Data acquisition and integration 3.

Global Navigation Satellite System

Bányai László

Data acquisition and integration 3.: Global Navigation Satellite System

Bányai László
Lector: Árpád Barsi

This module was created within TÁMOP - 4.1.2-08/1/A-2009-0027 "Tananyagfejlesztéssel a GEO-ért" ("Educational material development for GEO") project. The project was funded by the European Union and the Hungarian Government to the amount of HUF 44,706,488.

v 1.0

Publication date 2010

Abstract

Main characteristics of the first and second generation GNSS systems are summarised. Signal structures, data acquisition and basic positioning concepts are presented on the bases of GPS system experiences. Error budget, receiver types, observation and data processing methods are explained in details. Complementary systems of the first generation GNSS systems and the possible fields of practical application are also summarised. Last the perspectives of the second generation GNSS systems are revealed.

Table of Contents

List of Tables

Chapter 3. Global Navigation Satellite System

1. 3.1 Introduction

The main purpose of this part is to introduce and summarize the basic principles of global navigation satellite systems (GNSSs). Receiver types, observation and data processing methods are outlined. The presented positioning accuracies can help to choose from the proper devices and services for different fields of applications.

The overview of GNSS concepts is followed by a short description of existing systems and those that are under development in Chapter 3.2.

The elements of satellite positioning − signal structures, data acquisition and positioning concepts − are introduced according to the most frequently used global positioning system (GPS) in Chapter 3.3.

The error budget of satellite positioning and the expected accuracies are summarised in Chapter 3.4.

Chapter 3.5 introduces the basic receiver types in the following categories: navigational, geodetic, CORS and GIS data collectors.

The observation and data processing methods of different receiver categories are discussed in details in Chapter 3.6.

The integrity, availability and accuracy of first generation GNSS systems are improved by complementary systems; they are summarised in Chapter 3.7.

For the sake of completeness Chapter 3.8 lists the fields of possible applications.

The last Chapter summarises the most important features and the perspectives of the near future of GNSS applications.

2. 3.2 General description

2.1. 3.2.1 Prologue

The first successful Earth-orbiting artificial satellite, the Sputnik-1 of the former Soviet Union was launched on 4th October, 1957. This date is the beginning of the space research and the Space Age. The spherical satellite, with 0.58 m diameter and 83.6 kg mass, transmitted on two frequencies (20.005 and 40.002 MHz) with 1 W power. The orbit height was between 220 and 940 km above the sea level. During its 92 days of life time it revolved the Earth more than 1400 times before finally burnt in the atmosphere on 4th January, 1958.

According to the technologies of those times, the observation of the satellite's orbit was carried out by using a photographic method with the help of special cameras. However the observation of the Doppler shifts and the time signals emitted by this satellite opened an unpredictable perspective in the development of global navigational positioning systems, too.

2.2. 3.2.2 Overview of GNSS

In the case of Sputnik-1 all the elements of modern satellite navigation systems were present. The experiences generated a huge development of satellite positioning that is still in progress.

The general purpose of the Global Navigation Satellite System (GNSS) is the continuous supply of the users by enhanced positioning capacity and other required information. Such a system can be divided into three mean segments:

- space segment,

- control segment and
- user segment.

The **space segment** consists of several satellites distributed on properly designed global constellation. The satellites are equipped with radio transceivers, high precision atomic clocks (or frequency standard), on-board computers, solar panels for power supply and propulsion system for orbit maintenance and stabilisation.

The satellites are launched on nearly circular geocentric orbits, which can be characterised by:

- inclination (the angle between the orbit- and the equatorial plane),
- average height above the Earth surface and
- orbital period (or time of revolution).

The useful navigational orbit types are:

- polar orbit (inclination is very near to 90 degrees),
- semi-synchronous inclined orbit (approximately 12-hour period time, the configuration is repeated in every day),
- Geostationary orbit (inclination is very near to 0 degree and the period time is approximately 24-hour; observed from the Earth it appears as “motionless”).

The applied nearly semi-synchronous inclined orbits are (theoretically) evenly distributed along the equatorial plane, and more satellites are evenly distributed along one orbit plane to fulfil the planned navigational requirements.

Geostationary orbit is useful for both navigation and complementary telecommunication purposes as well.

Due to perturbing accelerations the Keplerian orbits of the satellites are only instantaneous. The motion of satellites can be described in the Earth Centred Inertial System by modified Keplerian orbit elements. For positioning purposes the satellite positions have to be transformed to Earth Centred Earth Fixed (ECEF) coordinate systems.

The task of the official **control segment** is the continuous monitoring of the satellites. Based on its observations, the master control station determines the orbits and the satellite clock errors. The orbits and clock parameters are predicted for the next time period and uploaded to the satellites. Along with the clock synchronisation the necessary satellite manoeuvre can be performed.

Parallel to the official control segment there are several local, regional or worldwide organizations that independently monitor the GNSS systems for civilian or scientific purposes. The post-processed precise orbits, clock parameters and the additional by-products can be used to enhance the accuracy and the integrity of the positioning systems. They can significantly contribute to different fields of geosciences as well.

The **user segment** means the various categories of permitted customers (military, state or civilian) and their different types of receiver.

Similarly to the satellites, these receivers have antenna, radio module, clock (or frequency standard) and built-in processor. They receive and decode the signals and the information, determine the positions and store the raw data if it is necessary. (Later it will be discussed in details.)

2.3. 3.2.3 Practical implementations

For the sake of historical fidelity we have to mention that the first operational navigation satellite system was the TRANSIT or NAVSAT − Navy Navigation Satellite System (NNSS) − popularly called as **NNSS Doppler** after its observation and data processing method. It was designed to update the inertial navigation systems used on US Navy’s Polaris atomic submarines.

The satellites were launched on individual polar orbits at approximately 1100 km above the Earth surface with a period of 106-109 minutes. The 4-6 orbits were almost evenly distributed along the equator (Fig. 3-1.).

Based on nominal 5 MHz quartz oscillator two stable coherent frequencies (≈150 MHz and ≈400 MHz) were generated. The predicted orbit elements, navigation and other military information were phase modulated on both frequencies. The timing signals were transmitted in every even minute. The application of two frequencies permitted corrections for ionospheric refraction.

Approximately in every hour one satellite was observed for 15-20 minutes that allowed the single navigational accuracy of 200 m. The positions were derived from the Doppler shift of the received frequencies, measured as Doppler counts.

The NNSS was opened to civil users in 1967. During the geodetic applications the relative accuracy was improved to the level of about 20 cm using the method of multi-location. This system did not fulfil continuous positioning requirement of GNSS systems, but it was proper for navy's navigations. The system was ceased in 1996. (There are no sufficient information on secret Soviet version of Doppler system called “Cikada”.)

The *first generation GNSS systems* are the NAVSTAR Global Positioning System (GPS) and the similar Russian version called GLONASS.

The **GPS** replaced the TRANSIT system. It was realised by the US Department of Defence. The civilian use was allowed by the US Congress. The system was declared fully operational in 1994.

Theoretically 24 satellites are planned on six semi-synchronous orbits (Fig. 3-2.). The orbit inclination is 55 degrees, the average height is 20200 km and the period is 12 hours. The orbits and the satellites are distributed in a way that at any time and from any place on the Earth surface at least four satellites can be simultaneously observed. Because of the persistent satellite upgrades and replacements the number of satellites is more than 24 at the present.

Two carriers are generated - L1 = 1575.42 MHz and L2 = 1227.60 MHz – which are based on 10.23 MHz fundamental frequency. Different measuring and data codes are modulated on the carriers. Code division multiple access (CDMA) method is applied. The different satellites are identified by the code series.

The original positioning concept is based on the measurement of the travel time of code series that allows faster and more frequent positioning than in the case of TRANSIT. The phase information proved to be the means of high precision relative positioning. (The details will be discussed later.)

The **GLONASS** system is very similar to the GPS. The originally military system was opened for civilian use in 2007.

Theoretically 24 satellites are planned on only three near semi-synchronous orbital planes. The orbit inclination is 64.8 degrees, the average height is 19100 km and the period is 11.3 hours.

Unlike the GPS coding, in this system the frequency-division multiplexing (FDM) method was chosen. All the satellites have different frequencies around L1=1.602 GHz and L2=1.246 GHz.

In spite of the fact that the GPS, GLONASS and other supplementary systems fulfil the requirements of global navigational demands, the **second generation GNSS**s means modernised and newly developed systems.

In the case of GPS, new measuring codes are planned for military and civilian for users. The introduction of the third additional L5=1176.45 MHz frequency improve the performance of the system.

The introduction of CDMA coding instead of FDM, and the use of 30 satellites instead of 24 upgrades the capacity of the GLONASS system.

