools for the establishment of national reference frames and studies of crustal movements.

The GPS technique provides facilities for a large number of different applications from navigation with an accuracy of one metre to studies of crustal movements with an accuracy of one millimetre. GPS-related geodetic research work is currently being undertaken at many Nordic research institutes, using various approaches. The goal of the Nordic Autumn School in Geodetic Applications of GPS was to bring together Nordic geodesists, geophysicists and other GPS-related scientists and a number of internationally outstanding GPS experts to review the latest developments and to exchange experiences. The large number of applicants for the available places in the Autumn School demonstrated the wide interest shown in the subject.

The NKG Autumn School was sponsored mainly by NorFA (Nordisk Forskerutdanningsakademi) but also by the geodetic institutes of Denmark, Finland, Norway and Sweden. The NKG Presidium would like to take this opportunity to thank NorFA and the geodetic institutes for their valuable support.

The success of the Autumn School can also be attributed to the high professional standard of the presentations given at Båstad. The NKG Presidium wishes to thank all the lecturers for agreeing to come and present their recent scientific findings and other material related to the GPS.

Båstad is an idyllic small town (at least when the Swedish holiday season is over) and affords ample opportunities for participants to meet and discuss GPS experiences over a beer or coffee. There were also other opportunities to meet outside the official programme, for example on the sea shore just outside the hotel or on the famous Båstad tennis courts.

Juhani Kakkuri
Chairman of the NKG Presidium
Masala, Finland,
August 18, 1997

CONTENTS

Basics of the GPS Technique: Observation Equations[§]

Geoffrey Blewitt

Department of Geomatics, University of Newcastle
Newcastle upon Tyne, NE1 7RU, United Kingdom

geoffrey.blewitt@ncl.ac.uk

Table of Contents

1. INTRODUCTION

The purpose of this paper is to introduce the principles of GPS theory, and to provide a background for more advanced material. With that in mind, some of the theoretical treatment has been simplified to provide a starting point for a mathematically literate user of GPS who wishes to understand how GPS works, and to get a basic grasp of GPS theory and terminology. It is therefore not intended to serve as a reference for experienced researchers; however, my hope is that it might also prove interesting to the more advanced reader, who might appreciate some "easy reading" of a familiar story in a relatively short text (and no doubt, from a slightly different angle).

2. GPS DESCRIPTION

In this section we introduce the basic idea behind GPS, and provide some facts and statistics to describe various aspects of the Global Positionining System.

2.1 THE BASIC IDEA

GPS positioning is based on trilateration, which is the method of determining position by measuring distances to points at known coordinates. At a minimum, trilateration requires 3 ranges to 3 known points. GPS point positioning, on the other hand, requires 4 "pseudoranges" to 4 satellites.

This raises two questions: (a) "What are pseudoranges?", and (b) "How do we know the position of the satellites?" Without getting into too much detail at this point, we address the second question first.

2.1.1 How do we know position of satellites?

A signal is transmitted from each satellite in the direction of the Earth. This signal is encoded with the "Navigation Message," which can be read by the user's GPS receivers. The Navigation Message includes orbit parameters (often called the "broadcast ephemeris"), from which the receiver can compute satellite coordinates (X,Y,Z). These are Cartesian coordinates in a geocentric system, known as WGS-84, which has its origin at the Earth centre of mass, Z axis pointing towards the North Pole, X pointing towards the Prime Meridian (which crosses Greenwich), and Y at right angles to X and Z to form a right-handed orthogonal coordinate system. The algorithm which transforms the orbit parameters into WGS-84 satellite coordinates at any specified time is called the "Ephemeris Algorithm," which is defined in GPS textbooks [e.g., Leick, 1991]. We discuss the Navigation Message in more detail later on. For now, we move on to "pseudoranges."

2.1.2 What are pseudoranges?

Time that the signal is transmitted from the satellite is encoded on the signal, using the time according to an atomic clock onboard the satellite. Time of signal reception is recorded by receiver using an atomic clock. A receiver measures difference in these times:

$$\text{pseudorange} = (\text{time difference}) \times (\text{speed of light})$$

Note that pseudorange is almost like range, except that it includes clock errors because the receiver clocks are far from perfect. How do we correct for clock errors?

2.1.3 How do we correct for clock errors?

Satellite clock error is given in Navigation Message, in the form of a polynomial. The unknown receiver clock error can be estimated by the user along with unknown station coordinates. There are 4 unknowns; hence we need a minimum of 4 pseudorange measurements.

2.2 THE GPS SEGMENTS

There are four GPS segments:

- the Space Segment, which includes the constellation of GPS satellites, which transmit the signals to the user;
- the Control Segment, which is responsible for the monitoring and operation of the Space Segment,
- the User Segment, which includes user hardware and processing software for positioning, navigation, and timing applications;
- the Ground Segment, which includes civilian tracking networks that provide the User Segment with reference control, precise ephemerides, and real time services (DGPS) which mitigate the effects of "selective availability" (a topic to be discussed later).

Before getting into the details of the GPS signal, observation models, and position computations, we first provide more information on the Space Segment and the Control Segment.

2.2.1 Orbit Design

The satellite constellation is designed to have at least 4 satellites in view anywhere, anytime, to a user on the ground. For this purpose, there are nominally 24 GPS satellites distributed in 6 orbital planes. So that we may discuss the orbit design and the implications of that design, we must digress for a short while to explain the geometry of the GPS constellation.

According to Kepler's laws of orbital motion, each orbit takes the approximate shape of an ellipse, with the Earth's centre of mass at the focus of the ellipse. For a GPS orbit, the eccentricity of the ellipse is so small (0.02) that it is almost circular. The semi-major axis (largest radius) of the ellipse is approximately 26,600 km, or approximately 4 Earth radii.

The 6 orbital planes rise over the equator at an inclination angle of 55° to the equator. The point at which they rise from the Southern to Northern Hemisphere across the equator is called the "Right Ascension of the ascending node". Since the orbital planes are evenly distributed, the angle between the six ascending nodes is 60°.

Each orbital plane nominally contains 4 satellites, which are generally not spaced evenly around the ellipse. Therefore, the angle of the satellite within its own orbital plane, the "true anomaly", is only approximately spaced by 90°. The true anomaly is measured from the point of closest approach to the Earth (the perigee). (We note here that there are other types of "anomaly" in GPS terminology, which are angles that are useful for calculating the satellite coordinates within its orbital plane). Note that instead of specifying the satellite's anomaly at every relevant time, we could equivalently specify the time that the satellite had passed perigee, and then compute the satellites future position based on the known laws of motion of the satellite around an ellipse.

Finally, the argument of perigee is the angle between the equator and perigee. Since the orbit is nearly circular, this orbital parameter is not well defined, and alternative parameterisation schemes are often used.

Taken together (the eccentricity, semi-major axis, inclination, Right Ascension of the ascending node, the time of perigee passing, and the argument of perigee), these six parameters define the satellite orbit. These parameters are known as Keplerian elements. Given the Keplerian elements and the current time, it is possible to calculate the coordinates of the satellite.

GPS satellites do not move in perfect ellipses, so additional parameters are necessary. Nevertheless, GPS does use Kepler's laws to its advantage, and the orbits are described in parameters which are Keplerian in appearance. Additional parameters must be added to account for non-Keplerian behaviour. Even this set of parameters has to be updated by the Control Segment every hour for them to remain sufficiently valid.

2.2.2 Orbit design consequences

Several consequences of the orbit design can be deduced from the above orbital parameters, and Kepler's laws of motion. First of all, the satellite speed can be easily calculated to be approximately 4 km/s relative to Earth's centre. All the GPS satellites orbits are prograde, which means the satellites move in the direction of Earth's rotation. Therefore, the relative motion between the satellite and a user on the ground must be less than 4 km/s. Typical values around 1 km/s can be expected for the relative speed along the line of sight (range rate).

The second consequence is the phenomena of "repeating ground tracks" every day. It is straightforward to calculate the time it takes for the satellite to complete one orbital revolution. The orbital period is approximately T = 11 hr 58 min. Therefore a GPS satellite completes 2 revolutions in 23 hr 56 min. This is intentional, since it equals the sidereal day, which is the time it takes for the Earth to rotate 360°. (Note that the solar day of 24 hr is not 360°, because during the day, the position of the Sun in the sky has changed by 1/365.25 of a day, or 4 min, due to the Earth's orbit around the Sun).

Therefore, every day (minus 4 minutes), the satellite appears over the same geographical location on the Earth's surface. The "ground track" is the locus of points on the Earth's

surface that is traced out by a line connecting the satellite to the centre of the Earth. The ground track is said to repeat. From the user's point of view, the same satellite appears in the same direction in the sky every day minus 4 minutes. Likewise, the "sky tracks" repeat. In general, we can say that the entire satellite geometry repeats every sidereal day (from the point of view of a ground user).

As a corollary, any errors correlated with satellite geometry will repeat from one day to the next. An example of an error tied to satellite geometry is "multipath," which is due to the antenna also sensing signals from the satellite which reflect and refract from nearby objects. In fact, it can be verified that, because of multipath, observation residuals do have a pattern that repeats every sidereal day. As a consequence, such errors will not significantly affect the precision, or repeatability, of coordinates estimated each day. However, the accuracy can be significantly worse than the apparent precision for this reason.

Another consequence of this is that the same subset of the 24 satellites will be observed every day by someone at a fixed geographical location. Generally, not all 24 satellites will be seen by a user at a fixed location. This is one reason why there needs to be a global distribution of receivers around the globe to be sure that every satellite is tracked sufficiently well.

We now turn our attention to the consequences of the inclination angle of 55°. Note that a satellite with an inclination angle of 90° would orbit directly over the poles. Any other inclination angle would result in the satellite never passing over the poles. From the user's point of view, the satellite's sky track would never cross over the position of the celestial pole in the sky. In fact, there would be a "hole" in the sky around the celestial pole where the satellite could never pass. For a satellite constellation with an inclination angle of 55°, there would therefore be a circle of radius at least 35° around the celestial pole, through which the sky tracks would never cross. Another way of looking at this, is that a satellite can never rise more than 55° elevation above the celestial equator.

This has a big effect on the satellite geometry as viewed from different latitudes. An observer at the pole would never see a GPS satellite rise above 55° elevation. Most of the satellites would hover close to the horizon. Therefore vertical positioning is slightly degraded near the poles. An observer at the equator would see some of the satellites passing overhead, but would tend to deviate from away from points on the horizon directly to the north and south. Due to a combination of Earth rotation, and the fact that the GPS satellites are moving faster than the Earth rotates, the satellites actually appear to move approximately north-south or south-north to an oberver at the equator, with very little east-west motion. The north component of relative positions are therefore better determined than the east component the closer the observer is to the equator. An observer at mid-latitudes in the Northern Hemisphere would see satellites anywhere in the sky to the south, but there would be a large void towards the north. This has consequences for site selection, where a good view is desirable to the south, and the view to the north is less critical. For example, one might want to select a site in the Northern Hemisphere which is on a south-facing slope (and visa versa for an observer in the Southern Hemisphere).

2.2.3 Satellite Hardware

There are nominally 24 GPS satellites, but this number can vary within a few satellites at any given time, due to old satellites being decommissioned, and new satellites being launched to replace them. All the prototype satellites, known as Block I, have been decommissioned. Between 1989 and 1994, 24 Block II (1989-1994) were placed in orbit. From 1995 onwards, these have started to be replaced by a new design known as Block IIR. The nominal specifications of the GPS satellites are as follows:

- Life goal: 7.5 years
- Mass: ~1 tonne (Block IIR: ~2 tonnes)
- Size: 5 metres
- Power: solar panels 7.5 m^2 + Ni-Cd batteries
- Atomic clocks: 2 rubidium and 2 cesium

The orientation of the satellites is always changing, such that the solar panels face the sun, and the antennas face the centre of the Earth. Signals are transmitted and received by the satellite using microwaves. Signals are transmitted to the User Segment at frequencies L1 = 1575.42 MHz, and L2 = 1227.60 MHz. We discuss the signals in further detail later on. Signals are received from the Control Segment at frequency 1783.74 Mhz. The flow of information is a follows: the satellites transmit L1 and L2 signals to the user, which are encoded with information on their clock times and their positions. The Control Segment then tracks these signals using receivers at special monitoring stations. This information is used to improve the satellite positions and predict where the satellites will be in the near future. This orbit information is then uplinked at 1783.74 Mhz to the GPS satellites, which in turn transmit this new information down to the users, and so on. The orbit information on board the satellite is updated every hour.

2.2.4 The Control Segment

The Control Segment, run by the US Air Force, is responsible for operating GPS. The main Control Centre is at Falcon Air Force Base, Colorado Springs, USA. Several ground stations monitor the satellites L1 and L2 signals, and assess the "health" of the satellites. As outlined previously, the Control Segment then uses these signals to estimate and predict the satellite orbits and clock errors, and this information is uploaded to the satellites. In addition, the Control Segment can control the satellites; for example, the satellites can be maneuvered into a different orbit when necessary. This might be done to optimise satellite geometry when a new satellite is launched, or when an old satellite fails. It is also done to keep the satellites to within a certain tolerance of their nominal orbital parameters (e.g., the semi-major axis may need adjustment from time to time). As another example, the Control Segment might switch between the several on-board clocks available, should the current clock appear to be malfunctioning.

2.3 THE GPS SIGNALS

We now briefly summarise the characteristics of the GPS signals, the types of information that is digitally encoded on the signals, and how the U.S. Department of Defense implements denial of accuracy to civilian users. Further details on how the codes are constructed will be presented in Section 3.

2.3.1 Signal Description

The signals from a GPS satellite are fundamentally driven by an atomic clocks (usually cesium, which has the best long-term stability). The fundamental frequency is 10.23 Mhz. Two carrier signals, which can be thought of as sine waves, are created from this signal by multiplying the frequency by 154 for the L1 channel (frequency = 1575.42 Mhz; wavelength = 19.0 cm), and 120 for the L2 channel (frequency = 1227.60 Mhz; wavelength = 24.4 cm). The reason for the second signal is for self-calibration of the delay of the signal in the Earth's ionosphere.

Information is encoded in the form of binary bits on the carrier signals by a process known as phase modulation. (This is to be compared with signals from radio stations, which are typically encoded using either frequency modulation, FM, or amplitude modulation, AM). The binary digits 0 and 1 are actually represented by multiplying the electrical signals by either +1 or −1, which is equivalent to leaving the signal unchanged, or flipping the phase of the signal by 180 °. We come back later to the meaning of phase and the generation of the binary code.

There are three types of code on the carrier signals:

- The C/A code
- The P code
- The Navigation Message

The C/A ("course acquisition") code can be found on the L1 channel. As will be described later, this is a code sequence which repeats every 1 ms. It is a pseudo-random code, which appears to be random, but is in fact generated by a known algorithm. The carrier can transmit the C/A code at 1.023 Mbps (million bits per second). The "chip length", or physical distance between binary transitions (between digits +1 and −1), is 293 metres. The basic information that the C/A code contains is the time according to the satellite clock when the signal was transmitted (with an ambiguity of 1 ms, which is easily resolved, since this corresponds to 293 km). Each satellite has a different C/A code, so that they can be uniquely identified.

The P ("precise") code is identical on both the L1 and L2 channel. Whereas C/A is a courser code appropriate for initially locking onto the signal, the P code is better for more precise positioning. The P code repeats every 267 days. In practice, this code is divided into 7 day segments; each weekly segment is designated a "PRN" number, and is designated to one of the GPS satellites. The carrier can transmit the P code at 10.23 Mbps, with a chip length of 29.3 metres. Again, the basic information is the satellite clock time or transmission, which is identical to the C/A information, except that it has ten times the resolution. Unlike the C/A code, the P code can be encrypted by a process known as "anti-spoofing" , or "A/S" (see below).

The Navigation Message can be found on the L1 channel, being transmitted at a very slow rate of 50 bps. It is a 1500 bit sequence, and therefore takes 30 seconds to transmit. The Navigation Message includes information on the Broadcast Ephemeris (satellite orbital parameters), satellite clock corrections, almanac data (a crude ephemeris for all satellites), ionosphere information, and satellite health status.

2.3.2 Denial of Accuracy

The U.S. Department of Defense implements two types of denial of accuracy to civilian users: Selective Availability (S/A), and Anti-Spoofing (A/S). S/A can be thought of as intentional errors imposed on the GPS signal. A/S can be thought of as encryption of the P code.

There are two types of S/A: epsilon, and dither. Under conditions of S/A, the user should be able to count on the position error not being any worse than 100 metres. Most of the time, the induced position errors do not exceed 50 metres.

Epsilon is implemented by including errors in the satellite orbit encoded in the Navigation Message. Apparently, this is an option not used, according to daily comparisons made between the real-time broadcast orbits, and precise orbits generated after the fact, by the International GPS Service for Geodynamics (IGS). For precise geodetic work, precise orbits are recommended in any case, and therefore epsilon would have minimal impact on precise users. It would, however, directly impact single receiver, low-precision users. Even then, the effects can be mitigated to some extent by using technology known as "differential GPS", where errors in the GPS signal are computed at a reference station at known coordinates, and are transmitted to the user who has appropriate radio receiving equipment.

Dither is intentional rapid variation in the satellite clock frequency (10.23 MHz). Dither, therefore, looks exactly like a satellite clock error, and therefore maps directly into pseudorange errors. Dither is switched on at the moment (1997), but recent U.S. policy statements indicate that it may be phased out within the next decade. As is the case for epsilon, dither can be mitigated using differential GPS. The high precision user is minimally effected by S/A, since relative positioning techniques effectively eliminate satellite clock error (as we shall see later).

Anti-Spoofing (A/S) is encryption of the P-code. The main purpose of A/S is prevent "the enemy" from imitating a GPS signal, and therefore it is unlikely to be switched off in the foreseeable future. A/S does not pose a signficant problem to the precise user, since precise GPS techniques rely on measuring the phase of the carrier signal itself, rather than the pseudoranges derived from the P code. However, the pseudoranges are very useful for various algorithms, particularly in the rapid position fixes required by moving vehicles and kinematic surveys. Modern geodetic receivers can, in any case, form 2 precise pseudorange observables on the L1 and L2 channels, even if A/S is switched on. (We briefly touch on how this is done in the next section). As a consequence of not having full access to the P code, the phase noise on measuring the L2 carrier phase can be increased from the level of 1 mm to the level of 1 cm for some types of receivers. This has negligible impact on long sessions for static positioning, but can have noticeable effect on short sessions, or on kinematic positioning. Larger degradation in the signal can be expected at low elevations (up to 2 cm) where signal strength is at a minimum.

3. THE PSEUDORANGE OBSERVABLE

In this section, we go deeper into the description of the pseudorange observable, an give some details on how the codes are generated. We develop a model of the pseudorange observation, and then use this model to derive a least-squares estimator for positioning. We discuss formal errors in position, and the notion of "Dilution of Precision", which can be used to assess the effect of satellite geometry on positioning precision.

3.1 CODE GENERATION

It helps to understand the pseudorange measurement if we first take a look at the actual generation of the codes. The carrier signal is multiplied by a series of either +1 or -1, which are seperated by the chip length (293 m for C/A code, and 29.3 m for P code). This series of +1 and -1 multipliers can be interpreted as a stream of binary digits (0 and 1).

How is this stream of binary digits decided? They are determined by an algorithm, known as a linear feedback register. To understand a linear feedback register, we must first introduce the XOR binary function.

3.1.1 XOR: The "Exclusive OR" Binary Function

A binary function takes two input binary digits, and outputs one binary digit (0 or 1). More familiar binary functions might be the "AND" and "OR" functions. For example, the AND function gives a value of 1 if the two input digits are identical, that is (0,0), or (1,1). If the input digits are different, the AND function gives a value of 0. The OR function gives a value of 1 if either of the two input digits equals 1, that is (0,1), (1,0), or (1,1).

The XOR function gives a value of 1 if the two inputs are different, that is (1,0) or (0,1). If the two inputs are the same, (0,0) or (0,1), then the value is 0.

What is XOR(A,B)? Remember this: *Is A different to B? If so, the answer is 1.*

- If $A \neq B$, then XOR(A,B) = 1
- If A = B, then XOR(A,B) = 0

The XOR function can be represented by the "truth table" shown in Table 1.

Input A	Input B	Output XOR(A,B)
0	0	**0**
0	1	**1**
1	0	**1**
1	1	**0**

Table 1. Truth table for the XOR function.

3.1.2 Linear Feedback Registers

Linear feedback registers are used to generate a pseudorandom number sequence. The sequence is pseudorandom, since the sequence repeats after a certain number of digits (which, as we shall see, depends on the size of the register). However, the statistical properties of the sequence are very good, in that the sequence appears to be white noise. We return to these properties later, since they are important for understanding the measurement process. For now, we look at how the register works.

Cycle, N	$A_N = XOR(A_{N-1}, C_{N-1})$	$B_N = A_{N-1}$	$C_N = B_{N-1}$
1	initialise: 1	1	**1**
2	XOR(1,1) = 0	1	**1**
3	XOR(0,1) = 1	0	**1**
4	XOR(1,1) = 0	1	**0**
5	XOR(0,0) = 0	0	**1**
6	XOR(0,1) = 1	0	**0**
7	XOR(1,0) = 1	1	**0**
8 (=1)	XOR(1,0) = 1 (pattern repeats)	1	**1**

Table 2. A 3 stage linear feedback register. The output is in column C.

Table 2 illustrates a simple example: the "3 stage linear feedback register." The "state" of the register is defined by three binary numbers (A, B, C). The state changes after a specific time interval. To start the whole process, the intial state of a feedback register is always filled with 1; that is, for the 3 stage register, the initial state is (1, 1, 1). The digits in this state are now shifted to the right, giving (blank, 1, 1). The digit (1) that is pushed off the right side is the output from the register. The blank is replaced by taking the XOR of the other two digits (1,1). The value, in this case, equals 0. The new state is therefore (0, 1, 1). This process is then repeated, so that the new output is (1), and the next state is (1, 0, 1). The next output is (1) and the next state is (1, 1, 0). The next output is (0), and the next state is (0, 1, 1), and so on.

In the above example, the outputs can be written (1, 1, 1, 0, ...). This stream of digits is known as the "linear feedback register sequence." This sequence will start to repeat after a while. It turns out that during a complete cycle, the feedback register will produce every possible combination of binary numbers, except for (0, 0, 0). We can therefore easily calculate the length of the sequence before it starts to repeat again. For a 3 stage register, there are 8 possible combinations of binary digits. This means that the sequence will repeat after 7 cycles. The sequence length is therefore 7 bits. More generally, the sequence length is:

$$L(N) = 2^N - 1$$

where N is the size of the register (number of digits in the state). For example, a 4 state linear feedback register will have a sequence length of 15 bits.

3.1.3 C/A Code

The C/A code is based on the 10 stage linear feedback register sequence, for which the sequence length is $L(10) = 2^{10}-1 = 1023$ bits. The C/A code really has a repeating sequence of 1023 bits; however the design is slightly more complicated than presented above. The C/A code is actually a "Gold code", which is derived by taking the XOR of the output from 2 linear feedback registers. Unique C/A codes can be generated for each satellite by selecting different pairs of cells from each register to define the output.

In summary, the C/A code is a unique Gold code on each satellite, which is a pseudorandom sequence of bits with a repeating sequence length of 1023. C/A bit transitions occur at 1.023 Mhz. Note that the fundamental frequency in the satellite is 10.23 Mhz, so this represents one transition every 10 cycles. At this rate of bit transitions, the full sequence of 1023 bits is transmitted in 1 ms. Therefore, the sequence repeats 1000 times per second. The chip length (distance between bit transitions) is 293 m. Therefore, the sequence repeats every 300 km.

3.1.4 P Code

The P code is also generated from a combination of two different registers, in such a way that it repeats every 266.4 days. Each 7 day section is assigned a "PRN code." Satellites are often identified by their PRN number; however, the user should beware that any given satellite can have its PRN code changed. Therefore, PRN codes should not be used in place of Satellite Vehicle Numbers (SVN) when talking about particular satellites. (For example, if someone writes software which identifies satellites using PRN numbers, there might be a problem in orbit modelling, for example, PRN 2 is assigned to a Block II satellite now, but to a Block IIR satellite next year). There are 38 possible PRN codes; given that there are 24 nominal satellites, some PRN codes are left unused. The PRN sequence is reset at Saturday midnight, defining the start of "GPS week."

3.1.5 GPS signal transmission and reception

Let us now summarise how the GPS signal is transmitted from space, and then received on the ground. The GPS signal starts in the satellite as a voltage which oscillates at the fundamental clock frequency of 10.23 Mhz. (If selective availability is on, this signal is then "dithered" so that the frequency varies unpredictably). This signal is then separately multiplied in frequency by the integers 154 and 120, to create the L1 and L2 carrier signals. The signals are then multiplied by +1 and −1 according the algorithms described above to generate the C/A code (on L1) and the P code (on both L1 and L2). These codes are unique to each satellite. Finally, the Navigation Message is encoded onto the signal. The signals are boosted by an amplifier, and then sent to transmitting antennas, which point towards the Earth. These antennas are little more than exposed electrical conductors which radiate the signal into space in the form of electromagnetic waves.

These electromagnetic waves pass through space and the Earth's atmosphere, at very close to the speed of light in a vacuum, until they reach the receiver's antenna. The waves create a minute signal in the antenna, in the form of an oscillating voltage. The signal is now pre-

amplified at the antenna, to boost the signal strength, so that it is not overcome by noise by the time it gets to the other end of the antenna cable. The signal then enters the receiver, which then measures it using a process known as "autocorrelation." It is beyond the scope of this paper to go into the details of receiver design, so our description will be kept at the level required to understand how the observable model can be developed.

3.2 AUTOCORRELATION TECHNIQUE

We have described how the GPS satellites construct the GPS signals. Actually, the receiver also generate GPS-like signals internally. The receiver knows precisely what the transmitted GPS signal is supposed to look like at any given time, and it generates an electronic replica, in synchronisation with the receiver's own clock. The receiver then compares the replica signal with the actual signal. Since the GPS signal was actually created in the satellite some time previously (about 0.07 seconds ago, due to the speed of light), the receiver's replica signal must be delayed in to match up the incoming signal with the replica signal. This time delay is actually what the receiver is fundamentally measuring. Clearly, this represents the time taken for the signal to pass from the satellite to the receiver, but it also includes any error in the satellite clock, and any error in the receiver clock. One can see that the time delay is therefore related to the range to the satellite. We return to this model later, and now turn our attention to how the receiver matches the two signals.

The time difference is computed by autocorrelation. The first bit from signal one is multiplied by the first bit of signal two. For example, if the first bits from the two signals both have values -1, then the result is $(-1) \times (-1) = +1$. Similarly, if both bits have values $+1$, then the result is $+1$. On the other hand, if the two bits disagree, the result is $(+1) \times (-1) = -1$. This process is repeated for the second pair of bits, and so on. The result can be written as a sequence of $+1$ (where the bits agree) and -1 (where the bits disagree). This sequence is then summed, and divided by the total number of bits in each signal. For example, if signal A can be written $(+1, -1, -1, +1, -1)$, and signal B can be written $(+1, +1, -1, -1, +1)$, then multiplication gives $(+1, -1, +1, -1, -1)$; the sum of which gives -1; then dividing by the number of bits (5) gives -0.2. Note that if the two signals matched perfectly, the result would be $+1$. If the two signals were completely random, we should expect a result close to zero.

This is why the GPS signals are designed to look random. When the two signals are not properly matched in time, the result of autocorrelation gives an answer close to zero; if the signals are matched in time, the result is close to $+1$ (but not exactly, since a real signal also has noise, so some bits are incorrect). One can see that the larger the number of bits that are compared, the better the resolution. This is because the random bits will average to zero better, the more bits we compare.

The Gold codes have the property that the autocorrelation is constant until we get to within one chip of the correct answer. Within that window of ± 1 chip, the autocorrelation function looks like an equilateral triangle, with a value of 1 at its peak (assuming no noise). We can therefore use the known triangular shape as a model to help us find the time displacement that maximises the autocorrelation. (More sophisticated receivers account for the fact that multipath distorts the shape of this triangle, and can thus reduce the effect of multipath).

Now that we have found the peak autocorrelation, the inferred time displacement between the two signals is multiplied by the speed of light. This observation is called the pseudorange. The pseudorange measurement is described schematically in Figure 1.

3.3 PSEUDORANGE OBSERVATION EQUATIONS

3.3.1 Simplified Pseudorange Model

Receivers record data at regular, specified intervals (say, every 30 seconds, as instructed by the receiver user). It is the reading of the receiver clock time T, which is used to say exactly when the measurement is sampled. Therefore, the value of T at a measurement epoch is known exactly, and is written to the data file along with the observation. (What is not known, is the true time of measurement). The actual observation to satellite s can be written:

$$P^S = (T - T^S)\, c$$

where T is the known reading of the receiver clock when signal is received, T^S is the reading of the satellite clock when the signal was transmitted, and c is the speed of light (in a vacuum) = 299792458 m/s.

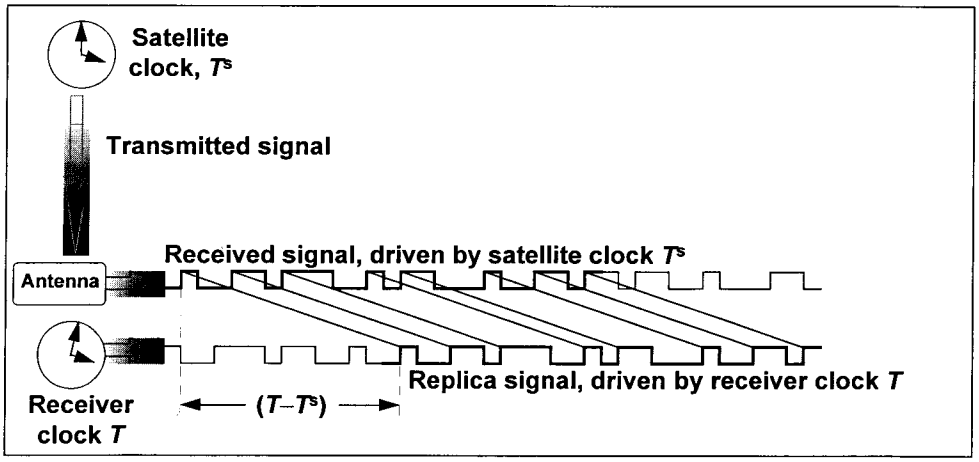


Figure 1: A schematic diagram showing how the GPS pseudorange observation is related to the satellite and receiver clocks.

The modelled observation can be developed by setting the clock time T equal to the true receive time t plus a clock bias τ, for both the receiver and satellite clocks:

$$T = t + \tau$$
$$T^S = t^S + \tau^S$$

Substitution gives the pseudorange as a function of the true time the signal was received:

$$\begin{aligned} P^S(t) &= ((t+\tau) - (t^S + \tau^S))c \\ &= (t - t^S)c + c\tau - c\tau^S \\ &= \rho^S(t, t^S) + c\tau - c\tau^S \end{aligned}$$

where $\rho^S(t, t^S)$ is the range from receiver (at receive time) to the satellite (at transmit time). This model is simplified; for example, it assumes the speed of light in the atmosphere is c, and it ignores the theory of relativity; but this simplified model is useful to gain insight into the principles of GPS. From Pythagoras Theorem, we can write:

$$\rho^S(t,t^S) = \sqrt{\left(x^S(t^S) - x(t)\right)^2 + \left(y^S(t^S) - y(t)\right)^2 + \left(y^S(t^S) - y(t)\right)^2}$$

The Navigation message allows us to compute the satellite position (x^S, y^S, z^S) and the satellite clock bias τ^S. Therefore we are left with 4 unknowns, the receiver position (x, y, z) and the receiver clock bias τ.

We note here one complication: that the satellite position must be calculated at transmission time, t^S. This is important, because the satellite range can change as much as 60 metres from the time the signal was transmitted, to the time the signal was received, approximately 0.07 seconds later. If the receive time were used instead, the error in computed range could be tens of metres. Starting with the receive time, t, the transmit time can be computed by an iterative algorithm known as "the light time equation," which can be written as follows:

$$\begin{aligned} t^S(0) &= t = (T - \tau) \\ t^S(1) &= t - \frac{\rho^S(t, t^S(0))}{c} \\ t^S(2) &= t - \frac{\rho^S(t, t^S(1))}{c} \\ &\text{M} \end{aligned}$$

where the satellite position (and hence the range $\rho^S(t, t^S)$) is calculated at each step using the Keplerian-type elements from the Navigation Message, and the algorithm is stopped once the computed range converges (i.e., don't change by more than a negligible amount). Although more rapidly converging methods have been implemented, the above method is probably the easiest to understand.

Note that the above algorithm starts with the true receive time, which requires the receiver clock bias. We usually don't know in advance what the bias is, but for most receivers it never gets larger than a few milliseconds (beyond which, the receiver will reset its clock). If we assume it is zero in the above computation, the error produced is a few metres, which is much smaller than typical point positioning precision of approximately 50 metres with S/A switched on. We can therefore safely ignore this effect for now, and return to it later when we discuss the more precise carrier phase observable.

We now look at our system of simplified observation equations from 4 satellites in view of the receiver. Using the above notation, we can write the pseudoranges to each satellite as:

$$P^1 = ((x^1 - x)^2 + (y^1 - y)^2 + (z^1 - z)^2)^{1/2} + c\tau - c\tau^1$$

$$P^2 = ((x^2 - x)^2 + (y^2 - y)^2 + (z^2 - z)^2)^{1/2} + c\tau - c\tau^2$$

$$P^3 = ((x^3 - x)^2 + (y^3 - y)^2 + (z^3 - z)^2)^{1/2} + c\tau - c\tau^3$$

$$P^4 = ((x^4 - x)^2 + (y^4 - y)^2 + (z^4 - z)^2)^{1/2} + c\tau - c\tau^4$$

(Note that in this and subsequent equations, the superscripts next to the satellite coordinates are meant to identify the satellite, and should not be confused with exponents). In the following section, we proceed to solve this system of equations for the 4 unknowns, (x, y, z, τ) using familiar least squares methods. Although this is not strictly necessary for 4 unknowns with 4 parameters, it does generalise the solution to the case where we have $m \geq 4$ satellites in view.

4. POINT POSITIONING USING PSEUDORANGE

4.1 LEAST SQUARES ESTIMATION

4.1.1 Linearised Model

We solve the point positioning problem by first linearising the pseudorange observation equations, and then using the familiar methods of least-squares analysis. For completeness, we summarise the linearisation procedure and the development of the least squares method specifically for the GPS point positioning problem. First, we assume we can write the actual observation to be the sum of a modelled observation, plus an error term:

$$\begin{aligned} P_{\text{observed}} &= P_{\text{model}} + \text{noise} \\ &= P(x,y,z,\tau) + v \end{aligned}$$

Next, we apply Taylor's theorem, where we expand about the model computed using provisional parameter values (x_0, y_0, z_0, τ_0), and ignore second and higher order terms.

$$\begin{aligned} P(x,y,z,\tau) &\cong P(x_0,y_0,z_0,\tau_0) + (x - x_0)\frac{\partial P}{\partial x} + (y - y_0)\frac{\partial P}{\partial y} + (z - z_0)\frac{\partial P}{\partial z} + (\tau - \tau_0)\frac{\partial P}{\partial \tau} \\ &= P_{\text{computed}} + \frac{\partial P}{\partial x}\Delta x + \frac{\partial P}{\partial y}\Delta y + \frac{\partial P}{\partial z}\Delta z + \frac{\partial P}{\partial \tau}\Delta\tau \end{aligned}$$

Note that the partial derivatives in the above expression are also computed using provisional values (x_0, y_0, z_0, τ_0). The residual observation is defined to be the difference between the actual observation and the observation computed using the provisional parameter values:

$$\begin{aligned}\Delta P &\equiv P_{\text{observed}} - P_{\text{computed}} \\ &= \frac{\partial P}{\partial x}\Delta x + \frac{\partial P}{\partial y}\Delta y + \frac{\partial P}{\partial z}\Delta z + \frac{\partial P}{\partial \tau}\Delta\tau + v\end{aligned}$$

This can be written in matrix form:

$$\Delta P = \begin{pmatrix} \frac{\partial P}{\partial x} & \frac{\partial P}{\partial y} & \frac{\partial P}{\partial z} & \frac{\partial P}{\partial \tau} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta\tau \end{pmatrix} + v$$

We get such an equation for each satellite in view. In general, for m satellites, we can write this system of m equations in matrix form:

$$\begin{pmatrix} \Delta P^1 \\ \Delta P^2 \\ \Delta P^3 \\ \text{M} \\ \Delta P^m \end{pmatrix} = \begin{pmatrix} \frac{\partial P^1}{\partial x} & \frac{\partial P^1}{\partial y} & \frac{\partial P^1}{\partial z} & \frac{\partial P^1}{\partial \tau} \\ \frac{\partial P^2}{\partial x} & \frac{\partial P^2}{\partial y} & \frac{\partial P^2}{\partial z} & \frac{\partial P^2}{\partial \tau} \\ \frac{\partial P^3}{\partial x} & \frac{\partial P^3}{\partial y} & \frac{\partial P^3}{\partial z} & \frac{\partial P^3}{\partial \tau} \\ \text{M} & \text{M} & \text{M} & \text{M} \\ \frac{\partial P^m}{\partial x} & \frac{\partial P^m}{\partial y} & \frac{\partial P^m}{\partial z} & \frac{\partial P^m}{\partial \tau} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta\tau \end{pmatrix} + \begin{pmatrix} v^1 \\ v^2 \\ v^3 \\ \text{M} \end{pmatrix}$$

The equation is often written using matrix symbols as:

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

which expresses a linear relationship between the residual observations **b** (i.e., observed minus computed observations) and the unknown correction to the parameters **x**. The column matrix **v** contains all the noise terms, which are also unknown at this point. We call the above matrix equation the "linearised observation equations".

4.1.2 The Design Matrix

The linear coefficients, contained in the "design matrix" **A**, are actually the partial derivatives of each observation with respect to each parameter, computed using the provisional parameter values. Note that **A** has the same number of columns as there are parameters, $n = 4$, and has the same number of rows as there are data, $m \geq 4$. We can derive the coefficients of **A** by partial differentiation of the observation equations, producing the following expression:

$$\mathbf{A} = \begin{pmatrix} \frac{x_0 - x^1}{\rho} & \frac{y_0 - y^1}{\rho} & \frac{z_0 - z^1}{\rho} & c \\ \frac{x_0 - x^2}{\rho} & \frac{y_0 - y^2}{\rho} & \frac{z_0 - z^2}{\rho} & c \\ \frac{x_0 - x^3}{\rho} & \frac{y_0 - y^3}{\rho} & \frac{z_0 - z^3}{\rho} & c \\ \mathrm{M} & \mathrm{M} & \mathrm{M} & \mathrm{M} \\ \frac{x_0 - x^m}{\rho} & \frac{y_0 - y^m}{\rho} & \frac{z_0 - z^m}{\rho} & c \end{pmatrix}$$

Note that **A** is shown to be purely a function of the direction to each of the satellites as observed from the receiver.

4.1.3 The Least Squares Solution

Let us consider a solution for the linearised observation equations, denoted $\bar{\mathbf{x}}$. The "estimated residuals" are defined as the difference between the actual observations and the new, estimated model for the observations. Using the linearised form of the observation equations, we can write the estimated residuals as:

$$\bar{\mathbf{v}} = \mathbf{b} - \mathbf{A}\bar{\mathbf{x}}$$

The "least squares" solution can be found by varying the value of **x** until the following functional is minimised:

$$J(\mathbf{x}) \equiv \sum_{i=1}^{m} v_i^2 = \mathbf{v}^{\mathrm{T}}\mathbf{v} = (\mathbf{b} - \mathbf{A}\mathbf{x})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\mathbf{x}).$$

That is, we are minimising the sum of squares of the estimated residuals. If we vary **x** by a small amount, then $J(\mathbf{x})$ should also vary, except at the desired solution where it is stationary (since the slope of a function is zero at a minimum point). The following illustrates the application of this method to derive the least squares solution:

$$\begin{aligned} \delta J(\bar{\mathbf{x}}) &= 0 \\ \delta\left\{(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})\right\} &= 0 \\ \delta(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) + (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})^{\mathrm{T}}\delta(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) &= 0 \\ (-\mathbf{A}\delta\mathbf{x})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) + (\mathbf{b} - \mathbf{A}\bar{\mathbf{x}})^{\mathrm{T}}(-\mathbf{A}\delta\bar{\mathbf{x}}) &= 0 \\ (-2\mathbf{A}\delta\mathbf{x})^{\mathrm{T}}(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) &= 0 \\ (\delta\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}})(\mathbf{b} - \mathbf{A}\bar{\mathbf{x}}) &= 0 \\ \delta\mathbf{x}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{b} - \mathbf{A}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{x}}) &= 0 \\ \mathbf{A}^{\mathrm{T}}\mathbf{A}\bar{\mathbf{x}} &= \mathbf{A}^{\mathrm{T}}\mathbf{b} \end{aligned}$$

The last line is called the system of "normal equations". The solution to the normal equations is therefore:

$$\bar{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}$$

This assumes that the inverse to $\mathbf{A}^T\mathbf{A}$ exists. For example, $m \geq 4$ is a necessary (but not sufficient) condition. Problems can exist if, for example, a pair of satellites lie in the same line of sight, or if the satellites are all in the same orbital plane. In almost all practical situations, $m \geq 5$ is sufficient. Alternatively, one parameter could be left unestimated (e.g., the height could be fixed to sea-level for a boat).

4.2 ERROR COMPUTATION

4.2.1 The Covariance and Cofactor Matrices

If the observations **b** had no errors, and the model were perfect, then the estimates $\bar{\mathbf{x}}$ given by the above expression would be perfect. Any errors **v** in the original observations **b** will obviously map into errors $\mathbf{v}_x$ in the estimates $\bar{\mathbf{x}}$. It is also clear that this mapping will take exactly the same linear form as the above formula:

$$\mathbf{v}_x = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{v}$$

If we have (a priori) an expected value for the error in the data, σ, then we can compute the expected error in the parameters. We discuss the interpretation of the "covariance matrix" later, but for now, we define it as the (square) matrix of expected values of one error multiplied by another error; that is, $C_{ij} \equiv E(v_i v_j)$. A diagonal element C_{ii} is called a "variance," and is often written as the square of the standard deviation, $C_{ii} \equiv E(v_i^2) = \sigma_i^2$. We can concisely define the covariance matrix by the following matrix equation:

$$\mathbf{C} \equiv E(\mathbf{v}\mathbf{v}^T).$$

Let us for now assume we can characterise the error in the observations by one number, the variance $\sigma^2 = E(v^2)$, which is assumed to apply to all m observations. Let us also assume that all observations are uncorrelated, $E(v_i v_j) = 0$ (for $i \neq j$). We can therefore write the covariance matrix of observations as the diagonal matrix, $\mathbf{C}_\sigma = \sigma^2\mathbf{I}$, where **I** is the $m \times m$ identity matrix:

$$\mathbf{C}_\sigma = \begin{pmatrix} \sigma^2 & 0 & \Lambda & 0 \\ 0 & \sigma^2 & & \mathrm{M} \\ \mathrm{M} & & \mathrm{O} & 0 \\ 0 & \Lambda & 0 & \sigma^2 \end{pmatrix}_{m \times m}$$

Under these assumptions, the expected covariance in the parameters for the least squares solution takes on a simple form:

$$
\begin{aligned}
\mathbf{C}_x &= \mathrm{E}\left(\mathbf{v}_x \mathbf{v}_x^{\mathrm{T}}\right) \\
&= \mathrm{E}\left(\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{v}\right)\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{v}\right)^{\mathrm{T}}\right) \\
&= \mathrm{E}\left(\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{v}\mathbf{v}^{\mathrm{T}}\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\right) \\
&= \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\,\mathrm{E}\left(\mathbf{v}\mathbf{v}^{\mathrm{T}}\right)\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \\
&= \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}} C_\sigma \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \\
&= \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\left(\sigma^2\mathbf{I}\right)\mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \\
&= \sigma^2\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \\
&= \sigma^2 \times \langle \text{cofactor matrix} \rangle
\end{aligned}
$$

Note that the "cofactor matrix" $(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}$ also appears in the formula for the least squares estimate, $\hat{\mathbf{x}}$. The "cofactor matrix" is also sometimes called the "covariance matrix," where it is implicitly understood that it should be scaled by the variance of the input observation errors. Since GPS observation errors are a strong function of the particular situation (e.g., due to environmental factors), it is common to focus on the cofactor matrix, which, like **A**, is purely a function of the satellite-receiver geometry at the times of the observations. The cofactor matrix can therefore be used to assess the relative strength of the observing geometry, and to quantify how the level of errors in the measurements can be related to the expected level of errors in the position estimates.

It should therefore be clear why **A** is called the "design matrix"; we can in fact compute the cofactor matrix in advance of a surveying session if we know where the satellites will be (which we do, from the almanac in the Navigation Message). We can therefore "design" our survey (in this specific case, select the time of day) to ensure that the position precision will not be limited by poor satellite geometry.

4.2.2 Interpreting the Covariance Matrix

The covariance matrix for the estimated parameters can be written in terms of its components:

$$
\begin{aligned}
\mathbf{C}_x &= \sigma^2\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1} \\
&= \sigma^2 \begin{pmatrix} \sigma_x^2 & \sigma_{xy} & \sigma_{xz} & \sigma_{x\tau} \\ \sigma_{yx} & \sigma_y^2 & \sigma_{yz} & \sigma_{y\tau} \\ \sigma_{zx} & \sigma_{zy} & \sigma_z^2 & \sigma_{z\tau} \\ \sigma_{\tau x} & \sigma_{\tau y} & \sigma_{\tau z} & \sigma_\tau^2 \end{pmatrix}
\end{aligned}
$$

As an example of how to interpret these components, if the observation errors were at the level of $\sigma = 1$ metre, the error in y coordinate would be at the level of σ_y metres; if the observation errors were $\sigma = 2$ metres, the error in y would be $2\sigma_y$ metres, and so on.

The off-diagonal elements indicate the degree of correlation between parameters. If σ_{yz} were negative, this means that a positive error in y will probably be accompanied by a negative error in z, and visa versa. A useful measure of correlation is the "correlation coefficient," which is defined as

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$$

The correlation coefficient is only a function of the cofactor matrix, and is independent of the observation variance, σ^2. Its value can range between -1 to $+1$, where 0 indicates no correlation, and $+1$ indicates perfect correlation (i.e., the two parameters are effectively identical). Several textbooks show that the "error ellipse" in the plane defined by the (z, y) coordinates (for example) can be computed using the elements σ_z^2, σ_y^2, and ρ_{zy}.

4.2.3 Local Coordinate Errors

Applications tend to focus on horizontal and vertical position. Also, height, h, tends to have largest error than horizontal coordinates. It is therefore more convenient to look at errors in local geodetic coordinates; that is to transform geocentric coordinates *(u,v,w)* to local topocentric coordinates *(n, e, h)*. For this, we have to transform the covariance matrix, using the laws of error propogation. Consider the rotation matrix G which takes us from small relative vector in geocentric system into the local system at latitude φ and longitude λ:

$$\Delta\mathbf{L} = \mathbf{G}\Delta\mathbf{X}$$

$$\begin{pmatrix} \Delta n \\ \Delta e \\ \Delta h \end{pmatrix} = \begin{pmatrix} -\sin\varphi\cos\lambda & -\sin\varphi\sin\lambda & \cos\varphi \\ -\sin\lambda & \cos\lambda & 0 \\ \cos\varphi\cos\lambda & \cos\varphi\sin\lambda & \sin\varphi \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}$$

Obviously, matrix $\mathbf{G}$ would also transform the errors in $\Delta\mathbf{X}$ into errors in $\Delta\mathbf{L}$:

$$\mathbf{v}_L = \mathbf{G}\mathbf{v}_X$$

We now derive how to transform the covariance matrix of coordinates from geocentric system to the local system. This procedure is sometimes referred to as the "law of propogation of errors":

$$\begin{aligned} \mathbf{C}_L &= E\left(\mathbf{v}_L \mathbf{v}_L^T\right) \\ &= E\left((\mathbf{G}\mathbf{v}_X)(\mathbf{G}\mathbf{v}_X)^T\right) \\ &= E\left(\mathbf{G}\mathbf{v}_X\mathbf{v}_X^T\mathbf{G}^T\right) \\ &= \mathbf{G}\,E\left(\mathbf{v}_X\mathbf{v}_X^T\right)\mathbf{G}^T \\ &= \mathbf{G}\mathbf{C}_X\mathbf{G}^T \end{aligned}$$

For future reference, the general form of the resulting equation $\mathbf{C}_L = \mathbf{G}\mathbf{C}_X\mathbf{G}^T$ is applicable to any problem involving an affine transformation (i.e., multiplication of a column vector by any rectangular matrix, **G**). Note that for this particular problem, $\mathbf{C}_X$ is really the 3×3 submatrix taken from the original 4×4 matrix (which also included coefficients for the clock parameter τ). The covariance matrix in the local system $\mathbf{C}_L$ can be written in terms of its components:

$$\mathbf{C}_L = \sigma^2 \begin{pmatrix} \sigma_n^2 & \sigma_{ne} & \sigma_{nh} \\ \sigma_{en} & \sigma_e^2 & \sigma_{eh} \\ \sigma_{hn} & \sigma_{he} & \sigma_h^2 \end{pmatrix}$$

We could then use this covariance, for example, to plot error ellipses in the horizontal plane.

4.2.4 Dilution of Precision

We can now define the various types of "dilution of precision" (DOP) as a function of diagonal elements of the covariance matrix in the local system:

$$VDOP \equiv \sigma_h$$
$$HDOP \equiv \sqrt{\sigma_n^2 + \sigma_e^2}$$
$$PDOP \equiv \sqrt{\sigma_n^2 + \sigma_e^2 + \sigma_h^2}$$
$$TDOP \equiv \sigma_\tau$$
$$GDOP \equiv \sqrt{\sigma_n^2 + \sigma_e^2 + \sigma_h^2 + c^2\sigma_\tau^2}$$

where, for example, *VDOP* stands for "vertical dilution of precision," H stands for horizontal, P for position, T for time, and G for geometric. As an example of how to interpret DOP, a standard deviation of 1 metre in observations would give a standard deviation in horizontal position of *HDOP* metres, and a standard deviation in the receiver clock bias of *TDOP* seconds. If *VDOP* had a value of 5, we could expect pseudorange errors of 1 metre to map into vertical position errors of 5 metres, and so on. As we have seen, the cofactor matrix and therefore the DOP values are purely a function of satellite geometry as observed by the receiver. A "good geometry" therefore gives low DOP values. A "bad geometry" can give very high DOP values. As a general rule, *PDOP* values larger than 5 are considered poor. If there are fewer than a sufficient number of satellites to produce a solution, or if 2 out of 4 satellites lie in approximately the same direction in the sky, then the cofactor matrix becomes singular, and the DOP values go to infinity. The above formulas assume that all 4 parameters (x, y, z, τ) are being estimated. Of course, if fewer than these are estimated, for example if height is not estimated, then the modified DOP values would get smaller, and they would no longer be generally infinity for only 3 satellites in view.

4.2.5 Mission Planning

Mission planning is the term used to describe the pre-analysis of the satellite geometry in advance of a survey. Essentially, it typically involves using commercial software to plot the

DOP values as a function of time at a given geographical location. Since most applications involve local to regional distances, it is not too important which station's location is used for this analysis, since the satellites will appear approximately in the same position in the sky for all stations. One thing that can vary a lot from station to station is the "elevation mask". Most software allow the user to specify which parts of the sky obstruct the view of the satellites (e.g., due to trees, buildings, or mountains). The elevation mask can substantially change the DOP values, so careful attention should be paid to this. Even if the elevation mask went down to the horizon, the user may wish to set it to 15 degrees all around, as research shows that data below 15 degrees is usually plagued by multipathing errors and other problems, such as cycle slips, and a low signal to noise ratio. As mentioned previously, the user might only be interested in horizontal position, where the height is known adequately in advance (e.g., for a boat at sea). Most software allow for DOP values to be computed under the assumption that height is fixed.

5. THE CARRIER PHASE OBSERVABLE

5.1 CONCEPTS

We now introduce the carrier phase observable, which is used for high precision applications. We start with the basic concepts, starting with the meaning of "phase", the principles of interferometry, and the Doppler effect. We then go on to describe the process of observing the carrier phase, and develop an observation model. Fortunately, most of the model can be reduced to what we have learned so far for the pseudorange. Unlike most textbooks, we take the approach of presenting the model in the "range formulism", where the carrier phase is expressed in units of metres, rather than cycles. However, there are some fundamental differences between the carrier phase and the pseudorange observables, as we shall see when we discuss "phase ambiguity" and the infamous problem of "cycle slips".

5.1.1 The Meaning of "Phase," "Frequency" and "Clock Time"

"Phase" is simply "angle of rotation," which is conventionally in units of "cycles" for GPS analysis. Consider a point moving anti-clockwise around the edge of a circle, and draw a line from the centre of the circle to the point. As illustrated in Figure 2, the "phase" $\varphi(t)$ at any given time t can be defined as the angle through which this line has rotated.

Phase is intimately connected with our concept of time, which is always based on some form of periodic motion, such as the rotation of the Earth, the orbit of the Earth around the Sun ("dynamic time"), or the oscillation of a quartz crystal in a wristwatch ("atomic time"). Even our reprentation of time is often based on rotation, such as the angle of the hands on the face of a clock. Angles of rotation give us our measure of "time." In this way, phase can be thought of as a measure of time (after conversion into appropriate units). We can write this formally as:

$$T(t) = k\left(\varphi(t) - \varphi_0\right)$$

where $T(t)$ is the time according to our clock at time t (whatever the clock may be), $\varphi_0 = \varphi(0)$ is so that the clock reads zero when $t = 0$, and k is a calibration constant, converting the units of cycles into units of seconds. Indeed, we can take the above equation as the *definition* of clock time. Whether of not this clock time is useful depends on the constancy of rate of change of phase. This brings us to the concept of frequency.

The "frequency," expressed in units of "cycles per second," is the number of times the line completes a full 360^0 rotation in one second (which of course, is generally a fractional number). This definition is somewhat lacking, since it seems to assume that the rotation is steady over the course of one second. One can better define frequency instantaneously as the first derivative of phase with respect to time; that is, the angular speed.

$$f \equiv \frac{d\varphi(t)}{dt}$$

We chose to treat phase as a fundamental quantity, and frequency as a derived quantity. For example, we can say that frequency is a constant, if we observe the phase as changing linearly in time. Constant frequency is the basis of an ideal clock. If the frequency can be written as a constant, f_0, then we can write the phase of an ideal clock as:

$$\varphi_{\text{ideal}} = f_0 t + \varphi_0$$

therefore

$$T_{\text{ideal}} = k f_0 t$$

Since we want our a clock second to equal a conventional second ($T_{\text{ideal}}=t$), we see that an appropriate choice for the calibration constant is $k = 1/f_0$, where f_0 is the nominal frequency of the oscillator. Going back to our original equation for clock time, we can now define clock time as:

$$T(t) = \frac{\varphi(t) - \varphi_0}{f_0}$$

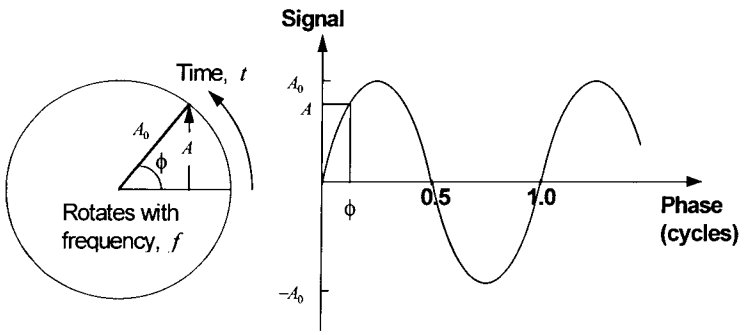


Figure 2: The meaning of phase.

5.1.2 How phase is related to a periodic signal

At time t, the height of point $A(t)$ above the centre of the circle in figure 2 is given by:

$$A(t) = A_0 \sin[2\pi\varphi(t)]$$

where A_0 is the radius of the circle. Since the concept of phase is often applied to periodic signals, we can call $A(t)$ the "signal" and A_0 the "amplitude of the signal". For example, in the case of radio waves, $A(t)$ would be the strength of the electric field, which oscillates in time as the wave passes by. Inverting the above formula, we can therefore determine the phase $\varphi(t)$ if we measure the signal $A(t)$ (and similarly, we could infer the clock time).

Note that, for an ideal clock, the signal would be a pure sinusoidal function of time:

$$\begin{aligned} A_{\text{ideal}} &= A_0 \sin 2\pi\varphi_{\text{ideal}} \\ &= A_0 \sin(2\pi f_0 t + 2\pi\varphi_0) \\ &= (A_0 \cos 2\pi\varphi_0)\sin 2\pi f_0 t + (A_0 \sin 2\pi\varphi_0)\cos 2\pi f_0 t \\ &= A_0^S \sin\omega_0 t + A_0^C \cos\omega_0 t \end{aligned}$$

where the "angular frequency" $\omega_0 \equiv 2\pi f_0$ has units of radians per second. For a real clock, the signal would the same sinusoidal function of its own "clock time," (but would generally be a complicated function of true time):

$$A(T) = A_0^S \sin\omega_0 T + A_0^C \cos\omega_0 T$$

We note that the nominal GPS signal takes on the above form, except that the signal is modulated by "chips", formed by multiplying the amplitudes A_0^S (for C/A code) and A_0^C (for P code) by a pseudorandom sequence of +1 or −1. The underlying sinusoidal signal is called the "carrier signal." It is the phase of the carrier signal that gives us precise access to the satellite clock time; therefore we can use this phase for precise positioning.

5.1.3 Carrier Beat Signal

The GPS carrier signal $G(t)$ from the satellite is "mixed" (multiplied) with the receiver's own replica carrier signal $R(t)$. The result of this mixing is shown in Figure 3.

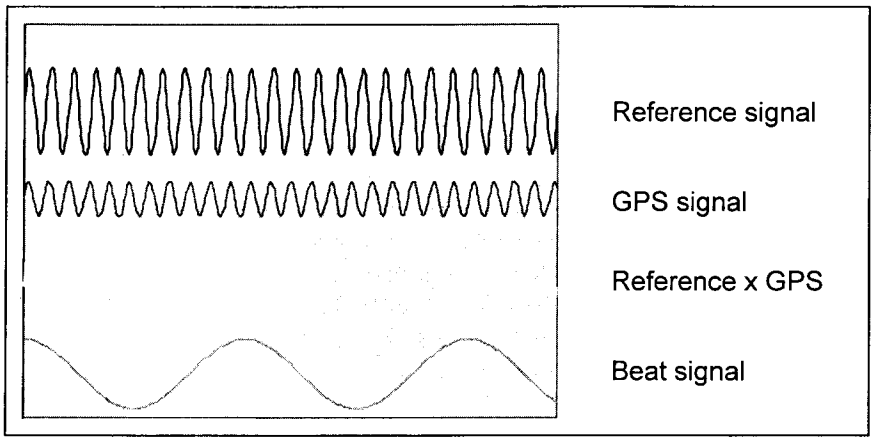


Figure 3: Producing a beat signal by mixing the carrier and replica signals

Mathematically, one can show that one would expect the result to be the difference between a low frequency signal and a high frequency signal:

$$\begin{aligned} R(t) \otimes G(t) &= G_0 \sin 2\pi\varphi_G(t) \times R_0 \sin 2\pi\varphi_R(t) \\ &= \frac{G_0 R_0}{2}\left[\cos 2\pi\left(\varphi_R(t) - \varphi_G(t)\right) - \cos 2\pi\left(\varphi_R(t) + \varphi_G(t)\right)\right] \end{aligned}$$

The high frequency component can be easily filtered out by the receiver electronics, leaving only the carrier beat signal.

$$\begin{aligned} B(t) &= \text{Filter}\left\{R(t) \otimes G(t)\right\} \\ &= \frac{G_0 R_0}{2}\cos 2\pi\left(\varphi_R(t) - \varphi_G(t)\right) \\ &\equiv B_0 \cos 2\pi\left(\varphi_B(t)\right) \end{aligned}$$

where we have introduced the carrier beat phase $\varphi_B(t)$, which is defined to be equal to the difference in phase between the replica signal and the GPS signal.

$$\varphi_B(t) \equiv \varphi_R(t) - \varphi_G(t)$$

By differentiating the above equation with respect to time, we find that the "beat frequency" is equal to the difference in frequencies of the two input signals.

$$f_B \equiv \frac{d\varphi_B}{dt} = f_R - f_G$$

We note that the above formulas apply even when the carrier phase is modulated with codes, provided the replica signal is also modulated (because the values of -1 will cancel when multiplying the two signals). If the codes are not known, it is possible to square both the

incoming signal and the replica signal prior to mixing. The problem with this is that squaring amplifies the noise, thus introducing larger random measurement errors.

5.1.4 Origin of the Phase Ambiguity

Our model of carrier beat phase not a complete picture, since we can arbitrarily add an integer number of cycles to the carrier beat phase, and produce exactly the same observed beat signal. Suppose we only record the fractional phase of the first measurement. We would have no way of telling which integer N to add to this recorded phase so that it really did equal the difference in phase between the replica signal and the GPS signal. This is fundamentally because we have no direct measure of the total phase of the incoming GPS signal. We can express this as follows:

$$\Phi + N = \varphi_R - \varphi_G$$

where we use a capital Greek symbol Φ to emphasise that it represents the phase value actually recorded by the receiver. Provided the receiver does keep track of how many complete signal oscillations there have been since the first measurement, it can attach this number of cycles to the integer portion of the recorded beat phase. However, there will still be an overall ambiguity N that applies to all measurements. That is, we can model N as being the same (unknown) constant for all measurements. If the receiver looses count of the oscillations (e.g., because the signal is obstructed, or because of excessive noise), then a new integer parameter must be introduced to the model, starting at that time. This integer discontinuity in phase data is called a "cycle slip."

5.1.5 Interpretation of the Phase Ambiguity

The reader might also be wondering if there is some kind of geometrical interpretation for N. It turns out that there is, but it does require some oversimplified assumptions. By definition, the unknown value of N can be written as:

$$N = (\text{integer portion of } \varphi_R - \varphi_G) - (\text{integer portion of } \Phi)$$

The second term is completely arbitrary, and depends on the receiver firmware. For example, some receivers set this value to zero for the first measurement. Let us assume this is true, and drop this term. For the sake of interpretation, let us now assume that the receiver and satellite clocks keep perfect time. Under these circumstances, the first term would equal the integer portion of the number of signal oscillations that occur in the receiver from the time the signal was transmitted to the time the signal was received. We can therefore interpret N as equal to the number of carrier wavelengths between the receiver (at the time it makes the first observation), and the satellite (at the time it transmitted the signal). Of course, we made assumptions about perfect clocks and the particular nature of the firmware; so we must beware not to take this interpretation too literally.

5.1.6 Intuitive Model: The Doppler Effect

How can phase be used to measure distance? One way hinted at above is that the phase essentially tells you the clock time. As we shall see in the next section, we can develop phase in almost the same way as the pseudorange model. Another intuitive way of looking at it is to consider the Doppler effect. We are all familiar with the acoustic version of the Doppler effect, as we hear a vehicle's at a higher pitch when it is approaching, and a lower pitch when receding. Can we use the Doppler effect to design a distance measuring device?

Imagine two perfect clocks; one is at a fixed point, the other is approaching in a vehicle. Let both clocks be generating a sinusoidal signal. The frequency difference between the reference signal, and the approaching signal, increases with the vehicle's speed of approach. Let us build a receiver to mix the two signals and measure the beat signal. The beat frequency would be a measure of the speed.

Let us count the cycles of the beat signal; or better yet, let us measure the phase (cycles plus fractional cycles) of the beat signal. Clearly, the beat phase would be measures the change in distance to vehicle. We can therefore (after appropriate unit conversion) write the intuitive equation:

$$\text{Beat phase} = \text{distance to vehicle} + \text{constant}$$

This demonstrates that, although beat phase can be used to precisely measure change in distance from one time to another, there is an unknown constant which prevents us from knowing the full distance. This can be seen by considering moving the reference observer 10 metres away from the original position, and then repeating the experiment. The Doppler effect is clearly exactly the same, and the number of cycles passing by would not change. The very first value of the measured beat phase will indeed be different, but this single measurement cannot be used to infer distance. For example, we have already discussed that don't know what integer number of cycles to attribute to the first beat phase measurement.

5.2 CARRIER PHASE OBSERVATION MODEL

5.2.1 Carrier Beat Phase Model

We now move towards a more rigorous treatment of the carrier beat phase observable, building on our concepts of phase and signal mixing. Our notation will change slightly in preparation for further development.

To summarise what we know already, the satellite carrier signal (from antenna) is mixed with reference signal generated by receiver's clock. The result, after high pass filtering, is a "beating" signal. The phase of this beating signal equals the reference phase minus the incoming GPS carrier phase from a satellite; however, it is ambiguous by an integer number of cycles. From this point on, "carrier beat phase" will be simply called "carrier phase" (but it should not be confused with the phase of the incoming signal!).

Observation of satellite S produces the carrier phase observable Φ^S:

$$\Phi^S(T) = \varphi(T) - \varphi^S(T) - N^S$$

where φ is the replica phase generated by the receiver clock, and φ^S is the incoming signal phase received from GPS satellite S. The measurement is made when the receiver clock time is T.

Now take the point of view that the phase of the incoming signal received at receiver clock time T is identical to the phase that was transmitted from the satellite at satellite clock time T^S.

$$\varphi^S(x,y,z,T) = \varphi^S_{\text{transmit}}(x^S, y^S, z^S, T^S_{\text{transmit}})$$

Of course, if we adopt this point of view, then we shall eventually have to consider the model of how long it takes a wavefront of constant phase to propagate from the satellite to the receiver, so that we may model the appropriate satellite clock time at the time of signal transmission, T^S. We return to that later.

As discussed previously, we can write clock time as a function of phase and nominal frequency:

$$T(t) = \frac{\varphi(t) - \varphi_0}{f_0}$$

We can therefore substitute all the phase terms with clock times:

$$\varphi(T) = f_0 T + \varphi_0$$
$$\varphi^S_{\text{transmit}}(T^S) = f_0 T^S_{\text{transmit}} + \varphi^S_0$$

Therefore, the carrier phase observable becomes:

$$\begin{aligned}\Phi^S(T) &= f_0 T + \varphi_0 - f_0 T^S - \varphi^S_0 - N^S \\ &= f_0\left(T - T^S\right) + \varphi_0 - \varphi^S_0 - N^S\end{aligned}$$

where we implicitly understand that the clock times refer to different events (reception and transmission, respectively).

We note that any term containing the superscript S are different for each satellites, but all other terms are identical. Receivers are designed and calibrated so that the phase constant φ_0 is identical for all satellites; that is, there should be no interchannel biases. Receivers should also sample the carrier phase measurements from all satellites at exactly the same time. (If the receivers have multiplexing electronics to save on cost, then the output should have been interpolated to the same epoch for all satellites). The time T^S will vary slightly from satellite to satellite, since the satellite transmission time must have been different for all signals to arrive at the same time. We also note that the last three terms are constant, and cannot be separated from each other. We can collectively call these terms the "carrier phase bias," which is clearly not an integer.

In preparation for multi-receiver and multi-satellite analysis, we now introduce the subscripts *A, B, C,* etc. to indicate quantities specific to receivers,, and we introduce superscripts *j, k, l,* etc. to identify satellite-specific quantities. We write the carrier phase observed by receiver *A* from satellite *j*:

$$\Phi_A^j(T_A) = f_0\left(T_A - T^j\right) + \varphi_{0A} - \varphi_0{}^j - N_A^j$$

Note that data should be sampled at exactly the same values of clock time (called "epochs") for all receivers, so all values of T_A are identical at a given epoch. However receivers clocks do not all run at exactly the same rate, therefore the true time of measurement will differ slightly from receiver to receiver. Also, note that each receiver-satellite pair has a different carrier phase ambiguity.

5.2.2 Range Formulation

It is convenient to convert the carrier phase model into units of range. This simplifies concepts, models, and software. In the range formulation, we multiply the carrier phase equation by the nominal wavelength.

$$\begin{aligned} L_A^j(T_A) &\equiv \lambda_0 \Phi_A^j(T_A) \\ &= \lambda_0 f_0\left(T_A - T^j\right) + \lambda_0\left(\varphi_{0A} - \varphi_0{}^j - N_A^j\right) \\ &= c\left(T_A - T^j\right) + \lambda_0\left(\varphi_{0A} - \varphi_0{}^j - N_A^j\right) \\ &\equiv c\left(T_A - T^j\right) + B_A^j \end{aligned}$$

where we still retain the name "carrier phase" for $L_A^j(T_A)$, which is in units of metres. We see immediately that this equation is identical to that for the pseudorange, with the exception of the "carrier phase bias," B_A^j which can be written (in units of metres):

$$B_A^j \equiv \lambda_0\left(\varphi_{0A} - \varphi_0{}^j - N_A^j\right)$$

Note that the carrier phase bias for (undifferenced) data is not an integer number of wavelengths, but also includes unknown instrumental phase offsets in the satellite and receiver.

We have not mentioned yet about any differences between carrier phase on the L1 and L2 channel. Although they have different frequencies, in units of range the above equations take on the same form. Actually, the clock bias parameters would be identical for both L1 and L2 phases, but the carrier phase bias would be different. The main difference comes when we develop the model in terms of the propagation delay, which is a function of frequency in the Earth's ionosphere.

5.2.3 Observation Model

We note that the first term in the carrier phase model is simply the pseudorange, and the second term is a constant. We have already developed a simplified model for pseudorange, so we can therefore write a model for carrier phase as follows:

$$\begin{aligned} L_A^j(T_A) &= c\left(T_A - T^j\right) + B_A^j \\ &= \rho_A^j(t_A, t^j) + c\tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j \end{aligned}$$

In the above expression, we have explicitly included the delay on the signal due to the troposphere Z_A^j and the ionosphere $-I_A^j$ (the minus sign indicating that the phase velocity actually increases). Models for the atmospheric delay terms are beyond the scope of this text.

The model for pseudorange can be similarly improved, with the small difference that the ionospheric delay has a positive sign.

$$\begin{aligned} P_A^j(T_A) &= c\left(T_A - T^j\right) \\ &= \rho_A^j(t_A, t^j) + c\tau_A - c\tau^j + Z_A^j + I_A^j \end{aligned}$$

This is because, from physics theory, any information, such as the +1 and −1 "chips" which are modulated onto the carrier wave, must travel with the "group velocity" rather than "phase velocity". According to the theory of relativity, information can not be transmitted faster than c. From the physics of wave propagation in the ionosphere, it can be shown that the group delay is (to a very good first order approximation) precisely the same magnitude, but opposite sign of the phase delay (which is really a phase "advance").

5.2.4 Accounting for Time-Tag Bias

Before proceeding, we return to the problem posed in our discussion of the pseudorange model, that we typically do not know the true time of signal reception t_A which we need to calculate the satellite-receiver range term $\rho_A^j(t_A, t^j)$ precisely. From section 3.3.1, the true time of reception can be written:

$$t_A = T_A - \tau_A$$

where the epoch T_A is known exactly, as it is the receiver clock time written into the data file with the observation (and hence called the "time-tag"). However, the receiver clock bias τ_A is not known initially, but could be as large as milliseconds. The problem is that, due to satellite motion and Earth rotation, the range will change by several metres over the period of a few milliseconds, so we must be careful to account for this for precision work (especially when using the carrier phase observable). For precision work (1 mm), we should use a value τ_A that is accurate to 1 μs.

There are various approaches to dealing with this in GPS geodetic software, which typically use some combination of the following methods:

- use values of the receiver clock bias computed in a first step using a pseudorange point position solution at each epoch;
- iterate the least-squares procedure, processing both carrier phase and pseudorange data simultaneously, and using estimates of the clock bias to compute the true receive time, and therefore the new range model;
- use an estimate $\tilde{P}^j$ of the true transmit time t^j to compute the satellite position.

$$\begin{aligned}\tilde{P}^j &= \tilde{T}^j - \tau^j \\ &= \left(T_A - P_A^j/c\right) + \tau^j\end{aligned}$$

where the satellite clock bias τ^j is obtained from the Navigation Message. One can then directly compute the range term and true receive time with sufficient precision, provided the approximate station coordinates are known to within 300 m (corresponding to the 1 μs timing requirement). Interestingly, this is the basis for "time transfer," since it allows one to compute the receiver clock bias using pseudorange data from only one GPS satellite. (For precise time transfer, two GPS satellites are always in operation with no S/A switched on.) As a method for computing range for precise positioning, this is not often used, perhaps for the reason that it is not a pure model, as it depends on pseudorange data and approximate positions.

