CHAPTER 2. PRINCIPLES OF EARTH OBSERVATION FROM SPACE

This chapter provides an overview of Earth observation from space, including its potential benefits and limitations. It describes basic concepts of orbits and the characteristics of Earth viewing from space. It also introduces the principles of remote-sensing.

2.1 ORBITS AND EARTH VIEWING FROM SPACE

The Earth can be observed from space from different orbits under various viewing conditions. The following issues are considered in this section:

(a) Satellite instrument field of view;
(b) Orbital period, geostationary orbit, observing cycle and repeat cycle;
(c) Orbital precession, Sun-synchronous orbits and drifting orbits;
(d) Elliptical orbits;
(e) Launchers and injection into orbit.

2.1.1 Satellite instrument field of view

The greatest advantage of observing from a satellite platform, rather than from the ground or a balloon, is the wide potential field of view (FOV). Satellite observing platforms usually orbit at a minimum height of 400 km. Often, they orbit much higher, some as far as the geostationary orbit (35,786 km). The FOV depends on the orbital height, the instrument configuration and the intended application. Those may limit the useful range of zenith angles (ζ) under which the Earth can be viewed. If the satellite FOV is characterized as the maximum ground distance potentially viewed from satellite height under a given zenith angle, the relationship set out in Figure 2.1 can be expressed as:

\[
\text{FOV} = 2R \delta \pi / 180 \sin (\zeta - \delta) = \frac{R}{H + R} \sin \zeta
\]

where \( R = 6371 \text{ km} \) (Earth’s radius), \( H = \) orbital height in km, and \( \delta = \) geocentric angle in degrees.

Table 2.1 presents values of the satellite FOV (in km) as a function of orbital height for typical values of zenith angle \( \zeta \). The corresponding geocentric angle \( \delta \) is also shown.

The potential satellite FOV may not be entirely covered by a single instrument. Either the sensing principle or the technological features of an instrument may set an upper limit to its FOV. For instance, radar altimeters can only operate in a nadir geometry. They therefore have

Figure 2.1. Field of view versus zenith angle \( \zeta \)
no proper FOV, except for the broadening of the beam due to diffraction. Very high-resolution
imagers usually have an FOV within a range of several tens of kilometres, as do synthetic aperture
radars (SARs).

The satellite motion enables the instrument to cover successive FOVs along the orbit. These
constitute a strip of observed Earth surface called a swath. The swath may be centred along
the sub-satellite track, or be parallel to it for side-looking instruments (e.g. SARs). For several
purposes (steerable pointing for emergencies, stereoscopy in association with successive orbits,
etc.), certain instruments with limited swath may tilt the swath to the side of the track within
what is called a field of regard. The swath width is a cross-track component of the actual FOV of
the instrument. The swath is not defined for instruments in geostationary orbit.

2.1.2 Orbital period, geostationary orbit, observing cycle and repeat cycle

The orbital height $H$ determines the orbital period $T$. The relationship is:

$$ T = a (1 + \frac{H}{R})^{\frac{3}{2}} $$

where $a = 84.47$ min ($T$ resulting in minutes).

The height, which corresponds to one sidereal day ($23$ h $56$ min $04$ s) is $35$ 786 km. A satellite
orbiting at this height is called geosynchronous. The orbit is called geostationary if the orbit lies
in the equatorial plane and is run eastward: the satellite appears steady compared to the Earth’s
surface on the nadir of the equatorial sub-satellite point.

For an inclined orbit with respect to the equatorial plane, the satellite will cross the Equator at
a certain longitude. After $T$ minutes, there will be another equatorial crossing at a longitude
displaced westward by the number of degrees that corresponds to the Earth’s rotation during the
orbital period. The difference of longitude (or space distance) between two successive equatorial
crossings in the same phase (descending or ascending) is called decalage. Together with the
instrument swath, decalage determines the time needed for a full Earth surface observation
(covering coverage) (Figure 2.2).

If the instrument swath is at least as large as the decalage, the coverage provided by the two
contiguous orbits is continuous. Therefore, the time needed for global coverage (observing
cycle) depends on the ratio between instrument swath and decalage.

Table 2.2 shows the period and corresponding decalage for the orbital height examples
in Table 2.1. In addition, observing cycles corresponding to several instrument swaths are
presented. Those swaths are associated with qualitative and quantitative use; that is with $70^\circ$ and

<table>
<thead>
<tr>
<th>Zenith angle for various applications</th>
<th>$H = 400$ km</th>
<th>$H = 600$ km</th>
<th>$H = 800$ km</th>
<th>$H = 35 786$ km</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\zeta = 90^\circ$ (horizon-to-horizon)</td>
<td>4 401 km</td>
<td>19.79°</td>
<td>5 326 km</td>
<td>23.95°</td>
</tr>
<tr>
<td>$\zeta = 85^\circ$ (telecommunications)</td>
<td>3 423 km</td>
<td>15.39°</td>
<td>4 322 km</td>
<td>19.43°</td>
</tr>
<tr>
<td>$\zeta = 70^\circ$ (qualitative use)</td>
<td>1 746 km</td>
<td>7.85°</td>
<td>2 405 km</td>
<td>10.82°</td>
</tr>
<tr>
<td>$\zeta = 60^\circ$ (quantitative use)</td>
<td>1 207 km</td>
<td>5.43°</td>
<td>1 707 km</td>
<td>7.68°</td>
</tr>
</tbody>
</table>
60° zenith angles, respectively. Observing time is halved for instruments capable of observation during the day and night. The decalage and observing cycle are not quoted for $H = 35,786$ km (geostationary altitude).

For a geostationary satellite, which continuously views the same area of the Earth’s surface, the observing cycle is only determined by instrument characteristics and may take a few minutes or less to complete, depending on the area scanned. Within the satellite’s area of coverage, geostationary observation is perfectly suited for continuous monitoring. For example, such monitoring is needed to detect instantaneous events like lightning strokes, or for high-frequency temporal sampling of rapidly evolving situations, such as active convection. However, the coverage excludes very high latitudes or any locations too far from the sub-satellite point. Table 2.1 shows that $\delta = 81.31^\circ$ is the maximum geocentric angle.

For a non-geostationary satellite, the orbit is said to have a repeat cycle if it overpasses the same track exactly after a given number of revolutions. During the timespan of a repeat cycle, the satellite track may shift from day to day, following a determined pattern that may exhibit certain periodicities called sub-cycles. Some sub-cycles may be of interest because distinct areas, relatively close to each other, are visited within short time intervals; other sub-cycles may be of interest because the areas covered are spatially adjacent.

