1. Introduction to GNSS

Richard B. Langley, Peter J.G. Teunissen, Oliver Montenbruck

This chapter is a primer on global navigation satellite systems (GNSSs). It assumes no prior knowledge of the systems or how they work. All of the key concepts of satellite-based positioning, navigation, and timing (PNT) are introduced with pointers to subsequent chapters for further details. The chapter begins with a history of PNT using satellites and then introduces the concept of positioning using measured ranges between a receiver and satellites. The basic observation equations are then described along with the associated error budgets. Subsequently, the various GNSSs now in operation and in development are briefly overviewed. The chapter concludes with a discussion of the relevance and importance of GNSS for science and society at large.

1.1 Early Satellite Navigation

We will introduce the basic concepts of the Navstar Global Positioning System (GPS) and the other global navigation satellite systems (GNSSs) in operation and under development but it will be helpful if we first view them in a historical perspective. Determining the positions of points on the Earth's surface using observations of distant objects has been carried out for hundreds of years. Reflecting mirrors on mountaintops gave way to using high-altitude flares and rockets. And, of course, celestial navigation using observations of the Sun, stars, and planets has been used for centuries. However, it was only with the dawning of the space age that it became possible to develop a global system for high accuracy positioning and navigation.

Sometimes we refer to these systems as space-based systems. They can be broadly classified into optical techniques and radio techniques. Both kinds of system were pioneered in the late 1950s and 1960s.

Optical techniques are those techniques that utilize the visible part of the electromagnetic spectrum and in addition to astronomical positioning using a theodolite or sextant, include ground-based imaging of orbiting satellites and satellite laser ranging (SLR). Although still an important source of information for satellite orbit determination and surveillance, imaging of satellites against background stars for geodetic positioning has been superseded by other techniques. However, SLR still plays a prominent role in geodetic positioning and celestial mechanics.

Several radio systems were developed for satellite tracking and orbit determination. In the United States, these systems included radar, the Goddard Space Flight Center Range and Range Rate (GRARR) system, and NASA's Minitrack system [1.1]. In addition to their role in orbit determination, the systems were utilized for tracking camera calibration and directly for geode-

tic positioning. The US Army's Sequential Collation of Range (SECOR) system, on the other hand, was developed specifically for positioning purposes.

But the most successful satellite-based positioning system and one that overlapped with the development of GPS, was Transit [1.2, 3]. Also known as the US Navy Navigation Satellite System, Transit was the world's first satellite-based positioning system to operate globally. The system evolved from the efforts to track the Soviet Union's Sputnik I, the first artificial Earth-orbiting satellite. By measuring the Doppler frequency shift of the 20 MHz radio signals received from the satellite at a known location, the orbit of the satellite could be worked out. And shortly thereafter, researchers determined that if the orbit of a satellite was known, then the position of a receiver could be determined from the shift. That realization led to the development of Transit, with the first experimental satellite being launched in 1959. Initially classified, the system was made available to civilians in 1967 and was widely used for navigation and precise positioning until it was shut down in 1996. The Soviet Union developed a similar system called Tsikada and a special military version called Parus [1.4, 5]. These systems are also assumed to be no longer in use – at least for navigation.

A series of Transit prototype and research satellites was launched between 1959 and 1964 with the first fully operational satellite, Transit 5-BN-2, launched on 5 December 1963. The first Oscar-class Transit satellite (NNS O-1, Fig. 1.1), was brought into orbit on 6
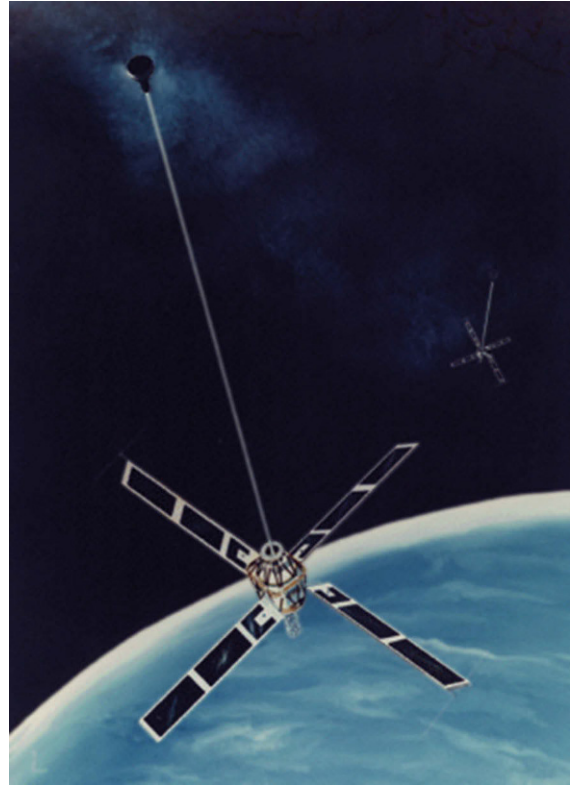

Fig. 1.1 US Navy Transit navigation satellite of the *Oscar* series (named after the phonetic code word for the letter *O*, or *operational*) (courtesy of US Navy)
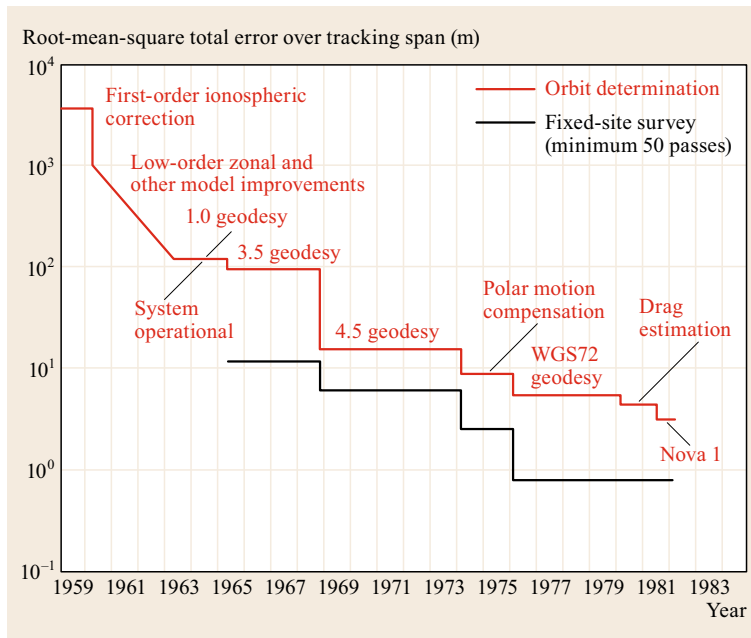

Fig. 1.2 Accuracy improvements over time with US Navy Transit satellites (after [1.3])

October 1964 and 24 operational satellites were subsequently launched. The last pair of Transit satellites, NNS O-25 and O-31, was launched on 25 August 1988.

Transit navigation required the measurements of the satellite signal's Doppler shift for a complete pass that could take up to about 18 min from horizon to horizon. At the conclusion of the pass, the latitude and longitude of the receiver, the position fix, could be determined. With five operational satellites, the mean time between fixes at a mid-latitude site was around 1 h. Eventually, as the orbits of the satellites became better determined, two-dimensional (2-D) position fix accuracies of several tens of meters were possible from a single satellite pass. By recording data from a number of passes over a few days from a fixed site on land, three-dimensional (3-D) accuracies better than one meter were possible and Doppler-based control points for mapping were established in many countries and the Canadian north, in particular, saw significant use of Transit for geodetic purposes.

The Transit satellites used signals on two different frequencies (150/400 MHz) to cancel out ionospheric delays, a concept that was later inherited by GPS and the other GNSSs. Besides its main use as a navigation system, Transit also provided early contributions to geodesy and helped to establish a new global reference frame (Fig. 1.2).

Transit was decommissioned at the end of 1996 with the advent of GPS and its superior performance. And the equivalent Russian satellite Doppler systems have essentially been replaced by the Global'naya Navigatsionnaya Sputnikova Sistema (Russian Global Navigation Satellite System, GLONASS), the second fully operational GNSS. These new systems were based on the concept of range measurements rather than Doppler observations and used a different constellation design offering continuous coverage. These new concepts enabled a notable increase in accuracy as well as instantaneous positioning around the globe.

1.2 Concept of GNSS Positioning

1.2.1 Ranging Measurements

GNSS signals are electromagnetic waves propagating at the speed of light. Signal frequencies in the radio spectrum between about 1.2 and 1.6 GHz (a part of the so-called L-band) have been selected for these signals since these enable measurements of adequate precision, allow for reasonably simple user equipment and do not suffer from attenuation in the atmosphere under common weather conditions. At the given frequencies, GNSS signals have a wavelength of about 19–25 cm. Similar to early radio navigation systems such as Transit, GNSSs provide signals on at least two different frequencies for compensation of ionospheric delays in their measurements.

A distinct feature of all GNSS signals is the modulation of the harmonic radio wave (termed the *carrier*) with a characteristic pseudorandom noise (PRN) code. This code is essentially a binary sequence of zeros and ones with no obvious pattern or regularity. The sequence is transmitted at a rate of typically 1–10 MHz, where higher rates imply a higher processing effort but promise more precise measurements. The PRN code is continuously repeated at intervals of a few milliseconds to seconds and facilitates measurements of the signal transmission time. In most GNSSs, the PRN sequence also serves as a unique fingerprint, which allows the receiver to distinguish individual satellites transmitting on the same frequency.

On top of the ranging code, the signal is also modulated with a low rate (e.g., 50 bits/s) navigation data stream (known as the broadcast navigation message) that provides information on the orbit of the transmitting satellite and the offset of its local clock from the GNSS system time.

The basic measurement made by a GNSS receiver is the time τ required for the GNSS signal to propagate from a satellite to the receiver. This can be obtained by tracking the PRN code modulation of the signal as illustrated in Fig. 1.3. Within the receiver, a local copy of the PRN sequence is generated, which is continuously compared and aligned with the signal received
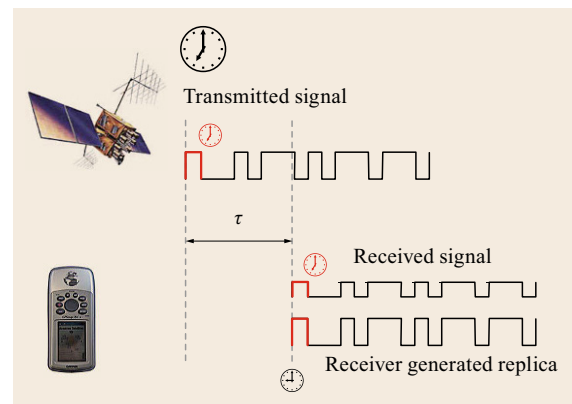

Fig. 1.3 Basic principle of pseudorange measurements

from the satellite. This *tracking loop* provides continuous measurements of the instantaneous code phase and hence the transmission time corresponding to the currently received signal (Chap. 13). By comparing this time with the local receiver time, the signal propagation time, and – upon multiplication by the speed of light – the distance or range from receiver to satellite are obtained.