The **Galileo Positioning System** of the European Union (EU) and the European Space Agency (ESA) will be the first non-military system, which is independent of GPS but at the same time the systems will supplement each other. Although the Galileo is in preparation, the concept and the time schedules may be changed according to the available budget in the frame of Public Private Partnership (PPP).

Originally 30 satellites were planned in three near semi-synchronous orbits. The orbit inclination is 56 degrees, the average height is 23600 km and the period is 14 hours. At the moment two test satellites are in orbit.

Four carrier frequencies of the system − E2(L1)=1572.42 MHz, E5a(L5)=1176.45 MHz, E5b=1207.14 MHz and E6= 1278.75 MHz − partly overlap with the GPS carriers. Ten different signals will be coded on the carriers using the CDMA method. The interference with GPS signals have to be avoided.

Based on different signals and carriers five different services with different navigational accuracies will be adopted. The first less accurate is free of charge; all the other more precise services have to be purchased.

The Chinese **COMPASS** (or Beidou-2) is similar to modernized GPS and Galileo systems. 30 satellites are planned in near semi-synchronous orbits. The orbit inclination is 55.5 degrees, the average height is 21150 km and the period is 12.6 hours. 5 additional satellites are planed on geostationary orbit (36000 km height). At the moment 5 satellites are operational.

The used carrier frequencies are B1=1.561098 GHz, B1-2=1.589742 GHz, B2=1.207.14 GHz and B3=1.26852 GHz. The signals are based on CDMA method.

The system is basically designed for military and state use. China officially joined to the Galileo system as well.

The three inclined orbit planes of GLONASS and Galileo together with the geostationary orbit of COMPASS (dashed line) are shown in Fig. 3-3.

Regional navigation systems (Beidoi-1 − China, IRNSS − India, QZSS − Japan) are not subjects of this study.

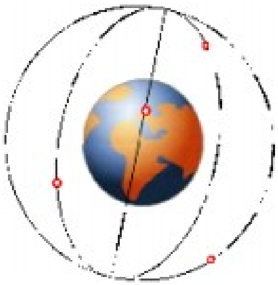

Figure 3-1.: Four polar orbits of NNSS

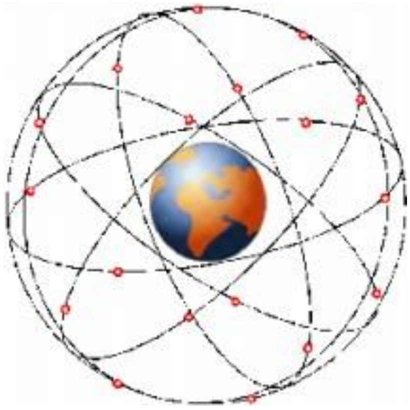

Figure 3-2.: The six orbit planes of GPS

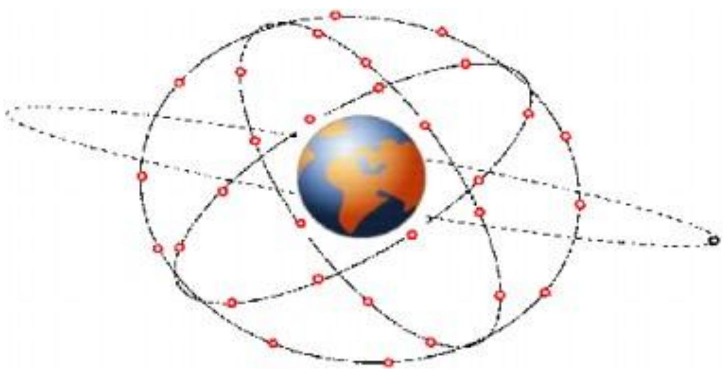

Figure 3-3.: The three orbit planes of GLONASS and Galileo, geostationary orbit of COMPASS

3. 3.3 Elements of satellite navigation

In spite of the evident differences the introduced navigation systems are very similar to each other. Therefore the well known and most frequently applied GPS system is used to introduce the main elements of satellite navigation. The basic ideas and equations can be adapted to all other systems.

The numerical qualification of the positioning performance is a common requirement; therefore the next measures are shortly introduced.

In the geodetic error theory usually the $\sigma_A = \sqrt{c^2 + \sigma_P^2}$ expression is used, where σ_A is the measure of accuracy and σ_P is the measure of precision, c is the average systematic error which cannot be distinguished from the estimated parameters. In this sense the estimated value (including c) and σ_P quantities are the analogues of the expected value and dispersion in probability theory; furthermore, they are the analogues of the mean and standard deviation in mathematical statistics. The σ_P which is estimated by least squares adjustment is frequently called mean squares error.

The probability that the errorless parameter is inside the $-3\sigma_A$ and $+3\sigma_A$ interval around the estimated parameter is 99.7%.

The adjustment theory is not subject of this section. In the following Chapters only σ notation will be used and classified as accuracy or precision.

3.1. 3.3.1 Signal structures

The carrier frequencies L1 and L2 are the multiples of the f_0=10.23 MHz fundamental frequency:

L1: f_1= 154×10.23 MHz = 1575.42 MHz , λ_1=0.19029 m,

L2: f_2= 120×10.23 MHz = 1227.60 MHz , λ_2=0.24421 m,

Where f_1 and f_2 are the frequencies and λ_1 and λ_2 are the corresponding wavelengths.

Different measuring codes (or ranging signals) and data codes (orbital, timing and other military information) are modulated on the carrier waves using binary biphase modulation. The principle is shown in Fig. 3-4. Using the code sequences -1 and +1 (according to binary 0 and 1) the phase is shifted by 180 degrees (if multiplied by -1) or not (if multiplied by +1). The ranging codes are classified as pseudo-random noise (PRN).

The frequency of C/A (coarse/acquisition or civil/access) code modulation is f_0/10=1.023 MHz. The PRN codes are repeated in every millisecond. The code length is 1023 chips (bites) and the chipping rate multiplied by the speed of light equals to approximately 300 m chip length. Every satellite has unique C/A code sequence.

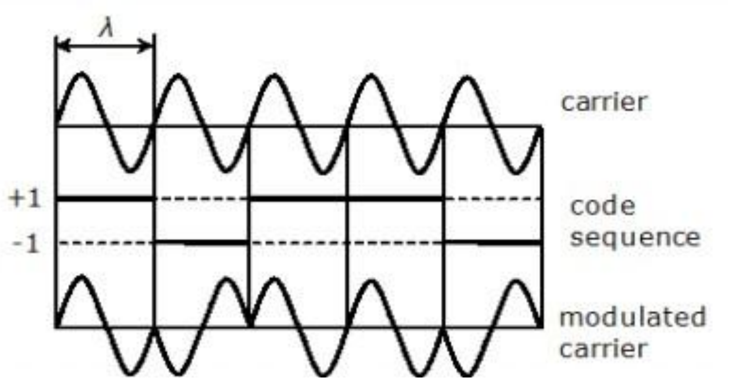

Figure 3-4.: The principle of binary biphase modulation

The frequency of P (precision or protected) code modulation is f_0=10.23 MHz. This code is repeated approximately in every 266.4 days. The chipping sequence is divided into 37 weeks and each satellite has its own weekly sequence. These sequences are repeated in every week. The chip length is approximately 30 m.

The secret P code was opened because of the first Gulf War. To encrypt the P code the Y=P·W code was introduced in the case of newly launched satellites. It is called anti-spoofing (A-S). The frequency of W modulation is $f_0/20$ Mhz.

The data code (D) modulation frequency is $f_0/204600=50$ Hz. The total message contains 1500 bites which are divided into five subframes. The subframes are divided into ten words. One word consists of ten bites.

The navigation messages contain the almanac data (less accurate orbital data), the broadcast ephemerides (BE) and the satellite clock corrections. Each satellite broadcasts its own BE ephemeris and the almanac of all other satellites.

The BE contains modified Keplerian orbit elements that are necessary for position computation. (The GLONASS provides three positions and three velocities in ECEF coordinate system quarter-hourly. This is equivalent to instantaneous Keplerian orbit.)

The GPS modulation can be represented by the equations

$$L1(t)=a_{1P}P(t)W(t)D(t)\cos(f_1 t)+a_{1C}C(t)D(t)\sin(f_1 t)$$
$$L2(t)=a_{2P}P(t)W(t)D(t)\cos(f_2 t)$$, (1)

where a indicates the amplitudes, C replaces the C/A code and t is the time. The C/A and P codes have a quadrature phase relationship on L1.

During GPS modernisation the C/A will be coded on L2 as well, and a new L5 carrier and safer military (M) code will also be introduced.

3.2. 3.3.2 Data acquisition

The receivers according to their purpose can generate the same carrier frequencies and measuring codes as the satellites. This is controlled by the receiver clock (frequency standard).