If the orbit is Sun-synchronous (see section 2.1.3), the existence of a repeat cycle means that a whole number of revolutions can be completed in exactly a whole number of days. The orbital period determines $N$, the number of orbits that the satellite runs in 24 h. This is normally not an integer. In order to obtain a repeat cycle of $m$ days, the orbital period is adjusted to ensure that $N$ multiplied by $m$ is an integer. $N$ can then be expressed in the following form, where $n$ and $\ell$ are, respectively, the quotient and remainder of the integer division of “$N \cdot m$” by $m$:

$$N = n + \ell/m$$

where $n$, $\ell$ and $m$ are integers ($\ell < m$).

Table 2.2. Period, decalage and observing cycle for the orbits indicated in Table 2.1

<table>
<thead>
<tr>
<th>Orbital parameter</th>
<th>$H = 400$ km</th>
<th>$H = 600$ km</th>
<th>$H = 800$ km</th>
<th>$H = 35,786$ km</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period $T$</td>
<td>92.6 min</td>
<td>96.7 min</td>
<td>100.9 min</td>
<td>23 h 56 min 4 s</td>
</tr>
<tr>
<td>Decalage</td>
<td>2,570 km</td>
<td>2,690 km</td>
<td>2,800 km</td>
<td>0 km</td>
</tr>
<tr>
<td>Observing cycle for $\zeta = 70^\circ$ (day only)</td>
<td>35 h</td>
<td>27 h</td>
<td>23 h</td>
<td>From instrument</td>
</tr>
<tr>
<td>Observing cycle for $\zeta = 60^\circ$ (day only)</td>
<td>51 h</td>
<td>38 h</td>
<td>31 h</td>
<td>From instrument</td>
</tr>
<tr>
<td>Observing cycle for $\zeta = 70^\circ$ (day and night)</td>
<td>18 h</td>
<td>13 h</td>
<td>11 h</td>
<td>From instrument</td>
</tr>
<tr>
<td>Observing cycle for $\zeta = 60^\circ$ (day and night)</td>
<td>26 h</td>
<td>19 h</td>
<td>16 h</td>
<td>From instrument</td>
</tr>
</tbody>
</table>

Figure 2.2. Decalage between two successive orbits, and instrument swath
Equation 2.3 also applies to non-Sun-synchronous orbits. However, in such cases, the repeat cycle $m$ is no longer expressed in solar days of 24 h but must account for a slight correction due to the drift of the orbit. Table 2.3 provides examples of repeat cycles and main sub-cycles for a number of orbits.

Table 2.3. Repeat cycles and main sub-cycles for a number of orbits

<table>
<thead>
<tr>
<th>Orbital height</th>
<th>Sun-synchronous orbits</th>
<th>Non-Sun-synchronous orbit</th>
</tr>
</thead>
<tbody>
<tr>
<td>909 km (e.g. Landsat 1–3)</td>
<td>705 km (e.g. Landsat 4–8)</td>
<td>832 km (e.g. SPOT)</td>
</tr>
<tr>
<td>791 km (e.g. Envisat)</td>
<td>820 km (e.g. Metop)</td>
<td>1 336 km (e.g. JASON)</td>
</tr>
<tr>
<td>Period</td>
<td>103.2 min</td>
<td>98.9 min</td>
</tr>
<tr>
<td>No. of orbits/day</td>
<td>13 + 17/18</td>
<td>14 + 9/16</td>
</tr>
<tr>
<td>Cycle</td>
<td>18 days</td>
<td>16 days</td>
</tr>
<tr>
<td>Revolutions/cycle</td>
<td>251</td>
<td>233</td>
</tr>
<tr>
<td>Main sub-cycle(s)</td>
<td>1 day</td>
<td>7 days, 2 days</td>
</tr>
</tbody>
</table>

Note:
^a In the case of the Joint Altimetry Satellite Oceanography Network (JASON), which is not Sun-synchronous, the figures refer to a day of 23 h 48 min in duration, i.e. 0.99156 of the duration of a solar day.

Figure 2.3. Schematic evolution of the orbital track of early Landsat over a repeat cycle ($N = 13 + 17/18$, repeat cycle: 18 days, 251 revolutions/cycle)
An orbit with a repeat cycle is a necessary feature if a certain location needs to be viewed at fixed intervals under identical conditions. This is true of altimetric measurements for geodetic application, or of high-resolution land observation imagers, used to detect local variations.

Repeat cycles may be useful when the instrument swath is substantially narrower than the decalage and global coverage cannot be achieved in a single day. In that case, the sequence of coverage over successive days can be arranged to follow a certain logic if requested. That logic might be to provide regular progression, or to avoid biases due to unsuitable sampling.

Figure 2.3 shows the pattern evolution of orbital passes for an orbit with a one-day sub-cycle (such as for early Landsat). As Figure 2.3 shows, the one-day sub-cycle ensures that each day, the covered strip is adjacent to the one that was observed on the previous day. The width of the covered strip can be tuned to the instrument swath so as to avoid any gaps. The drawback is that, after the first few daily visits over or close to the target area, the next sequence of visits occurs only after the completion of the repeat cycle.

With the current Landsat, the temporal evolution of the orbit tracks during the repeat cycle (16 days) results in two main sub-cycles, as shown in Figure 2.4. The two-day sub-cycle ensures a shorter temporal gap, but the seven-day sub-cycle provides a closer geographical match.

Although the concept of repeat cycles and sub-cycles stems from the requirements placed on the use of narrow-swath instruments, including those with nadir-only viewing, orbits with sub-cycles may also be useful for relatively wide-swath instruments. For example, sounding instruments may have a swath as wide as several thousand kilometres. (For example, the Advanced

![Figure 2.4. Schematic evolution of the orbital track of current Landsat over a repeat cycle (N = 14 + 9/16, repeat cycle: 16 days, 233 revolutions/cycle). Three sub-cycles are shown: seven days (westbound), the main one, providing the closest observations in space; two days (eastbound), for closer observations in time; and nine days (eastbound), of marginal interest, the sum of the first two sub-cycles.](image)
Microwave Sounding Unit (AMSU) or the Infrared Atmospheric Sounding Interferometer (IASI) have swaths of over 2,200 km. However, the quality of the products retrieved is higher when closer to the nadir sub-track. Therefore, there is an interest in ensuring that the coverage provides a fair blend of higher and lower quality data. Orbits for the National Oceanic and Atmospheric Administration (NOAA) satellites and Meteorological Operational (Metop) satellites have a five-day sub-cycle, as shown in Figure 2.5 for Metop.

