Chapter 2

GPS Signal Structure

The Global Positioning System (GPS) is based on the accurate measurement of time: the GPS signals contain time tags indicating the time of transmission, which is compared to the time of reception and the distance to the satellite is estimated. For single-receiver real-time positioning, the transmitters positions and corrections to the raw measurements are needed to be included in the signal received. The GPS signals carry information both in the data message (which has a rate of 50 bps) and in the modulation coding. The system uses two right-hand circularly polarized L-band signals at $f_1 = 1.57542$ GHz and $f_2 = 1.2276$ GHz; f_1 is modulated in phase and quadrature and f_2 only in phase (future GPS developments foresee the use of more frequencies). The GPS signals use the spread spectrum technique using a Code Division Multiple Access-Direct Sequence technique (CDMA-DS). The signals are encoded with uncorrelated codes to discriminate between transmitting satellites. They actually carry precise information about the time of transmission because they are synchronized with the stabilized clocks on-board the satellites. The use of two frequencies is motivated by the dispersive nature of the effect introduced by the ionosphere. Finally, because GPS was conceived as a military system, it provides two types of positioning service: Standard Positioning System (SPS) and Precise Positioning System (PPS). That is, a coarse positioning and a precise positioning. The latter can only be achieved using classified receivers that have information about the finest encryption of the signal.

In this chapter we present the GPS signal structure, the different effects that appear in the measurements and strategies for the data processing to estimate the ionospheric and tropospheric parameters.

2.1 Signal Modulation and Access Technique

The GPS signal can be written as:

$$s_1(t) = A_1 D(t) G^i(t) \cos(\omega_1 t + \phi_1) + A_1 D(t) P^i(t) \sin(\omega_1 t + \phi_1),$$

$$s_2(t) = A_2 P^i(t) \cos(\omega_2 t + \phi_2) \tag{2.1}$$

where $D(t)$ is the data message at 50 bps, G^i and P^i are the pseudo-random codes that identify the transmitter and $\omega_1 = 2\pi f_1$, $\omega_2 = 2\pi f_2$. The f_1 carrier is modulated in amplitude with the data message. This data message is broken into data frames with different periodicities and contains information about the position of the satellites, time and corrections needed to increase the accuracy. The spread spectrum technique is based on the reduction of the power density by enlarging the bandwidth of the signal. This is accomplished by multiplying the amplitude-modulating message with a BPSK pseudo-noise sequence. The Coarse Acquisition (C/A) code, represented in Equation 2.1 by $G^i(t)$, is a Gold code of 1023 bits, also called chips. The chip rate in the C/A code is $f_{chip} = 1.023$ MHz. Therefore, the C/A code has a periodicity of 1 ms. This leads to a processing gain of 43.1 dB. The chip rate of the C/A code is, in fact, a fundamental frequency for the GPS system.

The Precise code (P^i) has a chip rate of $f_{chip} = 10.23$ MHz and the sequence is formed multiplying two long subsequences $X1$ (15,345,000 chips) and $X2$ (15,345,037 chips), the former having a 1.5 s period. By shifting i chips the $X2$ code prior to the product, the sequence for satellite i is obtained. Since $X1$ and $X2$ are prime to each other, the P code, as the product of $X1$ and $X2$, would have a period of slightly more than 38 weeks ($15,345,000 \cdot 15,345,037/f_{chip}$) , but the sequences are reset to begin the week at the same epoch time. The processing gain is in this case 53 dB ([7]).

This precise code, however, is encrypted with an additional classified code (identified as Y code) of 500 KHz in order to prevent spoofing or jamming of the GPS signals. This is the Anti-Spoofing mode (A/S).

The frequencies f_1 and f_2 are defined to be integer multiples of the P chip rate: $f_1 = 154 \cdot 10.23$ MHz, $f_2 = 120 \cdot 10.23$ MHz. This allows the receiver to synthesize all the frequencies from the same local oscillator and with few steps. For instance, the MAGR receiver ([8]) uses $f_1 - f_2 = 34 \cdot 10.23$ MHz as the Intermediate Frequency. It is worth mentioning that the GPS system was conceived to allow a user to track the satellites with simple instrumentation (a hemispherical antenna is enough to detect the signal) and with no a priori knowledge of the location of the transmitters.

Good reviews of todays receiver technology can be found in the literature ([9], [10]) and we will here just summarize the operation mode.

When the receiver is powered up (cold start), it has no information of where it is located nor where the satellites are. Therefore, the incoming signals have an unknown Doppler shift due to the relative motion of the receiver and the satellites. In addition, the signal is buried