The coded carriers transmitted by visible satellites are received. These signals are mixed (multiplied) with the receiver generated signals. The mixed signal is split up into high and low frequency parts. The high frequency part is eliminated by filtering. The frequency of the lower part is the difference between the generated and received frequencies. This is the so-called beat frequency which is continuously changing according to the Doppler effects and includes all the coded and phase information of the original signals.

In the case of GPS the satellites are identified by their unique C/A codes. (In the case of GLONASS they are identified by their Doppler shifted unique frequencies.)

The receiver generated (respond) signals are continuously shifted to detect the maximal correlation between the received and generated code sequences. This is called as code correlation method. If the satellite and receiver clocks were errorless, this time shift would be equal to the travel time (τ) of the code sequence from the satellite to the receiver.

The measurement process always starts with C/A code correlations which is easier but less accurate than the Y(P) code correlation.

The measured time shift can be written as

$$\Delta t_R^S(t)=t_R(t)-t^S(t-\tau)$$, (2)

where $t_R(t)$ is the reception time in the receiver time frame and $t^S(t-\tau)$ is the transmission time in satellite time frame. (If $t_R(t)$ was equal to $t^S(t)$, $\Delta t_R^S(t)=\tau$ would hold.) Introducing the clock errors (Hoffmann-Wellenhof et al., 1997)

$$t_R(t)=t-dt_R(t)$$
$$t^S(t)=t-dt^S(t)$$ (3)

the measured time shift is

$$\Delta t_R^S(t) = \tau + dt^S(t-\tau) - dt_R(t)$$. (4)

Multiplying eq. (4) by the speed of light (*c*) we get a range equation

$$P_R^S(t) = \rho_R^S + c \cdot dt^S(t-\tau) - c \cdot dt_R(t)$$, (5)

where ρ_R^S is a distance between the satellite and the receiver.

Even if this range is corrected by the satellite clock error ($dt^S(t-\tau)$) provided by BE messages, it can be handled as only a pseudo-range because of the unknown receiver clock error ($dt_R(t)$).

The receivers usually have a cheap and less accurate quartz clocks therefore the continuous determination of this error (the time synchronisation) is a fundamental issue.

If the codes are known they can be removed from the received carriers. This procedure provides the so-called reconstructed carriers, which can be used for phase measurements.

The phase of the reconstructed carrier from the starting time (t_0=0) to the reception time (*t*) is

$$\varphi_R^S(t) = \int_0^t (f_R - f_S')\,dt$$, (6)

where f_S' is the received, Doppler shifted frequency. This phase is equal to the difference between the receiver and satellite generated phases

$$\varphi_R(t) = \int_0^t f\,dt = f \cdot t + \varphi_R(0)$$

$$\varphi^S(t) = \int_0^{t-\tau} f\,dt = f \cdot (t-\tau) + \varphi^S(0)$$

$$\varphi_R^S(t) = \varphi_R(t) - \varphi^S(t) = f \cdot \tau + \varphi_R(0) - \varphi^S(0) = f \cdot \frac{\rho_R^S}{c} + \varphi_R^S(0)$$, (7)

where $\varphi_R(0)$ and $\varphi^S(0)$ are unknown phases. They can not be separated, their presence is signed by $\varphi_R^S(0)$ which is not exactly an integer number because of the small code synchronisation errors.

Equations (7) are valid in the case of errorless clocks only. Substituting the time correction eq. (3) and $\tau = \frac{\rho_R^S}{c}$ travel time the phase equation is

$$\varphi_R^S(t) = f \cdot \frac{\rho_R^S}{c} + f \cdot dt^S(t-\tau) - f \cdot dt_R(t) + \varphi_R^S(0)$$. (8)

Multiplying eq. (8) by the wavelength ($\lambda = c/f$) we get again a pseudo-range equation

$$\Phi_R^S(t) = \rho_R^S + c \cdot dt^S(t-\tau) - c \cdot dt_R(t) + \lambda \cdot \varphi_R^S(0)$$, (9)

which has an additional phase ambiguity as well.

When the phase is locked, the receiver starts to count the number of full wavelengths (positive zero crossings) and measure the fractional part in time *t*.

The t_0 time of the wave-front coming from the satellite, that is locked the phase loop in the receiver, is not known. This phase ambiguity has to be handled during the data processing.

Different measurement techniques were developed to derive the code and phase pseudo-ranges at both carriers even if the P code is not known:

- *signal squaring* − L2 carrier is multiplied by itself, the Y(P) code sequences disappear, the wavelength will be $\lambda_2/2$ but it is nosier then the original wavelength, it can be used for phase measurements on L2,

- *cross correlation* − the unknown Y(P) code sequences of L1 and L2 carriers are cross correlated (similar to code measurements), the differential code delay and the full wavelength carrier can be reconstructed,

- *Z-tracking* − similar to the previous method, but the known P code is used to improve the performance, because P is similar to Y(P).

The receivers can measure the code and phase observables in the predetermined time sequence. Based on BE massages instantaneous position can be determined, and the raw observables can be stored. The different type of receivers has different measurement capacities.

According to the rule of thumb the accuracy of the code pseudo-ranges can be characterised by 1% of their chip length (C/A: $\sigma \cong 3$ m, Y(P): $\sigma \cong 30$ cm), while the phase pseudo-ranges can be characterised by 1% of their wavelength (L1: $\sigma \cong 1.9$ mm, L2: $\sigma \cong 2.4$ mm). In the case of modern receivers the accuracies are better. The code accuracies can be improved by the combination of simultaneous phase measurements which is called *phase smoothed code*.

3.3. 3.3.3 Positioning concepts

The concept of **single point positioning** is demonstrated in Fig. 3-5. where the receiver (R_1) measures four satellites (S^1, S^2, S^3 and S^4) at the same epoch.

Supposing that the code pseudo-ranges are sufficiently corrected by satellites clock corrections, the range equations (5) can be written as

$$P_1^1(t) = \rho_1^1 - c \cdot dt_1(t)$$
$$P_1^2(t) = \rho_1^2 - c \cdot dt_1(t)$$
$$P_1^3(t) = \rho_1^3 - c \cdot dt_1(t)$$
$$P_1^4(t) = \rho_1^4 - c \cdot dt_1(t) \text{, (10)}$$

where the distances can be given in 3D Cartesian coordinates

$$\rho_1^s(t) = ((X^s(t-\tau) - X_1(t))^2 + (Y^s(t-\tau) - Y_1(t))^2 + (Z^s(t-\tau) - Z_1(t))^2)^{1/2} \text{ . (11)}$$

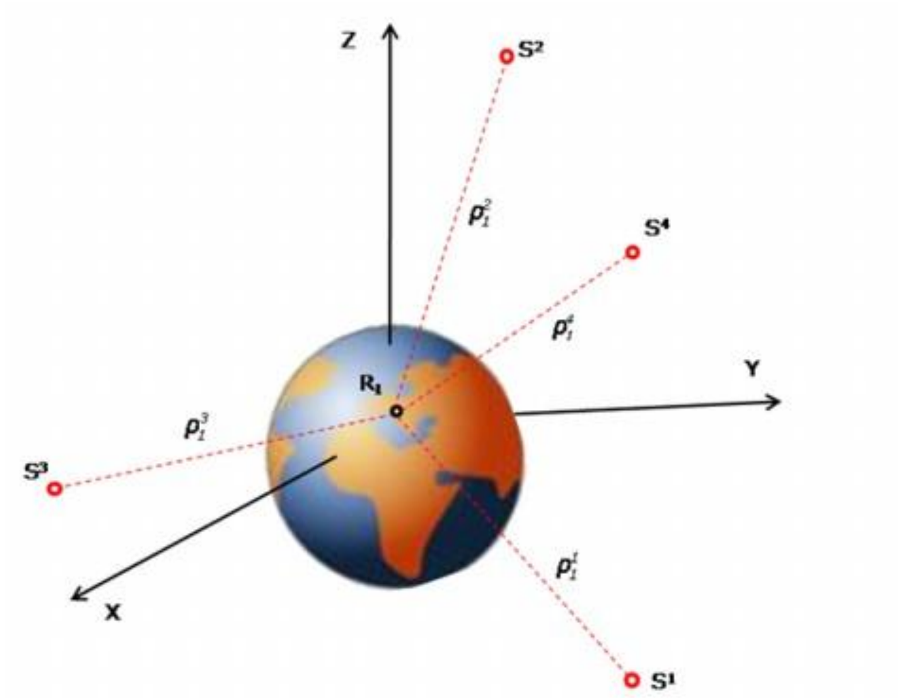

Figure 3-5.: The principle of single point positioning

The satellite positions have to be computed in transmission time from BE messages, consequently the four equations (10) contain four unknowns (X_1, Y_1, Z_1 and dt_1).