- one can take a modelling "short cut" to avoid iteration by expanding the range model as a first order Taylor series. Since this method often appears in the development of the observation equation in textbooks, we discuss it in more detail here.

5.2.5 A Note on the Range-Rate Term

The observation equation can be approximated as follows:

$$\begin{aligned}L_A^j(T_A) &= \rho_A^j(t_A, t^j) + c\tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j \\ &= \rho_A^j(T_A - \tau_A, t^j) + c\tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j \\ &\approx \rho_A^j(T_A, t'^j) - \dot{\rho}_A^j \tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j \\ &= \rho_A^j(T_A, t'^j) + \left(c - \dot{\rho}_A^j\right)\tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j\end{aligned}$$

where we see that the effect can be accounted for by introducing the modelled range rate (i.e., the relative speed of the satellite in the direction of view). The "prime" for the satellite transmit time t'^j (which is used to compute the satellite coordinates) is to indicate that it is not the true transmit time, but the time computed using the nominal receive time T_A. A first order Taylor expansion has been used. The higher order terms will only become significant error sources if the receiver clock bias is greater than about 10 ms, which does not usually happen with modern receivers. In any case, clock biases greater than this amount would result in a worse error in relative position due to the effect of S/A (see section 5.3.1).

Textbooks sometimes include a "range rate" term in the development of the phase observation model, even though, strictly speaking, it is unnecessary. After all, the first line line of the above equation is correct, and the lack of a priori knowledge of the receiver clock bias can easily be dealt with by least-squares iteration, or prior point positioning using the pseudorange. On the other hand, it is nevertheless instructional to show the above set of

equations, since it does illustrate that it is more correct to use $\left(c - \dot{\rho}_A^j\right)$ as the partial derivatives with respect to the receiver clock in the design matrix, rather than simply using c (section 4.1.2). This is crucial if one is not initialising clocks using point position solutions or iteration (as is typical, for example, with the GIPSY OASIS II software). It is not important if initialisation of τ_A is achieved with 1 μs accuracy.

In the expressions to follow, we shall not explicitly include the range rate term on the assumption that time-tag bias has been handled one way or another.

5.3 DIFFERENCING TECHNIQUES

5.3.1 Single Differencing

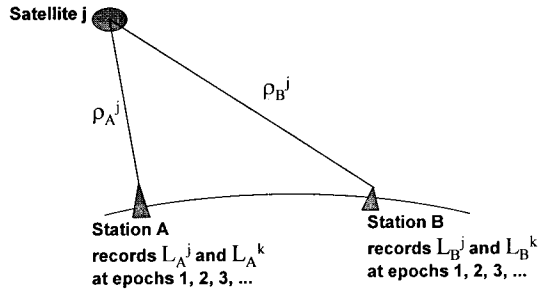


Figure 4: Single differencing geometry

The purpose of "single differencing" is to eliminate satellite clock bias. Consider the observation equations for two receivers, A and B observing same satellite, j:

$$L_A^j = \rho_A^j + c\tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j$$
$$L_B^j = \rho_B^j + c\tau_B - c\tau^j + Z_B^j - I_B^j + B_B^j$$

The single difference phase is defined as the difference between these two:

$$\begin{aligned}\Delta L_{AB}^j &\equiv L_A^j - L_B^j \\ &= \left(\rho_A^j + c\tau_A - c\tau^j + Z_A^j - I_A^j + B_A^j\right) - \left(\rho_B^j + c\tau_B - c\tau^j + Z_B^j - I_B^j + B_B^j\right) \\ &= \left(\rho_A^j - \rho_B^j\right) + \left(c\tau_A - c\tau_B\right) - \left(c\tau^j - c\tau^j\right) + \left(Z_A^j - Z_B^j\right) - \left(I_A^j - I_B^j\right) - \left(B_A^j - B_B^j\right) \\ &= \Delta\rho_{AB}^j + c\Delta\tau_{AB} + \Delta Z_{AB}^j - \Delta I_{AB}^j + \Delta B_{AB}^j\end{aligned}$$

where we use the double-subscript to denote quantities identified with two receivers, and the triangular symbol as a mnemonic device, to emphasise that the difference is made between two points on the ground. The geometry of single differencing is illustrated in Figure 4.

An assumption has been made, that the satellite clock bias τ^j is effectively identical at the slightly different times that the signal was transmitted to *A* and to *B*. The difference in transmission time could be as much as a few milliseconds, either because the imperfect receiver clocks have drifted away from GPS time by that amount, or because the stations might be separated by 1,000 km or more. Since selective availability is typically at the level of 10^{-9} (variation in frequency divided by nominal frequency), over a millisecond (10^{-3}s) the satellite clock error will differ by 10^{-12}s. This translates into a distance error of $10^{-12}c$, or 0.3 mm, a negligible amount under typical S/A conditions (however, it may not be negligible if the level of S/A were increased; but this effect could in principle be corrected if we used reference receivers to monitor S/A). Another point worth mentioning, is that the coordinates of the satellite at transmission time can easily be significantly different for receivers A and B, and this should be remembered when computing the term $\Delta\rho^j_{AB}$.

The atmospheric delay terms are now considerably reduced, and vanish in the limit that the receivers are standing side by side. The differential troposphere can usually be ignored for horizontal separations less than approximately 30 km, however differences in height should be modelled. The differential ionosphere can usually be ignored for separations of 1 to 30 km, depending on ionospheric conditions. Due to ionospheric uncertainty, it is wise to calibrate for the ionosphere using dual-frequency receivers for distances greater than a few km.

Although the single difference has the advantage that many error sources are eliminated or reduced, the disadvantage is that only relative position can be estimated (unless the network is global-scale). Moreover, the receiver clock bias is still unknown, and very unpredictable. This takes us to "double differencing".

5.3.2 Double Differencing

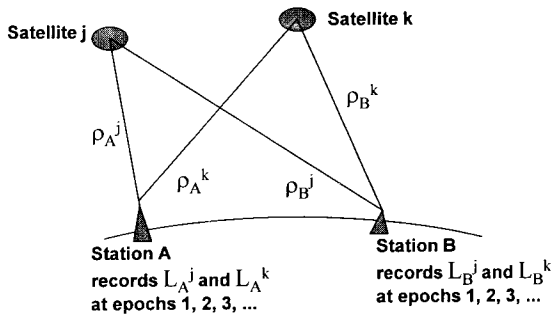


Figure 5: Double differencing geometry.

The purpose of "double differencing" is to eliminate receiver clock bias. Consider the single differenced observation equations for two receivers *A* and *B* observing satellites *j* and *k*:

$$\Delta L^j_{AB} = \Delta\rho^j_{AB} + c\Delta\tau_{AB} + \Delta Z^j_{AB} - \Delta I^j_{AB} + \Delta B^j_{AB}$$
$$\Delta L^k_{AB} = \Delta\rho^k_{AB} + c\Delta\tau_{AB} + \Delta Z^k_{AB} - \Delta I^k_{AB} + \Delta B^k_{AB}$$

The double difference phase is defined as the difference between these two:

$$\begin{aligned}\nabla\Delta L^{jk}_{AB} &\equiv \Delta L^{j}_{AB} - \Delta L^{k}_{AB} \\ &= \left(\Delta\rho^{j}_{AB} + c\Delta\tau_{AB} + \Delta Z^{j}_{AB} - \Delta I^{j}_{AB} + \Delta B^{j}_{AB}\right) - \left(\Delta\rho^{k}_{AB} + c\Delta\tau_{AB} + \Delta Z^{k}_{AB} - \Delta I^{k}_{AB} + \Delta B^{k}_{AB}\right) \\ &= \left(\Delta\rho^{j}_{AB} - \Delta\rho^{k}_{AB}\right) + \left(c\Delta\tau_{AB} - c\Delta\tau_{AB}\right) + \left(\Delta Z^{j}_{AB} - \Delta Z^{k}_{AB}\right) - \left(\Delta I^{j}_{AB} - \Delta I^{k}_{AB}\right) - \left(\Delta B^{j}_{AB} - \Delta B^{k}_{AB}\right) \\ &= \nabla\Delta\rho^{jk}_{AB} + \nabla\Delta Z^{jk}_{AB} - \nabla\Delta I^{jk}_{AB} + \nabla\Delta B^{jk}_{AB}\end{aligned}$$

where we use the double-superscript to denote quantities identified with two satellites, and the upside-down triangular symbol as a mnemonic device, to emphasise that the difference is made between two points in the sky. Figure 5 illustrates the geometry of double differencing.

A point worth mentioning, is that although the receiver clock error has been eliminated to first order, the residual effect due "time tag bias" on the computation of the range term (section 5.2.4) does not completely cancel, and still needs to be dealt with if the receiver separation is large.

Any systematic effects due to unmodelled atmospheric errors are generally increased slightly by approximately 40% by double differencing as compared to single differencing. Similarly, random errors due to measurement noise and multipath are increased. Overall, random errors are effectively doubled as compared with the undifferenced observation equation. On the other hand, the motivation for double differencing is to remove clock bias, which would create much larger errors.

One could process undifferenced or single differenced data, and estimate clock biases. In the limit that clock biases are estimated at every epoch (the "white noise clock model"), these methods become almost identical, provided a proper treatment is made of the data covariance (to be described later). It is almost, but not quite identical, because differencing schemes almost always involve pre-selection of baselines in a network to form single differences, and data can be lost by lack of complete overlap of the observations to each satellite. (This problem can be minimised by selecting the shortest baselines in the network to process, and by assuring that no more than one baseline be drawn to a receiver with a significant loss of data).

5.3.3 Double Differenced Ambiguity

The double difference combination has an additional advantage, in that the ambiguity is an integer:

$$\begin{aligned}\nabla\Delta B_{AB}^{jk} &= \Delta B_{AB}^{j} - \Delta B_{AB}^{k} \\ &= \left(B_A^j - B_B^j\right) - \left(B_A^k - B_B^k\right) \\ &= \lambda_0\left(\varphi_{0A} - \varphi_0{}^j - N_A^j\right) - \lambda_0\left(\varphi_{0B} - \varphi_0{}^j - N_B^j\right) - \lambda_0\left(\varphi_{0A} - \varphi_0{}^k - N_A^k\right) + \lambda_0\left(\varphi_{0B} - \varphi_0{}^k - N_B^k\right) \\ &= -\lambda_0\left(N_A^j - N_B^j - N_A^k + N_B^k\right) \\ &= -\lambda_0 \nabla\Delta N_{AB}^{jk}\end{aligned}$$

Hence we can write the double differenced phase observation equation:

$$\nabla\Delta L_{AB}^{jk} = \nabla\Delta\rho_{AB}^{jk} + \nabla\Delta Z_{AB}^{jk} - \nabla\Delta I_{AB}^{jk} - \lambda_0 \nabla\Delta N_{AB}^{jk}$$

From the point of view of estimation, it makes no difference whether we use a minus or plus sign for N, so long as the partial derivative has a consistent sign (which, for the above equation, would be $-\lambda_0$).

5.3.4 *Triple Differencing*

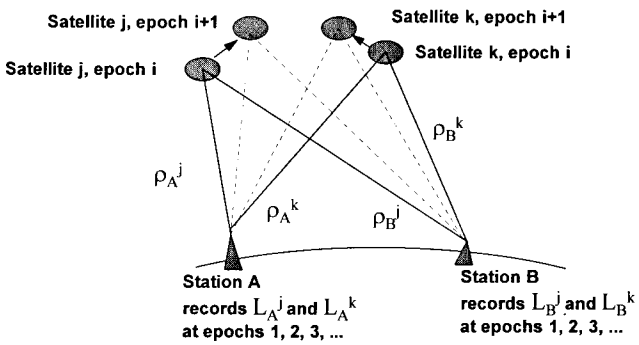


Figure 6: Triple differencing geometry

The purpose of "triple differencing" is to eliminate the integer ambiguity. Consider two successive epochs (i, i+1) of double differenced data from receivers A and B observing satellites j and k:

$$\begin{aligned}\nabla\Delta L_{AB}^{jk}(i) &= \nabla\Delta\rho_{AB}^{jk}(i) + \nabla\Delta Z_{AB}^{jk}(i) - \nabla\Delta I_{AB}^{jk}(i) - \lambda_0 \nabla\Delta N_{AB}^{jk} \\ \nabla\Delta L_{AB}^{jk}(i+1) &= \nabla\Delta\rho_{AB}^{jk}(i+1) + \nabla\Delta Z_{AB}^{jk}(i+1) - \nabla\Delta I_{AB}^{jk}(i+1) - \lambda_0 \nabla\Delta N_{AB}^{jk}\end{aligned}$$

The triple difference phase is defined as the difference between these two:

$$\begin{aligned}\delta(i,i+1)\nabla\Delta L_{AB}^{jk} &\equiv \nabla\Delta L_{AB}^{jk}(i+1) - \nabla\Delta L_{AB}^{jk}(i) \\ &= \delta(i,i+1)\nabla\Delta\rho_{AB}^{jk}(i) + \delta(i,i+1)\nabla\Delta Z_{AB}^{jk}(i) - \delta(i,i+1)\nabla\Delta I_{AB}^{jk}(i)\end{aligned}$$

where we use the delta symbol to indicate the operator that differences data between epochs. Figure 6 illustrates triple differencing geometry.

The triple difference only removes the ambiguity if it has not changed during the time interval between epochs. Any cycle slips will appear as outliers, and can easily be removed by conventional techniques. This is unlike the situation with double differencing, where cycle slips appear as step functions in the time series of data.

The disadvantage of the triple difference is that it introduces correlations between observations in time. Generally, increasing correlations in data has the property of decreasing the data weights. With triple differencing, the degradation in precision is substantial; so triple differenced data are inappropriate for precise surveys. On the other hand, it is a very useful method for obtaining better nominal parameters for double differencing (to ensure linearity), and it is a robust method, due to the ease with which cycle slips can be identified and removed.

It can be shown that triple difference solution is identical to the double differenced solution, provided just one epoch double differenced equation is included for the first point in a data arc, along with the triple differences, and provided the full data covariance matrix is used to compute the weight matrix. This special approach can provide tremendous savings in computation time over straightforward double differencing, while retaining robustness.

6. RELATIVE POSITIONING USING CARRIER PHASE

6.1 SELECTION OF OBSERVATIONS

6.1.1 Linear Dependence of Observations

We can usually form many more possible combinations of double differenced observations than there are original data. This poses a paradox, since we cannot create more information than we started with. The paradox is resolved if we realise that some double differences can be formed by differencing pairs of other double differences. It then becomes obvious that we should not process such observations, otherwise we would be processing the same data more than once. This would clearly be incorrect.

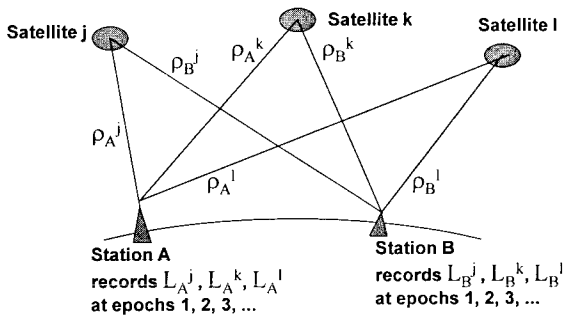


Figure 7: Double difference geometry with 3 satellites.

Figure 7 illustrates the simplest example of the problem. In this example, we have 3 satellites j, k and l, observed by two receivers A and B. If we ignore trivial examples (e.g., $L_{AB}^{jk} = -L_{AB}^{kj}$), there are 3 possible double differences that can be formed:

$$L_{AB}^{jk} = \left(L_A^j - L_B^j\right) - \left(L_A^k - L_B^k\right)$$
$$L_{AB}^{jl} = \left(L_A^j - L_B^j\right) - \left(L_A^l - L_B^l\right)$$
$$L_{AB}^{lk} = \left(L_A^l - L_B^l\right) - \left(L_A^k - L_B^k\right)$$

Note that we can form any one of these observations as a *linear combination* of the others:

$$L_{AB}^{jk} = L_{AB}^{jl} + L_{AB}^{lk}$$
$$L_{AB}^{jl} = L_{AB}^{jk} - L_{AB}^{lk}$$
$$L_{AB}^{lk} = L_{AB}^{jk} - L_{AB}^{jl}$$

The data set $\left\{L_{AB}^{jk}, L_{AB}^{jl}, L_{AB}^{lk}\right\}$ is therefore said to be *linearly dependent*. A linearly independent set must be used for least squares. Examples of appropriate linearly independent sets in this example are:

$$\left\{L_{AB}^{jk}, L_{AB}^{jl}\right\} = \Lambda^j \equiv \left\{L_{AB}^{ab} | a = j; b \neq j\right\}$$
$$\left\{L_{AB}^{kj}, L_{AB}^{kl}\right\} = \Lambda^k \equiv \left\{L_{AB}^{ab} | a = k; b \neq k\right\}$$
$$\left\{L_{AB}^{lj}, L_{AB}^{lk}\right\} = \Lambda^l \equiv \left\{L_{AB}^{ab} | a = l; b \neq l\right\}$$

6.1.2 The Reference Satellite Concept

The "reference satellite concept" involves using either set Λ^j, Λ^k or Λ^l throughout the data set. For example, double differences in set Λ^l all involve the satellite l. Any set is equally as valid, and will produce identical solutions provided the data covariance is properly constructed (see the next section). Obviously, the reference satellite itself has to have data at every epoch, otherwise data will be lost. This can cause problems for less sophisticated software. Typically, a reference satellite should be picked which has the longest period in view. A better algorithm is to select a reference satellite epoch by epoch.

Our simple example can be easily extended to more than 3 satellites. For example consider satellites 1, 2, 3, 4 and 5 in view. We can pick satellite 4 as the reference satellite; therefore our linearly independent set is:

$$\Lambda^4 \equiv \left\{L_{AB}^{ab} | a = 4; b \neq 4\right\}$$
$$= \left\{L_{AB}^{41}, L_{AB}^{42}, L_{AB}^{43}, L_{AB}^{45}\right\}$$

Note that for a single baseline (i.e. 2 receivers), the number of linearly independent double differenced observations is $s-1$, where s is the number of satellites being tracked.

6.1.3 The Reference Station Concept

However, if we have a network of more than 2 receivers, we must account for the fact that double differenced data from the set of all baselines are linearly dependent. We therefore introduce the "reference station" concept, where our set of double differences all include a common reference station. This guarantees linear independence. For example, consider satellites 1, 2, 3 and 4 being tracked by stations A, B, and C. If we pick our reference satellite to be 3, and reference station to be B, then our chosen set is:

$$\begin{aligned}\Lambda^3_B &\equiv \left\{ L^{ab}_{cd} \middle| a = 3; b \neq 3; c = B, d \neq B \right\} \\ &= \left\{ L^{31}_{BA}, L^{32}_{BA}, L^{34}_{BA}, L^{31}_{BC}, L^{32}_{BC}, L^{34}_{BC} \right\}\end{aligned}$$

Note that the number of linearly independent double differenced observations is $(s-1)(r-1)$, where s is the number of satellites being tracked, and r is the number of receivers. So, in our previous example, 3 receivers and 4 satellites gives 6 observations. This assumes that s satellites are observed by all stations. This may not be the case, either due to obstructions, receiver problems, or because the receivers are separated by such a large distance that the satellite is not above the horizon for some receivers.

If using the reference station concept, it is therefore best to choose a receiver close to the middle of a large network, with few obstructions, and no hardware problems, otherwise the set of double differences may not be as complete as it could be. The reference station concept is obviously not optimal, and is seriously problematic for large networks. A better strategy for large networks is to select short baselines that connect together throughout the entire network, being careful not to introduce linear dependent observations, by not including any closed polygons (such as triangles) in the network. In principle, there must be only one possible path between pairs of stations. An even better strategy would be to optimise this choice for every epoch.

6.1.4 Solution Uniqueness

It should be stressed that, if all stations were tracking the same set of satellites at all epochs, then the selection of reference station and reference satellite will not matter, since an identical solution will be produced whatever the selection. This assumes that the data weight matrix is properly constructed (as described below) and that no data outliers are removed.

The problem of linear dependence usually introduces a level of arbitrariness into the solutions due to violation of the above assumptions. This problem is also true even if the previously suggested improvements are made to the reference station concept, since the user typically has to make decisions on which baselines to process (even for more sophisticated software). This is somewhat unsatisfactory, since it is there generally no unique solution, However, experience shows that any reasonable selection will only produce small differences in the final solutions.

There is a way to produce a unique solution, and that is to process undifferenced observations, estimating clock parameters at each epoch. As stated previously, this will produce a solution identical to double differencing under ideal conditions. This class of software is not typically available commercially; however, it is should be stressed that double differencing software does not produce significantly inferior results for most situations. What is far more important is the quality of the observable models, the range of appropriate estimation options, and the ability to detect and deal with cycle slips and outliers.

6.2 BASELINE SOLUTION USING DOUBLE DIFFERENCES

6.2.1 Simplified Observation Equations

We now show how relative coordinates can be estimated between two receivers using the double differenced carrier phase data. We start by simplifying the observation equation, assuming that the relative atmospheric delay is negligible for short distances between receivers. We also drop the symbols "$\nabla\Delta$" of the previous section to simplify the notation. We shall therefore use the following simplified observation equation:

$$L_{AB}^{jk} = \rho_{AB}^{jk} - \lambda_0 N_{AB}^{jk}$$

6.2.2 General Procedure

Processing double differenced data from two receivers results in a "baseline solution." The estimated parameters include the vector between the two receivers, in Cartesian coordinates $(\Delta x, \Delta y, \Delta z)$ and may include parameters to model the tropospheric delay. In addition, the ambiguity parameters N_{AB}^{jk} for each set of double differences to specific satellite pairs (j, k) must be estimated.

The observation equations therefore require linearisation in terms of all these parameters (according to the process explained in section 4.1). Typically, one station is held fixed at good nominal coordinates, which quite often come from an initial pseudorange point position solution. We should mention, however, that due to S/A, point position solutions can have substantial errors (100 m) which may create significant errors in the double differenced observation model, and in the design matrix.

If we call the fixed station A, then estimating the baseline vector is equivalent to estimating the coordinates of station B. It is convenient to formulate the problem to estimate parameters (x_B, y_B, z_B). For example, consider a GPS survey between stations A and B, which observe satellites 1, 2, 3 and 4 for every epoch in the session, where we arbitrarily pick satellite 2 as the reference satellite. For every epoch i, we have the following linearly independent set of 3 double differenced observations:

$$\Lambda^2(i) \equiv \left\{ L_{AB}^{ab}(i) \middle| a = 2; b \neq 2 \right\}$$
$$= \left\{ L_{AB}^{21}(i), L_{AB}^{23}(i), L_{AB}^{24}(i) \right\}$$

We therefore have the parameter set $\left\{ x_B, y_B, z_B, N_{AB}^{21}, N_{AB}^{23}, N_{AB}^{24} \right\}$. If any cycle slips had occured and could not be corrected, then additional ambiguity parameters must be added to the list.

As in Section 3.4.1, the linearised observation equations can be expressed in the form

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{v}$$

where the residual observations are listed in the **b** matrix, which has dimensions $d \times 1$, where d is the number of linearly independent double differenced data. The design matrix **A** has dimensions $d \times p$ where p is the number of parameters, and the parameter corrections are contained in the **x** matrix, which has dimensions $p \times 1$. The observation errors are represented by the **v** matrix, which has the same dimensionality as **b**. We shall discuss the design matrix later on.

It is important to use a "weighted least squares" approach, because of correlations in the double differenced data. We shall not not derive the weighted least squares estimator here, but for completeness, the solution is given here:

$$\bar{\mathbf{x}} = \left(\mathbf{A}^T \mathbf{W} \mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{W} \mathbf{b}$$

where **W** is the data weight matrix, to be derived later on, and **b** is a vector containing the double-differenced residual observations.

The covariance matrix for the estimated parameters is given by:

$$\mathbf{C}_x = \left(\mathbf{A}^T \mathbf{W} \mathbf{A} \right)^{-1}$$

The covariance matrix can be used to judge whether the theoretically expected precision from the observation scenario is sufficient to allow ambiguities to be fixed to integer values. If ambiguity parameters can be fixed in the model, a theoretically more precise solution can be generated from the same data, but without estimating the ambiguities. This process will necessarily reduce the covariance matrix, lowering the expected errors in the station coordinates. This does not necessarily mean that the solution is better, but that it statistically ought to be better, assuming the integers were correctly fixed. The assessment of solution accuracy goes beyond the scope of this discussion, but basically one can compare results with previous results (using GPS, or even some other technique). In addition, how well the data are fit by the model is reflected in the standard deviation of the post-fit residuals.

6.2.3 The Design Matrix

The coefficients of the design matrix can be illustrated by looking at a single row, for example, corresponding to observation $L_{AB}^{24}(i)$:

$$A_{AB}^{24}(i) = \left(\frac{\partial L_{AB}^{24}(i)}{\partial x_B} \quad \frac{\partial L_{AB}^{24}(i)}{\partial y_B} \quad \frac{\partial L_{AB}^{24}(i)}{\partial z_B} \quad \frac{\partial L_{AB}^{24}(i)}{\partial N_{AB}^{21}} \quad \frac{\partial L_{AB}^{24}(i)}{\partial N_{AB}^{23}} \quad \frac{\partial L_{AB}^{24}(i)}{\partial N_{AB}^{24}} \right)$$
$$= \left(\frac{\partial \rho_{AB}^{24}(i)}{\partial x_B} \quad \frac{\partial \rho_{AB}^{24}(i)}{\partial y_B} \quad \frac{\partial \rho_{AB}^{24}(i)}{\partial z_B} \quad 0 \quad 0 \quad -\lambda_0 \right)$$

As an example of one of the partial derivatives for one of the coordinates:

$$\begin{aligned} \frac{\partial \rho_{AB}^{24}(i)}{\partial x_B} &= \frac{\partial}{\partial x_B}\left(\rho_A^2(i) - \rho_B^2(i) - \rho_A^4(i) + \rho_B^4(i) \right) \\ &= \frac{\partial \rho_A^2(i)}{\partial x_B} - \frac{\partial \rho_B^2(i)}{\partial x_B} - \frac{\partial \rho_A^4(i)}{\partial x_B} + \frac{\partial \rho_B^4(i)}{\partial x_B} \\ &= \frac{\partial \rho_B^4(i)}{\partial x_B} - \frac{\partial \rho_B^2(i)}{\partial x_B} \\ &= \frac{x_{B0} - x^4(i)}{\rho_B^4(i)} - \frac{x_{B0} - x^2(i)}{\rho_B^2(i)} \end{aligned}$$

6.2.4 Minimum Data Requirements for Least Squares

For a least squares solution, a necessary condition is that the number of data exceed the number of estimated parameters

$$d \geq p$$

where we allow for the "perfect fit solution" ($d = p$). Under the assumption that all receivers track the same satellites for every epoch, the number of linearly independent double differences is

$$d = q(r-1)(s-1)$$

where q is the number of epochs, r the number of receivers, and s is the number of satellites. Assuming no cycle slip parameters:

$$p = 3 + (r-1)(s-1)$$

where there are $(r-1)(s-1)$ ambiguity parameters. Therefore,

$$\begin{aligned} q(r-1)(s-1) &\geq 3 + (r-1)(s-1) \\ (q-1)(r-1)(s-1) &\geq 3 \end{aligned}$$

Now, we know that $s \geq 2$ and $r \geq 2$ for us form double differences. Therefore, we can deduce that $q \geq 4$ if we have the minimal geometry of 2 receivers and 2 satellites (only one double difference per epoch!). Obviously, this minimal configuration is very poor geometrically, and would not be recommended as a method of precise positioning.
Note that no matter how many receivers or satellites we have, q is an integer, and therefore under any circumstance, we must have at least $q \geq 2$. That is, we cannot do single epoch relative positioning, if we are estimating integer ambiguities. If we can find out the ambiguities by some other means, then single epoch relative positioning is possible. Otherwise, we have to wait for the satellite geometry to change sufficiently in order to produce a precise solution.

For a single baseline $r = 2$ with 2 epochs of data $q = 2$ (which we should assume are significantly separated in time), the minimum number of satellites to produce a solution is condition $s \geq 4$. Interestingly, this corresponds to the minimum number of satellites for point positioning. If a tropospheric parameter were also being estimated, the condition would be $s \geq 5$. Of course, these conditions can be relaxed if we have more than 2 epochs, however it is the end-points of a data arc which are most significant, since they usually represent the maximum geometrical change which we require for a good solution. In summary, one can achieve very good results over short distances with only 4 satellites, but over longer distances where the troposphere must be estimated, a minimum of 5 satellites is recommended (at least some time during the session).

6.3 STOCHASTIC MODEL

6.3.1 Statistical Dependence of Double Differences

We have seen how double differences can be linearly dependent. The problem we now address is that double differenced observations that involve a common receiver and common satellite are *statistically dependent*. For example, at a given epoch, double differences L_{AB}^{21}, L_{AB}^{23} and L_{AB}^{24} are correlated due to the single differenced data in common, L_{AB}^{2}. Any measurement error in this single difference will contribute exactly the same error to each of the double differences. Therefore, a positive error in L_{AB}^{21} is statistically more likely to be accompanied by a positive error in L_{AB}^{23}. As another example, if we are processing a network using a reference satellite j and reference receiver A, all double differences in the linearly independent set will be statistically dependent because of the data in common, L_{A}^{j}.

6.3.2 Data Weight Matrix for Double Differences

In a situation where data are correlated, weighted least squares is appropriate. To complete our description of how to compute a relative position estimate, we therefore need to explain how to compute the appropriate data weight matrix, **W**. The construction of **W** can be generally called the "stochastic model," which describes the statistical nature of our data (as opposed to the "functional model" described so far, from which the observables can be computed deterministically.)

(As an aside for more advanced readers, some software process undifferenced observations, estimating clock biases as "stochastic parameters" at every epoch. It should be emphasised that there is a equivalence between explicit estimation of "stochastic parameters," and the use of an appropriate "stochastic model" which, in effect, accounts for the missing parameters through the introduction of correlations in the data. In principle, any parameter can either be estimated as part of the functional model, or equivalently removed using an appropriate stochastic model. To go more into this would be beyond the scope of this text.)

The weight matrix is the inverse of the covariance matrix for the double differenced data:

$$\mathbf{W} = \mathbf{C}_{\nabla\Delta}^{-1}$$

which has dimensions $q(r-1)(s-1) \times q(r-1)(s-1)$.

We start by assuming a covariance matrix for undifferenced data (i.e., the actually recorded data), which has dimensions $qrs \times qrs$. Typically, this is assumed to be diagonal, since the receiver independently measures the signals from each satellite separately. We shall, however, keep the form general. So the problem is, given a covariance matrix for undifferenced data, how do we compute the covariance matrix for double-differenced data? This is achieved using the rule of propagation of errors, which we have already seen in section 4.2.3, where geocentric coordinates were mapped into topocentric coordinates using an affine transformation. By analogy, we can deduce that the covariance of double-differenced data can be written:

$$\mathbf{C}_{\nabla\Delta} = \mathbf{D}\mathbf{C}\mathbf{D}^{\mathrm{T}}$$

where **D** is the matrix which transforms a column vector of the recorded data into a column vector of double differenced data:

$$\nabla\Delta\mathbf{L} = \mathbf{D}\mathbf{L}$$

Clearly, **D** is a rectangular matrix with the number of rows equal to the number of linearly independent double-differenced data, and the number of columns equal to the number of recorded data. Using our previous assumptions, **D** has dimensions $q(r-1)(s-1) \times qrs$. The components of **D** must have values of either +1, −1, or 0, arranged such that we produce a linearly independent set of double differences (see section 6.1.1). To complete this discussion, the double differenced data weight matrix can be written:

$$\mathbf{W} = \left(\mathbf{D}\mathbf{C}\mathbf{D}^{\mathrm{T}}\right)^{-1}$$

6.3.3 Covariance Matrix for Estimated Parameters

As we have already seen, for weighted least squares we can write the computed covariance matrix for estimated parameters as:

$$\mathbf{C}_{\mathbf{x}} = \mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A}^{-1}$$

We can now write down the full expression for the computed covariance matrix, by substituting for the double differenced data weight matrix **W**:

$$\mathbf{C}_X = \left(\mathbf{A}^T\left(\mathbf{D}\mathbf{C}\mathbf{D}^T\right)^{-1}\mathbf{A}\right)^{-1}$$

As mentioned above, for the (undifferenced) data covariance **C** we often use a diagonal matrix, assuming a value for the standard deviation of an observation. Typical realistic values for this are several mm. Although the receiver can usually measure the phase with better precision than a mm, the post-fit residuals typically show several mm standard deviations, due to unmodelled errors such as multipath.

Even using such an inflated value for measurement precision might not produce a realistic covariance matrix for station coordinates. This is partly due to two effects: (i) unmodelled errors can be correlated with the parameters being estimated (an "aliasing effect"), and (ii) post-fit almost always show some degree of time-correlation (e.g., due to multipath). A simple, and often surprisingly effective way to deal with this problem, is to multiply the final coordinate covariance matrix by an empirical scaling factor, inferred "by experience," according to the brand of software being used, the observation scenario, and the estimation strategy used.

7. INTRODUCING HIGH PRECISION GPS GEODESY

7.1 HIGH PRECISION SOFTWARE

The observable model discussed so far has been very basic, as it glosses over advanced features that are important for high precision software. Several software packages have been developed since the 1980's that are capable of delivering high precision geodetic estimates over long baselines. This software is a result of intensive geodetic research, mainly by universities and government research laboratories.

Typical features of such software include:

- orbit integration with appropriate force models;
- accurate observation model (Earth model, media delay...) with rigorous treatment of celestial and terrestrial reference systems;
- reliable data editing (cycle-slips, outliers);
- estimation of all coordinates, orbits, tropospheric bias, receiver clock bias, polar motion, and Earth spin rate;
- ambiguity resolution algorithms applicable to long baselines;
- estimation of reference frame transformation parameters and kinematic modelling of station positions to account for plate tectonics and co-seismic displacements.

We can summarise the typical quality of geodetic results from 24 hours of data:

- relative positioning at the level of few parts per billion of baseline length;
- absolute (global) positioning at the level of 1 cm in the IERS Terrestrial Reference Frame (ITRF);

- tropospheric delay estimated to a few mm;
- GPS orbits determined to 10 cm;
- Earth pole position determined to 1 cm;
- clock synchronisation (relative bias estimation) to 100 ps.

Two features of commercial software are sometimes conspicuously absent from more advanced packages: (i) sometimes double differencing is not implemented, but instead, undifferenced data are processed, and clock biases are estimated; (ii) network adjustment using baseline solutions is unnecessary, since advanced packages do a rigorous, one-step, simultaneous adjustment of station coordinates directly from all available GPS observations.

Some precise software packages incorporate a Kalman filter (or an equivalent formulism). This allows for certain selected parameters to vary in time, according to a statistical ("stochastic") model. Typically this is used for the tropospheric bias, which can vary as a random walk in time. A filter can also be used to estimate clock biases, where "white noise" estimation of clock bias approaches the theoretical equivalent of double differencing.

Although many more packages have been developed, there are 3 ultra high-precision software packages which are widely used around the world by researchers and are commonly referenced in the scientific literature:

- BERNESE software, developed by the Astronomical Institute, University of Berne, Switzerland;
- GAMIT software, developed by the Massachussets Institute of Technology, USA;
- GIPSY software, developed by the Jet Propulsion Laboratory, California Institute of Technology, USA

There are several other packages, but they tend to be limited to the institutions that wrote them. It should be noted that, unlike commercial software packages, use of the above software can require a considerable investment in time to understand the software and how best to use it under various circumstances. Expert training is often recommended by the distributors.

7.2 SOURCES OF DATA AND INFORMATION

For high precision work, it is important to abide by international reference system standards and use the best available sources of data and ancilliary information. We therefore summarise two especially important international sources of data information for the covenience of the interested reader:

- IERS: International Earth Rotation Service
  - Central Bureau located at the Paris Observatory, France
  - Documented IERS Conventions for observation models and reference systems
  - IERS Annual Reports
  - IERS Terrestrial Reference Frame for reference station coordinates
  - Routine publication of Earth rotation parameters

- IGS: International GPS Service for Geodynamics

- Central Bureau located at the Jet Propulsion Laboratory, USA
- Documented IGS Standards for permanent GPS stations
- Oversees operation of global GPS network (~100 stations)
- Distributes tracking data and precise ephemerides
- Maintains on-line database with Internet access

8. CONCLUSIONS

Having read and understood this text, you should now understand the basics of GPS positioning observation models and parameter estimation. You should also have an appreciation of the difference between basic positioning, and the more advanced positioning using high precision software packages. If all has gone well, and you think the above statements are true, then you should now have a good background knowledge and an appropriate context to prepare you for more advanced material.

GPS Satellite Orbits, Orbit Determination, and the IGS

Markus Rothacher

Astronomical Institute
University of Berne
Switzerland

Nordic Research Course of the Nordic Geodetic Commission
(NKG)

Geodetic Applications of GPS

August 26 - 31, 1996

Båstad, Sweden

Contents

1 Introduction

There are presently two similar satellite systems available for global positioning and navigation as well as time transfer, namely

NAVSTAR GPS: **Na**vigation **S**atellite **T**iming **A**nd **R**anging **G**lobal **P**ositioning **S**ystem

GLONASS : **Glo**bal **Na**vigation **S**atellite **S**ystem

Both these systems have a very similar overall design (see e.g. [*Milliken and Zoller*, 1980] for GPS and [*Anodina*, 1988] for GLONASS). Although our main emphasis will be on the orbits of the GPS satellites, most of the topics presented in this lecture may easily be transformed to the case of GLONASS satellites.

The major characteristics of the two satellite systems are summarized in Table 1.

Characteristic	GPS	GLONASS
Altitude	20'200 km	19'100 km
Orbital Period	11 h 58 min	11 h 16 min
Orbit Inclination	55 deg (Block I) 63 deg (Block II)	65 deg
Number of Planes	6	3
Number of Satellites	24 (3 spares)	24 (3 spares)
Approximate Mass	815 kg	700 kg
Data Rate	50 bit/sec	50 bit/sec
PRN Codes[1]	Satellite-dependent	Satellite-indep.
Frequencies	Satellite-indep.	Satellite-dependent

[1] Pseudo-Random Noise Codes

Table 1: Characteristics of GPS and GLONASS Orbits.

The nominal satellite constellation of both systems consists of 24 satellites (21 + 3 spares). The satellites are revolving about the Earth in almost circular orbits at an altitude of about 20'200 km for GPS and 19'100 km for GLONASS. The GPS satellites have a revolution period of almost exactly half a sidereal day (11 hours 58 minutes) and are therefore in deep resonance with the rotation of the Earth itself. This leads to resonance phemonena (see Section 3.1) that are not present for the GLONASS satellites with a revolution period of 11 hours and 16 minutes. The GPS constellation has the nice characteristic, however, that the same configuration of satellites can be observed from a given point on the surface of the Earth at the same time on consecutive

days (more precisely: the configuration repeats itself almost perfectly after 23 hours and 56 minutes).

Whereas four GPS satellites are located in each of six orbital planes separated by 60 degrees in longitude (see satellite arrangement in Figure 1), the GLONASS satellites occupy only three different planes with nominally eight satellites in each plane at 45 degree intervals. Because the GLONASS constellation is changing very often, the positions of the satellites in their respective planes are not shown here.

Figures 2 and 3 show the arrangement of orbital planes as seen from space at 35 degree latitude for GPS and from above the North (or South) pole for GPS and GLONASS (pairs of GPS orbits coincide in this projection).

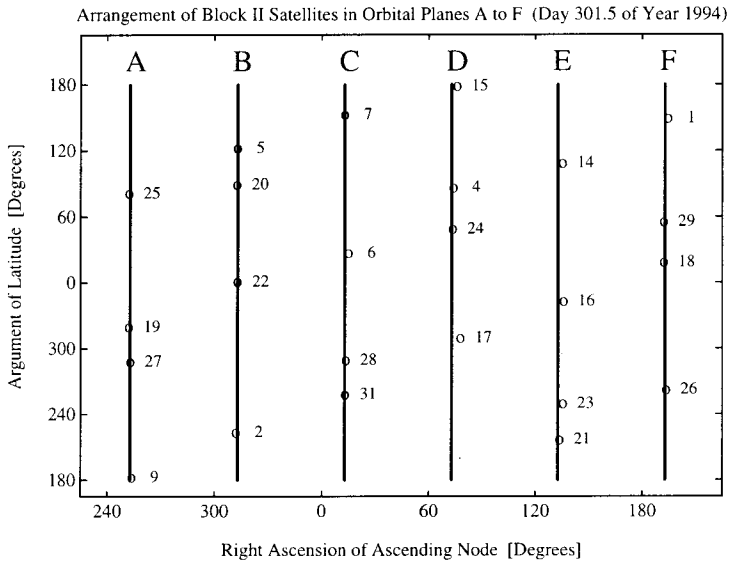


Figure 1: Arrangement of the GPS satellites in the orbital planes named A to F.

Both constellations have been designed in such a way that at least 4 satellites should be visible at any time from almost every point on the Earth's surface, the minimum number of satellites necessary to allow real-time navigation.

From Figures 2 and 3 it can easily be seen, however, that — due to the inclination of 55 degrees — no GPS satellites can be observed at an elevation above 53 degrees in the polar regions. At mid latitudes there is always a fairly large area in the north, where no satellites are ever visible. These weak points in the geometry of the GPS satellite constellation are less pronounced for GLONASS.

PRN 4, launched in February 1978, was the first GPS satellite in orbit and was part of a series of 11 Block I satellites forming the test configuration available in the eighties. In February 1989 the first Block II satellite was launched and the full constellation of

24 satellites was complete around November 1993. Since that time at least 24 satellites have always been active.

The situation is quite different in the case of GLONASS. The first GLONASS satellite was put into orbit in October 1982. The full set of 24 satellites became available around January 1996, but in view of the short life time of most GLONASS satellites and the economic situation in Russia, the future of the GLONASS system is not clear at all. More details about the GLONASS system may be found in the lecture notes by Börje Forssell "Technical Comparison Between the GLONASS and GPS Concepts" in this volume.

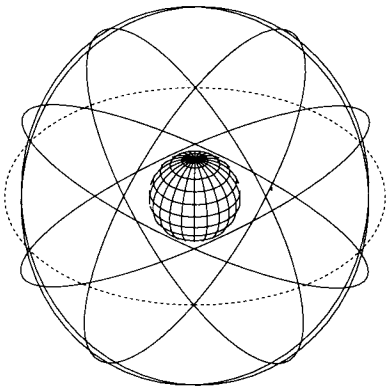


Figure 2: GPS orbits viewed from latitude $\beta = 35^{\circ}$.

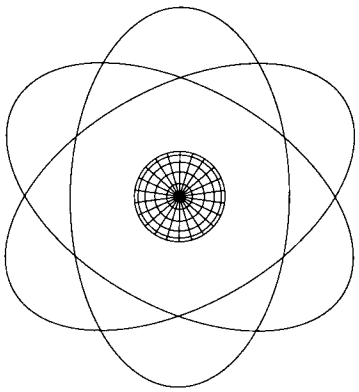


Figure 3: GPS orbits viewed from latitude $\beta = 90^{\circ}$.

Figure 4, finally, contains drawings of the various GPS spacecrafts (Block I, Block II, and Block IIR; from [*Fliegel et al.*, 1992]). It is quite obvious, that the modeling of the forces acting on such large and complex spacecrafts is a difficult task we will be looking into in the following sections.

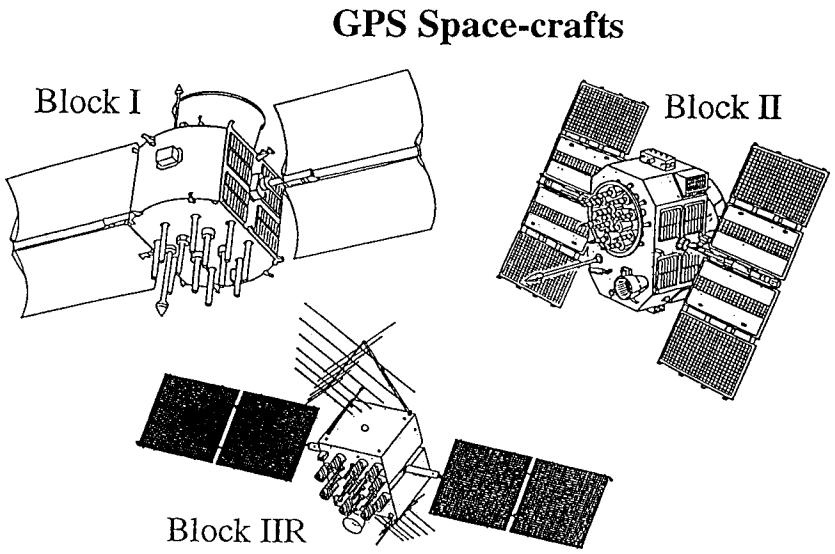


Figure 4: GPS satellites of different generations: Block I, Block II, and Block IIR.

We finish this introduction with a short overview of the sections of these lecture notes.

In the next section (Section 2) we will give an introduction into the laws governing the motion of a satellite around the Earth, starting with Kepler's laws and ending with the modern equations of motion of the satellite.

The Section 3 will discuss in detail the various forces and perturbations acting on the satellites.

The problem of the estimation of orbital parameters from observations is the topic of Section 4, where we will also have a look at the principles of numerical integration.

After studying (in Section 5) the quality of the GPS orbit types available today, we will give a short summary of the objectives, structure, and products of the IGS, the International GPS Service for Geodynamics, in Section 6.

Much of the material found in these sections is based on the publications [*Beutler*, 1990], [*Rothacher*, 1992], and [*Beutler et al.*, 1996]. There you will also find additional topics and information complementary to the material covered here.

2 The Keplerian Laws and the Equation of Motion

2.1 The Keplerian Laws and Keplerian Elements

During the years 1576–1597 (in Denmark) and 1599–1600 (in Prague) Tycho Brahe (1546–1601) observed the planets and the Sun with an unprecedented accuracy. Based on this long time series (24 years) of astronomical observations Johannes Kepler (1571–1630) found the famous three **Keplerian Laws** of planetary motion. The first two of these laws were published in his work **Astronomia Nova** [*Kepler*, 1609]. Ten years later he published the third law in **Harmonices Mundi Libri V** [*Kepler*, 1619].

The three Keplerian Laws are [*Danby*, 1989]:

1. The orbit of each planet is an ellipse, with the Sun at one of the foci.
2. Each planet revolves so that the line joining it to the Sun sweeps out equal areas in equal intervals of time.
3. The squares of the periods of any two planets are in the same proportion as the cubes of their mean distances to the Sun.

The second law, also known as the "Law of Areas", is illustrated in Figure 5.

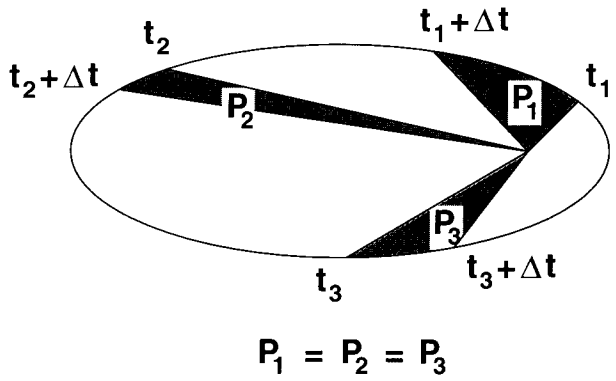


Figure 5: Illustration of Kepler's second law: the law of areas.

These three laws of planetary motion can easily be reformulated for the motion of a satellite around the Earth by replacing "Sun" with "Earth" and "planet" with "satellite".

According to Kepler's first law, the orbit of a satellite around the Earth is therefore an ellipse and may be described by its shape, given by the semi-major axis a and the numerical eccentricity e of the ellipse, and three angles that define the orientation of the ellipse in space, namely the inclination i of the orbital plane with respect to the

equator, the angle Ω between the direction to the equinox and the ascending node of the orbit, and the angle ω measured in the orbital plane between the node and the perigee, the point of the orbit nearest to the Earth's center. As a sixth parameter we need the information on when the satellite is at a specific position in its orbit, e.g., the time when the satellite passes through the perigee T_p. These six so-called **Keplerian Elements**, that are very often used to parameterize the orbit of a satellite, are listed in Table 2 and shown in Figures 6 and 7.

a	**Semi-major axis** of the ellipse, defining the size of the orbit.
e	**Numerical eccentricity** of the ellipse, describing the shape of the orbit.
i	**Inclination** of the orbital plane with respect to the equatorial plane (mean equator at a standard epoch, e.g. 2000.0).
Ω	**Right ascension of the ascending node**, i.e. the angle between the direction to the **vernal equinox** (X-direction in Figure 7) and the intersection line of the satellite's orbital plane with the equatorial plane.
ω	**Argument of perigee**, the angle (in the orbital plane) between the ascending node and the perigee (measured in the direction of the motion of the satellite).
T_p	**Perigee passing time**, the time when the satellite passes through the point nearest to the Earth, the **perigee**.

Table 2: The Keplerian elements.

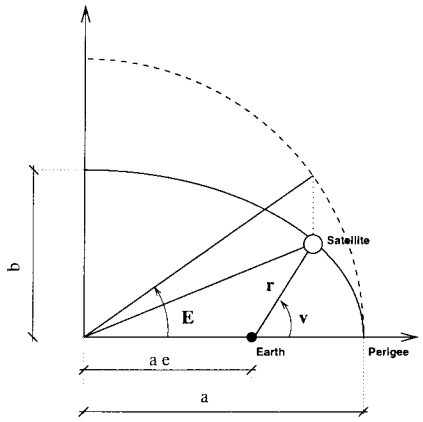


Figure 6: The Keplerian elements semi-major axis a and numerical eccentricity e. E and v are the eccentric anomaly and true anomaly, respectively.

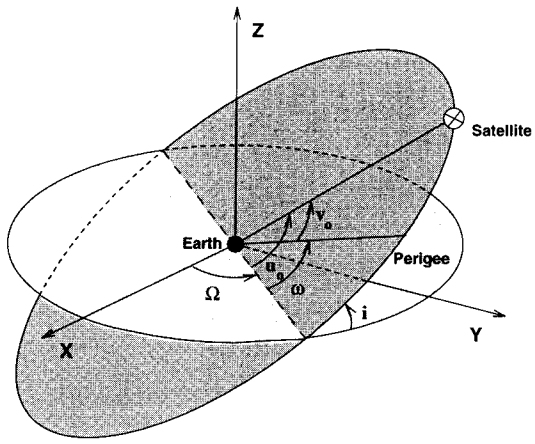


Figure 7: The Keplerian elements ascending node Ω, inclination i, and argument of perigee ω (see Table 2).

In reality the Keplerian laws are only correct to a first approximation. Because the Earth is not a point mass and because of the gravitational forces exerted by Sun and Moon, the satellites are not revolving around the Earth in perfect ellipses. But let us first show, how we may compute the position and velocity of a satellite starting with the Keplerian elements.

2.2 Satellite Position Computed from Keplerian Elements

Let us assume that we know the Keplerian elements describing the orbit of a satellite and that, in addition, we know the **revolution period** U of the satellite around the Earth. (In the next Section 2.3 we will see, how the revolution period may be computed from the masses of the satellite and the Earth, the gravity constant, and the semi-major axis of the satellite orbit; see Eqn. 20). From the revolution period U we obtain the mean angular velocity n of the satellite in its orbit around the Earth, also called the **mean motion** (in radian):

$$n = \frac{2 \cdot \pi}{U} \tag{1}$$

To compute the position at a given time t we have to introduce the **mean anomaly** M as a linear function in time (angle between the perigee and the fictitious position the satellite would have, if it were revolving with a constant mean motion n on a circle with radius a):

$$M = n \cdot (t - T_p) \tag{2}$$

Often the mean anomaly $M_0 = M(t_0)$ at a reference epoch t_0 is given instead of the perigee passing time T_p. In this case we compute the mean anomaly M at time t as

$$M = M_0 + n \cdot (t - t_0) \tag{3}$$

From the mean anomaly M we then obtain the **eccentric anomaly** E by using **Kepler's Equation**:

$$E = M + e \cdot \sin E \tag{4}$$

The eccentric anomaly is the angle between the direction from the center of the ellipse to the perigee (line of apsides) and the direction from the center to the point P'. As illustrated in Figure 8, P' is the projection of the point P normal to the semi-major axis onto the circle of radius a around the ellipse. Kepler's equation may be derived when applying the second law of Kepler to the time intervals $[T_p, t]$ and $[T_p, T_p + U]$ and the corresponding areas. Kepler's equation has to be solved iteratively to obtain the eccentric anomaly E.

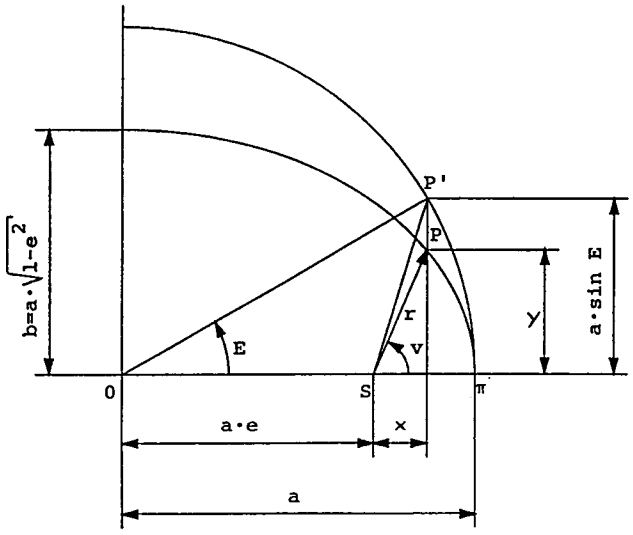


Figure 8: Relationships in the ellipse, eccentric anomaly E and true anomaly v.

In a coordinate system with the origin in the center of the ellipse, the x-axis pointing to the perigee (semi-major axis), the y-axis in the direction of the semi-minor axis, and the z-axis perpendicular to the orbital plane, we may write the position of the satellite as (see also Figure 8):

$$x = a \cdot (\cos E - e) \tag{5}$$

$$y = a \cdot \sqrt{1 - e^2} \cdot \sin E \tag{6}$$

$$z = 0 \tag{7}$$

The radius vector (distance satellite–center of the Earth) is then given by

$$r = \sqrt{x^2 + y^2 + z^2} = a \cdot (1 - e \cdot \cos E) \tag{8}$$

To compute the coordinates x', y', z' of the satellite in the equatorial system we have to rotate the coordinates x, y, z as follows:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R_3(-\Omega) \cdot R_1(-i) \cdot R_3(-\omega) \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{9}$$

where $R_i(\alpha)$ is a 3×3 matrix describing a rotation around the axis i by the angle α.

The velocity components $\dot{x}', \dot{y}', \dot{z}'$ of the satellite at time t may be computed by taking the first derivative of Eqn. 9 with respect to time:

$$\begin{pmatrix} \dot{x}' \\ \dot{y}' \\ \dot{z}' \end{pmatrix} = R_3(-\Omega) \cdot R_1(-i) \cdot R_3(-\omega) \cdot \begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix} \tag{10}$$

$\dot{x}, \dot{y}, \dot{z}$ are the time derivatives of Eqn. 5:

$$\dot{x} = -a \cdot \sin E \cdot \dot{E} = -n \cdot \frac{a^2}{r} \cdot \sin E \tag{11}$$

$$\dot{y} = a \cdot \sqrt{1 - e^2} \cdot \cos E \cdot \dot{E} = n \cdot \frac{a^2}{r} \cdot \sqrt{1 - e^2} \cdot \cos E \tag{12}$$

$$\dot{z} = 0 \tag{13}$$

where we substituted $\dot{E}$ obtained by taking the first time derivative of Kepler's equation (Eqn. 4):

$$\dot{E} = \frac{n}{1 - e \cdot \cos E} = n \cdot \frac{a}{r} \tag{14}$$

These are the formulas to compute – in the Keplerian approximation – the rectangular coordinates of the position and velocity vectors of a satellite at an epoch t, starting from the Keplerian elements. We have thus shown that the position and velocity are functions of the Keplerian elements (and of time t).

2.3 The Equation of Motion

To obtain a more precise and general description of the motion of a satellite we have to look at the next major step in the history of celestial mechanics: in 1687 Isaac Newton (1643–1727) published the well-known **laws of motion** and the **law of universal gravitation** in his book **Philosophiae Naturalis Principia Mathematica** [*Newton*, 1687]. He showed that the three laws of Kepler can be derived from the more fundamental laws of motion and the law of universal gravitation under the assumption, that only two bodies are present (two-body problem) and that these bodies have a spherically symmetric mass distribution.

Newton's laws of motion (see [*Danby*, 1989]) state:

1. Every particle continues in a state of rest or uniform motion in a straight lin unless it is compelled by some external force to change that state.

2. The rate of change of the linear momentum of a particle is proportional to the force applied to the particle and takes place in the same direction as that force.

3. The mutual actions of any two bodies are always equal and oppositely directed.

Interpreting the "rate of change of the linear momentum" as the first derivative in time of the momentum $\vec{p} = m \cdot \dot{\vec{r}}$ (product of mass and velocity of the particle) and assuming that the mass of the particle does not change in time, we may write the second law of motion as

$$m \cdot \ddot{\vec{r}} = \vec{F} \tag{15}$$

where

m ... Mass of the particle
$\vec{r}$ ... Position vector in inertial space
$\vec{F}$ ... Force acting on the particle or, more precisely, the vectorial sum of all forces

Leonard Euler (1707–1783) was the first to formulate the second law in a modern mathematical language. These so-called **equations of motion** are second order differential equations in time.

Newton's law of universal gravitation states that there is an attractive force $\vec{F}$ between two particles of masses m_1 and m_2 of the magnitude

$$F = \left|\vec{F}\right| = G \cdot \frac{m_1 \cdot m_2}{r^2} \tag{16}$$

where

G ... Newtonian gravitational constant
r ... Distance between the two particles

It can be shown that this formulation is valid if the two particles are point masses or if the mass distribution of the two bodies is spherically symmetric.

From Eqn. 16 we obtain the **equations of motion of an artificial Earth satellite** in their simplest form under the assumption that the mass distribution within the Earth is spherically symmetric and neglecting all other forces that might act on the satellite.

$$\vec{F} = m \cdot \ddot{\vec{r}} = -G \cdot \frac{m \cdot M}{r^2} \cdot \frac{\vec{r}}{r} \tag{17}$$

or simply

$$\ddot{\vec{r}} = -GM \cdot \frac{\vec{r}}{r^3} \tag{18}$$

where

m ... Mass of the satellite
M ... Total mass of the Earth
GM ... Product of G and M ($398.600415 \cdot 10^{12} m^3 s^{-2}$ [*McCarthy*, 1992])

$\vec{r}$ ... Position of the satellite with respect to the center of mass of the Earth

These are the equations of motion of the so-called **two-body problem**. The general solutions are known to be **conic sections**. In satellite geodesy, we are mainly interested in ellipses and we already know from Section 2.1, that an elliptic orbit may be characterized by the six Keplerian elements.

Let us briefly show how in principle Kepler's laws can be derived starting from the equations of motion (Eqn. 18).

We can easily verify that the vector with the components x', y', z' defined in Eqn. 9 is a solution of the equations of motion provided that we adopt the relationship

$$n^2 \cdot a^3 = GM \tag{19}$$

or using Eqn. 1:

$$\frac{a^3}{U^2} = \frac{GM}{4\pi^2} \tag{20}$$

This is in fact the equivalent to Kepler's third law in the Newtonian formulation. It is only true in this form, if the mass m of the satellite may be neglected compared to the mass M of the Earth. This is always the case for an artificial Earth satellite. For the Moon, however, the term GM must be replaced by the term $G \cdot (M + m)$.

Multiplying the equations of motion (Eqn. 18) by $\vec{r}$ using the vector product we get

$$\vec{r} \times \ddot{\vec{r}} = \vec{0} \tag{21}$$

this means that, because of

$$\frac{d}{dt}\left(\vec{r} \times \dot{\vec{r}}\right) = \underbrace{\dot{\vec{r}} \times \dot{\vec{r}}}_{=0} + \underbrace{\vec{r} \times \ddot{\vec{r}}}_{=0\,(Eqn.\ 21)} = \vec{0}, \tag{22}$$

the product $\vec{r} \times \dot{\vec{r}}$ is constant in time and that the motion takes place in one plane, the **orbital plane**, and we may write

$$\vec{r} \times \dot{\vec{r}} = \vec{h} \tag{23}$$

where $\vec{h}$ is a constant vector normal to the orbital plane. The second law of Kepler follows directly from Eqn. 23, if we interpret

$$\left|\vec{r} \times \dot{\vec{r}}\right| \cdot \Delta t = \vec{h} \cdot \Delta t = 2 \cdot A \tag{24}$$

as twice the area A swept out by the radius vector $\vec{r}$ in the time interval Δt. Eqn. 23 is also equivalent to the law of conservation of angular momentum.

If we denote the components of $\vec{h}$ in the equatorial system by h_1,h_2,h_3, we may easily compute the right ascension of the ascending node Ω and the inclination i of the orbital plane with respect to the equatorial plane:

$$\Omega = \arctan\left(\frac{h_1}{-h_2}\right) \tag{25}$$

$$i = \arctan\left(\frac{\sqrt{h_1^2 + h_2^2}}{h_3}\right) \tag{26}$$

We have shown now, that two of the Keplerian elements may be computed from the position vector $\vec{r}(t)$ and the velocity vector $\dot{\vec{r}}(t)$. This is also true for all the other Keplerian elements (see e.g. [*Beutler et al.*, 1996]): each of the elements $K_i \in \{a, e, i, \Omega, \omega, T_p\}$ may be written as a function of the position vector at time t and the velocity vector at time t (for an arbitrary time t):

$$K_i = f_{K_i}(\vec{r}, \dot{\vec{r}}; t) \tag{27}$$

On the other hand we have seen in Section 2.2 that the position and velocity vectors, $\vec{r}(t)$ and $\dot{\vec{r}}(t)$, may be computed from the Keplerian elements (Eqns. 1 to 14):

$$\vec{r}(t) = f_r(a, e, i, \Omega, \omega, T_p; t) \tag{28}$$

$$\dot{\vec{r}}(t) = f_{\dot{r}}(a, e, i, \Omega, \omega, T_p; t) \tag{29}$$

There exists, therefore, a one-to-one correspondence between the position and velocity vector at time t and the Keplerian elements:

$$a, e, i, \Omega, \omega, T_p \Longleftrightarrow \vec{r}, \dot{\vec{r}} \tag{30}$$

We all know that the equations of motion (Eqn. 18) given at the beginning of this section are only an approximation, because neither is the mass distribution in the Earth spherically symmetric nor are we allowed to neglect all other forces acting on the satellite. To account for all the forces acting on the satellite we rewrite the equations of motion in the following way:

$$\ddot{\vec{r}} = -GM \cdot \frac{\vec{r}}{r^3} + \vec{a}(t, \vec{r}, \dot{\vec{r}}, q_1, q_2, q_3, \ldots) \tag{31}$$

where the first term represents the central gravity term and $\vec{a}$ the total **perturbing acceleration** (discussed in the next section), or in short

$$\ddot{\vec{r}} = \vec{f}(t, \vec{r}, \dot{\vec{r}}, q_1, q_2, q_3, \ldots) \tag{32}$$

The equations of motion (Eqn. 31 or Eqn. 32) alone do not yet define a unique orbit. In addition we have to specify the so-called **initial conditions** at a time t_0

$$\vec{r}(t_0) = \vec{r}_0 \quad , \quad \dot{\vec{r}}(t_0) = \vec{v}_0 \tag{33}$$

to define a unique solution $\vec{r}(t)$ of the differential equation system (Eqn. 32). This second order differential equation system, in general, **cannot be solved analytically**, because the function $\vec{f}$ may be very complicated (including the Earth gravity field, tidal forces, solar radiation pressure, etc.). Therefore, **numerical integration algorithms** have to be used (see Section 4.5). $q_1, q_2, q_3, \ldots$ are **dynamical parameters** (e.g. solar radiation pressure parameters) which are not sufficiently well-known and have to be estimated together with the initial conditions in an **orbit determination** procedure (see Section 4).

Because of the perturbation term $\vec{a}$ the orbit can no longer be characterized by **one set** of Keplerian elements. The perturbation term $\vec{a}$ is, however, small compared to the two-body term (or central gravity term) and the solution of the unperturbed equations is therefore a good approximation to the perturbed solution, at least in the vicinity of the initial epoch t_0. It thus makes sense to introduce an **instantaneous ellipse** and to speak of the Keplerian elements as evolving in time. Let us assume that $\vec{r}(t)$ and $\dot{\vec{r}}(t)$ are the **true** position and velocity vectors for each time argument t as they are computed from the equations of motion (Eqn. 32). The **osculating elements** at time t are then defined as the Keplerian elements computed from $\vec{r}(t)$ and $\dot{\vec{r}}(t)$ according to Eqn. 25, Eqn. 26, and similar equations for the other elements, which are valid for the unperturbed two-body case. In this way we get time series of osculating elements $a(t)$, $e(t)$, $i(t)$, $\Omega(t)$, $\omega(t)$, and $T_p(t)$.

3 Forces Acting on GPS Satellites

Let us now have a closer look at the second term $\vec{a}$ on the right hand side of the equations of motion (Eqn. 31). This perturbing acceleration is composed of

$$\vec{a} = \vec{a}_{ns} + \vec{a}_{ms} + \vec{a}_{et} + \vec{a}_{rp} + \vec{a}_{rest} \tag{34}$$

where the various accelerations are due to:

$\vec{a}_{ns}$ ... Non-sphericity of the Earth gravity potential
$\vec{a}_{ms}$ ... Gravitational acceleration due to the Moon and the Sun
$\vec{a}_{et}$ ... Earth tidal potential
$\vec{a}_{rp}$ ... Solar radiation pressure

and $\vec{a}_{rest}$ denotes the sum of all the small accelerations ($< 10^{-9} m/s^2$) to be discussed in Section 3.5. The other terms will be the topic of the next sections. Table 3 summarizes the order of magnitude of the various perturbing accelerations.

3.1 Non-Central Part of the Earth Gravitational Potential

The most important perturbation of the elliptic motion of a satellite is caused by the non-sphericity of the Earth gravity field. The non-spherical part V_{ns} of the **Earth gravity potential** is usually represented by a spherical harmonics expansion in a geocentric Earth–fixed system:

$$V_{ns}(r, \lambda, \phi) = \frac{GM}{r} \cdot \sum_{n=2}^{\infty} \left(\frac{a_E}{r}\right)^n \cdot \sum_{m=0}^{n} P_{nm}(\sin\phi) \cdot (S_{nm} \cdot \sin(m\,\lambda) + C_{nm} \cdot \cos(m\,\lambda)) \tag{35}$$

where

r ... Geocentric distance of the satellite
ϕ ... Geocentric latitude of the satellite

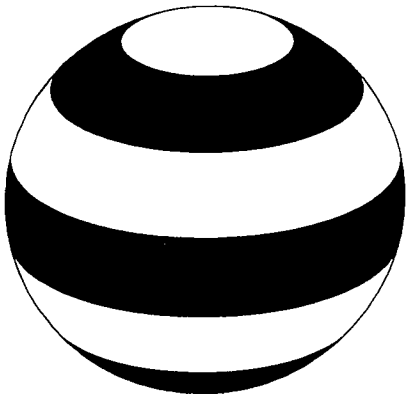
λ ... Geocentric longitude of the satellite
a_E ... Mean equatorial radius of the Earth
P_{nm} ... Associated Legendre functions of degree n and order m
C_{nm}, S_{nm} ... Denormalized geopotential coefficients

Perturbation	**Acceleration** m/s^2	**Orbit Error after one Day** (m)
Two-Body Term of Earth's Gravity Field	0.59	∞
Oblateness of Earth	$5 \cdot 10^{-5}$	10'000
Lunar Gravitational Attraction	$5 \cdot 10^{-6}$	3000
Solar Gravitational Attraction	$2 \cdot 10^{-6}$	800
Other Terms of Earth's Grav. Field	$3 \cdot 10^{-7}$	200
Radiation Pressure (direct)	$9 \cdot 10^{-8}$	200
Y-bias	$5 \cdot 10^{-10}$	2
Solid Earth Tides	$1 \cdot 10^{-9}$	0.3
Ocean Tides	$5 \cdot 10^{-10}$	0.04
Earth Albedo	$4 \cdot 10^{-10}$	0.03
Relativistic Effects	$3 \cdot 10^{-10}$	0.01

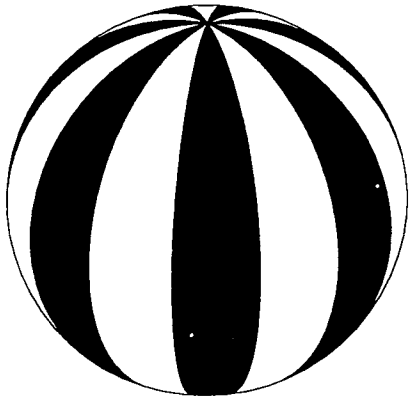
Table 3: Forces acting on GPS satellites (from [*Landau*, 1988]).

We distinguish between the **zonal terms** ($m = 0$), which only depend on latitude, the **sectorial terms** ($n = m$), which only depend on longitude, and the **tesseral terms** ($n > m \neq 0$), which depend on both, latitude and longitude. Examples are given in Figure 9.

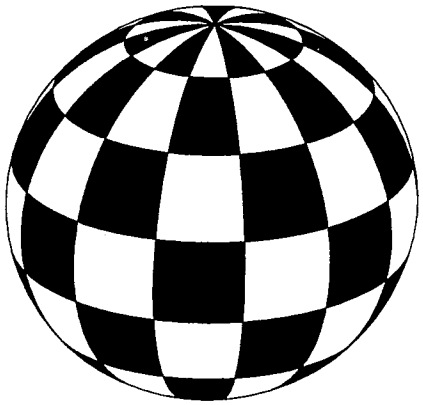
Terms with $n < 2$ are zero, since the origin of the Earth–fixed system is defined to coincide with the Earth's center of mass. Numerical values of the coefficients may be found in the IERS Standards [*McCarthy*, 1992].



(a) **Zonal** harmonics with $n = 6$, $m = 0$



(b) **Sectorial** harmonics with $n = m = 7$



(c) **Tesseral** harmonics with $n = 13$, $m = 7$

Figure 9: Examples for the Earth's gravity potential coefficients: sectors of equal sign in black and white, respectively.

The acceleration $\vec{a}_{ns}$ in the inertial frame is then given by

$$\vec{a}_{ns} = R \cdot D \cdot \vec{\nabla} V_{ns} \tag{36}$$

where $\vec{\nabla} V_{ns}$ is the gradient of the non-spherical part of the geopotential and D is the matrix containing the partial derivatives of the Earth–fixed coordinates (r, λ, ϕ) with respect to the geocentric coordinates (x, y, z) in the Earth–fixed frame. The exact

form of this matrix may e.g. be found in [*Landau*, 1988]. The matrix R, finally, is the transformation matrix from the Earth-fixed to the inertial coordinate system:

$$R(t) = P^T(t) \cdot N^T(t) \cdot R_3(-\Theta) \cdot R_1(y) \cdot R_2(x) \tag{37}$$

where

$R_i(\alpha)$ ... Matrix describing a rotation around axis i about the angle α
$P^T(t)$ ... Transpose of the precession matrix $P(t)$
$N^T(t)$ ... Transpose of the nutation matrix $N(t)$
Θ ... True Greenwich sidereal time
x, y ... Components of polar motion

More details on the transformation between the Celestial and Terrestrial Reference Frame (ICRF and ITRF) may be found in the IERS Standards [*McCarthy*, 1992].

The largest contribution to the acceleration $\vec{a}_{ns}$ comes from the Earth's potential coefficient C_{20} representing the oblateness of the Earth. It is, e.g., responsible for the precession of the line of nodes of a satellite orbit. Because GPS (and GLONASS) satellites are revolving in high altitude orbits, they are much less affected by the short wavelength terms of the geopotential than low orbiting satellites. It is usually sufficient to use an Earth potential model with terms up to degree and order 8 ([*Beutler et al.*, 1985]).

Due to the revolution period of almost exactly half a sidereal day, the GPS satellites (but not the GLONASS satellites) repeat their ground tracks daily. As a consequence there are resonance effects caused by tesseral and sectorial terms of the Earth's gravity field, e.g., a drift in the semi-major axis a of the GPS satellites of up to several meters per day. The changing semi-major axis changes the mean motion of the satellite and leads to a change in the satellite position within the constellation of typically 20 degrees in one year. Frequent manoeuvres of the satellites are therefore necessary to keep the constellation in place. For more details on these resonance phenomena we refer to [*Hugentobler*, 1995].