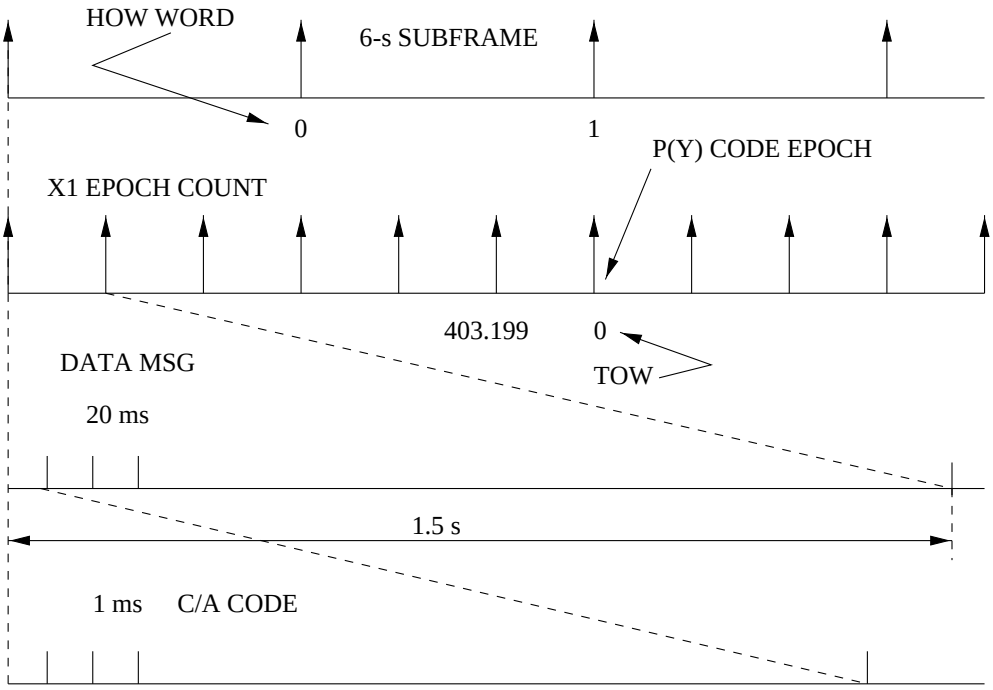

Figure 2.1: Timing relationships between C/A epochs, P epochs and data message (from [10]). HOW stands for Handover Word and is the number of 6 s intervals in the current week, TOW is the Time of Week which is the number of 1.5 s intervals of the current week. These two values indicate the time of transmission and are also aiding in the local generation of the P code, since it is reset every week.

below the noise level and needs to be raised by correlating with an aligned clean replica of the pseudo-random code. Therefore, the receiver starts to sweep across the different combinations of Doppler shifts and delays, until a peak of correlation is found. If this peak is kept (i.e. it is not a false alarm), then it locks onto that signal. The scanning is performed using the C/A code (due to its availability, easy generation, and wider chips) and using coarse Doppler bins *. After the navigation message is decoded, the receiver has information about the time (in GPS weeks and seconds) and about the location of other satellites. Thus, the search for new satellites is eased. Because the P code has a very long period, it is difficult to acquire without external information: once the data message is decoded and the receiver is locked to the C/A code, the receiver can obtain a good initial guess to correlate the P code based on the information of the handover word (HOW) (see Figure 2.1 for an illustration of the time synchronization).

The question of how the P code can be tracked without knowing the encrypting codes (Y), has been investigated almost from the start of the GPS. Different techniques try to overcome this lack of information. This will result in a worst Signal to Noise Ratio. Nevertheless,

*The frequency bins are defined to be a fraction of the predetection bandwidth, which in turn, is based on the expected uncertainty; if the search fails, both the bandwidth and the bins are resized

receiver technology has evolved to such a point that precise and accurate measurements can be obtained with civilian receivers. There are different techniques to cope with the P(Y) code. Modern receivers implement mainly two techniques: the *cross-correlation*, which makes use of the fact that both s_1 and s_2 from Equation 2.1 have the same encrypting code and thus, it can be removed by cross-correlating them and the *Z-tracking* technique, which correlates the incoming signal with the clean replica of the P code (without the Y encryption); the receiver estimates a Y code and then correlates it with the signal. These techniques represent a loss of 27 dB and 14 dB with respect to correlating with the real P(Y) code (see e.g. [11] for further details)

Finally, a GPS receiver dynamically combines the information from all the loops to keep them in a tracking mode. The measured ranges are corrected with the information from the navigation data (which include rough corrections for the satellite clock, ionospheric and tropospheric models and the satellites orbits) so that the receiver can solve for its position and time, as well as monitor its movements by making use of a Kalman filter.

2.2 GPS Observables

2.2.1 Primary Observables: Phase and Pseudorange

In this section we describe the *primary observables*, or the observables that are produced by the receiver at both frequencies. Linear combinations of these observables are used in the GPS data processing for different purposes and will be later described: carrier-phase smoothed pseudorange, wide-lane combination, ionospheric combination and ionospheric free combination.

The first primary observable to be obtained is the *pseudorange*. When the signal processor DLL finds the peak of correlation, it produces an observation of the code phase, that is, the signal transmit time for the local receive time. The difference between these times, scaled by the speed of light is the *pseudorange*. The precision of the range measurement is therefore dependent on the error with which the internal Delay Lock Loop (DLL) can track the code. The error is given by ([12])

$$\sigma_{code} \approx \frac{c}{R_c}\sqrt{\frac{B_L d}{2SNR_0}}, \tag{2.2}$$

where R_c is the chipping rate, B_L is the single-sided noise bandwidth of the loop and d is the Early-Late spacing in chips and SNR_0 is the carrier-to-noise ratio in Hz. The implementation of the DLL depends on the receiver. Analogous expressions can be found for the different cases. Equation 2.2, however, gives an correct order of magnitude of the precision. Typical