The geometric principle of the pseudo-range intersection can be understood easily when plane (2D) simplification is applied first (Fig. 3-6.). If the distances of unknown point P can be measured from the known points P_1 and P_2, than the intersections of the circles, which radii (r_1 and r_2) equal to the distances, give the P and the false P_{12} positions. P can be distinguished from P_{12} if we know on which side it is with respect to the P_1 P_2 line. If the third distance is measured (P_3, r_3) P can be determined unambiguously. The dashed circles differ from r_1, r_2 and r_3 in a constant value. The intersections in couples around P are F_{12}, F_{23} and F_{13}. The solution is the point which is in equal distance from the three arcs defined by F_{12}, F_{23} and F_{13}. This is just the point P.

In space (3D), the intersection of the two spheres around P_1 and P_2 is a circle. This circle intersects the sphere around P_3 in two points. The P is the nearest point to the Earth surface. It can be distinguished unambiguously using the fourth sphere around P_4. If the radii differ in constant value, the four possible triplets of the four spheres give four points near to the Earth surface (F_{123}, F_{124}, F_{134} and F_{234}). The solution P is in equal distance from the four spherical surfaces ($F\mathbf{1}_{23} F\mathbf{1}_{24} F\mathbf{1}_{34}$, $F_1\mathbf{2}_3$ $F_1\mathbf{2}_4$ $F\mathbf{2}_{34}$, $F_{12}\mathbf{3}$ $F_1\mathbf{3}_4 F_2\mathbf{3}_4$ and $F_{12}\mathbf{4}$ $F_{13}\mathbf{4}$ $F_{23}\mathbf{4}$).

The 3D Cartesian coordinates determined in ECEF system can be converted to ellipsoidal latitudes, longitudes and heights above ellipsoid (ϕ, λ and h). The GPS is defined in WGS−84 system. It can be transformed into any other systems and national grid coordinates, as well. The h=u+H is different from the H (height above geoid or mean sea level) in the geoid undulation u. According to the requirements the available undulations have to be taken into account.

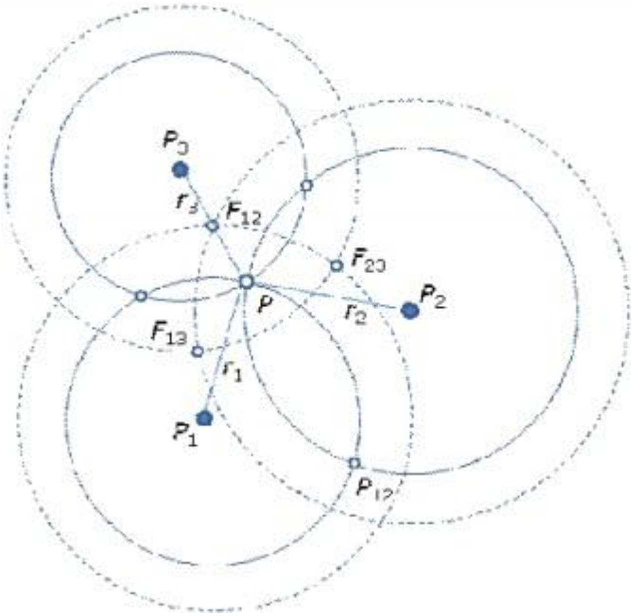

Figure 3-6.: The principle of pseudo range intersection in plane

If the height above ellipsoid can be fixed, three equations are enough for navigational purposes, the clock errors always have to be estimated.

Although there are methods to solve eq. (10) directly, the least square adjustment is preferred because usually more than four satellites can be observed simultaneously. In that case the equation (11) is linearised by first order Taylor series around the preliminary coordinates (X_{01}, Y_{01}, Z_{01}) and their unknown increments (x_1, y_1, z_1) are estimated to fulfil the equations in least squares sense. Neglecting the time dependence, the linearised form of (10) is

$$\begin{bmatrix} P_1^1-\rho_{01}^1 \\ P_1^2-\rho_{01}^2 \\ P_1^3-\rho_{01}^3 \\ P_1^4-\rho_{01}^4 \end{bmatrix} = \begin{bmatrix} \frac{\partial\rho_1^1}{\partial X_1} & \frac{\partial\rho_1^1}{\partial Y_1} & \frac{\partial\rho_1^1}{\partial Z_1} & -1 \\ \frac{\partial\rho_1^2}{\partial X_1} & \frac{\partial\rho_1^2}{\partial Y_1} & \frac{\partial\rho_1^2}{\partial Z_1} & -1 \\ \frac{\partial\rho_1^3}{\partial X_1} & \frac{\partial\rho_1^3}{\partial Y_1} & \frac{\partial\rho_1^3}{\partial Z_1} & -1 \\ \frac{\partial\rho_1^4}{\partial X_1} & \frac{\partial\rho_1^4}{\partial Y_1} & \frac{\partial\rho_1^4}{\partial Z_1} & -1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ c \cdot dt_1 \end{bmatrix}$$

$$\boldsymbol{b} = \boldsymbol{A}\,\boldsymbol{x} \qquad (12)$$

where the bold equation is a general matrix notation.

The following matrix is the analogue of the weight coefficient matrix of the least square adjustment

$$\boldsymbol{Q} = (\boldsymbol{A}^t\ \boldsymbol{A})^{-1} = \begin{bmatrix} Q_{xx} & Q_{xy} & Q_{xz} & Q_{xt} \\ Q_{xy} & Q_{yy} & Q_{yz} & Q_{yt} \\ Q_{xz} & Q_{yz} & Q_{zz} & Q_{zt} \\ Q_{xt} & Q_{yt} & Q_{zt} & Q_{tt} \end{bmatrix}, \quad (13)$$

but here the dimension of the unit weights is not taken into account.

The dilution of precision (DOP) values are defined as

- geometric − *GDOP*= $\sqrt{Q_{xx}+Q_{yy}+Q_{zz}+Q_{tt}}$

- position − *PDOP*= $\sqrt{Q_{xx}+Q_{yy}+Q_{zz}}$
- time − *TDOP*= $\sqrt{Q_{tt}}$
- horizontal − *HDOP*= $\sqrt{Q_{NN}+Q_{EE}}$
- vertical − *VDOP*= $\sqrt{Q_{UU}}$

The north, east and up component (Q_{NN}, Q_{EE}, Q_{UU}) can also be derived.

The smaller the DOP is, the larger the precision will be. The dimensionless values are always equal or larger than 1.0. (In the best case it would be equal to 1.0). Theoretically, more satellites provide smaller DOP, but the geometric distribution of satellites is even more important. In Fig. 3-5. three satellites are evenly distributed at the horizon (S1, S2 and S3) and the fourth (S4) is in the zenith direction, this would be the most ideal case of four satellites.

Imagine that near to R_1 receiver (Fig. 3-5.) there is another receives R_2, which also measures the same satellites. It can be supposed that the two estimated positions will have the same errors. If one of the receivers is placed at a known position, the differences can be used to correct the estimated position of another receiver.

Practically it would be very difficult to measure always the same satellites, therefore one of the receivers have to be placed at the known station where all the satellite above a horizon can be observed. It is called the base station. The distances between base station and visible satellites can be estimated by (11) using the BE data ($\rho_B^s(t)$). The difference of the computed distance and the measured pseudo-range eq. (5) is

$$\rho_B^s(t)-P_B^s(t)=-c\cdot dt^s(t-\tau)+c\cdot dt_B(t)$$. (14)

It can be used as range correction, which contains all the other common errors, too. Correcting the pseudo-range measurement (5) of the rover receiver (that moves from point to point) the next equation is derived

$$P_R^s(t)+(\rho_B^s(t)-P_B^s(t))=\rho_R^s+c(dt_B-dt_R)$$, (15)

where the satellite clock error is cancelled and the difference of the receiver clocks have to be estimated. All the other common systematic errors are cancelled, too, but the random part is exaggerated.

If the corrections can be passed to the rover, the necessary range corrections can be chosen. However, instead of real corrections, the value and the speed of change referring to the reference epoch is passed. Therefore the corrections can be extrapolated.

This method is called as **differential positioning** or **DGPS** because differential corrections are applied.

It was used successfully when the selective availability (S/A) was put in force to significantly degrade the precision of satellite clock corrections in the BE data.

The method, where the positions are determined with respect to the base receiver observations − placed at known position − is called **relative positioning**. This is based originally on direct measurement differences, which cancels the common systematic errors, but exaggerates the random effects. Therefore it is preferred in the proceedings of phase measurements, which have a lower noise level then the code pseudo-ranges.