3.2 Gravitational Effects of Sun and Moon and Solid Earth Tides

The acceleration $\vec{a}_{ms}$ of the satellite due to the gravitational attraction of the Sun and Moon may be written as:

$$\vec{a}_{ms} = -G \cdot M_M \cdot \left(\frac{\vec{r} - \vec{r}_M}{|\vec{r} - \vec{r}_M|^3} + \frac{\vec{r}_M}{|\vec{r}_M|^3} \right) - G \cdot M_S \cdot \left(\frac{\vec{r} - \vec{r}_S}{|\vec{r} - \vec{r}_S|^3} + \frac{\vec{r}_S}{|\vec{r}_S|^3} \right) \tag{38}$$

where

G ... Gravity constant
$\vec{r}$ ... Geocentric position vector of the satellite
$\vec{r}_S, \vec{r}_M$ ... Geocentric position vectors of the Sun and the Moon, respectively

M_S, M_M ... Masses of the Sun and the Moon

The gravitational perturbations due to the Moon and the Sun are not the full gravitational attraction exerted by these bodies, but only the corresponding "tidal" term, i.e. the difference between the force acting on the satellite and the force acting on the Earth. This explains the fact that there are two terms on the right hand side of Eqn. 38. This equation may be obtained by subtracting the equations of motion of the Earth (in an inertial frame) from the equations of motion of the satellite (in an inertial frame). The resulting equation is then referring to the geocenter (not an inertial frame).

The gravitational attraction of Sun and Moon also has an indirect effect on the satellite orbit due to the solid earth tides and ocean tides it causes: the tidal deformations change the Earth's gravity potential, which in turn acts on the satellite. Formulas for the tidal effects may e.g. be found in [*Rothacher*, 1992].

3.3 Solar Radiation Pressure

The acceleration $\vec{a}_{rp}$ of a GPS (or GLONASS) satellite due to the radiation pressure of the Sun is quite large (see Table 3) and is very difficult to model because of the complicated shape of these active satellites. It may be expressed as follows (see e.g. [*Rothacher*, 1992]):

$$\vec{a}_{rp} = \nu \cdot \left(P_s \cdot C_r \cdot \frac{A}{m} \cdot a_s^2 \cdot \frac{\vec{r} - \vec{r}_s}{|\vec{r} - \vec{r}_s|^3} \right) \tag{39}$$

where

ν ... Eclipse factor ($\nu = 1$ if satellite in sunlight, $\nu = 0$ if satellite in the Earth's shadow, $0 < \nu < 1$ if satellite in penumbra).

A ... Cross-section area of the satellite as seen from the Sun

m ... Mass of the satellite

a_s ... Astronomical unit (AU)

$P_s = S/c$ ... Radiation pressure for a completely absorbing object with $A/m = 1$ at the distance of 1 AU. S is the solar constant and c the velocity of light.

C_r ... Reflection coefficient

$\vec{r}, \vec{r}_s$ ... Geocentric coordinates of the satellite and the Sun, respectively

Numerical values for the quantities C_r, A, and m for GPS satellites may be found in [*Fliegel et al.*, 1992].

The acceleration $\vec{a}_{rp}$ always points in the direction Sun–satellite in this model. Whereas for a spherical satellite the ratio A/m remains constant, the total cross-section area A for a GPS satellite is constantly changing due to the changing attitude of the satellite. The pressure exerted by the solar radiation will therefore vary over one revolution as well as over the year (because of the changing orientation of the orbital plane with respect to the direction to the Sun). The most commonly used radiation pressure models for GPS satellites stem from [*Fliegel et al.*, 1992]. They were derived

by first computing the contributions to the radiation pressure from all the major surfaces of the GPS spacecraft taking into account the reflectivity properties of these surfaces (see Figure 10) and by then representing these results by relatively simple formulas in a spacecraft-fixed coordinate frame (see Figure 11).

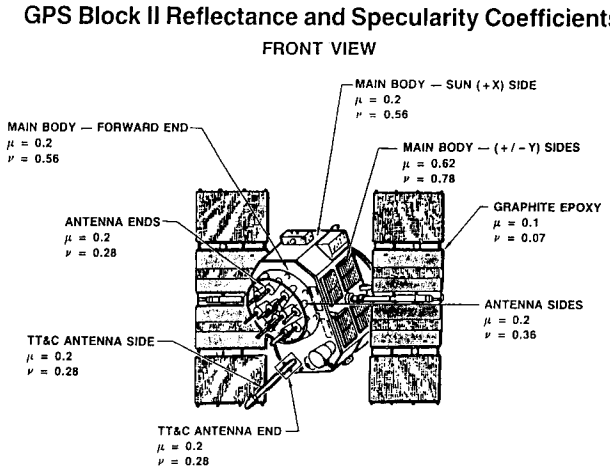


Figure 10: GPS Block II surfaces and their properties (from [*Fliegel*, 1993]).

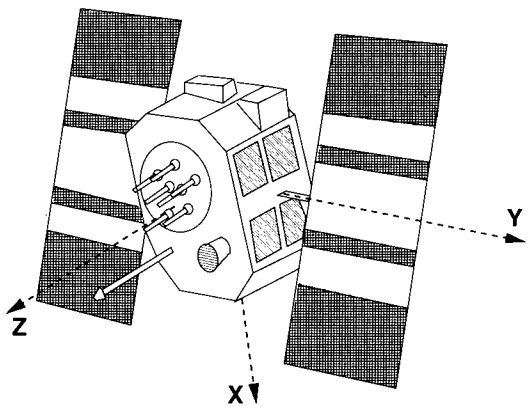


Figure 11: The satellite-fixed coordinate frame with the Z-axis pointing towards the center of the Earth, the Y-axis pointing into the direction of the solar panel axis, and the X-axis completing a right-hand system with Z and Y.

Assuming perfect attitude control the resulting force always lies in the (X, Z) plane and depends only on the angle β, the angle between the positive Z-axis and the direction from the Sun to the satellite. The models are called **Rock4** and **Rock42** for the Block I and Block II satellites, respectively. A distinction is made between the standard S-model and the T-model, which includes thermal re-radiation of the satellite and is recommended by the IERS Standards [*McCarthy*, 1992].

In practice the difference between the Rock4/Rock42 S- and T-models and the much simpler model, called Z-model, which assumes a constant acceleration in the direction Sun–satellite (see also Section 4.3), are very small, provided that either a direct radiation pressure parameter or a scaling parameter for the Rock models is estimated (two equivalent approaches). The differences are shown in Figures 12 and 13.

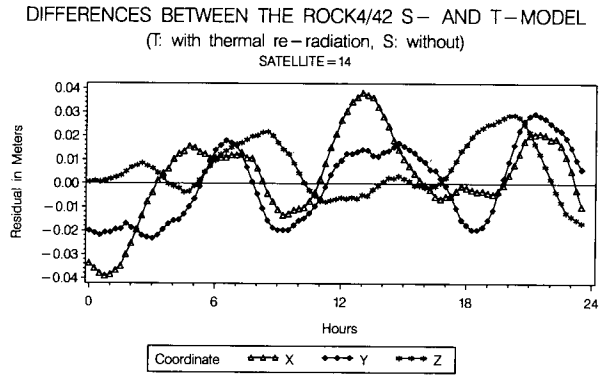


Figure 12: Differences between the Rock4/42 S- and T-model. A scaling parameter has been estimated for both models.

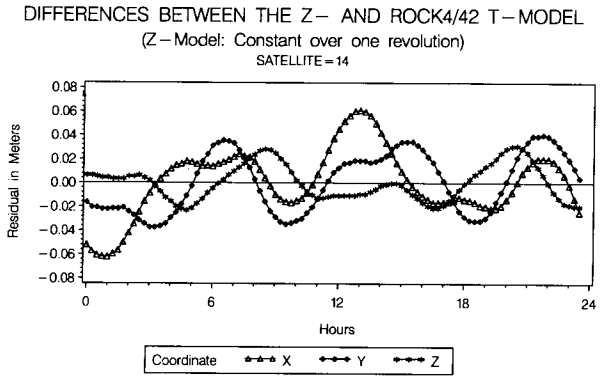


Figure 13: Differences between the Z- and Rock4/42 T-model. A scaling parameter has been estimated for both models.

The attitude control of the satellites is based on a feedback loop between solar sensors (on the panels) and momentum wheels. Although in theory the satellite's Y-axis should always be perpendicular to the direction Sun–satellite (to optimize the amount of energy collected by the solar panels), this is not true in practice. A small mis-orientation of the solar panels will cause a force acting in the direction of the Y-axis, called **y-bias**. Two possible causes are shown in Figures 14 and 15, taken from [*Fliegel*, 1993]).

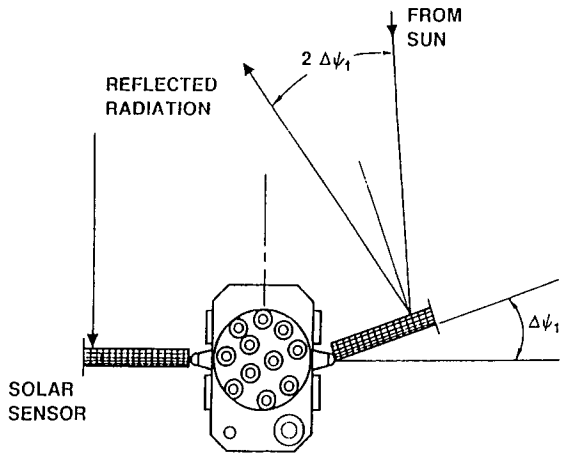


Figure 14: Possible cause of the y-bias (1) (from [*Fliegel*, 1993]).

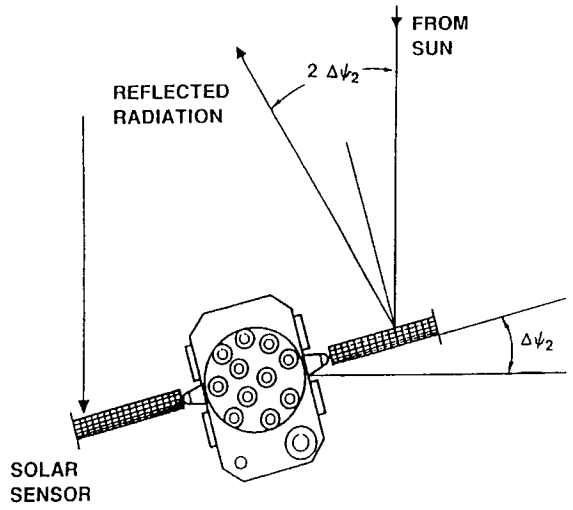


Figure 15: Possible cause of the y-bias (2) (from [*Fliegel*, 1993]).

The acceleration $\vec{a}_Y$ due to the y-bias may be written as:

$$\vec{a}_Y = \nu \cdot a_{Y0} \cdot \vec{e}_Y \tag{40}$$

where

ν ... Eclipse factor ($\nu = 1$ if satellite in sunlight, $\nu = 0$ if satellite in the Earth's shadow, $0 < \nu < 1$ if satellite in penumbra).
$\vec{e}_Y$ ... Unit vector in the direction of the solar panel axis
a_{Y0} ... Y-bias parameter

All the difficulties in modeling radiation pressure make it very clear, that at least two radiation pressure parameters — the direct radiation pressure coefficient (or a scaling factor of the Rock4/42 models) and the y-bias a_{Y0} have to be **solved for** for each satellite arc of one day or longer. More details on the estimation of such **dynamical** parameters will be given in Section 4.3 when discussing orbit determination. There we will also have a closer look at the values of radiation pressure parameters.

3.4 Eclipses

Because of their high altitude GPS satellites are almost permanently in the sunlight. Only if the direction Sun–satellite lies almost in one of the orbital planes, the satellites of this plane will pass through the Earth's shadow once per revolution. Such an "eclipse seasons" happens about twice per year for a specific orbital plane and typically lasts for a few weeks. One shadow passage has a duration of between 0 and 55 minutes. It is obvious that during the eclipse no solar radiation pressure is exerted on the satellite (eclipse factor $\nu = 0$, see e.g. Eqn. 39).

According to a simple **cylinder model** for the shadow of the Earth (see Figure 16) the eclipse factor ν in Eqns. 39 or 40 may be computed as:

$$\nu = \begin{cases} 0 \quad if & \cos\gamma = \dfrac{\vec{r} \cdot \vec{r}_s}{|\vec{r}| \cdot |\vec{r}_s|} < 0 \quad and \\ & h = |\vec{r}| \sqrt{1 - \cos^2\gamma} < a_E \\ 1 \quad else & \end{cases} \tag{41}$$

where

$\vec{r}$ ... Geocentric position vector of the satellite
$\vec{r}_s$ ... Geocentric position vector of the Sun
a_E ... Equatorial radius of the Earth

A more elaborate shadow model has been described in [*Landau*, 1988].

The y-axis or solar panel axis (see Figure 11) is oriented according to Sun sensors on the solar panels. As soon as the satellite enters the Earth's shadow, keeping the correct orientation of the y-axis becomes impossible. Since June 1994 (see [*Bar-Sever*, 1994]), the satellites are then rotating around the z-axis (pointing towards the center

of the Earth) with the maximum speed possible (about 0.1 deg/sec) during the shadow passage. The satellites then exits from the Earth's shadow with an almost **arbitrary orientation** of the solar panels. Until the satellite has regained its correct attitude, the mis-orientation of the solar panels causes orbit modeling problems.

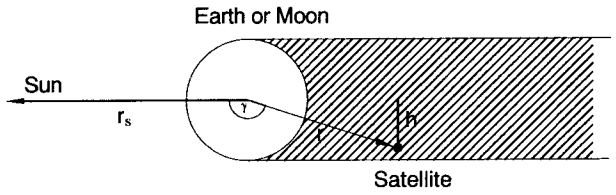


Figure 16: Simple cylinder model of the Earth's shadow.

The incorrect attitude of satellite PRN 28 during and after the eclipse can clearly be seen in the double-difference residuals (in narrow-lane cycles of 10 cm, epoch interval 30 sec) of a long baseline (Mas Palomas, Canary Islands - St. Johns, Canada) shown in Figures 17 and 18.

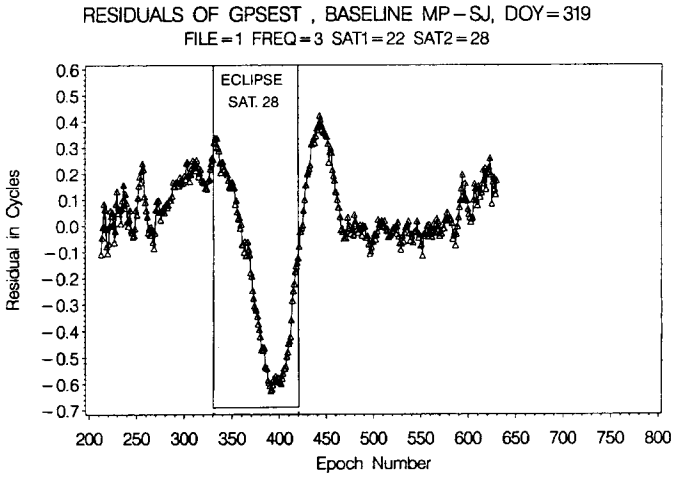


Figure 17: Mis-orientation of PRN 28 (deviation of the antenna phase center location from its nominal position) during and after the eclipse for day 319 of 1993, visible in the double-difference residuals of a long baseline.

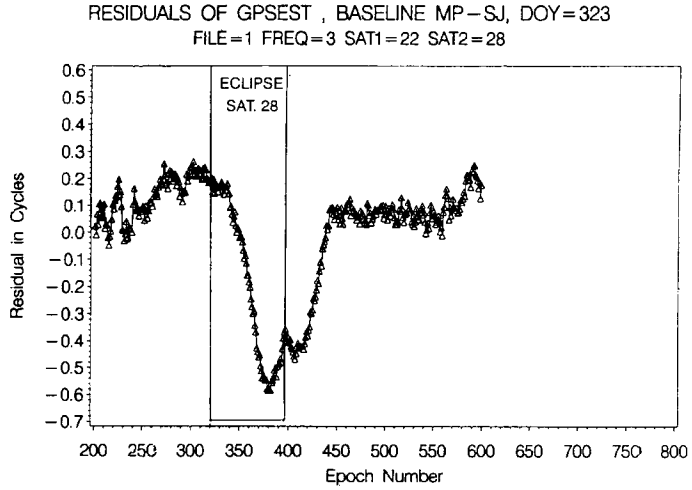


Figure 18: Mis-orientation of PRN 28 (deviation of the antenna phase center location from its nominal position) during and after the eclipse for day 323 of 1993, visible in the double-difference residuals of a long baseline.

The deviation of the satellite's orientation from the nominal attitude model can be seen in the residuals, because the satellite antenna phase center is not at the center of mass of the satellite, but offset by $(\Delta X, \Delta Y, \Delta Z) = (0.2100, 0.0000, 0.8540)$ meters and by $(0.2794, 0.0000, 1.0259)$ meters in the satellite-fixed frame for the Block I and Block II satellites, respectively.

Whereas in Figure 17 (day 319, 1993) the satellite keeps turning in the same direction after the end of the eclipse as during the eclipse, the direction of rotation is changing sign at the shadow exit four days later (day 323, 1993) shown in Figure 18.

To get a better idea of the orientation of the satellite during and shortly after the eclipse, we may estimate the satellite's antenna phase center location (relative to the nominal position). The resulting phase center variations in the X- and Y-direction, estimated in 6-minute intervals, are presented in Figures 19 and 20 for six consecutive days in 1993.

The Jet Propulsion Laboratory (JPL; see Section 6.1) routinely estimates the angular

velocities of the satellites, called yaw-rates, during eclipses [*Bar-Sever*, 1995].

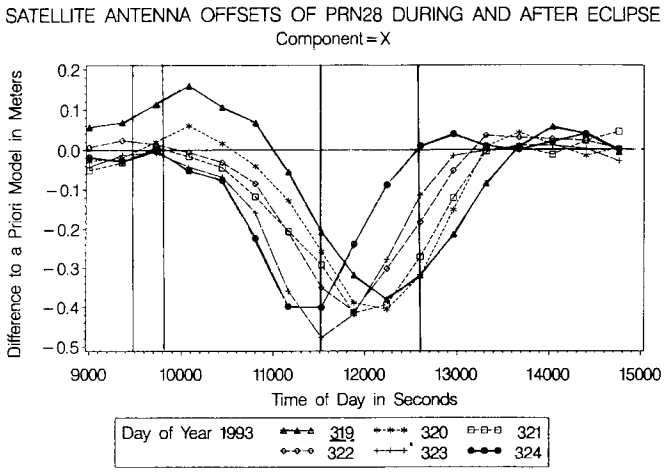


Figure 19: Antenna phase center offset in X-direction relative to the nominal attitude model estimated in 6-minute intervals.

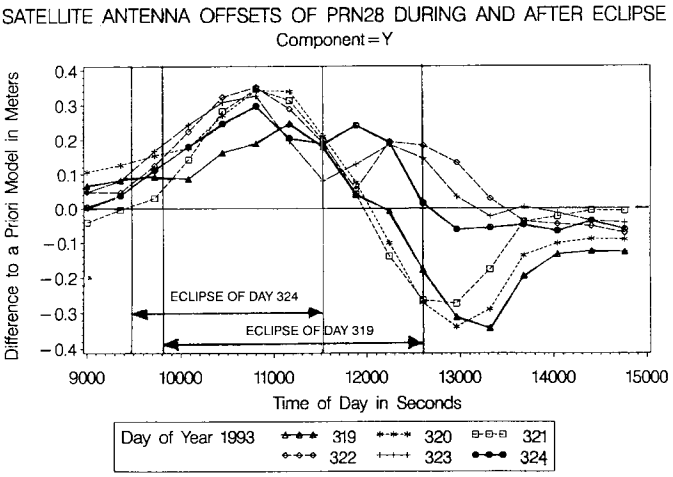


Figure 20: Antenna phase center offset in Y-direction relative to the nominal attitude model estimated in 6-minute intervals.

3.5 Other Perturbations

There are a few perturbation forces that are usually not modeled, because they are small compared to other unmodeled forces (momentum dumps, remaining radiation pressure effects, attitude problems, ...):

- Outgassing
- Albedo radiation pressure
- Ocean Tides
- Gravitational forces of the planets (e.g. Venus)
- Effects of general relativity

Outgassing:

The leaking of gas from a newly launched satellite is called **outgassing**. Estimates of the mass outgassed by a satellite vary from about 220 g in the first month after launch to about 10 g after several months. An effect that might be outgassing or a change in the reflectivity of the spacecraft's surfaces during the first few months in orbit can be seen in Figure 21, where the acceleration due to radiation pressure is changing very fast during the first days in orbit and only gradually becomes more stable.

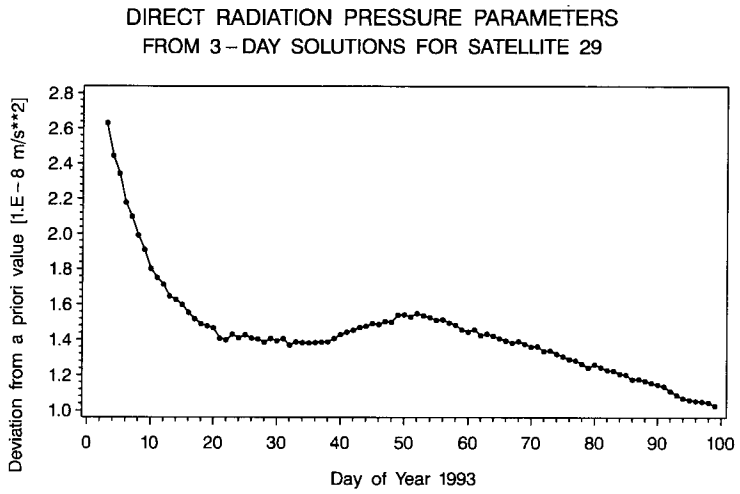


Figure 21: Estimated direct radiation pressure parameters for PRN 29 (from January 6 to March 19, 1993) shortly after the launch on December 18, 1992.

Earth's Albedo:

The Earth and its atmosphere reflect a large portion of the solar radiation received from the Sun back into space. The radiation pressure on the satellite due to the albedo of the Earth is very difficult to model (distribution of land, ocean, and clouds) and amounts to about 1–2 percent of the direct solar radiation pressure at the altitude of the GPS satellites or an acceleration of about $4 \cdot 10^{-10} m/s^2$ (see e.g. [*Rizos and Stolz*, 1985]).

Ocean Tides:

The changes in the gravity field due to ocean tides are difficult to model, too, since the ocean waves caused by Moon and Sun cannot propagate frictionless and also interact with the sea floor (shallow waters). The resulting acceleration on a GPS satellite is of the order of $5 \cdot 10^{-10} m/s^2$.

Gravitational Attraction by the Planets:

After the Moon and the Sun the planet Venus prövides the largest contribution, but the effect is very small (30 cm for a satellite arc of one week).

Effects of General Relativity:

The general relativistic perturbation due to the gravity field of the Earth may be found in [*Zhu et al.*, 1997]. The relativistic correction for the acceleration is about $3 \cdot 10^{-10} m/s^2$.

4 Orbit Determination

4.1 Statement of the Problem

Let us first re-state the equations of motion (see Eqns. 31, 32, and 33):

$$\ddot{\vec{r}} = -GM \cdot \frac{\vec{r}}{r^3} + \vec{a}(t, \vec{r}, \dot{\vec{r}}, q_1, q_2, \ldots, q_d) \tag{42}$$

$$\vec{r}(t_0) = \vec{r}(a, e, i, \Omega, \omega, T_p; t_0) = \vec{r}_0 \tag{43}$$

$$\dot{\vec{r}}(t_0) = \dot{\vec{r}}(a, e, i, \Omega, \omega, T_p; t_0) = \vec{v}_0 \tag{44}$$

where a, e, i, Ω, ω, T_p are the Keplerian elements at time t_0.

Orbit determination in its general sense is then the problem of determining the following $n = 6 + d$ unknown parameters p_i that define a unique (particular) solution of the above set of three equations:

$$\{p_1, p_2, \ldots, p_n\} = \{a, e, i, \Omega, \omega, T_p, q_1, q_2, \ldots, q_d\} \tag{45}$$

The six Keplerian elements at time t_0 define the **initial conditions** of the problem. Instead of the Keplerian elements it would be possible, too, to set up the components of the vectors $\vec{r}(t_0) = \vec{r}_0$ and $\dot{\vec{r}}(t_0) = \vec{v}_0$ as unknowns. The parameters $q_1, q_2, \ldots, q_d$ are the unknown **dynamical parameters** describing the force field. All the parameters p_i $(i = 1, 2, \ldots, n)$ have to be estimated using the GPS observations obtained from a global network of tracking stations (e.g. the global IGS network) in a certain time interval $[t_0, t_1]$. If we only consider the solution in this time interval, we speak of a **satellite arc** with an **arc length** $l = t_1 - t_0$.

When we have to determine the orbits of GPS satellites, we may consider most of the parameters $q_1, q_2, \ldots, q_d$ defining the force field to be known very accurately. The coefficients of the gravity field, e.g., are known with high precision from **Satellite Laser Ranging (SLR)** solutions. But it is also clear from the discussion of the various perturbing forces (Section 3), that it is not possible to assume **all** dynamical parameters to be known. At present at least **two dynamical parameters** have to be estimated (e.g. by the IGS analysis centers) for each GPS satellite when determining its orbit: the **direct radiation pressure parameter** (or a scaling parameter of the Rock4/42 model) and the **y-bias parameter** (see Section 3.3).

4.2 Principles of Orbit Determination

Let us now assume that we have an **a priori orbit** available already, e.g. derived from broadcast ephemerides. Else we would have to perform a so-called **first orbit determination** (having no information about the orbit except the actual observations) as e.g. developed by C. F. Gauss (1777–1855) to compute the orbits of minor planets (see [*Gauss*, 1809]). The a priori orbit $\vec{r}_0(t)$ **must** be a solution of the same equations of motion with the same set of parameters as those given in Eqns. 42, 43, and 44:

$$\ddot{\vec{r}}_0 = -GM \cdot \frac{\vec{r}_0}{r_0^3} + \vec{a}(t, \vec{r}_0, \dot{\vec{r}}_0, q_{10}, \ldots, q_{d0}) =: \vec{f} \tag{46}$$

$$\vec{r}_0(t_0) = \vec{r}(a_0, e_0, i_0, \Omega_0, \omega_0, T_{p0}; t_0) \tag{47}$$

$$\dot{\vec{r}}_0(t_0) = \dot{\vec{r}}(a_0, e_0, i_0, \Omega_0, \omega_0, T_{p0}; t_0) \tag{48}$$

where a_0,e_0,i_0,Ω_0,ω_0,T_{p0} and q_{i0} are the **a priori values** p_{i0} of the parameters p_i to be estimated.

We may now linearize the unknown orbit $\vec{r}(t)$ by developing it into a **Taylor series** which we truncate after the linear terms:

$$\vec{r}(t) = \vec{r}_0(t) + \sum_{i=1}^{n} \frac{\partial \vec{r}_0(t)}{\partial p_i} \cdot (p_i - p_{i0}) \tag{49}$$

We see that in this equation the unknown orbit $\vec{r}(t)$ is represented as a **linear function** of the unknown parameters p_i $(i = 1, 2, \ldots, n)$. That is all we need to set up a least squares algorithm. But how do we compute the partial derivatives in Eqn. 49 ? Let us

first introduce the following symbol for the partial derivative of the orbit with respect to one orbit parameter $p \in \{p_1, p_2, \ldots, p_n\}$:

$$\vec{z}(t) := \frac{\partial \vec{r}_0(t)}{\partial p} \tag{50}$$

Taking the first derivative of the equations of motion (Eqn. 46) with respect to the parameter p, we obtain a differential equation system for the partial derivatives $\vec{z}(t)$:

$$\ddot{\vec{z}} = A_0 \cdot \vec{z} + A_1 \cdot \dot{\vec{z}} + \vec{a}_p \tag{51}$$

where A_0 and A_1 are 3x3 matrices with elements defined by

$$A_{0,ik} = \frac{\partial f_i}{\partial r_{0,k}}, \qquad i,k = 1,2,3 \tag{52}$$

$$A_{1,ik} = \frac{\partial f_i}{\partial \dot{r}_{0,k}}, \qquad i,k = 1,2,3 \tag{53}$$

and

$$\vec{a}_p = \frac{\partial \vec{a}}{\partial p} \tag{54}$$

We used f_i here to denote the components of the vector function $\vec{f}$. All the partials have to be evaluated using the known a priori orbit $\vec{r}_0(t)$.

The Eqns. 51 are called the **variational equations** (one for each parameter p_i) belonging to the original equations of motions (Eqn. 46), also called the **primary equations** in this context. The initial conditions for the variational equations (Eqn. 51) may be obtained by taking the partial derivatives of the initial conditions (Eqn. 47) of the primary system with respect to the unknown parameter p:

$$\vec{z}(t_0) = \frac{\partial \vec{r}_0(t_0)}{\partial p}, \qquad \dot{\vec{z}}(t_0) = \frac{\partial \dot{\vec{r}}_0(t_0)}{\partial p} \tag{55}$$

For $p \in \{a, e, i, \Omega, \omega, T_p\}$ we have

$$\vec{a}_p = \vec{0}, \qquad \vec{z}(t_0) \neq \vec{0}, \qquad \dot{\vec{z}}(t_0) \neq \vec{0} \tag{56}$$

whereas for $p \in \{q_1, q_2, \ldots, q_d\}$ we have

$$\vec{a}_p \neq \vec{0}, \qquad \vec{z}(t_0) = \vec{0}, \qquad \dot{\vec{z}}(t_0) = \vec{0} \tag{57}$$

In summary we may say, that in an orbit estimation step we have to solve, in addition to the non-linear primary equations (Eqns. 46 and 47), one linear differential equation system (Eqns. 51 and 55) for each orbit parameter p_i to obtain the partial derivatives $\vec{z}(t)$ with respect to the orbit $\vec{r}(t)$. All these differential equation systems have to be solved using numerical integration methods (see Section 4.5).

4.3 Dynamical Parameters

Because the gravitational forces acting on GPS satellites are quite well-known (gravity field of Sun and Moon, the Earth's gravity field, tidal forces, ...) the only dynamical parameters q_i to be estimated in the orbit determination procedure belong to the solar radiation pressure model. The so-called **standard solar radiation pressure model**, used in GPS orbit determination since quite a long time, has the simple form:

$$\vec{a}_{rp} = \nu \cdot (\vec{a}_{ROCK} + a_{D0} \cdot \vec{e}_D + a_{Y0} \cdot \vec{e}_Y) \tag{58}$$

where

ν	...	Eclipse factor ($\nu = 1$ if satellite in sunlight, $\nu = 0$ if satellite in the Earth's shadow, $0 < \nu < 1$ if satellite in penumbra).
$\vec{a}_{ROCK}$	...	Acceleration according to the Rock4/42 models
a_{D0}	...	Direct radiation pressure parameter
a_{Y0}	...	Y-bias parameter
$\vec{e}_D$	...	Unit vector in the direction Sun–satellite
$\vec{e}_Y$	...	Unit vector in the direction of the solar panel axis

The unknown dynamical parameters in this model are the direct radiation pressure coefficient $q_1 = a_{D0}$ and the y-bias $q_2 = a_{Y0}$ (see Eqn. 40). When we compare Eqn. 58 with the more general model of Eqn. 39, we see that the parameter a_{D0} stands for

$$a_{D0} = P_s \cdot C_r \cdot \frac{A}{m} \cdot \frac{a_s^2}{|\vec{r} - \vec{r}_s|^2} \tag{59}$$

and

$$\vec{e}_D = \frac{\vec{r} - \vec{r}_s}{|\vec{r} - \vec{r}_s|} \tag{60}$$

It is clear that in the simple model according to Eqn. 58 no variations of the quantities on the right hand side of Eqn. 59 (e.g. the ratio $\frac{A}{m}$) are taken into account (a_{D0} is usually estimated as a constant over one satellite arc).

Let us illustrate the standard radiation pressure model by presenting the values for the direct radiation pressure coefficient a_{D0} and the y-bias a_{Y0} as they were estimated at the CODE Analysis Center of the IGS in Berne (see Section 6.1) from the global GPS data.

Figure 22 shows the estimates of the direct solar radiation pressure acceleration for PRN 19 — the direct radiation pressure acceleration given in the Rock4/42 a priori model has been added to the estimated values of a_{D0} to obtain the full direct acceleration — over a period of about two and a half years as computed by CODE. Estimates during eclipse seasons are not shown, because they are much noisier.

The mean acceleration due to solar radiation pressure is about $1 \cdot 10^{-7} m/s^2$. Because the variations due to the changing distance between the Sun and the Earth (ellipticity of the Earth's orbit) was not taken into account in Eqn. 58, we see the clear annual

period in the results. The maximum pressure is exerted in January (perihelion) and the minimum in June. The expected peak-to-peak variation is given by

$$\frac{max(a_{D0}) - min(a_{D0})}{1/2 \cdot (max(a_{D0}) + min(a_{D0}))} = \frac{1}{(1-e)^2} - \frac{1}{(1+e)^2} = 4 \cdot e \approx 0.067 \qquad (61)$$

After removing the annual signal there still remains a semi-annual variation (see Figure 23, solid line) in the series. The dots, finally, show the direct radiation pressure values after subtracting the best fitting annual and semi-annual amplitudes.

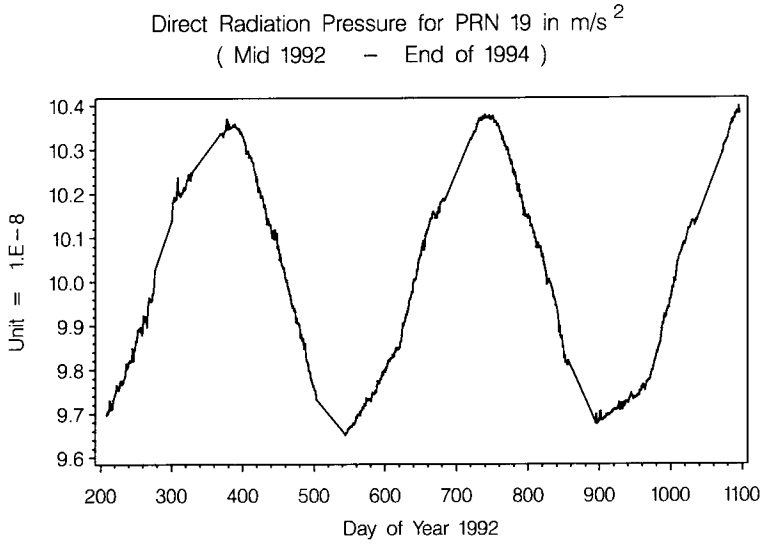


Figure 22: Direct radiation pressure values as estimated by the CODE Analysis Center over two and a half years.

The **mean values** of the direct radiation pressure parameters over a period of 2.5 years — after having removed the periodic variations — are plotted in Figure 24 for all GPS satellites. We see, that the different spacecraft shapes (Block I, Block II, and Block IIA; see Figure 4) cause different mean radiation pressure parameters. For satellites belonging to the same block the values are quite consistent. We should mention here that PRN 23 is an exception, because its solar panels were not correctly deployed.

The values of the y-biases, shown in Figure 25, are much smaller (by about a factor of 200) than the direct radiation pressure parameters. Similar to the direct radiation pressure parameters a semi-annual signature can be seen in the series. After mid 1994, there was a change in the attitude control of the satellites [*Bar-Sever*, 1994],which is probably responsible for the change in the behaviour of the y-bias values after this

epoch.

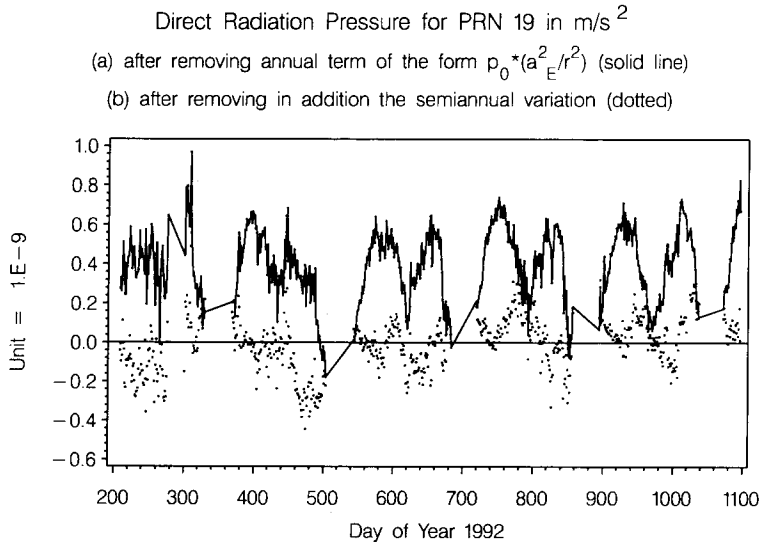


Figure 23: Direct radiation pressure parameters after removing an annual and a semi-annual variation.

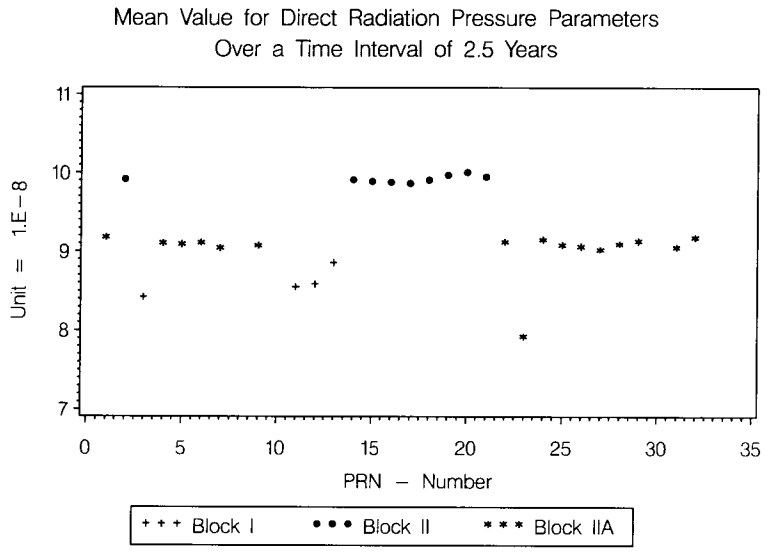


Figure 24: Mean Direct Radiation Pressure Values.

The y-bias values of all the satellites seem to be slightly negative (of the order of $-3 \cdot 10^{-10} m/s^2$).

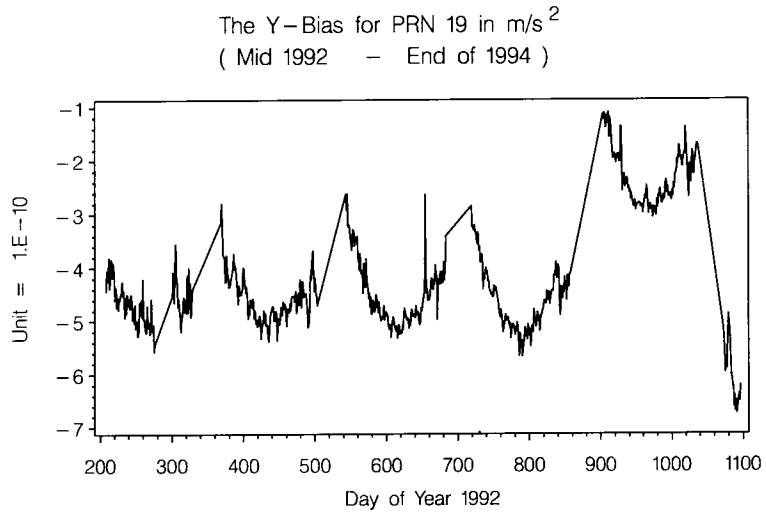


Figure 25: Y-bias values from 2.5 years of CODE solutions.

In view of the complicated shape of the GPS spacecrafts (see Figure 4) it was soon clear that the estimation of only two radiation pressure parameters (standard model) cannot be sufficient, especially for long arcs (longer than one day). An **extended radiation pressure model** has therefore been proposed by [*Beutler et al.*, 1994*b*]. This extended radiation pressure model, which is already used by a few IGS Analysis Centers, may be written in the following way:

$$\vec{a}_{rpr} = \nu \cdot (\vec{a}_{ROCK} + D(u) \cdot \vec{e}_D + Y(u) \cdot \vec{e}_Y + X(u) \cdot \vec{e}_X) \tag{62}$$

with

$$D(u) = a_{D0} + a_{DC} \cdot \cos(u) + a_{DS} \cdot \sin(u) \tag{63}$$

$$Y(u) = a_{Y0} + a_{YC} \cdot \cos(u) + a_{YS} \cdot \sin(u) \tag{64}$$

$$X(u) = a_{X0} + a_{XC} \cdot \cos(u) + a_{XS} \cdot \sin(u) \tag{65}$$

where a_{D0}, a_{DC}, a_{DS}, a_{Y0}, a_{YC}, a_{YS}, a_{X0}, a_{XC}, and a_{XS} are the nine parameters of the extended radiation pressure model. $\vec{e}_D$ is the unit vector Sun-satellite, $\vec{e}_Y$ the unit vector along the spacecraft's solar panel axis, and $\vec{e}_X$ is perpendicular to the two other unit vectors forming a right-hand system (see also Figure 26), and u ($\tilde{u}$ in Figure 26)

is the argument of latitude at time t for the satellite considered.

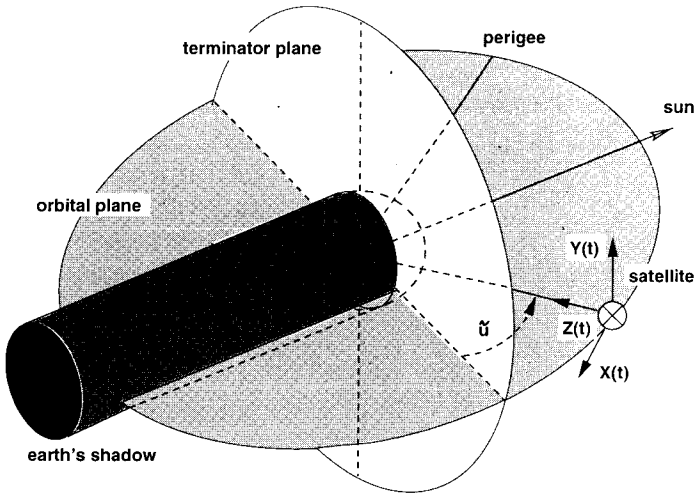


Figure 26: Definition of the coordinate system and angles used in the extended radiation pressure model.

This new model is also used by the IGS Analysis Center Coordinator for a quality check of the orbits submitted by the individual analysis centers before combining them into an official IGS orbit (see [*Beutler et al.*, 1993]). Figures 27 and 28 clearly show the superiority of the extended radiation pressure model, especially for long arcs. The extended model allows to fit a 2-week arc through the positions of a satellite (obtained from the daily CODE precise orbit files) with a position rms error of about 10 cm compared to an rms error of about one meter when using the standard model.

4.4 Stochastic Orbit Modeling

Even with the extended radiation pressure model not all modeling problems disappear. The model is of no help, if satellites are behaving in a strange way (panel mis-orientation, eclipses, momentum dumps, ...). Some orbit determination centers therefore allow for a **stochastic or pseudo-stochastic component** in their orbit model to absorb such effects. JPL, e.g., models the radiation pressure acceleration as a **stochastic process**, allowing the radiation pressure parameters to slowly change from epoch to epoch.

CODE and the GeoForschungsZentrum (GFZ) in Potsdam, on the other hand, allow for so-called **pseudo-stochastic pulses**, i.e. they estimate small velocity changes once per revolution. Both procedures result in an improvement of the satellite orbit quality by almost a factor of two. Figure 29 shows how the quality of the satellite orbits improve, when estimating pseudo-stochastic pulses once per revolution in the

radial (R), the along track (S), and the out-of-plane (W) orbit component for 1-day and 3-day solutions.

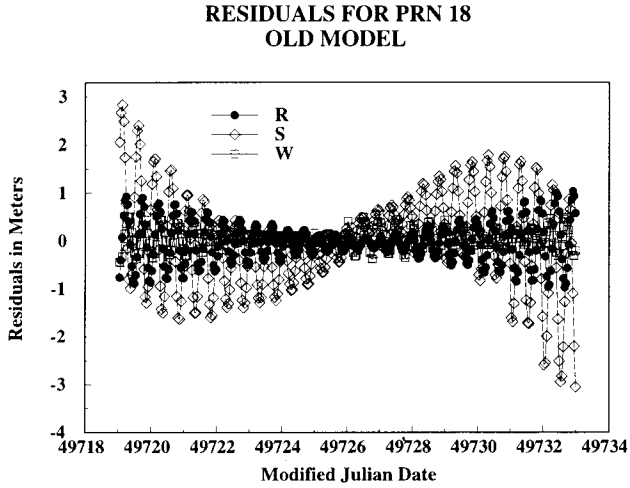


Figure 27: Residuals of an orbit fit with a 2-week arc through 2 weeks of orbit positions from CODE, estimating the two parameters of the standard radiation pressure model (Eqn. 58).

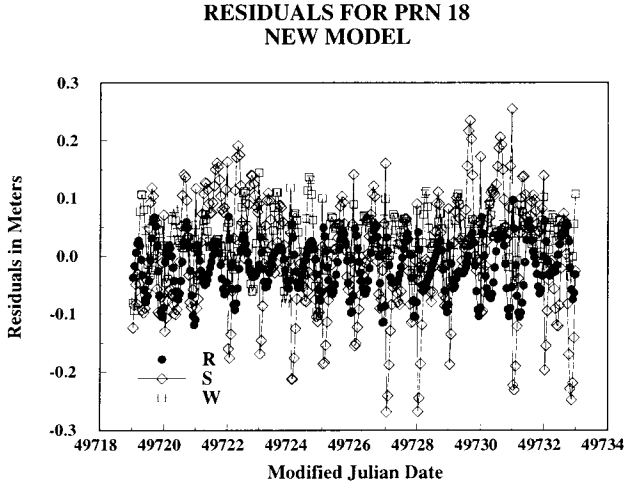


Figure 28: Residuals of an orbit fit with a 2-week arc through 2 weeks of orbit positions from CODE, estimating the nine parameters of the extended radiation pressure model (Eqn. 62).

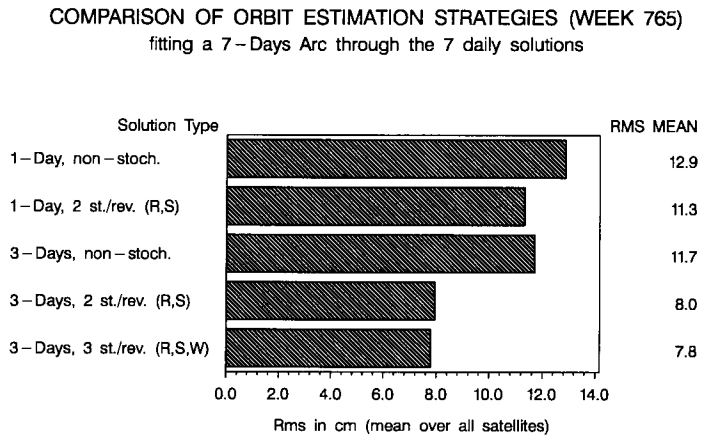


Figure 29: Improvement in the quality of the satellite orbits by estimating pseudo-stochastic pulses

The orbit differences between the JPL orbit and the CODE orbit for satellite PRN 7 during one day are plotted in Figure 30.

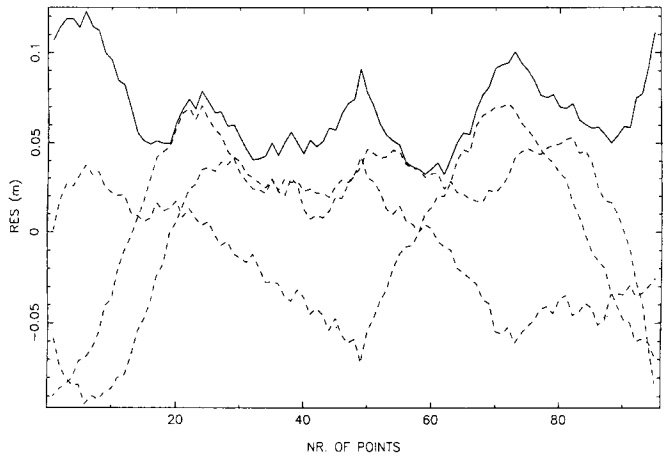


Figure 30: Differences between the JPL and CODE orbits over one day. The short period variations come from the stochastic model used by JPL. The pseudo-stochastic pulse estimated by CODE result in a discontinuity in the satellite's velocity in the middle of the plot.

The x-, y-, z-components of the orbit differences are drawn with dashed lines and the length of the vector difference with a solid line). Figure 29 shows the effect of the stochastic component in the JPL orbit modeling (short period variations) and the pseudo-stochastic pulse at the middle of the day (middle of the plot) in the CODE orbit.

4.5 Numerical Integration

Because numerical integration is an important part in the orbit determination process (in general only the two-body problem may be solved analytically), a short introduction into the principles of numerical integration will be given here. It is based on [*Beutler*, 1990], [*Rothacher*, 1992], and [*Beutler et al.*, 1996], which may also serve as further references.

Numerical integration can be understood as a part of approximation theory, where the "true" solution $\vec{r}(t)$ of the Eqns. 42, 43, and 44 is approximated in a certain time interval Δt by a polynomial $\vec{r}'(t)$ of degree q:

$$\vec{r}'(t) = \sum_{i=0}^{q} \vec{a}_i \cdot (t - t_0)^i \tag{66}$$

where

$\vec{a}_i$ ... 3-dimensional vector of polynomial coefficients of the fit
t_0 ... Origin of development in time

t_0 may be chosen arbitrarily. In the case of an initial value problem (initial conditions given at a starting epoch), t_0 is usually set to coincide with the starting epoch for which the initial values are given.

The problem of numerical integration now consists of the determination of the polynomial coefficients $\vec{a}_i$ $(i = 0, 1, 2, \ldots, q)$. This is achieved by setting up the following conditions:

1. The approximating function $\vec{r}'(t)$ has to satisfy the initial conditions of Eqns. 43 and 44. We thus get two linear (vectorial) condition equations for the unknowns $\vec{a}_0$ and $\vec{a}_1$:
$$\vec{a}_0 = \vec{r}_0, \qquad \vec{a}_1 = \vec{v}_0 \tag{67}$$

2. The approximating function $\vec{r}'(t)$ is then asked to satisfy Eqn. 42 at $q-1$ different time arguments t_j $(j = 1, 2, \ldots, q-1)$ in the integration interval Δt:
$$\ddot{\vec{r}}(t_j) = \sum_{i=2}^{q} i \cdot (i-1) \cdot \vec{a}_i \cdot (t_j - t_0)^{i-2} = \vec{f}(t_j, \vec{r}'(t_j), \dot{\vec{r}}'(t_j)) \tag{68}$$

This two sets of equations (Eqns. 67 and 68) give us a total of $2 + q - 1 = q + 1$ independent algebraic equations for the $q + 1$ unknowns $\vec{a}_i$.

The solution of a system of differential equations has thus been reduced to the solution of a system of non-linear algebraic equations. It can be shown that most of the classical methods of numerical integration are actually special approaches to solve this non-linear system of equations. This system may be solved by an iterative process, starting with approximate values for the coefficients $\vec{a}_i$ (e.g. from the Keplerian approximation) and then successively improving the coefficients from iteration I to iteration $I+1$ in the following way:

$$\ddot{\vec{r}}^{\,I+1}(t_j) = \sum_{i=2}^{q} i \cdot (i-1) \cdot \vec{a}_i^{\,I+1} \cdot (t_j - t_0)^{i-2} = \vec{f}(t_j, \vec{r}^{\,I}(t_j), \dot{\vec{r}}^{\,I}(t_j)) \qquad (69)$$

If a satellite orbit has to be integrated over a long time period, the entire time interval is subdivided into smaller intervals. A set of coefficients $\vec{a}_i$ $(i = 1, 2, \ldots, q)$ is then determined for each subinterval, where at the interval boundaries the satellite position and velocity are required to be continuous functions in time, thus connecting subsequent intervals.

The result of this procedure is not an ephemeris table of satellite coordinates, but one or more sets of coefficients $\vec{a}_i$, which allow us to compute $\vec{r}(t)$ or any of its derivatives with respect to time for any time within the integration interval Δt from the corresponding set of coefficients $\vec{a}_i$.

More details on the numerical integration method outlined here may be found in [*Beutler*, 1990].

In practice we normally use a polynomial degree of $q = 10$ and an integration interval of one hour for the numerical integration of a GPS satellite orbit. An orbit piece of one hour is then represented by one set of polynomial coefficients, e.g. $3 \cdot (q+1) = 33$ coefficients (3 dimensions).

5 Quality of GPS Orbits

The user of the GPS might, in general, not be as interested in the details of GPS orbit computation as the celestial mechanics, but would rather like to know, how accurate the available GPS orbits are. In the next sections we will discuss the various types of orbits available today and their accuracy.

The last part of this section will give an idea of the impact of orbit errors on the estimation of site coordinates.

5.1 GPS Orbit Types

Several different orbit sources are available today to the geodetic and geophysical GPS community. They vary considerably in quality and in the time delay, with which they become available. Table 4 gives an overview of these orbit types.

Orbit Type	Quality (m)	Delay of Availability
Broadcast Orbit	3.0 m	Real Time
DMA Precise Orbit	3.0 m	After 4-8 Weeks
IGS Predicted Orbit	.80 m	Real Time [1]
IGS Rapid Orbit	.10 m	After 24 Hours
IGS Final Orbit	.06 m	After 11 Days

[1] In a test phase since beginning of May 1996

Table 4: Accuracy and availability of GPS orbit types today.

The Broadcast Orbits

The Operational Control System (OCS) for the GPS became operational in September 1985. The Master Control Station, situated at Colorado Springs, is responsible for the overall satellite control, the determination, prediction, and dissemination of satellite ephemerides and clock information. The satellites are tracked by five monitor stations at Colorado Springs (USA), Hawaii (Pacific Ocean), Ascension Islands (Atlantic Ocean), Diego Garcia (Indian Ocean), and Kwajalein (Pacific Ocean, near Indonesia). The pseudorange data collected by these stations (during the most recent 12-24 hours) are used for an orbit improvement starting from an a priori orbit generated with the data of one week. The orbits thus obtained are then extrapolated for the next day (12-36 hours ahead) and for each 2-hour interval orbital elements are computed and uploaded to the GPS satellites. These broadcast orbits are then made available in the so-called **Broadcast Navigation Message** [*Dierendonck et al.*, 1978] in the form of pseudo-Keplerian elements and the time derivatives for some of these elements. Broadcast orbits refer to the WGS–84 (World Geodetic System–84). Because the broadcast orbits are available in real-time and directly transmitted from the satellites to the receivers, they are of great importance in all GPS applications, where high-precision orbits are not needed. As we see from Table 4, the broadcast orbits have a quality of about three meters. When comparing the broadcast orbits to the high precision orbits of the IGS, we should keep in mind that the broadcast orbits are **predicted orbits**. In view of the few sites used in their computation, they are of an amazing quality.

The DMA Precise Orbits

The so-called DMA precise orbits are generated by the Naval Surface Warfare Center (NSWC) together with the Defence Mapping Agency (DMA) approximately 4-8 weeks after the collection of the tracking data. In addition to the five Air Force monitor stations used for the Broadcast orbits, Quito (Ecuador), Buenos Aires (Argentina),

Smithfield (Australia), Hermitage (England), and Bahrain are part of the network processed. Approximately 8 days of data are fitted with one arc per satellite. The orbits thus derived have a quality of the order of a few meters and are not much used any longer since the availability of the IGS orbits.

The IGS Orbits

Since **June 21, 1992**, the start of the IGS Test Campaign, Analysis Centers of the International GPS Service for Geodynamics (IGS; see Section 6) are producing precise orbits from the data of the global IGS network (see Figure 37 in Section 6.1). Today there are seven such IGS Analysis Centers.

In contrast to the broadcast orbits, the IGS orbits are mainly based on the **phase observations** collected by the receivers in the global network. The IGS Analysis Centers submit their orbits in daily files, formatted according to the so-called **SP3-format**, a format originally defined by [*Remondi*, 1989]. Rectangular satellite positions (and satellite clock values) are given in the SP3-files in 15-minute intervals. The positions refer to a geocentric Earth-fixed reference frame, the ITRF (International Terrestrial Reference Frame) maintained by the IERS (International Earth Rotation Service).

Since the official start of the IGS on **January 1, 1994**, **IGS combined orbits** are generated by the Analysis Center Coordinator (Jan Kouba, NRC, Canada). A description of the combination algorithms may be found in [*Beutler et al.*, 1993] and [*Beutler et al.*, 1995]. These combined orbits are based on a weighted mean of the contributions of the individual Analysis Centers and proved to be very reliable. They are made available through the IGS Central Bureau Information System (CBIS; [*Gurtner and Liu*, 1995]) and through the global IGS Data Centers.

Today there are three different orbit products available from the IGS (see Table 4).

The **IGS final orbits** are the final and most accurate orbits of the IGS. They are available with a delay of 11 days, because the Analysis Centers wait for most of the data to arrive before they produce this final product.

The **IGS rapid orbits** are ready 24 hours after the observations. They are of almost as good a quality as the IGS final orbits and are also based on a combination of individual orbits delivered by the Analysis Centers (before 23 UT).

The **IGS predicted orbits**, finally, are not yet an official product of the IGS. During the present test phase three Analysis Centers (CODE, GFZ, and JPL) are generating orbit predictions of 24 and 48 hours. The orbits predicted ahead by 48 hours may be used **in real time** and are of encouraging quality. In the near future, the predicted orbits of the individual Analysis Centers will probably be combined in a similar way as the final and rapid orbits.

5.2 Quality of the IGS Orbits

Let us now try to assess the quality of the IGS final orbits. To get an idea of the accuracy of these orbits we may study:

- The weekly combination of the orbits of the individual Analysis Centers into one combined product.
- Check of the IGS orbits using Satellite Laser Ranging (SLR) data to the satellites PRN 5 and PRN 6, which carry retro-reflectors.
- The quality of the station coordinates resulting from a processing, where the IGS precise orbits are used. This aspect will be dealt with in Section 5.3.

The **seven** Analysis Centers are using **six** different software packages, each with its own strategies and models. Although all the centers are using the data of the same network (but not the same stations or number of stations), the deviation of the individual orbits from the combined IGS orbit should be a good indicator of the orbit consistency and quality. Figure 31 and Figure 32 show the development of the quality of the IGS final orbits over the time periods from September 1992 to December 1993 and from November 1993 to July 1996.

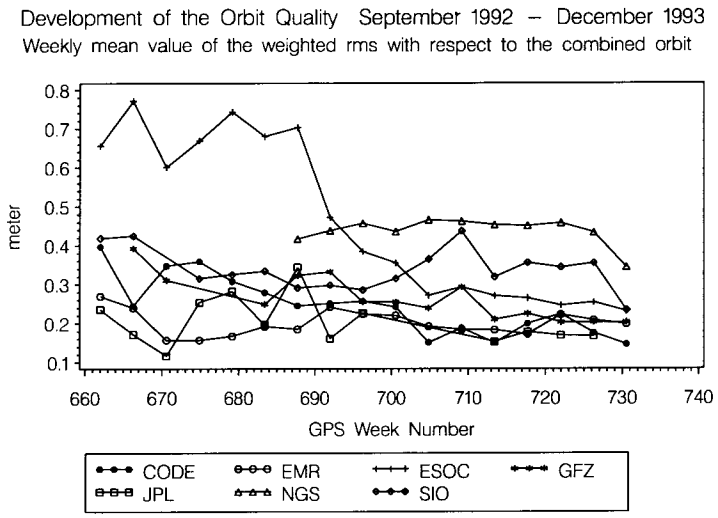


Figure 31: Development of the quality of the IGS final orbits (1).

Over the time span of 4 years we see a dramatic improvement in the orbit quality from about 0.5 meter in June 1992 to about 10 cm in June 1996. All the Analysis Centers are participating in this progress.

Two GPS satellites, **PRN 5 and PRN 6** (SVN 35 and SVN 36), are equipped with laser reflectors and can therefore be tracked by SLR stations. Accurate range measurements are thus available from an **independent space technique**. The residuals obtained when introducing the IGS orbits into an SLR adjustment may be used to

check the GPS orbit accuracy. In Figure 33 these residuals (for PRN 5 and PRN 6) are shown for almost the entire year 1995.

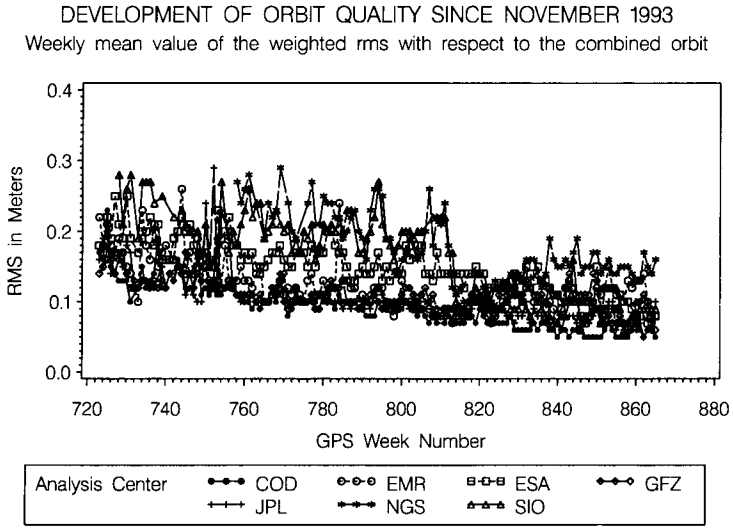


Figure 32: Development of the quality of the IGS final orbits (2).

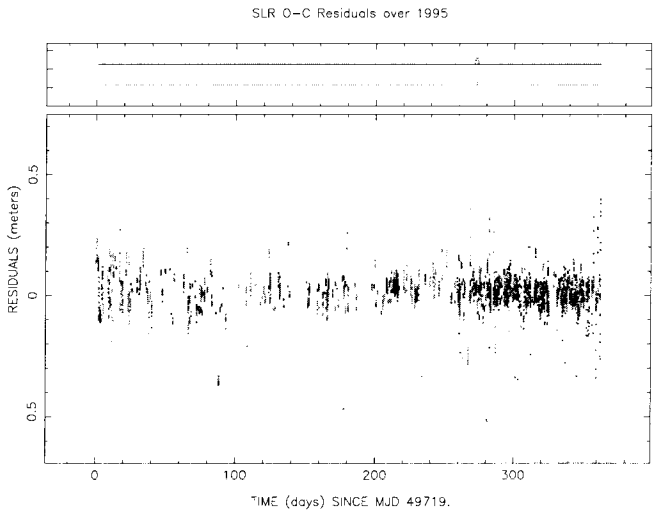


Figure 33: Residuals of SLR observations to the GPS satellites PRN 5 and PRN 6 for the year 1995 using the IGS final orbits as orbit information.

The overall scatter of the residuals is about 5 cm. When looking at the residuals in more detail, we detect that there are still systematic effects to be seen. An example of one day is given in Figure 34 showing the residuals of PRN 5 during its eclipse season. There is certainly still room for improvements in the orbit quality during these periods, that are difficult to model.

Table 5 summarizes the results obtained with SLR data when using the orbits generated by different Analysis Centers.

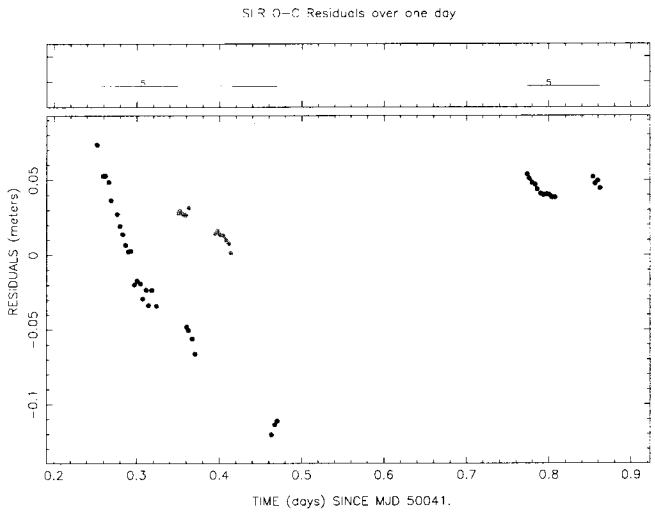


Figure 34: Residuals of SLR observations to the GPS satellite PRN 5 for one day during its eclipse season. The systematic residuals show the effect of the incorrectly modeled attitude of the satellite during eclipses.

Analysis Center	# Obs.	Mean Offset (m)	RMS (m) incl. Mean	RMS (m) w/o Mean
COD	2119	0.0090	0.0475	0.0467
EMR	2119	0.0060	0.0777	0.0775
GFZ	2089	-0.0198	0.0839	0.0815
IGS	2119	0.0033	0.0508	0.0507
JPL	2119	0.0012	0.0440	0.0440
SIO	2119	0.0029	0.0570	0.0569

Table 5: RMS of the residuals of SLR observations to the GPS satellites PRN 5 and PRN 6 when using GPS orbits of different Analysis Centers as "given".

Table 5 demonstrates that there is a good agreement between the orbit quality indicated by the orbit combination statistics and the residuals of the SLR measurements. The IGS orbits and the orbits of the best Analysis Centers are approaching the 5 cm rms level.

5.3 Impact of Orbit Errors on Site Coordinates

A very simple formula to estimate the effect of orbit errors on the determination of site coordinates, when processing GPS observations, has been given by [*Baueršíma*, 1983], Eqn. 84:

$$\frac{|\Delta\vec{b}|}{b} = \frac{|\Delta\vec{r}|}{r} \tag{70}$$

or

$$|\Delta\vec{b}| = \frac{b}{r} \cdot |\Delta\vec{r}| \tag{71}$$

where

$\Delta\vec{b}$ ... Error in the baseline vector due to the orbit errors
b ... Length of the baseline considered
$\Delta\vec{r}$ ... Error in the geocentric satellite position vectors
r ... Mean distance between the satellites and the receivers or about 25'000 km

The impact of orbit errors of various sizes on the site coordinates for different baseline lengths has been computed according to Eqn. 71 and put together in Table 6.

Orbit Error	Baseline Length	Baseline Error in ppm	Baseline Error in mm
25 m	1 km	1 ppm	1 mm
25 m	10 km	1 ppm	10 mm
25 m	100 km	1 ppm	100 mm
25 m	1000 km	1 ppm	1000 mm
2.5 m	10 km	.1 ppm	1 mm
2.5 m	100 km	.1 ppm	10 mm
2.5 m	1000 km	.1 ppm	100 mm
.25 m	100 km	.01 ppm	1 mm
.25 m	1000 km	.01 ppm	10 mm
.05 m	100 km	.002 ppm	- mm
.05 m	1000 km	.002 ppm	.5 mm

Table 6: Approximate errors in the baseline components according to Eqn. 71 for orbit errors between 5 cm and 25 m and baselines with a length between 1 km and 1000 km.

From Table 6 we may immediately conclude, that, in view of the accuracy of the IGS orbits of 5–10 cm, orbit errors are no issue any longer today: the impact of the residual orbit errors on the estimation of site coordinates — even for baselines as long as 1000 km — is below one millimeter. When using broadcast orbits with errors of about three meters, we will have no problem down to the millimeter as long as we process baselines below 10 km.

To demonstrate the difference between the processing of a long baseline (about 1000 km) with broadcast orbits and IGS final orbits, we include Figure 35 and Figure 36.

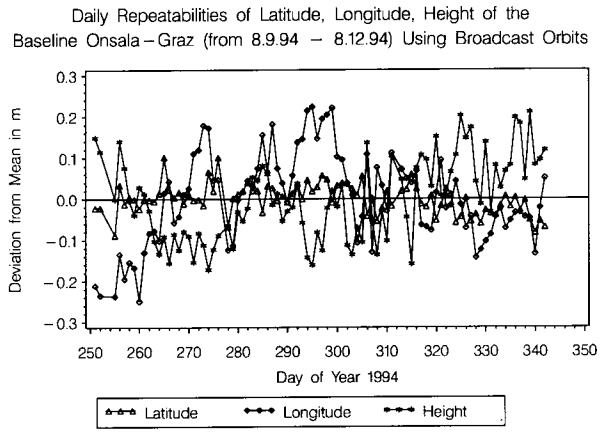


Figure 35: Daily baseline repeatability using broadcast orbits.

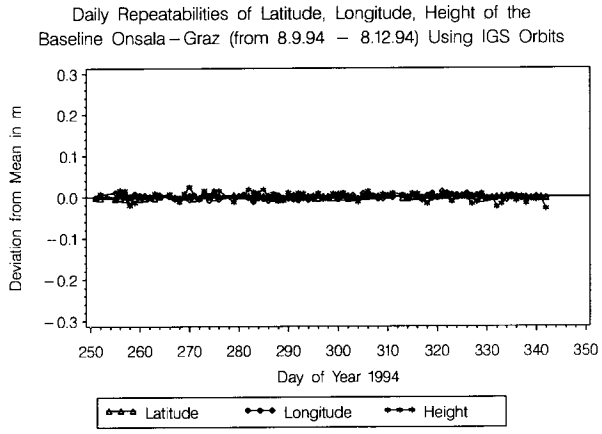


Figure 36: Daily baseline repeatability using IGS final orbits.

They show the repeatability of the estimated components of the baseline Onsala–Zimmerwald for a time interval of about three months using broadcast orbits (Figure 35) and IGS orbits (Figure 36), respectively, drawn with the same scale.

The horizontal repeatability, when making use of the IGS final orbits, is on the level of only a few millimeters. In height, the results are about a factor of three worse (around 1 cm), mainly due to the high correlation between the station heights and the troposphere zenith delay parameters, that have to be estimated to get more accurate heights.

As an example for the consistency of GPS station coordinate estimates resulting from long periods of data, Table 7 lists the coordinate residuals of a comparison of two GPS solutions from CODE for the European permanent stations, one solution computed from the GPS data of 1993, the other from the data of 1994.

Station	**North (mm)**	**East (mm)**	**Height (mm)**	**Mark**
NYAL 10317M001	-0.3	-1.2	2.1	
MASP 31303M001	-0.9	0.2	-0.1	
JOZE 12204M001	0.2	-2.3	11.2	M
BRUS 13101M004	-0.5	-0.2	-2.6	
GRAZ 11001M002	0.6	-0.3	2.7	
HERS 13212M007	0.6	0.1	-4.8	
KOSG 13504M003	0.4	0.3	-2.5	
MADR 13407S012	-0.9	1.0	4.5	
MATE 12734M008	1.0	0.0	-1.8	
TROM 10302M003	0.3	-0.2	1.8	
WETT 14201M009	0.5	-0.7	3.1	
ZIMM 14001M004	-0.5	-0.4	2.1	
ONSA 10402M004	-0.3	0.8	-3.4	
METS 10503S011	-0.1	0.5	-1.1	
RMS / Component	0.6	0.6	2.9	
Transformation				
Translation (mm)	-10.1	+6.8	-10.5	
Rotation (mas)	-0.2	+0.6	+0.3	
Scale f. (ppm)	.001	.001	.001	

Table 7: Helmert transformation between two yearly CODE solutions (solutions computed from one year of data in 1993 and one year of data in 1994, respectively) for the European IGS stations. The station marked with *M* has not been used to compute the RMS errors.

We see that the internal consistency of the GPS solutions has reached the level of about 1 mm in the horizontal position and about 3 mm in height. The level of consistency between the yearly solutions should not be taken for a measure of accuracy. There are still many effects (troposphere mapping function, antenna phase center variations, etc.) that may cause systematic biases in the GPS results not visible in the comparison above.

6 The International GPS Service for Geodynamics (IGS)

In this section we will give only a very short overview of the IGS, the International GPS Service for Geodynamics, because much material already exists documenting the main features of this international organization. Many details may, e.g., be found in [*Beutler et al.*, 1994*a*] and [*Beutler et al.*, 1994*c*] and much of the material presented here stems from these and similar publications. The IGS Central Bureau Information System (CBIS), containing all the relevant information about the various components of the IGS, is accessible through:

- World Wide Web: **http://igscb.jpl.nasa.gov**
- Anonymous FTP: **ftp://igscb.jpl.nasa.gov/igscb**
- E-Mail: **igscb@igscb.jpl.nasa.gov**

6.1 Objectives and Structure of the IGS

The IGS is an organization under the auspices of the IAG, the International Association of Geodesy. In the **IGS Terms of Reference**, contained e.g. in the IGS Colleague Directory, which is available at the IGS Central Bureau, we find the following objectives:

- IGS collects, archives, and distributes GPS observation data sets (...).
- These data sets are used by the IGS to generate (...):
  - High accuracy GPS satellite orbits
  - Earth rotation parameters
  - Coordinates and velocities of the IGS tracking stations
  - Satellite clock and station clock information
  - Atmosphere information
- The accuracies are sufficient to support current scientific objectives including the improvement of and the realization of the accessibility to the ITRF, the International Terrestrial Reference Frame.