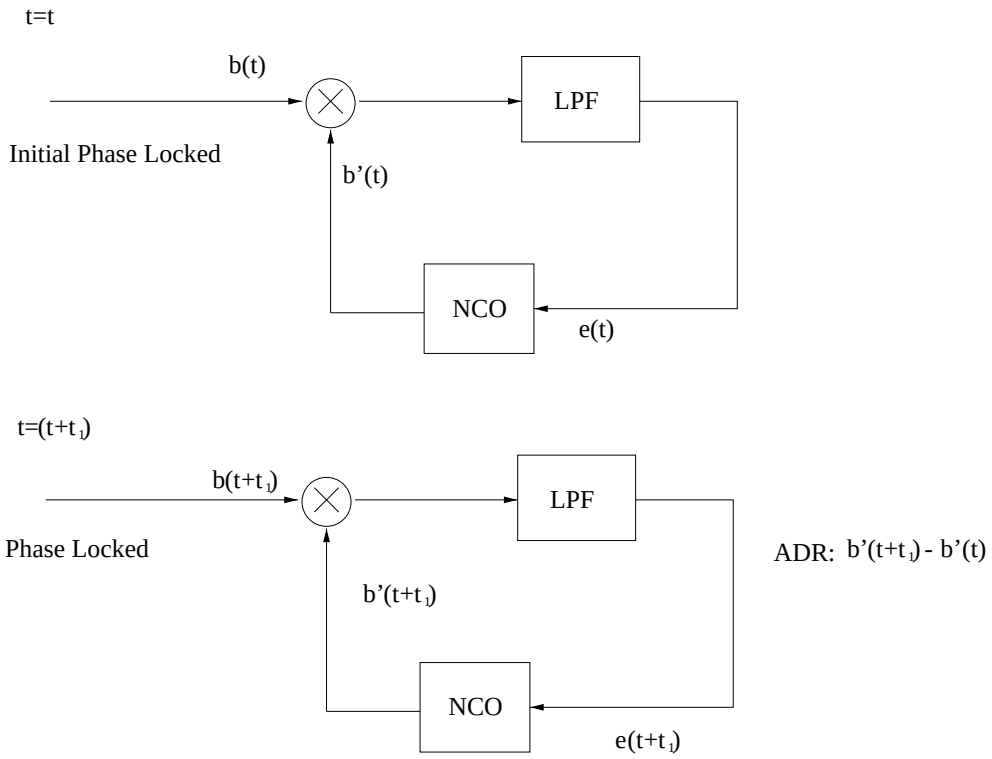

Figure 2.2: At the initial locking time, the incoming phase $b(t)$ is reproduced by the NCO as a $b'(t)$ to nullify the error. At subsequent times, the ADR is the difference of the present phase with the one at locking ($b'(t)$), therefore being a measurement of the relative motion. Actually, the ADR is formed using the error information, $e(t+t_1)$, but it is clearer to define it as $b'(t+t_1) - b'(t)$.

values are, for C/A code, $B_L = 1$ Hz (1-s loop time constant in the low-pass filter of the DLL implementation) $SNR_0 = 45$ dB-Hz and $d = 1$ (though this spacing increases the noise, it allows a better tracking of the dynamics of the signal), and then $\sigma_{code} \approx 1.04$ m.

This observable is termed as *pseudorange* because it contains not only the range but also the receiver and transmitter clock errors as well as additional effects such as extra delays introduced by the atmosphere. We devote Section 2.2.2 to analyze these effects.

The second observable is the *Accumulated Delta Range* (ADR): it is obtained by accumulating the commanded values given to the Numerically Controlled Oscillator (NCO) required to maintain the lock on the signal. It is, in fact, the integrated value of the Doppler frequency due to the relative motion of transmitter and receiver. It is often termed *carrier beat phase*. It is also common to refer to the ADR as *phase* and noted with L_i where i=1,2 is the frequency. Henceforth, we will use the term phase to indicate the ADR (see Figure 2.2 for a pictorial representation). The phase is a precise but inaccurate measurement. The precision

with which a PLL can track the phase is given by:

$$\sigma_{phase} \approx \frac{\lambda}{2\pi}\sqrt{\frac{1}{2SNR_0\tau}} \tag{2.3}$$

where, SNR_0 is the carrier-to-noise ratio in Hz and τ is the integration time or the inverse of the PLL bandwidth (see Table 2.1). However, there is an unknown initial bias (note that it is a measurement of relative motion). In addition, the phase tracking can be lost due to rapid movements of the receiver or unexpected change in the conditions in the signal path. This causes the receiver to loose cycles, leading to the *cycle slip* effect.

Observable	**SNR$_0$ (dB-Hz)**	R_c	**Chip Spacing**	τ	σ_{code}	σ_{phase}
C/A	45 dB-Hz	1.023 MHz	1	1 s	1.04 m	-
	45 dB-Hz	1.023 MHz	0.1	1 s	0.33 m	-
P	42 dB-Hz	10.23 MHz	1	1 s	0.165 m	-
	42 dB-Hz	10.23 MHz	0.1	1 s	0.05 m	-
L_1	45 dB-Hz	-	-	1 s	-	0.12 mm
	45 dB-Hz	-	-	0.5 s	-	0.2 mm
L_2	40 dB-Hz	-	-	1 s	-	0.3 mm
	40 dB-Hz	-	-	0.5 s	-	0.4 mm

Table 2.1: Errors in GPS signal tracking.

2.2.2 Modeling the Primary Observables

The *pseudorange* is a difference of transmit time and local receive time and hence includes the geometric distance covered by the radio signal, the receiver and transmitter clock errors, the extra delays produced by the non-zero refractivity of the atmosphere (neutral and ionosphere), instrumental delays, multipath and, finally, noise in the measurement itself. We can hence interpret the pseudoranges at both frequencies as follows:

$$\begin{aligned} P^p_{k,1} &= \rho^p_k - cdt_k + cdt^p + \frac{\alpha}{f_1^2} I^p_k + T^p_k \\ &\quad + hd_{k,1,P} + hd^p_{1,P} + M^p_{k,1,P} + \epsilon_{1,P} \end{aligned} \tag{2.4}$$

$$\begin{aligned} P^p_{k,2} &= \rho^p_k - cdt_k + cdt^p + \frac{\alpha}{f_2^2} I^p_k + T^p_k \\ &\quad + hd_{k,2,P} + hd^p_{2,P} + M^p_{k,2,P} + \epsilon_{2,P} \end{aligned} \tag{2.5}$$

where

- ρ^p_k is the geometric distance (on the order of 20,000 Km) between station k at the time of reception and satellite p at the time of transmission,