In the case of two receivers (R_1, R_2) and four satellites the next differences of phase observations (9) can be derived

$$\begin{aligned}
\Phi^1_{12}(t) &= \Phi^1_2(t)-\Phi^1_1(t) = \rho^1_2-\rho^1_1-c\cdot(dt_2(t)-dt_1(t))+\lambda\cdot(\varphi^1_2(0)-\varphi^1_1(0))\\
\Phi^2_{12}(t) &= \Phi^2_2(t)-\Phi^2_1(t) = \rho^2_2-\rho^2_1-c\cdot(dt_2(t)-dt_1(t))+\lambda\cdot(\varphi^2_2(0)-\varphi^2_1(0))\\
\Phi^3_{12}(t) &= \Phi^3_2(t)-\Phi^3_1(t) = \rho^3_2-\rho^3_1-c\cdot(dt_2(t)-dt_1(t))+\lambda\cdot(\varphi^3_2(0)-\varphi^3_1(0))\\
\Phi^4_{12}(t) &= \Phi^4_2(t)-\Phi^4_1(t) = \rho^4_2-\rho^4_1-c\cdot(dt_2(t)-dt_1(t))+\lambda\cdot(\varphi^4_2(0)-\varphi^4_1(0))
\end{aligned}$$, (16)

where the satellite clock errors are cancelled, and all the other satellite related effects can be significantly reduced. This is called *single difference*, where the measurements of the base station (R_1) are subtracted from the measurements of another receiver (R_2).

Subtracting the first row of (16) from the others

$$\begin{aligned}
\Phi^{12}_{12}(t) &= \Phi^2_{12}(t)-\Phi^1_{12}(t) = \rho^2_2-\rho^2_1-\rho^1_2+\rho^1_1+\lambda\cdot(\varphi^2_2(0)-\varphi^2_1(0)-\varphi^1_2(0)+\varphi^1_1(0))\\
\Phi^{13}_{12}(t) &= \Phi^3_{12}(t)-\Phi^1_{12}(t) = \rho^3_2-\rho^3_1-\rho^1_2+\rho^1_1+\lambda\cdot(\varphi^3_2(0)-\varphi^3_1(0)-\varphi^1_2(0)+\varphi^1_1(0))\\
\Phi^{14}_{12}(t) &= \Phi^4_{12}(t)-\Phi^1_{12}(t) = \rho^4_2-\rho^4_1-\rho^1_2+\rho^1_1+\lambda\cdot(\varphi^4_2(0)-\varphi^4_1(0)-\varphi^1_2(0)+\varphi^1_1(0))
\end{aligned}$$, (17)

the receiver clock errors are cancelled, and all the other receiver related effects can be significantly reduced. This is called *double difference*, where the single differences of the base satellite (S^1) are subtracted from the single differences of other satellites (S^2,S^3,S^4). The unique phase ambiguities can not be separated, only their combinations can be treated as unknowns

$$\begin{aligned}
\Phi^{12}_{12}(t) &= \Phi^2_{12}(t)-\Phi^1_{12}(t) = \rho^2_2-\rho^2_1-\rho^1_2+\rho^1_1+\lambda\cdot\varphi^{12}_{12}(0)\\
\Phi^{13}_{12}(t) &= \Phi^3_{12}(t)-\Phi^1_{12}(t) = \rho^3_2-\rho^3_1-\rho^1_2+\rho^1_1+\lambda\cdot\varphi^{13}_{12}(0)\\
\Phi^{14}_{12}(t) &= \Phi^4_{12}(t)-\Phi^1_{12}(t) = \rho^4_2-\rho^4_1-\rho^1_2+\rho^1_1+\lambda\cdot\varphi^{14}_{12}(0)
\end{aligned}$$. (18)

It's easy to understand that $\varphi^{12}_{12}(0)$, $\varphi^{13}_{12}(0)$ and $\varphi^{14}_{12}(0)$ combinations should be integer numbers as the consequence of calculating double difference.

If the coordinates of base station are fixed, the three equations (18) have six unknowns. Because the phase ambiguity combinations are constant during the continuous phase lock, theoretically at least two epochs

$$\begin{aligned}
&\Phi^{12}_{12}(t_1),\ \Phi^{13}_{12}(t_1),\ \Phi^{14}_{12}(t_1)\\
&\Phi^{12}_{12}(t_2),\ \Phi^{13}_{12}(t_2),\ \Phi^{14}_{12}(t_2)
\end{aligned}$$ (19)

are needed to determine the three coordinates of the rover and the three ambiguity combinations (instead of eight individual phase ambiguities).

In the practice more satellites, longer time span and least square adjustment is applied, where the correlations introduced by difference calculation have to be handled in rigorous way.

In the first step, the phase ambiguity combinations are estimated as real numbers. The estimation of the most probable integer numbers is called *phase ambiguity resolution*. The resolved values are treated as known quantities that significantly improve the geometric condition of eq. (18).

From mathematical point of view the same results can be provided without calculating differences. It is called *un-differenced relative positioning*.

The unknowns of eq. (9) can be solved if the clock and phase defects of the equations are properly constrained. The clock errors of one receiver (or one satellite) and the combination of properly chosen phase ambiguities have to be fixed.

For completeness, the measurements (19) can be time differentiated

$$\begin{aligned}
\Phi^{12}_{12}(t_2)-\Phi^{12}_{12}(t_1) &= (\rho^2_2-\rho^2_1-\rho^1_2+\rho^1_1)_{t_2}-(\rho^2_2-\rho^2_1-\rho^1_2+\rho^1_1)_{t_1}\\
\Phi^{13}_{12}(t_2)-\Phi^{13}_{12}(t_1) &= (\rho^3_2-\rho^3_1-\rho^1_2+\rho^1_1)_{t_2}-(\rho^3_2-\rho^3_1-\rho^1_2+\rho^1_1)_{t_1}\\
\Phi^{14}_{12}(t_2)-\Phi^{13}_{12}(t_1) &= (\rho^4_2-\rho^4_1-\rho^1_2+\rho^1_1)_{t_2}-(\rho^4_2-\rho^4_1-\rho^1_2+\rho^1_1)_{t_1}
\end{aligned}$$, (20)

which is called *triple difference*. Although all the nuisance unknowns are cancelled, it is usually used for error detection because it exaggerates the random effects significantly.

4. 3.4 Error budget of positioning

The measurement accuracies, the fundamental clock errors, the phase ambiguities and the geometric dilutions of precision have already been introduced. Other significant error sources are connected to satellites, signal propagation and receivers.

Because the satellite clocks travel along their orbits and the receiver clocks rotate around the Earth rotation axis − with respect to the inertial system − and they are at different equipotential surfaces, both the *special* and *general* **relativistic corrections** have to be applied as clock corrections.

Based on spherical Earth model the average *general* term is compensated by the nominal shift of the satellite frequency standards. The *special* corrections can be computed from average satellite velocity (spherical orbit) or from instantaneous velocities. The BE data contains all the necessary parameters. The relativistic effect on the Earth is less by two orders of magnitude.

The largest **clock errors** originate from the precision and the continuous ageing of satellite frequency standards, therefore it has to be estimated regularly by the control segment at an acceptable level.

All together the remaining satellite clock errors can be characterized by about $\sigma \cong 2$ m accuracy.

The next significant error source is related to the satellite orbit, which is treated as known quantity during the single positioning concept. The accuracy is significantly different in the cases of the predicted BE and the post processed precise ephemerides (PE). The **orbit error** depends on several factors (e.g. satellite perturbation models, observational data and computational procedures). Generally the BE orbit can be characterized by $\sigma \cong 2$ m accuracy. The predicted real time orbits can reach the $\sigma \cong 10$ cm precision level. The final PE orbit precision is $\sigma \cong 5$ cm or even better.

The largest effect is caused by **ionospheric refraction**, because the ionosphere is a dispersive medium in the frequency domain of GPS carriers. This effect is proportional to the total electron content (TEC) along the ray pass and reciprocally proportional to the square of carrier frequencies. The higher the frequency is, the lower the effect will be.

The codes travel at group velocity. The phase velocity is the same but it has a negative sign. Therefore the code ranges are longer, and the phase ranges are shorter than the geometric distances. The code delays, the phase advances.

According to daily and seasonal variations of TEC the effect is about 0.1-50 m in the vertical direction. (During an ionospheric storm it may be significantly larger.) The effect is larger in lower elevations because of the longer path in the ionosphere.

Practically 90% of the effects can be reduced by the measurements carried out at two carrier frequencies (that is the main reason of two carriers). The ionospheric errors can be characterized by $\sigma \cong 2$ m accuracy. The effectiveness of other models (e.g. Klobuchar model in the BE data) is about 50%, and the accuracy is about $\sigma \cong 5$ m. The effects can be reduced by DGPS and relative positioning methods.