Repeat cycles and sub-cycles are convenient for several reasons, but a few drawbacks should also be noted:

(a) If the instrument swath is too narrow with respect to the decalage and the number of orbital passes during the repeat cycle, some areas will never be observed. An extreme case is a nadir-only viewing instrument such as an altimeter.

(b) The day-to-day sequence of observations from a repeat-cycle orbit may introduce sampling biases in the observations (a spurious wavelength corresponding to a repeat cycle or sub-cycles).

(c) Maintenance of the repeat cycle/sub-cycles requires costly satellite orbit control systems.

Therefore, if all instruments on board have a sufficiently wide swath, a repeat cycle or sub-cycle is generally not carried out.
2.1.3 Orbital precession, Sun-synchronous orbits and drifting orbits

The orbital plane can lie in the Earth’s equatorial plane or be inclined by \( \varepsilon \) degrees (see Figure 2.6).

For \( \varepsilon = 90^\circ \), the satellite follows a meridian line and the orbit is polar. This is very convenient for observing the Earth’s surface pole-to-pole.

The gravity field acting on the satellite is perpendicular to the geopotential surface at satellite altitude, which is slightly ellipsoid like the geoid. Where \( \varepsilon \neq 90^\circ \), the effect of these forces is a precession of the orbital plane around the polar axis. The precession rate \( \alpha \) is computed as:

\[
\alpha = -10.02 \cos \varepsilon \left(1 + \frac{H}{R}\right)^{-\frac{3}{2}} \text{ (degree/day)}
\]  

(2.4)

For a purely polar orbit (\( \varepsilon = 90^\circ \)), the precession rate is thus zero. The orbital plane has an invariant orientation with respect to the fixed stars. However, as the Earth rotates around the Sun over one year, the illumination conditions of the surface, as viewed by the satellite, change every day by \( 360/365 \) degrees; that is 59 min. An area viewed in daylight at noon on day \( t_0 \) will be viewed in dawn conditions on day \( t = t_0 + 3 \) months (Figure 2.7, left panel). For measurements in daylight, this would mean different observing conditions day after day, with seasonally-dependent observation times. In particular, when the Earth–Sun direction becomes perpendicular to the orbital plane, the illumination in dawn–dusk conditions makes many measurements impossible.

The orbital inclination \( \varepsilon \) can be set in such a way that the precession rate exactly matches the yearly revolution of the Earth around the Sun. By imposing the value \( \alpha = 360/365 \) (degree/day) in equation 2.4, it is found that the inclination \( \varepsilon_0 \) must satisfy:

![Figure 2.6. Definition of inclined orbit](image)

![Figure 2.7. Left: Pure polar orbit with changing local solar time (LST) throughout the year; right: Sun-synchronous orbit with fixed LST throughout the year](image)
An orbit that satisfies that condition is called Sun-synchronous. The negative value of \( \varepsilon_0 \) indicates that the orbit is retrograde with respect to the Earth’s rotation. The local solar time of the areas overflown by the satellite at a given latitude is constant across the whole year (see Figure 2.7, right panel). Table 2.4 presents the \( \varepsilon_0 \) values of a number of Sun-synchronous orbits as a function of orbital height.

### Table 2.4. Inclination values of Sun-synchronous orbits as a function of orbital height

<table>
<thead>
<tr>
<th>Height (km)</th>
<th>( \varepsilon_0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>400</td>
<td>97.02°</td>
</tr>
<tr>
<td>600</td>
<td>97.78°</td>
</tr>
<tr>
<td>800</td>
<td>98.60°</td>
</tr>
<tr>
<td>1,000</td>
<td>99.47°</td>
</tr>
<tr>
<td>1,200</td>
<td>100.41°</td>
</tr>
<tr>
<td>1,400</td>
<td>101.42°</td>
</tr>
</tbody>
</table>

Note that the deviation from the polar axis increases with orbital height. This is a drawback for high Sun-synchronous orbits: the poles might not be observed unless the instrument swath is wide enough. However, for relatively low orbital heights, the orbital plane is near-polar.

The most important feature of a Sun-synchronous orbit – the fixed local solar time – may be a disadvantage for certain types of measurements. Diurnally-varying phenomena (e.g. convective clouds, precipitation, radiation budget, sea level affected by astronomical tides) display biased sampling if observed from a Sun-synchronous satellite (i.e. at a fixed local solar time).

In general, satellites for operational meteorology, land observation and oceanography, with the exception of geodetic-quality altimetry, use a Sun-synchronous orbit. Scientific missions focused on processes affected by diurnal variation, which require unbiased sampling, may favour non-Sun-synchronous (drifting) orbits.

### 2.1.4 Elliptical orbits

The previous sections are applicable to circular orbits, which are by far the most widely used in Earth observation, particularly for Sun-synchronous and geostationary orbits. However, both near-polar low Earth orbits (LEO) and geostationary Earth orbits (GEO) have several limitations.

A near-polar LEO satellite provides global but infrequent coverage. Even if the instrument swath is as large as the decalage, thus providing contiguous coverage by consecutive orbits, one satellite can cover the whole of the Earth’s surface twice a day at most (or even once a day if sensing can be performed either in daylight only or only in night-time conditions). If more frequent global coverage is needed, additional satellites in complementary orbital planes are necessary (see Table 2.5).

It is clear from Table 2.5 that any observing cycle shorter than, for instance, three hours, would be extremely demanding, as it would require a constellation of LEO satellites in coordinated orbits.

### Table 2.5. Number of LEO satellites needed to achieve a required observing cycle (assumed height: \( H = 800 \text{ km} \))

<table>
<thead>
<tr>
<th>Instrument swath</th>
<th>Observing capability</th>
<th>Required observing cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>24 h</td>
</tr>
<tr>
<td>2,800 km</td>
<td>Only in daylight</td>
<td>1 sat</td>
</tr>
<tr>
<td></td>
<td>Night and day</td>
<td>0.5 sat</td>
</tr>
<tr>
<td>1,400 km</td>
<td>Only in daylight</td>
<td>2 sat</td>
</tr>
<tr>
<td></td>
<td>Night and day</td>
<td>1 sat</td>
</tr>
</tbody>
</table>
The coverage improves substantially for high latitudes (see Figure 2.2). For example, coverage is twice as frequent at 60° latitude than at the Equator. In polar regions, the frequency of coverage becomes close to the orbital period $T$ (i.e., ~100 min) or sub-hourly with more satellites.