- cdt_k and cdt^p are the clock errors; the former can be initially off by any quantity, though for precise data processing needs to be within 1 ms; the latter is specified to be less than $1\mu s$ during 99.99 % of the time (see Section 2.2.4) ,
- $\frac{\alpha}{f_i^2} I_k^p$ is the ionospheric delay, as a function of frequency $(i = 1, 2)$; it is on the order of ten meters; $\alpha = 40.3$ m^3/(s^2 el), f is in Hz and I is the integrated electron density along the propagation path (in el/m^2),
- T_k^p is the tropospheric delay, on the order of two meters,
- $hd_{k,i,P}$ and $hd_{i,P}^p$ are the hardware delays for pseudorange and frequency f_i for the station k and the satellite p. They depend on the calibration procedure, but should be on the meter level and they are constants over long periods of time,
- M_k^p is the multipath term which depends on the DLL chip spacing and the precorrelation bandwidth; for P-code it can be as small as 0.15 C/A-chip delay (see [9]), and
- $\epsilon_{i,P}$ is the noise $(i = 1, 2)$ in the measurement of the pseudorange(σ_{code}), which is about 1 m.

Regarding the phase, the accumulated measurement can be written as:

$$\begin{aligned} L_{k,1}^p &= \rho_k^p - cdt_k + cdt^p + \frac{c}{f_1} N_{k,1}^p - \frac{\alpha}{f_1^2} I_k^p + T_k^p \\ &\quad + hd_{k,1,\Phi} + hd_{1,\Phi}^p + M_{k,1,P}^p + \epsilon_{1,\Phi} + \Psi_{k,1}^p \end{aligned} \tag{2.6}$$

$$\begin{aligned} L_{k,2}^p &= \rho_k^p - cdt_k + cdt^p + \frac{c}{f_2} N_{k,2}^p - \frac{\alpha}{f_2^2} I_k^p + T_k^p \\ &\quad + hd_{k,2,\Phi} + hd_{1,\Phi}^p + M_{k,2,\Phi}^p + \epsilon_{2,\Phi} + \Psi_{k,2}^p \end{aligned} \tag{2.7}$$

with the same terms as the ones described above, plus $N_{k,1}^p$ and $N_{k,2}^p$, which are unknown integer number of cycles (the ambiguity), and noting that the phase noise is in the order of 1 mm. Since the signal is circularly polarized, the phase wind up effect, $\Psi_{k,1}^p$, $\Psi_{k,2}^p$, caused by the relative rotation of the transmitting GPS satellite, p, and the receiver, k, should be considered in the modeling [13].

Note also that the presence of uncalibrated errors in the instrumentation prevents the estimation of the phase bias as an integer number. Usually, instrumentation biases and the integer number of cycles are merged into a single bias:

$$\begin{aligned} b_{k,1}^p &= \frac{c}{f_1} N_{k,1}^p + hd_{k,1,\Phi} + hd_{1,\Phi}^p \\ b_{k,2}^p &= \frac{c}{f_2} N_{k,2}^p + hd_{k,2,\Phi} + hd_{2,\Phi}^p \end{aligned} \tag{2.8}$$

2.2.3 Secondary Observables

2.2.3.1 Carrier-Phase Smoothed Pseudorange

The concepts of precision and accuracy actually define the main characteristics of the phase and the pseudorange respectively: the pseudorange is a true measurement of time differences (the delay plus the clock offset), but it is noisy, while the phase has a much lower noise but is a measurement of relative movement and therefore has an undetermined offset (see Figure 2.3). The *carrier-phase smoothed pseudorange* combines the primary observables to obtain the precision of the phase and the accuracy of the pseudorange. In fact, it is a combination that, starting from the phase time-series, first converts its dependency with the ionosphere to follow the pseudorange and then imposes its average be the same as the pseudorange, thus eliminating the unknown bias. We can write this process in two steps: first, we combine the phases as follows:

$$\begin{aligned}\tilde{L}^p_{k,1} &= L^p_{k,1} + \frac{2f_2^2}{f_1^2 - f_2^2}\left(L^p_{k,1} - L^p_{k,2}\right) \\ &= \tilde{\rho} - \frac{\alpha}{f_1^2} I^p_k + \frac{2f_2^2}{f_1^2 - f_2^2}\left(\frac{\alpha}{f_2^2} - \frac{\alpha}{f_1^2}\right) I^p_k \\ &= \tilde{\rho} + \frac{\alpha}{f_1^2} I^p_k \end{aligned} \tag{2.9}$$

$$\begin{aligned}\tilde{L}^p_{k,2} &= L^p_{k,2} + \frac{2f_1^2}{f_1^2 - f_2^2}\left(L^p_{k,1} - L^p_{k,2}\right) \\ &= \tilde{\rho} - \frac{\alpha}{f_2^2} I^p_k + \frac{2f_1^2}{f_1^2 - f_2^2}\left(\frac{\alpha}{f_2^2} - \frac{\alpha}{f_1^2}\right) I^p_k \\ &= \tilde{\rho} + \frac{\alpha}{f_2^2} I^p_k \end{aligned} \tag{2.10}$$

where in $\tilde{\rho}$ we have included the non-ionospheric terms of Equation 2.6. Then, we need to subtract the average value for its time series of $\tilde{L}^p_{k,i}$ and add the average value of $P^p_{k,i}$ and thus obtain the carrier-phase smoothed pseudorange, $\hat{P}^p_{k,i}$

$$\begin{aligned}\hat{P}^p_{k,1} &= \tilde{L}^p_{k,1} - <\tilde{L}^p_{k,1}> + <P^p_{k,1}> \\ \hat{P}^p_{k,2} &= \tilde{L}^p_{k,2} - <\tilde{L}^p_{k,2}> + <P^p_{k,2}> \end{aligned} \tag{2.11}$$

where $<>$ indicates time average over a phase-connected arc.