The IGS structure consists of the following elements:

- A global **network** of tracking stations. At present, the network consists of more than 100 stations (see Figure 37).
- **Data Centers**, hierarchically divided into operational, regional, and global data centers. The three global data centers are: CDDIS, IGN, and SIO (see Table 8).
- **Analysis Centers** generating daily global products without interruption. Seven analysis centers are contributing at present: CODE, EMR, ESA, GFZ, JPL, NGS, and SIO (see Table 8).
- An **Analysis Center Coordinator**, a position presently occupied by Jan Kouba, Natural Resources, Canada.
- The **Central Bureau** at JPL with Ruth Neilan as director.
- The international **IGS Governing Board** consisting of 15 members with Gerhard Beutler as chairman.

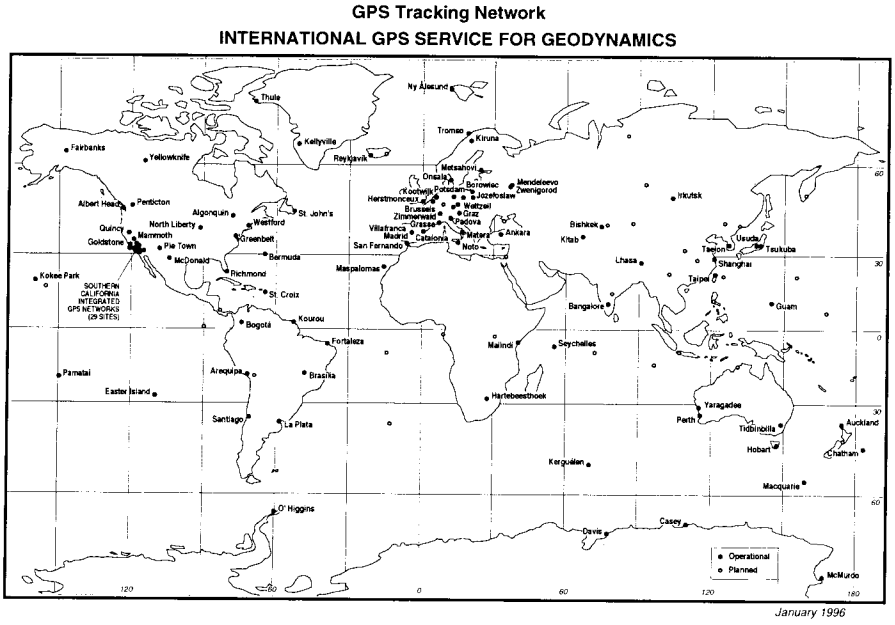


Figure 37: The IGS Network.

Table 8 summarizes some of the IGS institutions. A complete list of the more than 100 institutions participating in the IGS may be found at the CBIS.

Abbreviation	Institution	Function
CODE	at Astronomical Institute, Uni Berne Collaboration of AIUB, IfAG, IGN, and Swiss Federal Office of Topography	A
NRCan	Natural Resources, Canada (former EMR)	A
ESOC	European Space Agency, Germany	A
GFZ	Geoforschungszentrum, Germany	A
JPL	Jet Propulsion Laboratory, USA	A
NOAA	Nat. Oceanic and Atmosph. Adm., USA	A
SIO	Scripps Inst. of Oceanography, USA	A
UTX	University of Texas at Austin,USA	A
CDDIS	Goddard Space Flight Center, USA	D
IGN	Institut Géographique National, France	D
SIO	Scripps Institution of Oceanography, USA	D
OSU	Ohio State University, USA	C
NRCan	Natural Resources, Canada (former EMR)	C
JPL	Jet Propulsion Laboratory, USA	B

A : Analysis Center, D : Global Data Center, C : Coordinator, B : Central Bureau

Table 8: Some IGS institutions and their function.

A short history of the major events in the development of the IGS is given in Table 9.

Date	Event
Aug 89-Feb 90	*IAG General Meeting* in Edinburgh. Original Ideas by I.I. Mueller, G. Mader, W.G. Melbourne, B. Minster, and R.E. Neilan.
21-Jun-92	Start of *1992 IGS Test Campaign.*
27-Jul-92	Start of *Epoch-92* (2 weeks duration).
23-Sep-92	Official end of test campaign; data collection and processing continue.
01-Nov-92	Start of *IGS PILOT Service.* Regular weekly orbit comparisons performed by *IGS Analysis Center Coordinator.*
01-Jan-94	*Start of official IGS.* Production of combined IGS Orbits, Central Bureau Information System (CBIS) installed.
30-Jun-96	Major Review of Processing (Use of ITRF94, subdaily ERP terms, start of activities for new AACs, compliance to IERS Conventions 1996)

Table 9: Chronicle of important IGS events 1989–1996.

6.2 IGS Products and Quality

The IGS is generating quite a large palette of different products for the GPS user community. Table 10 contains a list of all the major products, with what delay they become available and what accuracy they have.

Product	Availability after	Precision
Satellite orbits		
Predicted	Real-time	50 cm
Rapid	1 day	10 cm
Final	10-14 days	6 cm
Satellite clocks		
Predicted	Real-time	150 ns
Rapid	1 day	0.5 ns
Final	10-14 days	0.3 ns
Pole (x- und y-component)		
Rapid	1 day	0.2 mas
Final	10-14 days	0.1 mas
Length of day (LOD)		
Rapid	1 day	60 ms/Tag
Final	10-14 days	30 ms/Tag
Station coordinates		
Weekly	4 weeks	3-5 mm

Table 10: The major products of the IGS, their availability and precision (from Jan Kouba, NRC).

Apart from these well-known products, some IGS Analysis Centers also extract information about the atmosphere from the global network. Figure 38 shows – as an example of the activities of the IGS in the field of global ionosphere modeling – an **single layer ionosphere model** produced by CODE. The density of the free electrons in the infinitesimal, single layer is developed into a series of spherical harmonics functions. ESA, JPL, UNB (University of New Brunswick), and DLR in Neustrelitz are producing ionosphere models, too. First attempts have been made to compare and combine the models of the individual groups.

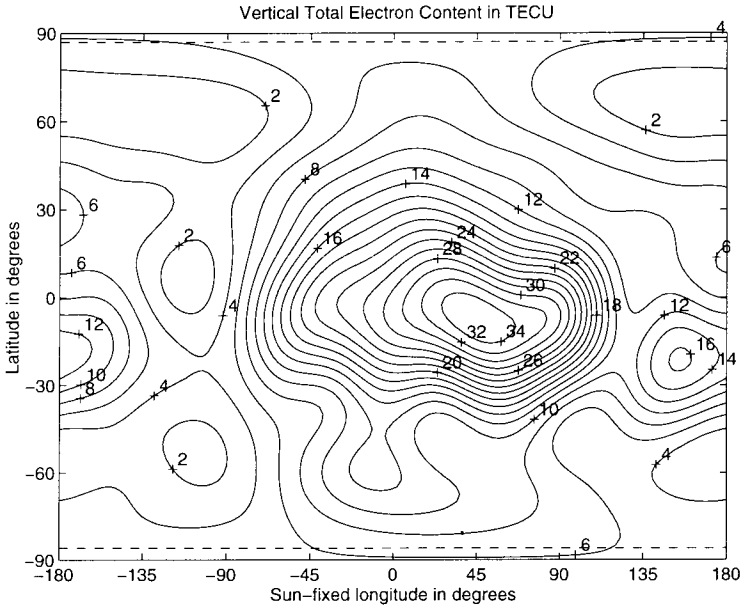


Figure 38: Global Ionosphere Model for Day 073 of Year 1996 from CODE.

All IGS Analysis Centers have to estimate tropospheric zenith delays to obtain good results for their global products. Some Analysis Centers use stochastic models (based on Kalman filters), others just solve for station- and time-specific troposphere parameters. [*Gendt and Beutler*, 1995] showed, that the consistency of troposphere estimates stemming from different Analysis Centers is relatively high. It is thus in principle possible to extract on a routine basis the precipitable water content for the entire IGS network with a high temporal resolution (two hours or finer). To subtract the dry troposphere delay from the total delay (to obtain the wet delay and subsequently the integrated precipitable water vapor (IPWV)), it is necessary, however, that high precision barometers are deployed in the IGS network. These issues are presently discussed within the IGS.

This small summary shows, that there are many very interesting developments going on in the IGS. It is clear, that IGS orbits allow to perform regional GPS analyses of highest accuracy without further orbit refinement. The results refer to the best available realization of the ITRF. The IGS pole positions (x- and y-component) are vital contributions to the earth rotation parameter series of the IERS (Bulletin A and B) today. Through the annual solutions of the IGS Analysis Centers the IGS also contributes to the realization and densification of the ITRF, a reference frame of most importance in the present interconnected world.

References

Anodina, T. (1988), The GLONASS System Technical Characteristics and Performance, International Civil Aviation Organization Special Committee on Future Air Navigation Systems (FANS), 4th Meeting, Montreal, May 2–20, FANS/4–WP/75.

Bar-Sever, Yoaz E. (1994), Improvement to the GPS Attitude Control Subsystem Enables Predictable Attitude During Eclipse Seasons, IGS Mail No. 591, IGS Central Bureau Information System.

Bar-Sever, Yoaz E. (1995), A new Model for GPS Yaw Attitude, in *Special Topics and New Directions*, edited by G Gendt and G. Dick, GeoForschungsZentrum, Potsdam, Germany, May 15–18 1995.

Baueršíma, I. (1983), NAVSTAR/Global Positioning System (GPS), II., Mitteilungen der Satelliten-Beobachtungsstation Zimmerwald, No. 10, Astronomical Institute, University of Berne.

Beutler, G. (1990), Numerische Integration gewöhnlicher Differentialgleichungenssysteme: Prinzipien und Algorithmen, Mitteilungen der Satelliten-Beobachtungsstation Zimmerwald, No. 23, Astronomical Institute, University of Berne.

Beutler, G., W. Gurtner, I. Bauersima, and R. Langley (1985), Modeling and Estimating the Orbits of GPS Satellites, in *Proceedings of the 1st International Symposium on Precise Positioning with the Global Positioning System, April 15–19*, vol. 1, edited by Clyde Goad, pp. 99–111, U.S. Department of Commerce, Rockville, Maryland.

Beutler, G., J. Kouba, and T. A. Springer (1993), Combining the Orbits of the IGS Processing Centers, in *Proceedings of the IGS Analysis Center Workshop*, edited by J. Kouba, pp. 20–56, Ottawa, Canada, October 12–14 1993.

Beutler, G., I. I. Mueller, and R. E. Neilan (1994*a*), The International GPS Service for Geodynamics (IGS): Development and Start of Official Service on January 1, 1994, *Bulletin Géodésique*, *68*(1), 39–70.

Beutler, G., E. Brockmann, W. Gurtner, U. Hugentobler, L. Mervart, and M. Rothacher (1994*b*), Extended Orbit Modeling Techniques at the CODE Processing Center of the International GPS Service for Geodynamics (IGS): Theory and Initial Results, *Manuscripta Geodaetica*, *19*, 367–386, April 1994.

Beutler, G., I. I. Mueller, R. E. Neilan, and R. Weber (1994*c*), IGS – Der Internationale GPS-Dienst für Geodynamik, *Zeitung für Vermessungswesen (DVW)*, *119*(5), 221–232, May 1994.

Beutler, G., J. Kouba, and T. A. Springer (1995), Combining the Orbits of the IGS Processing Centers, *Bulletin Géodésique*, *69*(4), 200–222.

Beutler, G., R. Weber, U. Hugentobler, M. Rothacher, and A. Verdun (1996), GPS Satellite Orbits, in *Lecture Notes International School "GPS for Geodesy"*, Springer-Verlag, Delft, The Netherlands.

Danby, J. M. A. (1989), *Fundamentals of Celestial Mechanics*, Willmann–Bell Inc., Second Edition, Second Printing, ISBN 0–943396–20–4, Richmond, VA.

Dierendonck, A. Van, S. Russel, E. Kopitzke, and M. Birnbaum (1978), The GPS Navigation Message, *Navigation: Journal of the Institute of Navigation*, *25*(2), 147–165.

Fliegel, H. F. (1993), Indirect Solar Pressure: Non-Equilibrium Force Models for GPS Satellites, presented at the American Geophysical Union Spring Meeting, Baltimore, Maryland, May 1993.

Fliegel, H. F., T. E. Gallini, and E. R. Swift (1992), Global Positioning System Radiation

Force Model for Geodetic Applications, *Geophysical Research Letters*, *97*(B1), 559–568.

Gauss, C. F. (1809), *Theoria motus corporum coelestium in sectionibus conicis solem ambientum*, Hamburgi, Perthes & Besser.

Gendt, G., and G. Beutler (1995), Consistency in the Troposphere Estimations using the IGS Network, in *IGS Workshop Proceedings on Special Topics and New Directions*, edited by G. Gendt and G. Dick, pp. 152–166, GeoForschungsZentrum, Potsdam, Germany, May 15–18 1995.

Gurtner, W., and R. Liu (1995), The Central Bureau Information System, in *IGS 1994 Annual Report*, edited by J.F. Zumberge *et al.*, pp. 43–57, IGS Central Bureau, Jet Propulsion Laboratory, Pasadena, California U.S.A., September 1995.

Hugentobler, U. (1995), Resonances for High Altitude Satellites, University Bern, University Press, in preparation.

Kepler, J. (1609), *Astronomia nova de motibus stellae Martis ex observationibus Tychonis Brahe*, Pragae.

Kepler, J. (1619), *Harmonices mundi libri V*, Lincii.

Landau, H. (1988), Zur Nutzung des Global Positioning Systems in Geodäsie und Geodynamik: Modellbildung, Software-Entwicklung und Analyse, PhD Thesis, Studiengang Vermessungswesen, Universität der Bundeswehr München, Neubiberg.

McCarthy, D.D. (1992), IERS Standards (1992), *IERS Technical Note 13*, Observatoire de Paris, Paris, July 1992.

Milliken, R. J., and C. J. Zoller (1980), Principle of Operation of NAVSTAR and System Characteristics, Global Positioning System, papers published in Navigation, reprinted by the Institute of Navigation, Washington D.C.

Newton, I. (1687), *Philosophiae naturalis principia mathematica*, Joseph Streater, Londini.

Remondi, B.W. (1989), Extending the National Geodetic Survey Standard GPS Orbit Formats, *Technical Report NOS 133 NGS 46*, NOAA, USA.

Rizos, C., and A. Stolz (1985), Force Modelling for GPS Satellite Orbits, in *Proceedings of the First International Symposium on Precise Positioning with the Global Positioning System*, vol. 1, pp. 87–98, Rockville, MD.

Rothacher, M. (1992), *Orbits of Satellite Systems in Space Geodesy*, Schweizerischen Geodätischen Kommission, Institut für Geodäsie und Photogrammetrie, Eidg. Technische Hochscule Zürich, Zürich.

Zhu, S. Y., E. Groten, R. S. Pan, H. J. Yan, Z. Y. Cheng, W. Y. Zhu, C. Huang, and M. Yao (1997), Motion of Satellites – the Choice of Reference Frames, Astrophysical Space Science Series, The Few Body Problem.

Modeling of the Earth Atmosphere in Space Geodetic Applications

Jan M. Johansson *

Introduction

GPS signals start traveling on their propagation path from a satellite to a receiver through what we call space, which essentially is a vacuum, before entering the Earth's atmosphere. The path distortions which radio signals undergo when traversing the atmosphere of the earth introduce uncertainties in the time of arrival of the signal due to bending and retardation along the propagation path. The Earth's atmosphere is usually divided into different layers dependent on the temperature variations as a function of the altitude (see Figure 1). The thermosphere and the mesosphere are the first atmospheric layers encountered relevant to the propagation of the signals. In the mesosphere the temperature is decreasing with increasing altitude up to a minimum of about -90° C at an altitude of about 80-90 kilometers. This minimum is called the mesopause. Above the mesopause the temperature is increasing again to about 1500–3000 K at an altitude of 500 kilometers. In this part of the atmosphere the air is very thin and the pressure is extremely low. Electromagnetic radiation and emission from the sun have the ability to ionize gases to produce free electrons and ions. This layer has a large impact on radio frequencies and is normally referred to as the ionosphere. Because the ionosphere is a dispersive medium for radio waves, which means that the propagation velocity depends on the frequency of the waves, one can almost completely eliminate the ionospheric propagation effects on GPS signals by simultaneously using two different carrier frequencies. After passing through the ionosphere, GPS signals travel through the stratosphere and

* *Onsala Space Observatory, Chalmers University of Technology, S-439 92 ONSALA, Sweden, phone +46 31 7725500, fax +46 31 7725590, e-mail: jmj@oso.chalmers.se*

the troposphere (Figure 1). These layers of the atmosphere are nondispersive at radio frequencies below about 30 GHz. In the lower stratosphere the temperature is constant with increasing altitude however in the upper stratosphere it starts again to increase reaching its maximum at the stratopause at an altitude of about 50 kilometers. The reason for the increase in the temperature within this layer is the natural presence of ozone. The ozone absorbs the ultraviolet emission from the sun leading to an change in the temperature. The amount of ozone in the stratosphere is small, only about 0.0001 percent of volume, but the absorption is tremendous. Most ozone is located in the interval 15–25 kilometers but the ozone at an altitude of about 50 kilometers is most important for the absorption processes. The small variations in temperature in the stratosphere prevents from circulations of the air in that layer. Pollution is consequently only removed over substantial periods of time. In the troposphere, which is the lowest part of the Earth's atmosphere, temperature decreases with an increase in altitude. Almost all of the activity normally referred to as "weather" takes place here. The decrease in temperature is about 0.6–0.7° C/100 meters. The thickness of the troposphere is not the same everywhere. It extends to a height of less than 9 kilometers over the poles and exceeds 16 kilometers over the equator. Even if the stratosphere extends from the upper boundary of the troposphere, called the tropopause, to a height of about 50 kilometers the bulk of the neutral atmosphere lies within the troposphere. The whole neutral atmosphere is therefore often referred to as the "troposphere".

The nondispersive nature of the troposphere delays the arrival of radio frequencies such as the carrier phase and the carrier modulation of both the L1 and L2 by the same amount. We cannot directly measure the tropospheric delay as we can ionospheric delay using the GPS signals themselves. Instead we must resort to modeling. The delay in the neutral atmosphere which is often considered to be the sum of a "dry" component, about 2.3 m in the zenith direction at sea-level, caused by induced dipoles mainly in N_2 and O_2, and a "wet" component caused by the permanent dipole moment of water vapor. Even though the wet term is smaller, it is highly variable, and can range from less than 1 cm to 40 cm or more in the zenith direction. Because of the difficulty in predicting line-of-sight water-vapor content from surface measurements, *e.g.*, a remote sensing approach may be used. In the following paragraph we discuss how to estimate or eliminate the effects of all parts of the atmospheric propagation delay.

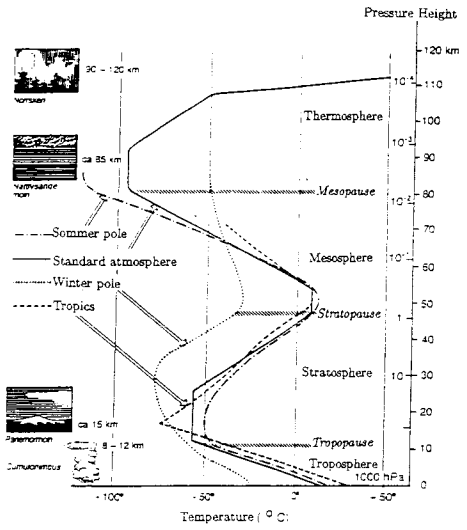


Figure 1: Temperature variations as a function of height for a standard atmosphere. Also shown are three different profiles, namely from a tropical climate and polar climate at winter and summer, respectively.

Definition of the excess propagation path.

Let us assume that we know the refractive index, n, in a three-dimensional atmosphere. The electrical path length L of a signal propagating along S is defined as

$$L = \int_S n\, ds \tag{1}$$

The path S is determined from the index of refraction in the atmosphere using Fermat's Principle, to wit: the signal will propagate along the path that gives the minimum value of L. The geometry is illustrated in Figure 2. The geometrical straight line distance, G, through the atmosphere is always shorter than the path S of the propagated signal. However, the electrical path length of the signal propagating along G is *longer* than that for the signal propagating along S. The reason is that it would travel a longer distance with lower velocity close to the surface of the earth where the index of refraction is largest. The difference between the electrical path length and the geometrical straight line distance is called "excess propagation path", "path delay", or simply "delay":

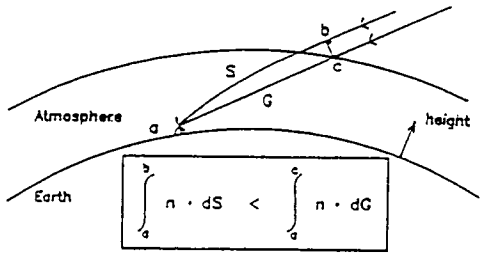


Figure 2: The signal from the GPS satellite will propagate to the antenna at the Earth's surface along the path S instead of the geometrically shorter path G, due to the decreasing refractive index with height in the atmosphere. The path S is determined using Fermat's Principle [*Elgered*, 1993].

$$\Delta L = \int_S n \, ds - G \tag{2}$$

We may rewrite this expression as

$$\Delta L = \int_S (n-1) \, ds + S - G \tag{3}$$

where $S = \int_S ds$. The $(S - G)$ term is often referred to as the "geometric delay" or the delay due to bending, denoted ΔL_g.

$$\Delta L_g \equiv S - G \tag{4}$$

In the case of a horizontally stratified atmosphere, we note that the two paths (S and G) are identical in the zenith direction and hence the geometric delay becomes equal to zero at this angle. The geometric delay is typically 3 cm at an elevation angle of 10° and 10 cm at 5°.

Alternatively, we may expand G in (2) to obtain

$$\Delta L = \int_G (n-1) \, ds - \int_G n \, ds + \int_S n \, ds \tag{5}$$

The corresponding correction in (5), where the integral of (n-1) is estimated along the straight line G instead of the curved path S, produces a slightly different definition for the "geometric delay":

$$\Delta L'_g \equiv \int_S n \, ds - \int_G n \, ds \tag{6}$$

The "geometric delays" in (4) and (6), corresponding to calculating ΔL according to (3) and (5), are different and care should be taken since the literature available on this subject is not consistent. The two corrections, ΔL_g and $\Delta L'_g$, are of approximately the same size but have different signs. In the following text we will use the delay and the corresponding geometric delay according to (3).

The Neutral Atmosphere.

In the zenith direction the tropospheric time delay results in an increase in measured apparent range of about 2.4 meters. The delay grows with increasing zenith angle and reaches 9.3 meters for a zenith angle of 75°. The zenith delay can be predicted easily to an accuracy better than 20 centimeters which means that tropospheric propagation effects are not a concern in low accuracy positioning with GPS. For positioning using the carrie-phase observations the tropospheric effect may be a severe limitation especially in the vertical component. An important rule of thumb is that an error of 1 centimeter in modeling of the tropospheric zenith delay can result in a vertical position error of about 3 centimeters. In this section we will discuss the physics behind the delays caused by the neutral atmosphere and define terms fundamental to this subject.

The delay of radio waves propagating through the neutral atmosphere

The propagation delay of the neutral atmosphere is often considered to be the sum of a "dry" component caused by induced dipoles in the molecules, and a "wet" component caused by the permanent dipole moment of water vapor. The dry term is normally about 2.3 m in the zenith direction at sea level. It can be accurately estimated in the zenith direction using surface measurements of the pressure at the ground. An uncertainty of 1 mbar in the ground pressure implies an uncertainty of approximately 2 mm in the dry zenith delay. In order to model its elevation dependence other surface measurements, such as the ground temperature, are often used (see, *e.g.*, *Hopfield*, [1969]; *Davis et al.*, [1985]; and *Lanyi*, [1984]).

The wet term is smaller and range from less than 1 cm to 40 cm or more in the zenith direction depending on the climate. The problem with the wet contribution is that the distribution of water vapor cannot be accu-

rately predicted and that it is highly variable. Much research has gone into the development of tropospheric refraction models [*Hopfield*, 1969; *Saastamoinen*, 1972; *Askne and Nordius*, 1987]. These models show that the wet component can be modeled to an rms accuracy of about 2-5 cm in the zenith direction from surface measurements. In order to predict the line-of-sight water-vapor content with higher accuracy a remote sensing approach can be used [see, *e.g.*, *Resch et al.*, 1985; *Davis* , 1986; *Elgered*, 1993].

The index of refraction is conveniently expressed in terms of the refractivity N defined as

$$N = 10^6 (n - 1) \tag{7}$$

A common expression for N is [*Thayer*, 1974]:

$$N = k_1 \frac{p_d}{T} Z_d^{-1} + k_2 \frac{e}{T} Z_w^{-1} + k_3 \frac{e}{T^2} Z_w^{-1} \tag{8}$$

where p_d is the partial pressure of the dry constituents of air in mbar, e is the partial pressure of water vapor in mbar (*i.e.* ,the total pressure $P = p_d + e$), T is the absolute temperature in K, and Z_d^{-1} and Z_w^{-1} are the inverse compressibility factors for dry air and water vapor respectively (corrections for the departure of air from an ideal gas)[*Owens*, 1967]. The first two terms are due to the induced dipole effect, whereas the third term is caused by the permanent dipole moment of the water-vapor molecule. The values of k_1, k_2, and k_3 can be estimated from laboratory experiments , *e.g.*, , those given by *Boudouris* [1963] which are: 77.593 ± 0.08 K/mbar, 72 ± 10 K/mbar, and $(3.754 \pm 0.03)\, 10^5$ K^2/mbar, respectively.

Of the dry atmospheric gases, carbon-dioxide (CO_2) shows the largest variations in concentration. The concentration of CO_2 shows an annual variation (peak to valley 6 ppm) and is currently increasing at the rate 1.7 ppm/year [*Keeling et al.*, 1989]. Measured variation in the atmospheric profiles of CO_2 [*Handbook of Geophys.*, 1985] corresponds to uncertainties well below the mm-level in the zenith delay.

The refractivity also has dispersive components [*Liebe*, 1985]. Their maximum values are found at the absorption spectral lines of the atmospheric gases—mainly O_2 and H_2O. The delay due to the dispersive components of the refractivity (in addition to the frequency independent part in (8)) is much less than 1 mm in the zenith direction for signals at frequencies below 30 GHz.

From refractivity to delay

It is now possible to derive an expression for the excess propagation path in terms of the refractive index as given by (8). We may call the first term in (8) the "dry" refractivity, and the second and third terms together the "wet" refractivity. However, the integral of the dry refractivity (dry delay) is difficult to determine from this definition because the hydrostatic equation is valid for the total pressure and not for the partial pressure of dry air only. It therefore common to define the tropospheric propagation delay to consist of two parts: the hydrostatic and the wet path delay [*Davis et al.*, 1985].

The hydrostatic delay in the zenith direction can be written;

$$\Delta L_h = (0.0022768 \pm 0.0000024\,(\mathrm{m/mbar}))\,\frac{P_\circ}{f(\Phi, H)} \tag{9}$$

where

$$f(\Phi, H) = (1 - 0.00266\cos 2\Phi - 0.00028H) \tag{10}$$

is used to model the variation of the acceleration due to gravity [*Saastamoinen*, 1972], Φ is the latitude, and H is the height in km of the station above the ellipsoid. The error introduced by the assumption of hydrostatic equilibrium will depend on the wind profile above the site but is typically of the order of 0.01% [*Holton*, 1972]. This corresponds to 0.2 mm in the zenith delay.

The hydrostatic delay in the zenith direction will consequently be determined with high accuracy from the total pressure at the ground level and an effective value of the acceleration due to gravity in the atmosphere above the site. Its elevation dependence can be defined by a "dry mapping-function" [*Davis et al.*, 1985; *Lanyi*, 1984; *Chao*, 1992] which is often defined to also include the small but significant effect of the geometric delay discussed at the beginning of this section.

The wet delay can then be written

$$\Delta L_w = 10^{-6}\left[(24 \pm 10)\int_S \frac{e}{T}\,Z_w^{-1}\,ds + (3.754 \pm 0.030) \times 10^5 \int_S \frac{e}{T^2}\,Z_w^{-1}\,ds\right] \tag{11}$$

Since the first term is only about 1% of the second term, (11) can be simplified without loss of accuracy by introducing a mean temperature T_m, defined as

$$T_m = \frac{\int_S \frac{e}{T} Z_w^{-1}\, ds}{\int_S \frac{e}{T^2} Z_w^{-1}\, ds} \tag{12}$$

whereupon we obtain

$$\Delta L_w = \left(1 + (6 \pm 3) \times 10^{-5} T_m\right) (0.3754 \pm 0.0030) \int_S \frac{e}{T^2} Z_w^{-1}\, ds \tag{13}$$

The term ΔL_w is much more difficult to estimate than the hydrostatic path delay because there is no valid hydrostatic-equilibrium equation relating the integral of water-vapor density with its partial pressure at the surface of the earth. For most sites and weather situations, it is very difficult to predict the altitude distribution of the water vapor from the ground value with high accuracy; ideally, complete information of the distribution of water vapor and the temperature in the troposphere are needed.

Examples of hydrostatic and wet delays in the zenith direction are shown in Figure 3 for three sites with different climates. The delays are calculated using meteorological profiles obtained from radiosonde launches made every 12 hours. We note that the variations in the hydrostatic delays increase with increasing distance from the equator (a well known meteorological effect [*Handbook of Geophys.*, 1985]) and that the wet delays are quite variable for all three sites.

The path delay due to condensed water

Even though the path delay is caused mainly by atmospheric gases, there is a possible contribution also from condensed water in the atmosphere—clouds and rain. However, this is nearly always less than one centimeter, even for paths of several kilometers through heavy clouds. More detailed information can be obtained from *Elgered* [1993].

Mapping Functions

The tropospheric delay experienced by the GPS signal is the integration all along the propagation path. Figure 4 shows a typical example of the variation of the tropospheric delays as a function of elevation angle to a satellite. Figure 4 indicates that it should be possible to express the tropospheric delay at a certain elevation angle as the product of the tropospheric zenith

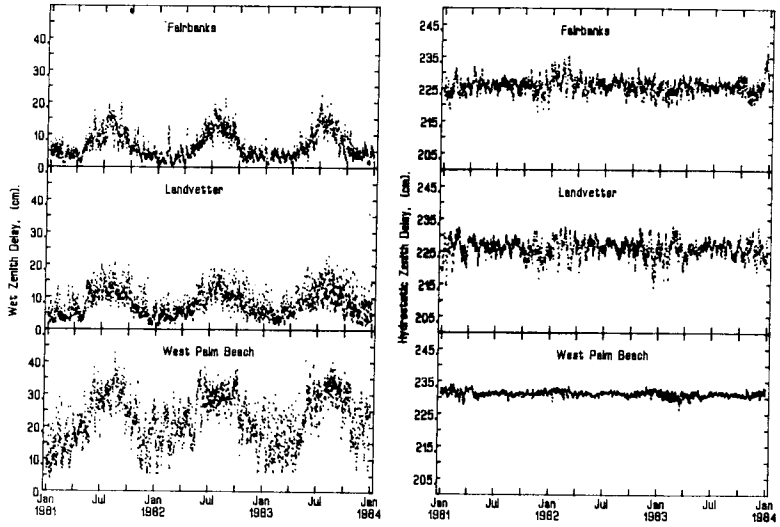


Figure 3: The wet and the hydrostatic delay in the zenith direction for West Palm Beach, Florida, Göteborg-Landvetter Airport, Sweden, and Fairbanks, Alaska. The delays are calculated using radiosonde data obtained twice daily at noon and midnight universal time (UT).

delay and a function that maps the increase in delay with an increasing zenith angle. For elevation angles down to 15° and assuming a plane earth with vertically homogeneous atmosphere, the excess path increases with a factor of $1/\sin(\epsilon)$ where ϵ is the elevation angle ("mapping function"). The same mapping function $m(\epsilon)$ can be used for both the dry and wet components and give the total delay $\Delta L_t(\epsilon)$ in the direction (ϵ) as

$$\Delta L_t(\epsilon) = m(\epsilon) \times (\Delta L_w(90°) + \Delta L_h(90°)) \tag{15}$$

where m is the mapping function. The great advantage being that one can model or estimate the zenith delay for each site as an unknown parameter in the least-squares adjustment. A commonly used mapping function is the one associated with the Saastamoinen tropospheric delay model. For a cut-off angle off 15°, the accuracy of the Saastamoinen mapping function is better than 5 millimeters.

More realistic mapping functions contain additional terms relating to the sphericity of the earth and to the vertical structure of the atmosphere, which contains geographic as well as thermodynamic parameters. New mapping

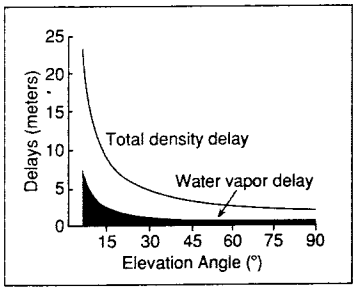


Figure 4: Propagation delays in meters as a function of elevation angle (from *Brunner and Welsch* [1993].

functions are being developed. Separate mapping functions are being used for the hydrostatic and wet delay, respectively. For elevation angles less than 15° several fairly recent mapping funtions have to be considered. The often used mapping functions are usually moderated by meteorological data in order to improve the fit to equation (15) [*Lanyi*, 1984], [*Ifadis*, 1986], and [*Herring*, 1992]. The time of the year and the station position are also used [*Niell*, 1996]. We can now define the equivalent zenith delay,

$$\Delta L_t(90°) = \frac{1}{m_w(\epsilon)} \Delta L_w(\epsilon) + \frac{1}{m_d(\epsilon)} L_h(\epsilon) \tag{16}$$

In general, the equivalent zenith delay is not the same as the wet delay in the zenith direction, rather, it is what the delay in the zenith direction would have been if the measurement acquired in ϵ had been taken in a homogeneous atmosphere.

Estimating the Tropospheric Delay

There are a number of methods to directly measure or estimate the tropospheric propagation delay. Below, we will briefly mention a few of them which has been used in conjunction with GPS obervations.

Surface Meteorological Measurements

Many meteorological models have been developed to calculate the zenith delays from meteorological observations made near the ground at the time of the GPS measurements. The dry delay is directly proportional to the barometric pressure at the antenna height. With a good pressure sensor one can accurately determine the dry zenith delay. A measurement error

of about 0.5 mbar measurement error, give rise to a 1 millimeter error in the zenith propagation delay. The zenith wet delay is only weakly related surface conditions. All known ground based meteorological model suffers from inaccuracies due the poor correlation between surface conditions and the humidity distribution above. The accuracy of existing models has been tested by the use of other techniques, such radiosondes and microwave radiometry, and found to be no better than 30 millimeters. In general , the tropospheric delay models using meteorological surface measurements have produced rather poor results, and in most cases worse, results compared with the results from the default model values that replaced the actual observed meteorological values. This again stress the fact about that the water vapor in the troposphere is poorly mixed. Furthermore, taking accurate surface measurements is difficult task. An error of 5% for the relative humidity or a temperature error of 2° C each cause a zenith delay error of 12 millimeters. This will than be enlarged at lower elevations.

Ground-based Microwave Radiometry

Improved accuracy for estimates of the wet delay can be achieved by the use of remote sensing techniques. An instrument which has almost all weather capability, is a microwave radiometer measuring the emission from the atmosphere on and off the water vapor spectral line centered near 22 GHz. This type of radiometer is normally referred to as a water-vapor radiometer (WVR). A variety of different WVRs have been built (see, *e.g.,*, *Guiraud et al.*, [1979]; *Resch et al.*, [1985]; *Elgered and Lundh*, [1983]). Since a WVR uses directional antennas, the measurments are taken along the line-of-sight. The accuracy in terms of the wet propagation delay is normally better than 1 centimeter. Also the presence of liquid water in clouds can be determined.

Typically a radiometer receives radiation from within a narrow cone of the field of view, defined by the design of the microwave feed horns. Narrow beam width is desirable since admittance of ground radiation must be eliminated. The sky is scanned by means of pointing maneuvers. Thus, the sky can be mapped by laying out a regular angular pattern. Alternatively, the instrument can be pointed at GPS satellites.

There are several applications. Delay estimates from WVR observations can be used directly as corrections in the geodetic parameter solutions. The demand on accuracy is very high in this case. Deriving parameters for water vapour variability is somewhat less crucial; they can be applied to down-

weight observations taken in unfavourable conditions, leading to reduced random and systematic error in a least-suqares fit.

Operating a WVR, however, implies also a substantial amount of data processing and experience with the technique. To consider WVR as a regular ancillary device at geodetic observation sites at present appears not really feasible. Size and delicacy of the instrument also argue against that prospect. Instead, a few of these instruments at fundamental stations, where they can run in parallel with several geodetic techniques has been and will probably also in the future be a more successful route.

Problems persist, however, during rain; the radiometer data will typically overestimate the delay, however with a very high variance. A comparison between a ground-based model and observations of wet delay obtained from a WVR are shown in Figure 5.

Estimating Zenith Propagation Path Delays.

The standard atmospheric model will fail to describe the actual meteorological conditions at a GPS site during a particular observation session. We can define the residual zenith delay as the difference between the actual zenith delay and the calculated from a standard model. It is common practice to also lump the dry and the wet component together. We now include the estimation of the unknown residual zenith delay for each GPS station in a least-squares adjustment of the carrier-phase observations. One unknown residual zenith delay is often estimated per site and session. Since the variation in the tropospheric propagation path delay across a GPS site can be as large as several centimeters per hour it is favorable to use an approach with several unknown residual zenith delays over the observation session. The delay due to the neutral atmosphere can therefore be modeled as a constant for a period of time. For example, in the Bernese Software [*Rothacher and Mervart*, 1996] it is possible to divide the observation session in sub-intervals and thereby estimate several constant zenith delays for each station. *Davis et al.* [1989] achieve high-precision results by solving for a constant zenith delay parameter.

Another approach is to use a stochastic model which treats the unknown residual delay as a time-varying parameter. One such model, which is highly effective is the random walk process. The zenith delay is modeled at every epoch as the sum of the previous zenith delay value plus the noise of purely random process called the process noise which constrains the delay

changes. The mathematical adjustment is carried out using a "Kalman filter". Promising results have been accomplished by the use of stochastic models which makes the assumption that the wet path delay is likely to vary within a limited range over a short time interval. The temporal evolution of the water vapour content and thus the zenith delay can be described most adequately in statistical terms as a random walk process. Hence the delay parameter can be estimated in a way similar to clock drift parameters. In a Kalman filter version of the observation equations the temporal random-walk like property of the process can be modelled with a good degree of realism. This is an approach taken in the GIPSY/OASIS-II program [*Webb and Zumberge*, 1993].

The accuracy of the mapping function at low elevation angles is crucial since these observations contribute most strongly in the parameter estimation owing to the long atmospheric path and large delay. Research on refining their formulation while keeping the number of characterizing parameters managably low is expected to lead to major improvements in the observation and analysis methods. This must be matched with a trade-off of the elevation limit below which observations will have to be discarded. If (and only if) the mapping function represents the actual situation accurately, the gain of accuracy of position solutions due to low-elevation observations is significant.

The fluctuations of water vapour constitute a highly nonstationary process. Temporal characteristics must be studied carefully in order to derive realistic Kalman filter parameters and to trim the solution parameters in order to obtain minimum bias results for geodetic positions [*Elgered et al.*, 1991]. Additional complication arises when spatial water vapour gradients occur. These originate primarily in weather fronts, but can also persist, e.g. across coastlines, during stable weather conditions. Implementation of gradient parameters in the geodetic solutions is currently considered [*Davis et al.*, 1993; *Jarlemark*, 1994]. Making use of low elevation observations are under consideration. Tests have been performed using data down to 5° (in some cases weighting the data). Promising results using new mapping functions, low elevation data, and horizontal gradients have been reported by several groups (*e.g.*, the groups behind the Bernese, GIPSY, and GAMIT software packages). The development of mapping functions and the tailoring of the Kalman filter parameters has been dependent on progress in microwave radiometry.

Normally, it is important to estimate one or several tropospheric pa-

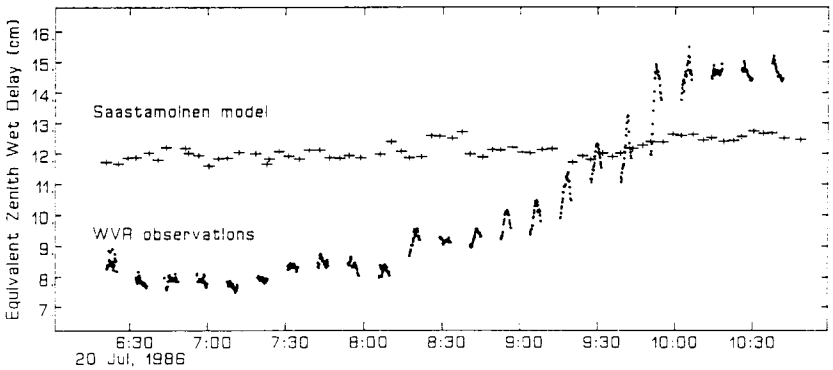


Figure 5: Microwave radiometer measurements of the wet delay during a warm front passage. Each group of data consists of an azimuth scan at a constant elevation of 30°. The scans start in the east, moves over south and west making one observation every 10° and end in the north. Note the total lack of correlation between the delay obtained from the ground-based model [*Saastamoinen*, 1972] and the delay estimates from the radiometer.

rameters also on short baselines (small stations separation). The absolute value of the tropospheric propagation delay may not be correct but the estimated relative propagation delay and the relative height difference between these stations will most likely improve. Shown in Figure 6 is a comparison study of wet delay estimates obtained from WVR, VLBI, and GPS data, respectively.

Ionospheric Propagation Effects and Dual-Frequency Observations

The ionosphere lies above 99.9% of the mass of the Earth's atmosphere and at an altitude of about between 50–1000 kilometers. Because the sun is the primary source of free electron production, the ionosphere varies to a large extent with the rotation of the Earth and with solar activity. The electron content in the ionosphere is created by a production-recombination mechanism involving ionization by ultraviolet and more energetic radiation from the sun and cosmic sources. The losses involve recombination of free electrons to neutral atoms and molecules. Ionization rates and ionizing flux attenuation are proportional to density, and recombination to density and

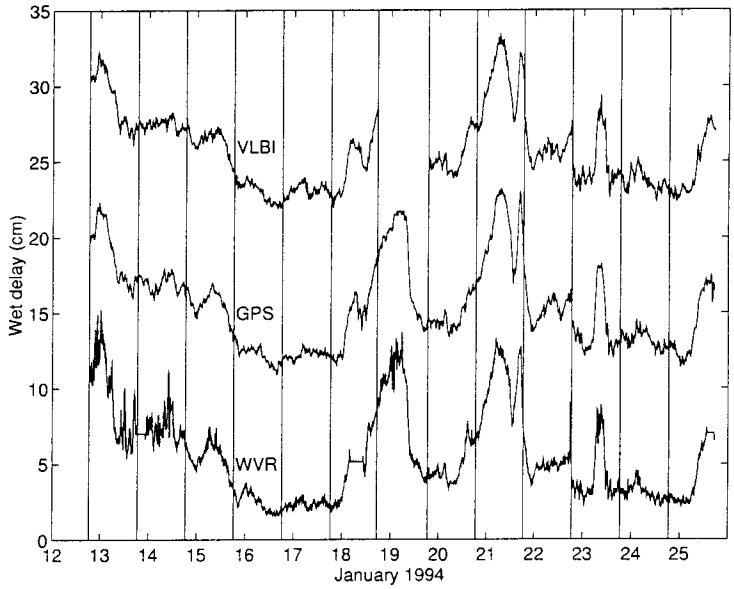


Figure 6: Estimated zenith wet delay for the Onsala site from VLBI, GPS, and WVR, where in order to make it easier to compare results, biases of +20 and +10 cm have been added to the VLBI (upper curve) and GPS (middle curve) results, respectively.

pressure.

The ionosphere is generally classified by the *D*, *E*, *F1*, and *F2* regions. The part of the ionosphere below an altitude of 90 kilometers is referred to as the *D* region. This region primarily affects radio waves through absorption effects, since the neutral density is so high that the electron-neutral collision frequency is substantial. The part 90–120 kilometers is referred to as the *E* region. The *E* region usually does not contribute to either absorption or refraction effects under normal conditions. The part of the ionosphere where the highest production rate occurs *i.e.* from 120 to 180 kilometers, is commonly referred to as the *F1* region. The part of the ionosphere between 180 and 1000 kilometers is referred to as the *F2* region. The *F1* and the *F2* regions usually merge into one region and cannot be distinguished on the basis of the electron density profile. The *F* region exert strong effects

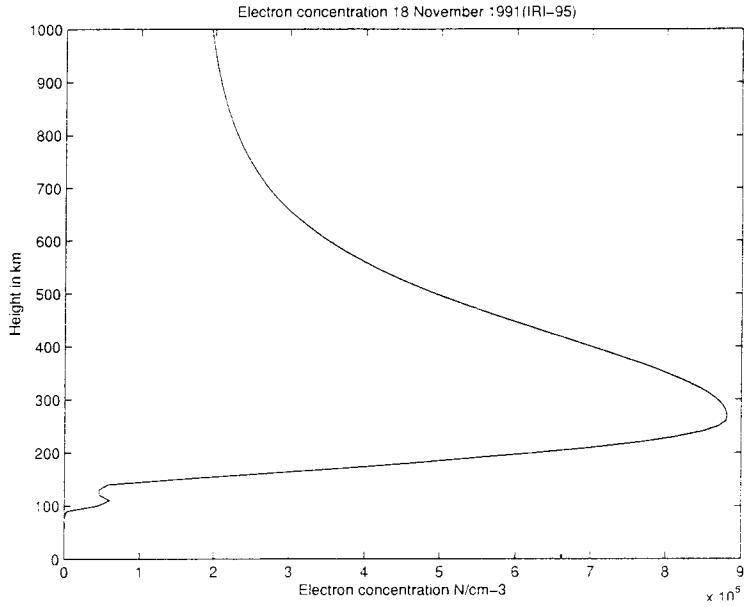


Figure 7: The ionospheric electron content as a function of height at latitude N57° and longitude E11° at 12:00 UTC. The profile is obtained from the IRI-95 model

on the propagation of radio waves, both as regards refraction, phase and group delays, polarization and even absorption. Scintillation effects usually originate in the *F* region because of spatially irregular distribution of ionization. The electron density in the *D* region is typically on the order of 10^8 to 10^9 m^{-3}, in the *E* region during the daytime electron densities of 10^{11} m^{-3} may occur. *F1* region electron densities are typically a few times 10^{12} m^{-3} during the daytime and typically a factor of three to four smaller during night-time conditions. When radio waves propagate through the ionosphere several effects occur. One effect is the retardation of the modulation carrier wave also known as the ionospheric group delay and the advance of the carrier, known as the ionospheric phase advance. The ionospheric effect is proportional to the total electron content (TEC). A problem for GPS observations is that the total amount of electrons in the ionosphere varies in time and space. TEC is a function of many periodic phenomenons such as the solar activity, the zenith angle, the season, the time of day, and also of the magnetic latitude. In Figure 7, we show an electron density profile.

The impact on GPS radio propagation is quite accurately modeled and measured using dual-frequency observations from each GPS satellite for elevation angles above 15°. Additionally, double differencing between sites that see roughly the same ionosphere minimizes ionospheric errors. Below elevation of 15°, other propagation errors such as multipath, tropospheric uncertainties, separation of the L_1 and L_2 carriers, and the ionospheric propagation path becomes significant. We will therefore start with the complete ionospheric magnetoplasma delay. GPS technology has to a large extent ignored the effect of the Earth's magnetic field upon delay and polarization. The plasma refractivity is due to mobility of free ions under the effect of the electromagnetic wave. *Davies* [1965] and *Budden* [1985] gives the complex index of refraction for phase in a magnetoplasma as

$$n^2_{phase,complex} = 1 - \frac{R_p^2}{1 - iR_c - \frac{R_H^2 \sin^2\theta}{2(1-R_p^2-iR_c)}\left(\frac{R_H^4 \sin^2\theta}{4(1-R_p^2-iR_c)^2)}\right)^2 + R_H^2\cos^2\theta} \tag{17}$$

where $R_p = \frac{f_p}{f}$ = electronic plasma to carrier frequency, $R_H = \frac{f_H}{f}$ = gyromagnetic to carrier frequency, $R_c = \frac{f_c}{f}$ = ion collision to carrier frequency, and where f is the carrier frequency, and the θ is the magnetic field and the Poynting vector for the propagation. The frequencies used in GPS are greater than collision frequencies by 4 orders of magnitude which makes it reasonable to ignore collisions. Since the plasma frequency dominates over the gyromagnetic frequency it has been common practice in space geodesy to also ignore the Earth's magnetic field. With these assumptions expression (17) can be approximated by

$$n^2_{phase} = 1 - \left(\frac{f_p}{f}\right)^2 = 1 - \left(\frac{n_e e^2}{\pi m_e}\right)\left(\frac{1}{f^2}\right) \tag{18}$$

where n_e is the electron density, e is the elementary charge, and m_e is the electron mass. The phase index of refraction is less than unity implies that carrier phase velocities exceeds the speed of light. The energy and information of the GPS signals are however, carried at the group velocity which is $\leq$ c. The group index of refractivity can be expressed in a similar way:

$$n^2_{group} = 1 + \left(\frac{f_p}{f}\right)^2 \tag{19}$$

still ignoring collisions and the Earth's magnetic field.

The refractivity, N, is defined as parts per million difference from unity as $N = (n-1)10^6$. The plasma refractivity is therefore:

$$N = \left(\frac{f_p}{f}\right)^2 \times 10^6 = \frac{40.3 \times 10^{-6} n_e}{f^2} \tag{20}$$

Using dual-frequency observation the ionospheric delays can be determined by measuring the offset in the two GPS carrier phases ($L_1 \approx 1575$ MHz and $L_2 \approx 1227$ MHz).

Keeping the assumption that for sufficiently high frequencies we can ignore the magnetic field and collisions in the calculations and obtain for the excess phase path (the difference between measured and geometric range)

$$\Delta L_{phase} = \int (n-1) ds \approx -\frac{1}{2} \int (\frac{f_p}{f})^2 ds = -\frac{1}{2} \int (\frac{f_p}{f})^2 \frac{dh}{\cos i} \tag{21}$$

where n is the refractive index and the plasma frequency f_p is equal to $\frac{\omega_p}{2\pi} = \frac{ne}{m\epsilon_o}$. The various fundamental quantities entering into the above definitions are m $= 9.11 \times 10^{-31}$ kg, e $= 1.60 \times 10^{-19}$ C, and $\epsilon = 8,859 \times 10^{-12}$ C/(V m). By inserting these fundamental constants we obtain the following useful relation between electron density and the plasma frequency

$$f_p = \frac{\omega_p}{2\pi} = 8.984 \sqrt{n_e} \tag{22}$$

The negative sign for the phase in equation (21) indicates that the phase is decreased by the presence of the plasma. The group delay, however, is increased by the same amount

$$\Delta L_{group} = \frac{1}{2} \int (\frac{f_p}{f})^2 \frac{dh}{\cos i} = \frac{40.3}{f^2} \int \frac{ndh}{\cos i} \tag{23}$$

The difference between day time and night time values of the integrated electron content could amount to one order of magnitude. This full change particularly during the morning, occurs in the course of 2 to 4 hours. A time rate of change of the phase path will cause a frequency shift which is proportional to the time rate of change of the excess phase path.

The state of the ionosphere is described by the electron density n_e $[electrons/m^3]$. By defining the TEC as

$$TEC = \int n_e(s) ds \tag{24}$$

and substituting TEC into equation (23) the ionospheric group delay can be approximately expressed as

$$\Delta L_{group} = \frac{40.3}{c f^2} \mathrm{TEC} \tag{25}$$

where c is the speed of light in meters per second, f is the frequency in Hz and TEC is the total electron content in electrons per square meter. The influence of smaller terms in the group delay expression and the signal bending effect are normally negligible [*Brunner and Gu*, 1991]. The daily variations in the zenith delay can be about 15 m. In geodetic applications dual-frequency observations are necessary in order to eliminate the ionospheric delay. From the equations above it is clear that the ionospheric delay may almost be eliminated if we form the ionospheric-free carrier phase as [see, *e.g.*,, *Blewitt*, 1989]

$$L3 = \frac{f_{L1}^2}{f_{L1}^2 - f_{L2}^2} L1 - \frac{f_{L2}^2}{f_{L1}^2 - f_{L2}^2} L2 \tag{26}$$

Radio propagation through the plasma from near zenith at the two carrier frequencies travels through less plasma, and does not bifurcate appreciably because it is aligned with the plasma gradient. The delay is therefore quite accurately evaluated with this expression still however keeping the assumption that we can ignore the effect of the magnetic field and free electron collisions. At lower elevations, however, the these assumptions along with unmodeled plasma gradient-induced ray curvature and unmodeled birefringence induced by the Earth's magnetic field will result in errors in the calculated ionospheric delay. A more comprehensive model including above deficiencies, has been developed by *e.g., Brunner and Gu* [1991]. They have reported improvement from 33 to 0.2 millimeter at 15° elevation angle for high solar activity (zenith TEC 1.4×10^{18}). By including an additional frequency of known phase relationship to L_1 and L_2 would enable much more accurate modeling of the ionosphere by utilizing a three-term nonlinear combination that would remove most of the bending and bifurcation effect allowing for observations at lower elevation angle. This may be desirable in order to improve the accuracy of the vertical and tropospheric estimates.

A global system of permanent sites can be used to routinously map the ionosphere. Such a ionosphere service is currently devised under IGS

for applications in research and as a support for single-frequency surveys [*Sardon et al.*, 1994].

Presently the sunspot number is low. The ionospheric activity is going to increase to a maximum in the year 2001. This could affect the precision of GPS based measurements, and solving for second order dispersion terms may become actual.

Effects on Geodetic Networks

Double-differences of GPS phase observations greatly reduce the effect of errors in satellite and receiver clocks and significantly reduces the atmospheric effect for short baselines where the GPS signals travel along nearly the same paths through the atmosphere. The positive correlation of the tropospheric effects is reduced if there are differences in station heights and larger station separations. One should always remember the inherent geometrical weakness in the GPS baseline results which usually makes the determination of the vertical baseline component worse by a factor of 3 compared with the horizontal baseline component if a elevation cut-off angle of 15° is used. Additionally a residual atmospheric delay error of 1 centimeter causes a 3-centimeter in the height relative height difference between baseline stations.

Theoretically, the use of a lower cut-off angle could reduce this high correlation between atmospheric parameters and heights and thereby allow for more accurate height determination. However, observations at elevation cut-off angles less than 15° will also include problems such as unmodeled parts of the tropospheric propagation delay, less accurate mapping functions, influence of the higher order ionospheric terms, and also effects due to signal scattering and multipath related the GPS antenna and its environment.

Air pressure loading

In analogy to the loading of the oceanic tide also air pressure variations induce deformation of the earth. The transient, aperiodic nature of the time evolution of the barometric field implies that predictions of the loading effects must be explicitly time-dependent and to be based on global pressure data.

The loading response is most efficient on continental areas. Where air pressure variations occur over water-covered areas, the water body will ad-

just in order to attain hydrodynamic equilibrium, meaning that the pressure along the bottom will become constant in the long-time limit.

Vertical displacements of the crust amount typically to 1 cm, but in extreme situations, particularly in the centre of large continents where loading is not compensated by oceans, up to 3 cm. Horizontal displacement of the crust occurs at roughly a 1:3 ratio with respect to the vertical ones. The precision ratios of space geodetic methods are typically similar, so both displacement directions are equally important.

Other elevation-dependent perturbations

GPS Antennas

It has been found that antenna-to-antenna phase differences can introduce range biases at the several centimeter level, which may limit the precision of the measurements [*Rocken*, 1992]. Differential phase errors due to GPS antennas will not only affect the precision in GPS networks with different types of antennas, but also in networks using identical antennas if the network covers a large spatial area (baseline lengths $\approx$ 1000 km) [*Schupler and Clark*, 1991; *Schupler et al.*, 1994].

The problem of antenna mixing was addressed at the IGS Analysis Center Workshop in Silver Spring, 1996. Two sets of phase calibration corrections (PCC) tables have been put together based on material presented by *Mader and MacKay* [1996], *Rothacher and Shär* [1996], and *Meertens et al.*, [1996a] to be used by the IGS Analysis Centers and others in the GPS community: (1) a set of "mean" phase center offsets and (2) a set of elevation-dependent PCC and offsets relative to the Dorne-Margolin T antenna.

Since the PCC values are all relative to the Dorne-Margolin T antenna some effects of antenna mixing still remain. Even with the same type of antenna the variation in the apparent phase center as a function of elevation angle will influence the results on longer baselines. Therefore the task of getting absolute calibration of the antennas through, e.g., chamber measurements or antenna simultation software may be essential for some applications even though these calibration values most likely will change when the antenna is deployed in the field.

Effects like these can of course be reduced by utilizing antennas less sensitive to scattering from external structures. One way to achieve this is to reduce the side- and back-lobe levels of the amplitude patterns by means

of well designed ground-planes. For this purpose new antenna designs have been proposed [see e.g., *Alber* [1996]; *Ware et al.*, [1997]; *Jaldehag*, [1995]; and *Clark et al.* [1996]. Futhermore, several groups are currently developing methods to perform absolute field calibration of antennas [see e.g., *Wübbena et al.*, [1996] and insitu calibration of antenna/pillar systems.

Antenna-Pillar System and the Signal

The total electromagnetic field of an antenna which radiates a signal in the presence of conducting structures may be expressed as a superposition of the transmitted field and the fields scattered (i.e., reflected and diffracted) by the structures. By reciprocity, the same is true for a receiving antenna. The significance of the scattered field depends on the degree of electromagnetic coupling between the antenna and the scatterer, that is, the distance to the scatterer and the size and reflectivity of the scatterer. Signal scattering affects both the amplitude and phase of the received GPS signal, presumably independently at each site in a network. This independence creates differential phase errors.

Scattering from structures in the vicinity of the antenna effectively changes the antenna phase pattern, and, thus, affects the precision of the carrier phase measurements of the GPS signal. In studies by *Elosegui et al.* [1995] and *Jaldehag et al.* [1996a] it was shown that estimates of the vertical component of baselines formed between sites using identical antennas were dependent on the minimum elevation angle of the data processed. Both studies found that the elevation-angle- dependent systematic effect was associated with non-identical pillar arrangements, causing differential phase errors due to scattering from structures associated with the mounting of the antenna to the pillar, and with the pillar itself. Even the most perfectly calibrated antenna the antenna phase pattern will change when attached to a pillar. Estimates of the vertical component of many baselines strongly depend on the minimum elevation angle (elevation cutoff angle) of the data analyzed. Offsets of several cm in the vertical component of many stations are evident when the elevation cutoff angle is changed.

Radomes - Protective Covers

At several permanent GPS sites located in areas with periodically severe environmental conditions (snow, rain) radomes have been employed. Some of the regional networks are located at higher latitudes and radomes have

been used to protect the antenna from snow accumulation. Until recently, most radomes in use have had a conical shape e.g. in order for the snow to slide off. Several groups have recently been investigating effects due to the excess signal path delay through the radome. Different radomes have been tested in an anechoic chambers [*Clark et al.*, 1996; *Meertens et al.*, 1996b] as well as in field tests [*Meertens et al.*, 1996b; *Jaldehag et al.*, 1996c]. All tests show that a conical cover may cause cm-level vertical errors when the tropospheric delay parameter is estimated since the excess signal path delay which will map into other parameters in the GPS software. Preliminary results of hemispheric radomes shows a smaller, 2 mm vertical offset, showing less elevation dependence. The influence on the tropospheric wet delay estimates will also be small. In general, the effect is more or less constant and may be calibrated or modeled.

Precipitation

Signal propagation delay during snow storms has been investigated by, e.g., *Tranquilla and Al-Rizzo* [1993] and *Tranquilla and Al-Rizzo* [1994] who demonstrated that due to the localized nature of many snow storms differential effects may cause systematic variations at the centimeter level in estimates of the vertical coordinate of site position. Systematic variations introduced by snow storms may, however, if short-lived (minutes to hours), be reduced to a high degree by data averaging. A potentially more serious effect of heavy snow precipitation is the accumulation of snow on the top of the GPS antenna and on its surroundings, such as on the top of the GPS pillar or, when present, on the radome covering the antenna. This accumulation may last for days, weeks, or months. *Webb et al.* [1995] reported variations on the order of 0.4 m in estimates of the vertical coordinate of site position. The variations were correlated with the accumulation of snow over the antenna. Variations at the several centimeter level in estimates of the vertical coordinate of site position strongly correlated with changes in the accumulation of snow on top of GPS antennas have also been observed by others [*Jaldehag et al.*, 1996b; *Bifrost Project*, 1996; *Meertens et al.*, 1996a]. The results indicate that the variations in the vertical coordinate of site position can be fully explained by reasonable accumulations of snow which retard the GPS signals and enhance signal scattering effects.

References

Alber, C., Millimeter Precision GPS Surveying and GPS Sensing of Slant Path Water Vapor, Ph.D. Thesis, Univ. of Colorado, Dec. 1996.

Askne, J., and H. Nordius, Estimation of tropospheric delay for microwaves from surface weather data, *Radio Sci.*, vol 22, pp. 379–386, 1987.

BIFROST project members, GPS Measurements to Constrain Geodynamic Processes in Fennoscandia, *EOS, Transactions, American Geophysical Union, 35*, 1996.

Blewitt, G., Carrier Phase Ambiguity Resolution for the Global Positioning System Applied to Geodetic Baselines up to 2000 km, *J. Geophys. Res., 94*, 10187–10203, 1989.

Boudouris, G., On the Index of Refraction of Air, the Absorption and Dispersion of Centimeter Waves by Gasses, *J. Res. Natl. Bur. Stand.*, 67D, pp. 631–684, 1963.

Brunner, F.K., and M. Gu, An Improved model for dual frequency ionospheric correction of GPS observations, *Manuscr. Geod.*, 16(3), 205–214, 1993.

Brunner, F.K., and W.M. Welsch, Effects of the Troposphere on the GPS Measurements, *GPS World*, January, 1993.

Budden, K.G., The Propagation of Radio Waves, The Theory of Low Power Radio Waves in the Ionosphere and Magnetosphere, Cambridge Univerisy Press, 1985.

Chao, C.C., A Model for Tropospheric Calibration from Daily Surface and Radiosonde Balloon Measurements, *Tech. Mem. 391-350*, Jet Propul. Lab., Pasadena, California, 1972.

Clark, T.A., B. Schupler, C. Kodak, and R. Allshouse, GPS Antennas: Towards Improving Our Understanding of Factors Affecting Geodetic Performance, *EOS Trans. AGU, 77*, Fall Meet. Suppl. 1996.

Davies, K., Ionospheric Radio Propagation, NBS Monograph 80, 1965.

Davis, J. L., T. A. Herring, I. I. Shapiro, A. E. E. Rogers, and G. Elgered, Geodesy by Radio Interferometry: Effects of Atmospheric Modeling Errors on Estimates of Baseline Lengths, *Radio Science, 20*, 1593–1607, 1985.

Davis, J.L., Atmospheric Propagation Effects on Radio Interferometry, Ph.D. Thesis, Massachusetts Inst. of Tech., 1986.

Davis, J. L., G. Elgered, A. E. Niell, and C. E. Kuehn, Ground-based measurement of gradients in the "wet" radio refractivity of air, *Radio Sci., 28*, 1003–1018, 1993.

Elgered G., and P. Lundh, A dual channel water vapor radiometer system, *Res. Rept. No 145*, Onsala Space Observatory, Chalmers Univ. of Tech., 1983.

Elgered, G., J.L. Davis, T.A. Herring, and I.I. Shapiro, Geodesy by Radio Interferometry: Water Vapor Radiometry for Estimation of the Wet Delay, *J. Geophys. Res.*, vol**96**, pp. 6541–6555, 1991.

Elgered, G., Tropospheric Radio Path Delay from Ground-Based Microwave Radiometry, in Janssen M (ed.): Atmospheric Remote Sensing by Microwave Radiometry, 215–258, Wiley & Sons, New York, 1993.

Elósegui, P., J. L. Davis, R. T. K. Jaldehag, J. M. Johansson, A. E. Niell, and I. I. Shapiro, Geodesy Using the Global Positioning System: The Effects of Signal Scattering on Estimates of Site Position, *J. Geophys. Res., 100*, 9921–9934, 1995.

Guiraud, F.O., J. Howard, and D.C. Hogg, A Dual Channel Microwave Radiometer for Measurement of Precipitable Water Vapor and Liquid, *IEEE Trans. Geosci.*

Electronics, **GE-17**, pp. 129–136, 1979.

Handbook of Geophysics and the Space Environment, Ed. A.S. Jursa, p. 21-21, 1985.

Herring T.A., Modeling Atmospheric Delays in the Analysis of Space Geodetic data, *Proc. of the Symp. Refraction of Transatmospheric Signals in Geodesy*, Eds. J.C. De Munch and T.A.Th. Spoelstra, Netherlands Geodetic Commission, Publ. on Geodesy, No. 36, pp. 157–164, 1992.

Holton, J.R., An Introduction to Dynamic Meteorology, Academic Press, New York, 1972.

Hopfield, H.S., Tropospheric Effects on Signals at Very Low Elevation Angles, Tech. Memo. TG1291, Applied Physics Laboratory, John Hopkins University, 1976.

Ifadis, I., The Atmospheric Delay of Radio Waves: Modeling the Elevation Dependence on a Global Scale, Lic. thesis, *Technical Rep. No. 38L*, Chalmers University of Technology, Göteborg, 1986.

Jaldehag, R. T. K., Space Geodesy Techniques: An Experimental and Theoretical Study of Antenna Related Error Sources, Ph. D. Thesis, Tech. Rept. No. 276, School of Electrical and Computer Engineering, Chalmers University of Technology, Gothenburg, Sweden, 1995.

Jaldehag, R. T. K., J. M. Johansson, P. Elósegui, J. L. Davis, A. E. Niell, B. O. Rönnäng, and I. I. Shapiro, Geodesy Using the Swedish Permanent GPS Network: Effects of Signal Scattering on Estimates of Relative Site Positions, *J. Geophys. Res.*, *101*, 17814–17860, 1996a.

Jaldehag, R. T. K., J.M. Johansson, J.L. Davis, and P. Elósegui, Geodesy using the Swedish Permanent GPS Network: Effects of snow accumulation on estimates of site positions, *Geophys. Res. Lett.*, 23, 1601–1604, 1996b.

Jaldehag, R.T.K., J.M. Johansson, and J.L. Davis, Environmental Effects on the Swedish Permanent Network, *EOS Trans. AGU, 77*(46), Fall Meet. Suppl., F453, 1996c.

Jarlemark P O J, Microwave radiometry for studies of variations in atmospheric water vapor and cloud liquid content, Techn. Rep. No. 181L, School of Electrical and Computer Engineering, Chalmers University of Technology, Göteborg, Sweden, 99 pp, 1994.

Keeling, C.D., R.B. Bacastov, A.F. Carter, S.C. Piper, and T.P. Whorf, *Am. Geophys. Un., Geophys. Monogr.*, 55, pp. 165–236, 1989.

Lanyi, G., Tropospheric delay effects in radio interferometry, *TDA progress report 42–78*, Jet Propulsion Laboratory, Pasadena, California, 1984.

Liebe, H.J., An update model for millimeter wave propagation in moist air, *Radio Sci.*, 20, 1069–1089, 1985.

Mader, G.L., and MacKay, J.R., Calibration of GPS Antennas, *Proc. IGS Analysis Center Workshop*, Silver Sping, MD, 1996.

Meertens, C., C. Rocken, J. Braun, M. Exner, B. Stephens, and R. Ware, Field and Anechoic Chamber Tests of GPS Antennas, *Proc. IGS Analysis Center Workshop*, Silver Sping, MD, 1996a.

Meertens C., J. Braun, C. Alber, J. Johnson, C. Rocken, T. Van Hove, B. Stephens, and R. Ware, Antenna and Antenna Mounting Effects in GPS Surveying and Atmospheric Sensing, *EOS Trans. AGU, 77*(46), Fall Meet. Suppl., F453, 1996b.

Niell A.E., Global Mapping Functions for Atmosphere Delay at Radio Wavelength, *J. of Geophysical Res.*, 101 B2, 3227–3246, 1996.

Resch, G.M., M.C. Chavez, N.I. Yamane, K.M. Barbier, R.C. Chandlee, Water Vapor Radiometry Research and Development Phase Final Report, *JPL Publication 85-14*, Jet Propulsion Laboratory, Pasadena, California, 1985.

Owens, J.C., Optical Refractive Index of Air: Dependence on Pressure, Temperature, and Composition, *Appl. Opt.*, 6, pp. 51–58, 1967.

Rocken, C., GPS antenna mixing problems, *UNAVCO Memo*, UNAVCO, Boulder, Colorado, 1992.

Rothacher, M., and Schär, S., Antenna Phase Center Offsets and Variations Estimated from GPS Data, *Proc. IGS Analysis Center Workshop*, Silver Sping, MD, 1996.

Rothacher, M., and L. Mervart, Eds. *Bernese GPS Software Version 4.0)*, Astronomical Institute, University of Berne, 1996.

Saastamoinen, J., Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, in *The Use of Artificial Satellites for Geodesy, Geophys. Monogr. Ser.*, 15, edited by S.W. Henriksen et al., pp. 247–251, American Geophysical Union, Washington, D.C., 1972.

Sardon E, Rius A, Zarraoa N, Ionospheric calibration of single frequency VLBI and GPS observations using dual GPS data, *Bull. Géodésique*, Vol 68, 230–235, 1994.

Schupler, B. R. and T. A. Clark, How different antennas affect the GPS observable, *GPS World*, 32–36, Nov./Dec., 1991.

Schupler, B. R., R. L. Allshouse, and T. A. Clark, Signal characteristics of GPS user antennas, *Navigation*, 41, No. 3, 277–295, 1994.

Thayer, G.D., An Improved Equation for the Radio Refractive Index of Air, *Radio Sci.*, 9, pp. 803–807, 1974.

Tranquilla, J.M. and H.M. Al-Rizzo, Theoretical and experimental evaluation of precise relative positioning during periods of snowfall precipitation using the global positioning system, *Manuscripta Geodaetica*, 18, 362–379, 1993.

Tranquilla, J.M. and H.M. Al-Rizzo, Range errors in Global Positioning System during ice cloud and snowfall periods, *IEEE Trans. Antennas and Propagat.*, 42, 157–165, 1994.

Ware, R., C. Alber, C. Rocken, and F. Solheim, Sensing integrated water vapor along GPS ray paths, *Geophys. Res. Lett*, 24, pp 417–420, 1997.

Webb F H, Zumberge J F, *An Introduction to GIPSY/OASIS-II Precision Software for the Analysis of Data from the Global Positioning System*, JPL Publ. No. D-11088, Jet Propulsion Laboratory, Pasadena, Cal., 1993.

Webb, F.H., M. Bursik, T. Dixon, F. Farina, G. Marshall, and R.S. Stein, Inflation of Long Valley Caldera from one year of continuous GPS observations, *Geophys. Res. Lett.*, 22, 195–198, 1995.

Wübbena, G., F. Menge, M. Schmitz, G. Seeber, and C. Völksen, A New Approach for Field Calibrations of Absolute Antenna Phase Center Variations, *Proc. ION GPS-96*, 1996.

TECHNICAL COMPARISON

BETWEEN

THE GLONASS AND GPS CONCEPTS

Börje Forssell

Department of Telecommunications
Norwegian University of Science and Technology
Trondheim

Geodetic Applications of GPS, Båstad, 26 - 31 August 1996

Introduction

GLONASS (Global Navigation Satellite System; in Russian: Global'naya Navigatsyonnaya Sputnikovaya Sist'ema) is similar to GPS in that it is a space-based navigation system providing global, around-the-clock, all-weather access to precise position, velocity and time to a properly equipped user.

GLONASS was inaugurated in 1982, four years after GPS. It reached full constellation with 24 satellites on 18 January this year, about half a year after GPS was declared having reached full operational status. During these 14 years of GLONASS development, a large number of satellites have been launched. The system is military, like GPS, but general information about it was for many years not as available as about GPS.