A shorter observing cycle may be obtained by leaving aside the Sun-synchronous feature and adopting a lower inclination, but the coverage is then no longer global. Low-inclination orbits are used for monitoring the intertropical zones.

The GEO orbit provides observations at a rate that is limited only by the instrument. However, a constellation of about six satellites around the Equator is needed to cover all longitude sectors up to a latitude of at least 55°; the highest latitudes cannot be covered.

Some of these limitations can be mitigated by adopting an elliptical orbit. On an elliptical orbit, the satellite speed changes along the orbit; it is minimal around the apogee allowing more time for acquiring measurements from the overflown area. Elliptical orbits are usually optimized for specific purposes, particularly for space science, such as to collect in situ measurements up to very high altitudes by physically passing through the ionosphere and plasmasphere.

One problem of elliptical orbits is that, since the argument of the perigee is affected by the secular perturbation, the apogee occurs at latitudes that change with time. The secular perturbation can be compensated for if the orbital inclination is $\varepsilon = \sin^{-1} \left(\frac{4}{5}\right)^{1/2} \approx 63.4^\circ$. In this case, the apogee region where the satellite dwells for most of the time is stable. In that position, measurements can be taken very frequently, in a quasi-geostationary fashion.

Two orbits of this kind have been used for telecommunication satellites and are planned to be used for Earth observation: Molniya (Figure 2.8), which has a 12 h period and an apogee at 39 800 km; and Tundra, which has a 24 h period and an apogee at 48 300 km. In the Molniya orbit, the satellite is nearly geostationary for about 8 h of the 12 h period. In the Tundra orbit, it is nearly geostationary for about 16 h of the 24 h period.

The Molniya and Tundra orbits only serve one hemisphere. In addition, the 8 h or 12 h quasi-geostationary observing area is centred on a specific local solar time. If all latitudes above 60° have to be covered 24 h a day, three Molniya satellites or two Tundra satellites are required. An interesting variant is the three-apogee orbit with a 16 h period and an apogee at 43 500 km. Table 2.6 presents the main features of Molniya and Tundra orbits. In the apogee position, which is useful for frequent sampling, the height of the satellite exceeds GEO height.
Launchers and injection into orbit

Satellites are injected into orbit by a launcher, which has to perform the following functions:

(a) To host the satellite in the fairing, where vital functions for the satellite are ensured. When in the fairing, the satellite is stowed in a compact configuration to minimize volume occupancy and to be protected against the effects of acceleration.

(b) To bring the satellite to orbit. In order to minimize the total mass brought to high altitudes, the launcher is generally structured by stages. The first stage, which is the heaviest since it has to provide the maximum thrust for lift-off, is separated early. The fairing is released at an appropriate altitude. A further one or two stages are fired and separated in sequence.

(c) To release the satellite. For satellites in LEO circular orbits, the launcher releases the satellite directly on the final orbit. For elliptical orbits, the launcher releases the satellite at the perigee and provides it with a last acceleration to acquire the energy corresponding to the intended orbit.

When in orbit, the satellite deploys its solar panels and starts autonomous operations. One of the operations is to reach final orbit by activating its propulsion system. In the case of a geostationary orbit, the satellite is released at a perigee in an elliptical orbit whose apogee is 35 786 km, and is equipped with an apogee boost motor. The motor, which uses solid, hybrid or more often, liquid propellant (a liquid apogee motor), is fired at the apogee to provide the acceleration necessary to circularize the orbit (see Figure 2.9).

Table 2.6. Main features of Molniya and Tundra orbits (apogee and perigee heights can be slightly adapted to need)

<table>
<thead>
<tr>
<th>Orbit type</th>
<th>Inclination</th>
<th>Period</th>
<th>Apogee</th>
<th>Perigee</th>
<th>Coverage (from one satellite)</th>
<th>Sats for hemispheric coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molniya</td>
<td>63.4°</td>
<td>12 h</td>
<td>~39 800 km</td>
<td>~1 000 km</td>
<td>Visible over 2 positions for ~8 h</td>
<td>3</td>
</tr>
<tr>
<td>Tundra</td>
<td>63.4°</td>
<td>24 h</td>
<td>~48 300 km</td>
<td>~24 000 km</td>
<td>Visible over 1 position for ~16 h</td>
<td>2</td>
</tr>
</tbody>
</table>

Figure 2.9. Achievement of GEO
Launching a satellite is a complicated and costly exercise. In order to optimize cost-effectiveness, a launch is often shared by several satellites. In this case, the task of the last stage is extended to release the various satellites at different times to separate the orbits. The respective platform propulsion systems will thereafter achieve the transfer to the final orbits.

Constellations of minisatellites, whose individual launches would be extremely uneconomical, are a good example of the usefulness of multiple launches. One effective launching strategy is known as the Walker Delta Pattern. The orbital inclination \( \varepsilon \) of all satellites must be the same. Either the launcher releases the satellites at different altitudes, or the satellite propulsion systems spread them into different altitudes. Therefore, each orbit will have a different precession rate, according to equation 2.4, and the orbital phases will differentiate as time elapses. When the orbits of all the satellites are appropriately spaced, the platform propulsion systems will bring each of the satellites to the desired altitude \( H \). Equation 2.4 shows that, in practice, the strategy works well for relatively low inclinations and relatively low altitudes, otherwise the time needed to deploy the constellation becomes too long. For example, the Constellation Observing System for Meteorology, Ionosphere and Climate (COSMIC) (six satellites, \( H = 800 \text{ km}, \varepsilon = 71^\circ \)) took one year to deploy.

### 2.2 PRINCIPLES OF REMOTE-SENSING

Earth observation from space is mainly performed by exploiting electromagnetic radiation. The few exceptions are in situ measurements at platform level (of gravity field, magnetic field, electric field and charged particle density in the solar wind). This section deals with remote-sensing of the Earth and focuses on:

(a) The electromagnetic spectrum and the ranges used for remote-sensing;
(b) The basic laws of interaction between electromagnetic radiation and matter;
(c) Observations in the atmospheric windows;
(d) Observations in absorption bands;
(e) Limb sounding and radio occultation;
(f) Active sensing.