If the alignment process is good enough the carrier-phase smoothed pseudorange can be used as the main observable; otherwise, a phase bias has to be computed. This alignment, however, is only valid when the time series contain one single phase-connected arc: if there is a cycle slip in between, the offset is no longer constant and the mean value is no longer representative.

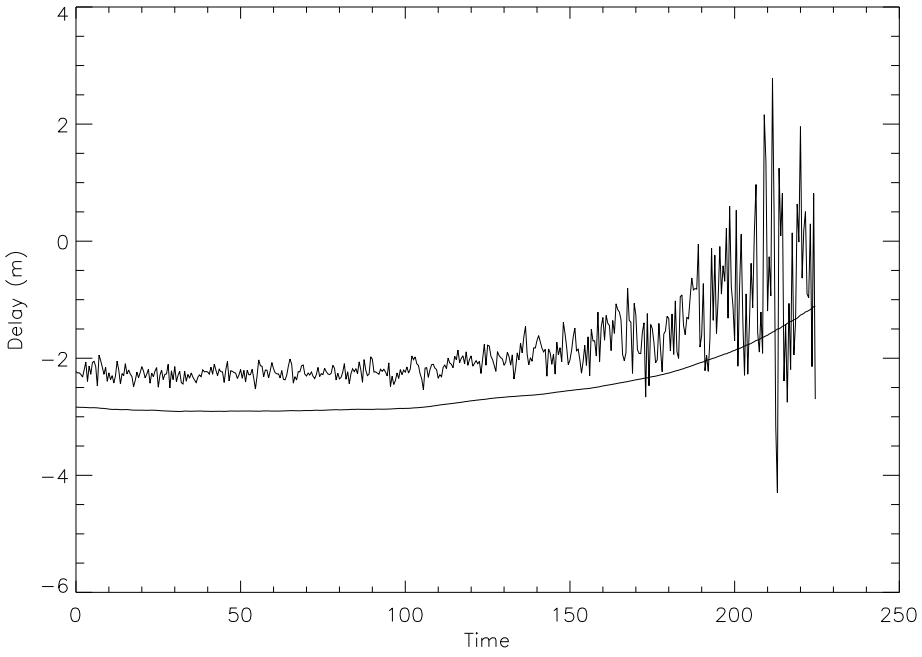

Figure 2.3: Qualitative example of noises on the carrier measurement and the pseudorange, and the bias between them (not to scale). For clarity, we have been plotted the observable minus the geometric range.

2.2.3.2 Cycle Slip Detection: Wide-laning and Narrow-laning

The wide laning process consists of a linear combination of the primary observables to yield a signal with a relatively large wavelength ($\lambda_\delta = c/(f_1 - f_2) \approx 86.2$ cm), so that cycle slips will be more noticeable (see [14]):

$$\begin{aligned} L^p_{\delta k} &= (f_1 L^p_{k,1} - f_2 L^p_{k,2})/(f_1 - f_2) \\ P^p_{\delta k} &= (f_1 P^p_{k,1} + f_2 P^p_{k,2})/(f_1 + f_2) \end{aligned} \tag{2.12}$$

The running mean and scatter (σ) of the difference between P and L are calculated using the data sorted by time. Subsequent estimates of the difference have to lie within 4σ. When two consecutive outliers are found, a potential cycle slip is detected. The wide lane combination, however, does not distinguish in which frequency the cycle slip happened (or both) and does not detect simultaneous and equal in magnitude cycle slips in both frequencies. To this end, the narrow-lane combination or ionospheric combination (see Section 2.2.5.1) is formed since it will contain all the wide-lane cycle slip information plus these minor cases.

2.2.4 Clock errors: Selective Availability

GPS satellites use atomic clocks (cesium and rubidium oscillators) with stabilities of about 1 part in 10^{13} over a day. However, GPS implements an artificial dither on their clocks in order to reduce the precision of the measurements. This is called Selective Availability (the term Selective Availability also refers to the intentionally introduced degradation of the satellites broadcast ephemeris, but for precise measurements, precise post-processed orbits are used, which are free from such degradation). Measurements of S/A reveal that the power spectrum of such dithering is mostly confined to below 0.5 Hz ([15], [16]). Therefore clock corrections that consider SA have to be on the order of 1 s or faster to be interpolated to higher frequencies. Clock corrections at 300 s rate are routinely computed at different processing centers [†] when simultaneously solving for the precise orbits of the satellites. Such processing requires a ground network of stations relatively far from each other to effectively decorrelate the satellite clock error from other effects. When such corrections are not available, it is necessary to process a network of GPS receivers at a time, even if we are interested in only one position. This is particularly true for precise positioning of a station; for a coarse positioning of a single station, this error represents the bound precision of some 300 m (in fact, SA is agreed to produce a maximum error of 300 m in the horizontal position for 99.99% of the time [15] and is usually around 50 m). In Section 2.3 we discuss data processing strategies to cope with satellite clock instabilities.

The local receiver clock can initially be off by any quantity. However, almost all receivers adjust their local clock to the GPS timing once they have enough satellites in view to compute its position (at least 4 satellites). The way how this adjustment is implemented depends on the receiver design, and it should be considered when modeling the clock.

2.2.5 Atmospheric effects

The atmosphere can be divided into two main parts: ionosphere and neutral atmosphere. The former extends from 60 Km up and includes free electrons due to the ultraviolet radiation from the sun; it has a dispersive effect that increases the group delay (appears in the pseudorange measurements) and advances the phase (apparent in the phase measurements). The neutral atmosphere is the region below 60 Km and its effect on the electromagnetic waves is mainly induced by the atmospheric gases; it produces an extra delay in both phase and group measurements and has no dependence on frequency at L-band.