The **tropospheric refraction** is independent from the frequency domain of GPS carriers, therefore equally delay the code and phase measurements. It can be handled by the principle of geometric optics. The refraction depends on pressure, temperature and humidity of the troposphere along the ray path. It can be divided into “dry” and more variable “wet” components, which can be estimated using the surface meteorological data. In the case of standard data the dry component is about 2.3 m and wet is 0.3 m in the vertical direction. In lower elevations they are significantly larger. The tropospheric errors are usually characterized by $\sigma \cong 1$ m accuracy, although it can be modelled more accurately.

The most dangerous error source is the **multipath** effect. If the coherent signal fronts reach the antenna in two different ways the phenomenon of interference is experienced. The first is the direct shortest and the second is the reflected longer path. The code interference larger than one code chip length and the wave interference larger than one carrier wave length can be filtered effectively.

The multipath can be recognised easily on the time series of the observations, because it is repeated in every day at the same time as a consequence of semi-synchronous orbits. Several methods and more resistant antenna types were developed to reduce these errors. The code multipath errors may be characterized around $\sigma \cong 1$ m accuracy.

The high precision applications require the knowledge of antenna **phase centre offsets and variations**. The phase centre, from which the wave fronts leave the satellite transponder, and the phase centre of the antenna, where the wave front is received, is continuously changing as function of carrier frequencies, azimuth and elevation angle of the signal path. The average phase centre and the variations depending on the directions can be determined in laboratory with respect to geometrically defined reference point. The orbits refer to the centre of gravity of the satellites, which is different from the average phase centre.

The **cycle slip** is the interruption of continuous counting of the full wavelengths, which may be caused by large noise level or obstructions. The short slips can be resolved or new phase ambiguities have to be introduced during the relative data processing.

The user range error (URE) can be estimated taking into account the relevant accuracies

$$\sigma_{URE} = \sqrt{\Sigma \sigma^2} \quad . (21)$$

The errors of parameters estimated by single point positioning are qualified by

$$\sigma = \sigma_{URE} \cdot DOP \quad . (22)$$

Because the *DOP* values are ≥ 1.0, the precisions are "diluted" by the geometric distribution of the used satellites.

There are several estimations of URE values in the literature. The error sources of single positioning are summarized in Table 3-1.

Table 3-1. táblázat Approximate error sources of navigational applications (m). Approximate error sources of navigational applications (m)

Error sources	C/A code (L1)	common effects	Y(P) codes (L1, L2)
satellite orbit		2	
satellite clock		2	
ionospheric refraction	5		2
tropospheric refraction		2	
multipath		1	
receiver code noise	3		0.3
σ_{URE}	6.9		4.1

The global standard expectations of GPS standard positioning service (SPS) − C/A code − and the precise positioning service (PPS) − Y(P) code − are given in (http://www.pnt.gov/public/docs/2007/ppsps2007.pdf) for horizontal and vertical components:

PPS: $3\sigma_H = 6.3$ m , $3\sigma_V = 13.6$ m , SPS: $3\sigma_H = 36.0$ m , $3\sigma_V = 77.0$ m .

The DGPS positions can be estimated by $3\sigma = 2\text{-}10$ m.

The double differenced phase observables can provide the position accuracies at cm level or even better. It is widely accepted that the modern geodetic receivers can provide

$\sigma_H = 3$ mm , $\sigma_V = 6$ mm

geodetic precisions under normal conditions (< 4 km baseline, < 15 minutes observation time). The prospects of manufacturers always quantify the accuracies of their products.

5. 3.5 Basic receiver types

The data acquisition methods, the positioning concepts, the error budget and the purpose of the applications are closely related to different receiver types. Based on the purpose and accuracy requirements, two main receiver types: navigational and geodetic receivers can be distinguished.

The desired accuracy of **navigational receivers** is usually not better than 0.5-1.0 m. Therefore they are definitely based on code pseudo-ranges. *C/A code and P(Y) code receivers* can be distinguished. The raw observational data are not stored. Only the derived positions can be stored if it is desired. Over the instantaneous position determination they can be used for special navigational purposes as well. The predefined routes, which are given by breakpoints, can be followed where the alterations are continuously monitored. Additionally, *hand-held* and *built-in* receivers can be distinguished. The built-in receivers can be applied in the means of land, maritime and air transportations.

The highest navigational accuracy can be provided by differential, or DGPS positioning method. In that case additional modules are needed to receive the necessary corrections (see later).

In consequence of huge progress in the electronics (telecommunication, integrated circuits, processors, memory cards and graphic interfaces) the receivers became more accurate and more miniaturised.

The next step of the development was the inclusions of graphic displays and interfaces which made it possible to integrate the digital maps. This opened a new perspective of the navigations.

The miniaturised GPS circuits and antennas can be attached or integrated in different devices of other purposes e.g. hand-held computers or smart (or mobile) phones (assisted GPS − A-GPS).

This concept led to the new type of the so-called **GIS data collector**, where special graphic computer, GIS (Geographic Information System) software and GPS module is integrated for real-time land use.

The manufactures designed individual navigation receiver pairs, one of them to pass and the other to receive DGPS corrections. The data can be broadcasted on URH radio frequencies or via cellular phone modems.

The accuracy requirements of **geodetic receivers** are usually about 1-5 cm (or even better). It requires the full ranges of code and phase measurements and the application of relative data processing technique to benefit the high accuracy of phase measurements. The raw data can be stored for post-processing.

The size and the weight of these receivers were dramatically reduced, as well, while the accuracy was increased. The modern receivers can be equipped with additional devices to receive the data, which are necessary for nearly real-time relative data processing. Such a receiver can be integrated in geodetic total stations (originally designed for digital angle and distance measurements) to determine the positions for geodetic mapping purposes. In the case of total station the graphic representations and digital mapping is the trend of the new developments.

Similarly to navigational receivers, there are individual geodetic receiver pairs, too, that passes and receives the necessary data for relative data processing.

Among the geodetic devices one “subspecies” can be specified, which have the same accuracy requirements, but instead of field accessories, it is designed as **CORS receiver** (continuously operating reference station). They are practically internet servers, and transfer the continuous observation data stream from a permanent station to different data and processing centres to maintain specific requirements and services.

In the future the differences in size and in the data acquisition capacity of the satellite sensors will disappear. The navigational and geodetic receivers can be distinguished only by different antenna designs and special additional services, which determine the differences in accuracy and price.

6. 3.6 Observation and data processing methods

Regardless of the two basic receiver types, the *observation methods* can be classified in three basic categories:

- **static** – the receivers are standing on Earth surface,
- **kinematic** – (after a short standing period) the receivers are in continuous movement with respect to the Earth surface,
- **semi-kinematic** – the standing and moving periods are alternating.

According to the time of data processing two *computational categories* can be distinguished:

- **real-time processing** – in (nearly) the same time with data acquisitions, the computations are carried out in rover receivers,
- **post-processing** – the collected raw date are processed later with different software.

Based on three receiver types (navigational, geodetic and CORS), three basic positioning concepts (single, DGPS and relative), three observation and two computational categories – all together 54 combinations – can be distinguished from statistical point of view.

There are, of course, several unreasonable combinations, while the others can be further divided into subtypes. In the following the most important combinations will be summarised according to the receiver types.

Navigational receivers:

The static application of navigational receivers is based on equations (12). The position series can be averaged after the predefined number of position, or the average is updated continuously, and can be stopped if the precision is no more improved. The Y(P) receivers are more accurate than the C/A receivers. The DOP values indicate the geometric condition. If the PDOP is larger than 7-9, the position cannot be accepted. This procedure belongs to real-time processing.

The kinematic application of navigational receivers is based also on equations (12), but the position series refer to different geographical points. The performance of the receiver can be improved in real-time by Kalman-filtering. The velocity and the acceleration of position series can be used to predict the next position, which is updated by the next code measurements. The Kalman-filtering (Starng and Borre, 1997) is not subject of this material.

There are procedures, which automatically recognise that the receiver is standing or moving, and apply the corresponding methods. It can be classified as semi-kinematic procedure.

Instead of eq. (12) the three observation categories can be solved by DGPS processing concept eq. (15), too.

Because the raw observations are not stored, traditional post-processing is not possible. However the stored position series can be smoothed during the practical applications.

Geodetic receivers

The geodetic accuracy requirements can be provided only by phase measurements and relative data processing methods.

The strategies of static post-processing can be distinguished as short, medium and long baseline solutions (Kleusberg and Teunissen, 1996). It is usually started by single positioning of code measurements to have better preliminary parameters.

In the case of short baselines (< 4 km) usually standard tropospheric models are used and their residual and the ionospheric effects are cancelled by calculating double differences of L1 or L1 + L2 phase measurements.