#### 2.2.1 The electromagnetic spectrum and the ranges used for remote-sensing

The spectrum of electromagnetic radiation as observed from space (with nadir viewing) is shown in Figure 2.10. The displayed range (0.2 \( \mu \text{m} \) to 3 cm (or 10 GHz)) includes all that is used for remote-sensing from space. The variation in wavelength is due to the interposed atmosphere, with transmissivity ranging from 1 (atmospheric window) to 0 (full opacity due to total atmospheric absorption).

![Figure 2.10. Spectrum (transmissivity) of electromagnetic radiation as observed from space with nadir viewing; range: 0.2 \( \mu \text{m} \) to 3 cm](image)
Table 2.7 presents definitions of the subdivisions of the spectrum that are generally accepted, though not standardized. In addition to the commonly used wavelength $\lambda$ and frequency $\nu$, the wave number $\nu^* = 1/\lambda$, mostly used in spectroscopy, is also quoted.

A finer subdivision of the MW range and nearby FIR, used for radar but also by extension, for passive radiometry, is provided in Table 2.8.

The overall spectrum shown in Figure 2.10 comprises five distinct regions, each with rather different features.

Table 2.7. Bands of the electromagnetic spectrum exploited for remote-sensing

<table>
<thead>
<tr>
<th>Spectrum subdivision</th>
<th>Wavelength range</th>
<th>Wave number $\nu^* = 1/\lambda$</th>
<th>Frequency $\nu = \frac{c}{\lambda}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UV Ultraviolet</td>
<td>0.01–0.38 $\mu$m</td>
<td>26 320–1 000 000 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>B Blue</td>
<td>0.436 $\mu$m</td>
<td>22 935 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>G Green</td>
<td>0.546 $\mu$m</td>
<td>18 315 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>R Red</td>
<td>0.700 $\mu$m</td>
<td>14 285 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>VIS Visible</td>
<td>0.38–0.78 $\mu$m</td>
<td>12 820–26 320 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>NIR Near infrared</td>
<td>0.78–1.30 $\mu$m</td>
<td>7 690–12 820 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>VNIR Visible and near infrared (VIS + NIR)</td>
<td>0.38–1.3 $\mu$m</td>
<td>7 690–26 320 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>SWIR Short-wave infrared</td>
<td>1.3–3.0 $\mu$m</td>
<td>3 330–7 690 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>SW Short wave</td>
<td>0.2–4.0 $\mu$m</td>
<td>2 500–50 000 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>LW Long wave</td>
<td>4–100 $\mu$m</td>
<td>100–2 500 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>MWIR Medium-wave infrared</td>
<td>3.0–6.0 $\mu$m</td>
<td>1 665–3 330 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>TIR Thermal infrared</td>
<td>6.0–15.0 $\mu$m</td>
<td>665–1 665 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>IR Infrared (MWIR + TIR)</td>
<td>3–15 $\mu$m</td>
<td>665–3 330 cm$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>FIR Far infrared</td>
<td>15 $\mu$m–1 mm</td>
<td>10–665 cm$^{-1}$</td>
<td>300–20 000 GHz</td>
</tr>
<tr>
<td>Sub-mm Submillimetre wave (part of FIR)</td>
<td>0.1–1 mm</td>
<td>10–100 cm$^{-1}$</td>
<td>300–3 000 GHz</td>
</tr>
<tr>
<td>Mm Millimetre wave (part of MW)</td>
<td>1–10 mm</td>
<td>1–10 cm$^{-1}$</td>
<td>30–300 GHz</td>
</tr>
<tr>
<td>MW Microwave</td>
<td>0.1–30 cm</td>
<td>0.033–10 cm$^{-1}$</td>
<td>1–300 GHz</td>
</tr>
</tbody>
</table>

Table 2.8. Bands used in radar technology (according to the American Society for Photogrammetry and Remote-sensing)

<table>
<thead>
<tr>
<th>Band</th>
<th>Frequency range</th>
<th>Wavelength range</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>220–390 MHz</td>
<td>77–136 cm</td>
</tr>
<tr>
<td>UHF</td>
<td>300–1 000 MHz</td>
<td>30–100 cm</td>
</tr>
<tr>
<td>L</td>
<td>1–2 GHz</td>
<td>15–30 cm</td>
</tr>
<tr>
<td>S</td>
<td>2–4 GHz</td>
<td>7.5–15 cm</td>
</tr>
<tr>
<td>C</td>
<td>4–8 GHz</td>
<td>3.75–7.5 cm</td>
</tr>
<tr>
<td>X</td>
<td>8–12.5 GHz</td>
<td>2.4–3.75 cm</td>
</tr>
<tr>
<td>$K_\alpha$</td>
<td>12.5–18 GHz</td>
<td>1.67–2.4 cm</td>
</tr>
<tr>
<td>$K_c$</td>
<td>18–26.5 GHz</td>
<td>1.1–1.67 cm</td>
</tr>
<tr>
<td>$K_u$</td>
<td>26.5–40 GHz</td>
<td>0.75–1.18 cm</td>
</tr>
<tr>
<td>V</td>
<td>40–75 GHz</td>
<td>4.0–7.5 mm</td>
</tr>
<tr>
<td>W</td>
<td>75–110 GHz</td>
<td>2.75–4.0 mm</td>
</tr>
</tbody>
</table>
In the UV region, atmospheric absorption is strong, mainly due to the major air constituents (nitrogen (N₂) and oxygen (O₂)) and trace gases (the most important being ozone (O₃)). The Earth’s surface cannot be observed in that spectral region. The radiation source for remote-sensing consists of reflected solar radiation.

The VIS, NIR and SWIR regions, from 0.4 to 3.0 μm and, in some cases, up to 4.0 μm, can be sensed by means of reflected solar radiation. This range includes several transparent regions (windows) and many absorption bands (see Figure 2.11).

In the MWIR and TIR regions, from 4 to 15 μm, the radiation source consists of the thermal emission from the Earth’s surface and the atmosphere, which is mainly driven by water vapour and carbon dioxide absorption/emission. These are important contributors to the greenhouse effect. This thermal emission combined with the main atmospheric window enables the planet’s thermal equilibrium to be maintained at agreeable values (see Figure 2.12).

Figure 2.11. Atmospheric spectrum in the range 0.4 to 4.0 μm. It includes several windows and absorption bands from carbon monoxide (CO, about 2.3 μm), carbon dioxide (CO₂, about 1.6, 2.1 and 2.8 μm), methane (CH₄, about 2.3 and 3.4 μm), several oxygen bands (O₂, mainly about 0.77 μm), some nitrogen (N₂) and ozone (O₃) bands, and many important bands of water vapour (H₂O, mainly 0.94, 1.13, 1.37, 1.8 and 2.7 μm). Also shown is the molecular continuum, which prevents using UV for Earth surface and low atmosphere sensing from space.