[†]Jet Propulsion Laboratory, JPL, and the Center for Orbit Determination in Europe, CODE, are two organizations that daily produce precise orbits and clock corrections for the GPS system. Both centers are coordinated through the International GPS Service, IGS

2.2.5.1 Ionosphere

As it was shown in Equations 2.4 and 2.6, there is a dispersive term in the measured delay due to the ionosphere. This delay can be related to the refraction index as [17]

$$d_{ion} = \int_{r_i} (n-1) ds. \tag{2.13}$$

For frequencies above 100 MHz, the phase and group index of refraction can be related to the ionospheric electron density as [5]:

$$\begin{aligned} n_{phase} &= 1 - \frac{\alpha}{f^2} N_e \\ n_{group} &= 1 + \frac{\alpha}{f^2} N_e \end{aligned} \tag{2.14}$$

where, to first order, $\alpha = 40.3\ \mathrm{m}^3/(\mathrm{s}^2\ \mathrm{el})$ and N_e is the electronic density in el/m^3. The Faraday rotation affects the polarization states, and is used in polarimetry systems as a measure of the electron contents. In GPS signals we will combine the received measurements at both frequencies to extract the ionospheric information.

The primary observables can be combined in the *ionospheric combination* as follows, to isolate the delay due to the ionosphere:

$$\begin{aligned} L^p_{k,I} &= L^p_{k,1} - L^p_{k,2} = \alpha_I I^p_k + b^p_{k\Delta} \\ P^p_{k,I} &= P^p_{k,2} - P^p_{k,1} = \alpha_I I^p_k + q^p_{k,2} - q^p_{k,1} \end{aligned} \tag{2.15}$$

where, for the GPS frequencies $\alpha_I = \alpha(1/f_2^2 - 1/f_1^2) = 1.05 \cdot 10^{-17}$ in $\mathrm{m}^3/(\mathrm{el})$, when I is given in el/m^2, $b^p_{k\Delta} = b^p_{k,1} - b^p_{k,2}$, and $q^p_{k,1}$ and $q^p_{k,2}$ are the frequency dependent errors in the pseudorange. These two measurements are again combined to obtain the precision of the phase measurement and the accuracy of the pseudorange in the carrier-phase smoothed pseudorange. However, the terms due to the instrumental delays cannot be completely canceled and they will become part of the tomographic solution as we will see.

2.2.5.2 Neutral Atmosphere

The neutral atmosphere induces a non-dispersive extra delay, T^p_k, on both the group and the phase. This delay is due to the presence of gases in the atmosphere. It can be related to the neutral refractivity N by:

$$T^p_k = \int_{s.l.} 10^{-6} N\ dl + \mathcal{S} - \mathcal{G}, \tag{2.16}$$

$$N = 77.6\frac{P}{T} + 3.776 \cdot 10^5 \frac{P_w}{T^2} + 1.4W = N_h + N_w + 1.4W, \tag{2.17}$$

where $10^{-6}N$ can be expressed in mm/km, *s.l.* stands for the slant path through the atmosphere, $\mathcal{S}$ is the path length along the slant path, $\mathcal{G}$ is the straight-line path length, P is the total atmospheric pressure in mbar (1 mbar = 1 hPa), P_w is the water vapor pressure in mbar, T is the atmospheric temperature in K ([18], [19]), and W is the liquid water vapor in the atmosphere in grams per cubic meter. The latter term is generally neglected except for very particular cases ([20]). The term $\mathcal{S} - \mathcal{G}$ is the bending term which is about 1 cm or less for elevations greater than 15^o and in general will not be considered [21].

The total atmospheric delay can be split into two components: the *hydrostatic delay*, due to the dry gases in the troposphere and the non-dipole component of water vapor, and the *wet delay*, due to the dipole component of water vapor ([19]). The refractivity is also divided into hydrostatic N_h and wet components N_w, as shown in Equation 2.17. The hydrostatic delay participates with about the 90% of the total (about 2.3 m at sea level when projected to the zenith) tropospheric delay and it is well modeled with surface measurements ([22], [23], [24]). The wet refractivity, however, has a minor effect (on the order of 20 cm in the zenith) but it is not well modeled from surface measurements. In order to process the data to estimate the tropospheric delay, one must first remove the ionospheric delay using a combination of the primary observables, the *ionospheric free combination* [12] :

$$L^p_{k,c} = \frac{f_1^2 L^p_{k,1} - f_2^2 L^p_{k,2}}{f_1^2 - f_2^2}$$
$$P^p_{k,c} = \frac{f_1^2 P^p_{k,1} - f_2^2 P^p_{k,2}}{f_1^2 - f_2^2} \tag{2.18}$$

This combination cancels almost exactly the ionospheric contribution but, in return, increases the uncorrelated noise by a factor of $\sqrt{f_1^4 + f_2^4}/(f_1^2 - f_2^2) \approx 3$.

There have been a number of studies that prove the capability of GPS to sense water vapor in the atmosphere (see [21], [25], [26], [27], [28], [29] as some examples). This requires a precise processing of the GPS data.

2.3 GPS data processing

The primary information are the observables described in Section 2.2.2 plus the information read from the navigation message, in which satellite positions and correction terms are broadcast. It is clear from the equations of the observables, that there are many effects with different weight in the total measurements and these cannot be fully corrected for with the single receiver information mainly because broadcast ephemeris of the satellites are precise only up to 100 m, the transmitted corrections for the satellite clocks do not account for SA, which can be on the order of 300 m, and the atmospheric corrections are based on large scale

models, which may leave some meters of error. This translates into the nominal GPS precision of 300 m for an unclassified handy receiver.