The first release from the Soviet Union of detailed GLONASS information occurred at the International Civil Aviation (ICAO) special committee meeting in Montreal in May 1988 [1]. Since the break-down of the Soviet Union the information flow has increased steadily, and now there is little difficulty to get information as required. It could be mentioned that the Russian Space Forces, i.e. the system operator, even have their home page on the Internet [2].

Both GLONASS and GPS (L1 - C/A-code) have been officially offered for free use by the international navigation community for the next 10 -15 years.

Below, technical comparisons are made between GPS and GLONASS. It is generally assumed that the reader has broad knowledge of GPS and its principles of operation.

Similarities between GPS and GLONASS.

In comparing the two systems, one is generally more struck by the many similarities than by the differences. Both systems consist of a nominal number of 24 satellites in near-circular inclined orbits about 20 km above the earth's surface.The orbital time is approximately half a day.

Every satellite in the two systems transmits pseudo-random codes bi-phase modulated onto two L-band carriers. A 50-bit/s navigation message is also modulated onto the carriers. This message contains necessary information to the users such as satellite almanacs, time, health, etc.

The codes are used for time-of-arrival determination by the receiver, thus enabling it to measure the travel time from the satellite and compute the corresponding distance. With knowledge of satellite positions, the receiver computes the four unknowns in the general case: three spatial coordinates and time (really: the difference between receiver and satellite system time).

Depending on user requirements, both systems can be used analogously for more precise positioning by code and phase differential methods, and for time transfer.

Differences

Orbits

GLONASS has three orbital planes, 19100 km above the earth and inclined 64.8°, with eight satellites in each. The ascending nodes are separated by 120° at the equator, and the satellites are evenly distributed in each orbit, i.e. 45° orbital phase (spacing). There is also a displacement of 30° for satellites in neighbouring planes.

The GLONASS inclination of 64.8° (compared to 55° for GPS) makes GLONASS DOP values slightly lower at high latitudes.

All satellites have the same nominal orbit period of 11h 15.73 min. The orbit has a ground-track repeat every 17 orbits, i.e. eight days less 32.56 min. This means that in 24 hours less Δt = 4.07 min. a satellite completes 17/8 orbits, i.e. two whole revolutions plus 45°. Consequently, two satellites in the same plane but 45° apart in orbital phase appear at exactly the same orbital position on two consecutive days less Δt. Thus, the user on the ground sees both satellites at the same elevation and azimuth with that time interval. In the course of eight days, all the satellites in this plane appear at that position with the time interval of Δt, and then the whole cycle repeats. By examining the other two planes, one observes that the satellites in those planes also appear at this very position with the same interval.

Launches and satellites

The GLONASS satellites [3] are launched three at a time by four-stage Proton rockets from Tyuratam (Baikonur Cosmodrome) in Kazakhstan. Each launch is aimed at placing the satellites in their final orbit within one minute of the desired orbital period. For many years, an average of two launches a year has been common. This makes the total number of satellites launched since 1982 exceed 70. The reason for this large (compared to GPS) number is a short lifetime for the satellites. They were originally designed to live only one year, but this has been successively increased, and at present the designed lifetime is five years.

In order to replace malfunctioning satellites, GLONASS maneuvering capability is used to shift healthy satellites to wanted positions. A relocation is usually accomplished within two just weeks.

A GLONASS satellite has a mass of 1400 kilos (a GPS Block IIA has 930 kilos). It is a cylindrical pressure vessel about 3 m in height, carrying a payload platform with twelve element antennas and two solar panels. The three-axis stabilised spacecraft is equipped with a propulsion system for initial orbit acquisition and positioning, station-keeping and relocation. Attitude-control sensors include a geomagnetic sensor on the tall boom at the top of the craft. Laser corner-cube reflectors are carried in order to aid in precise orbit determination and geodetic research.

Frequencies

Like GPS, GLONASS transmits in two L-band channels. Whereas GPS uses one frequency for L1 and another for L2, the same for all satellites, each GLONASS satellite originally had its unique frequency for both, according to the formulas [4]

$f_{L1} = 1602\ \text{MHz} + m \cdot \Delta f_{L1}$, where $m = 1, \ldots\ldots, 24$, $\Delta f_{L1} = 562.5\ \text{MHz}$, (1)
and
$f_{L1} / f_{L2} = 9/7$. (2)

However, there is a lack of frequencies in this band, and both radio astronomers and mobile satellite communication service providers are potential users of parts of the band 1610 - 1626.5 MHz. For this reason, the Russians have agreed to gradually shrink the GLONASS band in accordance with the following schedule [4]:

- until 1998: stop transmission of frequencies in the radioastronomy band (1610.6 - 1613.8),
- 1998 - 2005: only carrier frequencies between 1602 and 1609.25 MHz will be used,
- after 2005: carrier frequencies between 1598.0625 and 1605.5 MHz will be used.

System operation with this reduced number of frequencies can be accomplished by making antipodal satellites use the same frequencies.

Codes and modulation

Like GPS, GLONASS uses spread-spectrum codes for distance determinations and jamming protection. However, as different GLONASS satellites transmit on different frequencies (frequency-division multiplex, FDM), the same codes, one "C/A" and one "P", are used for all satellites without interference problems. The code rates are half those of GPS, i.e. 511 kbits/s for "C/A" and 5.11 Mbits/s for "P". The codes are simple maximum-length sequences, generated according to polynomials $1 + X^5 + X^9$, and $1 + X^3 + X^{25}$, respectively. The latter code is truncated every second. At "L2", only the high-rate code is carried (GLONASS-M satellites (see below) transmit both codes on both carriers), at "L1", both, one in phase and the other in quadrature (as GPS).

The codes are BPSK modulated onto the carriers together with 50-baud data of the navigation message.

The navigation message

This message contains information similar to, although not identical with, that of GPS. Data are sent in lines of 2-s duration, assembled in subframes of 15 lines lasting 30 s, and frames of five subframes, lasting 2.5 min.

A subframe contains the almanacs of five satellites. The ephemerides are given with validities in hours and quarter-hours. Whereas GPS ephemeris data are given as Kepler parameters with additional corrections, GLONASS data are given as satellite instantaneous position and velocity in an earth-centered, earth-fixed rectangular coordinate system. The data are encoded at half-hour intervals, and intermediate data are intended to be calculated by the receiver using interpolation procedures and acceleration terms provided. Resolutions are 0.5 m in position and 1 mm/s in velocity.

Contrary to the ephemerides of a satellite, almanacs are given as Kepler parameters in GLONASS (as in GPS). GLONASS almanacs are much more accurate than GPS ones, and consequently need more frequent updates. This is done about every day (in contrast to intervals of several days in GPS).

Both GPS and GLONASS transmit satellite clock corrections to system time and system-time corrections to a national UTC time reference. GPS satellite corrections are given as clock offset, frequency offset and frequency rate-of-change, whereas GLONASS only transmits clock and frequency offsets. Information of GLONASS system time deviation from the Russian UTC reference (in the order of a few tens of ns, no consideration of leap seconds) is given as a single time offset parameter, whereas the equivalent GPS information also contains the time derivative.

It has been indicated by the Russians that future GLONASS satellites will give the time difference between US (Naval Observatory) and Russian UTC. At the beginning of 1996, GPS time was within about 20 nanoseconds from UTC, whereas GLONASS time was off by about 25 microseconds [8]. UTC (Russia) differs by about 7 μs from UTC.

The GPS message gives a number of coefficients enabling the single-frequency user to calculate an average ionospheric correction term for every satellite. The GLONASS message contains no such information.

Clocks

GPS satellites carry two rubidium and two cesium clocks. GLONASS satellites carry three cesium standards. One of the reasons for the comparatively short lifetime of GLONASS satellites has been onboard clock failures. GLONASS clock performance is steadily improving, and although the average lifetime has not yet reached the GPS level, long-term stabilities are comparable. Lately, even better performance has been achieved by GLONASS clocks.

Coordinate system

As well known, GPS uses the World Geodetic System 1984 (WGS-84). Although well established globally, this system is not used by GLONASS whose reference coordinate system is called PZ-90. (PZ is short for Russian "Parameters of the Earth".) It was formely called SGS-90 (SGS = Soviet Geodetic System). SGS-90 is a slight modification of its predecessor, SGS-85 (see below).

The PZ-90 datum uses the Krasovsky ellipsoid with its fundamental point at the Pulkovo observatory south of St.Petersburg. There exists a transformation formula between the two datums according to the Helmert or Bursa-Wolf transformation [5]:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} + (1+s)\begin{bmatrix} 1 & \sigma & -\phi \\ -\sigma & 1 & \varepsilon \\ \phi & -\varepsilon & 1 \end{bmatrix}\begin{bmatrix} u \\ v \\ w \end{bmatrix} \qquad (3)$$

where x, y, z are the coordinates in PZ-90 and u, v, w are the coordinates in WGS-84. The two coordinate frames are brought substantially into coincidence by a rotation of 0.06" around the z-axis, i.e. $\sigma = 0.06" = 0.29$ μrad, and $\varepsilon = \phi = 0$, and a two-meter displacement along the z-axis, i.e. $\Delta z = -2.0$ m and $\Delta x = \Delta y = 0$. The scale coefficient is $s = 0.23 \cdot 10^{-6}$. (It should be added that the difference between SGS-90 and SGS-85 can be seen in the transformation above where for SGS-85 σ was ten times larger, i.e. 0.6", Δz was - 4 m and s was zero.)

The rms difference between PZ-90 and WGS-84 has been found to be well below 20 meters. A general observation is that the coincidence between positions measured by GLONASS and GPS (without consideration of coordinate system differences) is better within the borders of the former Soviet Union, probably because all the GLONASS control stations are in that area, whereas corresponding GPS stations are more evenly distributed around the world.

Selective availability

There is no deliberate performance reduction for GLONASS "C/A"-code users. It has also been stated repeatedly by the system providers that nothing of that kind will be introduced. Thus, GLONASS accuracies are generally comparable to those of GPS without SA.

Receivers

So far, commercial availability of GLONASS receivers has been very limited. This is kind of a chicken-and-egg situation: as long as the receiver market is small, prices stay high, and the market stays small while prices are high. Another reason is perhaps lack of general confidence in the system itself and in the economic ability and willingness by the responsible Russian authorities to fulfill their commitment. This situation has started to change after the system was declared operational, but still there is a very long way to go before GPS and GLONASS market volumes can be compared. Even in Russia, GPS receivers far outnumber GLONASS ones.

Now that GLONASS receiver availability is slowly increasing and prices slowly decreasing, there is a tendency among professional users to look at GLONASS for applications where GPS with SA is too inaccurate, or where combined use of the

systems is necessary to meet other requirements, e.g. integrity and/or availability. Some manufacturers have also started to make combined receivers.

Use of signals from both systems in the same receiver configuration is of course very advantageous — except for cost and complexity. Because of the double number of satellites, integrity and availability requirements are much more easily met, and more robust solutions are achieved. However, system differences between coordinate and time references must be taken into account. This means that coordinate transformations (see above) must be performed, and the time difference between US and Russian UTC must be determined. This difference is an additional unknown, which requires an additional equation, i.e. range measurement.

Every GLONASS receiver naturally contains a frequency synthesiser producing all the frequencies needed to receive the different carrier frequencies from the satellites. This functional requirement is taken care of by the manufacturer and usually does not bother the user of code-measurement techniques. However, the situation gets more complicated when the user wants to perform carrier-phase double-difference measurements as this means using differences between phases measured at two different frequencies. The problem can be solved [6] by e.g. introducing a "synthetic" wavelength λ^s based on Eq. (1) above and letting

$$\lambda^s = \lambda_i / k_i = \lambda_j / k_j \qquad (4)$$

where λ_i and λ_j are the two carrier wavelengths (i.e. $\lambda_i = c/ f_{L1}$) and k_i and k_i are integers. Inserting Eq. (1) in Eq. (4)

$$k_i = 2848 + m_i \qquad (5)$$

because 1602/0.5625 = 2848. With this modification, double-difference solutions can be obtained in exactly the same way as with GPS.

Measurement results

A general observation, as already indicated above, is that GLONASS measurements give position accuracies as GPS with SA off (Figs. 1 - 3). A number of different publications confirm that (e.g. [7]). The figures also show that ignoring coordinate system differences (i.e. using WGS-84 positions as references) when using GLONASS means introducing biases of some metres (Fig.1). Another observation is that GLONASS has lower DOP values at high latitudes (Fig. 4).

Summary and conclusions

A description of GLONASS has been given with emphasis on a comparison with GPS. A summary of this comparison is given in Table 1.

Parameter	*GLONASS*	*GPS*
Number of spacecraft (S/C)	21 + 3	21 + 3
Number of orbital planes	3	6
Orbit inclination	64.8°	55°
Orbit altitude	19100 km	20180 km
Period of revolution	11 h. 15 min.	12 h.
Ephemeris representation	position, velocity and acceleration in earth-centered earth-fixed coord.	Kepler parameters
Datum	PZ-90	WGS-84
Time reference	UTC(Russia)	UTC(NO)
Almanac:		
length	152 bits	120 bits
duration	12.5 min.	2.5 min.
	day of validity	week of validity
	channel number	S/C identifier
	eccentricity	eccentricity
	inclination	inclination
	equator time	almanac time
	validity of almanac	health
	equatorial longitude	right ascension
	--	RA rate of change
	period of revolution	sq. root of semi-major axis
	argument of perigee	argum. of perigee
	--	mean anomaly
	luni-solar term	—
	time offset	time offset
	--	frequency offset
Signalling	FDMA	CDMA
Carrier frequencies:		
L1	1598.0625 - 1604.25 MHz	1575.42 MHz
L2	7/9 L1	60/77 L1
Type of PN-code	ML	Gold
No. of code elements:		
C/A	511	1023
P	33554431	$2.35 \cdot 10^{14}$
Code rate:		
C/A	0.511 Mbits/s	1.023 Mbits/s
P	5.11 Mbits/s	10.23 Mbits/s
Crosscorrelation interference	-48 dB	-21.6 dB
Navigation message:		
rate	50 bits/s.	50 bits/s.
modulation	BPSK Manchester	BPSK NRZ
total length	2.5 min.	12.5 min.
subframe length	30 s.	6 s.

TABLE 1. Comparison of GPS and GLONASS parameters.

GLONASS is today a mature system, although constantly being improved. This year launching of a new generation of satellites ("GLONASS-M") has commenced. The new satellites carry improved cesium clocks, but the system providers are even aiming at using masers in the satellites in the future. The design lifetime of the new satellites is now five years.

Moving the system frequency band downwards and compressing it has also started and will be completed in 2005 when all carriers will be below 1604.25 MHz. Band compression is achieved by use of same frequencies for antipodal satellites. Other means to reduce out-of-band interference with other systems include better filtering.

Other improvements to be implemented are inclusions of additional information in the navigation message, e.g. propagation time differences between "L1" and "L2" in each satellite and differences between GPS and GLONASS system time. Transmission of the "C/A"-code at "L2" is also being discussed. This would greatly improve user ionospheric correction possibilities.

As more receiver manufacturers and users are taking interest in GLONASS, the question of harmonising GPS and GLONASS time and coordinate references arises. A recommendation has been given by the Consultative Committee for the Definition of the Second (CCDS) suggesting among other things that [8]

- the reference times (modulo 1 second) be synchronised as close as possible to UTC,
- the coordinate reference frames be transformed in conformity with the International Terrestrial Reference Frame (ITRF) maintain by the International Earth Rotation Service (IERS),
- both GLONASS and GPS receivers are used at timing centres.

(The WGS-84 is within about half a metre of the ITRF.)

References

[1] Anodina, T: *The GLONASS system technical characteristics and performance.* ICAO FANS Working Paper No. 75, May 1988.

[2] Russian Space Forces' Home Page, *http://mx.iki.rssi.ru/SFCSIC/SFCSIC_main.html*

[3] Johnson, N.L.: *GLONASS spacecraft.* GPS World, pp. 51, Nov. 1994.

[4] Ivanov, N., Salischev, V., Vinogradov, A.: *Ways of GLONASS system advancing.* Proceedings of the ION GPS-95, pp. 991, Sept. 1995.

[5] Letter from the Internavigation Research & Technical Centre (Moscow) to the Chairman of RTCM-104, 1996.

[6] Rossbach, U., Hein, G.: *DGPS/GLONASS carrier phase solutions.* Proceedings of the DSNS-96, paper no. 47; St.Petersburg, May 1996.

[7] MIT/Lincoln Lab. Home Page, *http://vega.atc.ll.mit.edu*

[8] Allan, D.W.: *Harmonizing GPS and GLONASS.* GPS World, pp. 51, May 1996.

APPENDIX

Printouts from INTERNET

Status and Almanac Data

Please register your email address to receive timely notice of significant *TrueGNSS* news and events.

Current GNSS Satellite Status

- GLONASS Constellation Status (text/nagu.txt or ftp://truegnss.com/pub/status/nagu.txt) or http://mx.iki.rssi.ru/SFCSIC/NAGU.TXT) contains official Russian military information about the GLONASS constellation status.
- INIC GLONASS Status (text/inic.txt or ftp://truegnss.com/pub/status/inic.txt) contains GLONASS status information from the Russian civilian Intergovernment Navigation & Information Center (INIC).
- GPS Constellation Status (ftp://ftp.navcen.uscg.mil/ftp/gps/status.txt) contains official US Coast Guard information about the GPS constellation status.

Current GNSS Satellite Almanacs

GLONASS almanacs are transmitted from GLONASS satellites and continuously received at 3S Navigation. These almanacs describe the orbit of each GLONASS satellite and indicate which satellites are transmitting valid signals. A snapshot of the live GLONASS almanac is placed on this web site once each week in four different formats.

- The Leeds-format GLONASS almanac (text/glo-led.txt or ftp://truegnss.com/pub/almanacs/glo-led.txt) is a tabular listing of the contents of the GLONASS almanac message in a format originally presented by Professor Peter Daly at the University of Leeds.
- The AL3 format GLONASS almanac (text/glo-al3.txt or ftp://truegnss.com/pub/almanacs/glo-al3.txt) is used by the System Effectiveness Model Version 3.5 (SEM3.5) orbit prediction program from ARINC.
- The Yuma format GLONASS almanac (text/glo-yum.txt or ftp://truegnss.com/pub/almanacs/glo-yum.txt) is used a number of US DoD orbit prediction programs.
- The raw GLONASS almanac (text/glo-dat.txt or ftp://truegnss.com/pub/almanacs/glo-dat.txt) contains almanac data bits transmitted by the GLONASS satellites printed in the output format of 3S Navigation *TrueGNSS* receivers.
- The zip compressed GLONASS almanacs (text/glo-alm.zip or ftp://truegnss.com/pub/almanacs/glo-alm.zip) contains all four of the above almanac files (glo-led, glo-al3, glo-yum and glo-dat) compressed by pkzip Version 2.04g.
- An archive of past GLONASS almanac data is available at the FTP site:

ftp://truegnss.com/pub/almanacs/archive.

GPS Almanacs are available from the US Coast Guard in two formats: the AL3 format GPS almanac (ftp://ftp.navcen.uscg.mil/ftp/gps/current.al3) and the Yuma format GPS almanac (ftp://ftp.navcen.uscg.mil/ftp/gps/current.alm).

Home ¦ Company Profile ¦ Products ¦ Technical Data ¦ FAQ ¦ Contact 3S ¦ Links

Contents of "Status and Almanac Data" was last revised on November 12, 1997

Coordinational Scientific Information Center of the Russian Ministry of Defence

UPDATED

Events

Fast access

CSIC RSF page.

International Cooperation page.

GLONASS page.

The all questions connected with GLONASS system using should be addressed to:

Coordinational Scientific Informational Center of Russian Space Forces

E-mail: sfcsic@mx.iki.rssi.ru
Mail: R U S S I A, 117279, P/O BOX 14

The Links you might be interested in:

Keith Peshak's articles dedicated to satellite navigation

Up to Previous Page

Global Navigation Satellite System
(G L O N A S S)

 *INTRODUCTION to GLONASS.*

 *GENERAL*information on GLONASS.

UPDATED  *Documents and other publications.*

 *GLONASS status.*

 *GLONASS deployment.*

 *Notices Advisories to GLONASS Users (NAGU).*

 *GLONASS almanac*

 *Ephemerides.*

 National *Time and Frequency Service Bulletin*

 *GLONASS archive.*

 *Where one can receive further information about GLONASS?*

Up to main document.

Coordinational Scientific Information Center of Russian Space Forces

COORDINATIONAL SCIENTIFIC INFORMATION CENTER
OF RUSSIAN SPACE FORCES
(CSIC RSF)

The mission of the Coordinational Scientific Information Center is to plan, manage and coordinate the activities on

- use of civil-military space systems (navigation, communications, meteorology etc.);
- realization of Russian and international scientific and economic space programs;
- realization of programs of international cooperation;
- conversional use of military space facilities, as well as to provide the scientific-informational, contractual and istitutional support of these activities.

- **CSIC composition.**
- **CSIC materials.**

Up to main document.

Russian Space Forces
in international cooperation orbit

Russian Space Forces play a leading role in the space activity. Today Russian Space Forces:

- is a general order of the space systems for military and dual purposes;
- accomplish all Russian and international space programs which are carrying out in Russia;
- posses three Russian cosmodromes: *Plesetsk*, *Svobodny* (under construction), *Bayconur* (territory of Kazakhstan).

Up to main document.

1.STATUS Information Group

GLONASS Constellation Status
(January 08, 1998)

GLONASS number	Cosmos number	Plane/ slot	Frequ. chann.	Launch date	Intro date	Status	Outage date
758	2275	3/18	10	11.04.94	04.09.94	unusable	23.12.97
760	2276	3/17	24	11.04.94	18.05.94	operating	
767	2287	2/12	22	11.08.94	07.09.94	operating	
770	2288	2/14	9	11.08.94	04.09.94	unusable	20.11.97
775	2289	2/16	22	11.08.94	07.09.94	operating	
762	2294	1/4	12	20.11.94	11.12.94	operating	
763	2295	1/3	21	20.11.94	15.12.94	operating	
764	2296	1/6	13	20.11.94	16.12.94	operating	
765	2307	3/20	1	07.03.95	30.03.95	operating	
766	2308	3/22	10	07.03.95	05.04.95	operating	
777	2309	3/19	3	07.03.95	06.04.95	withdrawn	24.12.97
780	2316	2/15	4	24.07.95	26.08.95	operating	
781	2317	2/10	9	24.07.95	22.08.95	operating	
785	2318	2/11	4	24.07.95	22.08.95	operating	
776	2323	2/9	6	14.12.95	07.01.96	operating	
778	2324	2/9	11	14.12.95		spare	
782	2325	2/13	6	14.12.95	18.01.96	operating	

Note: All the dates (DD.MM.YY) are given at Moscow Time (UTC+0300)

2.SUMMARY Information Group

SUBJ:GLONASS STATUS 08 JANUARY 98
1.SATELLITES, PLANES, SLOTS AND CHANNELS

Plane 1/ slot:	01	02	03	04	05	06	07	08
Channel:	--	--	21	12	--	13	--	--
Plane 2/ slot:	09	10	11	12	13	14	15	16
Channel:	06	09	04	22	06	--	04	22
Plane 3/ slot:	17	18	19	20	21	22	23	24
Channel:	24	--	--	01	--	10	--	--

2.CURRENT ADVISORIES (ADV) AND FORECASTS (FCST)
INFORMATION IS REPEATED FOR ONE MONTH AFTER EVENT CONCLUDES.
ALL THE DATES (DD.MM.YY) ARE GIVEN AT MOSCOW TIME (UTC+0300)

A. FORECASTS
NAGU-MSG.DATE-TIME-SL/CH-TYPE-SUMMARY
110-971216-16.12.97-1100-18/10-FCST-UNUSABLE 23.12/1140- UNDEFINED

B. ADVISORIES
NAGU-MSG.DATE-TIME-SL/CH-TYPE-SUMMARY
067-970717-17.07.97-1100-19/03-ADVS-UNUSABLE 17.07/0350-UNFINISHED
105-971121-21.11.97-1100-14/09-ADVS-UNUSABLE 20.11/2148-UNFINISHED
106-971201-01.12.97-1100-10/09-ADVS-UNUSABLE 29.11/0446-29.11/0530
107-971203-03.12.97-1100-06/13-ADVS-UNUSABLE 03.12/0500-03.12/0528
108-971211-11.12.97-1100-15/04-ADVS-UNUSABLE 10.12/2213-11.12/0305
109-971211-11.12.97-1100-17/24-ADVS-UNUSABLE 11.12/0832-11.12/0916
111-971222-22.12.97-1100-12/22-ADVS-UNUSABLE 19.12/1204-UNFINISHED
112-971224-24.12.97-1100-18/10-ADVS-UNUSABLE 23.12/1140-UNFINISHED
113-971224-24.12.97-1100-12/22-ADVS-PUT INTO OPERATION 23.12/1801
114-971229-29.12.97-1100-15/04-ADVS-UNUSABLE 27.12/0217-27.12/0317
115-980108-08.01.98-1100-17/24-ADVS-UNUSABLE 07.01/1400-07.01/1500

C. GENERAL: NO IMPACT, INFORMATIONAL PURPOSE ONLY
NAGU-MSG.DATE-TIME-SL/CH-TYPE-SUMMARY

3.NAGU Information Group

067-970717
NOTICE ADVISORY TO GLONASS USERS (NAGU) 067-970717

SUBJ:19/03 (777) UNUSABLE SINCE 17.07/0350 MT
1.CONDITION: 19/03 (777) UNUSABLE SINCE 17.07/0350 MT
(UTC+0300) UNTIL FURTHER NOTICE DUE TO MAINTENANCE
2.USERS ARE REMINDED TO UPDATE ALMANACS IF NECESSARY
3.POC:CSIC RSF AT +7-095-333-81-33

105-971121
NOTICE ADVISORY TO GLONASS USERS (NAGU) 105-971121
SUBJ:14/09 (770) UNUSABLE SINCE 20.11/2148 MT
1.CONDITION: 14/09 (770) UNUSABLE SINCE 20.11/2148 MT
(UTC+0300) UNTIL FURTHER NOTICE DUE TO MAINTENANCE
2.USERS ARE REMINDED TO UPDATE ALMANACS IF NECESSARY
3.POC:CSIC RSF AT +7-095-333-81-33

106-971201
NOTICE ADVISORY TO GLONASS USERS (NAGU) 106-971201
SUBJ: 10/09 (781) UNUSABLE 29.11/0446-29.11/0530 MT
1.CONDITION: 10/09 (781) WAS UNUSABLE SINCE 29.11/0446
UNTIL 29.11/0530 MT (UTC+0300) DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33

107-971203
NOTICE ADVISORY TO GLONASS USERS (NAGU) 107-971203
SUBJ: 06/13 (764) UNUSABLE 03.12/0500-03.12/0528 MT
1.CONDITION: 06/13 (764) WAS UNUSABLE SINCE 03.12/0500
UNTIL 03.12/0528 MT (UTC+0300) DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33

108-971211
NOTICE ADVISORY TO GLONASS USERS (NAGU) 108-971211
SUBJ: 15/04 (780) UNUSABLE 10.12/2213-11.12/0305 MT
1.CONDITION: 15/04 (780) WAS UNUSABLE SINCE 10.12/2213
UNTIL 11.12/0305 MT (UTC+0300) DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33

109-971211
NOTICE ADVISORY TO GLONASS USERS (NAGU) 109-971211
SUBJ: 17/24 (760) UNUSABLE 11.12/0832-11.12/0916 MT
1.CONDITION: 17/24 (760) WAS UNUSABLE SINCE 11.12/0832
UNTIL 11.12/0916 MT (UTC+0300) DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33

110-971216
NOTICE ADVISORY TO GLONASS USERS (NAGU) 110-971216
SUBJ: FORECAST OUTAGE 18/10 (758) 23.12/1140- UNDEFINED
1.CONDITION: 18/10 (758) IS SHEDULED TO BE UNUSABLE
ON 23.12/1140 UNTIL FURTHER NOTICE DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33

111-971222
NOTICE ADVISORY TO GLONASS USERS (NAGU) 111-971222
SUBJ:12/22 (767) UNUSABLE SINCE 19.12/1204 MT
1.CONDITION: 12/22 (767) UNUSABLE SINCE 19.12/1204 MT
(UTC+0300) UNTIL FURTHER NOTICE DUE TO MAINTENANCE
2.USERS ARE REMINDED TO UPDATE ALMANACS IF NECESSARY
3.POC:CSIC RSF AT +7-095-333-81-33

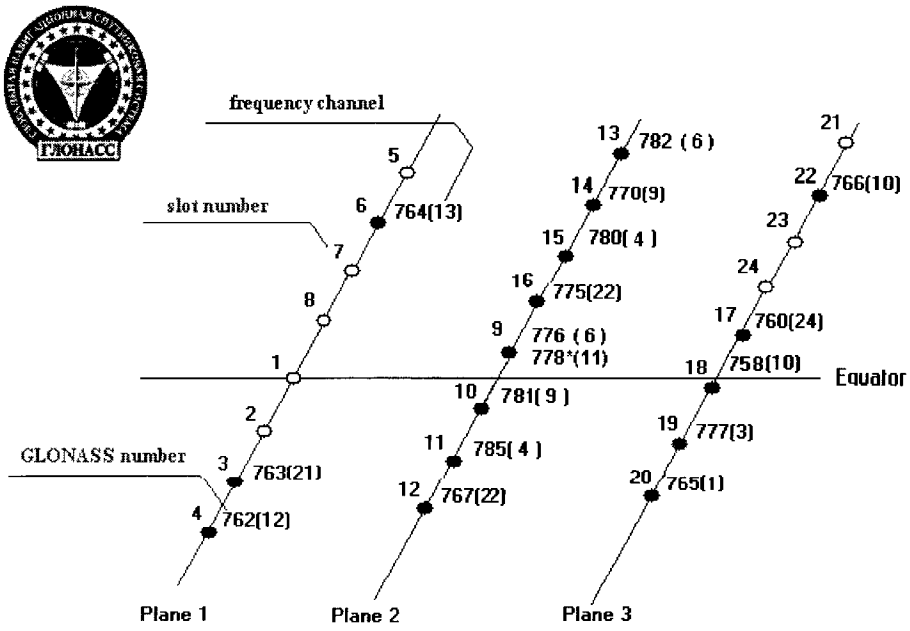
112-971224
NOTICE ADVISORY TO GLONASS USERS (NAGU) 112-971224
REF: NAGU 110-971216
SUBJ:18/10 (758) UNUSABLE SINCE 23.12/1140 MT
1.CONDITION: 18/10 (758) UNUSABLE SINCE 23.12/1140 MT
(UTC+0300) UNTIL FURTHER NOTICE DUE TO MAINTENANCE
2.USERS ARE REMINDED TO UPDATE ALMANACS IF NECESSARY
3.POC:CSIC RSF AT +7-095-333-81-33

113-971224
NOTICE ADVISORY TO GLONASS USERS (NAGU) 113-971224
SUBJ: 12/22(767) PUT INTO OPERATION 23.12/1801 MT

1.CONDITION: 12/22(767) WAS PUT INTO OPERATION ON
23.12/1801 MT (UTC+0300)
2.USERS ARE REMINDED TO UPDATE ALMANACS IF NECESSARY
3.POC:CSIC RSF AT +7-095-333-81-33

114-971229
NOTICE ADVISORY TO GLONASS USERS (NAGU) 114-971229
SUBJ: 15/04 (780) UNUSABLE 27.12/0217-27.12/0317 MT
1.CONDITION: 15/04 (780) WAS UNUSABLE SINCE 27.12/0217
UNTIL 27.12/0317 MT (UTC+0300) DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33

115-980108
NOTICE ADVISORY TO GLONASS USERS (NAGU) 115-980108
SUBJ: 17/24 (760) UNUSABLE 07.01/1400-07.01/1500 MT
1.CONDITION: 17/24 (760) WAS UNUSABLE SINCE 07.01/1400
UNTIL 07.01/1500 MT (UTC+0300) DUE TO MAINTENANCE
2.POC:CSIC RSF AT +7-095-333-81-33



778*(11) -spare

GLONASS Deployment (October 24, 1997)

The all questions connected with GLONASS system using should be addressed to:

R U S S I A, Moscow, Profsoyuznaya ul., 84/32,
Coordinational Scientific Informational Center of Russian Space For- ces

E-mail: sfcsic@mx.iki.rssi.ru
Mail: R U S S I A, 117279, P/O BOX 14

Up to main document.

- About GLONASS **(Updated 15 December 1997)**
- GLONASS Status Report **(Updated 15 December 1997)**

	GLONASS	GPS	GPS+GLONASS
CONSTELLATION	● (11k GIF)	● (19k GIF)	
HEALTH / AVAILABILITY	● (12k GIF)	● (13k GIF)	
POSITION ESTIMATES (Scatter Plots)			
Single Receiver	● (10k GIF)	● (13k GIF)	● (13k GIF)
Local-Area Differential	● (13k GIF)	● (13k GIF)	● (13k GIF)
Wide-Area Differential			
US. Coast Guard Differential		● (13k GIF)	
VELOCITY ESTIMATES (Scatter Plots)			
Single Receiver	● (13k GIF)	● (17k GIF)	● (13k GIF)
POSITION ESTIMATES (Data)	● (Text)	● (Text)	● (Text)
ALMANACS (Yuma Format)	● (Text)	● (Text)	
SATELLITE VISIBILITY			● (11k GIF)

- GLONASS Developments **(Updated 15 December 1997)**
- Recent MIT/LL Reports on GPS & GLONASS **(Updated January 1998)**
- GPS+GLONASS User Equipment **(Updated 15 December 1997)**

For comments and suggestions send email to *satnav@atc.ll.mit.edu*

CARRIER PHASE KINEMATIC POSITIONING: FUNDAMENTALS AND APPLICATIONS

M.E. Cannon
Department of Geomatics Engineering
The University of Calgary
Calgary, Alberta
CANADA

ABSTRACT

GPS kinematic techniques have proven to be an effective means to accurately position a wide variety of moving platforms to accuracies on the order of a few centimetres. The completion of the satellite constellation as well as improvements in receiver technologies and processing algorithms has fueled tremendous interest which has resulted in many new applications of GPS carrier phase methods. The following paper discusses the various measurements which are used to obtain cm-level accuracies and presents a general algorithm for integer ambiguity resolution. Issues concerning the implementation of precise kinematic positioning in real-time are then outlined. Some applications of kinematic GPS are given and these include structural monitoring, precision farming and water level profiling. The paper concludes with a discussion on GPS attitude determination which is a special case of kinematic positioning. Airborne results with wing flexing are used to illustrate the capability of GPS to provide attitude parameters.

1.0 INTRODUCTION

There has been a tremendous growth in the interest in precise GPS kinematic positioning over the past few years which has mainly been due to the maturing of this technique from a purely research system to one which can be used in operational environments. Early feasibility tests clearly showed the capability of GPS to provide cm-level accuracies for a number of different platforms and environments, see Krabill et al. (1989), Mader & Lucas (1989), Henderson & Leach (1990), Colombo (1991). With the completion of the satellite constellation and improvements in receiver technologies and processing algorithms, the performance of on-the-fly (OTF) ambiguity resolution has increased and commercial systems for post-mission and real-time kinematic positioning have been developed.

Kinematic positioning has been accepted within the broad GPS community as the case when the precise carrier phase measurement is used in the positioning model. Figure 1 gives an overview of the various positioning modes of GPS and

the associated accuracies when operating under reasonable conditions (i.e. favorable geometry and relatively low multipath). As can be seen, when using the carrier phase measurement, accuracies on the order of a few cm can be obtained when integer ambiguities are resolved, whereas decimetre accuracies are generally achievable if so-called 'floating' ambiguities are estimated. This second case is typically for longer monitor-remote separations (i.e. greater than 25 km) or under significant multipath or unstable atmospheric conditions.

As Figure1 indicates, the achievable accuracy when using the code observable in DGPS mode may approach near-decimetre levels of accuracies when using narrow correlator-type technology as well as carrier smoothing. Improvements in code correlation technology are discussed in the following section and encompass improved multipath elimination technologies as well as measurement resolution. Single point GPS performance has also been improved for the post-mission case by utilizing precise orbit and clock information which are computed using International GPS Service (IGS) data typically by federal government agencies (e.g. Natural Resources Canada). In some cases, the availability of precise orbit and clock data has been sufficient to meet accuracy requirements previously met by differential processing.

The following paper discusses the various measurements which are used to obtain cm-level accuracies and presents a general algorithm for integer ambiguity resolution. Issues concerning the implementation of precise kinematic positioning in real-time are then outlined. Some applications of kinematic GPS are given. The paper concludes with a discussion on GPS attitude determination which is a special case of kinematic positioning. Airborne results with wing flexing are used to illustrate the capability of GPS to provide attitude parameters.

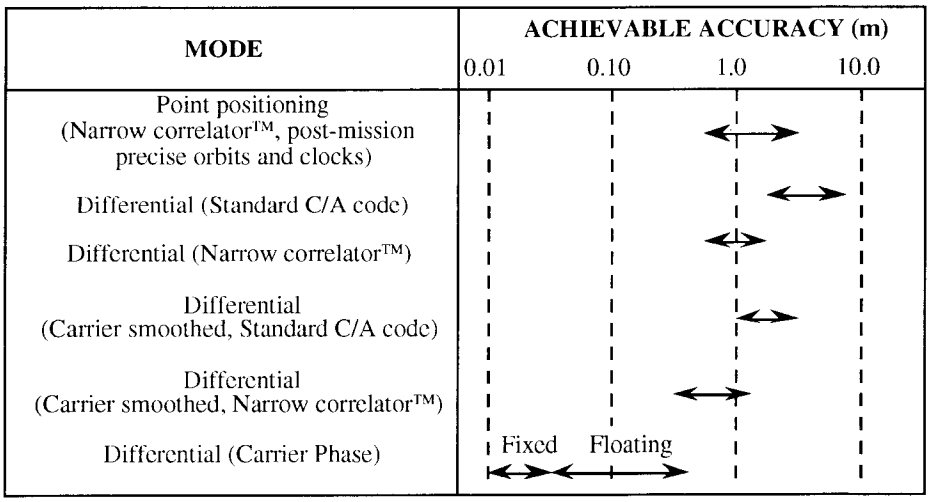


Figure 1
GPS Operating Modes and Achievable Accuracies.

2.0 KINEMATIC POSITIONING DATA AND MODELS

2.1 Measurement Data and Observation Equations

The L1 carrier phase measurement has the inherent advantage that it can be measured extremely accurately, typically in the range of 0.2 mm for a high end geodetic receiver, to 1-2 mm for a low cost system. The main reason for the increased noise in the lower cost receivers is that the tracking bandwidth is wider to be able to track the lower quality oscillator that is used. Zero baseline test, i.e. one antenna and two receivers, can be used to test these noise values.

The L1 C/A code data can be measured with an accuracy of 0.1 - 1.5 m, depending on the tracking software implemented in the receiver. For example, standard C/A code receivers will measure the code to about 1.5 m while narrow-correlator type receivers can measure towards the 0.1 m level (Fenton et al., 1991). Accurate code data can play an important role during the integer ambiguity resolution phase of kinematic positioning as will be discussed in Section 3. Once integer ambiguities are resolved, code measurements become less critical, except to maintain quality assurance.

Although kinematic positioning can be performed using single frequency data, many commercial systems employ the use of dual frequency receivers which obtain the L2 carrier phase from a codeless or semi-codeless technique (e.g. Ashjaee & Lorenz, 1992). The main advantage of having dual frequency data is that it allows the use of the widelane, i.e. the difference in phase between L1 and L2, which forms a new measurement with a wavelength of 86 cm. Clearly, it is much easier to resolve an integer ambiguity of 86 cm compared to the L1 wavelength of 19 cm. The disadvantage of using this L2 data is that it is derived without full use of the Y code, which means that there is a degradation in measurement signal strength, hence quality. This degradation is a function of the specific technique which is used, but can vary between 14 and 30 db compared to the Y code L2 carrier phase. A practical problem that this may introduce is that the L2 data is more susceptible to interference than the L1 data. Nevertheless, the widelane measurement is extremely effective in reducing the time to ambiguity resolution. Due to the fact that the widelane is much more noisy compared to L1, most commercial OTF systems use the widelane to resolve ambiguities quickly and then switch over to use L1 to achieve the highest level of accuracy.

Virtually all OTF processing systems utilize the double difference observable since it eliminates receiver clock errors and greatly reduces spatially correlated errors (see Cannon, 1991 for further details). The equation is given as

$$\Delta\nabla\Phi = \Delta\nabla\rho + \Delta\nabla d\rho + \lambda\Delta\nabla N - \Delta\nabla dion + \Delta\nabla dtrop + \varepsilon\Delta\nabla\Phi \qquad (1)$$

where	$\Delta\nabla$	is the double difference operator,
	Φ	is the carrier phase measurement (m),
	ρ	is the range between the receiver, x_r, and the satellite, x^s, i.e., $\mid x^s - x_r \mid$ (m)
	dp	is the orbital error (m),
	λ	is the measurement wavelength (m),
	N	is the integer ambiguity (cycles),
	dion	is the ionospheric error (m),
	dtrop	is the tropospheric error (m),
and	ε	is the measurement error (m).

The spatially correlated atmospheric and orbital errors are virtually eliminated when the monitor and rover receivers are close together, say within 20-30 km. However, once the separation starts to exceed these distances, integer ambiguity resolution may require significantly more time, and at some limit will not be possible. It should also be noted that the ε term includes multipath which is not spatially correlated so it does not reduce in differential mode. This implies that there can be situations where the monitor-rover separation are very close, but due to multipath, the integer ambiguities cannot be resolved.

2.2 Processing Strategies

There are two modes of processing double difference carrier phase data, one is called fixed ambiguity processing and the other is called floating ambiguity processing. In the second case, carrier phase ambiguities are estimated to be real numbers and are included in the system state vector. Over time, it is expected that these real numbers will approach the true integer ambiguity. Depending on the system, there may or may not be an attempt to actually fix the ambiguities to real numbers. Two examples where this may apply are for long monitor-rover separations where integer ambiguities cannot generally be found (say > 60 km) and secondly for a single frequency system which does not have the advantage of widelaning. Many systems that operate in fixed ambiguity mode, process data in floating mode until the integer ambiguities can be found. In floating mode, achievable accuracies are on the order of several cm to a few decimetres, while fixed ambiguity system can ultimately deliver cm accuracies.

3.0 CARRIER PHASE INTEGER AMBIGUITY RESOLUTION

Carrier phase ambiguity resolution is the key component to any kinematic positioning system, since it is only after the integer ambiguities are resolved that the highest level of accuracy can be obtained. The time required to resolve integer ambiguities is generally a function of the following:

(1) Measurement used (i.e. L1 versus L1-L2)
(2) Satellite geometry
(3) Separation between monitor and rover receivers
(4) Environment (i.e. multipath and signal blockage)
(5) Receiver characteristics

As discussed in Section 2.1, dual frequency receivers provide the capability of using the widelane observable during the ambiguity resolution process. This alone can reduce the time to resolution by a factor of five to ten (Lachapelle et al., 1992). Satellite geometry is also critical since it directly impacts the information content of the data. At least five satellites are needed for ambiguity resolution so that there is some measurement redundancy, however better performance can be achieved if there are six to eight satellites visible above ten or 15 degrees elevation. The separation between the monitor and rover receiver will influence the magnitude of residual atmospheric and orbital errors remaining in the data after differencing and at long distances (e.g. >50 km), ambiguity resolution will require more time and at some point may not be possible.

Multipath is especially a problem for ambiguity resolution since it does not cancel out in the differencing process, thus a systematic error with a period of a few to several minutes remains in the data.

Receiver characteristics are important when selecting a unit for kinematic applications. Noise and signal tracking capability are two of the more important characteristics, with signal tracking being especially critical when operating in sub-optimal environments (e.g. around foliage).

3.1 Steps for Integer Ambiguity Resolution

The following is intended to give an overview of the concepts of integer ambiguity resolution. Many investigators have developed techniques that vary from the steps below, however the fundamental concepts generally remain the same. The six main steps to integer ambiguity resolution are:

At the first epoch....

(1) Code DGPS solution to an approximate position
(2) Computation of search range for each integer ambiguity

At the first and following epochs...

(3) Calculation of measurement residuals for each integer combination
(4) Statistical testing of squared residuals to reject outliers

(5) Statistical testing of remaining squared residuals to determine if integer identification is possible

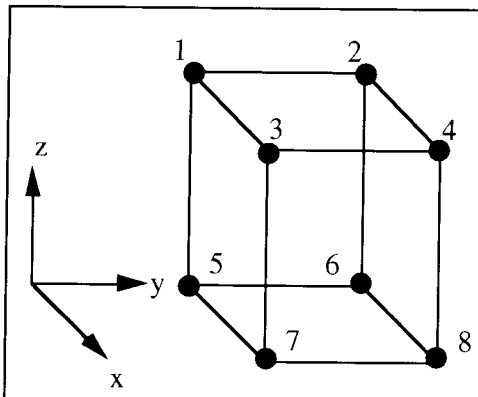
(6) If (5) fails, sum squared residuals with previous epoch, go to next epoch and then repeat steps (3) - (5)

The differential code position is used to define an approximate solution for the rover receiver and its uncertainty is then used to bound the search space for potential integer ambiguity combinations. For this reason, it is important to estimate an accurate DGPS which will ultimately improve computational efficiency. In most OTF implementations, a narrow correlator - type receiver is utilized. To give an example of the correlation between DGPS position accuracy and search efficiency, if the position accuracy is ± 5 m, the number of L1 ambiguities to be considered could reach tens of thousands whereas a position accuracy of ± 2 m would give less than six thousand if three double difference integer ambiguities are searched (corresponding to four satellites). It is important to note here that it is really only required to search three double difference ambiguities since the remaining integer ambiguities are dependent on these values. This concept of selecting four primary satellites was first introduced by Hatch (1990) as a means to improve computational efficiency and entails the use of the four satellites with the best geometry (i.e. lowest PDOP). When these are combined with the remaining redundant secondary satellites, residuals can be computed to determine the fit of the particular integer combination.

The integer ambiguity search range is defined as those integers which give double differenced positions that fall within the bounds defined from the code differential position. Since this search space must contain the true position, the estimated standard deviations output from the DGPS solution are conservatively scaled (e.g. to at least 3 σ). One quick means of defining the ambiguity search range is to define the search space as a cube and then to compute the ambiguities which would result at each of the eight nodes of the cube. These would typically be real numbers, but could be rounded to the next integer as shown in the Figure 2.

Once the range of integer ambiguities has been computed, the algorithm must then select which integer ambiguity combination is in fact the correct one. This selection process is based on the goodness of fit of a particular integer ambiguity combination to the data. One of the difficulties during the identification phase is to balance the reliability of the selection (i.e. ensuring that the correct result has been obtained) with the time required to make the decision. An analogy that can be used to illustrate this point is shown in Figure 3 where a series of points are drawn. If three lines are fit to the data on the left of the vertical line, all three lines look like they fit the data about the same. The fit in this case can be a simple visual check of the residuals (differences between the data points and the line) in each case. It is only after more data is considered to

the right of the vertical line, is the correct solution clear. This is the same for integer ambiguity resolution - if only a short data span is considered then many integer ambiguity combinations will fit the data. Only after the satellites have moved in their orbit, thus changing the geometry, will the information content be sufficient to isolate the correct integer ambiguity combination.



'Real' ambiguities are computed at each node:

$$\Delta\nabla N_k^{ij} = \Delta\nabla\rho_k^{ij} - \Delta\nabla\Phi^{ij} \quad k = 1..8$$

and the largest and smallest values are retained. The range in integers for each ambiguity is then:

$$\Delta\nabla N_{min}^{ij} \text{ to } \Delta\nabla N_{max}^{ij}$$

Min and max 'real' ambiguities are rounded to the farthest integer to be conservative

Figure 2
Computation of Integer Ambiguity Search Range

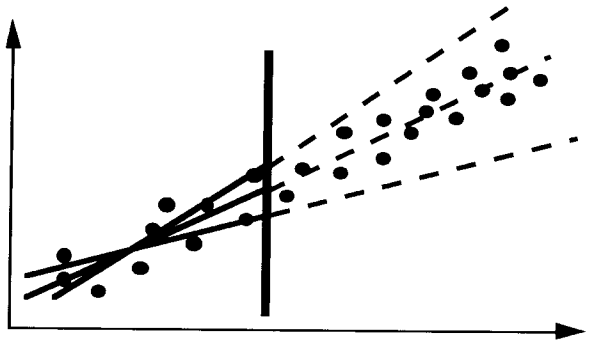


Figure 3
Integer Ambiguity Analogy

The fundamental value that is used in ambiguity identification is the sum of squared measurement residuals, $\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}$, where $\mathbf{v}$ is the vector of residuals and $\mathbf{C}_l$ is the covariance matrix of the measurements. This implies that there must be measurement redundancy so at least five satellites must be continuously tracked during the ambiguity resolution phase. Since the best fit of a least squares adjustment is the one in which the sum of squared residuals is minimized, it is

reasonable to select the integer ambiguity combination which meets this criterion. If the errors in the observations are Gaussian and the integer ambiguity combination is correct, $\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}$ will have a Chi-square distribution (Koch, 1989), and can therefore be tested using the following:

$$\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v} \leq \chi^2_{f,1-\alpha} \tag{2}$$

where $\chi^2_{f,1-\alpha}$ is the Chi-square percentile corresponding to the degrees of freedom f, and the confidence level 1-α.

If a particular integer ambiguity combination fails to satisfy eqn. (2), it is rejected from the pool of potential integer combinations. It should be noted that the test is dependent on the selection of the measurement accuracy (i.e. embedded in C_l), so this value should be chosen carefully and conservatively. Typical values are 0.5-2 cm for an L1 phase measurement and 2-4 cm for a widelane observable.

Potential integer ambiguity combinations which satisfy eqn. (2) are retained and further tested with respect to one another. The correct integer ambiguities should give the smallest $\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}$, however it must also be sufficiently smaller than the other the values for the other combinations so there is statistical certainty that the correct decision is being made. There are a number of approaches which can be used for this test, and a simple one is just to compute the ratio between the smallest and second smallest $\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}$, i.e.

$$\text{Ratio} = \frac{\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}_{\text{second minimum}}}{\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}_{\text{minimum}}} \tag{3}$$

If this ratio is greater than a pre-defined threshold, then the integer ambiguities associated with the smallest $\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}$ will be considered correct.

In most cases, eqn. (3) cannot be fulfilled in one measurement epoch. In fact this can often lead to incorrect integer ambiguity selection. In order to improve the reliability of selection, the integer ambiguity combinations passing eqn. (2) can be retained for the next epoch and the new $\mathbf{v}^T\mathbf{C}_l^{-1}\mathbf{v}$ are summed with the previous values. In this way, information can be accumulated over time and the ratio test is then performed on the summed residuals.

Once ambiguities are resolved, continuous checking must be performed to ensure that the correct solution has in deed been found. One method to do this is

to monitor the magnitude of the measurement residuals and if they become too large, then the ambiguity resolution process is restarted.

There have been a number of investigations into improvements in ambiguity resolution over the past several years. These studies include Chen (1993), Tuenissen (1994), Landau & Euler (1???). Each of these techniques are similar in the sense of using the residuals as an indicator for selection of integer ambiguities. Performance in terms of efficiency and reliability may differ between the methods however.

4.0 REAL-TIME IMPLEMENTATION

A number of groups have developed and tested real-time kinematic systems, see for example Frodge et al. (1994) and Blomenhofer et al. (1994). Both of these systems attempt to determined the integer ambiguities using dual frequency data. Ford and Neumann (1994) describe the RT20 which is a system using single frequency carrier phase data which estimates the ambiguities to be real numbers.

When implementing a kinematic positioning system in real-time, GPS data (or alternatively corrections) must be transmitted to the rover receiver using a communications link. The type of link used will depend on the distances between the monitor and rover receivers as well as the general operating conditions (e.g. shading). Many applications use UHF or spread spectrum links which can give line-of-sight coverage of about 30 km and 5 km, respectively. The spread spectrum data links do not require a license to operate and repeaters can also be used to extend the usable range. Other services are also being developed, such as FM subcarrier, to provide carrier phase data to a wider user base. Although not currently in place, it is anticipated that there will eventually be satellite services which transmit carrier phase data.

Most commercial groups that have developed real-time kinematic systems have used proprietary messages to transmit the raw GPS data or corrections. RTCM standards for the transmittal have been developed, and it is expected that manufacturers will adopt these standards once real-time kinematic systems become more widespread (RTCM,1994).

A real-time system under development at The University of Calgary is described in Lan and Cannon (1996). In this system, spread spectrum data links are used to transmit single frequency raw data (RTCM message types 18 and 19) or corrections (message types 20 and 21). To date, the system has been tested with NovAtel model 3951 receivers, but it is currently being upgraded to handle data from the new MiLLenium dual frequency sets. Tests conducted to date include static tests over short baselines as well as land kinematic tests conducted on the UofC test range and in conjunction with the precision farming project.

Accuracies at the few cm have been achieved. Typical times to resolve integer ambiguities vary between 10-60 seconds.

5.0 KINEMATIC POSITIONING APPLICATIONS AND RESULTS

Three applications of kinematic GPS are described below and include precision farming, structural monitoring and water level profiling. Further kinematic positioning results can be found in Tiemeyer et al. (1994), Shi and Cannon (1995), Shi (1994) and Lachapelle et al. (1992).

5.1 Precision Farming

In the past, farmers generally treated an entire field with the same quantity and type of fertilizer, without regard for possible variations in topography, salinity and soil type within the field. By applying a variable amount and type of fertilizer as a function of location within the field, the overall crop productivity can be improved. This is called *precision farming*. The fertilizer application is determined from a prescription map derived from information previously collected. As the fertilizer prescription is applied to smaller parcels of land, the overall economic gain is improved since local conditions are taken into account. The process begins with the harvesting process where the crop yield is monitored as a function of location. The resulting yield map is used, together with other soil information such as salinity, to prepare a prescription map which is used during the fertilizing process to optimize spreading.

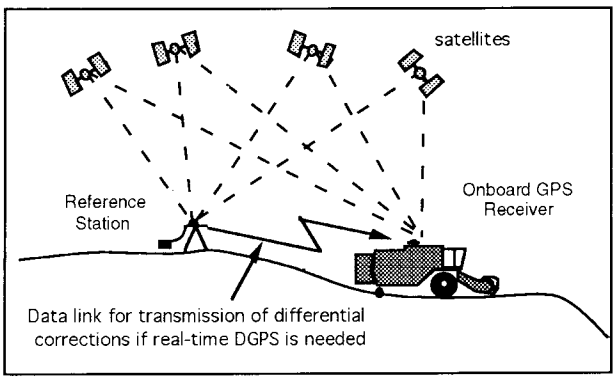


Figure 4
Positioning of Combine Harvester with DGPS

The University of Calgary, together with Alberta Agriculture, is currently developing a system which can be used to measure the variability of crop production as a function of location (Lachapelle et al., 1994a). The system consists of several components which include GPS receivers for positioning, a yield monitoring system which outputs the instantaneous Bu per acre, an EM conductivity meter for salinity measurements, and soil samples for

determination of soil types and nutrients. The DGPS/yield monitoring data can be collected under normal combining operations, see Figure 4 for the concept. As the fertilizer moves along a pre-determined trajectory, the fertilizer prescription for the current cell is retrieved from the Geographic Information System (GIS).

The precise positioning system needed to assign spatial coordinates to the field measurements must have an accuracy of 1 m or better and should be available in real-time on a practically continuous basis (< 1 second). GPS can meet, in differential mode, these requirements using relatively low cost user equipment. Position updates are available at a rate of several times per second if required. Two 10-channel C/A code narrow correlator spacing NovAtel GPSCard™ sensors were selected as they have shown to provide sub-metre accuracy in previous field tests using a robust carrier phase smoothing of the code approach.

Four test fields across Alberta ranging from 80 to 200 acres have been selected for study over a four year period. These areas vary in soil type as well as topography and salinity conditions. The project began during the 1993 harvest and is expected to continue until 1997. Positioning results are presented in Table 1 for one of the test sites which consisted of gently rolling hills. The reference station was installed near the field and the moving platform was operating within a few km from the reference station. The crop was harvested on September 21. On November 9, soil samples were taken at various locations using an All-Terrain-Vehicle (ATV). The ambiguities could be resolved OTF on September 21 and November 9 using typically 10 to 15 minutes of data with a choke ring groundplane at the reference station only. This is to be expected when using single frequency data. On each day, a minimum of six satellites were tracked and the theoretical PDOP was always smaller than 3. Data was collected at a rate of 1 Hz. Although cycle slips were occasionally detected on low satellites, the tracking stability of the GPSCard™ was fully satisfactory.

Table 1 gives comparisons between coordinates at crossover points during each of the test days, as well as between the two test days. The results show that by using carrier smoothing, a repeatable accuracy of about 0.5 m or better can be achieved, whereas accuracies below 15 cm can be reached using carrier phase kinematic processing. Although the OTF results exceed the current positional requirements, the agricultural community is interested in these levels of accuracy to generate topographic information.

The use of GIS and GPS in agriculture is expanding rapidly and the precision farming case reported here is only one of several applications which include salinity measurements (Cannon et al. 1994), aerial and terrestrial crop spraying and seeding, and animal tracking. In the future, the extension of this technology to unmanned farm vehicles is projected to occur.

Table 1
RMS Height Differences at Crossover Points using Carrier Phase Smoothing of the Code and OTF Solutions.

Date	Carrier Smoothing (m)	OTF (m)
Sept. 21	0.25	0.07
Nov. 9	0.56	0.08
Sept. 21 vs Nov. 9	0.43	0.14

5.2 Structural Monitoring

The use of precise GPS techniques for structural monitoring has become increasingly widespread as an alternative to measurements using accelerometers, laser interferometers or electronic distance measurement (EDM) instruments. For many applications, kinematic GPS processing algorithms must be used due to the relatively high frequency of the movement. One such example is the deformation monitoring of a tall structure as described in Lovse et al. (1994). Figure 5 shows the Calgary Tower which is a 160 m structure built in 1968 to a design specification of 165 mm movement in a 160 km/h wind.

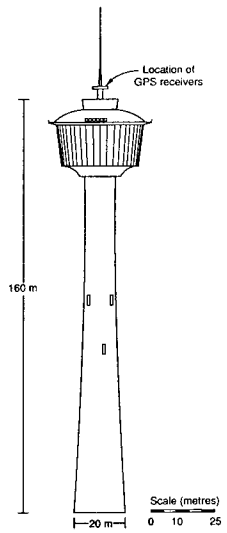


Figure 5
Calgary Tower

In order to measure the actual measurement of the tower, a GPS receiver was placed near the base of the main communication antenna above the observation deck. The GPS reference receiver was situated in a stable location on the roof of

an apartment building approximately 1 km from the tower. Single frequency receivers were used in the project since the baseline separation was relatively short.

Data was collected for 15 minutes at a 10 Hz rate. A high data rate is required in these types of applications since the deformation may have a frequency of up to several Hz. During the particular session in which the data was recorded, the winds were typically 60 km/h from the west, however gusts of up to 100 km/h were noted. The GDOP varied between 2 and 3, and there were 8 satellites observed above a 10 degree elevation cutoff.

In order to process the data in post-mission, OTF techniques were utilized. About 3 minutes of data was needed to resolve the integer ambiguities after which the double difference measurement residuals were below 5 mm. Figure 6 shows a representative plot of the tower movement over a 1 minute time interval.

The movement of the tower is larger in the north-south direction even though the wind was from the west. This is due to the eddies that are produced on the north and south sides of the tower which create zones of slightly different pressure which tend to push the higher pressure side and pull on the lower pressure side. As well, the wind is out of phase with the east-west movement of the tower which dampens movement in that direction.

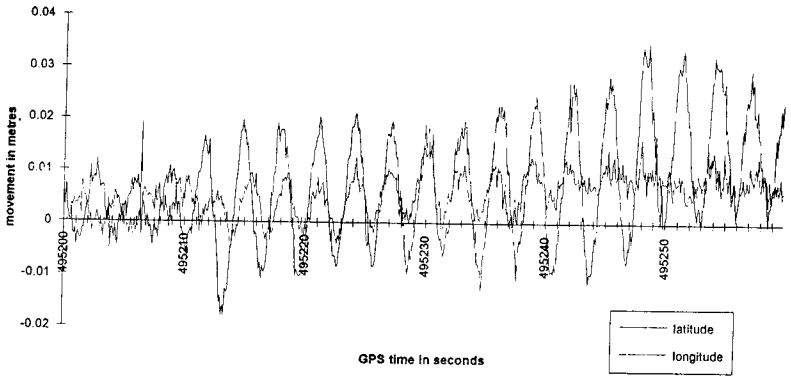


Figure 6
Movement of the Calgary Tower.

Although there is no external reference to check the motion as depicted in Figure 6, a visual analysis shows that the movement is regular and within the

expected range of such a structure. The north-south motion had an amplitude of approximately ± 15 mm and an east-west component of ± 5 mm.

5.3 Water Level Profiling

Precise knowledge of water levels is essential for tidal studies and other hydrographic purposes such as the establishment of chart datums. GPS offers the possibility of determining water level profiles with cm-level accuracy using carrier phase measurements on-the-fly. Accurate Bench Marks (B.M.'s) can also be established along the shores if an accurate geoid model is available. This is an important development since levelling operations may be prohibitive in remote areas such as the McKenzie River in Northern Canada. A feasibility study was conducted by The University of Calgary in conjunction with the Canadian Hydrographic Service (CHS) to obtain water level profiles and establish B.M.'s along an 80 km segment of the Fraser River, British Columbia. A first order levelling line along the river was used to assess the accuracy of the GPS-derived orthometric heights. Figure 7 shows a map of the area.

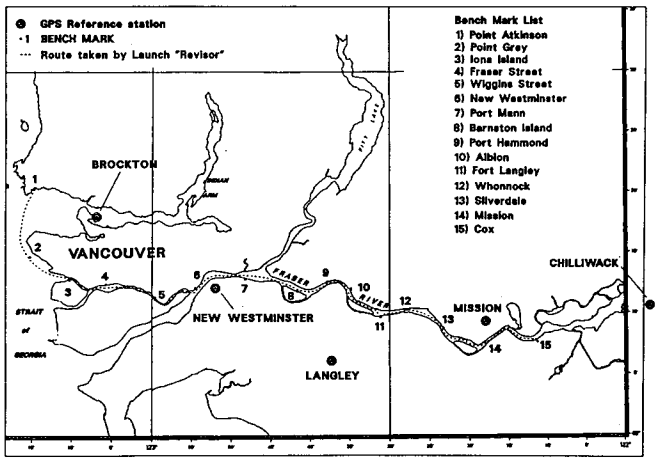


Figure 7
Fraser River GPS Water Level Profiling Survey

Results from the test are given in Lachapelle et al. (1993) and show that an accuracy of about 5.5 cm (RMS) could be achieved between the GPS-derived orthometric heights and the levelled B.M. height when kinematic OTF techniques are utilized with dual frequency Ashtech receivers. Part of the error budget is due to the height transfer method (accurate to about 2 cm) as well as the geoid. Overall, the feasibility study was successful and the GPS-derived water level profiles are being analysed to determine the extent of tidal and other effects. Table 2 gives a summary of the statistics derived from the comparisons.

Table 2
Summary Statistics of GPS-derived and Levelled Heights

RMS agreement between successive GPS height determinations ($\Delta T \leq 200$ s) at each B.M. visit.	5.2 cm
RMS agreement between successive B.M. visits (several hours $< \Delta T < 3$ days).	5.5 cm
RMS agreement between GPS-derived and levelled (B.M.) heights.	9.3 cm
Geoid undulation bias (ΔN) estimated by comparing GPS and levelled heights.	7.1 cm
RMS agreement between GPS-derived and levelled (B.M.) heights (Geoid bias removed)	6.0 cm

6.0 ATTITUDE DETERMINATION

The ability of GPS to provide accurate aircraft attitude components has been demonstrated using several platforms and a variety of operational conditions. The use of GPS-derived attitude for both photogrammetric and remote sensing applications are some of the main driving forces to test the feasibility of the system. Many of the tests that have been conducted to date utilize a dedicated attitude determination system comprising of a multi-channel receiver which has a bank of channels dedicated to each of the supported antennas, which is typically four (e.g. Ferguson et al., 1991; Schwarz et al., 1992). The advantage of this type of system is that all channels are driven from the same oscillator which means that all carrier phase measurements have a common clock offset to provide an additional degree of freedom in the determination of attitude. Results using these dedicated attitude determination systems show that arcminute-level accuracies can be achieved depending on antenna separation, e.g. van Graas and Braasch (1991), Cohen and Parkinson (1992) and Schade et al. (1993).

An alternative to the above is to use a so-called non-dedicated attitude determination system composed of three or more independent GPS receivers mounted on the platform. One advantage of such a system is flexibility since the receivers can be used for a variety of applications in addition to attitude determination, e.g. Sun (1994). Cost-effectiveness may also be gained through the utilization of low-cost GPS receivers which output the carrier phase observable. Marine tests conducted with this type of system confirm an achievable accuracy of 1 to 2 arcminutes for antenna separations of up to 40 m (Lachapelle et al., 1994b) and several arcminutes for shorter separations (Lu et al., 1993; McMillan et al., 1994).

6.1 Computation of Attitude Parameters

Roll, pitch and heading components can be determined using carrier phase measurements from three or more antennas (Lu et al.,1993). These antennas are generally placed as far apart as possible in order to maximize the achievable accuracy. The two main error sources that are of concern in GPS attitude determination are multipath effects as well as carrier phase noise. Antenna phase centre stability may also be a factor.

In order to define the platform attitude, a body frame, must be realized by three antennas. This is shown on Figure 8 for a four antenna case. The body frame can be measured directly using a theodolite or can be determined by GPS initialization, which is typically more convenient. Pre-measured distances between the GPS antenna pairs can be used as constraints in the attitude determination algorithm to eliminate incorrect carrier phase integer ambiguities during the search phase. This is very powerful constraint and greatly improves the time to ambiguity resolution. Refer to Lu (1995) for details on carrier phase integer ambiguity resolution for the attitude determination case.

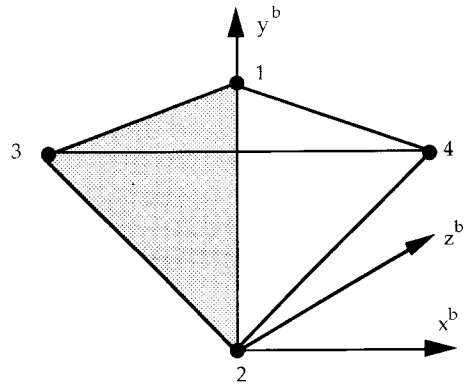


Figure 8
Body Frame Defined by Three GPS Antennas

Once integer ambiguities are resolved, precise interstation vectors between the various antenna pairs can be coimputed. Attitude components are then estimated via a least squares approach using the interstation vectors between antennas as quasi-observables. Suppose $r_i^b = (x_i^b, y_i^b, z_i^b)^T$ are the body-frame coordinates of the i-th antenna which were previously estimated. The measurements are $r_i^n = (x_i^n, y_i^n, z_i^n)^T$, the local level coordinates of the i-th antenna, which are determined from the differential GPS carrier phase solution. These coordinates satisfy the following equation

$$\begin{pmatrix} r_2^b \\ r_3^b \\ r_4^b \end{pmatrix}^T = R_n^b(\varphi,\theta,\psi) \begin{pmatrix} r_2^n \\ r_3^n \\ r_4^n \end{pmatrix}^T \qquad (4)$$

where $R_n^b(\varphi,\theta,\psi)$ is the transformation matrix between the body-frame coordinates and the local-level frame coordinates, and

$$R_n^b(\varphi,\theta,\psi) = \begin{pmatrix} c(\psi)c(\varphi)-s(\psi)s(\theta)s(\varphi) & s(\psi)c(\varphi)+c(\psi)s(\theta)s(\varphi) & -c(\theta)s(\varphi) \\ -s(\psi)c(\theta) & c(\psi)c(\theta) & s(\theta) \\ c(\psi)s(\varphi)+s(\psi)s(\theta)c(\varphi) & s(\psi)s(\varphi)-c(\psi)s(\theta)c(\varphi) & c(\theta)c(\varphi) \end{pmatrix} \quad (5)$$

where c() is a cosine function and s() is a sine function. When there are three antennas on the platform, a unique solution is generated, whereas a fourth antenna provides redundancy. These equations can be solved using a least squares adjustment model by minimizing the cost function

$$J(\varphi,\theta,\psi) = \left\| (r^b - R(\varphi,\theta,\psi)r^n) \right\|^2 \quad (6)$$

The least squares method has many advantages over other methods such as a direction computation of attitude (Lu et al., 1993). It can easily accommodate more antennas and attitude is less effected by multipath from a single antenna since it is based on a least squares fit of all antenna positions. Further details on the methodology used in the attitude determination algorithms are given in Lachapelle et al. (1994) and Lu (1995).

6.2 Aircraft Results

A series of flight tests were conducted by The University of Calgary and Sandia National Laboratories during February 1 to 4, 1994 near their facility located on Kirtland Air Force Base in Albuquerque, New Mexico. The tests used a DeHavilland Twin Otter aircraft operated for Sandia by the Department of Energy's Ross Aviation. Three single frequency GPS antennas were installed on the aircraft, one on each wing and one near the tail, in addition to the dual frequency GPS antenna already in place just aft of the cockpit. The locations of the four GPS antennas, as measured by a theodolite, are shown in Figure 9.

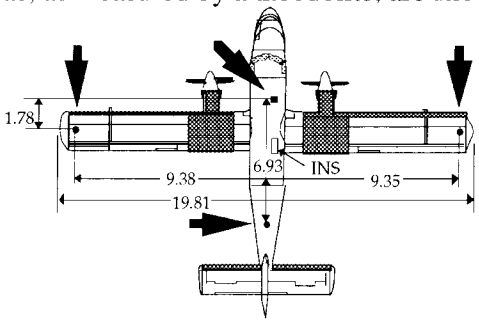


Figure 9
Aircraft Antenna Locations

The suite of test equipment in the aircraft included: 1) four NovAtel GPSCard™ receivers, each housed in a portable personal computer; 2) a Honeywell ring

laser gyro assembly (RLGA) inertial measurement unit (IMU); 3) a Sandia Airborne Computer (SANDAC) to implement the navigation equations; and 4) a Texas Instruments (TI) embedded P-code GPS receiver integrated with the SANDAC. The RLGA IMU and SANDAC were mounted on the floor of the aircraft just forward of the main cabin door. The RLGA has long term drift rates commensurate with 1 to 2 nautical miles per hour navigation accuracy, with individual gyro specifications of: 0.01 deg/hr bias, random walk less than 0.008 deg/$\sqrt{h}$ and scale factor less than 1 part-per-million.

Four portable computers containing the GPSCards™ were mounted in the flight racks, with each receiver connected to one of the four GPS antennas. The TI embedded P-code receiver was connected to the dual frequency forward fuselage GPS antenna in parallel with one of the NovAtel receivers. The TI receiver's 1 pulse per second interrupt was used to time tag the SANDAC/RLGA navigation and attitude measurements to GPS time to an accuracy of a few milliseconds.

A static test was performed in order to compute the relative positions between the four aircraft antennas before the flight tests commenced. In the following, results from the February 1 test are reported where the aircraft was flown near the airport under low dynamic conditions. During this flight, the maximum pitch of the aircraft was about 10 degrees while the absolute maximum roll of the aircraft was about 40 degrees. GPS data was logged at 1 Hz while the IMU data was logged at 4 Hz.

8.2.1 Wing Flexure Model

Due to wing flexure of the aircraft, the body-frame defined above is not a fixed rigid body frame. Since the frame is changing with the wing flexure, the derived attitude is relative to a different coordinate frame. Wing flexure is especially prevalent during the take-off and landing phases.

In order to obtain attitude with respect to one fixed coordinate frame, the wing flexure has to be removed before attitude is computed. A wing flexure model was considered here. Wing flexure is constrained in the z- component in the body frame. That is

$$r_i^b = r_i^{b0} - B_f f \tag{7}$$

where

$$B_f = (0,0,1)^T \tag{8}$$

and f is a scalar amount which is estimated in the least square adjustment.