Overall, the GNSS signals enable three basic types of measurements:

- *Pseudorange*: A measure of the difference between the receiver clock at signal reception and the satellite clock at signal transmission (scaled by the speed of light). Except for the asychronicity of the two clocks and some other delays, the pseudorange measures the satellite–receiver distance, the precision of which is in the dm-range.
- *Carrier phase*: A measure of the instantaneous beat phase and the accumulated number of zero-crossings obtained after mixing with a reference signal of the nominal frequency. Changes in carrier phase over time reflect the change in (pseudo)range but are substantially (≈ 2 orders) more precise. In case of interrupted tracking the accumulated cycle count is lost and the carrier-phase measurements exhibit a cycle slip.
- *Doppler*: The change in the received frequency caused by the Doppler effect is a measure of the range-rate or line-of-sight velocity.

Pseudorange, carrier-phase, and Doppler observations provide the basic measurements for computing position and velocity as well as the offset of the receiver time with respect to the GNSS system time scale.

They are complemented by information on the orbit and clock offsets of the individual GNSS satellites, which is transmitted as part of the broadcast navigation message and allows the receiver to compute the position and velocity of the transmitting satellite at the signal transmission time. To provide such information with adequate accuracy, the GNSS operator must be able to determine and to predict the satellite orbit (Chap. 3) ahead of time, so that it can be uploaded to the satellite for subsequent broadcasting to the users. Likewise GNSS relies on highly stable onboard clocks, whose time offsets can be accurately predicted. Rubidium or cesium atomic frequency standards or even hydrogen masers are used for this purpose (Chap. 5), which deviate by only 10^{-13} to 10^{-14} from their nominal frequency over time scales of a day.

Before addressing pseudorange- and carrier-phase-based positioning in more detail, we first discuss the basic principles of range-based positioning using distance measurements.

1.2.2 Range–Based Positioning

As mentioned, the basic measurement made by a GNSS receiver is the time τ_r^s required for the GNSS signal to propagate from a satellite antenna s to a receiver antenna r. Since the signal travels at the speed of light, c, this time interval can be converted to a distance or range, simply by multiplying it by c

$$\rho_r^s(t) = c\tau_r^s \, . \tag{1.1}$$

Let us assume that the clock in the receiver is synchronized with the clock in the satellite, and that the atmosphere (ionosphere and troposphere), which slightly delays the arrival of the signal and about which we will talk later, does not exist. Furthermore, let us assume there is no measurement noise; that is, no random perturbation to the measurement, something that invariably affects all measurements to a greater or lesser degree. Under these ideal and simplified circumstances, the observation equation for the observed range takes the form

$$\begin{aligned}\rho_r^s(t) =& ||\boldsymbol{r}_r(t) - \boldsymbol{r}^s(t-\tau)|| \\ =& \Big[(x_r(t) - x^s(t-\tau))^2 + (y_r(t) - y^s(t-\tau))^2 \\ &+ (z_r(t) - z^s(t-\tau))^2 \Big]^{-\frac{1}{2}} , \end{aligned} \tag{1.2}$$

with $\boldsymbol{r}_r = (x_r, y_r, z_r)^\top$ being the unknown position vector of the receiver antenna (possibly moving and therefore a function of time) and $\boldsymbol{r}^s = (x^s, y^s, z^s)^\top$ that of the satellite (known from the navigation message transmitted by the satellite). Typically, both vectors are referred to an Earth-centered, Earth-fixed (ECEF) coordinate frame. Examples of such frames include versions of the World Geodetic System (WGS) 84 and the International Terrestrial Reference Frame (ITRF) (Chaps. 2 and 36).

With a single range measurement (1.2), we would know that the position of the receiver antenna must lie somewhere on a sphere, centered on the satellite, with a radius equal to the measured range; call it ρ_r^1. If we simultaneously make a range measurement to a second satellite, then our receiver must also lie on a sphere, of radius ρ_r^2, centered on this satellite. The two spheres will intersect, with the loci of intersection points forming a circle. Our receiver must lie somewhere on this circle, which is therefore called a line of position. A third simultaneous range measurement, ρ_r^3, gives us a third sphere, which intersects the other two at just two points. One of these points can be immediately dismissed as being the location of our receiver, since it lies far out in space. So, the simultaneous measurement of the ranges to three satellites is sufficient to determine

a position fix in three dimensions – at least in principle (Fig. 1.4).

Computationally, the receiver position solution $(x_r, y_r, z_r)^\top$ of the simultaneous range equations

$$\begin{aligned}
\rho_r^1 &= \sqrt{(x_r - x^1)^2 + (y_r - y^1)^2 + (z_r - z^1)^2} \\
\rho_r^2 &= \sqrt{(x_r - x^2)^2 + (y_r - y^2)^2 + (z_r - z^2)^2} \\
\rho_r^3 &= \sqrt{(x_r - x^3)^2 + (y_r - y^3)^2 + (z_r - z^3)^2}
\end{aligned}$$ (1.3)

is usually obtained through an iterative linearization approach. Switching to vector formalism $\boldsymbol{p} = [\rho_r^1, \rho_r^2, \rho_r^3]^\top$ and dropping the indices, the system (1.3) of three nonlinear range equations can be approximated to the first order as

$$\boldsymbol{p} = \boldsymbol{p}_0 + \mathbf{A}\Delta\boldsymbol{x} ,$$ (1.4)

where $\boldsymbol{p}_0$ is the vector of computed range values based on given satellite coordinates (x^i, y^i, z^i), $i = 1, 2, 3$, and an initial estimate $\boldsymbol{x}_0 = (x_{r,0}, y_{r,0}, z_{r,0})^\top$ of the receiver's position. Furthermore,

$$\mathbf{A} = \begin{pmatrix}
\dfrac{\partial \rho_r^1}{\partial x_r} & \dfrac{\partial \rho_r^1}{\partial y_r} & \dfrac{\partial \rho_r^1}{\partial z_r} \\[2mm]
\dfrac{\partial \rho_r^2}{\partial x_r} & \dfrac{\partial \rho_r^2}{\partial y_r} & \dfrac{\partial \rho_r^2}{\partial z_r} \\[2mm]
\dfrac{\partial \rho_r^3}{\partial x_r} & \dfrac{\partial \rho_r^3}{\partial y_r} & \dfrac{\partial \rho_r^3}{\partial z_r}
\end{pmatrix}$$ (1.5)

with $\partial p_r^i / \partial x_r = (x_{r,0} - x^i) / \rho_{r,0}^i$ $(i = 1, 2, 3)$, is the design matrix, and $\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$ is the increment to the initial vector of receiver coordinates that is to be determined. Note that the matrix $\mathbf{A}$ reflects the relative geometry of the satellites and the receiver. Solving for $\Delta\boldsymbol{x}$, we have

$$\Delta\boldsymbol{x} = \mathbf{A}^{-1}(\boldsymbol{p} - \boldsymbol{p}_0) = \mathbf{A}^{-1}\Delta\boldsymbol{p}$$ (1.6)
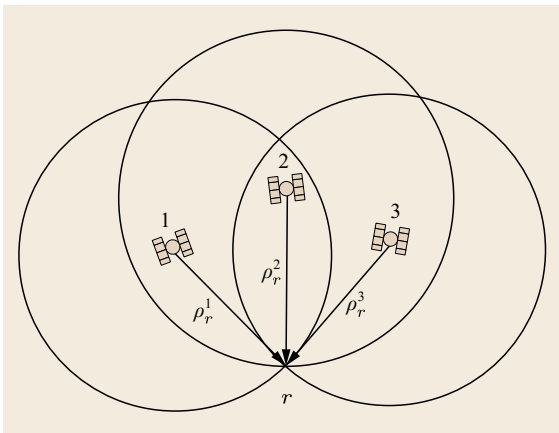
Fig. 1.4 Positioning through intersecting spheres

and then

$$\boldsymbol{x} = \boldsymbol{x}_0 + \Delta\boldsymbol{x} .$$ (1.7)

Depending on the closeness of the approximate ranges, $\boldsymbol{p}_0$, to the measurements, $\boldsymbol{p}$, several iterations are, in general, required to arrive at final values for $\boldsymbol{x}$.

1.2.3 Pseudorange Positioning

So far we assumed that the clock in the GNSS receiver was synchronized with the clocks in the satellites. This assumption, however, is fallacious. When a GNSS receiver is switched on, its clock will in general be mis-synchronized with respect to the satellite clocks, by an unknown amount. Furthermore, the clocks in the satellites are synchronized with each other and to a master time scale, called the system time, only to within about a millisecond. The range measurements the receiver makes are biased by the receiver and satellite clock errors, dt_r and dt^s, and are therefore referred to as *pseudo*ranges

$$p_r^s = \rho_r^s + c(dt_r - dt^s) .$$ (1.8)

A timing error of a millisecond would result in an error in position of about 300 km, clearly an intolerable amount. It would be possible to better synchronize the satellite clocks by frequently sending them adjustment commands from the ground, but it has been found that clocks actually keep better time if they are left alone and the readings of the clock corrected. The GNSS operators monitor the satellite clocks and determine the offsets and drifts with respect to system time. These parameters are subsequently uploaded to the satellites and transmitted as part of a navigation message broadcast by the satellites. A GNSS receiver uses these satellite clock offset values to correct the measured pseudoranges.

However, we then still have the receiver clock error dt_r to deal with. Because of this error, the three spheres with radii equal to the measured pseudoranges corrected for the satellite clock offsets will not intersect at a common point. But, if the receiver clock error dt_r, can be determined, then the pseudoranges can be corrected and the position of the receiver determined. The situation, compressed into two dimensions, is illustrated in Fig. 1.5.

So now we actually have four unknown quantities or parameters in our pseudorange observation equation

$$p_r^s = \rho_r^s + c\,dt_r .$$ (1.9)

They are the three coordinates of the receiver antenna position (x_r, y_r, z_r) and the receiver clock offset dt_r.
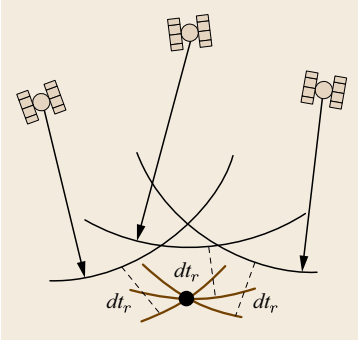
Fig. 1.5 Determination of receiver clock offset dt_r and true user position (intersection of *brown lines*) from the intersection of spheres centered on the satellites. Pseudoranges are shown by arcs of *black lines*

We thus need at least four simultaneous pseudoranges to estimate the three receiver coordinates and the receiver clock offset (measured in units of distance). With $x = (x_r, y_r, z_r, dt_r)^\top$ and the four-by-four design matrix

$$A = \begin{pmatrix} \dfrac{\partial \rho_r^1}{\partial x_r} & \dfrac{\partial \rho_r^1}{\partial y_r} & \dfrac{\partial \rho_r^1}{\partial z_r} & 1 \\ \dfrac{\partial \rho_r^2}{\partial x_r} & \dfrac{\partial \rho_r^2}{\partial y_r} & \dfrac{\partial \rho_r^2}{\partial z_r} & 1 \\ \dfrac{\partial \rho_r^3}{\partial x_r} & \dfrac{\partial \rho_r^3}{\partial y_r} & \dfrac{\partial \rho_r^3}{\partial z_r} & 1 \\ \dfrac{\partial \rho_r^4}{\partial x_r} & \dfrac{\partial \rho_r^4}{\partial y_r} & \dfrac{\partial \rho_r^4}{\partial z_r} & 1 \end{pmatrix} \tag{1.10}$$

the same iterative procedure as described before can then be applied.