Geodetic and atmospheric studies require a higher precision. This is accomplished through an accurate modeling of all the effects and combining information obtained at different receivers located over a wide area. This processing yields precise orbits for the transmitters, clock corrections and allows steady receivers to accurately obtain tropospheric delays over long time series. In this section we will discuss two approaches for precise GPS data processing, the characterization of each variable and the processing required for ionospheric and tropospheric studies (see [30] for a discussion on the equivalence of processing methodologies).

2.3.1 Differential GPS vs. Precise Point Positioning

There are two basic trends in the GPS data processing: 1) differencing the observations to cancel out common effects, and 2) direct use of observations, estimating the contribution of all the effects. Along this thesis they will be referred as *differencing* and *direct* approaches This is reflected in the two different types of software packages available for the processing, of which the most representative are GAMIT ([31]) and Bernese ([32]) for the differencing approach and GIPSY-OASIS II ([33]) for the direct approach. The pros and cons for each of them are the core of an on-going debate and will be here summarized.

2.3.1.1 Differential GPS Processing

The differential GPS stems from the fact that given two ground stations observing two satellites, there will be common errors in the observations. By differencing the observations, one can remove these errors from the equations and hence reduce the number of parameters to solve for. Let us now consider the observations for stations k and m and satellites p and q as were expressed in Equations 2.6 and 2.7; if we now subtract the observations tracked at the same stations, we will end up with

$$\begin{aligned} L_k^{pq} &= \Delta^{pq}\rho_k + \Delta^{pq}cdt + \frac{c}{f_i}\Delta^{pq}N_{k,i} - \frac{\alpha}{f_i^2}\Delta^{pq}I_k + \Delta^{pq}T_k \\ &\quad + \Delta^{pq}hd_{1,\Phi} + \Delta^{pq}M_{k,i,P} + \epsilon_{i,m,\Phi,s.d.} + \Delta^{pq}\Psi_{k,i} \end{aligned} \tag{2.19}$$

$$\begin{aligned} L_m^{pq} &= \Delta^{pq}\rho_m + \Delta^{pq}cdt + \frac{c}{f_i}\Delta^{pq}N_{m,i} - \frac{\alpha}{f_i^2}\Delta^{pq}I_m + \Delta^{pq}T_m \\ &\quad + \Delta^{pq}hd_{1,\Phi} + \Delta^{pq}M_{m,i,P} + \epsilon_{i,m,\Phi,s.d.} + \Delta^{pq}\Psi_{m,i} \end{aligned} \tag{2.20}$$

where $\Delta^{pq}\Phi \equiv \Phi^p - \Phi^q$, $s.d.$ stands for single difference, and we have removed the errors due to the station: instrumental offsets and station clock errors. There is, however, an increase

in thermal noise and in multipath noise. These two equations can be further subtracted to obtain the double difference:

$$L_{km}^{pq} = \Delta_{km}^{pq}\rho(t) + \frac{c}{f_i}\Delta_{km}^{pq}N_i - \frac{\alpha}{f_i^2}\Delta_{km}^{pq}I(t) + \Delta_{km}^{pq}T(t) + \Delta_{km}^{pq}M_{i,P} + \epsilon_{i,km,\Phi,d.d.} + \Delta_{km}^{pq}\Psi_i \quad (2.21)$$

(*d.d.* stands for double difference) where satellite errors are removed. The terms left here are the ones to solve for. The major advantage of such a processing is the reduction of unknowns (particularly the clock behaviour, which is difficult to model) and the speeding of the processing; in addition, it can be shown that such a differencing [34] eases the determination of the phase ambiguity. On the other hand, in the differenced observable, the noise is increased, and a high correlation between the same parameters of different stations makes their estimation difficult.

In the direct approach one needs to model and *solve* for all common errors. It implies a larger number of unknowns but it keeps the noise to the original values and the solution is easily interpreted. In addition, it provides *nuisance parameters* (e.g. instrumental biases): noise in the measurements or badly tracked data can fall in these extra parameters in the solution, leaving the rest almost intact. This approach, however, can only be followed if there are enough data as to provide information to discriminate between all the parameters.

2.3.1.2 Precise Point Positioning

As a particular application of the direct approach, it was developed at the Jet Propulsion Laboratory (JPL) a strategy that permits precise estimation using single receiver data and ancillary data processed somewhere else: the Precise Point Positioning, PPP ([35]). It has been mentioned before that the orbits and clock corrections broadcast by the satellites are not precise enough. For this reason, there are different organizations that process world-wide tracked data and estimate the orbits of the GPS satellites to precisions of 10 cm. The most widely used orbits come from the IGS and these orbits are a combination of solutions from different processing centers (JPL and CODE for instance). In addition, some centers provide corrections for the satellite clocks, accounting for the SA, to the same level of precision. The availability of orbits and clocks at the data processing stage allows to obtain accurate and precise positioning of steady or roving receivers without need for close-by stations to perform differences or to combine data for the estimation of common parameters: the PPP strategy decouples the unknowns for each station.

2.3.1.3 Characterization of the variables in Precise Point Positioning

We routinely follow the Precise Point Positioning approach and the estimation processing contained in the GISPY-OASIS II software for the estimation of tropospheric parameters. For the ionosphere, we use the GIST data preprocessor, to be described later. The input parameters in the GOA II process are precise satellite orbits, precise clock corrections, earth rotation parameters, tides and ocean loading effects and antenna phase patterns. The variables to estimate are:

- Station Position in XYZ coordinates: they are estimated as constants throughout the time span of the data (for steady receivers).

- Station Clock Error: while an atomic-quality clock could be modeled as a bias and a drift, this does not apply for the usual clock inside GPS receivers and it is best and most robust to model the clock as a stochastic process with white noise characteristics, to account for possible glitches and other random behaviour of the clock. This modeling is possible thanks to the huge amount of data (see [34] and [30] for a nice discussion on the topic).