When considering all four antennas, the body frame coordinates and the local level coordinates should satisfy the following relation

$$\begin{pmatrix} r_2^b \\ r_3^b \\ r_4^b \end{pmatrix}^T = \begin{pmatrix} r_2^{b0} \\ r_3^{b0} \\ r_4^{b0} \end{pmatrix}^T - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} f \quad . \tag{9}$$

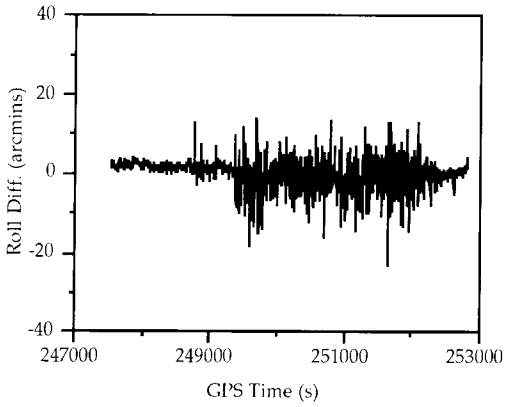
The solution is obtained by minimizing the cost function

$$J(\varphi,\theta,\psi,f) = \left\| (r^{b0} - Bf) - R(\varphi,\theta,\psi) r^n \right\|^2 . \tag{10}$$

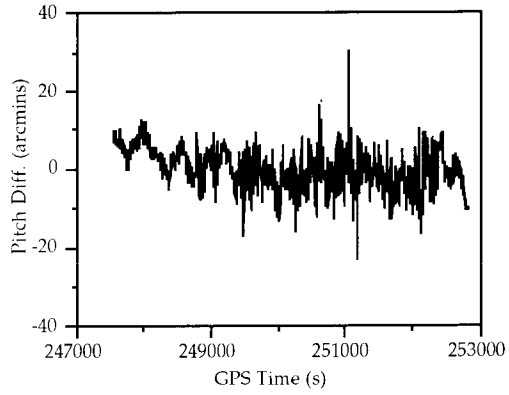
It should be noted that when wing flexure is estimated, at least four antennas are required on the platform since an additional parameter is estimated in addition to roll, pitch and heading. Refer to Cohen et al. (1993) for a similar approach to flexure modelling.

8.2.2 GPS Attitude Accuracy

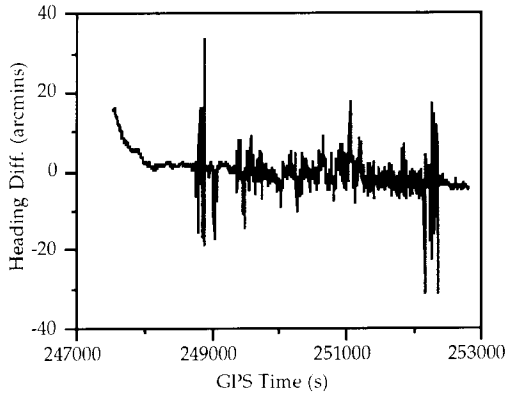
Figure 10 illustrates the agreement between the GPS and INS attitude angles for the flight test once the wing flexure model is applied. The estimated wing flexure was at the level of 12 cm. Overall, the agreement between the two systems is consistent within 10 arcminutes, however at times 248750 s and 252250 s there are fluctuations in the heading agreement at the level of approximately ± 20 arcminutes. These fluctuations occur when the aircraft makes a sharp turn on the ground before take-off and after landing. The RMS agreement between the GPS and INS attitude angles is 4.1 arcminutes for roll, 5.3 arcminutes for pitch and 4.9 arcminutes for heading. Similar results were obtained in for the other test days. See Cannon and Sun (1996) for details.



(a)



(b)



(c)

Figure 10
Differences Between Attitude from a GPS Multi-Antenna System and an INS When Using a Wing Flexure Model

CONCLUSIONS

This paper outlined the various measurements and modes of operation for GPS kinematic surveying. Carrier phase integer ambiguity on-the-fly was discussed in detail as it forms the basis of any kinematic positioning system. Performance of the OTF technique is dependent on the receivers used, monitor-rover separation, environmental conditions as well as satellite geometry. In particular, significant performance differences will be seen when using dual frequency receivers compared to single frequency units.

An overview of real-time carrier phase processing was presented. For these applications, the selection of communications link is critical since the data rate and reliability will dictate the level of accuracy and availability that is achieved. Theoretically, performance in real-time should mirror that obtained in post-mission.

Several applications of GPS kinematic positioning were given and these include structural monitoring, water level profiling as well as precision farming. In each case, cm-level accuracies could be obtained after integer ambiguities were resolved. These applications are meant to give an overview of the various uses of the system. It is expected that the number of applications will grow exponentially over the coming years as the cost of the technology decreases.

REFERENCES

Ashjaee, J and R. Lorenz (1992), Precision GPS Surveying After Y-Code, Proceedings of ION GPS 92, Albuquerque, New Mexico, September 16 - 18, pp. 657-659.

Blomenhofer, H., G.W. Hein, E.T. Blomenhofer and W. Werner (1994), Development of a Real-time DGPS System in the Centimetre Range, Proceedings of PLANS' 94, IEEE, New York, pp. 532-539.

Cannon, M.E. (1991), Airborne GPS/INS with an Application to Aerotriangulation, Report No. 20040, Department of Surveying Engineering, The University of Calgary.

Cannon, M.E. and H. Sun (1996), Assessment of a Non-Dedicated GPS Receiver System for Precise Airborne Attitude Determination, Photogrammetry and Remote Sensing, Journal of the ISPRS, Vol. 51, pp. 99-108.

Chen, D. (1993), Fast Ambiguity Search Filter (FASF): A Novel Concept for GPS Ambiguity Resolution, Proceedings of GPS-93, The Institute of Navigation, Alexandria, VA, pp. 781-787.

Cohen, C.E. and B.W. Parkinson (1992), Aircraft Applications of GPS-Based Attitude Determination, Proceedings of ION GPS-92, Albuquerque, September 16-18, pp. 775-782.

Cohen, C.E., B.D. McNally and B.W. Parkinson (1993), Flight Tests of Attitude Determination Using GPS Compared Against an Inertial Navigation Unit, Navigation, Journal of the Institute of Navigation, Vol. 41, No. 1, pp. 83-97.

Colombo, O. L. (1991), Errors in Long Distance Kinematic GPS, Proceeding of ION GPS 91, Albuquerque, New Mexico, Sept. 11 - 13, pp. 673-680.

Fenton, P.C., W.H. Falkenberg, T.J. Ford, K.K. Ng and A.J. Van Dierendonck (1991), NovAtel's GPS Receiver - the High Performance OEM Sensor of the Future, Proc. of ION GPS-91, Albuquerque, Sept. 11-13, pp. 49-58.

Ferguson, K., J. Kosmalska, M. Kuhl, J.M. Eichner, K. Kepski and R. Abtahi (1991), Three Dimensional Attitude Determination with the Ashtech 3DF 24 Channel GPS Measurement System, Proceedings of the ION NTM, Phoenix, January 22-24, pp. 35-41.

Ford, T. and J. Neumann (1994), NovAtel's RT20 - A Real-time Float Ambiguity Positioning System, Proceedings of the ION GPS-94, Salt Lake City, pp. 1067-1076.

Frodge, S., V. Labrecque, R. Barker (1995), Performance of the Real-time On-the-Fly GPS Positioning System on Board a Dredge, Proceedings of the ION National Technical Meeting, Anaheim, January 18-20, pp. 505-512.

Hatch, R. (1990), Instantaneous Ambiguity Resolution, Proceedings of IAG International Symposium No. 107 on Kinematic Systems in Geodesy, Surveying and Remote Sensing, Springer Verlag, New York, pp. 299-308.

Henderson, T and M. Leach (1990), An Assessment of the Absolute Accuracy of Long-baseline Kinematic Vehicle, Proceeding of ION GPS 90, Colorado Springs, Colorado, Sept. 19-21, pp. 91-100.

Krabill, W.B. C.F. Martin and R.N. Swift (1989), Applying Kinematic GPS to Airborne Laser Remote Sensing, Proceedings of ION GPS-89, Colorado Springs, Colorado. September 27 - 29, pp. 39 - 43.

Lachapelle, G., C. Liu, G. Lu, W. Qiu and R. Hare (1993), Water Level Profiling with GPS, Proceedings of the ION GPS-93, Salt Lake City, September 22-24, pp. 1581-1587.

Lachapelle, G., M.E. Cannon, H. Gehue, T. Goddard and D. Penney (1994a), GPS System Integration and Field Approaches in Precision Farming, Navigation, Vol 41, No. 3, pp. 323-335.

Lachapelle, G., G. Lu, and B. Loncarevic (1994b), Precise Shipborne Attitude Determination Using Wide Antenna Spacing, Proceedings of the International Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation - KIS94, Banff, August 30-September 2, pp. 323-330.

Lachapelle, G., M.E. Cannon and G. Lu (1992), A comparison of P Code and High Performance C/A Code GPS Receivers for On The Fly Ambiguity Resolution, Bulletin Géodésique, Springer Verlag, New York, Vol. 67, No. 3, pp. 185-192.

Lan, H. and M.E. Cannon (1996), Development of a Real-time Kinematic System: System Design, Performance and Results, Proceedings of the ION National Tehnical Meeting, Santa Monica, January 22-24, pp.605-613.

Landau, H. and H.J. Euler (19???),

Lovse, J.W., W.F. Teskey, G. Lachapelle and M.E. Cannon (1995), Dynamic Deformation Monitoring of a Tall Structure Using GPS Technology, Journal of Surveying Engineering, ASCE, Vol. 121, No. 1, pp. 35-40.

Lu, G. (1995), Development of a GPS Multi-Antenna System for Attitude Determination, Report No. 20073, Department of Geomatics Engineering, The University of Calgary.

Lu, G., M.E. Cannon, G. Lachapelle, and P. Kielland (1993), Attitude Determination in a Survey Launch Using Multi-Antenna GPS Technologies,

Proceedings of the ION National Technical Meeting, San Francisco, January 20-22, pp. 251-260.

Mader, G.L. and J.R. Lucas (1989), Verification of Airborne Positioning Using Global Positioning System Carrier Phase Measurements, Journal of Geophysical Research, Vol. 94, No. B8, pp. 10175 - 10181.

McMillan, J.C., D.A.G. Arden, G. Lachapelle, and G. Lu (1994), Dynamic GPS Attitude Performance Using INS/GPS Reference, Proceedings of the ION GPS94, Salt Lake City, September 21-23, pp. 675-682.

RTCM (1994), Recommended Standards for Differential NAVSTAR GPS Service, Version 2.1, RTCM SC-104, Washington, DC.

Schade, H., M.E. Cannon and G. Lachapelle (1993), An Accuracy Analysis of Airborne Kinematic Attitude Determination with the NAVSTAR/Global Positioning System, SPN Journal, Vol. 3, No. 2, pp. 90-95.

Schwarz, K.P., A. El-Mowafy and M. Wei (1992), Testing a GPS Attitude System in Kinematic Mode, Proceedings of ION GPS92, Albuquerque, September 16-18, pp. 801-809.

Schwarz, K.P., M.E. Cannon, R.V.C. Wong (1989), A Comparsion of GPS Kinematic Models for the Determination of Position and Velocity along a Trajectory, Manuscripta Geodaetica, Vol. 14, pp.345-353.

Shi, J. (1994), High Accuracy Airborne Differential GPS Positioning Using a Multi-Receiver Configuration, Report No. 20061, Department of Geomatics Engineering, The University of Calgary.

Shi, J. and M.E. Cannon (1995), Critical Error Effects and Analysis in Carrier Phase-based Airborne GPS Positioning Over Large Areas, Bulletin Geodesique,??????

Sun, H. (1994), Integration of INS with Multiple GPS Antennas for Airborne Applications, Proceedings of the ION GPS94, Salt Lake City, September 21-23, pp. 1401-1409.

Tiemeyer, B., M.E. Cannon, G. Lachapelle and G. Schanzer(1994), High Precision Aircraft Navigation with Emphasis on Atmospheric Effects, Proceedings of PLANS' 94, IEEE, New York, pp. 394-401.

Tuenissen, P. (1994), A New Method for Fast Carrier Phase Ambiguity Estimation, Proceedings of the IEEE PLANS, Las Vegas, April.

Van Graas, F. and M. Braasch (1991), GPS Interferometric Attitude and Heading Determination: Initial Flight Test Results, Navigation, Journal of The Institute of Navigation, Vol. 38, No. 4, pp. 297-316.

Three Years of Continuous Observations in the SWEPOS Network

Jan M. Johansson Hans-Georg Scherneck *

Abstract

The SWEPOS system of permanently operating GPS stations, established in 1993, has been designed, devised and furnished as a joint effort between the National Land Survey of Sweden and Onsala Space Observatory, Chalmers University of Technology. Here we describe the operations within SWEPOS with geophysical purpose to detect crustal motions in Fennoscandia. For this purpose a project named BIFROST was created; BIFROST stands for Baseline Inferences for Fennoscandian Rebound Observations, Sea-level and Tectonics. We show solutions of site positions obtained from 1000 days of operation of SWEPOS. We determine their variations in time, discerning them from plate or frame orientation, and discuss a number of perturbation effects. First results are presented, indicating movements which generally support the notion of a dominating displacement pattern due to the postglacial rebound of Fennoscandia. However, deviations exist. In order to discern regional movements of a presumably tectonic origin the coverage of the region must be extended, both concerning the areas that neighbor Sweden and array densification within the country. We foresee observing operations of at least ten years if deformation rates of 0.1 mm/yr are to be concluded at a 95 percent confidence level.

Background

The Swedish permanent GPS network (SWEPOS) is the result of proposals, originally presented by different organizations, merged together to support

*both authors at Onsala Space Observatory, Chalmers University of Technology, S-439 92 ONSALA, Sweden, phone +46 31 7725500, fax +46 31 7725590, Johansson, e-mail: jmj@oso.chalmers.se

a large variety of GPS applications. In the early 90's the National Land Survey (NLS) of Sweden introduced the idea of establishing an experimental network of permanent reference sites for GPS. The intentions were to facilitate test projects regarding routine surveying with GPS. A network of five stations, not necessarily equipped with continuously operating GPS receivers, was established in 1992.

Independently, a project initiated by the Smithsonian Astrophysical Observatory (SAO) and Onsala Space Observatory (OSO) received funding from the NASA program Dynamics of the Solid Earth (DOSE). This project, initially proposed in March 1991, targeted on the use of annual GPS field measurements to provide accurate determination of the three-dimensional pattern of crustal motion over a wide geographic area in Fennoscandia for the investigation of glacial isostatic adjustment, mantle viscosity, ice models, and sea level rise.

In 1992, the Onsala Space Observatory submitted to the Swedish Council for Planning and Coordination of Research (FRN) a proposal to establish a permanently operating GPS network as part of the geophysical projects and for further improvements of the space geodesy techniques and atmospheric remote sensing using GPS. With strong support from NLS funding were obtained for the purchase of 15 high precision GPS receivers. Additional funds were also provided by Knut and Alice Wallenberg's Foundation.

Through an extensive collaborative effort between OSO and NLS, merging of the individual plans became possible. The first sites to have continuously operating GPS receivers were those established by NLS and used during the test period of the International GPS Service for Geodynamics (IGS). Additional sites were established during the summer of 1993 and an experimental phase began in August 1993. In December 1994, TERACOM-Svensk Rundradio AB joined the SWEPOS project. The differential pseudorange corrections produced at the sites could then be broadcast via the FM-network with an almost complete national coverage. This service, called EPOS, provides means for real-time horizontal position accuracy of about 2 m.

In this paper we describe the present status (August, 1996) of the network, the data analysis, and products. Finally, we describe different applications of the network and present results from a couple of research project based on the data from the SWEPOS and other permanent networks in northern Europe.

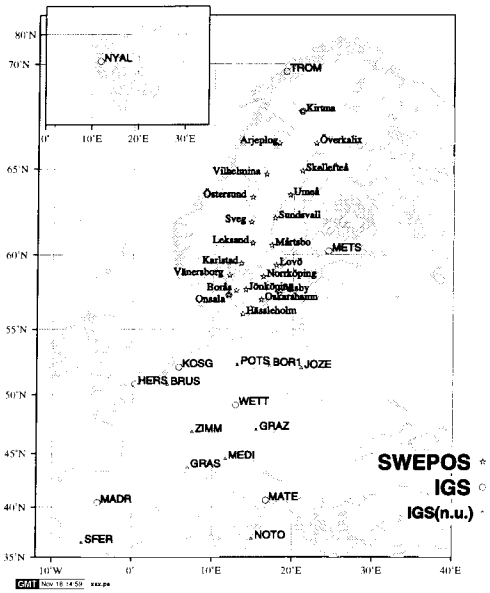


Figure 1: The SWEPOS Permanent Swedish GPS Network and the European IGS sites regularly included ("n.u."—not included) in the solutions derived for the BIFROST project

SWEPOS sites design and equipment

Here we describe the actual realization in SWEPOS of the antenna, its monumentation and other site facilities in conjunction with the instrument cabin. We will discuss the cases where problems have been detected, the solutions we devised and problems that might still persist. A map of the network is shown in Figure 1.

The SWEPOS pillar

The SWEPOS network is furnished with a standard 3 m tall concrete steel-enforced pillar of circular cross-section. The standard pillar design is due to the National Land Survey (cf. Figure 2). The Division of Structural Mechanics (Doc Ola Dahlblom) at Lund Technical University has assisted in its development with studies on the thermally induced deformations. The foot of the pillar consists of a 0.75 x 0.75 m concrete plate. The pillar's steel

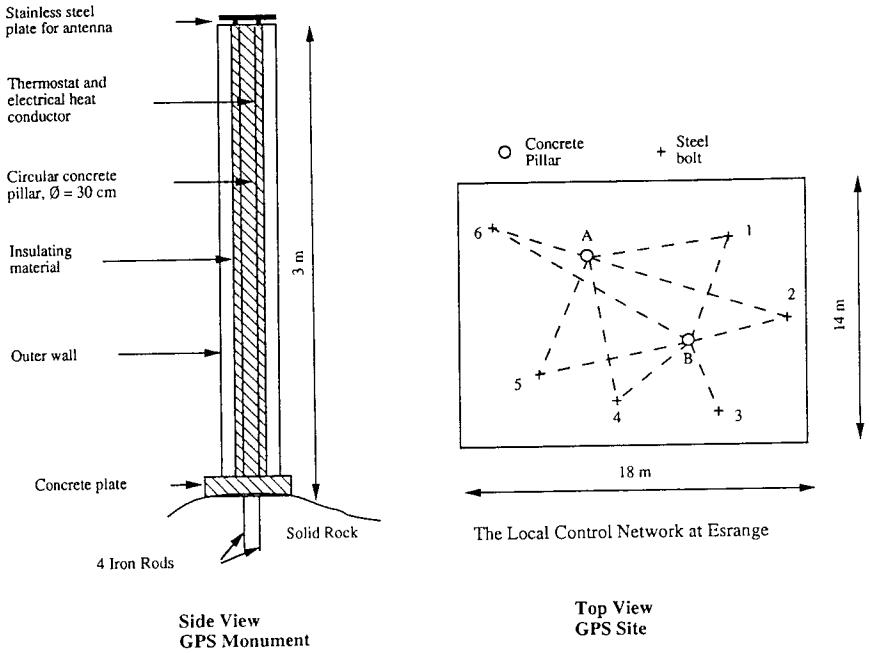


Figure 2: The SWEPOS standard pillar (left). Layout of the Kiruna site near Esrange with two monuments (A, B) and local reference marks. The latter serve as control points for surveying monument movements.

enforcement continues through the bottom plate into holes drilled into the bedrock. The Jönköping pillar is a 1 m lowered version of the standard pillar (for air traffic safety). Lovö and Mårtsbo monuments rest on pre-existing pillars with a rectangular cross-section. Onsala continues to use a 1 m tall pillar with a square cross-section (cf. Fig. 3).

Under an insulating outer cover that extends over the height of the pillar and consists of a helically wound corrugated plastic sheet, rockwool insulating material aids in attaining a homogenous temperature field throughout the pillar. The temperature level is controlled attaching a heating wire to the pillar, helically wound. A temperature sensor is fit into a small cavity inside the pillar and connects to a thermostat unit in the instrument cabin (on-off regulator, powered with AC from the mains).

The pillar set temperature is 15°. Present temperatures exceeding this limit may deform the pillar. At 40° the length change is still below one millimeter. Figure 4 shows a photograph of the Vänersborg site, which is furnished in the standard design.

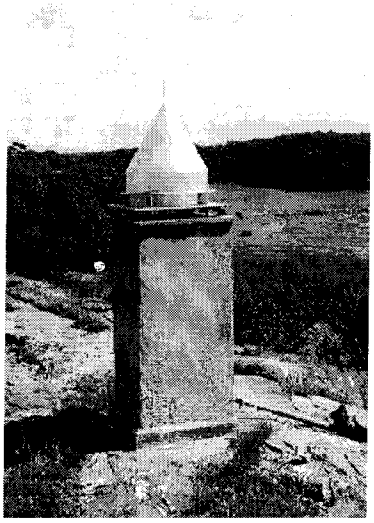


Figure 3: SWEPOS monument at Onsala, old radome. Eccosorb material in-line with the antenna bottom plate has been removed for the photo.

Figure 4: The SWEPOS station at Vänersborg with a SWEPOS standard pillar and conical radome.

Local control network

At ten to twenty metres distance from the monument, geodetic markers (stainless steel studs) are driven into the surface of the bedrock. Usually six such markers, as regularly distributed as possible with regard to healthy bedrock accessibility, form a local network within which eventual deformations of the top of the antenna monument can be detected.

In a control survey, the antenna is removed and replaced with a theodolite. The horizontal and vertical angles to the steel studs are observed. The detection level for horizontal and vertical movement of the pillar within the control network is 0.1 mm.

The National Land Survey carries out the control measurements. Leksand has been used as a control site where the local network has been observed with a few-month intervals during 1993–1995. In general the monuments have been measured and re-measured once in two years. Preliminarily, the surveys seem to indicate movements of one millimeter or less.

The SWEPOS antennas

All sites are equipped with Dorne-Margolin choke ring antennas. Where stations are furnished with the standard pillar a profoundly homogeneous geometry in the aspect of EM propagation, diffraction and scattering has been achieved.

The shorter pillar and the special antenna mount at Onsala has caused some concern as to the wave scattering conditions in its near field. Also, the site is just outside our laboratories so that testing alternate arrangements is less time consuming. Our experiments have converged to the following solution: Scattering perturbations originating from the mount and backplane structure has been attenuated with the help of EM wave absorbing material (Eccosorb AN-W). A $61 \times 61 \times 5.7$ cm square piece has been put directly above the metallic back plate, yielding satisfactory performance (Jaldehag et. al. 1996a).

The SWEPOS radome

After a first winter of experience with a fiberglass radome manufactured by Delft University, a new design was developed in order to avoid snow/ice accumulation. Large (40 mm) vertical variations in the vertical component determinations were observed primarily at the inland sites of northern

Sweden. Jaldehag et al. (1996b) found correlation with precipitation and the local temperature at ground-level. The study suggested that snow was accumulating around the top of the pillar and the mounting elements of the Delft radome which together present many surfaces for snow and ice to attach and settle on. The OSO model was developed from this experience and eliminates edges due to its straight inverted icecream cone design with an aperture wide enough to extend over the top surface of the pillar with an overhang (cf. Fig. 4).

Owing to the electrical properties of the radomes, the transition from the Delft to the OSO model induces a detectable, virtual change in the site position, generally less than 20 mm in the vertical and smaller than that in the horizontal. It was found that this change in position is different at each site and therefore must be determined individually. All materials have some effect on a electromagnetic wave. Radomes appears to delay and refract the GPS-signal in a similar way as snow (Jaldehag et al., 1996b). Independent tests show that a conical cover may cause cm-level vertical errors when the tropospheric delay parameter is estimated.

In June 1996, all con-shaped radomes in the SWEPOS network were replaced with hemispheric plexi-glass radomes. These are mounted so that the center of the sphere coincides with the antenna phase center. The employed hemispheric radome show less elevation dependence since the satellite signal travels the same distance through the radome material independent of elevation angle. The influence on the tropospheric wet delay estimates and subsequently, the vertical component will only be on the 1-2 mm level. We also assume that differential effects due to the excess signal path delay through the radome are canceled out as identical radomes are employed in local or regional type of network.

The SWEPOS receivers

At present all SWEPOS sites house two receivers. Employed models are eight channel TurboRogue and/or twelve channel Ashtech Z-12. Table 1. shows the currently deployed receiver/antenna combinations at all sites. At stations that house two receivers a power splitter connects both receivers to the same antenna.

In the first phase (1993–94) 15 sites were equipped with TurboRogue receivers, all of which connected to DM antennas. At the time AS was switched on by the US DoD, the TurboRogue was the only receiver model

capable of full wavelength L2-measurements. Some stations were equipped with Ashtech P-12 receivers of which the larger part was connected to Ashtech antennas. In order to achieve homogeneity in receiving conditions, which was a primary concern, DM antennas replaced the Ashtech ones in 1995. From July 1, 1995, all 20 (21) stations could be included in the data analysis.

Sites that are used for Differential GPS require the full RTCM capability and Doppler measurements. Also, the sampling rates preferred in this application are high. The receiver matching this purpose is the Ashtech Z-12. The increased number of simultaneously visible satellites has motivated to favour the Ashtech model.

Table 4: Summary of site equipment

site[a]	Latitude[b] [°]	Longitude[b] [°]	Height[b] [m]	receiver(s)[c]	pillar[c]	cable[e] [m]	antenna[f]	remark
Arjeplog (ARJE)	66.31	18.12	489	2 A	3-m CC	15	DM-T	
Borås (BORA)	57.71	12.89	219	A + TR	3-m CC	60	DM-TA	
Hässleholm (HASS)	56.09	13.72	114	A + TR	3-m CC	12	DM-T	DGPS
Jönköping (JONK)	57.75	14.06	260	A + TR	1-m CC	22	DM-T	
Karlstad (KARL)	59.44	13.51	114	A + TR	3-m CC	17	DM-T	DGPS
Kiruna (KIRU)	67.88	21.06	498	A + TR	3-m CC	15	DM-T	DGPS
Leksand (LEKS)	60.72	14.88	478	2 A	3-m CC	20	DM-T	
Lovö (LOVO)	59.34	17.83	80	2 A	3-m RA	10	DM-TA	DGPS
Mårtsbo (MART)	60.60	17.26	75	A + TR	3-m RA	25	DM-T	DGPS
Norrköping (NORR)	58.59	16.25	41	2 A	3-m CC	26	DM-TA	DGPS
Onsala (ONSA)	57.40	11.93	46	A + TR	1-m SQ	20	DM-B	DGPS
Oskarshamn (OSKA)	57.07	16.00	150	2 A	3-m CC	42	DM-TA	
Östersund (OSTE)	63.44	14.86	490	2 A	3-m CC	10	DM-T	
Överkalix (OVER)	66.32	22.77	223	2 A	3-m CC	18	DM-TA	
Skellefteå (SKEL)	64.88	21.04	81	A + TR	3-m CC	7	DM-T	DGPS
Sundsvall (SUND)	62.23	17.66	31	2 A	3-m CC	18	DM-T	DGPS
Sveg (SVEG)	62.02	14.70	491	A + TR	3-m CC	10	DM-T	DGPS
Umeå (UMEA)	63.58	19.51	54	A + TR	3-m CC	20	DM-T	
Vänersborg (VANE)	58.69	12.03	170	A + TR	3-m CC	15	DM-T	
Vilhelmina (VILH)	64.70	16.56	450	A + TR	3-m CC	19	DM-T	DGPS
Visby (VISB)	57.65	18.37	80	A + TR	3-m CC	12	DM-T	DGPS

[a] Four-letter abbreviations.
[b] Approximate WGS-84 coordinates.
[c] A - Ashtech-Z12 and TR - SNR-8000 TurboRogue.
[d] Height and cross-section of pillar; CC - circular; RA - rectangular; SQ - square
[e] Antenna cable length.
[f] Type of antenna mount is shown in annotated figure.

Instrument cabin

The standard SWEPOS instrument cabin has a size of 3×2 m and houses two satellite receivers, one data collecting computer for each receiver, one telecommunication modem for each receiver, one modem for remote control of electrical power supply to all instruments, Uninterrupted Power Supply (UPS) including batteries, heater control of the antenna pillar. Thermal insulation helps to maintain the ambient temperature range of the equipment, and ventilation safeguards for lead-acid rechargeable batteries. Considering an outdoor temperature range between -40°and +35°, the additional heating effect due to insulation on the cabin's surfaces, and the power dissipation of the apparatus (15 W/receiver; 40 W/PC; 40 W/LapTop; 50 W/modem) necessary heating and cooling devices have been installed. Power mains and telephone installation must be properly protected against transients, primarily those induced by lightning strikes. Attendance or visual inspection is done on the average four times per year.

Data flow in SWEPOS

The National Land Survey (NLS) hosts the operational center of SWEPOS, responsible for the downloading, RINEX-conversion, and filing of data from the SWEPOS sites. Data sampling rate is 15 s and the elevation cutoff level is 4° or 5° in the TurboRogue or Ashtech case, respectively.

To offload the data the site is dialed up from the control center in an automated process. One day's load of data is transferred at a time (2.5 Mbyte in compressed form) through a 19,200 baud high-speed modem, consisting of the following data types: Pseudorange measurements from C/A-code and from the P-code on both L1 and L2 frequencies; carrier phase observations on L1 and L2; Doppler frequency observations; and satellite broadcast ephemeris.

Data from the IGS incorporated into the standard solution concerns the following sites: Tromsö and Ny Ålesund (Norway), Metsähovi (Finland), Herstmonceux (UK), Brussells (Belgium), Kootwijk (Holland), Madrid (Spain), Matera (Italy), Potsdam and Wettzell (Germany). This data is acquired regularly via Internet.

Ancillary data bases specifying the reference sites setup, local ties between monuments etc. are also provided in the IGS archives.

Solving geodetic parameters

The dual-frequency GPS phase and pseudorange data are processed at the OSO regional processing center using the 2nd release of GIPSY software developed at Jet Propulsion Laboratory (JPL) (e.g., Webb and Zumberge, 1993, and references therein). Selected periods of the SWEPOS data are also processed using the Bernese Software ver. 4.0 (Rothacher et al., 1996). This redundant procedure may reveal erroneous data and possible modeling discrepancies.

The data from about 40 continuously operating GPS stations are processed. All processing is performed automatically, i.e., noninteractively. For the standard data analysis an elevation cutoff-angle of 15° is used for all sites giving the lowest uncertainties in the estimation of horizontal and vertical baseline components (Jaldehag et al., 1996a).

Improved satellite orbits and earth orientation parameters are readily available from the IGS processing centers. For the our standard analysis we have adopted a weighted combination of the estimated orbits from the seven analysis centers. With the present distribution of tracking stations, models, and processing techniques, the accuracy of the IGS orbit determination is known to be approximately 10 centimeter, or better.

Data processing utilizes a regional "no-fiducial" technique wherein the coordinates of site position have only weak a priori constraints. The coordinates of the sites are estimated as bias terms with a priori uncertainties of 10 m (IGS sites with well determined coordinates) or 1 km (regional sites). Constraints are thereafter applied to transfer the results into a terrestrial reference frame.

The zenith values, one for each site, of the tropospheric propagation delay are estimated as random walk bias terms. The a priori signal propagation delay is calculated based on a standard atmosphere and the latitude and height of the site.

The parameters estimated in the standard analysis are:

- stations clocks (white noise parameter)
- satellite clocks (white noise parameter)
- phase ambiguities (white noise parameter)
- stations coordinates (constant bias)
- tropospheric delay (random walk parameter)

- ionospheric delay (calculated from dual frequency observations)

Tides, earth orientation parameters and satellite positions are not estimated. However, at the post-processing stage also these parameters are investigated.

Differential GPS

In passing we shortly mention the Differential GPS service based on the SWEPOS network, which is a joint venture of NLS, OSO and Teracom Swedish Radio Broadcasting. The latter organisation is the commercial provider of the service, which is termed Epos. Those SWEPOS stations that serve as DGPS reference points transfer their data derived from C/A-code measurements continuously over dedicated telephone lines to the central TV tower in Stockholm for further distribution by Teracom as part of the RDS signal of the P3 program on the public FM band. This data consists of pseudorange corrections according to the RTCM-104 industry standard (Radio Technical Commission for Maritime Services) at a rate of about 0.5 Hz for all satellites with elevations above the horizon. Special FM receivers and subscription to the service are required to perform the RDS decoding. Relative positions of roving surveyors can be obtained in real-time regularly at 2 m accuracy level (95% confidence).

More advanced developments make use of the carrier frequency, which can provide accuracies below 1 m, optimistically as good as 0.1 m. For the part of Sweden, a service in this context is currently under development.

Results From Three Years of Operation

As of current, more than there years of SWEPOS operation and daily analyses of SWEPOS data and within the BIFROST project have resulted in a large number of repeated independent determinations of positions and baseline variations. BIFROST stands for Baseline Inferences for Fennoscandian Rebound Observations, Sea-level and Tectonics. The main final products from the analyses are

- estimates of site positions and variance/covariance between the estimates in the ITRF geocentric reference frame;
- estimates of baseline components between the sites and variance/covariance between the estimates.

- estimates of the tropospheric delay parameters for each site.

For item 1 and 2 a reference solution is selected. The differences of successive site position determinations are then displayed in e.g.,. local coordinates North, East, Vertical.

Site motion analysis

The following items can be addressed and conclusions, although still preliminary, can be expected: Can rates of change of site position rather than baseline components be estimated, or, conversely, is the degree of covariance of SWEPOS and IGS site positions in the daily solution so great that useful information can only be extracted from differential movement, i.e.,. baseline determinations ? Second, does an assessment of site position evolution confirm expectations on monument stability or, conversely, do we find signatures of random walk as proposed in Johnson and Agnew (1995) ?

Perturbing effects on site positions

While a permanent network has a number of advantages, primarily that antennas remain in the same place, and while continuous operation and data processing provides an excellent statistical basis for analysis, certain limitations exist, which require solutions or awareness. This has consequences also in the final stages of data analysis and interpretation.

If the permanent network is simultaneously used as a geodetic reference network, additional requirements arise. The specification of absolute position, which is more difficult in vertical component due to geometric dilution, must be as neutral as possible with respect to equipment used in e.g.,. high definition land surveys. Some trade-off of performance for a pure crustal deformation purpose is inevitable.

Unattended stations in remote, cold areas are exposed to the problem of snow and ice deposition on the antenna itself or on protection surfaces (Jaldehag et al., 1996a). Radomes are necessary to cover and protect the antenna assembly, implying consequences for both snow and ice deposition and antenna receiving conditions. The problem turned out to be nontrivial. In the long term, a sacrifice on the data available for the analysis might be more worthwhile to accept than overloading the project with complicated safeguard measures.

Inhomogeneous antenna diagrams may occur due partly to scattering off objects in the immediate environment (parts of the mounting assembly, pillar) and nearby surfaces (roofs, trees), but due also to elevation angle dependent transmission properties of e.g.,. antenna radomes. The effects of these antenna heterogeneity patterns are systematic offsets of the phase center from the nominal reference point, varying with the observation angles. Errors can occur in the range of tens of millimeters (Elósegui et al., 1995). For applications aiming to determine changes in position this may become negligible if the distribution of satellite viewing angles can be considered invariable. The serious implication is that a decision on a certain elevation cutoff angle cannot be revised after some years into the project as re-processing of the data accumulated thus far will become more and more infeasible.

The final choice of radome, to be implemented during autumn 1996, emphasizes a more uniform antenna diagram, trading-off data quality in the case of observations taken under snow; they may have to be discarded. The temporal pattern is easy to identify on the basis of the observations themselves, but also more sophisticated rejection criteria based on local meteorological data appear feasible. The level of these perturbations may reach several centimeters (Jaldehag et al., 1996a).

Analysis of GPS site position solutions

For the time being we determine preliminary results of site position rates by simultaneous least-squares fit of

- a box car train, i.e.,. bias terms that allow discontinuity of site position at known instances
- one slope for the whole scope of each site position component, conceptually representing the motion
- annual, semi-, ter-, and quater-annual sinusoids and cosinusoids that absorb some of the climatic problems, of which snow effects are the most important group.

The climatic signatures regularize the data to some degree. However, a fit is only reasonable if the box car sections are long enough to yield acceptable levels of parameter correlation. This is the case if the data spans

more than one year. Signal separation has maximum impact if the data coverage (including the effects of variable data weights) is heterogeneous.

The data to which the model is fitted consists of the post-processed time series of site position estimates, separately component by component and site by site (see Figures 6 and 7). To this end, the gipsy solutions are first projected into the ITRF94 frame performing a free network adjustment with respect to the IGS sites that participated in the GPS analysis. This provides the long-term compatibility of the solutions with respect to the IGS orbits. In order to obtain site motion with respect to a rigid frame that co-rotates with the plate (essentially with the frame designated by the participating IGS sites), we estimate this motion in another least-squares fit.

We estimate only three frame rotation parameters, the reason being that solving also translation parameters makes the frame motion more susceptible to a tilt around an axis that does not run through the geocenter, thus absorbing some of the vertical movement of the European IGS sites; to some part this might perturb the vertical estimates of the SWEPOS stations more than what the remaining weakness in our solution might contribute, which is long-term motion of the geocenter in the ITRF94 frame. Likewise, the scale factor of the frame is kept fixed since the rebound area undergoes areal strain, and we wish to preserve this component of the deformation in the station data.

One important modification of the rigid frame motion is needed in order to avoid another bias in the rate determination: The GPS orbits prior to July 1, 1995, relate to the ITRF93N frame. Thus, we *add* the differential motion of the new versus the old frame to the site positions of prior to this date. The difference amounts to $r \times \omega$ = [2.31, -1.09, 0.08] mm/yr.

Rate estimates

Displaying the site position rates on one map (Figure 5 right frame) we show the horizontal results in the form of motion vectors together with their 95 percent confidence limits. This figure comprises more than 1000 daily SWEPOS solutions. The results confirm largely the pattern predicted by e.g.,. Mitrovica et al. (1994). The left frame in Figure 5 shows the vertical rate in a color-coded contour plot.

We notice, however, that one of the stations in the central uplift area has a much larger vertical rate. Also, the observed horizontal rates appear

greater than the predictions by about a factor of two. However, conclusions at this stage would be highly preliminary. Considering an expected lifetime of the project of ten years, quantitative comparison with the large number of modeling results and attempts of parameter inversion are kept for the future.

We have excluded stations where the amount of data and the total time span are small. Due to the radome changes the introduction of jumps yields a nonnegligible degree of correlation between biases and rates. Also, the seasonal signatures cannot be reduced if the length of data branches is less than one year.

Comparison with tide gauge and levelling data

In Figure 8 we show the vertical rates determined at fifteen SWEPOS and two IGS sites versus the results from mareograph analysis and geodetic levelling (Ekman, 1996). This data type will be denoted ML henceforth. At Ny Ålesund we use the revised estimate of Breuer and Wolf (1995). The ML data for the two IGS sites have been taken from the vertical projection of the rates given in the ITRF94. The rates of change of the ML data represent relative land uplift. In the central uplift area it is less then the vertical motion of the crust by an amount corresponding to the rebound of the geoid. The rate uncertainty does not yet allow to resolve details of the interrelation, specifically the long-wavelength enhancement of the geoid change as compared with the solid surface. Therefore, a straight line fit will do. Excluding the apparently anomalous Skellefteå rate from the fit, we find a ML rate retardation of 29 percent. This appears large compared to even extreme models. Quite on the opposite side of the scale, Ekman and Mäkinen (1996) propose a value of only on the order of 5 percent.

In comparison with the ML rates, the GPS rates at Skellefteå and Vänersborg appear anomalous, causing the slope of GPS versus ML to steepen. Both stations have less data—they came online in April and June 1994, respectively, and the estimates of offsets, rates, and sinusoids have still a high degree of correlation. Thus, we expect a future result to settle at a slope which is closer to unity, unless Skellefteå continues to be affected by a local problem. Considering the short distance between Skellefteå and Umeå, and even more so between the GPS station and the Furuögrund tide gauge, on which the ML estimate is based, the possibility to find an explanation within the realm of glacial rebound theory and ice load is unlikely.

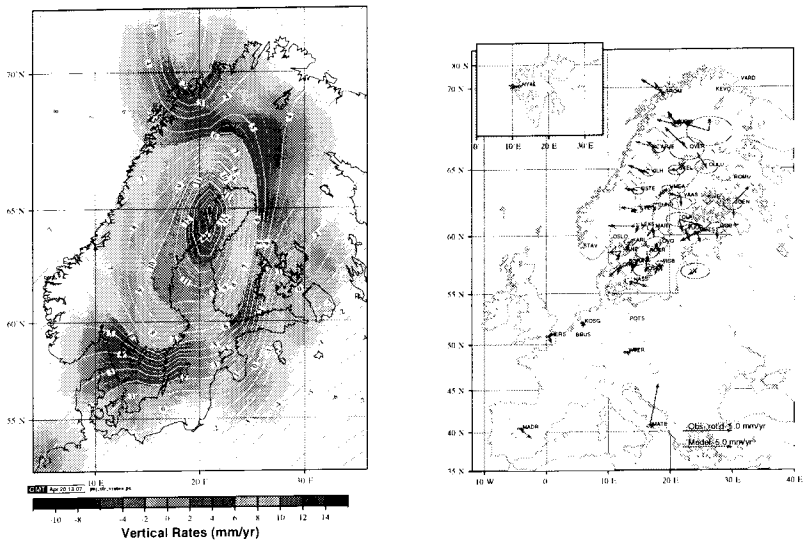


Figure 5: Left frame shows a contour plot of vertical rate estimates based on more than 1000 days of BIFROST GPS operation and data analysis. The right frame shows the horizontal rate estimates operations, uncertainties are 0.5 (0.7) mm/yr in the horizontal (vertical) except at the sites in Finland which have larger errors due less data analyzed. Also shown are the predicted motions from a geophysical model described by Mitrovica et al., [1994].

Most probably, the effect of the radome change is overestimated.

The intercept of the regression line at zero crustal rate, diminished by the geoid rate at that node, would under ideal circumstances indicate the amount of water level fall independent of glacial isostasy. The geoid rate at the node can be assumed to be less than 0.1 mm/yr. The vertical rates estimated by GPS, however, may be offset by a translation of the ITRF94 with respect to the geocenter. Assuming the latter term to be negligible, our estimate of the nonisostatic water level rate is -1.4 $\pm$0.3 mm/yr, i.e.,. our finding has the opposite sign compared to Douglas (1991), who inferred

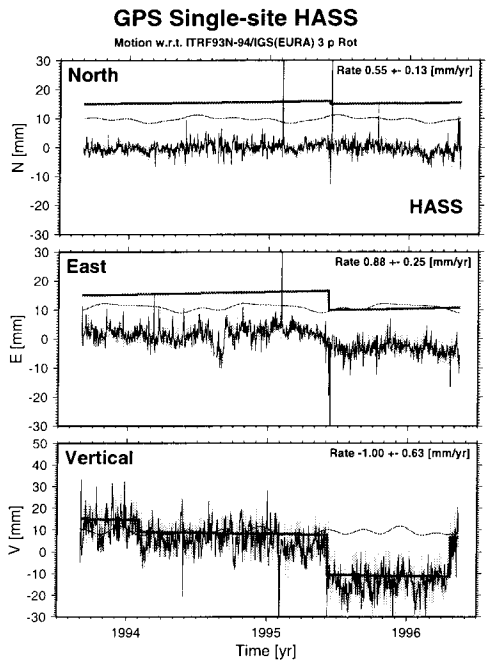


Figure 6: Single site solution, Hässleholm. Results from daily solutions are shown as the noisy thin line on a gray background signifying the 95% confidence limit. The positions are shown after free network adjustment and alignment with a rigid frame that co-rotates with the the European subset of the ITRF94. Station model least-squares fit assumes for each component a constant linear rate. Additional systematic features that are included in the model are position offsets. Their start and stop times are defined from known changes of the antenna mount or radome replacements. Slope and offset terms are combined in the thick line. Seasonal oscillations included in the fit are shown as a thin, wiggling line.

+2 mm/yr for the North Sea. In all, remaining systematic errors, is the probable cause of the inverse sea level signature. In fact, in a more recent analysis, with one additional year of data, we find rates of the same size and sign as Douglas (1991).

Other Applications of the SWEPOS Network and Analysis Products

Investigations on the use of the continuously observing GPS network for ionospheric and tropospheric research are undertaken. The possibility to use

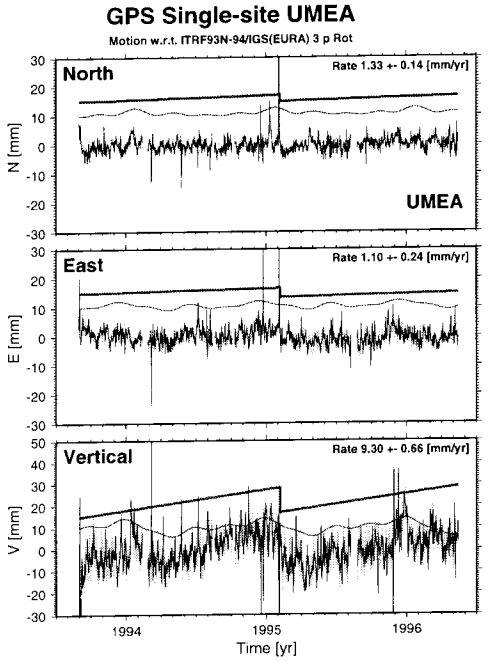


Figure 7: Single site solution, like in Fig. 6, for Umeå, however.

data from space geodetic techniques, such as GPS, to estimate atmospheric water vapor content has opened up new applications in association with weather forecasting and climate studies [e.g.,. Bevis et al., 1992 and Rocken, 1993]. The continuously operating network, SWEPOS, serves also as a data bank for detailed investigations of sources of errors in space geodesy [Elosegui et al., 1995,Jaldehag et al., 1996a Jaldehag et al., 1996b]. We have used GPS and VLBI data to estimate the tropospheric propagation delay due to water vapor and simultaneously obtained measurements by ground-based microwave radiometer (Water Vapor Radiometer - WVR) (see Figure 9) [Carlsson et al., 1995]. We believe that the accuracy of all these three techniques are now comparable and that it is realistic to believe that they all can be further improved during the next few years.

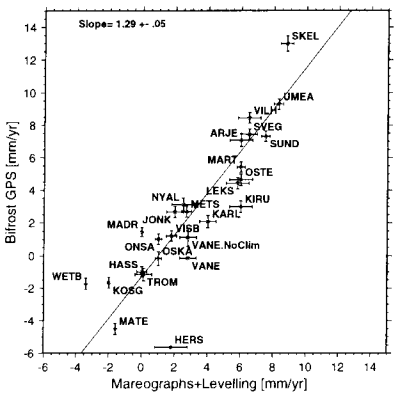


Figure 8: GPS vertical crustal rates from SWEPOS analysis versus rates of land emergence determined by Ekman (1995) using Mareograph and Precise Levelling data. A line of regression (solid) is fit considering one sigma limits of both data types. Where horizontal error bars are missing the corresponding data have not been included in the fit; these estimates have been taken from the ITRF94 station catalogue.

Conclusions

Vertical movements in Fennoscandia determined from continuous observations with the SWEPOS permanent GPS network in the central region of postglacial rebound in Europe confirm the rebound pattern obtained with viscoelastic earth models. A comparison with traditional geodetic measurement shows agreement around a linear regression term with a normalized $\chi^2 = 5$ (20 degrees of freedom). We find a signature in the linear regression between the data types that has the same sign as the geoid rebound term, however in excess of the amplitude estimate of Ekman and Mäkinen (1996) by a factor of three. After 1000 days into the project, the precision of single site position rates is regularly at the $2\sigma \simeq 95$ percent level of 0.7 mm/yr (vertical) and 0.15 mm/yr (horizontal) with a slight degradation in the east direction.

Repeatabilities of single site positions are regularly found at 7 mm (vertical) and 3 mm (east and north); in some cases, north repeatability is found as low as 2 mm. To obtain these results we compare the daily position estimates with a fitted model that, besides uniform velocities, contains sinusoids of seasonal character and offsets to account for changes in the an-

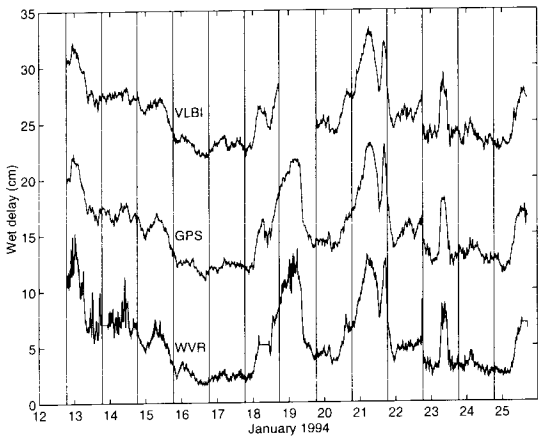


Figure 9: Estimated wet propagation paths using a WVR (true scale), GPS (offset of 10 cm added for clarity), and VLBI (offset of 20 cm added). The vertical lines indicate the start and stop time of the individual 24 hour VLBI experiments.

tenna setup. We discern seasonal signatures predominantly in the vertical; they are almost certainly related to snow conditions during winter. Residual systematic errors are most probably related to these seasonal perturbations and, in addition, to unaccounted processes in the neutral atmosphere and the ionosphere.

Future

We will be careful to maintain SWEPOS, replacing equipment after careful consideration. Experience shows that too frequently occurring replacements and rearrangements can render segments of recorded data and processed results quite useless. However, many types of replacements are needed, especially those that make the system more homogeneous. The balance is delicate.

Currently BIFROST daily analysis with the GIPSY/OASIS-II software and the Bernese software can process ~45 sites within reasonable time. As data processing power is developing, utilizing distributed processing and subnetting techniques, the processing of ~100 stations in one merged solution appears feasible. We priorities new stations in Sweden in areas where crustal stability appears to be lower. Data from new stations outside Sweden will be included, commencing this year, in the course of cooperating

Telespazio, Matera, Italy.

Ekman, M., 1996. A consistent map of the postglacial uplift of Fennoscandia, *Terra Nova,* **8,** 158–165.

Ekman, M., and Mäkinen, J., 1996. Recent postglacial rebound, gravity change and mantle flow in Fennoscandia, *Geophys. J. Int.*, **126,** 229–234.

Elósegui, P., J. L. Davis, R. T. K. Jaldehag, J. M. Johansson, A. E. Neill, and I. I. Shapiro, 1995. Geodesy Using the Global Positioning System: The Effects of Signal Scattering on Estimates of Site Position, *J. Geophys. Res.*, **100**, 9921–9934.

Douglas, B.C., 1991. Global Sea Level Rise, *J. Geophys. Res.,* **96**, 6981–6992.

Jaldehag, R.T.K, Johansson, J.M., Rönnäng, B.O., Elósegui, P., Davis, J.L., Shapiro, I.I., and Neill, A.E., 1996a. Geodesy using the Swedish Permanent GPS Network: Effects of signal scattering on estimates of relative site positions, *J. Geophys. Res.*, **101,** 17,841–17,860

Jaldehag, R.T.K, Johansson, J.M., Davis, J.L., and Elósegui, P., 1996b. Geodesy using the Swedish Permanent GPS Network: Effects of snow accumulation on estimates of site positions, *Geophys. Res. Letters,* **23,** 1601–1604.

Johnson, H. O., and Agnew, D. C., 1995. Monument motion and measurement of crustal velocities, *Geophys. Res. Letters,* **22,** 2905–2908.

Mitrovica, J.X., Davis, J.L., Shapiro, I.I., 1994. A spectral formalism for computing three-dimensional deformations due to surface loads, 2. Present-day glacial isostatic adjustment, *J. Geophys. Res.,* **99,** 7075–7101.

Rothacher, M., G. Beutler, W. Gurtner, E. Brockmann, and L. Mervart, 1996. *Bernese GPS Software (Version 4.0), 1996 Documentation,* Astronomical Institute, University of Berne.

Webb F H, Zumberge J F, 1993. *An Introduction to GIPSY/OASIS-II Precision Software for the Analysis of Data from the Global Positioning System,* JPL Publ. No. D-11088, Jet Propulsion Laboratory, Pasadena, Cal.

Wessel, P. and Smith, W.H.F., 1995. New version of the Generic Mapping Tools released, *EOS Trans. American Geophys. Union,* **76,** 329.

projects, goaded to within the Nordic Geodetic Commission and including the three Baltic states.

Building on our successful Fennoscandian GPS project, we have proposed a densification of this network. Densification of the existing GPS network will enable us to obtain viscosity estimate with higher resolution in the mantle, and increases our ability to investigate ice models errors, and to use corrected tide-gauge measurements to estimate global sea-level change. We are currently investigating the possibility of placing these GPS sites at some of the over 100 permanent meteorological stations of the Swedish Meteorological and Hydrological Institute (SMHI) since these sites are already provided with power and communications.

Acknowledgement

We would especially like to thank our colleagues in the geogroup at Onsala Space Observatory, the National Land Survey of Sweden, and the BIFROST group for valuable contributions to the manuscript and to the project. This research was in part supported by the EC Environment and Climate Research Programme, the Swedish Natural Research Council, and the Swedish National Space Board. We have used GMT graphics software due to Wessel and Smith (1995).

References

Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R.A., and Ware, R.H., 1992. GPS Meteorology: Remote Sensing of atmospheric water vapor using the Global Positioning System, *J. Geophys. Res.*, **97**, 15787–15801.

BIFROST Project Members, 1996. GPS measurements to constrain geodynamic processes in Fennoscandia, *EOS Trans. American Geophys. Union,* **77,** p.337+339.

Breuer, D., and Wolf, D., 1995. Deglacial land emergence and lateral upper-mantle heterogeneity in the Svalbard archipelago—I. First results for simple load models, *Geophys. J. Int.,* **121,** 775–788.

Carlsson, T.R., Elgered, G., and Johansson, J.M., 1995. External wet tropospheric corrections during a two-week-long VLBI campaign. in Lanotte, R., and Bianco, G. (eds.) *Proceedings of the 10th Working Meeting on European VLBI for Geodesy and Astrometry*, pp. 148–153,

STUDIES ON A NEW METHOD FOR GPS BASE AMBIGUITY RESOLUTION BY COMBINED PHASE AND CODE OBSERVABLES

by

Lars E. Sjöberg
Royal Institute of Technology
Department of Geodesy and Photogrammetry
S-100 44 Stockholm, Sweden

0. ABSTRACT

Firstly, we present the new solutions for GPS phase base ambiguities by Sjöberg (1996). We show that they work well for three baselines (in the static mode).

Secondly, the BLUE (Best Linear Unbiased Estimator) and the BLE (Best Linear Estimator) of the unknown u, composed of the satellite-to-receiver range plus a clock bias, are studied. For example, it is shown that in the special case of no ionosphere bias, the variance of the BLE is merely 6% of that of the BLUE (which is the same as the wellknown ionosphere-free linear combination).

Thirdly, we study the propagation of a gross error in the widelane integer ambiguity into gross errors in the base ambiguity solutions. It is shown that small odd errors can easily be detected, while the even errors are more difficult to discover.

1. INTRODUCTION

A limiting problem in the effective use of the Global Positioning System (GPS) in the centimetre level of accuracy or even better is the accurate and reliable determination of integer phase ambiguities of the carrier wave. It is wellknown that the so-called widelane ambiguity (N_w) can most successfully be determined from a combination of phase and code observables. A remaining problem is to separate the base ambiguities N_1 and N_2, corresponding to the L_1 och L_2 observables, from $N_w = N_1 - N_2$.

One solution to this problem was presented by Sjöberg (1996). He started from the full set of phase and code observation equations for a dual frequency GPS receiver:

$$\text{and} \quad \begin{cases} \phi_i = uf_i - \dfrac{\mu}{f_i} + N_i & ; \ i = 1,2 \\ P_i = uf_i + \dfrac{\mu}{f_i} & ; \ i = 1,2 \, , \end{cases} \tag{1}$$

where ϕ_i (i=1,2) are the carrier phase observables and P_i (i = 1,2) are the code phase observables on L_1 and L_2 with frequencies f_1 and f_2, μ is the unknown ionosphere effect and, finally, $u = \rho / c + \Delta\delta$, i.e. it is a combination of the range (ρ) from satellite to receiver, velocity of the signal (c) and the bias difference between satellite and receiver clocks ($\Delta\delta$). The instantaneous, unique solutions for N_1 and N_2 from the system (1):

$$\begin{aligned} \hat{N}_1 &= \phi_1 + gP_2 - hP_1 \\ \hat{N}_2 &= \phi_2 - gP_1 + hP_2 \, , \end{aligned} \tag{2}$$

where

$$g = 2f_1f_2 / (f_1^2 - f_2^2)$$

and

$$h = (f_1^2 + f_2^2) / (f_1^2 - f_2^2) \, ,$$

have large standard errors and are therefore impractical. On the contrary, the widelane ambiguity estimator

$$\hat{N}_w = \phi_1 - \phi_2 - \frac{f_1 - f_2}{f_1 + f_2}(P_1 + P_2)$$

has a low standard error. Sjöberg (1996) suggested to take advantage of the fixed widelane solution $\hat{N}_w$ to attain improved estimates of N_1 and N_2 by the formulas

$$\hat{\hat{N}}_1 = \frac{\left\{g^2 + h^2\nu^2 - hg(\nu^2+1) + \nu^2\kappa\right\}\left(N_w + \hat{N}_2\right) + \left\{\nu^2 g^2 + h^2 - hg\left(\nu^2+1\right) + \kappa\right\}\hat{N}_1}{\left(\nu^2+1\right)(1+\kappa)} \tag{3a}$$

and

$$\hat{\hat{N}}_2 = \hat{\hat{N}}_1 - N_w \ , \tag{3b}$$

where

$$\nu = f_1 / f_2 = \lambda_2 / \lambda_1$$

and

$$\kappa = \mathrm{Var}\{\phi_i\} / \mathrm{Var}\{P_i\}$$

λ_i = carrier wavelength for frequency f_i.

The solutions (3) have several times lower standard errors than the original estimates $\hat{N}_1$ and $\hat{N}_2$. For short baselines, where the ionosphere effect is negligable, the solutions become

$$\hat{\hat{N}}_1 = \frac{\hat{N}_1 - \nu\left(N_w + \hat{N}_2\right)}{1-\nu} \tag{4a}$$

and

$$\hat{\hat{N}}_2 = \hat{\hat{N}}_1 - N_w \ , \tag{4b}$$

(where we have neglected some minor terms of order $\kappa = 10^{-4}$). The standard errors of (4a) and (4b) are very small, assuring that N_1 and N_2 can be resolved as soon as N_w has been fixed. A successful application of this method is demonstrated in Figs. 1-6 for the estimation of the widelane and L1 base ambiguities in static mode for double difference phase observables on a 10 m, 20 m and a 800 m baseline. (The next step will be to test the method in real-time applications.)

The purpose of this paper is to take advantage of the resolved ambiguities N_1 and N_2 for satellite-to-receive range determination and also to consider some reliability aspects of the estimated unknowns.

2. SATELLITE-TO-RECEIVER RANGE DETERMINATION

The range from the GPS receiver to satellite is closely related with point positioning. Once the base ambiguities N_1 and N_2 have been fixed, the question arises how the satellite-to-receiver ranges are best estimated from the dual frequency phase and code observables. Thus we return to the observation equations (1), where N_1 and N_2 have now been fixed. There are two remaining unknowns (u and μ) and four equations. (Notice that ρ and $\Delta\delta$ cannot be separated from these equations.)

The least squares solution for u becomes

$$\hat{u} = \left\{(1+\bar{v}^2)(1+\kappa)\left(\tilde{\Delta}_1 + \tilde{\Delta}_2 + \kappa(R_1+R_2)\right) - (1-\kappa)(1+\bar{v})\left(\tilde{\Delta}_1 + \tilde{\Delta}_2\bar{v} - \kappa(R_1+R_2)\right)\right\} / D \qquad (5a)$$

where

$$D = (1-\bar{v})^2 + 2\kappa(3+2\bar{v}+\bar{v}^2) + \kappa^2(1-\bar{v})^2 \qquad (5b)$$

$$\tilde{\Delta}_i = \lambda_i\phi_i - \lambda_i N_i + \varepsilon_{i1} \qquad ; i = 1,2 \qquad (5c)$$

$$R_i = \lambda_i P_i + \varepsilon_{i2} \qquad ; i = 1,2 \qquad (5d)$$

$$\bar{v} = (f_1 / f_2)^2$$

$\varepsilon_{i1}, \varepsilon_{i2}$ = random errors of observables.

Furthermore, the variance of $\hat{u}$ becomes

$$\sigma^2_{\hat{u}} = \sigma^2_{\phi\lambda}(1+\bar{v}^2)(1+\kappa) / D \ , \qquad (6)$$

where we have assumed that $\sigma_{\phi\lambda} = \lambda_i \sigma_{\phi_i}$ is constant and σ_{R_i} = constant, for i = 1,2. As $\kappa = \left(\sigma_{\phi_\lambda} / \sigma_R\right)^2$ is about 10^{-4} for modern GPS receivers, (with $\sigma_{\phi\lambda} = 3$ mm, and $\sigma_R = 60$ cm ,) all terms including κ and κ^2 can be neglected, yielding

$$\hat{u} = \frac{f_1^2}{f_1^2 - f_2^2}\tilde{\Delta}_1 - \frac{f_2^2}{f_1^2 - f_2^2}\tilde{\Delta}_2 \qquad (7a)$$

and

$$\sigma_{\hat{u}} = \frac{\sigma_{\phi\lambda}}{f_1^2 - f_2^2}\sqrt{f_1^4 + f_2^4} \ . \qquad (7b)$$

This solution is nothing but the ionosphere-free linear combination. (Obviously the code observables are too poor to effect the solution.) This solution can be regarded as the best linear unbiased estimator, the BLUE.

If we give up the unbiasedness, we may derive the best linear estimator (the BLE), minimizing the mean square error, cf. Sjöberg (1990):

Definition: The BLE of $\tilde{u}$ minimizes the mean square error

$$m^2 = E(\tilde{u} - u)^2 \ . \qquad (8)$$

Introduce the general estimator

$$\tilde{u}(c) = (k_2 + c)\tilde{\Delta}_1 - (k_1 + c)\tilde{\Delta}_2 \ , \qquad (9)$$

where

$$\tilde{\Delta}_i = \Delta_i + \varepsilon_{il} \qquad ; i = 1,2$$

$$k_2 = \frac{f_1^2}{f_1^2 - f_2^2} \quad \text{and} \quad k_1 = \frac{f_2^2}{f_1^2 - f_2^2} \ .$$

We notice that $\tilde{u}(0)$ is the ionosphere-free linear combination (7a). The mean square error of $\tilde{u}(c)$ becomes

$$m^2(c) = \mathrm{Var}\{c\} + \mathrm{Bias}^2\{c\} = \left\{(k_2 + c)^2 + (k_1 + c)^2\right\}\sigma_{\phi\lambda}^2 + c^2(\Delta_1 - \Delta_2)^2 \qquad (10)$$

Differentiating (10) with respect to c we find the minimum at

$$\hat{c} = -(k_1 + k_2)\sigma_{\phi\lambda}^2 / F \ , \tag{11a}$$

where

$$F = 2\sigma_{\phi\lambda}^2 + (\Delta_1 - \Delta_2)^2 \ . \tag{11b}$$

Inserting $\hat{c}$ into (9) we thus obtain the BLE

$$\hat{u} = k_2^* \ \tilde{\Delta}_1 + k_1^* \ \tilde{\Delta}_2 \ , \tag{12a}$$

where

$$k_1^* = \left\{\sigma_{\phi\lambda}^2 - k_1(\Delta_1 - \Delta_2)^2\right\} / F \quad \text{and} \quad k_2^* = \left\{\sigma_{\phi\lambda}^2 - k_2(\Delta_1 - \Delta_2)^2\right\} / F \tag{12b}$$

The variance and bias of this estimator are

$$\mathrm{Var}\left\{\hat{u}\right\} = \left\{\left(k_1^*\right)^2 + \left(k_2^*\right)^2\right\}\sigma_{\phi\lambda}^2 \tag{12c}$$

$$\mathrm{Bias}\left\{\hat{u}\right\} = \hat{c}(\Delta_1 - \Delta_2) \ . \tag{12d}$$

Also, it can be shown that (cf. Sjöberg, 1990)

$$m^2(\hat{u}) = \mathrm{Var}\{\hat{u}\} - \left(\frac{f_1^2 + f_2^2}{f_1^2 - f_2^2}\right)^2 \frac{\sigma_{\phi\lambda}^4}{F} \ , \tag{13}$$

i.e. the mean square error of the best solution is smaller than the variance of the ionosphere free linear combination (7b). The last term of (13) is the gain in the BLE over the estimator in (7b). In the extreme case of no ionosphere effect, i.e. $\Delta_1 - \Delta_2 = 0$, this term becomes

$$\text{gain} = \left(\frac{f_1^2 + f_2^2}{f_1^2 - f_2^2}\right)^2 \frac{\sigma_{\phi\lambda}^2}{2} \ . \tag{14}$$

As

$$\mathrm{Var}\{\hat{u}\} = \frac{f_1^4 + f_2^4}{(f_1^2 - f_2^2)^2}\,\sigma_{\phi\lambda}^2 \qquad (15)$$

the gain is about 94%, i.e. the mean square error of the BLE,

$$m^2\{\hat{\hat{u}}\} = \sigma_{\phi\lambda}^2 / 2 \qquad (16)$$

is merely 6% of $\mathrm{Var}\{\hat{u}\}$! This result indicates the great advantage of the Best Linear Estimator.

3. RELIABILITY ASPECTS

Let us assume that a gross error dN_w was committed when fixing the widelane ambiguity to N_w. For GPS observations κ is negligible. It follows from formulas (4) that dN_w propagates into the following errors of $\hat{\hat{N}}_1$ and $\hat{\hat{N}}_2$:

$$dN_1 \approx \frac{\lambda_2}{\lambda_2 - \lambda_1} dN_w \qquad (17a)$$

and

$$dN_2 \approx \frac{\lambda_1}{\lambda_2 - \lambda_1} dN_w \ . \qquad (17b)$$

As λ_1 = 19,0 cm and λ_2 = 24,4 cm, formulas (17) can be written

$$dN_1 \approx 4{,}5\, dN_w \qquad (18a)$$

and

$$dN_2 \approx 3{,}5\, dN_w \qquad (18b)$$

Also $\hat{\hat{N}}_1$ and $\hat{\hat{N}}_2$ have small standard errors, and it follows from (18) that a small odd integer ambiguity error dN_w can easily be detected, while even gross errors dN_w cannot be easily detected in this way. However, as the factors in (18 a,b) are not exactly 4,5 and 3,5 but rather 4,4364 and 3,5185, it means that large integer errors, odd or even, dN_w can probably be detected in this way from the real valued estimates of N_1 and N_2 or their sum. Cf. Table 1 below.

Table 1 **Estimated float ambiguities as a function of integer error dN_w**

dN_w	N_1	N_2	$N_1 + N_2 = N_w$	N_w
0	10	2	12	8
-3	-3.309	-8.556	-11.865	5.247
-2	+1.127	-5.037	-3.910	6.164
-1	+5.564	-1.519	4.045	7.083
1	14.436	5.519	19.955	8.917
2	18.873	9.037	27.910	9.836
3	23.309	12.556	35.865	10.753
4	27.746	16.074	43.820	11.672

4. CONCLUDING REMARKS

The base ambiguities are preferably determined by the new method presented in chapter 1. The correct fixing of the widelane ambiguity (N_w) is critical for this approach. If N_w is in error by one (or a few) odd cycle(s) (dN_w), this gross error is easily detected by the method discussed in 3.

Furthermore we have shown that there could be a considerable gain in using the BLE compared to the BLUE for the satellite-to-receiver range.

ACKNOWLEDGEMENTS

The numerical results presented in Fig 1, carried out by Kjell Almgren, are cordially acknowledged.

REFERENCES

Sjöberg, L.E. (1990): The best linear combinations of L1 and L2 frequency observables in the application of Transit/Doppler and GPS. **Manuscripta Geodaetica (1990), 15,** pp. 17-22.

Sjöberg, L.E. (1996): A new method for GPS phase base ambiguity resolution by combined phase and code observables. Submitted to J. of Geodesy.

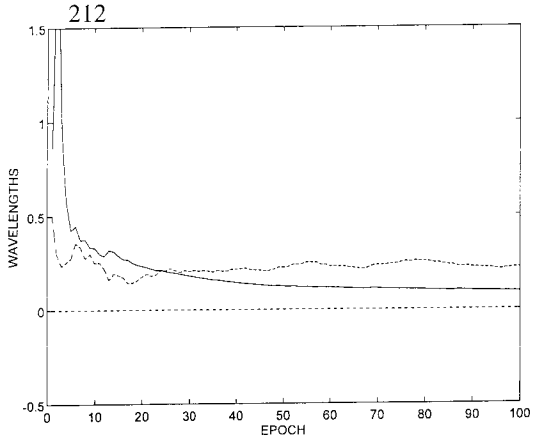


Fig 1a. The figure shows the estimated double difference base ambiguity N_w vs. its true value (dotted line) and its standard error (solid line). The double difference is determined for a 20 m baseline in an open field and two satellites at elevations 67° and 59°.

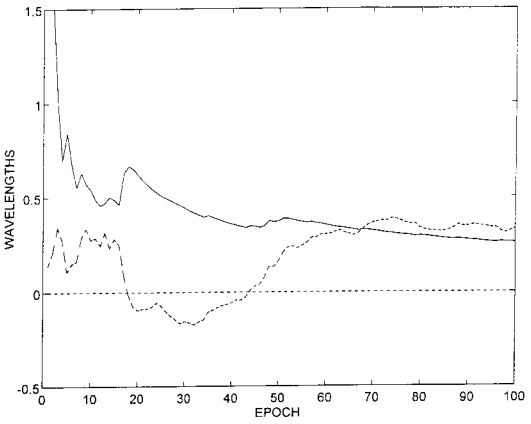


Fig 1b. As fig 1a, but lower satellite at elevation 39°.

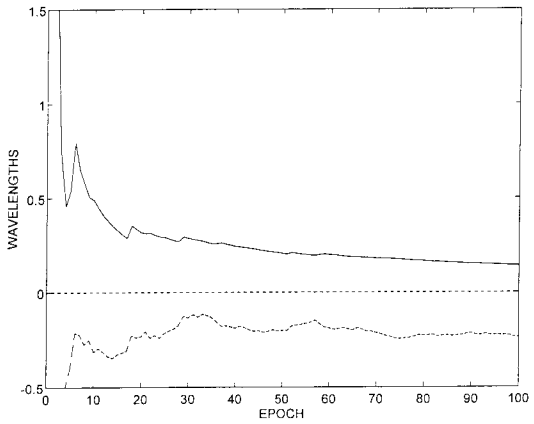


Fig 1c. As fig 1a, but lower satellite at elevation 20°.

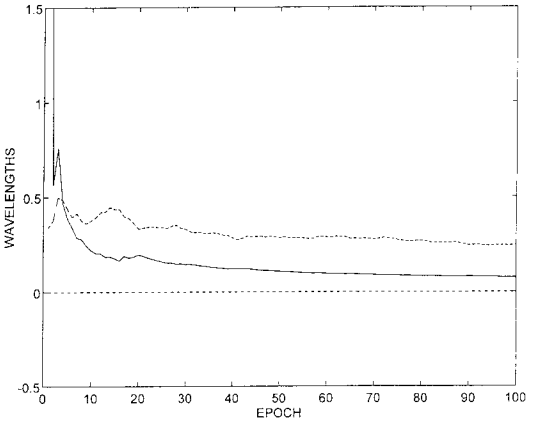


Fig 2a. As fig 1a, but the baseline length is about 800 m and the satellite elevations are 68° and 54°.

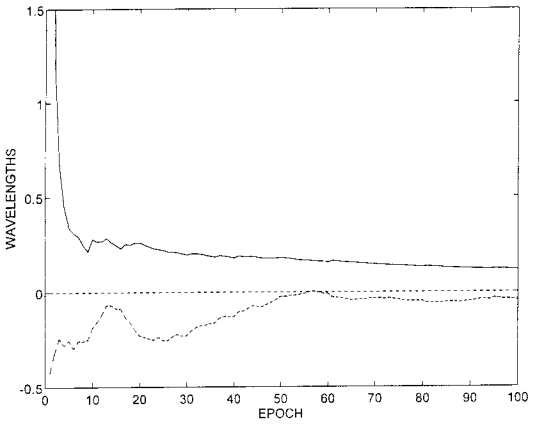


Fig 2b. As fig 2a, but lower satellite at elevation 37°.