Figure 2.12. Atmospheric spectrum in the range 3.33 to 16.67 μm. The main atmospheric windows are in the ranges 3.7 to 4.0 μm and 10 to 12 μm. There are large absorption bands from water vapour (H₂O) and carbon dioxide (CO₂). Other species are: ozone (O₃, about 9.7 μm), methane (CH₄, about 7.7 μm), carbon monoxide (CO, about 4.6 μm) and nitrous oxide (N₂O, about 4.5 and 7.7 μm).
The next spectral region, the Far IR, ranges from 15 \( \mu m \) to 1 mm (or 300 GHz). It is fully opaque because of the water vapour continuum. In that region, which is difficult to explore because of the lack of efficient detection techniques, there are absorption lines of several important species, such as hydroxyl radical (OH), known as a “cleaner” of the atmosphere, and hydrogen chloride (HCl), a reservoir species that releases ozone-aggressive chlorine. Hydroxyl radical and hydrogen chloride are only observable in the FIR (e.g. about 120 \( \mu m \approx 2500 \) GHz and 480 \( \mu m \approx 625 \) GHz, respectively).

In the MW range, from 1 to 300 GHz, the radiation source consists of the thermal emission from the Earth’s surface and the atmosphere. The atmospheric spectrum, starting from 2 GHz and extending to submillimetre frequencies up to 1000 GHz, is shown in Figure 2.13.

In the portion of the MW range where the atmosphere is more transparent, (i.e. at frequencies below about 100 GHz), the wavelengths exceed 3 mm and so are much larger than cloud drop size, except for precipitating clouds. Therefore, the MW range is used for observing the Earth’s surface or atmospheric properties in nearly all weather conditions.

Active sensing is conditioned by technology and, in the case of MW, by radio-frequency spectrum regulations. Radar makes use of MW, while lidar makes use of optical wavelengths where suitable sources (crystals) are available. Table 2.9 presents a few commonly used radar
frequencies and lidar wavelengths. As a comparison, the table also lists frequencies used by
the global navigation satellite system (GNSS) and associated radio-occultation sensing of the
atmosphere (see Part III, Chapter 3, 3.2.7).

### 2.2.2 Basic laws of interaction between electromagnetic radiation and matter

The macroscopic properties of a condensed body at thermodynamic equilibrium that is not
affected by chemical or nuclear reactions are summarized in terms of electromagnetic radiation
by three coefficients linked by the following equation:

\[
\rho (\lambda, T, \zeta, \varphi) + \tau (\lambda, T, \zeta, \varphi) + \varepsilon (\lambda, T, \zeta, \varphi) = 1
\]

In this equation, \(\rho(\lambda, T, \zeta, \varphi)\) denotes reflectivity that is the ratio of the backscattered radiation
\(I(\lambda, T, \zeta, \varphi)\) to the incident radiation \(I(\lambda)\) that crosses the body; and \(\varepsilon(\lambda, T, \zeta, \varphi)\) is the fraction of \(I(\lambda)\) that is absorbed by the body – it is
called emissivity for reasons explained below. The three coefficients depend on the radiation
wavelength \(\lambda\), the body temperature \(T\), and the observing geometry \(\zeta, \varphi\) (zenith and azimuth
angles, respectively).

A body that is not reflecting and is totally opaque to radiation on any wavelength (where
\(\rho = \tau = 0\), and thus \(\varepsilon = 1\)) is called a black body. It radiates at any temperature, \(T\) and over the full \(\lambda\)
or \(\nu\) spectrum, according to Planck’s law:

\[
B(\lambda, T) = \frac{2\pi\lambda^2}{h c} \frac{1}{e^{\frac{h\lambda}{kT}} - 1} \quad \text{or} \quad B(\nu, T) = \frac{2\pi\nu^3}{c^2} \frac{h}{e^{\frac{h\nu}{kT}} - 1}
\]

where:

\(h = 6.626 \cdot 10^{-34}\) J s (which is the Planck constant);
\(c = 2.99793 \cdot 10^8\) m s\(^{-1}\) (which is the speed of light in a vacuum); and
\(k = 1.38044 \cdot 10^{-23}\) J K\(^{-1}\) (which is the Boltzmann constant).

\(B(\lambda, T)\) (or \(B(\nu, T)\)) is the radiative power per unit surface over the hemisphere per unit of
wavelength (or frequency). The power radiated per unit of solid angle is \(B/\pi\). The Planck function
for temperatures of 6 000 K and 273.16 K, which are representative of the Sun and the Earth’s
surfaces respectively, is shown in Figure 2.14.

<table>
<thead>
<tr>
<th>Observation</th>
<th>Instrument</th>
<th>Frequency or wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radar</td>
<td>Sea-surface wind</td>
<td>C band (~5.3 GHz) or K(_u) band (~13.4 GHz)</td>
</tr>
<tr>
<td>Ocean topography</td>
<td>Altimeter</td>
<td>C band (~5.3 GHz) + K(_u) band (~13.6 GHz)</td>
</tr>
<tr>
<td>Cloud and precipitation</td>
<td>Rain radar, cloud radar</td>
<td>K(_u) band (~13.6 GHz) and/or K(_a) band (~35.5 GHz) or W band (~94 GHz)</td>
</tr>
<tr>
<td>Imagery</td>
<td>Synthetic aperture radar</td>
<td>L band (~1.3 GHz) or C band (~5.4 GHz) or X band (~9.6 GHz)</td>
</tr>
<tr>
<td>Lidar</td>
<td>Clear-air wind</td>
<td>UV-lidar (355 nm)</td>
</tr>
<tr>
<td></td>
<td>Doppler lidar</td>
<td>UV-lidar (355 nm), VNIR-lidar (532 + 1 064 nm)</td>
</tr>
<tr>
<td></td>
<td>Aerosol, cloud top</td>
<td>UV-lidar (355 nm), VNIR-lidar (532 + 1 064 nm)</td>
</tr>
<tr>
<td></td>
<td>Ice-sheet topography</td>
<td>Altimeter</td>
</tr>
<tr>
<td>Radio occultation</td>
<td>Atmospheric refraction in LEO</td>
<td>L band: ~1 580 + ~1 250 + ~1 180 GHz (GPS, GLONASS and Galileo)</td>
</tr>
</tbody>
</table>
The power radiated over the full spectrum is:

\[
W(T) = \int_0^{\infty} B(\lambda, T) \cdot d\lambda = \sigma \cdot T^4
\]

Stefan-Boltzmann law \hspace{1cm} (2.8)

where \(\sigma = 5.6681 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}\).