What if signals from more than four satellites are available? Because of the unmodeled errors (e.g., atmospheric delays) as well as the residual errors in the modeled terms, it is beneficial to use simultaneous pseudoranges to all m available satellites for estimating the receiver coordinates and clock offset. This requires use of a nonlinear least-squares (or related Kalman filter) estimation procedure (Chap. 22)

$$\Delta x = (A^\top W A)^{-1} A^\top W \Delta p, \tag{1.11}$$

where A now has dimensions of $m \times 4$ and W is a weight matrix, which reflects the uncertainty in the observations and any correlations that may exist among them. This weight matrix may be written as

$$W = Q_{pp}^{-1}, \tag{1.12}$$

in which Q_{pp} is the covariance matrix of the pseudorange errors. In general, the solution of a nonlinear problem must be iterated to obtain the result. However, if the linearization point is sufficiently close to the true solution, then only one iteration is required (Chaps. 21 and 22). The processing of measurements

can take place in real time within the receiver or the raw measurements can be stored for postprocessing by a standalone computer.

1.2.4 Precision of Position Solutions

The precision with which the receiver's coordinates and clock offset can be determined is described by the covariance matrix of the solution Δx. This covariance matrix, denoted as Q_{xx}, follows, with (1.12), from applying the error propagation law, also known as the variance propagation law, to (1.11) as

$$\begin{aligned} Q_{xx} &= [(A^\top W A)^{-1} A^\top W] \\ &\quad \cdot Q_{pp} \cdot [(A^\top W A)^{-1} A^\top W]^\top \\ &= (A^\top Q_{pp}^{-1} A)^{-1}. \end{aligned} \tag{1.13}$$

The diagonal elements of this matrix are the estimated receiver coordinate and clock-offset variances, and the off-diagonal elements (the covariances) indicate the degree to which these estimates are correlated. This equation represents a fundamental relationship widely used for actual measurement analysis as well as for experiment and system design studies. It allows one to examine the effect a particular design (through design matrix A) or measurement capability (through measurement covariance matrix Q_{pp}) will have on specified parameters without actually making any measurements.

In GNSS-related studies, for example, we might use the equation to answer a variety of questions:

- What is the behavior of the estimated parameter covariance matrix as a function of particular satellite configurations?
- How do various model errors propagate into the receiver coordinates as a function of satellite configurations?
- What is the tolerance value that a particular model error should not exceed to achieve a desired positioning accuracy?

User Equivalent Range Error
Such analyses simplify if we assume that the measurement and residual model errors are uncorrelated and the same for all observations with a particular standard deviation (σ). Then $Q_{pp} = \sigma^2 I_m$ (in which I_m is the identity matrix of order m) and the expression for the covariance matrix of x simplifies to

$$Q_{xx} = \sigma^2 (A^\top A)^{-1}. \tag{1.14}$$

When we combine satellite clock and ephemeris error, atmospheric error, receiver noise, and multipath – all expressed in units of distance – we obtain a quantity

known as the total *user equivalent range error* (UERE), which we can use for σ. UERE can further be divided into two parts [1.6], namely:

- The *signal-in-space (user) range error* (SISRE or SISURE), which comprises errors related to the space and control segment (primarily broadcast satellite orbit and clock errors), and
- The *user equipment error* (UEE), which includes the remaining contributions specific to the user's receiver and environment.

The total UERE can then be written as

$$\text{UERE} = \sqrt{\text{SISRE}^2 + \text{UEE}^2} \, . \tag{1.15}$$

An overview of key contributions to the UERE budgets is provided in Table 1.1. The given values are mainly for illustration, since it is difficult to provide universally valid error bounds in most cases. Among others, UEE contributions may differ widely among individual receivers and sites. In the case of GPS single-frequency positioning, the total UERE is typically in the neighborhood of a few meters, with the actual value dominated by ionospheric and multipath effects [1.7]. Dual-frequency positioning, with the capability to remove almost all of the ionospheric delay from the pseudorange observations, can result in even smaller UEREs.

Dilution of Precision

A simple scalar indicator of the overall quality of the least-squares solution is given by the square root of the sum of the parameter estimate variances

$$\begin{aligned} \sigma_{\text{G}} &= \sqrt{\sigma_{\text{E}}^2 + \sigma_{\text{N}}^2 + \sigma_{\text{U}}^2 + \sigma_{dt}^2} \\ &= \sigma \, \text{tr} \{ (\mathbf{A}^\top \mathbf{A})^{-1} \} \, , \end{aligned} \tag{1.16}$$

Table 1.1 Representative magnitudes of individual contributions to the GNSS user equivalent range error (after [1.8–10]) for estimates of the individual contributions

Error source	Contribution 1σ (m)
SISRE	
Broadcast satellite orbit	0.2–1.0
Broadcast satellite clock	0.3–1.9
Broadcast group delays	0.0–0.2
UEE	
Unmodeled ionospheric delay	0–5
Unmodeled tropospheric delay	0.2
Multipath	0.2–1
Receiver noise	0.1–1
UERE	**0.5–6**

in which σ_{E}^2, σ_{N}^2, and σ_{U}^2 are the variances of the east, north, and up components of the receiver position estimate, respectively, and σ_{dt}^2, is the variance of the estimated receiver clock offset. If the solution algorithm is parameterized in terms of geocentric Cartesian coordinates, it is a straightforward procedure to transform the solution variance matrix to the local coordinate frame to get the north, east, and up components.

The elements of matrix $\mathbf{A}^\top \mathbf{A}$ are a function of the receiver–satellite geometry and as the trace of its inverse is typically greater than 1, it amplifies σ, or dilutes the precision, of the position determination. This scaling factor is therefore usually called the geometric dilution of precision (GDOP). The GDOP becomes the position-DOP (PDOP), if one leaves out the contribution of the receiver clock, and it further reduces to the horizontal-DOP (HDOP), if one also leaves out the contribution of the up-component. In a likewise manner one can obtain the vertical-DOP (VDOP).

It turns out that DOP values depend on the volume of the polyhedron formed by the tips of the receiver–satellite unit vectors. The larger the volume, the smaller the DOPs. If the tips lie in a plane, the DOP factors are infinitely large. In fact, no position solution is possible with this receiver–satellite geometry as the matrix $\mathbf{A}^\top \mathbf{A}$ is singular: the solution cannot distinguish between an error in the receiver clock and an error in the position of the receiver. DOP values are smaller and hence solution errors are smaller when the satellites used in computing the solution are spread out in the sky.

High DOP values can sometimes occur even for all-in-view receivers operating at mid-latitudes. In some environments, such as heavily forested areas or urban canyons, a GNSS receiver's antenna may not have a clear view of the whole sky because of obstructions. If it can only receive GNSS signals from a small region of the sky, the DOPs will be large, and position accuracy will suffer. Being able to track more satellites can help in such situations, and a multi-GNSS receiver may provide acceptable accuracies. New receiver technology permitting use of weaker GPS signals, even those present inside buildings, will also be beneficial.

While DOP and UERE are highly useful concepts to understand GNSS positioning errors and their dependence on the geometric distribution of tracked satellites as well as individual pseudorange error sources, readers should keep in mind that the common rule-of-thumb

$$\text{Navigation error} = \text{DOP} \times \text{UERE} \tag{1.17}$$

is only a rough approximation and limited to random error propagation. Efforts to better account for system-

atic and random errors and to arrive at a more realistic description of real postioning errors are, for example, presented in [1.11].

1.2.5 GNSS Observation Equations

So far, we have been working with a rather simplified version of the pseudorange observation equation (1.8) and (1.9). In actuality, however, there are a number of additional error sources that must be taken into consideration and modeled in the observation equation. We therefore have to modify the pseudorange observation equation as follows

$$p_{\mathrm{r}}^{\mathrm{s}} = \rho_{\mathrm{r}}^{\mathrm{s}} + c(dt_{\mathrm{r}} - dt^{\mathrm{s}}) + T_{\mathrm{r}}^{\mathrm{s}} + I_{\mathrm{r}}^{\mathrm{s}} + e_{\mathrm{r}}^{\mathrm{s}} \, . \qquad (1.18)$$

Here, dt_{r} and dt^{s} are the receiver and satellite clock offset from GNSS system time as before, $T_{\mathrm{r}}^{\mathrm{s}}$ is the neutral atmosphere (troposphere) propagation delay, $I_{\mathrm{r}}^{\mathrm{s}}$ is the ionospheric propagation delay, and $e_{\mathrm{r}}^{\mathrm{s}}$ represents unmodeled errors including receiver noise, multipath, and other small effects (Chaps. 13–15). For a more detailed discussion of the basic pseudorange observation equation, see Chap. 19.

The carrier-phase observation equation is similar to the pseudorange equation

$$\varphi_{\mathrm{r}}^{\mathrm{s}} = \rho_{\mathrm{r}}^{\mathrm{s}} + c(dt_{\mathrm{r}} - dt^{\mathrm{s}}) + T_{\mathrm{r}}^{\mathrm{s}} - I_{\mathrm{r}}^{\mathrm{s}} + \lambda M_{\mathrm{r}}^{\mathrm{s}} + \epsilon_{\mathrm{r}}^{\mathrm{s}} \, . \qquad (1.19)$$

In addition to the previously defined terms, λ is the carrier wavelength, $M_{\mathrm{r}}^{\mathrm{s}} = N_{\mathrm{r}}^{\mathrm{s}} + \delta_{\mathrm{r}} - \delta^{\mathrm{s}}$ is the sum of the integer carrier-phase ambiguity $N_{\mathrm{r}}^{\mathrm{s}}$ (in cycles) and the instrumental receiver and satellite phase delays $\delta_{\mathrm{r}} - \delta^{\mathrm{s}}$ (in cycles), $\epsilon_{\mathrm{r}}^{\mathrm{s}}$ represents unmodeled phase errors including receiver noise, multipath, and other small effects.

Several terms in the carrier-phase observation equation are nominally identical to those in the pseudorange observation equation, including the geometric range, receiver and satellite clock offsets, and the tropospheric propagation delay. The magnitude of the ionospheric term in both equations is the same, however its sign is negative in the carrier-phase equation. This relates to the fact that the phase of the carrier is advanced during the signal's passage through the ionosphere, as opposed to the pseudorange, which suffers a delay. And the ionospheric phase advance is frequency dependent like the pseudorange delay. For a more detailed discussion of the basic carrier-phase observation equation, see Chap. 19.