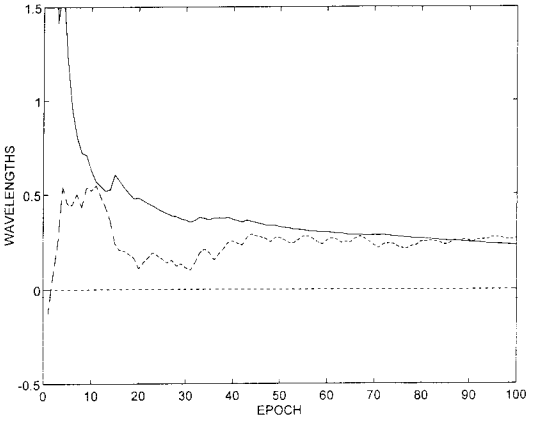


Fig 2c. As fig 2a, but lower satellite at elevation 18°

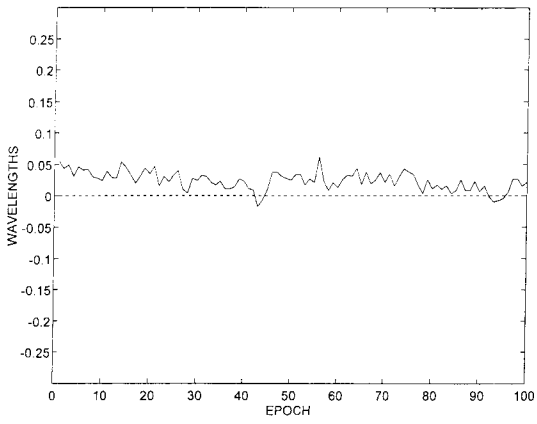


Fig 4a. The figure shows the N_1 ambiguity when N_w is fixed at its correct value, the x-axis is the correct value for N_1. Same double difference as 1a.

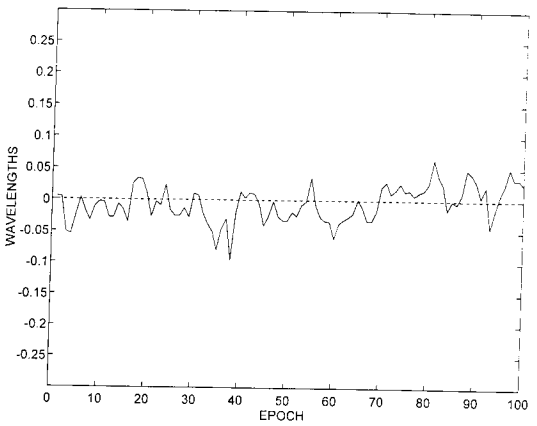


Fig 4b. As fig 4a, but same double difference as 1b.

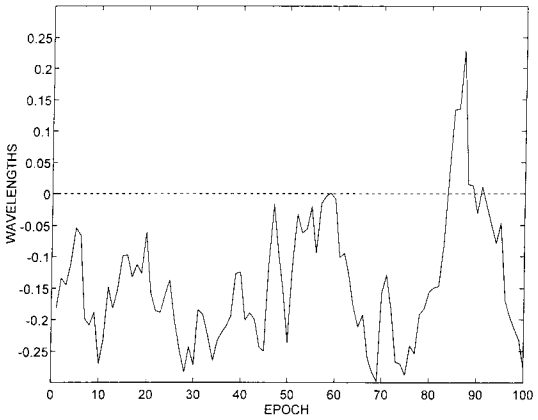


Fig 4c. As fig 4a, but same double difference as 1c.

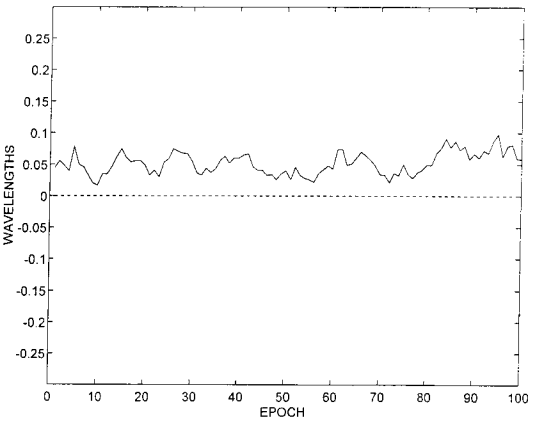


Fig 5a. As fig 4a, but same double difference as 2a.

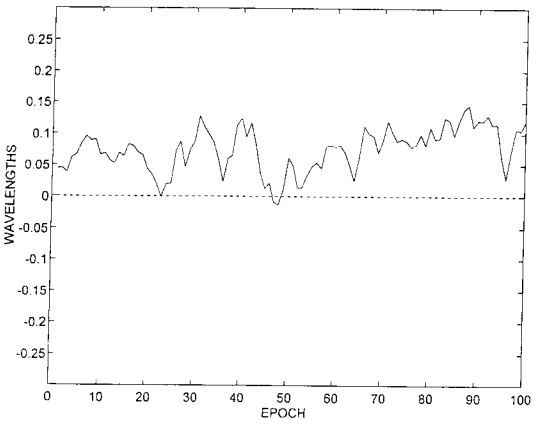


Fig 5b. As fig 4a, but same double difference as 2b.

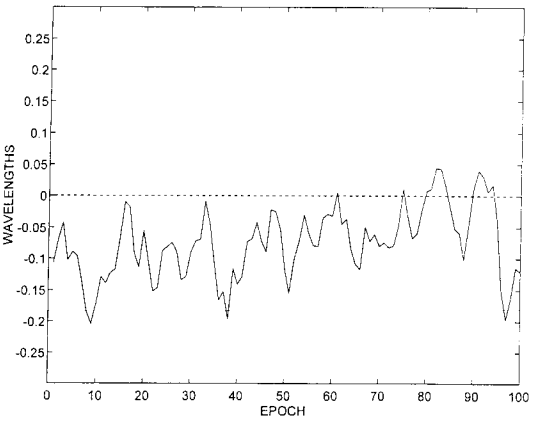


Fig 5c. As fig 4a, but same double difference as 2c.

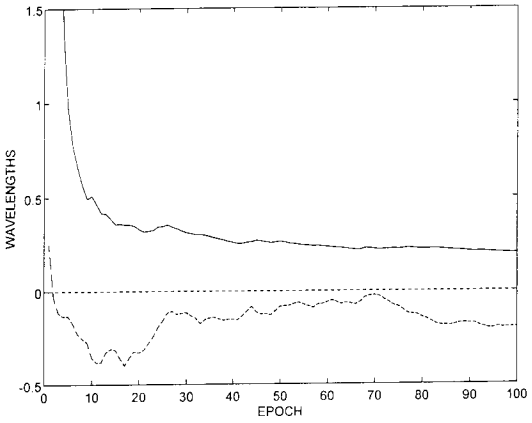


Fig 3a. As fig 1a, but the baseline is located at a roof with a metallic staircase, lenght of baseline 10 m, satellite elevations 65° and 56°.

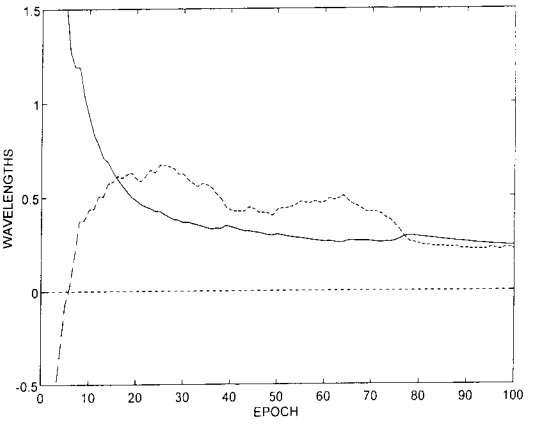


Fig 3b. As fig 3a, but lower satellite at elevation 40°

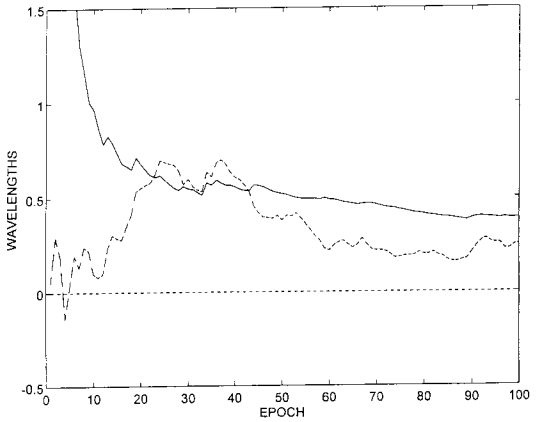


Fig 3c. As fig 3a, but lower satellite at elevation 22°

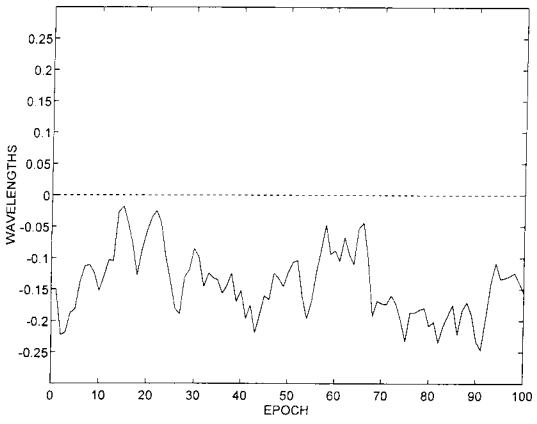


Fig 6a. As fig 4a, but same double difference as 3a.

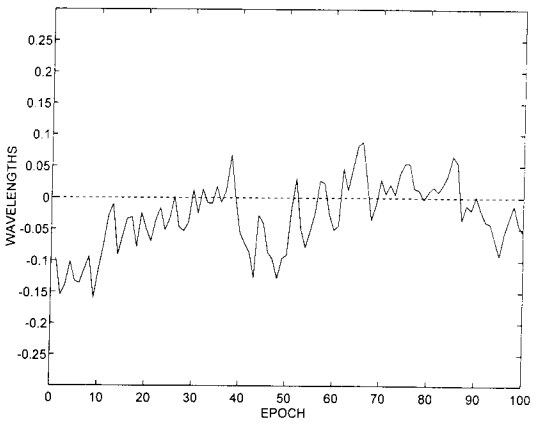


Fig 6b. As fig 4a, but same double difference as 3b.

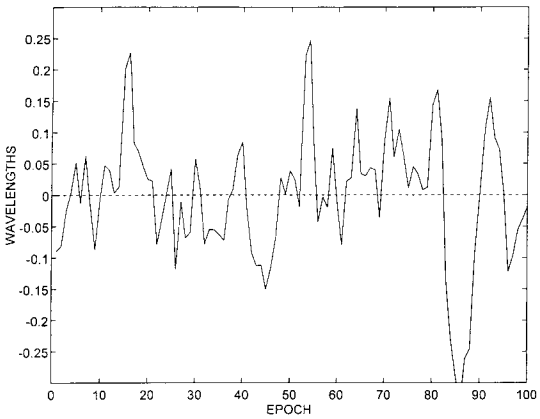


Fig 6c. As fig 4a, but same double difference as 3c.

USE OF THE GLOBAL POSITIONING SYSTEM FOR GEODYNAMIC RESEARCH IN FINLAND

Martin Vermeer

Abstract

We describe the activities of the Finnish Geodetic Institute in the field of geodynamic research and related activities, such as atmospheric studies.

A link with physical geodesy is formed by sea level and geoid related studies, in which GPS nowadays also plays an integral role.

1 Introduction

In these lecture notes we present the field of GPS activity as currently existing in Finland at the Finnish Geodetic Institute. Focal points of the Finnish Geodetic Institute's GPS research work are:

- The realization of a national EUREF implementation on the Finnish territory; this we will not discuss here.
- Use of GPS for research into geodynamics, especially with a view on understanding the postglacial rebound and properties of the Earth's mantle, particularly in the context of global climate change. This is a highly international field of research requiring worldwide and interdisciplinary collaboration.

In addition to this of course we have an interest in using our GPS resources to the widest extent possible, also in collaborative contexts with other sciences such as meteorology (tropospheric water vapour), atmospheric science (troposphere, ionosphere), solid Earth geophysics etc.

In this article we shall concentrate on the use of GPS for geodynamics, sea level and atmospheric studies.

2 Usability of GPS for geodynamics

2.1 Technical properties of GPS

While it is possible to use GPS for precise geodynamics studies, there are a number of limitations and constraints caused by the physical nature of the measurement process. One has to be well aware of these.

1. GPS measures with *microwaves*. Geodetic GPS is carrier phase interferometry, which means that the measurement accuracy to be expected is at best only some fraction of the wavelength. The wavelength being 19-24 cm, this means, for a fraction of 0.1%, some ±2 mm.
2. Microwave propagation is *wave propagation*. This means that the signal can reach the antenna through a number of different paths, interfering at the antenna. This problem is known as *multipath*. Because the antenna is always located close to the Earth's surface, one should always expect this problem to occur to some extent. Reflecting metal surfaces such

as car roofs may worsen conditions. In current geodetic GPS, multipath is the limiting factor for precision, which is around ±1 cm *irrespective of baseline length.* Below we will discuss methods for reducing the multipath problem.

3. Microwaves are very sensitive to *water vapour.* On the other hand, it is possible to *estimate* the atmosphere's water vapour content by making use of the circumstance that there are several satellites simultaneously observable, standing on differing elevations in the sky. The principle of this estimation process will be explained below.

4. GPS measurement quantity is *pseudo-range*, i.e. range contaminated by unknown receiver clock error. The unknown receiver clock parameters have to be estimated or eliminated (by constructing differences) from the observation equations; this tends to weaken them. Especially in a global geometry, where the number of satellites simultaneously visible from stations far apart may be as few as zero, this will seriously weaken the whole network geometry.

   One way of reducing this problem is to connect the GPS reciever to an accurate atomic time standard. This has been done on some stations and is common practice on VLBI stations. However, these devices are very expensive.

5. A closely related problem is that of *orbit predictability* of the GPS satellites. Because of solar radiation pressure and the irregular shape and reflection properties, it is not possible to accurately predict — or model — the movement of the satetellites in their orbits even over one revolution. This contrary to e.g. laser ranging satellites, which are heavy and compact and thus move very predictibly. This means that *global* GPS networks tend to be relatively weaker than SLR networks. The problem also does not occur for VLBI, as the radio sources used by this technique stand still in the sky.

2.2 Experiences with GPS precision over various baseline lengths

In an unpublished study we investigated the effect of baseline length on baseline vector error. The Bernese group (Beutler *et al.*, 1989) quote a precision equal to

$$\delta r[\mathrm{mm}] = 0.7\sqrt{d[\mathrm{km}]},$$

where δr stands for the inter-session three-dimensional variability, computed from coordinate variabilities by

$$\delta r = \sqrt{\delta x^2 + \delta y^2 + \delta z^2},$$

and session length is 24 hours.

This formula resembles in form that used to express the precision of levelling measurements. Also the level of precision is very similar to that of precise levelling (which thus, despite its age, is a remarkably precise technique). The physical reasons behind the formula, in the case of GPS, are not quite clear, but presumably some error accumulation process over distance.

We computed a seven point network covering Finland, measured with TurboRogues for five successive 24-hour sessions in August 1993 at a sampling interval of 30 s, in order to gain experience with this kind of computation. We found that the Bernese rule of thumb held very well:

$$\delta r[\mathrm{mm}] = 0.421\sqrt{d[\mathrm{km}]},$$

more or less independent of precise solution strategy chosen. In the below Fig. 1 we display the inter-session variability for adjustment using "correct" inter-baseline correlation modelling. M. Poutanen (personal comm.) found that adjustment using per-baseline correlations and an overdetermined vector network in practice gives nearly the same results, a finding confirmed by this calculation.

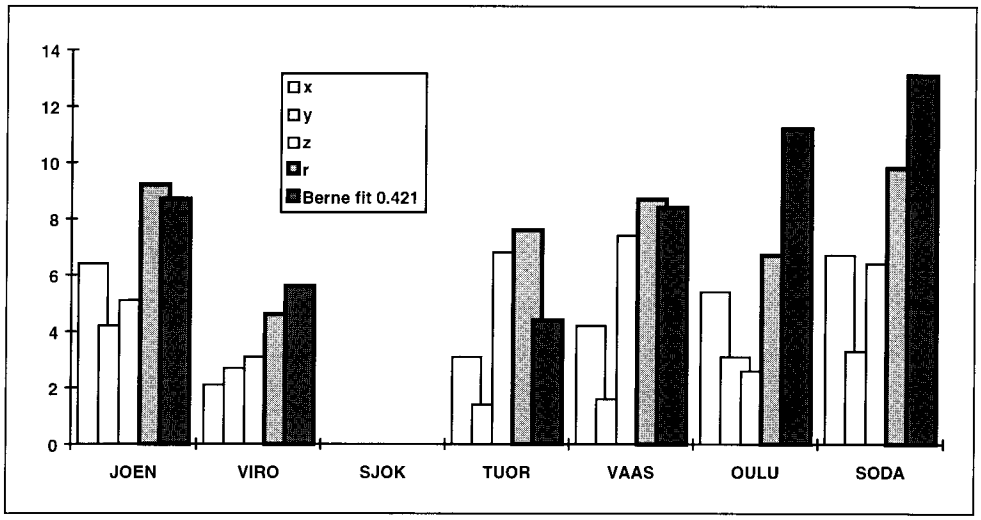


Fig. 1: Inter-session coordinate variability RMS compared with Berne rule-of-thumb. Correct inter-baseline correlations.

Ambiguity resolution with Bernese proceeds in this way, that first the wide-lane (L5) ambiguities are resolved, and then the L1/L2 ambiguities to the extent possible. It was found that it is possible to resolve them wrongly, which shows up as grossly enlarged inter-session variability. This problem can be suppressed by enlarging the GPSEST sigma menu's "at most within" parameter from the standard 3σ to, e.g., 6σ.

Lesson: *Always* derive inter-session variability figures! No other accuracy measure will do, certainly not formal (internal) adjustment precision results.

On very short baselines (several km, cf. Vermeer and Paunonen 1994) it was found that millimetre precision requires resolution of the ambiguities. For longer baselines, complete resolution is no longer possible and for 24 hour sessions does not appear to offer any precision advantage anyway (M. Poutanen, personal comm.). It is possible that the technique of QIF ambiguity resolution (Mervart 1995) here produces improvement, there are some indications for this.

2.3 Parameters affecting GPS precision

2.3.1 Resolving for ionosphere on short baselines

It has been found that, on short baselines, it is better not to try to eliminate the effect of the ionosphere, as it will only be a small differential effect anyway. Elimination requires the construction of a so-called *ionosphere free linear combination*, also called L3:

$$L_3 = L_1 - \left(\frac{f_1}{f_2}\right)^2 L_2 .$$

It turns out that the noise level, in metric measure, of this linear combination is some three times higher than that of the constituent observables. Therefore it is better in this case to use the L1 and L2 observations separately, to increase the volume (weight) of the observations. The critical baseline length below which this is worthwhile, is about 15 km under typical conitions.

2.3.2 *Estimating the troposphere*

Geodetic GPS software allows the estimation, as unknowns, of tropospheric delay parameters, also called *tropospheric zenith delays*. This possibility is afforded by the fact that tropospheric delay is proportional to $1/\cos z$, z being the elevation of the GPS satellite, and by the large number of satelllites in the sky at any time under the fully deployed GPS constellation of today.

In a small network it will not be possible to estimate absolute tropospheric delay parameters for all stations. This problem is singular, due to the very similar observation geometry at each station. This is solved by *fixing* the tropospheric delay parameter for one station, after which all the others become estimable. These estimated parameters are then of course only *relative* delay parameters. In a large (e.g. Finland-sized) network, it is possible to estimate absolute tropospheric delay parameters.

Another possibility to obtain absolute tropospheric delays is fixing them using a *water vapour radiometer*. These are expensive instruments and the only one existing in the Nordic area is in Onsala, Sweden.

The Bernese software allows the estimation of tropospheric delay parameters. One should enter the statistics i.e. standard deviation of the tropospheric delay parameter itself, and of the amount of change in it allowed from epoch to epoch. Epoch interval is typically several hours.

With Bernese *one should not forget* to specify the use of a standard atmospheric model e.g. Saastamoinen (1973) or Hopfield! If one forgets this, the software will estimate *total* tropospheric delays (of order 2.3 m) rather than the (mm or cm level) *anomalous* delays, as one should. Especially in networks with significant height differences this mistake will produce large deformations!

The tropospheric zenith delay consists of two parts: The wet delay (water vapour) and the dry delay (the bulk mass effect of all gases). It is known that water vapour, in relation to its mass, is much more effective than other gases in delaying microwaves. This is due to the permanent dipole nature of water molecules, important also for its role in supporting life processes.

We can write for the refractivity of wet air according to the Essen formula (Laurila, 1983):

$$N = \frac{77.62}{T} p + \left(\frac{3.718 \cdot 10^5}{T^2} - \frac{12.92}{T}\right) e = 77.62 \frac{p}{T}\left(1 + 4790 \frac{e}{pT} - 0.1665 \frac{e}{p}\right), \qquad (1)$$

where the term in p and the ones in e are called *dry* and *wet refractivity*, respectively. T is temperature in Kelvin, p air pressure, e water vapour partial pressure, both in mbar.

We may write for the pressure change *in mbar*:

$$dp = -g\rho dh = -0.01g\frac{m}{22.414}\frac{273.15}{T}\frac{p}{1013}dh,$$

where g is ambient gravity, 9.8 ms^{-2}, and m average molecular weight:

$$m = \frac{29.0(p-e)+18.0e}{p} = 29.0 - 11.0\frac{e}{p},$$

for an atmosphere consisting only of nitrogen, oxygen and water vapour. All this yields

$$dp = -\left(0.0342\frac{p}{T} - 0.0130\frac{e}{T}\right)dh = -0.0342\frac{p}{T}\left(1 - 0.38\frac{e}{p}\right)dh \Rightarrow$$

$$\Rightarrow dh \approx -29.24\frac{T}{p}\left(1 + 0.38\frac{e}{p}\right)dp, \qquad (2)$$

exploiting the smallness of e / p.

Now the integrated tropospheric delay becomes (in the zenith):

$$\Delta^{Trop} = 10^{-6}\int_0^\infty Ndh =$$

$$= 10^{-6}\int_0^\infty 77.62\frac{p}{T}\left(1 + 4790.\frac{e}{pT} - 0.1665\frac{e}{p}\right)\cdot -29.24\frac{T}{p}\left(1 + 0.38\frac{e}{p}\right)dp$$

$$= -0.00227\int_{p_0}^0\left(1 + \frac{e}{p}\left[\frac{4790.}{T} - 0.1665 + 0.38\right]\right)dp \approx$$

$$\approx 0.00227\left(p_0 - \int_0^\infty \frac{e}{p}\left[\frac{4790.}{T} + 0.2135\right]\cdot -0.0342\frac{p}{T}dh\right) =$$

$$= 0.00227\left(p_0 + \int_0^\infty\left[163.8\frac{e}{T^2} + 0.0073\frac{e}{T}\right]dh\right)$$

For a dry atmosphere we find the expression

$$\Delta_d^{Trop} = 0.00227 p_0,$$

a simple function of ground level total pressure p_0.

Now using the "flat Earth" (i.e. horizontally layered atmosphere) approximation and ignoring ray path curvature, we derive straightforwardly:

$$\Delta^{Trop} = \frac{0.00227}{\cos z}\left(p_0 + \int_0^\infty\left[\frac{163.8}{T^2} + \frac{0.0073}{T}\right]edh\right). \qquad (3)$$

(In the above derivation we assumed a constant approximate value for gravity. More accurate calculation would require use of the true gravity at the latitude considered, at one atmospheric scale height above ground. Also the true composition of the dry atmosphere including argon

and varying amounts of carbon dioxide, should be considered, and also the elevation dependency law, or *mapping function*, could be refined.)

Compare the above expression to the one given by Saastamoinen (1973):

$$\Delta^{Trop} = \frac{0.002277}{\cos z}\left[p_0 + \left(\frac{1255}{T} + 0.05 \right) e_0 - B \tan^2 z \right] + \delta_R ,$$

(with B and δ_R being tabulated corrections depending in station height only (B) or on station height and zenith distance z (δ_R). Clearly some *hypothesis* concerning the behaviour of e with height, or with pressure, has been used to derive this.

We thus see that the *dry* tropospheric delay is to good approximation directly proportional to the mass of the column of air above the observation station, which can be measured directly and precisely using a *barometer*.

Unfortunately such a straightforward and inexpensive way to determine the *wet* delay is not available. In the Saastamoinen formula above, the term containing the humidity e has been derived *for a standard atmospheric model* and will often bear little relation to physical reality within the atmosphere. Other refraction formulas (models) exist, but suffer from the same weakness where the wet part is concerned.

For *measurement* of water vapour profiles we already mentioned the costly water vapour radiometer possibility; another one is launching weather balloons for humidity profiling. Note that measuring humidity at the GPS station *is no good* as there is no reason why humidity measures in the boundary layer close to the ground should in any way be representative of water vapour content *anywhere* higher up!

Equipping permanent GPS stations with registering barometers is an attractive possibility and is being planned for the Finnish permanent GPS network. Also in the field using barometers is a possibiliy and was widely practiced before surface meteo measurements came into disrepute. We would like to bring this possibility under renewed consideration, however, at least for air pressure. Best would be a digital barometer integrated with the GPS receiver and its registration system. We are soon moving into the 21st century!

Once air pressure values are available, we obtain the value of an integral over the partial water vapour pressure (also called precipitable water vapour) as follows (and note that, to be able to estimate the actual water vapour content, we need an average temperature value for the lower troposphere!):

$$\Delta^{Trop}_{Zenith} - 000227\, p_0 = \int_0^{\infty} \left[\frac{0.3718}{T^2} + \frac{0.0000166}{T} \right] e dh .$$

GPS software should then be modified to transparently apply the dry correction from the air pressure readings and allow estimation of these *wet-only* anomalous tropospheric zenith delays, which have (probably) a much nicer and more modest statistical signature than the total tropospheric signal. Its variation could even be completely ignored over baselines of considerable length.

2.3.3 *Effect of antenna phase delay patterns*

The measured phase of the carrier wave depends on the direction from which the radio wave reaches the antenna, i.e. the elevation and azimuth of the satellite in the sky as seen from the receiver location. The strongest dependence is upon elevation, but also azimuth dependence cannot be ignored. The patterns of phase delay differ between L1 and L2, between different receiver brands, between antennas of different type number belonging to the same receiver and (for some antenna types) even between individuals!

There has been a substantial body of publications concerning these antenna phase delay patterns. We can mention Schupler and Clark (1991); Schupler *et al.* (1994); Dunn and Young (1992); Rocken (1992); and many more. One effect that was noticed especially, was that trying to estimate tropospheric zenith delay parameters using different antennas, without accounting for the different phase delay diagrams, may produce height errors of some 20-30 mm.

For short baselines the satellite direction is virtually the same on both ends of the baseline so, if identical antennas are used, the problem of the phase delay pattern does not occur. For somewhat larger baseline length only a small, differential effect comes in. If antenna types are mixed, however, the problem enters in its full severity. This is also the case for long (continental scale and longer) baselines, where antenna orientations (local vertical) are completely different. All these situation require *accurate antenna calibration tables* to be available.

Laboratory calibrations (e.g. Dunn and Young, 1992 and Schupler *et al.*, 1994) meet with difficulties. Even if the problem of accurately positioning the antenna to be calibrated with respect to the microwave source can be solved, there remains the problem of reflections off laboratory structures. Laboratory calibration should be done in an *anechoic chamber*, and such tests are quite costly.

The Bernese group (Rothacher, personal comm.) has been working on this problem along the lines of *field calibration*. In this way, only *relative* calibration tables can be constructed, i.e. one antenna type, e.g. Dorne-Margolin, should be taken as "known" and the relative values found be added to the absolute (anechoic chamber) values of the Dorne-Margolin. The relative (inter-antenna) accuracy of calibration can be quite good: a manuscript by Mader and MacKay (1995) demonstrates successful calibration of elevation-only dependent phase correction values for seven different antenna types.

Yet, if possible one should use antennas of the same type in one campaign, and this type should preferably be Dorne-Margolin Choke ring, because of their superior multipath suppression properties.

2.3.4 *Minimizing multipath*

Multipath, the reflection of signal from the Earth's surface and objects on it, and interference of these reflected signals with the directly travelling signal, is a serious limiting factor on the accuracy of geodetic GPS. The problem can be minimized by antenna design; the so-called *choke-ring antennas* provide a way to suppress signal coming from below the antenna's horizon.

The most popular design of choke-ring antenna is called Dorne-Margolin type "T"; originally designed at Jet propulsion Labs and manufactured by Allen Osborne Inc. (the builders of the

Rogue series of geodetic receivers), now functionally equivalent models are being marketed by several other GPS firms, including Ashtech and Trimble. These are the antennas used throughout the Nordic area for permanent GPS reference stations.

One other possible method to cope with multipath is to put the antenna on a very high mast, far away from the Earth's surface. Making this mast completely rigid is practically impossible. Alternatively one could try to *measure* (i.e. continuously monitor) the position of the top of the pole relative to a fixed point down on the Earth, and correct for its sway.

One could even use GPS for this measurement. Remember that for a baseline of only several metres, both ionosphere and troposphere can be completely ignored and only one frequency need be used. This means high precision, low sensitivity to multipath, and no need to use satellites down to just above the horizon. The idea, which may be called the "virtual antenna" concept, is sketched in the picture below (Fig. 2).

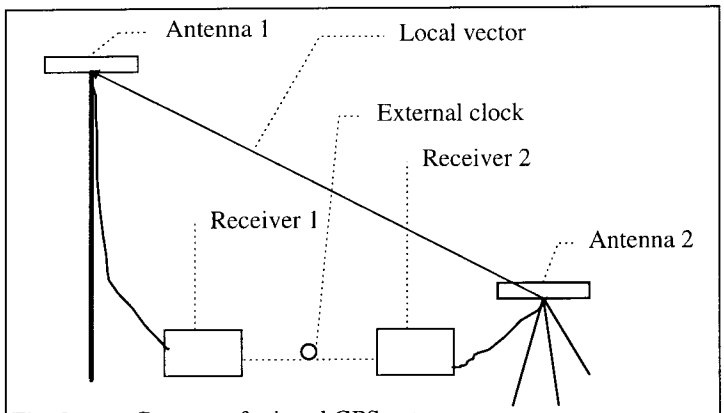


Fig. 2: Concept of virtual GPS antenna.

This picture actually gives a *test set-up*, where two existing geodetic receivers are used to test the method. In actual measurements one would use two such set-ups, one at every site. The height of the "high" antenna could be 5 m or more, depending on vegetation.

In GPS measurements there are no less than *three unknowns* dependent on satellite elevation: vertical station position, station clock offset, and tropospheric zenith delay. Determining all three for a given baseline requires the availability of satellites on three significantly different elevations above the horizon. Even then, estimates of these three unknowns will be of low precision and rather strongly correlated.

The dependence on satellite elevation of these three unknowns are:

Station clock:	1
Troposphere:	1 / cos z ("*mapping function*")
Vertical posision:	cos z

It has been observed (e.g. Mader and MacKay, 1995) that under these conditions small errors are amplified, e.g. the effect of different antenna phase delay diagrams on both ends of the vector may easily produce 20 to 30 mm vertical position errors.

Now for short vectors it is possible to eliminate two of these unknowns: the troposphere (being the same on both ends of the vector, or the difference being precisely computable) and the station clock unknown (by using a common external oscillator on both ends). If this is done, only one unknown remains, the vertical component of the vector. For estimating this, and the two horizontal components, three satellites would suffice, separated by 90° in the sky.

Not only the sensitivity of the vector solution to multipath will be reduced by an order of magnitude or so, also the requirements to be placed on the site (low horizon!) can be substantially relaxed.

The above test set-up can be streamlined for operational use, by mounting the mast on the back of a pick-up and putting all the receiver electronics needed in one box. Instead of using two geodetic-quality receivers, one can use one such receiver plus an inexpensive single-frequency receiver for local vector monitoring, and a beam splitter. And of course correcting for this vector to the reference point of the "low" antenna should be done in real time on-site. But we can safely leave all that to the GPS developers, if the concept is found viable!

2.4 Determination of deformation parameters

In analysing network deformations, one should always do the adjustment of the GPS observations in the form of *free network adjustment*. Any attempt to introduce "known coordinates" from a number of reference points will lead to an overconstrained network, and any crustal deformations affecting the reference points will enter the network as a whole and deform it in an intractable way.

In Finland, one can do a free network adjustment by e.g. keeping only the reference station Sjökulla fixed. Properly this is done by first obtaining Sjökulla's coordinates in a recent ITRF frame in the epoch of observation; then doing the GPS adjustment e.g. with Bernese; and then transforming back to a common reference frame and epoch in which the study of possible deformations is to be done.

One may ask if this complicated procedure is really necessary. Can one not simply do the GPS adjustment by keeping Sjökulla fixed on one single set of coordinate values? We investigated this problem empirically by performing a small test, where we perturbed all three coordinates of the fixed points in a six point network (Metsähovi and Sjökulla) with the same amount, chosen to be first 1 mm, then 10 mm, and finally 100 mm. In a small network, the effect of such a perturbation would be to parallelly shift all point coordinates in the network with exactly the same amount; but Finland is not small. Our network has a size of 800 km, which subtends an angle of 2° seen from the GPS satellite. Below we tabulate the deviation from "small network behaviour" for these various simulated perturbances.

Table 1: Anomalous effect of perturbing fixed station coordinates. Unit: mm. One 24 hour session (day 236) only.

	Perturbance 100 mm			Perturbance 10 mm			Perturbance 1 mm		
Tuorla	-1.4	+0.7	-0.4	0	+0.1	0	0	0	0
Virolahti	+0.9	-0.9	+0.2	0	-0.1	0	0	-0.1	0
Vaasa	+0.3	+2.0	-1.8	0	+0.2	-0.1	0	0	0
Oulu	+2.2	+1.6	-2.3	+0.2	+0.1	-0.2	0	0	0
Joensuu	+2.2	-0.9	-0.5	+0.3	-0.1	-0.1	+0.1	0	0
Sodankylä	+3.3	+2.5	-3.2	+0.3	+0.2	-0.3	0	0	0

We see clearly that already a coordinate perturbance of 0.1 m leads to several mm change in the *relative* computed locations of the network points. The lesson is clear: The coordinates of the fixed points have to be correct (i.e. in the same datum as the satellite orbits used) within just a few cm.

This means that a datum connected to the Eurasian plate already loses its usefulness for computation after a few years. The computation must be done in some ITRF-datum; ETRF coordinates can only be produced afterwards by transformation.

Research into the best way to extract the deformation information is ongoing (M. Poutanen, personal comm.). An attractive alternative is the use of invariant so-called "shape elements" such as (in the plane) horizontal angles and ratios of distances (Baarda, 1973), which can be expressed as complex numbers. Another possibility, used by Chen (1991) is the conversion of inter-point distances in the network into deformation tensor elements.

3 Crustal motions in the Fennoscandian area

In the Fennoscandian area occur the following types of crustal motion which may be studied by GPS:

- The continental drift, moving the whole Eurasean plate as a unit across the surface of the Earth. This motion is currently well known and routinely observed by the global IGS (International GPS Geodynamics Service) network.
- The postglacial rebound or land uplift. This movement, traditionally studied by sea level registration (mareographs) and repeated precise levellings, is actually three-dimensional and excellently suited to be studied by GPS.

  Study of the current and historical rebound motion allows one to build models of the ease with which the mantle material flows, i.e. its *viscosity profile* with depth (ignoring lateral inhomogeneities) or even its spatial viscosity distribution (much more difficult). Cf. Ekman (1992) for a review article.
- Intra-plate geotectonic motions. It is known that the crust in the Baltic Shield area is pervaded by numerous deep fault lines within the lower crust, at many tens of kilometres depth. Along these lithospheric fault lines, slow motions are believed to take place (under one centimetre per year, i.e. an order of magnitude less than plate tectonics). As the fault lines are not quite straight, small enclosed areas called "stress ovals" may be squeezed up (Gotland) or down (Vänern). If a GPS station is located on such a piece, it may undergo anomalous vertical motions easily confused with postglacial rebound. This stresses the need for adequate spatial resolution.
- Earth tidal motion, including ocean and atmospheric loading.

4 Local GPS geodynamics studies

4.1 The test network in Nuottavaara

In the Nuottavaara geodynamic GPS test network repeated measurements with GPS have been performed in 1991, 1992 and 1995 (Poutanen and Ollikainen 1995). In 1991 and 1992 Ashtech P-12 receivers were used; in 1995, Ashtech Z-12 receivers.

In this small test area a network in the form of a tetragon was established; of the four points, two were located on each side of the known line of fracture. All these points were in the bedrock or connected reliably to nearby bedrock. Baseline lengths were from 1500 to 3000 m.

Processing was originally done with Ashtech's GPPS software, and repeated with Bernese. Within the network, the L1 and L2 carrier phase observations were both used but without forming the ionosphere-free L3 linear combination. This is preferable (lower noise level) for a small network like this, as discussed above: the ionospheric refraction will be the same on both ends of every baseline. Also no tropospheric zenith delay parameters were estimated.

Baseline standard deviations found were generally under 1 cm and no significant movements of the network points were found during the monitoring period. This may well reflect the obvious fact that even along an active fault line segment, tectonic motions will only occur once every so many (i.e. *many!*) years, in other words, a sampling problem.

We intend to continue these repeat measurements also in the future.

4.2 The Central Finland measurements

These measurements done in 1991 and processed by Ruizhi Chen (1992) were aimed at detecting larger-scale geodynamic movements in the Central Finland area. Processing of the GPS measurements was done together with classical terrestrial measurements known to be very sensitive to atmospheric refraction. Because of this circumstance, and because of the use of Ashtech L-XII receivers of an older type, one should unfortunately say that this study does not represent the state of the art any more.

However, when later GPS measurements will be made in the same area in the EUREF densification framework, it may be possible to combine these with the older GPS measurements to obtain a better way of studying possible geotectonic movements in this area.

4.3 The monitoring measurements in Olkiluoto, Romuvaara and Kivetty

In the permanent GPS sites Olkiluoto, Romuvaara (Konginkangas) and Kivetty (Kuhmo), repeated local geodynamic GPS measurements are being conducted in order to detect possible movements in the Earth's crust. This work is done in commission for the Finnish power companies, as these are candidate areas for the underground storage of nuclear waste.

These measurements are done at half-yearly intervals; the size of the local networks involved is several kilometres (Chen and Kakkuri 1994). Tests have shown that, with precautions, millimetre precision can be achieved in principle over a one kilometre baseline. The precautions needed are related to the multipath or ground reflection problem: one should make sure that the antenna is at least 2 m above the ground, that no reflecting objects are present in the vicinity of the antenna, and that the duration of measurement is at least six hours in order to average out any differences in satellite geometry. Unfortunately it is in practice impossible to make the satellite geometry identical for the measurement periods being compared.

5 Regional GPS geodynamics studies

5.1 The postglacial rebound and GPS

Only during the last few years has there been a realistic possibility to monitor the postglacial rebound by means of GPS. The same period has seen the establishment of permanent GPS monitoring networks in the countries located in the Fennoscandian land uplift area. This situation promises important progress in coming years in understanding the mechanism of the postglacial rebound, including important mechanical properties (viscosity) of the Earth's mantle.

5.2 The DOSE (Dynamics of the Solid Earth) project

The DOSE project, in which also the Finnish Geodetic Institute has actively participated, is an international collaboration aiming at better understanding the mechanical properties of the Earth's mantle as expressed in the postglacial rebound in Fennoscandia. Partners in this particular project, that goes under the name BIFROST, are the Cambridge Astrophysical Observatory, the Physics Department of Toronto University, the Onsala Space Observatory in Sweden, and the Finnish Geodetic Institute.

In 1993, the first GPS campaign was organized; at that time the permanent GPS network was only in its first stages of deployment and most of the Finnish sites were occupied by Turbo Rogues received on loan from the UNAVCO (University NAVSTAR Consortium) in the USA.

In 1994, the second GPS campaign took place; many of the Finnish permanent GPS stations were already operational, but Oulu and Vaasa still had to be provided with a temporary Turbo Rogue. This second campaign had as its objective the attachment of tide gauges to the permanent GPS station framework. For the first time, also the Riga Observatory (Botanical Garden) was along as a site, in collaboration with the University of Latvia's Astronomical Institute.

5.3 FinnNet, the Finnish Permanent GPS network

This important geodynamic GPS resource has already been mentioned; we describe it here in detail. Establishment of the network has progressed towards completion during the years 1992-1996. It is a multipurpose permanent GPS monitoring facility comprising 12 stations covering the Finnish territory. The stations are depicted in Fig. 3 (circles).

All stations except Sodankylä are on bedrock, the antennas being mounted on 2.5 m tall steel grid masts (Fig. 4). In Sodankylä, a concrete pedestal was cast in a deep pit extending below the frost line. The soil is sandy, so cryoturbation should not be a problem there. In Oulu, an existing steel tower of 8 m height was adapted and provided with an invar wire height

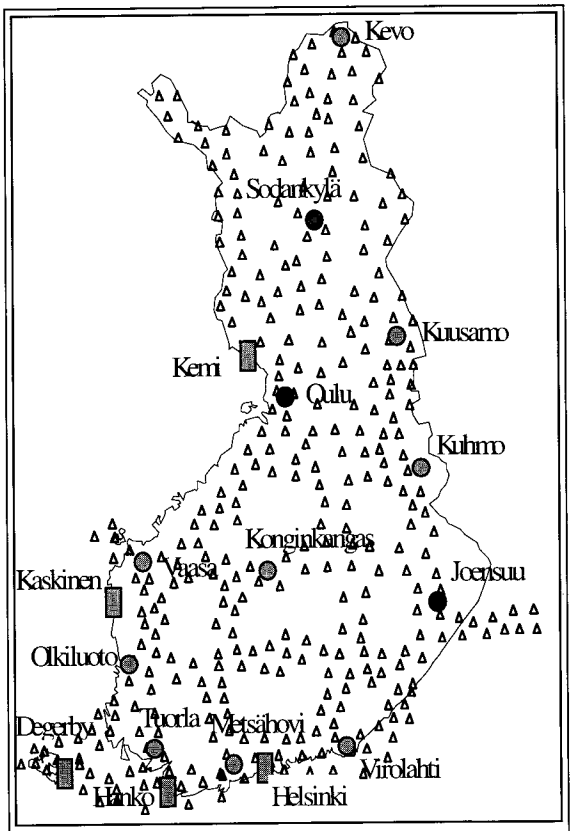


Fig. 3. The Finnish permanent GPS network *FinnNet* (circles) and the EUVN mareograph pillars (rectangles). Black stations support Differential GPS RTCM correction dissemination by the Finnish Broadcasting Company.

The small triangles mark the primary triangulation network.

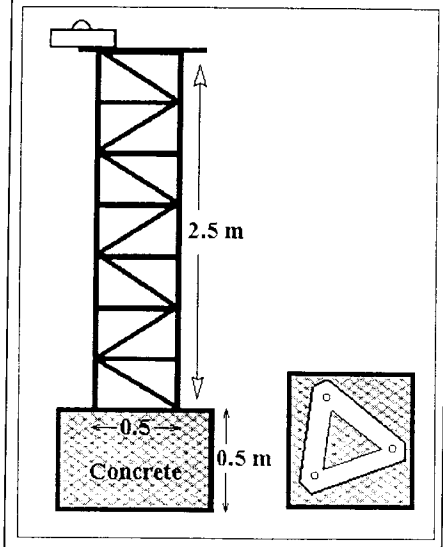


Fig. 4. A typical GPS antenna mast.

control system similar to the one in Metsähovi.

Part of the stations are provided with 2 by 3 m wooden instrument cabins; on part, existing buildings could be used.

For geodynamics studies, the incoming data of the permanent network (sampling rate 30 s) is routinely processed using the Bernese software and precise ephemeris once these are available. This processing routine is nearly operational. Working routines exist already for data download and archiving on writable CD-ROM disks.

We intend, as already discussed earlier, to process the Finnish data as a *free network* first, in order to better detect possible crustal motions. Processing of daily, fortnightly and yearly data should already give values for the station motion due to postglacial rebound. Also tectonic motions should jump to the eye from these results. The free network coordinate solution will be formatted as so-called SINEX files (Software Independent Exchange) and made available, for the EUREF stations concerned, for EUREF processing, currently planned to take place at the NKG computing centre at Onsala (J. Johansson, personal comm.). Also geodynamics studies will be undertaken there, using data from all the Nordic countries.

5.4 The planned EUREF densification network

The local test measurements described in Chapter 3 are for small to medium size GPS networks specially selected for geodynamic interest, either because crustal motion are known to take place there, or because the need to know of they do is especially strong. One should be aware, however, that intra-plate crustal tectonic motions are theoretically expected to be few and far in between.

For this reason, a better strategy might be to perform repeated measurements over a contiguous area, such as the whole territory of Finland, and repeat these at regular intervals. This would also allow statistical studies to be performed.

One network covering the whole of Finland, and offering a *very* good temporal resolution, is of course the permanent GPS network FinnNet. However, the number of stations is only twelve, so what is gained in temporal resolution is again lost in spatial resolution.

A *densification* of the permanent station network by campaign measurements is in progress; this so-called EUREF densification will measure 50 stations in 1996 and another 50 in 1997. Most of the stations selected are primary triangulation stations, for reasons of historical continuity, meaning that approximately every *third* triangulation station will become a GPS site (i.e. every site more or less suitable for GPS measurement). Of course, at some time in the future (e.g. after ten years) the network would have to be measured again, and then possible movements in the Earth's crust will be detected. The point spacing of this network is some 80

km, providing a good resolution for detecting also Earth surface displacements caused by deep fault zones.

6 GPS, sea level and the geoid

6.1 The Baltic Sea Level studies

6.1.1 The project

The Baltic Sea level (BSL) project aims at unifying the vertical datums around the Baltic Sea by a combination of GPS and geoid determination, as well as studying the level of the Baltic Sea and its possible changes by connection to mareographs. The project was a Special Study Group (no. 5.147) of the International Association of Geodesy (IAG), and presently continues as a Subcommission of Special Commission 8: Sea Level and Ice Sheet Variations. This Subcommission is called, as before, Studies of the Baltic Sea, and is chaired by J. Kakkuri.

In total, two GPS campaigns have been organized in the BSL framework: the one of 1990 (Kakkuri, ed., 1994) and the one of 1993 (Kakkuri, ed., 1995). A precursor was the Åland Campaign of 1987 (Kakkuri and Vermeer, 1988).

Besides GPS measurements, also a precise *geoid* is needed to be able to connect different vertical datums by GPS. A geoid determination effort for the Nordic area including the Baltic Sea is nearing completion in Copenhagen (Forsberg, personal comm.). In the meantime, a Baltic Sea geoid was already computed in Helsinki (Vermeer 1995).

6.1.2 The Sea Surface Topography from GPS

Another interesting approach, providing the *sea surface topography* (SST) is first deriving the absolute location in space of mean sea level using GPS and local tie levelling to a nearby mareograph, and then subtracting the geoid model available. This gives the sea surface topography in probably the most direct and hypothesis-free way possible. Of course one should remember that, if the geoid has been fitted to GPS/levelling points, all the systematic errors of the levelling network used will be present in the geoid also.

This method was used in Kakkuri and Poutanen (1996). Of course the situation was ideal around the Baltic Sea: good levelling was available to correct the gravimetrically derived geoid, which is not the case e.g. around the Mediterranean. Globally the method can be used if a good global geoid is available, e.g. the one produced by a future gradiometric satellite mission (GOCE, see e.g. ESA 1996). In this way also the height systems of the continents can be connected to each other, probably on the decimetre level (Rummel and Xu, 1991).

6.2 The EUVN (European Vertical GPS Reference Network) inititative

This initiative (Ihde *et al.*, 1996) originated from a EUREF decision (Gubler and Hornik, eds., 1994) to become more active in the field of vertical reference systems. Since the Boulder meeting of the International Union of Geodesy and Geophysics, of which the IAG is a member, EUREF (Subcommission for the European Reference Frame) and EULN (European United Levelling Network) were merged into one Subcommission named EUREF.

In a way, the EUVN proposal can be seen as the extension to all of Europe of the original Baltic Sea Level Project idea. We may take this to mean that the concept of using GPS to

unify vertical datums has come to age: imitation is the sincerest form of flattery. We have come a long way since the Åland GPS campaign of 1987 (Kakkuri and Vermeer, 1988)!

In Finland we will integrate the proposed EUVN measurements with the third BSL GPS campaign, to take place during the same time, May 21-28, 1997. We shall do this in such a way, that the overall framework over Europe and the Baltic Sea area will be formed by EUVN stations, permanently monumented and occupied for the full 8 x 24 hours to achieve maximum precision especially in the vertical coordinate; but additionally we will occupy with roving GPS receivers interlying mareograph stations for shorter times (1 to 2 times 24 hours) to connect them to the larger-scale EUVN framework. Due to the shorter vectors, these points will not suffer any loss in precision.

The Finnish mareograph stations provided with monuments (pillars) are: Helsinki, Hanko, Degerby Åland, Kaskinen and Kemi. Fig. 5 gives an idea of what a typical pillar looks like. They are designed to meet, as far as possible, the requirements formulated in Gurtner (1995).

6.3 Sea level studies and global change

The absence of a global vertical reference system is felt most painfully by students of global change, especially the possible change in sea level due to warming of the climate due to the greenhouse effect. This sea level rise, due to a combination of continental ice sheet melting and increase in average sea water temperature, can only be studied properly in a global context. This is the aim of the GLOSS programme (Global Sea Level Observing System) initiated by the Intergovernmental Oceanographic Commission (IOC) in 1990. It aims at the establishment of some 300 scientific-quality tide gauges worldwide. Positioning these tide gauges regularly with the aid of GPS, as EUVN will do in the European area, is a valuable addition to this programme (Baker *et al.*, 1996).

7 GPS and solid Earth tides

7.1 Theoretical studies

A theoretical study was made by Poutanen *et al.* (1996) into the way the permanent part of the solid Earth tide should be corrected for in processing GPS observations. It was concluded that, contrary to the practice in gravity, height and geoid work, GPS coordinates do not refer to the *zero crust* (see e.g. Ekman 1989), but to the *non-tidal crust*, involving the choice of a conventional value for the Love number k of the solid Earth. This is an unfortunate incompatibility between standards, which is difficult to correct now.

7.2 Potential for observational work

Models of the solid Earth tide as currently used in GPS work are global and somewhat simplified. Because the true Earth is more complex than this, one may expect a development in the direction of empirically determined tidal parameters, different for every site.

This is already the common practice for gravimetric, Verbaandert-Melchior and long water tube tilt meter work (e.g., Kääriäinen and Ruotsalainen, 1989). We are quickly approaching the point where also GPS observations continued over long periods can be used to estimate at least some of the tidal coefficients, or rather, their deviations from the standard values, which will be a few centimetres at most, and less for most sites.

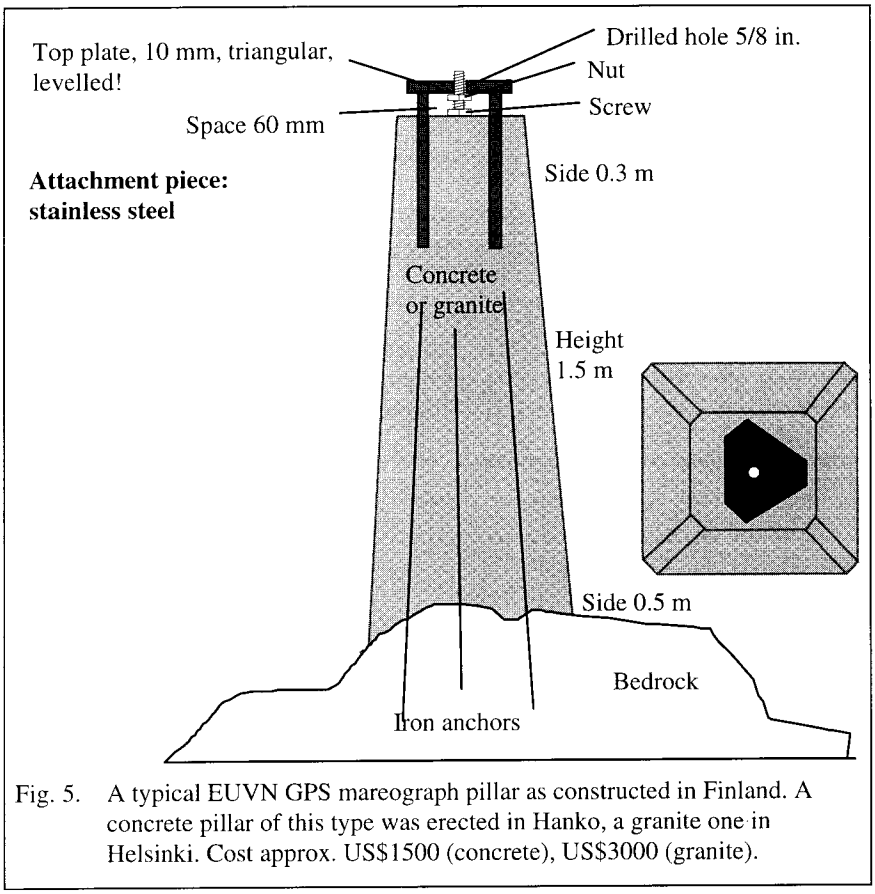


Fig. 5. A typical EUVN GPS mareograph pillar as constructed in Finland. A concrete pillar of this type was erected in Hanko, a granite one in Helsinki. Cost approx. US$1500 (concrete), US$3000 (granite).

However, it may be difficult to separate some of the tidal parameters from polar motion parameters and one should beware, in formulating theoretical models, of introducing singularities (M. Rothacher, personal comm.).

7.3 Ocean and atmospheric loading

In coastal regions, solid Earth tidal motion receives a clear signature from ocean tidal loading, an effect observable in precise gravimetry and clinometry. Also the atmosphere undoubtedly influences on the coordinates of GPS points (Scherneck, 1994). None of these effects have been unambiguously observed in GPS data, but they must be taken into account in precise geodynamic GPS processing.

If air pressure sensors are installed in permanent GPS stations, these can also be used to routinely estimate air pressure crustal loading corrections to the GPS coordinates. Possibly elasticity parameters can be estimated at the same time, yielding a semi-empirical technique.

8 Conclusions

We have described the various aspects of the study of geodynamics with the aid of GPS as practiced at the Finnish Geodetic Institute. It is a broad field and we have not attempted to

provide any very deep analysis of the various aspects. We have concentrated on outlining the theoretical approaches and practical solutions that we have arrived at.

Acknowledgement. Dr. Ruizhi Chen critically read the manuscript and provided some corrections.

References

BAARDA W (1973): S-transformations and criterion matrices. Netherlands Geodetic Commission Publ. On Geodesy, New series Vol. 5 No. 1.

BAKER TF, PL WOODWORTH, G BLEWITT, C BOUCHER, G WÖPPELMANN (1996): A European Network for Sea Level and Coastal Land Level Monitoring. Subm. *Journal of Marine Systems.*

BEUTLER G, I BAUERSIMA, S BOTTON, W GURTNER, M ROTHACHER, T SCHILDKNECHT (1989): Accuracy and biases in the geodetic application of the Global Positioning System. *Man. Geodaetica* 14 pp. 28-35.

CHEN R (1991): On the horizontal crustal deformations in Finland. Report 91:1 Finnish Geodetic Institute.

CHEN R (1992): Processing the 1991 GPS Campaigns in Central Finland for crustal deformation studies. Report 92:3 Finnish Geodetic Institute.

CHEN R, J KAKKURI (1994): Feasability study and technical proposal for long-term observations of bedrock stability with GPS. Report YJT-94-02, Nuclear Waste Commission of Finnish Power Companies, Helsinki.

DUNN CH, L YOUNG (1992): TOPEX GPS Antenna Phase Calibration. Jet Propulsion Laboratory interoffice memorandum 335.9-92-016, 29 April, 1992.

EKMAN M (1989): Impacts of geodynamic phenomena on systems for height and gravity. *Bull. Géod.* 63 pp. 281-296.

EKMAN M (1992): Postglacial rebound and sea level phenomena with special reference to Fennoscandia and the Baltic Sea. In: Geodesy and Geophysics, Lecture notes NKG Autumn School, Korpilampi Finland. Publication 115 Finnish Geodetic Institute.

EUROPEAN SPACE AGENCY (1996): Gravity Field and Steady-State Ocean Circulation Mission. Coordinator Chris J. Readings. Reports for Assessment: The Nine Candidate Earth Explorer Missions, ESA SP-1196(1), Noordwijk.

GUBLER E, H HORNIK (ED.) (1994): Proceedings EUREF Symposium Warsaw, Poland, 8-11 June 1994. Munich.

GURTNER W (1995): Guidelines for EUREF Tide Gauge Campaign — Draft Proposal. In: Proceedings EUREF Symposium Helsinki, Finland, 3-6 May 1994, eds. E. Gubler and H. Hornik, Munich.

IHDE J, W SCHLÜTER, W. GURTNER, G. WÖPPELMANN, BG HARSSON, J ADAM (1996): Proposal for the European Vertical GPS Reference Network (EUVN). Annex to a circular letter by EUREF.

KAKKURI J, ed. (1994): Final results of the Baltic Sea level 1990 GPS campaign. Research works of the SSG 5.147 of the International Association of Geodesy. Report 94:2 Finnish Geodetic Institute.

KAKKURI J, ed. (1995): Final results of the Baltic Sea level 1993 GPS campaign. Research works of the SSG 5.147 of the International Association of Geodesy. Report 95:2 Finnish Geodetic Institute.

KAKKURI J, M POUTANEN (1996): Geodetic determination of the Surface Topography of the Baltic Sea. Subm. *Marine Geodesy.*

KAKKURI J, M VERMEER (1988): The Åland GPS levelling experiment. Proceedings, Integrated Geodesy Symposium, Sopron, Hungary 16-20 May 1988, pp. 27-35.

KÄÄRIÄINEN J, H RUOTSALAINEN (1989): Tilt measurements in the underground laboratory Lohja 2, Finland, in 1977-1987. Publ. 110 Finnish Geodetic Institute.

LAURILA S (1983): Electronic surveying in practice. Wiley-Interscience, New York.

MADER GL, JR MACKAY (1995): Calibration of GPS Antennas. Draft June 11, 1995. Http://www.grdl.noaa.gov/GRD/GPS/Projects/ANTCAL/antcal_paper.html.

MERVART L (1995): Ambiguity Resolution Techniques in Geodetic and Geodynamic Applications of the Global Positioning System. Inauguraldissertation der Phil.- Nat. Fakultät der Univ. Bern.

POUTANEN M, M OLLIKAINEN (1995): GPS measurements at the Nuottavaara post glacial fault. Report 95:6 Finnish Geodetic Institute.

POUTANEN M, M VERMEER, J. MÄKINEN (1996): The permanent tide in GPS positioning. *Journal of Geodesy* 70:8, pp. 499-504.

ROCKEN C (1992): GPS Antenna Mixing Problems. UNAVCO Memo, November 12, 1992.

RUMMEL R, PL XU (1991): A quality investigation of global vertical datum connection. Netherlands Geodetic Commission Publ. On Geodesy, New series No. 34.

SAASTAMOINEN J (1973): Contributions to the theory of atmospheric refraction. Part II. Refraction corrections in satellite geodesy. *Bulletin Géodésique* 107 pp. 13-34.

SCHERNECK H-G (1994): Realistic Modelling of Deformations due to Air Pressure Variations. In: Proc. 12th General Meeting Nordic Geodetic Commission, Ullensvang, Norway 30 May - 3 June 1994.

SCHUPLER BR, TA CLARK (1991): How different antennas affect the GPS observable. *GPS World* November/December 1991.

SCHUPLER BR, RL ALLSHOUSE, TA CLARK (1994): Signal Characteristics of GPS User Antennas. Preprint of article accepted for *Navigation.*

VERMEER M (1995): Experiences in computing the BSL95A Baltic Sea gravimetric geoid. In: Two new geoids determined at the FGI. Report 95:5 Finnish Geodetic Institute.

VERMEER M, M PAUNONEN (1994): The vector connecting the fundamental points Metsähovi and Sjökulla. In: Proceedings EUREF Symposium Warsaw, Poland, 8-11 June 1994, eds. E. Gubler and H. Hornik, Munich.

Geoid Information and GPS - a Review and Nordic Status

Rene Forsberg
Kort og Matrikelstyrelsen, Rentemestervej 8
DK-2400 Copenhagen NV, Denmark
E-mail: rf @ kms.min.dk

Table of Contents

1. Introduction

Geoid determination - the classical primary objective of physical geodesy - has become an area of increasing importance due to the advance of GPS techniques, and the eventual replacement of traditional levelling by GPS. Currently the goal of geoid determinations is to provide accurate geoid surfaces at the cm to dm accuracy level, a big challenge which has progressed rapidly in recent years due to developments in global models, regional data availability and the development of new, more efficient computation methods.

The basics for the recent surge of interest in the geoid and geoid determination is the simple equation

$$h = H + N \tag{1}$$

relating heights h above the ellipsoid to the orthometric (sea-level) height H and geoid height N. However, the problem of obtaining heights from GPS ellipsoidal heights is not as simple as it might look at a first glance: All three quantities in (1) generally relate to different *datums*: h is typically in a local WGS-84 system, which might be anywhere from 1 cm to several m away from the best "global" ellipsoidal heights (as e.g. expressed in recent ITRF coordinates); H refers to a local height datum based on sea-level (which might be different from the global average sea level by up to 0.5-1 m due to local sea surface topography); and regional gravimetric geoid models refer to an implicit global system which is different from the two other. Equation (1) therefore only "fits" to within some 0.5 m typically .. unless the geoid is *constrained* to local GPS/levelling data. Only in this case can we talk about dm- or

cm-geoids - a common source of confusion for the GPS surveyor not familiar with physical geodesy and the inherent nature of the errors in geoid models.

Fig. 1. GRIM3-L1 spherical harmonic model geoid (from Torge: Gravimetry, 1989). Unit: m.

In this lecture I will address some of these issues will be addressed, and the related issues of different "types" of geoids - the *classical geoid* N and the *quasi-geoid* ζ - which relates directly to the definition of sea-level heights (*orthometric heights* or *normal heights*), classical research topics which nowadays have become practically relevant when trying to push the accuracy of local geoids toward the cm-level. The use of geoid information will be demonstrated by examples from the Nordic area. First, however, a brief review will be given of the fundamental geoid definitions, and the gravimetric geoid determination methods currently in use for regional or continental geoids.

2. Geoid and quasigeoid determination

The geoid is defined as the equipotential surface of the earth geopotential most closely corresponding to average sea level.

On the oceans the actual sea-surface deviates from the geoid by the effects of tides, currents, and variations in temperature and salinity, giving rise to *sea-surface topography*, the permanent part of which may be up to 1-2 m. The seperation of ocean sea-surface ellipsoidal height (which can be measured by satellite radar altimetry or GPS at tide gauges) and the geoid is of prime importance in oceanography, since the difference can give the information on currents etc.

On land the geoid is defined as the equipotential surface *inside* the topographic masses - i.e. the height the water would have above a reference ellipsoid in a narrow channel or tunnel, dug from the ocean into the a point P* at the zero level below the point P in question (Fig. 2). Because thc topography itself generates a gravitational field, the geoid is disturbed by the topography, and requires thus - at least in principle - a detailed knowledge of the bulk density

of the topographic masses for the precise computation of the geoid.

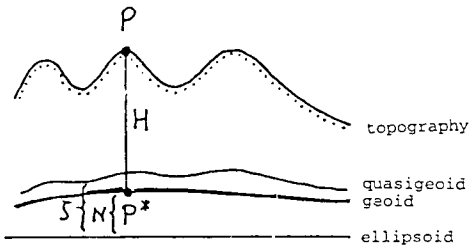


Fig. 2. Geoid and quasigeoid.

To circumvent this problem a modified surface - *the quasigeoid* - may be used, in a theory originally developed in a comlicated expansion theory by russian geodesists in the 1950's (Molodesky et. al., cf. Heiskanen and Moritz, 1967). The quasi-geoid may loosely be viewed as the *"geoid at the surface of the topography"*, and often the word "geoid" is used freely for any of the surfaces. Luckily a simple approximation links the two types of geoid, as outlined below, so in practice the question of geoid or quasigeoid is not a major issue.

2.1. The mathematical description of the gravity field

The details of the mathematical treatment of the gravity field is given in textbooks like Heiskanen and Moritz (1967). The geopotential, W, is conventionally split into an *ellipsoidal* and an *anomalous* potential

$$W = U + T \tag{2}$$

where U is the potential function of a selected reference ellipsoid (which may be computed at any point in space from a unique mathematical solution based on given ellipsoid parameters a, J_2, GM and ω), and T the residual *anomalous* potential. The geoid height relates directly to T through *Bruns formula*

$$N = \frac{T}{\gamma} \tag{3}$$

where γ is normal gravity, which may be considered nearly a constant (≈ 9.8 m/s^2) over local regions.

The potential T fullfills the Laplace equation (i.e., is a harmonic function) outside the masses

$$\nabla^2 T = 0 \tag{4}$$

whereas the potential inside the mass (e.g. below the topography on the continents) fullfils the Poisson equation

$$\nabla^2 T = -4\pi G\rho \tag{5}$$

where ρ is the density and G the newtonian gravitational constant. T inside the mass is thus *not* a harmonic function.

The anomalous potential is in principle a *spatial* function $T = T(r,\phi,\lambda)$, where r radial distance and (ϕ, λ) the latitude and longitude, respectively. The geoid may be viewed as the result of evaluating Bruns' formula (1) at sea level, whereas the evaluation of (1) at the topographic surface yields the quasigeoid ζ (in the original derivations by Molodensky et. al. a special surface - *the telluroid* - was used to rigorously define the quasigeoid, but it is more straightforward just to assume Bruns formula as the definition point for the quasigeoid).

The fundamental geoid determination formula is Stokes' formula

$$N = \frac{R}{4\pi\gamma} \iint_\sigma \Delta g \; S(\psi) \; d\sigma \tag{6}$$

which relates the geoid height to the gravity anomalies as an integral extending over the complete surface of the earth. The function S of spherical distance ψ is Stokes' function

$$S(\psi) = \frac{1}{\sin(\frac{\psi}{2})} - 6\sin\frac{\psi}{2} + 1 - 5\cos\psi - 3\cos\psi\log(\sin\frac{\psi}{2}+\sin^2\frac{\psi}{2}) \tag{7}$$

Stokes' formula is derived assuming T is a harmonic function (4), and thus assumes the topography to be computationally removed in some way.

The corresponding equation for the quasigeoid is Molodensky's formula (here to first order)

$$\zeta = \frac{R}{4\pi\gamma} \iint_\sigma (\Delta g + g_1) \; S(\psi) \; d\sigma \tag{8}$$

where g_1 is the first term of the Molodensky series, which when applied in the integral may be approximated by the classical gravimetric terrain correction, cf. Heiskanen and Moritz (1967, sec. 8) or Moritz (1980).

The relationship between N and ζ is quite simple. For a point P at the surface, an ellipsoidal height from GPS fullfills

$$h_{ellipsoidal} = N + H = \zeta + H^* \tag{9}$$

where H is the orthometric height, and H* the normal height.

The two different height systems (orthometric "Helmert" heights and normal heights) are both

used as basic definition of heights in first-order national levelling. The basic observable in geodetic levelling is the geopotential number C (i.e., the potential difference W_0-W) at a surface point P, with practical approximative formulas for the heights

$$H_{Helmert} = \frac{C}{\bar{g}} \approx \frac{C}{g_P + 0.0424[mgal/m]H}$$
$$H^* = \frac{C}{\bar{\gamma}} \approx \frac{C}{\gamma_o - 0.1543[mgal/m]H^*} \qquad (10)$$

Examples of countries which use orthometric heights are USA, Canada, and Norway, while normal heights are used e.g. in Russia, Germany, Sweden and Finland. From (10) a formula for the relationship between N and ζ may be found

$$\zeta - N = H_P - H_P^* \approx - \frac{g_P - \gamma_o + 0.1967[mgal/m]H}{\gamma_o} H = - \frac{\Delta g_B}{\gamma_o} H \qquad (11)$$

where Δg_B is the Bouguer gravity anomaly. In Norway, e.g., with $\Delta g_B \approx 100$ mgal and H = 1000 m, this yields a difference of 10 cm, a quantity which must now be considered significant in practice. It follows from (11) that the quasigeoid locally mirrors the topography, and thus is less smooth than the classical geoid.

2.2. Geoid determination in practice

For the practical determination of the geoid, the solution is typically split into three parts

$$\zeta = \zeta_1 + \zeta_2 + \zeta_3 \qquad (12)$$

where the first part is the global field, coming from a spherical harmonic expansion of the geopotential, the second part is the contribution from the local topography, and the third part the "residual" contribution from local gravity.

The global spherical harmonic solution is found from a set of spherical harmonic coefficients C_{nm} and S_{nm} by

$$\zeta_1 = \frac{GM}{R\gamma} \sum_{n=0}^{N} \left(\frac{R}{r}\right)^n \sum_{m=0}^{n} (C_{nm} cosm\lambda + S_{nm} sinm\lambda) P_{nm}(\sin\phi) \qquad (13)$$

Current models, complete to degree 360, include the OSU91A model and the new DMA/NASA model to be released soon (Kenyon et al., pers.comm., 1996; for a review see Rapp, 1994).

The global models are based on a combination of satellite tracking and surface gravity data (0.5° mean values). In the oceans the surface gravity data are generated from satellite altimetry. The new DMA/NASA model represent a significant improvement over earlier

models due to the release of much new gravity data from formerly unclassified sources (Fig. 3), as well as improved satellite tracking data. Currently the global models gives the geoid to within ±30-50 cm in regions of good data coverage, but in regions with insufficient data and many mountainous regions errors might be much worse (up to several m).

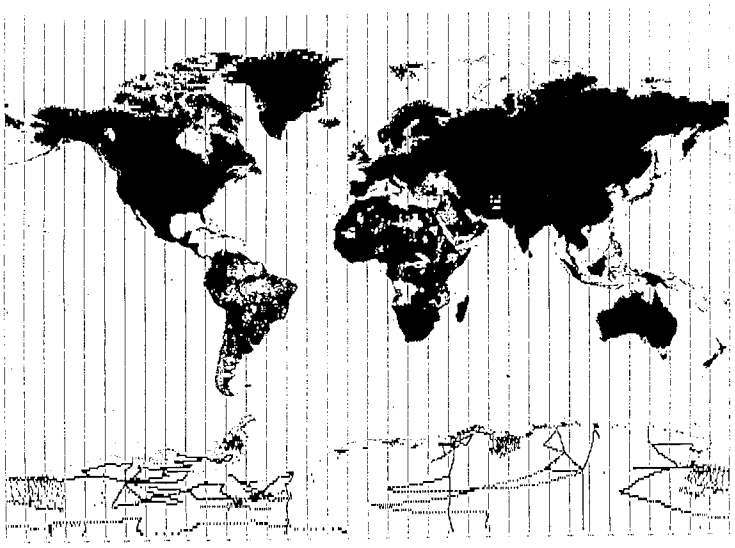


Fig. 3. Location of surface gravity data in land areas now available for the joint DMA/NASA geopotential model project. The oceans are covered by satellite altimetry. From Rapp (1994).

The influence of the topography (ζ_2) may either be computed directly by a remove-restore technique (remove topographic effects from gravity data, and restore effects on predicted geoid), or implicitly utilized in the gridding process of randomly located gravity data (e.g., the use of Bouguer anomalies and average terrain heights to infer average free-air anomalies).

The direct computation of terrain effects require the evaluation of volume integrals over the topographic masses, e.g. for gravity residual terrain model (RTM) effects of form

$$\Delta g_{RTM} = G\rho \int_{-\infty}^{\infty}\int\int_{z=h_{ref}(x,y)}^{z=h(x,y)} \frac{z-h_P}{[(x_Q-x_P)^2+(y_Q-y_P)^2+(z_Q-h_P)^2]^{3/2}}\, dx_Q dy_Q dz_Q \qquad (14)$$

The evaluation of such integrals may be done by subdividing the topographic masses into prisms, or using techniques based on Fast Fourier Transformation (FFT), for details see e.g. Forsberg (1984) and Schwarz et al. (1990). In mountainous regions the terrain effects on the geoid may be large - for Norway e.g. on the order of 0.5-1 m, this giving a signal comparable to the influence of local gravity (ζ_3).

The residual gravity ζ_3-term may be evaluated by various methods. Stokes' or Molodensky's

integrals may either be directly implemented on computers (currently used, e.g., for national geoids of Russia, Australia and France), or the Stokes' integration may be implicitly performed using FFT methods (used, e.g., in the current geoids of USA, Canada, Europe and the Nordic area).

The FFT methods use in principle the simple relationship between gravity and geoid in the wavenumber domain, or uses the convolution theorem for rapid integrations. In planar approximation residual gravity Δg_3 may be transformed into geoid by

$$\zeta_3 = \frac{1}{\gamma} \mathcal{F}^{-1}\left[\frac{1}{\sqrt{(k_x^2+k_y^2)}-2/R} \mathcal{F}(\Delta g_3)\right] \qquad (15)$$

where $\mathcal{F}$ is the two-dimensional Fourier transform and k the wavenumber, for details see, e.g., Schwarz et al. (1990). The Fourier transform methods may be formulated as working on (ϕ,λ)-grids on a sphere, as originally pointed out by Strang van Hees (1992), and later refined into virtually exact Stokes' formula evaluations by the multi-band FFT method of Forsberg and Sideris (1993) and the 1-D FFT method of Haagmans (1994).