The two curves in Figure 2.14 illustrate that the Sun’s and Earth’s surfaces have very different radiative powers. However, after scaling down the Sun curve by the square of the distance between the Sun and Earth, the two integrated areas become comparable, which reflects the Earth’s radiative balance. If the upper curve in the figure is scaled, it is easy to see that the solar radiation on Earth is very small for \(\lambda > 4 \mu\text{m}\), while Earth radiation is negligible for \(\lambda < 3 \mu\text{m}\). There are significant amounts of solar and Earth radiation in the narrow interval between 3 and 4 \(\mu\text{m}\).

A major difference between the two curves in Figure 2.14, after scaling, is the wavelength \(\lambda_{\text{max}}\) where the maximum emission occurs. This is given by:

\[
\lambda_{\text{max}} = \frac{b}{T}\quad \text{with } b = 0.0028981 \text{ m K} \quad \text{Wien law} \hspace{1cm} (2.9)
\]

Because of the double-logarithmic scale in Figure 2.14, it is difficult to appreciate how sharp the Planck function is around \(\lambda_{\text{max}}\). In the case of solar radiation \((T = 6,000 \text{ K})\) the peak emission occurs around \(\lambda_{\text{max}} = 0.5 \mu\text{m}\) and most of that power lies in the 0.2–3.0 \(\mu\text{m}\) range. In the case of terrestrial radiation \((T = 273.16 \text{ K}, \text{i.e. } T = 0^\circ\text{C})\) the peak is around \(\lambda_{\text{max}} = 10 \mu\text{m}\) and most of the power lies in the 3–50 \(\mu\text{m}\) range.

Another interesting feature of the Planck function is that, when moving from \(\lambda_{\text{max}}\) to shorter wavelengths, radiative power dramatically decreases, whereas when moving to longer wavelengths, the decrease is more gradual (approximately two thirds of power occurs at wavelengths longer than \(\lambda_{\text{max}}\)). For very long wavelengths, or low frequencies, such as those in the MW range, where the argument of the exponential in the Planck function is \(\hbar \nu / kT \ll 1\), the term \(e^{\hbar \nu / kT}\) becomes \(\approx 1 + \hbar \nu / kT\) and the Planck function thus reduces to:

\[
B(\nu, T) = \frac{2\pi \hbar}{c^2} \frac{\nu^2}{e^{\hbar \nu / kT} - 1} T \quad \text{Rayleigh-Jeans law} \hspace{1cm} (2.10)
\]

Because of this relationship, radiation measurements in the MW range can be considered as temperature measurements and can be expressed as a brightness temperature \((T_B)\) in temperature units. While the total radiative power changes with temperature in proportion to \(T^4\) according to the Stefan-Boltzmann law (equation 2.8), that power changes in a linear way in the MW portion of the spectrum. Conversely, when moving towards shorter wavelengths, the Planck
function is increasingly dependent on temperature. Terrestrial radiation varies approximately with $T^4$ in the main TIR window around 11 $\mu$m and with $T^2$ in the window around 3.7 $\mu$m. This is an interesting feature for remote-sensing, since it means that sensitivity to high temperatures is higher at 3.7 $\mu$m. Conversely, these shorter wavelengths are less sensitive to low temperatures.

For example:

(a) For a body at 220 K (such as a cloud top in the upper troposphere), the radiation at 11 $\mu$m is ~1,200 times greater than at 3.7 $\mu$m; but at 300 K (surface), it is only ~130 times;

(b) The sensitivity to temperature changes, (i.e. $(\partial B/\partial T)/B$) is three times higher at 3.7 $\mu$m than at 11 $\mu$m.

As a consequence, 3.7 $\mu$m is well suited for surface observations, is optimal for fire detection, and is less useful for high clouds.

The relationship represented by equations 2.7–2.10 is only valid for a black body ($\varepsilon = 1$). For a common body, the relationship can be established through the following conceptual experiment. If a number of bodies in an isolated system only exchange radiation among themselves, it could be assumed that, after a transient time, each of them would reach thermodynamic equilibrium, where the radiation each absorbs according to its absorption coefficient $\varepsilon(\lambda, T)$ is equal to the power it radiates $P(\lambda, T)$. That is, $P(\lambda, T)/\varepsilon(\lambda, T) = \text{constant}$.

Assuming that one of the bodies is a perfect black body, this yields:

$$P(\lambda, T) = \varepsilon(\lambda, T) \cdot B(\lambda, T)$$  \hspace{1cm} (Kirchhoff principle)

This shows that the absorption coefficient introduced in equation 2.5 also controls the body’s emission properties, hence the name emissivity. Equation 2.11 indicates two important consequences:

(a) At any wavelength and temperature, a body cannot radiate more than a black body at the same wavelength and temperature;

(b) A body can radiate only at wavelengths at which it can also absorb.

Emissivity $\varepsilon$ is a function of wavelength and, to a lesser extent, temperature. For certain bodies, $\varepsilon$ may be constant over large portions of the spectrum. If it is constant over the whole spectrum, the body is called grey. The shape of the radiated power $P(\lambda, T)$ is then exactly like $B(\lambda, T)$, although it is damped by a factor $\varepsilon$. The Wien law (equation 2.9) applies unchanged. The Stefan-Boltzmann law becomes $W(T) = \varepsilon \cdot \sigma \cdot T^4$.

The Kirchhoff principle also applies to gaseous materials. Therefore, spectral lines of atmospheric gases are generally (but not always) relevant to both absorption and emission.

### 2.2.3 Observations in the atmospheric windows

#### 2.2.3.1 Emerging radiation

Atmospheric windows are spectral regions where the atmosphere is nearly transparent. There is no region where the atmosphere is fully transparent. All regions have some residual disturbance from species that have a continuum, the most common of which is water vapour in the IR, MW and, to some extent, the SW ranges. Another factor, particularly in SW, is scattering from dry air molecules (mostly $N_2$ and $O_2$) and aerosols. Ultimately, the most transparent windows are:

(a) In the SW (Figure 2.11): 0.5–0.9 $\mu$m, 1.6–1.7 $\mu$m and 2.0–2.3 $\mu$m;

(b) In the IR (Figure 2.12): 3.5–4.0 $\mu$m and 10–12 $\mu$m;

(c) In the MW (Figure 2.13): 80–100 GHz, 25–50 GHz and below 20 GHz.
Equation 2.5 indicates that the coefficients $\rho$ (reflectivity), $\tau$ (transmissivity) and $\varepsilon$ (emissivity) not only depend on radiation wavelength $\lambda$ and body temperature $T$, but also on the geometric condition (zenith angle $\zeta$ and azimuth angle $\phi$) of the satellite platform with respect to the body.