The receiver or postprocessing software will use the above forms of the observation equations to compute the receiver coordinates or any other relevant GNSS parameters. Models exist for describing the neutral atmosphere propagation delay, and the ionospheric propagation delay can be corrected using the coefficients of a model included in the broadcast navigation message or by combining simultaneous pseudorange or carrier-phase measurements made on two transmitted frequencies (Chaps. 38 and 39).

1.3 Modeling the Observations

To accurately determine the receiver coordinates or any other relevant GNSS parameters, we must model the right-hand sides of the observation equations (1.18) and (1.19) to match as accurately as possible the receiver's observations of the satellites. This requires knowledge of the satellites' positions at the time of signal transmission, the offset of their clocks, the atmospheric propagation delays, ambiguities for the carrier-phase observations, plus perhaps some smaller contributions such as instrumental delays. If the available information is not sufficiently accurate, we may be able to estimate residual effects from the observations themselves and so better match the right-hand and left-hand sides of the equations.

In the next few sections, we will overview some of the information required for a receiver or external software to process GNSS observations. More detailed descriptions are provided in subsequent chapters of the Handbook.

1.3.1 Satellite Orbit and Clock Information

The predicted position of a satellite's antenna phase center can be derived from the Keplerian-like orbit parameters given in the broadcast navigation message. The predicted offset of the active satellite clock from system time is also derived from the navigation message. The combined accuracy of these terms, SISRE, is typically $0.5-2\,\mathrm{m}$ for current global and regional navigation satellite systems. For GPS, the Global Positioning Systems Directorate has reported that the annual root-mean-square signal-in-space range error (SISRE) across all healthy satellites dropped from $1.6\,\mathrm{m}$ in 2001 to $0.7\,\mathrm{m}$ in 2014 [1.12] and further improvement has been made after replacing the old Block IIA satellites with the latest generation of Block IIF satellites in early 2016. SISRE values of about $0.5\,\mathrm{m}$ can also be expected for Galileo based on the performance of the preoperational constellation in the 2015/2016 timeframe. The

performance of GLONASS, the second fully operational GNSS, is presently about three times worse but likewise expected to improve with the ongoing modernization of the space and ground segments.

Significantly more accurate satellite orbit and clock information is available from various sources, such as satellite-based augmentations systems, the international GNSS service, and private service providers, for both real-time and postprocessing applications (Chaps. 3, 33, and 34).

1.3.2 Atmospheric Propagation Delay

GNSS signals travel through the Earth's atmosphere to receivers on or near the ground (Chap. 6). The signals are refracted, changing their velocity – both speed and direction of travel. Measured pseudoranges and carrier phases are biased by meters to tens of meters. The biases are determined by the integrated effect of the refractive index – the ratio of the speed of propagation of an electromagnetic wave in a vacuum to that in a medium – all along the signal raypath. In general, the refractive index will be a function of the characteristics of the medium including the type and densities of the medium's constituents as well as the frequency of the wave and possibly external factors such as ambient magnetic fields.

The atmospheric propagation delays can be separated into those due to the electrically neutral atmosphere and those due to the ionosphere (Fig. 1.6).
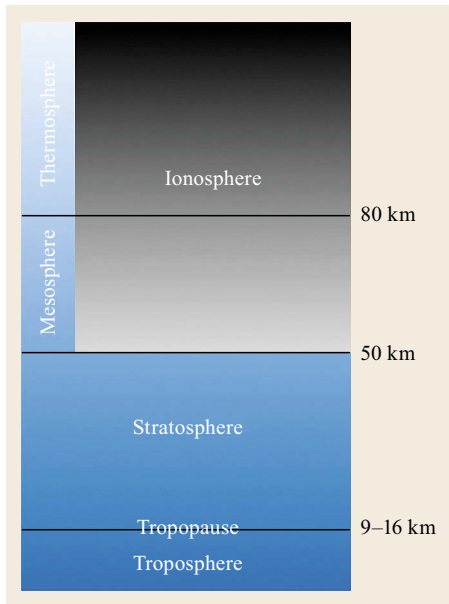

Fig. 1.6 Structure of the Earth's atmosphere

Neutral Atmosphere

The neutral atmosphere is that part of the atmosphere that is electrically neutral and stretches from ground level up to a height of 50 km and beyond (Chap. 6). It is what we colloquially refer to as the air. Air is made up of nitrogen, oxygen, carbon dioxide, as well as some other atoms and molecules including water vapor. With the inclusion of water vapor, we refer to the medium as moist air. The refractive index n of a parcel of moist air is a function of its temperature, the partial pressures P_d of the dry constituents (nitrogen, oxygen, etc.) and the partial pressure e of water vapor

$$n = n(T, P_d, e) . \tag{1.20}$$

Note that air is essentially a nondispersive medium, with n independent of frequency throughout most of the radio spectrum and including the frequencies used by all GNSSs. This also means that the effect on pseudoranges is identical to that on carrier-phase measurements.

At sea level, values of the refractive index of air are close to 1.0003, becoming smaller with increasing height. A more useful quantity is refractivity, $N = 10^6(n-1)$, with sea level values near 300. Since the bulk of the neutral atmospheric effect occurs in the lowestmost part of the atmosphere – the troposphere – the effect is often termed the tropospheric propagation delay.

How much delay is imparted to a GNSS signal by the neutral atmosphere? This depends on the location and height of the receiver as well as weather conditions and also the elevation angle (and, to a lesser degree, azimuth) at which the signal arrives at the receiver. The total delay experienced by the received signal is referred to as the slant delay. It is modeled by mapping the delay of a hypothetical signal arriving from directly overhead – the zenith direction – to the actual signal slant path using a mapping function (also called an obliquity factor). Typically, at sea level, zenith delay is about 2.4 m, rising to more than 24 m at an elevation angle of 5°.

Various neutral atmosphere delay models exist. Many receivers use the Radio Technical Commission for Aeronautics (RTCA) MOPS (Minimum Operational Performance Standards model) [1.13] (a slightly simplified version of the University of New Brunswick's UNB3 model, the predecessor to UNB3m [1.14, 15]) mandated for use by satellite-based augmentation system (SBAS) user equipment. Several more sophisticated models exist (Chap. 6). No model is perfect, though, and there is some advantage in estimating residual delays from the GNSS data itself.
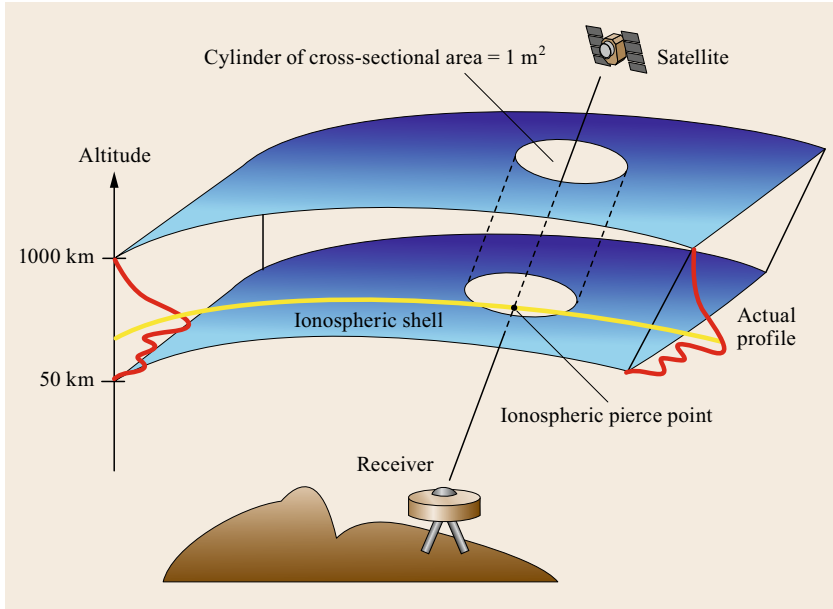
Ionosphere

The ionosphere is that region of the Earth's atmosphere in which ionizing radiation (principally from solar extreme ultraviolet (EUV) and x-ray emissions) cause electrons to exist in sufficient quantities to affect the propagation of radio waves (Sect. 6.3). It extends from about 50–1000 km or more, above which we have the plasmasphere (also known as the protonosphere).

The ionosphere is a dispersive medium for radio waves: the refractive index is a function of frequency. It also means that pseudorange and carrier-phase observations are affected differently, with a phase refractive index and a pseudorange (group delay) refractive index. This results in an ionospheric delay for pseudoranges (increased value) and an ionospheric phase advance for carrier-phase measurements (decreased value). In addition to the radio frequency of the signal, the refractive indices are a function of electron density. The Earth's magnetic field also plays a minor role.

Again, it is the integrated effect of ionospheric refractive index all along the raypath that determines the pseudorange delay and the carrier-phase advance. It turns out that, to first order, the magnitudes of the pseudorange delay and the carrier-phase advance are the same; they just differ in sign. To an excellent approximation, the magnitude, I, in m is given by

$$I = 40.3 \frac{\text{TEC}}{f^2} , \qquad (1.21)$$

where TEC is the total electron content and is the total number of electrons in a column of one-meter-square cross-section centered on the signal raypath and stretching from the receiver to the satellite. In this equation, TEC is given in electrons/m^2, and the frequency, f, is given in Hz. Typical TEC values measured near the Earth's surface range from about 10^{16} to 10^{19} with the actual value depending on geographic location, local time, season, solar EUV flux, and magnetic activity. And, as with neutral atmosphere delays, we have slant and zenith delays, corresponding to slant and zenith TEC.

Since the ionospheric effect is to a very good approximation inversely proportional to the square of the frequency, by linearly combining simultaneous measurements (either pseudoranges or carrier phases) on two frequencies such as the GPS L1 and L2 frequencies, an observable virtually free of ionospheric effects can be constructed and used for position determinations (Sect. 20.2.3). This approach does require, however, a dual- or multifrequency receiver.

If only single-frequency observations are available, then the sum of the pseudorange and phase measurement can be taken to eliminate the ionospheric delay, or alternatively a model can be used to account for the ionospheric biases as much as possible. The GPS navigation message includes values of the parameters of a simple ionospheric model known as the broadcast or *Klobuchar* model [1.17], named after its developer *Jack Klobuchar*. This model permits an estimate of the zenith ionospheric delay to be computed at a receiver's location at a particular time of day and is driven by recent solar conditions as interpreted by the GPS control

segment. The zenith delay is then mapped into a slant delay assuming a thin shell model for the ionosphere, where all of the electron content is assumed to occur in a shell at a particular height above the Earth's surface (Fig. 1.7). The broadcast model assumes a shell height of 350 km.