- Atmospheric effects: the ionospheric delay is canceled using the ionospheric-free combination and the tropospheric delay is split into zenith component plus horizontal gradients. These parameters are modeled as random walk stochastic processes (see next section).

- Phase bias: a constant bias that accounts for the offset between the pseudorange and the phase. Each bias is constant only between phase breaks.

2.3.2 Atmospheric modeling in the GPS data processing

The atmosphere introduces delays (Section 2.2.5.2) that can be, for the zenith direction, of some 2.5 m in the ionospheric free combination and of some tens of meters in each frequency. This effect has to be removed for geodetic studies and extracted for atmospheric studies. In particular, for tomography, it is needed to compute the *slant delays*: the delay due to the ionosphere and the troposphere in the line-of-sight of the transmitting satellite. In this section we describe the parameters obtained when processing GPS data. Because we are dealing with only one pair satellite-station, the indices k and p are omitted for clarity.

2.3.2.1 Ionospheric Data

The ionospheric combinations of phase and pseudorange L_I and P_I already contain the information about the status of the ionosphere along the line of sight. They can be combined into carrier-phase smoothed measurements and reduce the level of noise. This noise is on the order of 10 cm, after the detection of all cycle slips and partition of the observations into arcs. Some of these arcs may be too short and should not be processed because there are not enough data to yield a good alignment [17].

2.3.2.2 Tropospheric Data

While for the ionosphere, a linear combination of observables provide the slant observation, the slant delay due to the neutral atmosphere cannot be directly estimated. In the regular GPS data processing, the slant tropospheric delay (L_t [‡]) is usually modeled as a zenith component plus horizontal gradients and discriminating between hydrostatic and wet components [36] (see Appendix A for a derivation of the expression):

$$\begin{aligned} L_t = m_h(e)\left[L_z^h + \cot e\left(\vec{L}_h^G \cdot \hat{\rho}(\phi)\right)\right] \\ +m_w(e)\left[L_z^w + \cot e\left(\vec{L}_w^G \cdot \hat{\rho}(\phi)\right)\right] \end{aligned} \tag{2.22}$$

where L_z^h and L_z^w are the hydrostatic and wet zenith delays, e and ϕ are the satellite elevation and azimuth as seen from the station, respectively; $\hat{\rho}(\phi)$ is the horizontal unit vector defining the direction of the projection of the ray on the horizontal plane; $\vec{L}_h^G$ and $\vec{L}_w^G$, are the hydrostatic and wet delay gradients; m_h and m_w are the hydrostatic and wet mapping functions. There have been different mapping functions proposed over time: the most simple case considers a flat Earth and simply relates the slant delay to the zenith through $\mathcal{M}(e) = 1/sin(e)$; however, it was soon realized that the Earth's curvature needed to be accounted for and this led to the Marini mapping function [37], which has been improved by adding parameters to account for the bending of the rays across standard or empirical atmospheres. Among the most widely used mapping functions are the ones from Niell [38] and Lanyi [39]. The former considers variations with seasons and location and the latter does not fully separate the wet and the hydrostatic components. In our processing we normally use the Niell mapping function.

The delay gradients are defined to be ([23], and using the notation of [40]; see appendix

[‡]in the literature it is usually written as L_t because it is understood that it is the effect measured on the phase, due to its higher precision; rigorously it is the carrier-phase smoothed pseudorange

A for a derivation of the expressions)

$$\vec{L}^G = 10^{-6} \int_0^\infty z\vec{\psi}(z)dz, \tag{2.23}$$

where $\vec{\psi}(z) = \nabla_{\vec{\rho}} N(\vec{\rho}, z)|_{\vec{\rho}=0}$ is the horizontal gradient of the refractivity, and z is the height above the surface.

The splitting between hydrostatic and wet components is due to the possibility to model the hydrostatic zenith component from surface measurements ([24]) and thus be removed from the total measurement.

$$L_z^h = (2.2779 \pm 0.0024)\frac{P_0}{f(\lambda, H)} \tag{2.24}$$

where P_0 is the surface pressure in milibars and $f(\lambda, H)$ is a function that accounts for the dependence of the gravity field on the latitude λ and the height above the ellipsoid H

$$f(\lambda, H) = 1 - 0.00266\cos 2\lambda - 0.00028H \tag{2.25}$$

The removal of the hydrostatic gradients is also necessary if we want to obtain the slant wet delays (SWD); these gradients, however, cannot be discriminated from the wet gradients in the GPS data processing and we need either numerical weather prediction (NWP) models or simplifying assumptions to extract them. In the regular processing using GOA II, hydrostatic and wet gradients are estimated as a unique horizontal gradient and only the zenith hydrostatic delay (ZHD or L_z^h) is a priori removed; the total tropospheric delay will be modeled as [29]:

$$L_t = m_w(e)L_z^w + m_\Delta(e)\cot e\left[\vec{L}^G \cdot \hat{\rho}(\phi)\right] \tag{2.26}$$

where m_Δ is the mapping function for the hydrostatic and wet gradients, and L_z^w and $\vec{L}^G$ are considered to be stochastic process (random walk). This decomposition allows to estimate three parameters (zenith plus the two components of the gradients) at a given time from the slant measurements of all the satellites in view (6 as a mean value). The output of the GPS data processing is the time series of the total zenith delay (the L_z^h is added back to the output solution) and the total gradients. They need to be combined to obtain the slant wet delays. We leave for Chapter 5 the description of how SWD are formed[§].

[§]Note that SWD can also be obtained by carefully modeling all the effects but the atmospheric effects and consider the SWD to be the post-fit residuals. This, however, would not allow to impose a smooth time variation on the tropospheric parameters and also, would neglect their correlations with other parameters