The double differences are not sufficient to reduce the ionospheric effects on medium length baselines (≈ 4-40 km). Therefore the ionosphere free linear combinations of L1 and L2 have to be applied, but in that case the phase ambiguity combinations loose their integer property. In the meantime, taking into account the features of wide lane and narrow lane linear combinations, the L1 and L2 ambiguities can be resolved and ionosphere free solution can be performed. The random errors are however exaggerated.

In the case of long baselines (> 40 km) the tropospheric effects can not be reduced effectively by double differences. Therefore, in addition to the previous solution, residual tropospheric unknowns are estimated.

Kinematic observations can also be processed if the phase ambiguity combinations can be resolved on the bases of the short starting period, and they can be treated as known quantities during the moving period. In the case of modern receivers there are methods to solve the ambiguities even on the fly (OTF).

The combination of the previous solutions, the semi-kinematic observations can also be processed if stop and go flags are registered in the data files. It is frequently called as Stop & Go method.

The post-processing computations can be executed usually by commercial software provided by the receiver manufacturers.

There are geodetic receivers which can be used for real-time relative processing, if the base station (or equivalent) data can be passed to the rover. It is popularly called as Real-Time (semi) Kinematic or **RTK** method.

If high accuracy orbits, satellite clock corrections, ionospheric and tropospheric parameters can be determined and extrapolated by regional data processing centres, furthermore, the accuracy of code pseudo-ranges will be comparable to the carrier wavelength, the difference between equations (9) and (5) will be

$$\Phi_R^S(t) - P_R^S(t) = \lambda \cdot \varphi_R^S{}_0 \quad . (23)$$

This can be used to determine the starting, individual phase ambiguities. In that case the code pseudo-ranges can be replaced by more accurate phase-pseudo ranges, and single point positioning can be applied for static, kinematic or semi-kinematic observations as well. It seems to be the new trend in the near future.

CORS receivers

In this case the static code and phase observations are post-processed. The baselines of different networks can be classified as long or even very long baselines. Special combinations of different observations and procedures are developed to reduce the error effects and provide the desired products of the networks. According to the purposes unique scientific software are used and constantly improved.

The control networks are devoted to produce the satellite orbits and clock corrections. In that case not only the ground positions, but the satellite positions of eq. (11) have to be handled as unknowns, too. Even if all the ground positions are known, it is not reasonable to determine the time series of satellite positions only from range measurements. These series are replaced in the inertial coordinate system by the unknown three starting positions and the three starting velocities of the observed orbit arc. These unknowns are equivalent to the instantaneous Keplerian orbit. The known perturbing accelerations are taken into account using the solution methods of second order differential equations.

The transformation of the orbits into the ECEF system requires the knowledge of the geodynamic phenomenon: precession, nutation, polar motion and Earth rotation as well. Additional unknown corrections can be assigned to the perturbation and geodynamic models to improve the solutions.

All the ground positions and clock errors can not be estimated unambiguously; therefore sufficient numbers of positions and clock unknowns has to be fixed or constrained. The Kalman filtering can also be used to predict the orbit positions.

7. 3.7 Complementary systems

The PPS concept of the GPS system was designed for military purposes. Only the less accurate SPS was opened for civilian users. However the civilian designers and scientist achieved − a not expected − huge development in the field of data acquisition and data processing.

To degrade the real-time positioning capacity of the civil SPS users the selective availability (S/A) was introduced. In practice the satellite clock corrections were artificially destroyed and the degradation of BE orbit was planned.

The answer of civilian user was not so late. Private or scientific control networks were established to determine the degraded elements. The DGPS concept significantly improved the C/A code navigations, which made the S/A concept highly inefficient. The S/A was stopped in 2000.

The civilian applications also need performance requirements: integrity (system alerts and fast reaction) availability and accuracy of the satellite systems. Therefore different autonomous *augmentation* systems were created to complement the performance of the first generation GNSS systems.

The **IGS** (recently − International GNSS Service) is the first and largest international organisation of scientist, which provides precise data for scientific applications. The cooperation of more than 400 selected permanent stations of worldwide, regional and national networks collect the GPS and GLONASS observations. The stations are arranged into regions and there are several data centres which collect the data in hourly (or still in daily) packages via Internet. The regional processing centres can provide hourly, daily and weakly solutions. The weekly solutions are combined to produce final orbits, clock corrections and different geodynamic parameters: tectonic movements, polar coordinates and change of the length of day (LOD).

The IGS contributes to the maintenance of the International Terrestrial Reference System (ITRS). The relevant permanent networks are given in ITRF (International Terrestrial Reference Frame). The coordinates and the determined velocities refer to the reference epoch. The velocities describe the instantaneous trend of tectonic movements.

The ionospheric vertical TEC values and the zenith tropospheric delays (ZTD) are also produced by the regional data centres as important by-products.

The estimated and stored data can be downloaded from the IGS home page free of charge. The accuracy categories, the latency, the actualization, the sampling rates and all the other information can be found at the IGS webpage (http://igscb.jpl.nasa.gov/).

The activity of the **EUREF EPN** (European Permanent Network) is very similar to the IGS. The EPN maintains the European Terrestrial Reference System (ETRS) and the ETRF (European Terrestrial Reference Frame). The ETRF is connected to the Euro-Asian tectonic plate. The transformation parameters between ITRF and ETRF are known. The details of the EUREF EPN organisation can be found in the webpage (http://www.epncb.oma.be/).

There are other civilian organizations and/or services where the collected data are pre-processed and the required data can be broadcasted in real time (or later) to the different customers, who pay for the services.

The broadcasting can be based on available regional Radio, cellular phone or satellite telecommunication services. The cellular phone modems using GPRS (General Packet Radio Service) technology can connect the customers via internet protocol (NTRIP) to the GNSS services. It can be used only in highly developed regions.

The data messages are standardised by the RTCM (Radio Technical Commission for Maritime) international organisation (with more than 120 members). All the data required by DGPS or precise relative data processing − including code and phase measurements − have been standardised (http://www.rtcm.org/).

As an example the Hungarian active GNNS Network (**GNSSnet**) is introduced. The data and computational centre receive the real time observations from the Hungarian and from some neighbouring stations in every second. The data are prepared and stored for different DGPS and relative (RTK) applications.

The observations of selected IGS and EPN stations are forwarded to the relevant data centres. The real time geodetic (RTK) users can receive the data via internet and cellular phone modems with a few second delays.

According to the user's receivers and the subscription different data can be acquired:

- data from the nearest or user selected station (the benefit of the network is not used),
- data from the nearest station together with error parameters (estimated from all the stations),

- data from virtual station (it is defined near to the receiver, all the stations are used),
- data from the nearest station and their differences from three other stations (equivalent to data of four neighbouring stations).

Some other representative augmentation systems are also introduced.

Satellite-Based Augmentations Systems (**SBAS**) − where geostationary satellites broadcast the DGPS corrections − are:

- Wide Area Augmentation System (**WAAS**) is maintained by the US Federal Aviation Administration to improve the integrity and accuracy of GPS based aircraft navigations. The observations are collected by North American stations. The master station sends the DGPS corrections to the WAAS geostationary satellite.
- **WAAS EGNOS** (European Geostationary Navigation Overlay Service) of the European Commission and ESA is developed to supplement the GPS, GLONASS and Galileo systems. The DGPS corrections determined from European stations are sent to three geostationary satellites.
- **OmniSTAR** Corporation broadcasts DGPS corrections for eight important regions of the Earth.

The Ground-Based Augmentations Systems (**GBAS**), Ground-Based Regional Augmentation system(**GRAS**) and Local Area Augmentation System(**LAAS**) − where ground radio transmitters broadcast the DGPS correction on VHF or http://en.wikipedia.org/wiki/Ultra_high_frequencyUHF bands to the users − are:

- US LAAS for precision landing approach of Aircrafts.
- United States Coast Guard (USCG) for maritime usage.
- United States Nationwide DGPS (NDGPS) for transportation.
- Canadian DGPS mainly for maritime usage.

The concept of pseudolite (pseudo-satellite) is the possible augmentation of ground based systems in the future. The pseudolites broadcast the same signals as the satellites from the known ground positions. It may be used for indoor navigation, too.

8. 3.8 Practical applications

In the previous Chapters there are several references to different fields of possible applications. From general point of view, GNSS can be used by all the application fields which require position information on and around the Earth. In this Chapter the main applications are consecutively summarised without the military applications.

Geodynamics

This field of applications is connected to the scientific control networks. It was introduced by the activity of **IGS** in more details.

The **geodynamic phenomenon** of Earth rotations and global tectonic movements are continuously estimated and the basic geometric frames of **geodetic sciences** (ITRS and IRTF) are maintained. This field needs the highest accuracy and precision.

Aeronomy

The continuous monitoring of regional vertical TEC values significantly contributes to the ionospheric investigations.

Meteorology

The zenith tropospheric delays (ZTD) as the by-products of regional networks can be used to predict the precipitation water (or integrated humidity) above receivers.