A Klobuchar-like model is also used within the BeiDou system, while a version of the NeQuick model [1.18] has been selected as an alternative for the Galileo system. NeQuick describes the 3-dimensional electron density distribution as a function of a small set of ionospheric activity parameters provided in the navigation message and an extensive set of static, seasonal coefficients. The ionospheric slant delay is then obtained by integrating the electron density along the ray path between satellite and receiver. While computationally more demanding, the NeQuick model requires a smaller set of broadcast parameters and achieves an improved overall correction performance compared to the GPS Klobuchar model [1.10].

1.4 Positioning Modes

There are a variety of GNSS positioning (and navigation) modes of differing degrees of complexity and precision and accuracy. These range from the standard single-frequency pseudorange-based approach used by most consumer receivers including those in mobile phones to high-integrity methods for safety-of-life applications to sophisticated multifrequency carrier-phase-based techniques capable of centimeter to subcentimeter accuracies for demanding applications like machine control and scientific studies. In this section, we will briefly outline some of the approaches with deeper discussions appearing in later chapters (Chaps. 21, 23, 26, and 35).

1.4.1 Precise Point Positioning

Precise point positioning (PPP) is an advanced version of the single-point positioning (SPP) technique that we discussed earlier. PPP (Chap. 25) uses carrier-phase measurements as the primary observable with pseudorange measurements playing a secondary role. The PPP processing algorithm is similar to that for pseudorange-measurement positioning, except that effects down to the centimeter-level and lower must be modeled or estimated. This means that satellite constellation precise orbits and clock offsets (Chap. 34) must be used such as those provided by the International GNSS Service (IGS). Typically, a dual-frequency GNSS receiver is used with dual-frequency code and phase measurements linearly combined to remove first-order ionospheric effects. The carrier-phase ambiguities are estimated (resolving them to integer values if possible) as well as residual tropospheric propagation delay after applying an a priori model. Subtle effects, such as Earth tides, ocean tide loading, satellite and receiver antenna offsets, and carrier-phase windup are also modeled.

The performance of PPP can be measured in terms of accuracy, precision, convergence period (the time required for a position solution to converge below a certain accuracy threshold), availability, and integrity. Best effort PPP is generally bias free, so there is little difference between accuracy and precision statistics. PPP can provide few-centimeter-level 1-sigma accuracies in each coordinate (north, east, and up) for a static site after convergence whereas decimeter-level accuracies can be achieved for a moving platform or a static site with data processed in *kinematic mode* (independent epoch-by-epoch position fixes). Accuracies, although good, may not be sufficient for some applications. The convergence period for achieving a decimeter-level solution is typically up to about 30 min under normal conditions with some newer procedures (multifrequency, multiconstellation) achieving convergence periods of 20 min or less. Continuous availability of accurate PPP fixes depends on the environment. Signal blockages by trees and buildings can reduce availability but multiconstellation observations can be a big help in this regard. Integrity measures for PPP are currently limited.

PPP can be used for processing data from either static (stationary) sites or kinematic (moving) platforms
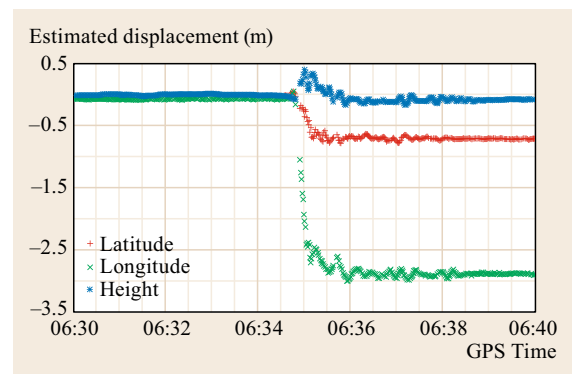

Fig. 1.8 Estimate of co-seismic displacement at IGS station CONZ following the 8.8 magnitude Chilean earthquake of 27 February 2010 (after [1.19])

and its uses include establishing and updating reference-station coordinates for crustal-deformation monitoring (Fig. 1.8), precise orbit determination of low-Earth-orbiting satellites, ocean buoy positioning for tsunami detection with main commercial applications in precision farming, seafloor mapping, marine construction, and airborne mapping. Application of PPP is expanding to atmosphere remote sensing, precise time transfer, land surveying, construction, and military uses.

1.4.2 Code Differential Positioning

The advantage of differential positioning (Chap. 26) over SPP is that with differential techniques certain effects are eliminated or largely reduced (e.g., orbit errors and atmospheric delays with dependence on the spatial correlation). There are two basic kinds of code differential positioning: measurement domain (typically covering a local/regional area and known as differential GPS/GNSS or DGPS/DGNSS) and state-space domain (typically covering a wide area and known as wide-area GPS/GNSS or WADGPS/WADGNSS). The best example of a measurement-domain DGPS is that implemented by coast guard agencies around the world for maritime navigation (Sect. 29.4). The best example of state-space-domain DGPS is that of SBASs, the first of which was the US Federal Aviation Administration's (FAA's) Wide Area Augmentation System (WAAS). SBAS is discussed in Sect. 1.5 and in detail in Chap. 12.

The measurement domain techniques provide composite measurement corrections to the user without estimating individual error components whereas state-space domain techniques provide individual error corrections such as satellite orbit and clock and ionospheric propagation delay. These corrections are determined at reference stations and transmitted to users using radio beacons.

The combined corrections account for navigation message satellite orbit and clock error, tropospheric propagation delay, ionospheric propagation delay and, in the past, GPS Selective Availability. A user requires a GNSS receiver with an integrated beacon receiver or a separate beacon receiver connected to the GNSS receiver by a serial communications link such as RS-232. The corrections are datum dependent. In North America, user-computed positions will be in the North American Datum (NAD) 83 system, not WGS 84.

Position accuracy generally degrades with increasing distance from the beacon transmitter site. Official accuracy is stated as $10\,m$ (horizontal at 95%) within the coverage area but typically, the error of a DGPS position is $1-3\,m$. Accuracy will be affected by user multipath and DOP. The error is often seen as a bias in positioning, resulting in a position offset. The scatter of the coordinates is likely to remain close to constant. A general rule of thumb is an additional $1\,m$ error per $100\,km$. However, accuracy is worse during strong ionospheric disturbances due to gradients.

1.4.3 Differential Carrier Phase

Differential carrier-phase positioning is a classic technique dating to the early 1980s. The procedure combines data from one (or more) reference stations with user data (Chap. 26). Observations on the same satellite at the same epoch are differenced between receivers (single difference) and then single differences are differenced between pairs of satellites (double difference). This procedure eliminates residual satellite and receiver clock errors, and reduces satellite orbit error and atmospheric propagation delay errors (Chap. 20). Accuracies at the decimeter level and better can be obtained. In principle, a similar approach could be taken using pseudoranges (alone) but with much lower resulting accuracies. See Chap. 26 for a more detailed description.

Single Differencing

Let us take a look at the differencing operations in detail starting with single differencing. Differencing carrier-phase measurements on one satellite s from two receivers, 1 and 2, gives us

$$\varphi_{12}^s = \rho_{12}^s + cdt_{12} + T_{12}^s - I_{12}^s + \lambda M_{12}^s + \epsilon_{12}^s , \qquad (1.22)$$

where $\varphi_{12}^s = \varphi_2^s - \varphi_1^s$ is the between-receiver single-difference carrier phase, with a similar notational convention for the terms on the right-hand side (1.22). Note that the satellite clock error has disappeared because it is identical for both receivers. Likewise, the satellite phase bias disappeared from $M_{12}^s = \delta_{12} + N_{12}^s$.

Differencing measurements made on two satellites, s and t, with receiver 1 gives us

$$\varphi_1^{st} = \rho_1^{st} + cdt^{st} + T_1^{st} - I_1^{st} + \lambda M_1^{st} + \epsilon_1^{st} , \qquad (1.23)$$

where $\varphi_1^{st} = \varphi_1^t - \varphi_1^s$ is the between-satellite single-difference carrier phase. Note that the receiver clock error has disappeared because it is identical for measurements made on all satellites at the same epoch. Likewise, the receiver phase bias disappeared from $M_1^{st} = \delta^{st} + N_1^{st}$.

Double Differencing

If we difference carrier-phase measurements between receivers and then difference the resulting values between satellites, we have the double-difference observ-

able

$$\varphi_{12}^{st} = \rho_{12}^{st} + T_{12}^{st} - I_{12}^{st} + \lambda N_{12}^{st} + \epsilon_{12}^{st} , \qquad (1.24)$$

where $\varphi_{12}^{st} = \varphi_{12}^{st} - \varphi_1^{st} = \varphi_{12}^{t} - \varphi_{12}^{s}$. Note that both receiver and satellite clock error have disappeared as well as the corresponding instrumental phase biases. As a result, the real valued parameters, M_{12}^s in (1.22) and M_1^{st} in (1.23), have been replaced by the integer parameter N_{12}^{st} in (1.24).

In processing double differences (between a pair of receivers and a pair of satellites), we must model the double-difference geometric range, the double-difference tropospheric delay, the double-difference ionospheric effect (which can be neglected in case of sufficiently short baselines), and the double-difference phase ambiguity, here now an integer. Least squares or a Kalman filter approach is used to estimate the coordinates of the user receiver along with the *nuisance* parameters in a similar approach to that used for processing undifferenced pseudorange or carrier-phase observations. Redundancy in the system of observation equations is used to check the validity of the assumed models (Chaps. 22 and 24).

The integer ambiguities N_{12}^{st} for all satellite pairs in an observation set are estimated together with the coordinates of the user receiver and (optionally) a residual tropospheric delay error. The ambiguities are initially estimated as floating-point numbers (the so-called *float solution*); procedures are available for fixing some or all of the ambiguities at their correct integer values (the so-called *fixed solution*) (Chap. 23). Successful ambiguity resolution depends on several factors including baseline length (shorter is better), the number of satellites in view (more is better), continuous tracking of satellites, low dilution of precision values, the degree of multipath (less is better), the number of frequencies observed (two are better than one), and length of observing session (longer is better). With modern receivers and techniques, ambiguity resolution can be carried out with just a few tens of seconds of observations, even if the user receiver is in motion (*on the fly*). Fixed solutions generally provide more accurate results. A common ambiguity fixing technique is LAMBDA (Least-squares Ambiguity Decorrelation Adjustment) [1.20] (Chap. 23).

Real-Time Kinematic Positioning

In real-time kinematic (RTK) positioning, a GNSS reference station transmits carrier-phase and pseudorange data over a radio link to a roving station. Either single- or dual-frequency GNSS receivers can be used, with the dual-frequency systems typically affording faster ambiguity resolution and higher positioning accuracies over

longer distances. The receivers must incorporate data radios (or be wired to external radios), typically operating in the very high frequency (VHF, 30–300 MHz) or ultrahigh frequency (UHF, 300 MHz and 3 GHz) parts of the radio spectrum. The reference station transmits pseudorange and carrier-phase measurements and ancillary data. Radio Technical Commission for Maritime Services (RTCM) SC-104 2.x or 3.x [1.21] data protocols are typically used although proprietary data formats also exist.