The advantage of the FFT methods is that they give the results directly on grid form, and that very large grids may be transformed in a very short time (Nordic geoid gravity grids of e.g. 1000 x 1000 point transform in just a few minutes on a workstation). Therefore the majority of the time spent today on geoid computations is used on data preparation, evaluation and gridding, rather than the actual geoid computation.

Major regional geoids computed by FFT methods include the NKG89 geoid of the Nordic area (Forsberg, 1990), the EGG95 of Europe (Denker, 1996), GEOID95 of the US (Milbert, 1996) and many other models, cf. the recently published report "Geoids of the World", a special issue of Bulletin d'Information, International Geoid Service, Milano, 1996. The grid sizes for such models are typically in the range of 3-5 km, and grid dimensions may be several 1000's. The models, however, do depend strongly on the data entering the solutions, and residual geoid errors may be large, reflecting a.o minute systematic errors in the terrestrial gravity data and terrain corrections. Most models are computed as either geoid or quasigeoid models, with a few models providing the option for predicting both quantities.

Apart from the Stokes' and FFT methods, a completely different class of methods may be used - the operational methods of least-squares collocation and point mass modelling (the latter may be viewed as a special case of collocation). Least-squares collocation is a very powerful tool, capable of utilizing all kinds of gravity field data in a near-optimal way, for a review see Tscherning (1994). Therefore the method is especially suitable when data of different types (e.g. satellite altimetry, gravity and/or deflections of the vertical) need to be utilized together. However, due to the large computational burden it is rarely used for large-scale regional geoid solutions.

The practical computations of geoid grids may be carried out by different software packages, with most investigators developing own software, with some key software modules (e.g., for spherical harmonic models) often being shared. An example of a comprehensive software package, covering all aspects of geoid determination from the computation of spherical

harmonic models, computation of terrain effects and the handling of residual gravity or satellite altimetry by collocation, FFT or Stokes' integration methods is the GRAVSOFT package of KMS and the University of Copenhagen (Tscherning et al., 1992). Common to most practical computations is that the end result is a geoid grid, which typically is delivered together with a simple interpolation routine to users.

2.3. The nature of gravimetric geoid errors

A gravimetric geoid of a large region will have errors due to the inherent differences in datum systems, and geoid errors due to errors in the gravity or terrain data used for the geoid prediction. Due to the long-wavelength nature of the geoid prediction process, even minute errors in e.g. the reference level of used gravity data may have a large impact on the geoid prediction. As a rule of thumb, the following formula may be used to estimate geoid errors due to errors in gravity

$$\delta N = \frac{\delta g}{\gamma} s \tag{16}$$

This formula gives the maximal geoid effect of a constant gravity error δg over a circular zone with radius s. For s = 100 km and δg = 1 mgal this gives a geoid error of 10 cm. Errors at the 1 mgal level are quite likely for many older gravity surveys, especially when data have not been transformed properly from the older Potsdam datum to an absolute gravity datum, or tied to inaccurate or erronous gravity base stations.

Other potentially serious sources of gravity data biases are the use of digital terrain data. These errors can be very difficult to quantify. One type of errors relate to the use of the classical terrain correction, which give a one-sided effect (it is always positive). By using insufficient height data the correction may be underestimated, yielding a bias which can easily be at the 1-mgal level or more in rugged topography.

A more serious avenue for errors is the use of DTM mean heghts to estimate mean free-air (or, rather, Faye anomalies) from gridded Bouguer anomalies by

$$\overline{\Delta g_{FA}} = \overline{\Delta g_{BA}} + 2\pi G\rho\overline{h_{DTM}} \tag{17}$$

This is common practice, and is implicitly also done for rigoruous use of terrain reductions such as the RTM method. In this case any systematic error in the DTM heights will propagate directly to the free-air anomalies, used for the geoid prediction. An error of 3 m in height corresponds to 1 mgal in this case. Therefore DTM's need to be carefully evaluated when used for geoid determination, especially in lowlands where the potential geoid accuracy is the highest.

At the longest wavelengths (100's to 1000's of km) long-wavelength gravity errors can be diminshed by the use of modified Stokes' kernels in (6) or (8), essentially "overriding" the influence of local gravity data with the influence of the spherical harmonic reference models (used, e.g., in the European geoid determination, cf. Denker, 1996). The trade-off between the

impact of local surface gravity data versus the spherical harmonic model is not a straightforward task, as the error spectrum of the spherical harmonic model is usually not known with sufficient accuracy, and the errors can be expected to be non-stationary. The correct weighting of "local" versus reference models and long-wavelength satellite information is therefore an area where long GPS/levelling lines can be most useful.

Locally lack of sufficient gravity data coverage will cause more random geoid errors. A hand rule for these errors can be obtained from Kaula's rule, which is a classical "handrule" for the average "global" variation of the gravity field spherical harmonic degree-variances:

$$\sigma_n = \sum_{m=0}^{n} (C_{mn}^2 + S_{mn}^2) \approx 10^{-10} \frac{2n+1}{n^4} \tag{18}$$

In terms of geoid variance, Kaulas rule will give the following estimate for the wavenumbers above a certain harmonic degree (n > N)

$$\sigma_N^2 \approx (\frac{64}{N})^2 [cm^2] \tag{19}$$

The average r.m.s. variation below 55 km half-wavelength (N = 360) is thus 18 cm, below 10 km only 3 cm. This shows that geoid variations locally can be expected to be quite small.

A more realistic error simulation by collocation is shown in Table 1 (from Forsberg, 1993). In this example error propagation is carried out for a gravity field model with statistical parameters similar to the "average" of Scandinavia (with terrain effects removed), and the results are expressed as both "absolute" geoid errors σ_N (which are large due to errors assumed for the spherical harmonic reference model), an errors in geoid *differences* $\sigma_{\Delta N}$ for various (GPS) baseline lenghts.

From the table and Kaulas' rule it can be inferred that gravity data spaced around 10 km corresponds to a 3 cm geoid, while a 1 cm geoid should have gravity data density somewhere around 2-3 km in the best cases for short baselines. To obtain the same geoid acccuracy for longer baselines it is essential to constrain the geoid to GPS/levelling, as discussed in the next section (in the Scandinavian error simulation example GPS geoid control spaced 50-100 km seemed to be sufficient, cf. ibid., 1993).

Table 1. R.m.s. geoid errors $\sigma(N)$ and geoid difference erors $\sigma(\Delta N)$ as a function of baseline lengths, when predicting geoid heights from regularly spaced gravity in a square of sidelength "block size".

Block size	Grid spacing	$\sigma(N)$ (m)	$\sigma(\Delta N)$ (meter) 0.1°	0.25°	0.5°	1.0°	$\sigma(\Delta N)$ (ppm) 0.1°	0.25°	0.5°	1.0°
0.5°	0.1°	0.342	0.049	0.172	–	–	4.5	6.2	–	–
1.0°	0.1°	0.276	0.033	0.067	0.183	–	3.0	2.4	3.3	–
2.0°	0.1°	0.206	0.027	0.039	0.073	0.229	2.4	1.4	1.3	2.1
4.0°	0.1°	0.139	0.024	0.027	0.039	0.071	2.2	0.9	0.7	0.6
2.0°	0.4°	0.370	–	0.201	0.313	0.495	–	7.2	5.6	4.5
2.0°	0.2°	0.235	0.060	0.106	0.135	0.295	5.4	4.9	2.4	2.7
2.0°	0.1°	0.206	0.027	0.039	0.073	0.229	2.4	1.4	1.3	2.1
2.0°	0.05°	0.202	0.017	0.034	0.067	0.203	1.5	1.2	1.2	1.8

3. Improvement of gravimetric geoids by GPS levelling

By using geoid information from GPS-levelling, especially the long-wavelength geoid errors can be supressed, and the inherent datum differences eliminated. However, it is essential when computing GPS geoid heights by

$$N_{GPS} = h_{GPS} - H_{levelling} \tag{20}$$

that both levelling and GPS heights are as error-free as possible, otherwise these errors will creep into the "fitted" geoid. Common sources of GPS heighting errors are ionospheric biases and - especially - errors in antenna heights. Similarly errors in levelling can be systematic, are generally not well-known, and are to a large degree dependent on the levelling practices of the particular country.

The fitting of a gravimetric geoid - typically available in grid form - to a set of GPS geoid heights entails modelling the difference signal

$$\epsilon = N_{GPS} - N_{gravimetric} \tag{21}$$

and adding the modelled ε-correction to the gravimetric geoid. In this way a new geoid grid is obtained which is "tuned" to the levelling and GPS datum in question.

The simplest models of the geoid difference is as a constant bias only, or polynomials like

$$\epsilon = a_1 \;;\; \epsilon = a_1 + Na_2 + Ea_3 \;;\; \epsilon = a_1 + Na_2 + Ea_3 + NEa_4 + N^2a_5 + E^2a_6 \; etc. \tag{22}$$

where N and E are northing and easting coordinates. A special type of such regression function, which have been found to work well in practice, is the 4-parameter "Helmert" model

$$\epsilon = \cos\phi\cos\lambda a_1 + \cos\phi\sin\lambda a_2 + \sin\phi a_3 + Ra_4 \tag{23}$$

Applying this model is equivalent to applying a 7-parameter Helmert coordinate transformation, where the the unknowns a_1 to a_3 corresponds to coordinate shifts ΔX, ΔY, ΔZ, and a_4 to the scale factor (the geoid will to first order be invariant to coordinate system rotations). This kind of regression should of course not be interpreted as a rigorous coordinate transformation, since the parameters will absorp long-wavelength geoid errors as well.

Polynomial style fits like (22)-(23) have the problem that ε can obtain large unrealistic values in data voids or outside the GPS coverage. Therefore collocation is a much more suitable method for modelling the residuals, preferably in connection with a low order regression like (23), to ensure residual geoid correction quantities with zero mean.

In the collocation process a covariance function must be assumed for the residual geoid errors ε' (after fit of e.g. bias or 4-parameter model) as a function of distance s

$$C(s) = cov(\epsilon', \epsilon') \tag{24}$$

Such a covariance function will be characterized by zero variance C_0 and correlation length $s_{1/2}$ (distance where covariance function attains half its top value), which in turn determine the degree of fit and the smoothness of the interpolated geoid error. A quite simple covariance model will usually be sufficient. In the GEOGRID collocation program of the GRAVSOFT software a second order Markov model (which models Kaula's rule quite well) is used

$$C(s) = C_0 \, (1+\alpha s) \, e^{-\alpha s} \tag{25}$$

where the constant α is the only quantity to be specified by the user, with C_0 automatically adapted to the data. Other investigators have used other similar simple models, for an example see Fig. 4.

The geoid error correction is obtained by the familiar collocation predictor

$$\epsilon_i' = C_{ix} \, [C_{xx} + C_{nn}]^{-1} \, x \tag{26}$$

where $x = \{\varepsilon_1 \, .. \varepsilon_n\}$ is the vector of observed geoid errors with apriori variance matrix n. In the selection of correlation length and noise of observed errors the user has a quite large degree of freedom to select either a strong fit to the GPS data, or a more relaxed fit, diminishing the impact of possible errors in the GPS levelling data. In the end the choice of function will in practice often be quite intuitive, as only rather few GPS levelling geoid control data are typically available.

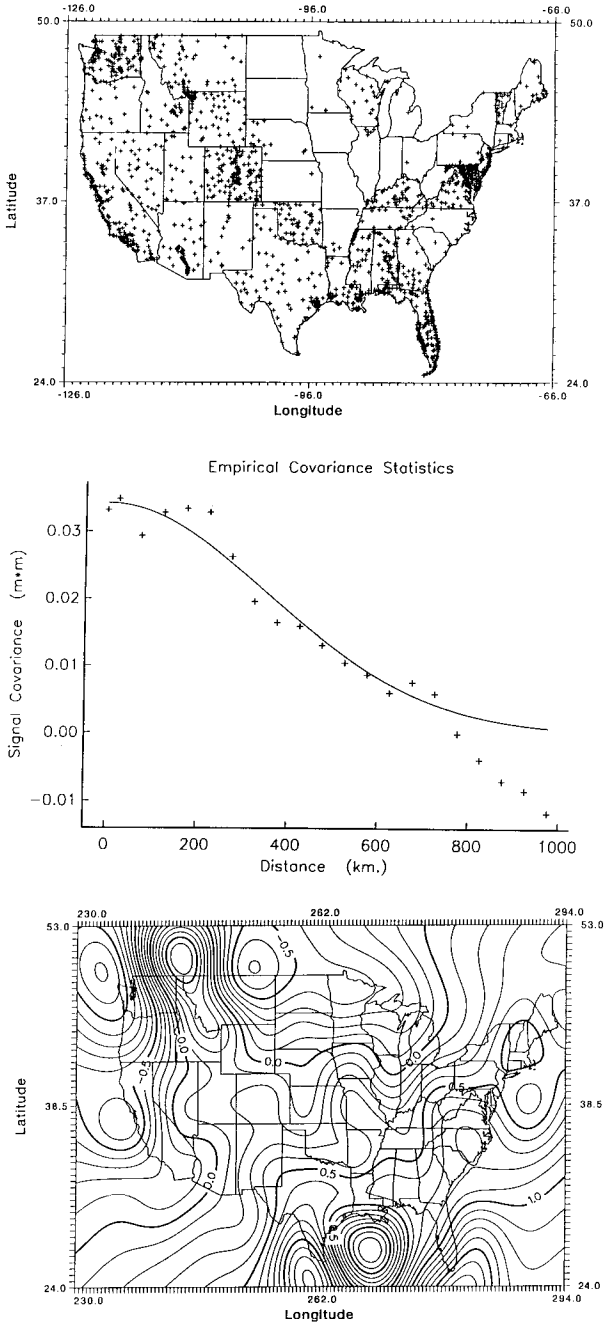


Fig. 4. Example of geoid errors in the US FFT geoid G9501 when compared to GPS levelling. Top: Location of GPS/levelling points, center: Empirical covariance function of errors (ε), bottom: plot of the modelled ε-corrector after fitting a gaussian covariance function. From Milbert (1996).

4. The Nordic area geoid

The geoid determination project in the Nordic region (recently expanded with the Baltic countries) have been a continuing process for over a decade by now. The work has been carried out within the framework of a Working Group under the Nordic Geodetic Commission (NKG). A first blocked collocation solution was published by Tscherning and Forsberg (1986), followed by a series of later FFT models, including the currently officially adopted model NKG-89 (Forsberg, 1990).

The by far most time consuming work relating to the gravimetric geoid determination has been the collection and validation of available gravity data, and inclusion of data into the joint data base at KMS (programmed by and operated in close cooperation with Dag Solheim, Statens Kartverk, Norway). Currently the data base holds more than 900.000 gravity observations covering Scandinavia and surrounding regions from Greenland to the Urals. In recent years significant new data have been entered into the data base especially from Norway and Sweden (geological survey data and offshore data), from the Baltic republics (point data from digitized maps and station sheets) as well as data from East Germany, Poland and Russia. The data coverage is shown in Fig. 5.

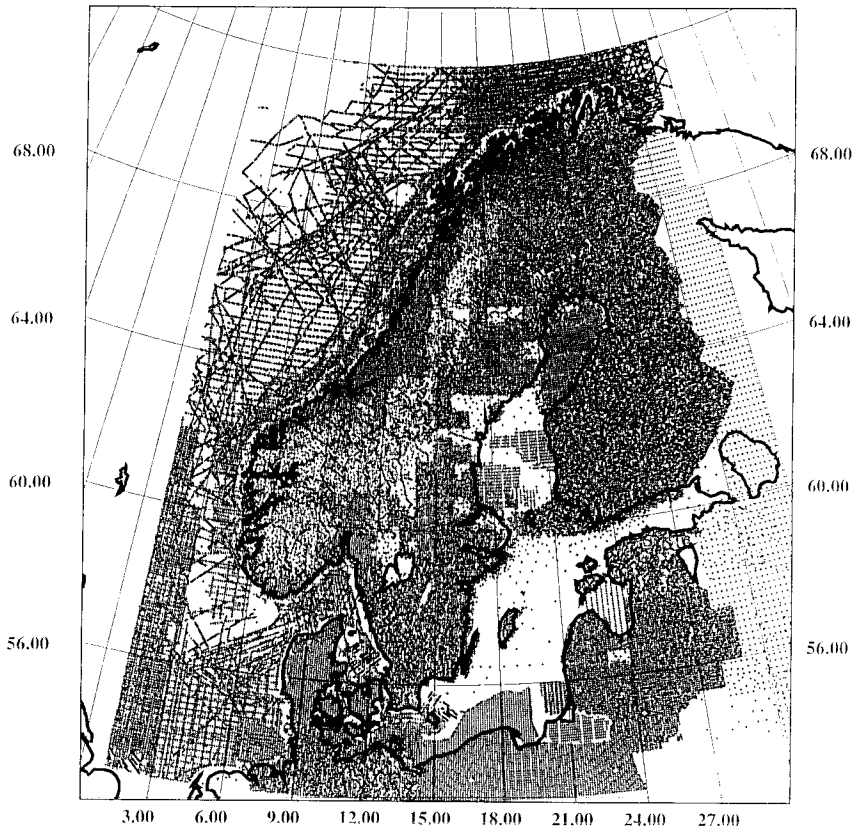


Fig. 5. Gravity data coverage of Nordic and Baltic area, status August 1996.

Detailed digital terrain models now also covers the complete region at various resolutions. For the project the following DTM data were available:

Norway: 100 x 100 m
Denmark: 200 x 200 m
Sweden: 500 x 500 m
Finland and Baltic area: 1 x 1 km data (GLOBE DTED averages).

The terrain data reduction currently used is a RTM remove-restore terrain reduction, with a resolution of the mean height surface of appr. 100 km, for details see Forsberg (1984).

Compared to earlier geoid solutions, much improved spherical models are now available, and early validation versions of the DMA/NASA spherical harmonic models (EGM-X01 or EGM-X05) are currently used in the ongoing experimental Nordic geoid computations. Since the final DMA/NASA is not yet released, the "final" new version of the Nordic geoid can not yet be released, either.

As an example of a current solution, the solution of April 1996 will be presented (internally named "fa.geoid"). For this solution available gravity data was extrated from the data base, and thinned to a pixel spacing of 0.05° x 0.1° grid, yielding a total of 75325 points selected in the area 53°-72°N, 3°-33°E. The data reductions include reduction for the reference field (EGM-X01), using a "sandwich" interpolation routine to taking into account the spatial nature of the reference model (including the effect of the atmosphere), and terrain reduction by the RTM method. The statistics of the data reduction are shown in Table 2.

Table 2. Statistics of gravity data reductions in Nordic area (unit: mgal)

	mean	std.dev.	min	max
Original data	-1.2	25.6	-136.7	193.2
Δg - ref.field	-0.8	16.5	-151.7	122.9
Δg - ref - RTM	0.6	9.8	-60.4	77.3

As can be seen from Table 2, the use of the spherical harmonic reference model and the RTM terrain reductions smooth the data significantly. However, the mean value of the gravity data is not zero, and this may affect the FFT solution, which has difficulty properly handling such biases (as has most other methods). Indeed, quite large discrepancies occur in the solution depending on whether this bias is eliminated or not.

To augment the sparse gravity coverage in the southern Baltic Sea, gravity data was predicted from geodetic mission ERS-1 satellite altimetry. The ERS-1 altimetry was reduced for sea-slope effects using the sea surface topography model of Ekman and Mäkinen (1994), and the altimetry data subsequently edited, cross-over adjusted and gridded to yield an estimate of the geoid surface, which was subsequently inverted to gravity using Wiener filtering inverse Stokes' FFT (Forsberg and Solheim, 1988), the same technique used for the global gravity map of Andersen and Knudsen (1995). The outcome of this process is a gravity grid, which

has subsequently been "draped" upon the existing gravity data using collocation techniques, producing a smooth, merged gravity data set across the southern Baltic. Fig. 6 shows the reduced free-air anomalies of the Baltic from ERS-1 prior to draping.

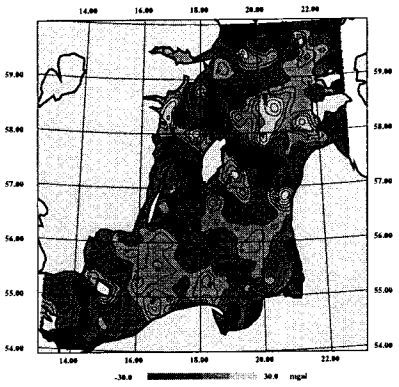


Fig. 6. Reduced Baltic free-air anomalies from ERS-1, 5 mgal contour interval.

From the merged surface gravity data and ERS-1 gravity data, a grid was predicted using rapid collocation (GEOGRID). On this grid terrain effects were restored prior to FFT using (17), yielding essentially a grid of reduced Faye-anomalies (terrain-corrected free-air anomalies), which was then converted into height anomalies by multi-band spherical FFT (SPFOUR, using 5 bands and 50% zero padding). The reference spherical harmonic reference effects were finally restored using a (spatial) grid interpolation routine, referring to the surface of the topography. Fig. 7 and Fig. 8 shows the contributions to the geoid from the FFT step (i.e., reflecting the impact of the local data), and the final geoid.

The final geoid model is by the above procedure a *quasigeoid*, and to obtain a classical geoid the correction (11) must be added. This correction can be up to 20 cm in the mountains of Norway. This correction is available in a special grid file.

To evaluate the gravimetric geoid, several GPS nets observed mostly on levelling benchmarks are available. The GPS data below have all been transformed into (approximately) ITRF93 GPS datum and the UELN European levelling datum (East European levelling have been corrected by 15 cm, cf. Ihde, 1993). Fig. 9 shows the location of the GPS points, and Table 3 lists the fits of the GPS data to the geoid for different GPS/levelling surveys both directly and after detrending with the 4-parameter model (23).

It appears from Table 3 that a surprising good geoid fit is obtained over many areas, and that except for the expected datum bias (which is around 64 cm) the "fa.geoid" fits to within 10 cm, with a much better fit in some areas such as Denmark (and southern Finland) where the gravity coverage is very dense.

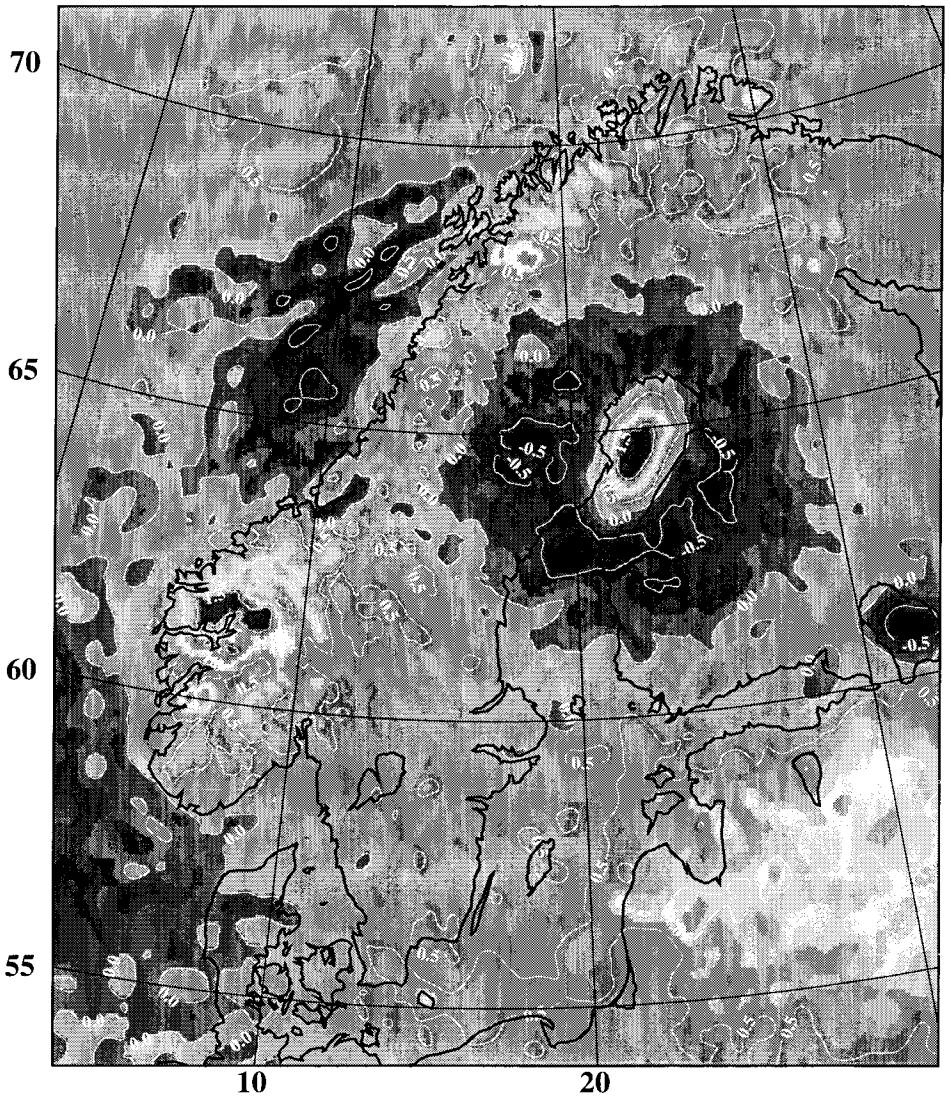


Fig. 7. FFT contribution to the geoid in the most recent solution. An apparent large contribution from the northern Bothnian Bay appears to be a problem in the used EGM-X05 reference field, which has a large deviation from satellite altimetry in this region. This problem should be overridden by the local gravity data in the FFT geoid. Contour interval 0.5 m.

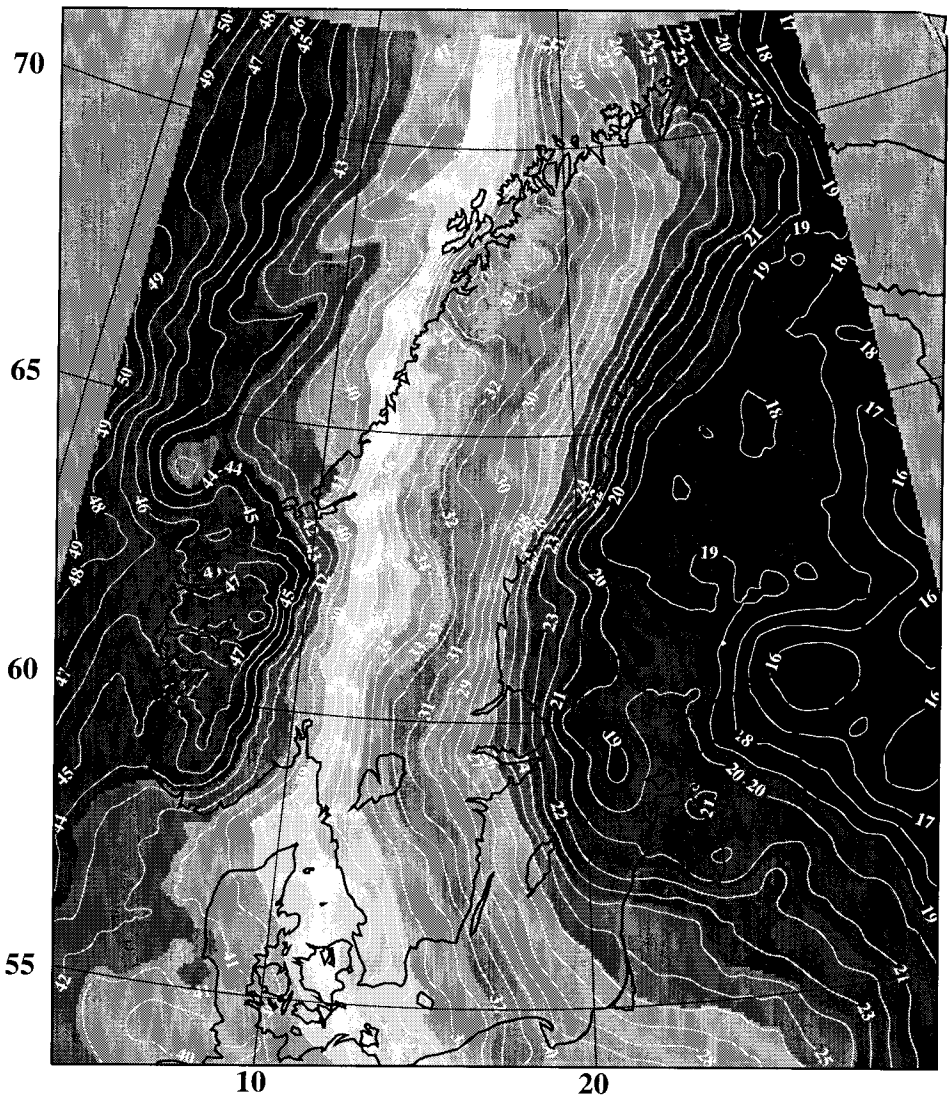


Fig. 8. Best current Nordic geoid model ("fa.geoid"). Contour interval 1 m. The geoid is available to users in grid form with associated interpolation program "geoid". Final distribution awaits updates with improved handling of the long-wavelength biases and the final release of the DMA/NASA reference model.

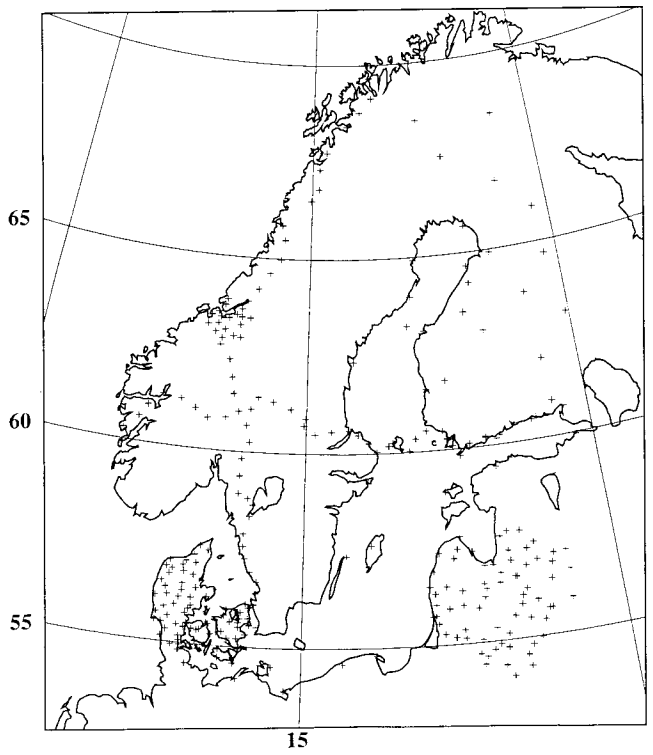


Fig. 9. N_{GPS} points in Nordic area

Table 3. Fit of various GPS/levelling data sets in ITRF93/UELN datum to geoid (unit: m).

GPS data set / *source*	Mean diff.	Std.dev.	Std.dev. after 4-par fit
IfE profile (Germany-Tromsø, 46 pts., *from Torge*)	-0.632	0.079	0.055
SWET profile (Bergen-Russian border, 34 pts., *from Poutanen, FGI*)	-0.634	0.118	0.045
Finland (42 points, *from Poutanen*)	-0.638	0.142	0.085
Baltic Sea Level Project (30 points, *Poutanen*)	-0.558	0.102	0.074
Lithuania (36 points, *from Parseliunas)*	-0.604	0.087	0.085
Latvia (26 points, *from Kaminskis)*	-0.676	0.136	0.094
Denmark-REFDK (67 points, *from Madsen)*	-0.689	0.024	0.019
Trondheim area (20 points, *from Simensen)*	-0.691	0.076	0.069
All data (301 points)	-0.643	0.104	0.102

5. Example of geoid fitting - Denmark

To give an example of local fitting of the geoid to GPS, the REFDK net of Denmark will be used as an example. The 4-parameter model was used to fit the REFDK N_{GPS}-results to a detailed ("model95") and less detailed ("fa.geoid") Nordic geoid model. The REFDK GPS net is a very accurate GPS net, with long occupation times (up to a week per point), processed by Bernese Software. Great care has been taken to ensure antenna heights are reliable (B. Madsen, pers.comm.). The GPS points are mainly on first-order levelling benchmarks, except for a few islands.

The geoid heights were fitted with collocation, assuming an error correlation length of 50 km, and a GPS/levelling data noise of 1 cm. The 4-parameter model was removed prior to collocation. Fig. 10 shows the total geoid difference signal (ε), and Table 4 outlines the results of the fitting.

As seen from the table, a fit of around 1 cm is obtained, and the maximum error amounts to 2.4 cm. It appears that 5 km resolution is not sufficiently dense, as the 2.5 km grid seems to pick up more local details in the geoid. The error numbers of the fitted geoids have been supported by independent GPS resurveys of levelling points in Jutland, which have shown maximum errors around 15 mm (K. Madsen, pers.comm.). The cm-geoid for the danish area is therefore allready now very close to being a reality.

Table 4. GPS/levelling geoid fits to the REFDK network of Denmark. Units: m.

Geoid model	Mean fit	Std.dev.	Std.dev. after 4-parameter fit	Std.dev. after 4-par fit and collocation
"fa.geoid" (5 km resolution)	-0.563	0.023	0.019	0.009
"model95" (2.5 km resolution)	-0.614	0.047	0.022	0.007

6. Conclusions

The advances in geoid determination techniques and reference field modelling now routinely allow large-scale geoids to be determined with accuracies which range from some 0.3-0.6 m in a global sense, with best local errors of 10 cm over 1000 km-baselines, and down to 1 cm on short baselines (10's of km). These numbers refer to regions of reasonably good gravity coverage.

The understanding of the nature of the geoid errors, and the awareness of the different geoid definitions, is essential for a good utilization of GPS for height determination.

When GPS/levelling networks of high quality exists together with dense gravity (less than 5 km spacing), it appears feasible to constrain a national geoid to the 1-2 cm level. GPS might thus replace most of the lower order levelling in that case. If a country only has e.g. a 10 km average gravity grid spacing geoid elevations might only be determinable to around 3-5 cm in the best cases.

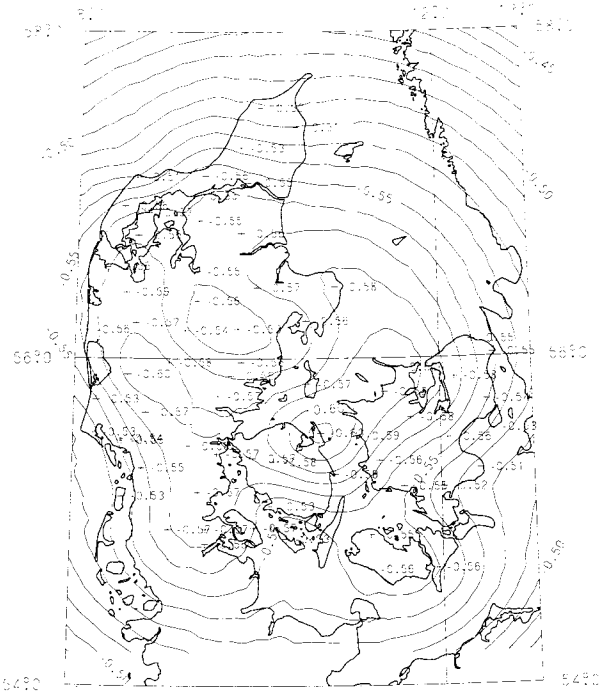


Fig.8. Modelled geoid errors from the combination of a gravimetric geoid and GPS/levelling in Denmark. 1 cm contour interval.

For mountainous regions less accurate results are typically obtainable, and going below a 10 cm geoid will probably be very difficult, and require very careful evaluation of all systematic errors relating available gravity and terrain data, as well as a very careful review of methods.

Acknowledgements

The Nordic geoid project is a cooperation between many individuals and agencies from the Nordic countries. J. Kamiskis, Latvia, D. Solheim, Norway and several others (M. Ekman and L.Haller, Sweden; M. Vermeer and M. Poutanen, Finland; E. Parseliunas, Lithuania) have contributed significantly to the present investigations. The various national geodetic agencies of the Nordic countries and neighboring countries are thanked for putting their data at the disposal of the present project.

References

Andersen, O. B., P. Knudsen: Global gravity field from the ERS-1 geodetic mission. Earth observation quaterly, 47, ESRIN, 1995.

Denker, H.: The European quasigeoid EGG95. In: New Geoids of the World, IGeS Bulletin 4, Milano, 1996.

Ekman, M., J. Mäkinen: Mean sea surface topography in a unified height system for the Baltic Sea area. Proc. 12th General meeting of the NKG, Ullensvang, Norway, pp. 244-258, 1994.

Forsberg, R.: A Study of Terrain Reductions, Density Anomalies and Geophysical Inversion Methods

in Gravity Field Modelling. Reports of the Department of Geodetic Science and Surveying, No. 355, The Ohio State University, Columbus, Ohio, 1984.

Forsberg, R. and D. Solheim: Performance of FFT methods in local gravity field modelling. Proc. Chapman Conf. on Earth's gravity field, Ft. Lauderdale, Florida, pp. 100-103, 1988.

Forsberg, R.: A new high-resolution geoid of the Nordic area. In: Determination of the geoid. IAG symposium 106, Milano, Proc. published by Springer Verlag, 1990.

Forsberg, R.: Modelling the fine-structure of the geoid: Methods, data requirements and some results. Surveys in Geophysics, vol. 14, pp. 403-418, 1993.

Forsberg, R., M. G. Sideris: Geoid computations by the multi-band spherical FFT approach. Manuscripta Geodaetica, 18, pp. 82-90, 1993.

Forsberg, R.: Terrain effects in geoid computations. In: Lecture notes, International School for the determination and use of the geoid, International Geoid Service, Milano, 1994.

Haagmans, R., E. de Min, M. Van Gelderen: Fast evaluation of convolution integrals on the sphere using 1D FFT, and a comparison with existing methods for Stokes' integral. Manuscripta Geodaetica, vol. 18, pp. 227-241, 1993.

Heiskanen, W. A., H. Moritz: Physical Geodesy. Freeman, San Francisco, 1967.

Ihde, J.: Some remarks on geodetic reference systems in Eastern Europe in preparation for a uniform European geoid. Bulletin Geodesique, vol. 67, pp. 81-85, 1993.

Milbert, D.G.: Improvement of a high resolution geoid model in the United States by GPS height on NAVD88 benchmarks. In: New Geoids of the World, International Geoid Service Bulletin 4, 1996.

Moritz, H.: Advanced physical geodesy. Herbert Wichmann Verlag, Karlsruhe, 1980.

Rapp, R. H.: The use of potential coefficient models in computing geoid undulations. In: Lecture notes, International School for the determination and use of the geoid, International Geoid Service, Milano, 1994.

Schwarz, K.-P., M. G. Sideris, R. Forsberg: Use of FFT methods in Physical Geodesy. Geophysical Journal International, vol. 100, pp. 485-514, 1990.

Strang van Hees, G.: Stokes formula using fast Fourier techniques. Manuscripta Geodaetica, vol. 15, pp. 235-239, 1990.

Tscherning, C. C., R. Forsberg and P. Knudsen: The GRAVSOFT package for geoid determination. Proc. 1st continental workshop on the geoid in Europe, Prague 1992, pp. 327-334.

Tscherning, C. C., R. Forsberg: Geoid determination in the Nordic countries from gravity and height data. Boll. di Geodesia e Sc. Aff., vol. XLVI, pp. 21-43, 1986.

Tscherning, C.C.: Geoid modelling by least-squares collocation using GRAVSOFT. In: Lecture notes, International School for the determination and use of the geoid, International Geoid Service, Milano, 1994.

Processing of data from the space geodesy systems VLBI, GPS, SLR, PRARE and DORIS *

Per Helge Andersen

Norwegian Defence Research Establishment
P.O. Box 25, N-2007 Kjeller, Norway
Email: Per-Helge.Andersen@ffi.no

Abstract.

This paper gives an overview of the principles of the high-precision space geodetic observation methods VLBI (Very Long Baseline Interferometry), GPS (Global Positioning System), SLR (Satellite Laser Ranging), PRARE (Precise Range And Range-rate Equipment), and DORIS (Doppler Orbitography and Radio positioning Integrated by Satellite). Results from analyses of such data using the GEOSAT software are presented. The GEOSAT software for geodetic and geodynamic applications of tracking data from satellites and VLBI data from distant radio sources has been under development at the Norwegian Defence Research Establishment (NDRE) since 1982. Among the space geodesy softwares currently on the market, GEOSAT is unique in that one and the same program can, in addition to satellite tracking data, also handle VLBI data. The justification for building the satellite and VLBI applications into the same program, is that both types of application share a common terrestrial reference frame with all the complex motions that geodynamics imposes on it.

With the GEOSAT software it is possible to combine VLBI and any type of satellite tracking data in a simultaneous solution for maximum information extraction from the data. The paper presents results from analyses in a simultaneous combined mode with VLBI, GPS, and SLR data.

1 Introduction

Dynamics of the solid Earth covers a broad spectrum of frequencies from plate tectonic motion over hundreds of millions of years to glacial rebound over periods of tens of thousand of years, to daily variations in the Earth's rotation. Tectonic motion is of particular interest since it is the underlying cause of much of the volcanic and seismic activity which may affect mankind in a serious manner.

Modern space geodetic methods including VLBI, SLR and GPS have been used recently to determine accurate coordinates and velocities of the observing stations. The determined relative motion between the stations can be compared with those predicted by geologic models. However, it should be noted that the motion determined by space-geodetic methods are average values over some years, while most geologic models give average values over several million years. Any disagreement between the space-based results and the geologic models does therefore not necessarily imply an error in either of the models.

Station coordinates and velocities can, with the current space-based techniques be determined with an accuracy of 1-2 cm and a few mm/year. Such results rely on the ability to accurately determine the orbits of the satellites. Two main factors play an important role in the orbit determination process:

*Lecture given at the course on "Geodetic Applications of GPS" for young scientists in the Nordic countries arranged by The Nordic Geodetic Commision, Båstad, 26 - 31 August 1996.

The accuracy and distribution in space and time of the tracking data and the dynamical model for the satellite motion. The latter is usually the limiting factor since errors still exist in the models for calculating the instantaneous geopotential and the satellite surface forces. Also the measurement models contain significant and sometimes even dominating errors.

We have already mentioned the VLBI, SLR and GPS techniques. The Russian multi-satellite system GLONASS is expected to play a significant role in future tectonic studies. Since this system with respect to data analysis, is quite similar to the analysis of data from GPS, we will not consider this technique here. Lunar laser ranging (LLR) is equivalent to the SLR technique except that the Moon is used as the target instead of an artificial satellite in orbit around the Earth. Due to the distance between the Earth and the Moon, specific requirements must be satified for a laser station in order to track the Moon successfully. Since only a few existing lasers have this capability, we consider the LLR technique to be of only secondary importance with respect to applications involving precise positioning. The method will therefore not be treated here. Two other techniques that we will consider are the german PRARE system and the french DORIS system. These systems are expected to play an important role in precise orbit determination and station positioning in the future.

2 Space geodetic observation techniques

Three measurement principles involving radio signals in the microwave spectrum (e.q., 1-20 GHz) exist: Systems involving natural radio sources (VLBI), one-way satellite-based systems (GPS, GLONASS and DORIS) and two-way satellite-based systems (PRARE). In addition, we have satellite (SLR) or lunar (LLR) laser ranging operating in the optical spectrum.

The microwave systems have all-weather capabilities, while SLR and LLR are sensitive to the weather conditions. The tropospheric correction for the microwave techniques contains a wet and a dry part. The SLR measurement must only be corrected for the dry effect. The wet correction is only 10 % of the dry correction, but it is very difficult to model it accurately enough and it is usually the limiting factor in high-precision applications.

Another correction for microwave techniques is due to signal propagation through the ionosphere. Most of this effect can usually be eliminated using two-frequency measurements. The optical frequencies used in SLR and LLR are too high for the ionosphere to influence the signal propagation.

For the one-way systems VLBI, GPS, GLONASS and DORIS, the clocks involved will influence the measurements in a significant manner. In order to be able to isolate the geometric information inherent in the measurements, the clock contributions must either be reduced or removed by differencing or filtering techniques, or modelled accurately. The latter strategy is only realistic if hydrogen-maser clocks are applied since the clock time scale must be modelled within typically 10 picoseconds of the actual clock motion. For a two-way technique like PRARE, the clock requirements are much less stringent. A high quality quartz oscillator will suffice, since the clock only needs to measure the time of signal flight accurately. For a geodetic satellite this implies that the oscillator must have a stability of better than 10^{-10} s/s within 0.1 seconds which is easily satified by a quartz oscillator.

2.1 VLBI

The primary Very Long Baseline Interferometry (VLBI) geodetic observable is the measured time interval between the arrival of a signal from a radio source at one end of the baseline and its arrival at the other end. This interval is called the delay and its time derivative is the delay rate observable.

The basic principle is that each observing station receives and records on magnetic tape the signals from a radio source and the time of arrival of the signals in terms of the local time maintained by a very precise hydrogen-maser clock. After the observing session which typically lasts 24 hours, the tapes are sent to a correlator facility. In the correlator the delay can be found by cross-correlating the signals from pairs of stations and determining the point of maximum correlation. The stations are observing

frequencies in the S- and X-bands to be able to compensate for the ionospheric disturbance. Between 20 and 30 VLBI stations with antennas usually in the range 4 to 40 meters are available today and a few of them are mobile.

Actually there are two types of VLBI delay observables: The group delay, with a precision of 10 to 50 picoseconds (3 to 15 mm), is unambiguous and derived from the tracking of the phases of the signals at many frequencies within the S- and X-bands. The phase delay which is a factor 10 to 40 more precise than the group delay is unfortunately only known within one wavelength ($\sim$ 4 cm) and can only be applied if the baseline is very short (some kilometers). As the quality of the technology improves we may expect that in the future the group delay will be sufficiently precise to recover the phase delay ambiguities directly. The phase delay-rate does not really add any new geometric information and is only applied for obtaining statistical information on the variability of the troposphere.

2.2 GPS and GLONASS

The two military satellite systems, the Global Positioning System, GPS[8], and GLONASS[7] are very similar general-purpose navigation systems consisting of constellations of 24 satellites in orbits approximately 20000 km above the surface of the Earth and with an orbital period of 12 hours.

The satellites are continuously broadcasting dual frequency carrier signals (with frequencies 1.2 and 1.6 GHz) modulated with pseudorandom noise (PN) ranging codes (containing the broadcast orbit information) to permit one-way pseudorange and carrier phase measurements. The receiver on the ground (or in another satellite) generates its own signals with the PN-codes which are compared with the received signals from the satellites. The shift in time giving maximum autocorrelation between the station generated signal and the satellite signal is determined for each frequency. This pseudorange measurement consists mainly of the geometric range between the satellite and the ground station and clock terms due to the relative drift of the oscillators in the satellite and the station. If the receiver is a code receiver, the known code is removed from the received signal to recover the carrier from the satellite signal. The measured phase is the phase difference between satellite carrier signal and the receiver generated carrier signal. If the receiver is codeless, the carrier is retrieved by squaring the signal. The pseudorange measurement precision for the best geodetic receivers and antennas is a few decimeters, and the measurement precision of the phase measurement is a few millimeters. The phase measurement can be applied in the data analysis as a range measurement if the ambiguous integer number of wavelengths can be resolved. The process of estimating this integer quantity is usually denoted as *ambiguity resolution*. It is the application of the phase measurements that makes it possible to apply the GPS and the GLONASS systems for high-precision station positioning.

One of the main advantages of the GPS and GLONASS systems is that a receiver at an altitude below 3000 km at any instant can track simultaneously four or more satellites. This simulaneity enables a quasi-geometric positioning technique. Data from more than a hundred continuously tracking worldwide distributed high-quality GPS code receivers are available.

2.3 SLR and LLR

The concept of satellite laser ranging can simply be described as sending a short pulse to be reflected from a satellite, with the round-trip time-of-flight accurately measured. The measurement accuracy is currently about 1 cm for the best stations. Almost 10 mobile and approximately 30 fixed SLR stations are available globally.

The geodetic satellites LAGEOS I & II, ETALON I & II, STARLETTE, STELLA, AJISAI, designed specifically for high-precision orbit determination, and the altimeter missions SEASAT, ERS-1 and TOPEX/POSEIDON are all equipped with laser retroreflectors. The geodetic satellites generally have a small area-to-mass ratio to reduce the effects of surface forces, such as drag and radiation pressure. The satellite bodies are usually spherical in order to simplify the modelling of the remaining surface forces

and they are, with the exception of STELLA, STARLETTE and AJISAI, placed in high altitude orbits to reduce the effects of the poorly known short-wavelength part of the geopotential.

Lunar laser ranging (LLR) is a special application of SLR where the natural satellite, the Moon, is tracked instead of an artificial Earth satellite. The laser pulse is reflected from retroreflectors placed at the surface of the Moon by the US and Russian lunar missions. Only a small number of laser stations available are capable of getting pulse returns from the Moon. Thus the network of these LLR stations can not play an important role in precise geodetic positioning.

2.4 PRARE

The Precise Range And Range-rate Equipment PRARE[9] is a spaceborne two-way, double-frequency range and range-rate measurement system flown for the first time on the ERS-1 satellite. Unfortunately, the system failed to work. An improved version of the space segment was launched successfully with the Russian Meteor 3 satellite in January 1994.

PRARE will also be flown on the ERS-2 satellite to be launched early in 1995. The PRARE ground station network for ERS-2 will consist of approximately 30 globally distributed ground stations.

The principle of operation is as follows: Two signals of frequencies 2.2 GHz (S-band) and 8.5 GHz (X-band) are sent from the satellite to the Earth. Both signals are modulated with a pseudorandom noise (PN) code for the distance measurement containing broadcast information for the ground station operation. The time delay in the reception of the two simultaneously transmitted signals is measured in the ground station for later ionospheric correction of the data. The received X-band signal in the ground station is transposed to 7.2 GHz, coherently modulated with the regenerated PN-code (containing the ionospheric correction) and retransmitted to the satellite receiver. The PN-code received by the satellite is fed into a correlator to determine on board the two-way signal delay or the two-way range between the satellite and the ground station. Furthermore, the received carrier frequency is fed into a Doppler counter to measure the velocity of the satellite relative to the ground station. Four independent correlators and four Doppler counters allow simultaneous measurements with four ground stations. Geometric position and orbit determination are therefore possible with this tracking system.

The measurement accuracy of PRARE is estimated to 3 to 7 cm for X-band ranging (one measurement per second) and 0.1 mm/s for X-band Doppler (integration interval of 30 seconds). The main error source is the tropospheric refraction which is expected to contribute 2 to 5 cm. The first tests with real data from Meteor 3 indicate that the "single-shot" ranging precision is around 2 cm. The normal point precision seems to be better than 1 cm.

2.5 DORIS

The Doppler Orbitography and Radio Positioning Integrated by Satellite, DORIS[10], is a one-way French satellite tracking system, so far successfully flown on the SPOT 2 and the Topex/Poseidon satellites. The tracking network consists of around 50 globally distributed Doppler beacons giving an almost continuous tracking of the satellites.

The principle of DORIS is the following: Each ground station transmits two signals of frequencies 0.4 and 2.0 GHz to support calibration of the ionosphere. The transmitted signals contain digitally encoded meteorological and calibration data. A Doppler counter in the space segment measures, with an integration interval of 9 seconds, the relative velocity between the satellite and the ground beacons with an accuracy of 0.3 mm/s. The DORIS receiver observes the individual ground beacons in sequence and can therefore observe only one beacon at a time.

SOFTWARE SYSTEMS FOR SPACE GEODESY							
Software	Institution	VLBI	GPS	SLR	Other	Lsq	Kf
GEOSAT	NDRE	x	x	x	P,D,E,SST	x	x
CALC/SOLVE/SOLVK	GSFC,MIT	x				x	x
K-3	CRL	x				x	
MODEST	JPL	x				x	
OCCAM	SK	x					x
BAHN/GPSOBS	ESOC		xdd	x	D	x	x
BERNESE	AIUB		xdd			x	
CGPS22	GSC		xdd			x	
DIPOP	UNB		xdd	x		x	
DOGS	DGFI		x	x	P,D	x	
EPOS.P.V3	GFZ		x			x	
GAMIT	MIT,SIO		xdd			x	
GAS	IESSG		xdd			x	
GEODYN II	GSFC		xdd	x	P,D,E,SST	x	
GEONAP	UH		x			x	
GEPARD	GFZ		x			x	
GIPSY/OASIS II	JPL		x		D,SST		x
GLOBK	MIT		xdd				x
MSODP/TEXGAP	UTCSR		xdd	x	D	x	
MSOP	NAL		xdd			x	
OMNIS	NSWC		x				x
PAGE3	NGS		xdd			x	
SHAGAP	SO		xdd			x	
TOPAS	FAF		x				x
COSMOS	NAL			x		x	
GIN-DYNAMO	GRGS			x	D	x	
KIEV-GEOD-3	GAOUA			x		x	
POTSDAM-5	GFZ			x		x	
SATAN	RGO			x		x	
SODAPOP	IESSG			x		x	
UTOPIA	UTCSR			x	D	x	
ZOOM	CNES			x	D	x	

Table 1: The table shows the major packages presently available for high-precision station positioning and which types of data can be analyzed. P: PRARE, D: DORIS (or doppler), E: Ephemeris, SST: Satellite-to-satellite ranging or doppler. xdd means technique available with doubled-differenced data, and Lsq stands for least squares and Kf for Kalman filter. The information is based on personal communication with the authors or taken from the literature.

3 Major software systems

Table 1 lists the major software packages (known to the author) applied in high-precision space geodesy today. It can be seen that GEOSAT[2] is the only software that can be applied for analysis of all the high-precision techniques VLBI, GPS, SLR, PRARE, DORIS and satellite-to-satellite tracking data.

GEOSAT, CALC/SOLVE/SOLVK (which only takes VLBI data), and BAHN/GPSOBS (which only

takes satellite tracking data) are the only systems that have both least squares and filtering capabilities. The latter technique makes it possible to represent model parameters as stochastic processes. This has proven to be valuable in connection with parameters for modelling the tropospheric and clock behaviour. Also satellite surface force scaling parameters may benefit from such a strategy.

The many software systems for analysis of GPS data may solve the clock problems with either differencing the data twice (denoted as xdd in Table 1) to eliminate their contribution to the observable, or to use undifferenced data and model the clocks as a white noise process in which a new clock correction is estimated at every measurement epoch. This is easily performed with a filtering approach. For the analysis of differenced data from a global tracking network, the selection of an optimal set of independent double differences is not at all trivial. This problem disappears with the application of undifferenced data which also will have less noise. GEOSAT uses only undifferenced data even though an earlier version used double differenced data.

MSODP is a multi-satellite version of the single-satellite software UTOPIA which, together with the GSFC software GEODYN and the newest GEODYN II version, for the last 20 years have had a leading role in SLR analysis. GEODYN is the software used for many years for generating the high quality GSFC gravity models. The GEODYN and UTOPIA programs have been a great inspiration in connection with the developement of the GEOSAT software. With the GEODYN software the measurements are converted to geometric range and range-rate already at the start of the processing by correcting for the ionosphere, troposphere, center-of-mass, clocks etc. The advantages of this are reduction in computer time and easier combination of different measurement types. In high-precision applications with especially GPS, the measurements themselves are used to estimate improved values for parameters related to measurement corrections. In GEOSAT the raw and uncorrected measurements are used and corrected for in each iteration. The advantage of this procedure is that the improved corrections easily can be applied at any stage of the iteration process.

4 The GEOSAT software

GEOSAT as currently under development at NDRE, is a state-of-the-art software system for high-precision analysis of satellite and radio source tracking data for geodetic and geodynamic applications. GEOSAT is a multi-station, multi-satellite, and multi-measurement technique system, designed to provide a flexible tool for accuracy analysis related to geodetic studies. The software can be applied either in an estimation mode, a simulation mode or in an error analysis mode.

GEOSAT is a library of subroutines written in the FORTRAN language and presently consisting of approximately 150000 statements. The code is highly portable and can be used on any workstation running under the UNIX operating system. There are virtually no limits on the maximum numbers of satellites and of tracking stations that can be included; these will be limited only by the external storage size and processing time available on the computer.

GEOSAT has a sophisticated mathematical model[2] with equations of motion and measurements formulated either in a solar barycentric or a geocentric frame of reference and corrected for relativistic effects. In this software, the most precise reference frames, dynamical models, and measurement models available are used and updated continually as better information comes along.

Sophisticated versions of Bayesian weighted least squares and Kalman filtering techniques are available, and many types of model parameters (listed in Table 2), can either be considered in an error analysis or solved-for in an analysis of real data. Most of these parameters, including dynamical parameters, can be represented by stochastic models (white or colored noise, or random walk).

GEOSAT can be applied in the analysis of any of the existing high-precision satellite tracking data types available today in addition to VLBI data. Table 3 shows the validation status of each technique and the datasets applied in the validation tests. Table 4 presents a summary of the results from the analyses. The quality of these results are comparable with that obtained by the leading international institutions within each technique.

Station coordinates and velocities
Radio source positions
Universal time (UT1)
Length of day (LOD)
Polar motion, nutation, and precession
Solid Earth tidal parameters
Ocean tidal amplitudes and phases
Ocean loading parameters
Tidal variations in Earth orientation
Tidal variations in geocenter
Satellite orbital elements
Satellite dynamic scaling parameters
Earth gravity field
General relativity parameters
Tropospheric zenith delay
Atmospheric gradients
Mapping function coefficients
Atmospheric pressure load coefficients
Range biases and clock errors

Table 2: GEOSAT solve-for and consider parameters.

GEOSAT MEASUREMENT TECHNIQUES			
Technique	Status	Satellite system	Campaigns
VLBI	Validated	Radio sources	CONT94, ERDE, R&D
GPS	Validated	GPS	CONT94, GIG91
SLR	Validated	LAGEOS1 & 2	Oct. 92 - Sept. 93
	Validated	ETALON1 & 2	Oct. 92 - Sept. 93
	Validated	ERS-1	Selected arcs
	Validated	ERS-2	Selected arcs
	Validated	METEOR-3	Selected arcs
PRARE	Validated	METEOR-3	Selected arcs of range and doppler data
	Validated	ERS-2	Selected arcs of range and doppler data
DORIS	Validated	TOPEX	CONT94
XALT	Validated	ERS-1	Selected arcs
	Validated	ERS-2	Selected arcs
Ephemeris	Validated	GPS	IGS-orbits
	Under testing	ERS-1	GEODYN II orbits
WVR	Validated		CONT94/Onsala

Table 3: Measurement techniques presently available in GEOSAT, or to be implemented in the future, validation status, and satellite systems involved in the tests. Ephemeris means that it is possible to treat orbits generated by an external software as measurements either to compare orbital models or to generate a reference orbit for later applications in the filter. SST means satellite-to-satellite tracking measurements, IFSAR stands for interferometric use of synthetic aperature radar data, and XALT means crossover radar altimetry data.

GEOSAT SOFTWARE VALIDATION		
Technique	Satellite system	Precision of results
VLBI[3]	Radio sources	Baseline length repeatability: 0.8 mm + 0.7 ppb Polar motion and nutation: 0.1 - 0.5 mas UT1: 0.01 - 0.04 ms Source coordinates: 0.1 - 0.3 mas
GPS[1]	GPS	Baseline length repeatability: 1 mm + 1.4 ppb Satellite orbit precision: 10 cm in each coordinate
SLR[4]	LAGEOS/ETALON	Station coordinates: 1 -2 cm Polar motion: 0.2 mas UT1: 0.02 ms
SLR/PRARE/XALT[5]	ERS-1/2	SLR residuals: 5 - 8 cm PRARE range residuals: 5 - 8 cm PRARE doppler residuals: $\sim$ 0.3 mm/s XALT residuals: 11 - 18 cm Radial orbit precision : 5 - 7 cm Along-track orbit precision : $\sim$ 20 cm Cross-track orbit precision : $\sim$ 10 cm
DORIS[6]	TOPEX	DORIS residuals: 1.5 mm/s Cannon-ball satellite model

Table 4: Status of the GEOSAT software validation.

5 Combination of techniques

The different techniques can be combined to get improved and perhaps more realistic estimates and error estimates of geodetic parameters. The usual but in general non-optimal approach, is to perform a separate analysis for each of the techniques and then combine the aposteriori state vectors and variance-covariance matrices from each of the solutions to obtain a multi-technique solution.

Assume, as an example, that we have data from a number of 24-hour VLBI sessions, and that some of the sites also are equipped with GPS receivers which are connected to the same hydrogen-maser clocks as the VLBI electronics. The usual way of analyzing such a combined data set is to reduce separately the GPS and the VLBI data on a session-by-session basis and finally determine a combined solution. In the analysis of both of the datatypes, clock and tropospheric parameters must be estimated from the data in order to obtain results of high quality. This standard strategy does not take into account the fact that the VLBI and GPS signals, in principle, are experiencing the same atmosphere and Earth orientation, and that the station clocks will influence the two types of measurements in an, in principle, equivalent manner.

Another related problem is that different softwares are applied in the analysis of data from the different techniques. We have already mentioned that GEOSAT is the only single software capable of reducing all the existing types of high-precision data. This means in practice that slightly inconsistent models might have been used in the analyses, implying that the results will be given in different reference systems which complicate comparisons.

On the session level, an improved approach would be to analyze the VLBI and GPS datasets for each session simultaneously with the same clock and tropospheric model parameterization and with common estimates for these parameters. The two techniques will complement each other, especially with respect to the troposphere, since GPS acquires signals simultaneously from many spatial directions and typically down to 15 to 20 degrees while VLBI often measures all the way down to an elevation of 2 to 3 degrees.

PARAMETER OBSERVABILITY							
Technique	Nut	UT1	LOD	Xp	Yp	Geoc	Pos
VLBI	xx	xx	xx	xx	xx		xx
GPS			xx	xx	xx	x	xx
GLONASS			xx	xx	xx	x	xx
SLR			x	x	x	xx	xx
LLR			x	x	x	x	x
PRARE			xx	xx	xx	xx	xx
DORIS			x	x	x	x	x

Table 5: Parameter observability for the different techniques. x (and xx) means good (or extra good) observability either in accuracy or in temporal resolution.

The addition of SLR would, in principle, help in the decoupling of the tropospheric parameters from the others, since this technique does not sense the wet part of the troposphere.

We may expect that some extra parameters must be introduced to soak up systematic differences, but this would certainly give new insight into the sources of such errors. In fact, this simultaneous strategy is the only practical way of detecting and correcting systematic technique-dependent errors.

The observability of different important geodetic parameters is technique-dependent. Table 5 shows that VLBI is the only technique capable of determining the nutations, precession, and the siderial angle (represented by UT1) and thus realize a long-term stable celestial reference frame. The satellite-based techniques cannot be used to determine UT1 since this quantity is indistinguishable from the right ascension of the ascending node. The advantage of SLR, and to a certain extent GPS, is the ability to determine the location of the geocenter. The sampling interval, redundancy and measurement simultaneity and precision of GPS make it possible to estimate the motion of the instantaneous Earth rotation axis relative to a terrestrial reference frame with high precision and resolution in time. The combination of VLBI with other techniques might in some cases make it possible to resolve the ambiguities of some of the VLBI phase delay measurements, which would give access to phase measurements with a precision of around 1 mm or better. This should certainly increase the possibility of performing high-precision EOP analysis with high resolution in time.

The simultaneous analysis of different datatypes, due consideration of the physical interrelations, and presentation of results in a common reference system, are the main ideas behind the developement of the GEOSAT software.

6 A combined VLBI, GPS and SLR data analysis

A very important practical aspect of a combined analysis is that it is possible to determine geodetic and geodynamical parameters in a common reference frame. Furthermore, the GPS satellite ephemerides will be given in a long-term stable celestial reference frame with the orientation realized by the radio sources present in the VLBI dataset. The combined processing of VLBI and GPS can reduce the network-dependency seen in some of the VLBI-results[8] especially with datasets containing misbehaving stations.

The combined analysis included VLBI (the R& D network) and GPS data (the IGS network), and SLR data for LAGEOS 1 & 2, from January 13-15 and from January 20, 1994. Data from 2 collocated VLBI/GPS/SLR stations, 4 VLBI/GPS stations, 28 GPS-alone stations and 12 SLR-alone stations were applied in the analysis.

No steps were taken in the analysis to optimize the relative data weighting between the three techniques. Typically 40000 GPS pseudorange measurements with a weight based on σ = 1 m, 40000 GPS

REPEATABILITY				
$\sigma = a + bL$	South	East	Height	Length
a(mm)	1.7	2.1	4.2	2.3
b(ppb)	0	0	0	0.18

Table 6: Coordinate repeatability of collocated stations with high-precision eccentricity vectors in a combined VLBI/GPS/SLR analysis of data from the CONT94 campaign. Two of the stations, Ft. Davis (MDO1) and Wettzell (WETT), had also some passes of SLR data.

GPS ECCENTRICITY CHECK			
Station	Comp	Diff (mm)	σ (mm)
ONSA	X	0.8 ± 1.5	3
	Y	3.6 ± 0.5	3
	Z	0.5 ± 1.4	3
WETT	X	-2.8 ± 1.3	3
	Y	-3.8 ± 1.7	3
	Z	-1.3 ± 0.6	3
KOKB	X	0.8 ± 1.5	3
	Y	0.6 ± 1.3	3
	Z	-0.5 ± 2.0	3
MDO1	X	-0.5 ± 1.9	12
	Y	2.2 ± 2.4	12
	Z	-0.6 ± 2.2	12

Table 7: The difference between VLBI to GPS eccentricity vectors estimated by GEOSAT and the vectors applied by IERS in the derivation of ITRF'93. The given error estimates are the daily repeatabilities. The last column shows the precision of the ground tie.

phase measurements with a weight based on σ = 7 mm, 6000 VLBI group delay measurements with a weight based on σ = 7 mm, and 300 SLR measurements with a weight based on σ = 5 cm, were analyzed for each day.

All parameters were estimated simultaneously in a free network mode. One common set of station coordinates were derived for each collocated station with contributing data from all techniques when available. Table 6 presents the coordinate repeatability of collocated stations with high-precision eccentricity vectors. It can be seen that the repeatability in baseline length is almost independent of the distance between the stations.

Eccentricity vectors from VLBI to GPS and/or SLR were also estimated with an *a priori* sigma in accordance with the given precision of the ground tie. The difference between VLBI to GPS eccentricity vectors estimated by GEOSAT and the vectors applied by IERS in the derivation of ITRF'93 is shown in Table 7.

Individual estimates of troposphere and clock parameters were determined for each technique in the analysis. Some of the results are presented in Fig. 1 which shows that it should certainly be possible to apply a common tropospheric parameter representation for the collocated stations. It should also be possible to estimate a common clock drift parameter at each collocated station. The clock estimates from the GPS data show a peculiar periodic behaviour relative to the VLBI estimated clock solutions. Since the period is around 12 hours, it is expected that this phenomena is closely related to the orbital motion of the GPS satellites with the same revolution period. This effect is presently being investigated.

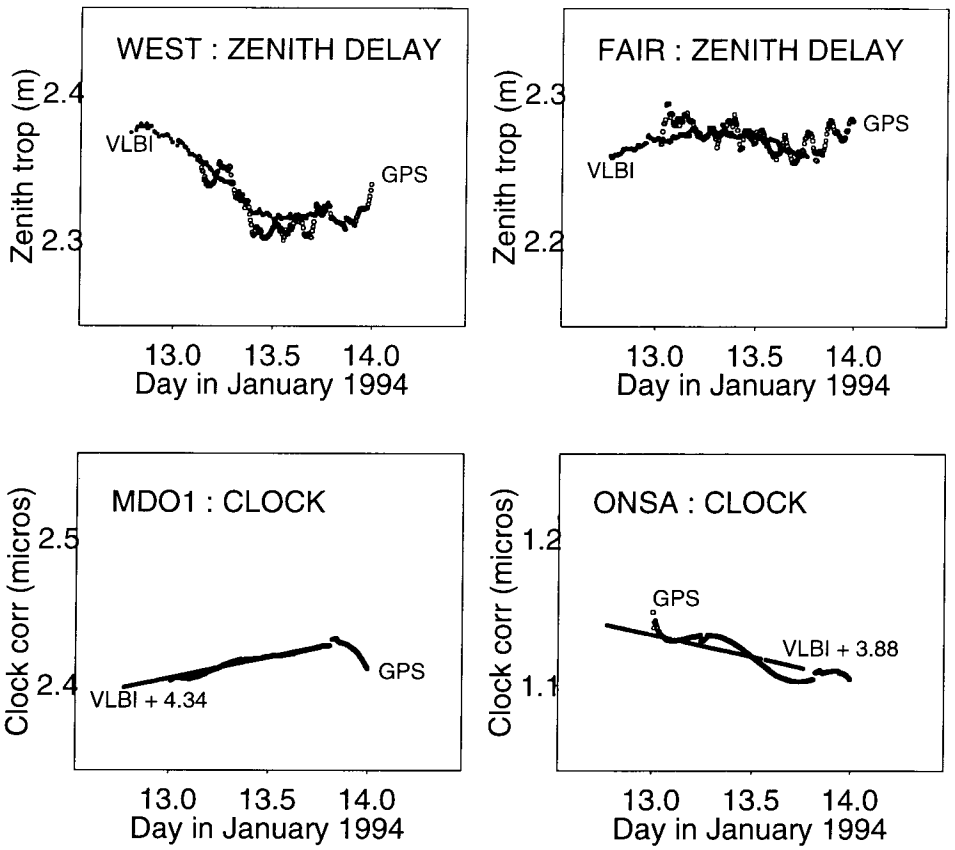


Figure 1: Total tropospheric zenith delay and clock corrections relative to the Westford clock (in milliseconds), estimated with GPS and VLBI data.

The next step will be to repeat the combined analysis with a common solution, not only for the station coordinates, but also for some of the troposphere and clock parameters. In that analysis I will also include WVR (Water Vapor Radiometry) data from the Onsala station.

7 Conclusions

The principles of VLBI, GPS, SLR, PRARE and DORIS have been described. Results from analyses of such data with the GEOSAT software have been presented. The quality of the results is comparable with that obtained by the leading international institutions within each technique.

The combined multi-technique analysis strategy has been demonstrated with very promising results. The solutions for the antenna phase center coordinates of the different techniques seem to be consistent at the few mm-level for stations with precisely determined ground ties (better than 12 mm, 1σ, in each direction).

Acknowledgements

The author is grateful to Statens kartverk for the continuous financial support of the project during many years.

References

1. Andersen, P. H., S. Hauge and O. Kristiansen (1993). GPS relative positioning at the level of one part per billion. Bulletin Geodesique, Vol. 67, No. 1, pp. 91-106.

2. Andersen, P. H. (1994). High-precision Station Positioning and Satellite Orbit Determination, PhD. thesis, the University of Oslo, NDRE/PUBL-95/01094, The Norwegian Defence Research Establishment.

3. Andersen, P. H., S. Rekkedal (1995). VLBI data analysis with the GEOSAT Software. Bulletin Geodesique, Vol 69, No. 3, pp. 125-134.

4. Andersen P. H. (1995). Measuring rapid variations in Earth orientation, geocenter and crust with satellite laser ranging. Bulletin Geodesique, Vol. 69, No. 4, pp. 233-243.

5. Andersen, P. H., K. Aksnes, and H. J. Karlsen (1996). ERS-1 & 2 Tandem Mission Orbit Determination. Presented at the XXI General Assembly of the European Geophysical Society, The Hague, The Netherlands, 6-10 May.

6. Andersen P. H. (1996). WEAG EUCLID CEPA 9 RTP9.1, WP 32100, Final report. Orbit determination with GPS, laser, DORIS and PRARE.

7. Anodina, T. G. and M. T. Prilepin (1989). The GLONASS system. In proceedings of the 5th International Geodetic Symposium on Satellite Positioning, Vol. 1, March 13-17, Las Cruces, New Mexico. U. S. Dept. of Commerce, Rockville, MD.

8. Blewitt, G. (1993). Advances in Global Positioning System Technology for Geodynamics Investigations: 1978-1992. JPL preprint. To appear in Crustal Dynamics Project AGU Monograph.

9. Hartl, Ph., Schäfer, W., Reigber, Ch. and Wilmes, H. (1986). PRARE system description. Proceedings of ESA special Workshop on Solid Earth Science and Application mission of Europe (SESAME), Chiemsee, Germany , 4-6 March. ESA SP-1080.

10. Lefebvre, M. (1989). DORIS system. CSTG Bulletin number 3, DGFI, Munich.