Geodetic survey and mapping

The main tasks of the geodetic survey are the establishment of the geodetic networks, the determination of the measurement points which are used for cadastral and topographic mapping, furthermore to set out the representative points of the planned engineering constructions.

The local (or regional) deformation networks and the higher order geodetic networks require the highest accuracy, which can be provided by *static* (30 minute or even longer) or fast static (< 30 minute) observation periods. The post-processed and redundant baselines can be additionally adjusted in the geodetic networks.

In Hungary a dense GPS network (OGPSH) was established to provide a frame for GNSS applications.

The lower order geodetic benchmarks and the measurement points can be determined by semi-kinematic (stop & go) method under some seconds depending on the distance of the OGPSH or equivalent base stations.

The Hungarian active GNNS Network made possible the real-time (semi) kinematic (RTK) applications. It can be applied in geodetic setting out, too.

In the future the national geodetic frames can be maintained by active GNSS networks together with additional, properly chosen traditional horizontal, vertical and gravimetric benchmarks.

In the areas where the use of precise GNSS is limited they have to be applied together with traditional geodetic devices.

Photogrammetry and remote sensing

In this field of data capture the GNSS applications are twofold. The positions of the sensors carried by Aircrafts (or spacecrafts) can be continuously navigated and determined. The control or tie points of the image processing can be measured on the ground.

Geographic Information System (GIS)

Digital geodetic maps, photogrammetric and remote sensing materials together with relevant attribute data can be stored and analysed by GIS software. These data can be actualised by additional GNSS measurement. If the accuracy demand is less than 0.5-1 m GIS data collectors can also be used.

Civil engineering

All the fields of civil engineering can benefit from the GNSS techniques which apply surveying engineering techniques in the phases of preparation, construction and operations:

- railway construction and geometry control,
- road construction,
- architecture construction, stability and deformation control,
- water management and hydrology.

Geology and geophysics

The geological and geophysical exploration measurements also require position information. These data are graphically represented on derived maps; therefore GIS tools can also be used even in the field. DGPS or geodetic accuracies are required.

Mining engineering

The GNSS technology can partly contribute to the geometric demands of mining activities. The setting out of the working areas of open-cast mining and the determination of peat extraction may be assisted by GNSS. Another field is the monitoring of surface subsidence caused by underground mining. The surface reference points of the underground traversing can also be determined by GNSS. The large accuracy demand of the height determination is the limiting factor of the mining applications; it can be improved by integrated spirit levelling.

Agriculture

One of the most demanding fields is called as precision agriculture. The GNSS assisted tractors and machines can precisely follow the predefined trajectories of planting, fertilising and plant protection. The amount of fertilizers or sprayed fluids can be provided in function of the geographic positions.

Forestry and wildlife management

Although the accuracy of GNSS application is limited under forest canopy there are several possibilities of the utilization. In the high precision applications the GNSS and the traditional methods are usually combined, because the density of the geodetic benchmarks is less satisfactory in woodlands. The DGPS accuracy is sufficient for forest cultivation and protection purposes. The small collar navigational receivers with internet connection can be used e.g. to study the natural habitat and geographic territory of deer populations.

Nature conservation and environment protection

These interdisciplinary fields deal with the consequences of the above mentioned human activities to maintain the natural and human resources. The GIS and the DGPS technique can be used effectively in the field of survey, registration and the control measurements.

Direct navigational applications

The “navigation” originates from the roman phrase “*navigare necesse est”* which means: to sail is necessary.

- *Maritime navigation* − The first satellite navigation system was created originally for maritime purposes. The method of Dead reckoning − measuring direction, velocity and time from known positions − was updated by satellite positioning. While the SPS service is sufficient for ocean and sea navigations, the approach of coast lines and harbours needs DGPS accuracies (0.5 − 1.0 m). The traditional coast guard radio systems can be replaced by DGPS services (e.g. USCG GBAS).

- *Aviation* − Aircraft navigation can be divided into several phases of departure, route flight, airport approach and landing. The different phases require different accuracies. There are automatic Flight Management System (FMS) that can control the full flight. Integrated navigation systems can be used in different phases: INS (Inertial Navigation System), DGPS and other radio systems. In route flight the GPS is still an auxiliary device. However the US LAAS is designed for Precision landing approach using DGPS technology. In the future the full navigation will be based on satellite systems (IGSANS − Integrated Global Surveillance and Navigational System).

- *Ground transportation* − Most of the users belong to this category. Two large categories can be distinguished: private users, as well as state or civilian firms and organisations. The first category uses the hand-held or car navigation GNSS devices to follow the predefined routes, which is aided by maps of road systems. In the case of second category there is an additional demand, the so-called fleet management. The positions of the fleet members are continuously sent to the control centres. The systems can be used by:

  - public transport,
  - railways,
  - waste collection services,
  - ambulance services,
  - fire and protection services,
  - police, etc.

Holiday, sport and hobby

There are several additional possibilities to civilian users to apply cheap hand-held navigation receivers for different free time occupations.

9. 3.9 Summary and perspectives

In this part the existing first and the second generation GNSS systems were shortly introduced. At the same time, the basic positioning concepts, observation and data processing methods of different receiver types were emphasised in details. The error sources that affect the positioning accuracy together with the necessary complementary systems were also discussed. Hopefully, those who study this material carefully can distinguish the DGPS and the relative data processing methods (e.g. RTK).

The GNSS is an amazing device, but not Aladdin's Magic Lamp. Summarising the several aspects of GNSS applications, users − who have no thorough knowledge of geodetic positioning − can select the most appropriate and most economic receivers and services for the solutions of their own practical or scientific problems.

The future perspectives were shortly mentioned in the case of second generation GNSS systems. It was stated that the GNSS systems should provide additional information by side the most important geometric positions, too. Except the Galileo, all the other systems are originally devoted to military purposes. The civilian users get the necessary information mainly from different augmentation systems.

In the case of new concepts the augmentations will be highly integrated in the satellite systems together with regional control and data uploading systems. For example, the five accuracy and application services of the planned Galileo system are mentioned as the near future's perspective:

- Open service (OS),
- Safety of Life Service (SoL),
- Commercial Service (CS),
- Public Regulated Service (PRS),
- Search and Rescue (SAR).

The data can be forwarded to the users by integrated geostationary, or even directly by the positioning satellites passing over the region. It can also be used on such territories, which do not possess developed electronic telecommunication infrastructures, too.

The activities of the IGS organization demonstrated that the regional data and computational centres can provide large accuracy estimated data (orbits, clocks, ionospheric and tropospheric delays) which is required by precise real-time positioning. In that case, it will not be reasonable to distinguish the different receiver types, observational and data processing methods.

Self-testing questions:

1. Which are the three basic segments of GNSS systems? (Chapter 3.2.2)
2. Which orbit types are used by GNSS systems? (Chapter 3.2.2)
3. Which are the first and second generation GNSS implementations? (Chapter 3.2.3)
4. What is the biphase code modulations used for? (Chapter 3.3.1)
5. Which are the two basic GNSS observables? (Chapter 3.3.2)
6. Which are the three basic positioning concepts? (Chapter 3.3.3)
7. Specify the most important error sources! (Chapter 3.4)
8. Which are the four distinguished receiver types? (Chapter 3.5)
9. Classify the observation and data processing possibilities! (Chapter 3.6)
10. Which are the basic complementary systems? (Chapter 3.7)
11. Which are the most important fields of possible applications? (Chapter 3.8)
12. Which are the main perspectives in the future? (Chapter 3.9)

Bibliography:

Hoffmann-Wellenhof B,Lichtenegger H,Collins J: *GPS Theory and Practice. Fourth, revised edition.*, Springer Wien, New York., 1997.

Leick A.: *GPS SatelliteSurveying. Second Edition.*, John Wiley&Sons, Inc., New York, Chichester, Toronto, Brisbane, Singapore., 1995.

Kleusberg A. - Teunissen P J G Eds.: *GPS for Geodesy, Lecture Notes in EarthSciences*, Springer-Verlag., 1996.

Seeber G.: *Satellite Geodesy. 2nd Edition.*, De Gruyter, 2003.

Strang G. - Borre K.: *Linear Algebra, Geodesy and GPS.*, Wellesley-Cambridge Press, UK., 1997.

Global Positioning System Precise Positioning System Performance Standard (February 2007): http://www.pnt.gov/public/docs/2007/ppsps2007.pdf

IGS Home page: http://igscb.jpl.nasa.gov/

EUREF EPN Home page: http://www.epncb.oma.be/

RTCM Home page: http://www.rtcm.org/

ESA Home page: http://www.esa.int/

Information-Analytical Centre - GLONASS: http://www.glonass-ianc.rsa.ru/