Transmission modes vary and include narrowband frequency modulation (FM) with frequency-shift-keying and packetized data transmission. Both 2 and 35 W transmitters are commonly available and for these licensing is typically required. However, license-free, low-power transmitters can also be used. In any case, VHF/UHF data links are limited to line of sight and transmitting and receiving antennas should be as high as possible. The maximum theoretical range is given by [1.22, 23]

$$d(\text{km}) = 3.57\sqrt{k}\left[\sqrt{h_t(\text{m})} + \sqrt{h_r(\text{m})}\right]. \qquad (1.25)$$

with h_t and h_r the height of transmitter and receiver, respectively, and where k varies with refractivity (typically between 1.2 and 1.6); for example, for a transmitting antenna at 30 m above the terrain and a receiving antenna at 2 m, the maximum range is 28 km. Any obstructions along the propagation path will affect the signal's range. Whether or not a received signal can be successfully used depends on several factors including receiver sensitivity.

Reference station data can also be transmitted via the Internet using, e.g., the Network Transportation of RTCM Internet Protocol (NTRIP; [1.24]) and accessed by a hardwire or wireless link such as a mobile phone.

To reduce latency effects, transmission protocols typically use data differences with reconstruction of reference station data at the rover. Cycle slips must be detected and corrected in real time and ambiguities must be resolved quickly, even if the receiver is in motion (*on the fly*). Several techniques exist. Use of GLONASS data in addition to GPS data can result in greater availability, faster ambiguity fixing, and higher positioning accuracies.

PPP–RTK, a combination or merging of PPP with state-space RTK can have significant advantages in ambiguity resolution, in convergence time, and in accuracy and practical systems have been implemented by industry and research organizations (Chaps. 25 and 26). All individual GNSS error components, derived from the RTK monitoring network, are determined and delivered using state-space representation (SSR). These include orbits, clocks, code (pseudorange) biases, ionosphere

(for single-frequency receivers), troposphere, and carrier-phase biases. In principle, the concept can be applied to small, regional, and global networks. RTCM and IGS have active committees developing standards for PPP–RTK and real-time PPP for delivering prototype research and commercial services.

1.5 Current and Developing GNSSs

GNSSs and regional navigation satellite systems (RNSSs) commonly consist of three components:

- The *space segment* comprises a constellation of satellites orbiting above the Earth's surface that transmit ranging signals on at least two frequencies in the microwave part of the radio spectrum.
- The *control segment* is responsible for maintaining the health of the system by monitoring the broadcast signals and computing and uploading to the satellites required navigation data. It consists of a group of globally (or locally)-dispersed monitoring stations, ground antennas for communicating with the satellites, and a master control station with a backup facility at a different location.
- The *user segment* consists of GNSS receiving equipment both civil and military. This includes receivers on the ground, at sea, in the air, and even in space.

GNSS constellations typically adopt a specific orbital configuration: MEO (medium-altitude Earth orbit) satellites for global coverage; IGSOs (inclined geosynchronous orbits) and GEOs (geostationary orbits) as supplements in regional systems.

MEO satellites are often evenly distributed in inclined near-circular orbits overequally spaced orbital planes, forming a constellation known as a *Walker* constellation [1.25]. The geometry of a specific Walker constellation is described by the triplet $t/p/f$, where t denotes the total number of satellites, p the number of equally spaced planes, and f the phase difference between the adjacent orbital planes. To determine the angle between satellites in adjacent planes, the parameter f should be multiplied by $360°/t$. A simple Walker constellation of eight satellites in two orbital planes is illustrated in Fig. 1.9.

Currently, there are six GNSSs/RNSSs in operation. The four GNSSs are: GPS (US), GLONASS (Russia), BeiDou (China), and Galileo (EU); and the two RNSSs: QZSS (Japan) and IRNSS/NavIC (India). For an overview summary, see Table 1.2.

1.5.1 Global Navigation Satellite Systems

GPS

The Global Positioning System (GPS; Chap. 7) is the US GNSS, which provides free positioning and timing services worldwide. It was originally developed for the US military and was made free for civilian purposes very early in the experimental phase of GPS. The launch of the first Block I Navstar GPS satellite occurred on 22 February 1978, followed by the declaration of the Initial Operating Capability in December 1993 with 24 operational satellites in orbit, and the Full Operational Capability in June 1995. GPS is maintained by the US government and is freely accessible by anyone with a GPS receiver. GPS provides two different positioning services: the Precise Positioning Service (PPS) on the GPS L1 (1575.42 MHz) and L2 (1227.6 MHz) frequencies both containing an encrypted precision (P) code ranging signal (known as the Y-code) with a navigation data message for authorized users, and the Standard Positioning Service (SPS) on the GPS L1 frequency containing a coarse/acquisition (C/A) code and a navigation data message for civilian users. The GPS modernization program began in 2005 with the launch of the first IIR-M satellite. Since that moment on, two new signals have been transmitted: L2C for civilian users and a new military signal (M-code) at the L1 and L2 frequencies to provide better jamming resistance than the Y-code. Moreover, a new radio frequency link L5 (1176.45 MHz) for civilian users has been introduced. This signal, available since
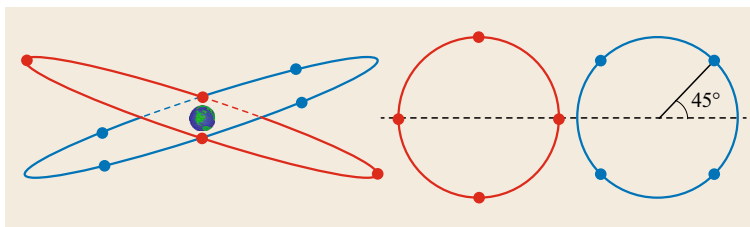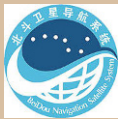

Fig. 1.9 Schematic illustration of an 8/2/1 Walker constellation

Table 1.2 An overview of the global and regional satellite-based navigation systems. Logos reproduced with permission of IAC PNT and FGUP TSNIIMASH (GLONASS), China Satellite Navigation Office (BeiDou), QZS System Services Inc. (QZSS), and Indian Space Research Organization (IRNSS/NavIC). The Navstar GPS logo is in the public domain

System	GPS	GLONASS	BeiDou	Galileo	QZSS	IRNSS/NavIC
Orbit	MEO	MEO	MEO, IGSO, GEO	MEO	IGSO, GEO	IGSO, GEO
Nominal number of satellites	24	24	27, 3, 5	30	3, 1	4, 3
Constellation	6 planes 56° inclination	Walker (24/3/1) 64.8° inclination	Walker (24/3/1) 55° inclination	Walker (24/3/1) 56° inclination	IGSOs with 43° inclination	IGSOs with 29° inclination
Services	SPS, PPS	SPS, PPS	OS, AS, WADS, SMS	OS, CS, PRS	GCS, GAS, PRS, EWS, MCS	SPS, RS
Initial service	Dec 1993	Sep 1993	Dec 2012	2016/2017 (planned)	2018 (planned)	2016 (planned)
Origin	USA	Russia	China	Europe	Japan	India
Coverage	Global	Global	Global	Global	East Asia Oceania region	$-30° < \phi < 50°$ $30° < \lambda < 130°$
Frequency (MHz)	L1 1575.42 L2 1227.60 L5 1176.45	L1 1602.00 L2 1246.00 L3 1202.025	B1 1561.098 B2 1207.14 B3 1268.52	E1 1575.42 E5a 1176.45 E5b 1207.14 E6 1278.75	L1 1575.42 L2 1227.60 L5 1176.45 E6 1278.75	L5 1176.45 S 2492.028

SPS: Standard Positioning Service; PPS: Precise Positioning Service; OS: Open Service; AS: Authorized Service; WADS: Wide Area Differential Service; SMS: Short Message Service; CS: Commercial Service; PRS: Public Regulated Service; GCS: GPS Complementary Service; GAS: GPS Augmentation Service; EWS: Early Warning Service; MCS: Message Communications Service; PS: Precision Service; RS: Restricted Service

the launch of the Block IIF satellites (beginning in May 2010) will be interoperable with that of Galileo, QZSS, and IRNSS/NavIC.

GLONASS

The former Soviet Union developed the *Global'naya Navigatsionnaya Sputnikovaya Sistema* or GLONASS (Chap. 8). The first GLONASS satellite was launched on 12 October 1982. By early 1996, a fully operational constellation of 24 satellites was in orbit. Unfortunately, the full constellation was short lived. Russia's economic difficulties following the dismantling of the Soviet Union hurt GLONASS. By 2002 the constellation had dropped to as few as seven satellites, with only six available during maintenance operations. With support of the Russian government, GLONASS was reborn, and on 8 December 2011, full operational capability (FOC) was again achieved and has been subsequently maintained.

The GLONASS satellites are categorized into three different generations: first generation GLONASS I/II (started in 1982), second generation GLONASS-M (started in 2003), and third generation GLONASS-K (started in 2011). All GLONASS satellites launched since December 2005 have been GLONASS-M satellites with the exception of two GLONASS-K1 satellites, launched on 26 February 2011 and 30 November 2014. GLONASS uses frequency division multiple access (FDMA) for its signals. Originally, the system transmitted the signals within two bands: Ll, 1602−1615.5 MHz, and L2, 1246−1256.5 MHz, at frequencies spaced by 0.5625 MHz at L1 and by 0.4375 MHz at L2. GLONASS-K satellites include, for the first time, code division multiple access (CDMA) signals accompanying the legacy FDMA signals. GLONASS-K1 as well as the latest GLONASS-M satellites transmit a CDMA signal on a new L3 frequency (1202.025 MHz).

Galileo

The Galileo system (Chap. 9) is a joint initiative of the European Commission (EC) and the European Space Agency (ESA). The first two In-orbit Validation (IOV) satellites were launched on 21 October 2011, and the

third and fourth IOV satellites were launched on 12 October 2012. The first two full-operational-capability satellites, were launched on 22 August 2014, into wrong orbits due to an upper rocket stage anomaly. Up to the end of 2015, six further satellites have been launched. In all, when the constellation is fully developed, there will be 30 Galileo satellites with 24 designated as primary and six spares.

Galileo satellites transmit three levels of service in three frequency bands using CDMA. The Open Service (OS) and the Public Regulated Service (PRS) are transmitted in the E1 frequency band centered on 1575.46 MHz (the same as the GPS L1 frequency) and the PRN ranging codes are modulated onto the carrier using binary offset carrier (BOC) techniques with each satellite, like GPS, assigned separate codes. The Commercial Service (CS) signal and the PRS are transmitted in the E6 frequency band centered on 1278.75 MHz using binary phase-shift keying (BPSK) and BOC modulation, respectively. Data and data-less (pilot) signals are transmitted in the E5 frequency band centered on 1191.795 MHz using BOC modulation. Data and pilot signals are also available on E1 and E6. The signals are separated into an E5a and an E5b component and either can be tracked separately or together. The various signals also contain navigation messages supplying the necessary information for acquiring Galileo signals and for determining receiver positions and time.

BeiDou

China fielded a demonstration regional satellite-based navigation system known as BeiDou (Chinese for the *Big Dipper* asterism and pronounced *bay-dough*) following a program of research and development that began in 1980 (Chap. 10). The initial constellation of three GEO satellites was completed in 2003. A fourth GEO satellite was launched in 2007. The initial regional BeiDou system (BeiDou-1) has been replaced by a global system known as BeiDou-2 (or simply BeiDou and formerly known as Compass). The BeiDou Navigation Satellite System (BDS) as it is officially now known will eventually include five GEO satellites, 27 MEO satellites, and five IGSO satellites. BeiDou-2 was declared operational for use in China and surrounding areas on 27 December 2011. FOC for this area was declared on 27 December 2012. The system will provide global coverage by 2020. As of February 2016, 21 BeiDou satellites have been launched. Some of the BeiDou satellites launched from 30 March 2015 onward are a new version termed BeiDou Phase 3 or simply BeiDou-3.

The satellites transmit two levels of service, an open service and an authorized service primarily for the Chinese government and military using three frequency bands. The bands and the central frequencies for the satellites now in use, the BeiDou-2 satellites, are B1 at 1561.098 MHz, B2 at 1207.14 MHz, and B3 at 1268.52 MHz. BeiDou-3 will transmit modernized signals in the L1/E1 and L5/E5 bands as well as the BeiDou B3 band. For compatibility it is also foreseen that they will transmit the B1 open service signal of the BeiDou-2 system.

1.5.2 Regional Navigation Satellite Systems

QZSS

The Quasi-Zenith Satellite System (QZSS; Chap. 11) will use multiple satellites in inclined orbits, placed so that one satellite always appears near zenith above Japan, well known for its high-rise cities where the signals from GPS satellites can be easily blocked. The design provides high-accuracy satellite positioning service covering almost all of the country, including urban canyons and mountainous terrain. The IGSO satellites will be supplemented with one GEO satellite. The start of full QZSS service is planned for 2018. QZSS Phase One is validating technological enhancement of GPS availability, performance, and application using the first QZSS satellite, Michibiki. Michibiki was launched on 11 September 2010 and is in full operation. Phase Two will demonstrate full system capability using at least three QZSS satellites, including Michibiki. Future plans call for a seven-satellite constellation.

The satellites will generate and transmit their own signals, compatible with modernized GPS signals. QZSS also transmits GPS corrections and availability data, the L1-SAIF (Submeter-class Augmentation with Integrity Function) signal, and so is also considered as a satellite-based augmentation system satellite. Altogether, Michibiki transmits six signals with structures similar to and compatible with GPS and Galileo signals: L1-C/A (1575.42 MHz), L1C (1575.42 MHz); L2C (1227.6 MHz), L5 (1176.45 MHz), L1-SAIF 1575.42 MHz), and LEX (L-band Experiment, 1278.75 MHz) a QZSS experimental signal for a high precision (3 cm level) service, sharing the frequency of the Galileo E6 signal.

IRNSS/NavIC

The Indian government has developed the Indian Regional Navigation Satellite System (IRNSS) as an independent system serving India and the surrounding area (Chap. 11). In April 2016, IRNSS was renamed to NavIC, a Hindi word for sailor or navigator as well as an acronym for *Navigation with Indian Constellation*. The area of coverage is from 30° south to 50° north in latitude and 30° east to 130° east in longi-

Table 1.3 An overview of the SBASs

System	WAAS	SDCM	EGNOS	MSAS	GAGAN
Orbit	GEO	GEO	GEO	GEO	GEO
Nominal number of satellites	3	3	4	1	3
Longitudes	133° W, 107° W, 98° W	16° W, 95° E, 167° E	15.5° W, 5° E, 25° E, 31.5° E	145° E	55° E, 83° E, 93.5° E
Date of being operational	July 2003	–	October 2009	September 2007	February 2014
Origin	USA	Russia	Europe	Japan	India
Service area	CONUS, Alaska, Canada, Mexico	Russia	Europe	Japan	India
Frequency (MHz)	L1 1575.42 L5 1176.45	L1 1575.42	L1 1575.42 L5 1176.45	L1 1575.42	L1 1575.42 L5 1176.45

tude. IRNSS provide two types of service: the SPS, which is an open service for all users, and the Restricted Service (RS), which is an encrypted service available only to authorised users. IRNSS is expected to provide a real-time pseudorange-based position accuracy of better than 20 m in the primary service area. The IRNSS constellation consists of three GEO satellites as well as two pairs of IGSO satellites.

The first satellite in the constellation, an IGSO satellite, IRNSS-1A, was launched on 1 July 2013. The second IGSO satellite, IRNSS-1B, was launched on 4 April 2014. The first GEO satellite, IRNSS-1C, was launched on 15 October 2014. IRNSS-1D, the third IGSO satellite, was launched on 28 March 2015. The constellation was completed by launches of IRNSS-1E, 1F, and 1G in 2016. The satellites transmit navigation signals at 1176.45 and 2492.028 MHz in the L- and S-bands respectively. The SPS and RS are transmitted on both frequencies. The SPS uses BPSK modulation while the RS uses BOC modulation with data and pilot channels.

1.5.3 Satellite–Based Augmentation Systems

In addition to the GNSS/RNSSs, there are also SBASs, which use geostationary communications satellites to provide differential correction data and integrity information to GNSS users in real time using a *bent path* from a ground station through the satellite to a user's equipment. The systems use a state-space-domain approach in which corrections for GNSS satellite orbit and clock data along with ionospheric propagation delays are provided.

Currently, four SBASs are in full operation: the FAA's WAAS, the European Geostationary Navigation Overlay Service (EGNOS), Japan's Multifunctional Transport Satellite (MTSAT) Satellite-based Augmentation System (MSAS), and India's GPS-aided GEO Augmented Navigation System (GAGAN). In addition, Japan's Quasi-Zenith Satellite System has an augmentation component as already mentioned. Russia's System for Differential Correction and Monitoring (SDCM) is currently in development. See Table 1.3 for an overview and Chap. 12 for a detailed description of SBASs.

1.6 GNSS for Science and Society at Large

GNSS is used for many types of applications, covering the mass market, professional and safety-critical applications as well as a whole range of scientific applications (Chaps. 29, 30, and 32). There are therefore literally hundreds of applications of GNSS, from the everyday to the exotic. And with the advent of next generation GNSSs, many more applications are expected to emerge.

According to a recent market study [1.26], the global GNSS market is expected to grow from roughly 50B Euros in 2013 to more than 100B Euros in 2023, when considering only the core revenue, i. e., the value of the chipsets sold. Enabled revenues covering the entire end-user equipment are projected to even grow from 200B Euros to almost 300B Euros in the same period. As illustrated in Fig. 1.10, the GNSS market is clearly
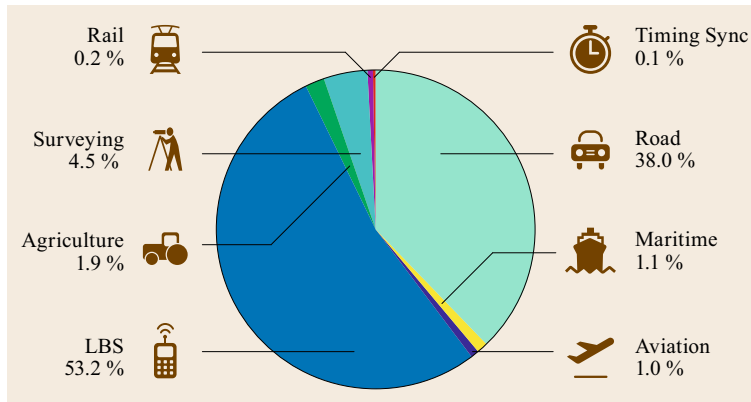
Fig. 1.10 Distribution of cumulative global revenue from GNSS chipset sales projected for the 2013–2023 period (after [1.26], courtesy of European GNSS Agency)
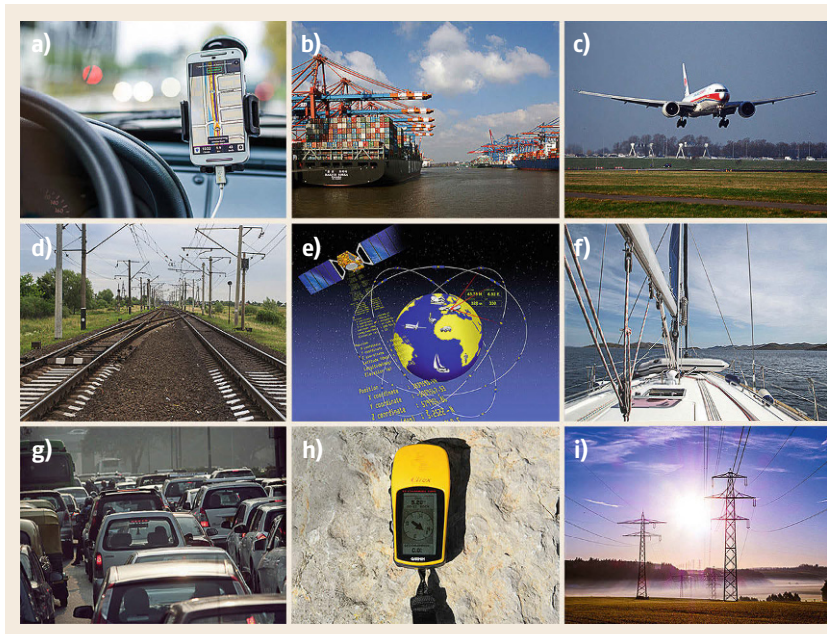

Fig. 1.11a−i Examples of everyday GNSS applications; from car navigation, to the landing of aircraft, to electrical power grid maintenance. (Photos courtesy of pixabay.com (a−d), (f−i) and ESA, J. Huart (e))

dominated by personal navigation devices (covered by location based services, LBS) and in-car navigation systems, which constitute more than 90% of the global core revenue. High-precision and specialized GNSS equipment for surveying and agriculture as well as maritime and aviation use, in contrast, contributes less than 10% of the overall chipset sales.

By far the most common use of GNSS is for navigation, which includes navigation for people who are hiking and geocaching (a treasure-hunting game); navigation of cars and other vehicles; ocean navigation for marine vessels and channel dredging; and aircraft control, as in approach and landing at airports (Fig. 1.11). Some of these applications, requiring higher accuracy, involve the use of carrier-phase measurements. GNSS is also used for tracking of people, vehicles, vessels,

aircraft, and assets, where GNSS-determined positions are reported via a supplementary communication channel such as that provided by a mobile phone.

One of the earliest high-precision applications of GNSS was in surveying and geodesy (Chaps. 35 and 36) (Fig. 1.12). Through the use of carrier-phase measurements, accurate coordinates of lot boundaries and geodetic markers can be established. Later on these carrier-phase-based techniques found their way into machine control, attitude determination, and precision agriculture [1.27]. Because GNSS also provides precise time (Chap. 41), it is also used to synchronize timing systems worldwide, permitting very accurate time-tagging of financial trades, for example. GNSS timing is widely used in the telecommunication industry including synchronization of mobile phone networks. It is also

Fig. 1.12a–i Examples of high-precision GNSS applications; from surveying and geodesy to machine guidance and precision agriculture (Photos courtesy of Position Partners ((**a**), *left*); Leica Geosystems ((**a**), *right*); M. Gottlieb, University NAVSTAR Consortium (UNAVCO) (**b**); pixabay.com (**c,f**) TU Delft (**d**); B. Morris (**e**); Trimble (**g**); V. Janssen (**h**); Deere & Co. (**i**))
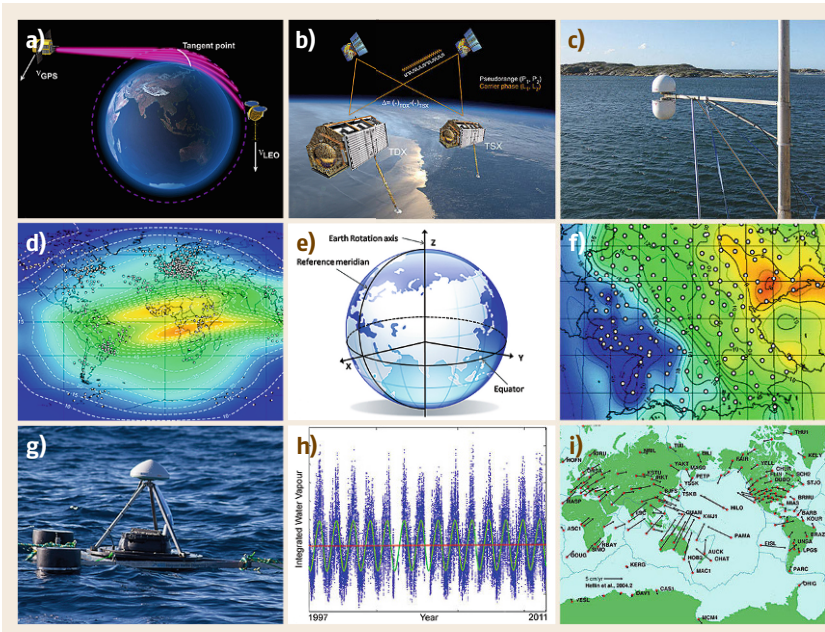

Fig. 1.13a–i Examples of scientific GNSS applications; from atmospheric sensing and reference frame studies to tectonic monitoring (Photos courtesy of UCAR 2007 (**a**); P. Kuss, DLR, NASA (**b**); J. Löfgren (**c**); N. Jakowski, DLR (**d**); J. Legrand & C. Bruninx, ROB (**e**); G. Dick, GFZ (**f**); IMOS (**g**); G. Elgered, Chalmers (**h**); NASA, JPL-Caltech (**i**))

used for electrical power grids [1.28] to synchronize the phase of alternating current and for power-line fault isolation.

GNSSs are also important for Earth system studies and global environmental Earth observation (Fig. 1.13). Advanced GNSS receivers are operated continuously or periodically at geophysically interesting sites (Chap. 37). Long-term tectonic plate motion is mea-

sured using GNSS [1.29] and networks of receivers are used to assess Earth surface activities for monitoring landslide and volcano activity or the study of land uplift following the last ice age. By monitoring crustal motions with extensive GNSS networks of continuously operating tracking stations, researchers have the long-term goal of being able to make accurate earthquake predictions, and thereby save lives.

GNSSs are also an invaluable instrument for atmospheric sensing (Chaps. 38 and 39). As GNSS signals are affected by their passage through the ionosphere and the lower atmosphere, an appropriate analysis of the received signals can be used to map the ionosphere's variable electron content [1.30] and the amount of water vapor in the troposphere [1.31]. Several national weather services use GNSS-determined water-vapor measurements to improve their weather forecasts. And GNSSs also help us to understand the processes taking place in the ionosphere. This is important as bad weather in the ionosphere may severely disturb our communication, navigation, and power systems. By looking at the effects that earthquakes and tsunamis have on ionospheric electron density [1.32, 33], GNSS ionospheric sensing may become an important component in tsunami warning systems as well.

During the last few years, GNSS data has also proven its potential for climate monitoring. Several studies have demonstrated the potential of both ground-based and radio occultation GNSS data for the accurate measurement of atmospheric water-vapor content and temperature structure [1.34, 35]. Determination of global temperature trends are important when it comes to the question of global warming. And since GNSS also provides an accurate positioning tool for investigating how ice masses move and change, it also contributes to gaining a better understanding of the linkage between global warming and the melting of our glaciers and the shrinkage of the Arctic ice cover.

References

1.1   J.R. Vetter: Fifty years of orbit determination: Development of modern astrodynamics methods, Johns Hopkins APL Tech. Dig. **27**(3), 239–252 (2007)

1.2   T.A. Stansell: The Navy Navigation Satellite System: Description and status, Navigation **15**(3), 229–243 (1968)

1.3   R.J. Danchik: An overview of Transit development, Johns Hopkins APL Tech. Dig. **19**(1), 18–26 (1998)

1.4   P. Daly, G.E. Perry: Recent developments with the Soviet Union's VHF satellite navigation system, Space Commun. Broadcast. **4**, 51–61 (1986)

1.5   P. Daly, G.E. Perry: Update on the behaviour of the Soviet Union's VHF satellite navigation system, Space Commun. Broadcast. **5**, 379–384 (1987)

1.6   G. Seeber: *Satellite Geodesy: Foundations, Methods and Aplications* (Walter de Gruyter, Berlin 2003)

1.7   K. Kovach: New user equivalent range error (UERE) budget for the modernized Navstar Global Positioning System (GPS), Proc. ION NTM, Anaheim (2000) pp. 550–573

1.8   Global Positioning System Standard Positioning Service Performance Standard (US Department of Defense, Washington DC 2008)

1.9   O. Montenbruck, P. Steigenberger, A. Hauschild: Broadcast versus precise ephemerides: A multi-GNSS perspective, GPS Solutions **19**(2), 321–333 (2015)

1.10   R. Prieto-Cerdeira, R. Orus-Peres, E. Breeuwer, R. Lucas-Rodriguez, M. Falcone: The European way: Performance of the Galileo single-frequency ionospheric correction during in-orbit validation, GPS World **25**(6), 53–58 (2014)

1.11   D. Milbert: Dilution of precision revisited, Navigation **55**(1), 67–81 (2008)

1.12   S. Whitney: Global Positioning System status, Proc. ION GNSS+, Tampa (2015) pp. 1193–1206

1.13   Minimum Operational Performance Standards for Global Positioning/Wide Area Augmentation System Airborne Equipment (RTCA, Washington DC 2006)

1.14   R. Leandro, M. Santos, R.B. Langley: UNB neutral atmosphere models: Development and performance, Proc. ION NTM 2006, Monterey (ION, Virginia 2006) pp. 564–573

1.15   R.F. Leandro, R.B. Langley, M.C. Santos: UNB3m_pack: A neutral atmosphere delay package for radiometric space techniques, GPS Solutions **12**(1), 65–70 (2008)

1.16   A. Komjathy: Global Ionospheric Total Electron Content Mapping Using the Global Positioning System, Ph.D. Thesis (Univ. New Brunswick, Fredericton 1997)

1.17   J.A. Klobuchar: Ionospheric time-delay algorithm for single-frequency GPS users, IEEE Trans. Aerosp. Electron. Sys. **23**(3), 325–331 (1987)

1.18   European GNSS (Galileo) Open Service Ionospheric Correction Algorithm for Galileo Single Frequency Users, Iss. 1.2 (European Commission, 2016)

1.19   S. Banville, R.B. Langley: Instantaneous cycle-slip correction for real-time PPP applications, Navigation **57**(4), 325–334 (2010)

1.20   P.J.G. Teunissen: The Least-squares Ambiguity Decorrelation Adjustment: A method for fast GPS integer ambiguity estimation, J. Geod. **70**(1), 65–82 (1995)

1.21   RTCM Standard 10403.2 Differential GNSS Services, Version 3 with Ammendment 2 (RTCM, Arlington 2013)

1.22   C. Haslett: *Essentials of Radio Wave Propagation* (Cambridge Univ. Press, Cambridge 2008)

1.23   A.W. Doerry: *Earth Curvature and Atmospheric Refraction Effects on Radar Signal Propagation* (Sandia National Laboratories, Albuquerque NM 2013), Sandia Report SAND2012-10690

1.24   G. Weber, D. Dettmering, H. Gebhard, R. Kalafus: Networked transport of RTCM via internet protocol (Ntrip) – IP-streaming for real-time GNSS applications, Proc. ION GPS, Long Beach (ION, Virginia 2005) pp. 2243–2247

1.25   J.G. Walker: Satellite constellations, J. Br. Interplanet. Soc. **37**, 559–572 (1984)

1.26   European GNSS Agency: *GNSS Market Report*, 4th edn. (Publications Office of the European Union, Luxembourg 2015)

1.27   J.V. Stafford: Implementing precision agriculture in the 21st century, J. Agric. Eng. Res. **76**(3), 267–275 (2000)

1.28   A. Carta, N. Locci, C. Muscas, S. Sulis: A flexible GPS-based system for synchronized phasor measurement in electric distribution networks, IEEE Trans. Instrum. Meas. **57**(11), 2450–2456 (2008)

1.29   K.M. Larson, J.T. Freymueller, S. Philipsen: Global plate velocities from the Global Positioning System, J. Geophys. Res. Solid Earth **102**(B5), 9961–9981 (1997)

1.30   M. Hernandéz-Pajares, J.M. Juan, J. Sanz: New approaches in global ionospheric determination using ground GPS data, J. Atmos. Sol. -Terr. Phys. **61**(16), 1237–1247 (1999)

1.31   M. Bevis, S. Chiswell, T.A. Herring, R.A. Anthes, C. Rocken, R.H. Ware: GPS meteorology: Mapping zenith wet delays onto precipitable water, J. Appl. Meteorol. **33**(3), 379–386 (1994)

1.32   E. Calais, J.B. Minster: GPS detection of ionospheric perturbations following the January 17, 1994, Northridge earthquake, Geophys. Res. Letts. **22**(9), 1045–1048 (1995)

1.33   A. Komjathy, Y.-M. Yang, X. Meng, O. Verkhoglyadova, A.J. Mannucci, R.B. Langley: Review and perspectives: Understanding naturalhazards–generated ionospheric perturbations using GPS measurements and coupled modeling, Radio Sci. **51**(7), 951–961 (2016)

1.34   T. Nilsson, G. Elgered: Long-term trends in the atmospheric water vapour content estimated from ground-based GPS data, J. Geophys. Res. **113**(D19101), 1–12 (2008)

1.35   R.A. Anthes: Exploring earth's atmosphere with radio occultation: Contributions to weather, climate and space weather, Atmos. Meas. Tech. **4**, 1077–1103 (2011)