In order to simplify the discussion, the following conditions are assumed: vertical viewing from the satellite (Figure 2.15), flat surfaces and radiation towards the zenith (irradiance).

One component, $\tau(\lambda, T) \cdot I(\lambda)$, is transmitted radiation through the body. It can be found where the body is not totally opaque and where there is a radiation source below it in the opposite hemisphere compared to the satellite.

The component $\varepsilon(\lambda, T) \cdot B(\lambda, T)$ is emitted radiation, expressed through the Kirchhoff principle (equation 2.11). It is always present unless the body is at absolute thermal zero. It is not present in spectral regions, where the body does not absorb.

The component $\rho(\lambda, T) \cdot I(\lambda, \theta)$ is reflected radiation. It is found where there is a radiation source in the same hemisphere as the satellite. Figure 2.15 indicates the Sun but, in the MW range, where solar radiation is virtually zero, the full hemispheric sky radiates as a black body, at temperature $T = 2.725$ K, where the maximum power occurs at $\lambda_{\text{max}} = 1.9$ mm (160 GHz). The notation $I(\lambda, \theta)$ indicates that the power of the incoming radiation depends on the angle of incidence $\theta$. Normally for the Sun, $I(\lambda, \theta) = S(\lambda) \cdot \cos \zeta$, where $S(\lambda)$ denotes incoming power with the Sun at its zenith.

Taking account of all components, the radiation reaching the instrument on the satellite can be expressed as:

$$I(\lambda, T) = \tau(\lambda, T) \cdot I(\lambda) + \varepsilon(\lambda, T) \cdot B(\lambda, T) + \rho(\lambda, T) \cdot I(\lambda, \theta)$$

There is considerable variation along the spectrum, so that the significance of the measurement is reasonably stable only within narrow bandwidths around a specific wavelength (channels). However, some general information is also contained in a wider range of wavelengths (VIS + NIR + SWIR; MWIR + TIR; MW).

2.2.3.2 **Measurements in the visible, near-infrared and short-wave infrared range**

In this range, there is no thermal emission from the Earth ($B = 0$). Focusing first on the Earth’s surface (land and ocean), the transmitted radiation is zero, since there is no source below the Earth’s surface. Also when considering that reflectivity is nearly independent of body temperature, equation 2.12 reduces to:

$$I_{\text{sw}}(\lambda) = \rho(\lambda) \cdot S(\lambda) \cdot \cos \zeta$$

(2.13)
Since the solar spectrum and observing geometry are known, the information carried by a measurement in a SW channel is uniquely associated to surface reflectivity. Many geophysical variables (vegetation parameters, ocean colour, texture of land) may be estimated by measuring reflectivity at several wavelengths. However, clouds are the objects that are most in evidence in SW. Equation 2.13 is not strictly correct for a cloud surface, since radiation from the underlying surface can transmit through them. However, this effect has a limited impact given:

(a) The total transmissivity through (downward and upward) the atmosphere;
(b) That the cloud is low;
(c) The originating radiation source (the Sun) is stronger than any surface below the cloud;
(d) The reflectivity of the underlying surface is low (except for sand desert, snow and ice);
(e) The reflectivity of clouds $\rho$ is generally higher than any other terrestrial surface.

Therefore, equation 2.13 is an approximation applicable to most clouds, the exceptions being thin clouds such as cirrus, especially on the bright background of sand deserts, ice and snow.

Using SW-reflected radiation for quantitative purposes is not an easy exercise, since reflectivity is generally anisotropic. The simplest case occurs when the body, whatever the direction of the incoming radiation, and for any azimuth, homogeneously redistributes the reflected radiation from the zenith according to a cosine-law. This is called Lambertian reflection. Fortunately, most Earth surfaces observed from space at relatively large scale appear rather flat and rough, so that Lambertian diffusion may be a good approximation. In many cases, however, the body exhibits a bidirectional reflectance distribution function that should be measured a priori by observations under different viewing directions and for different directions of the incoming radiation. The final computation of the irradiance towards space requires hemispheric integration.

Using SW for Earth observation in atmospheric windows requires spectral sampling at several wavelengths (channels) since at any one wavelength, several bodies may have signatures, and one body may have a signature at several wavelengths. A multichannel capability is therefore necessary to distinguish and simultaneously retrieve the different properties of different bodies. For instance, clouds and snow have identical reflectance in VIS at 0.65 $\mu$m, but very different reflectance at 1.6 $\mu$m. In addition, channel bandwidths must be appropriate to their purpose. The most stringent are for ocean colour ($\Delta \lambda \approx 10$ nm), and then vegetation ($\Delta \lambda \approx 20$ nm), whereas for other surface features and for clouds, bandwidths of several tens of nanometres are sufficient.

Another feature that affects the quantitative use of VIS + NIR + SWIR is polarization: specular reflection tends to be privileged, with damping of the vertical component of the electric field. The Stokes vector, which in the SW range consists of three components (polarization in three directions with phase differences of 120°), fully describes the electric field, and so provides important information about the body’s properties. Multipolarization is important for the observation of those bodies that do not have strong multispectral signatures. Typical examples would be aerosols and cirrus clouds (elongated ice crystals).

2.2.3.3 Measurements in the medium-wave infrared and thermal infrared range

In the 4–15 $\mu$m range, solar radiation is virtually zero. On the Earth’s surface (land and ocean), transmitted radiation is also zero. Furthermore, considering that emissivity is nearly independent from body temperature, equation 2.12 reduces to:

$$ I_{\text{IR}}(\lambda, T) = e(\lambda) \cdot B(\lambda, T) $$

(2.14)

Equation 2.14 is also approximately valid for clouds, since the transmissivity of clouds in IR is rather low (with the exception of thin cirrus).
The emissivity of most land surfaces, and certainly the ocean, is close to 1, with small variance. Therefore the information acquired by a measurement in an IR channel is closely associated with the Planck function (equation 2.7) or for a given wavelength, with the body’s temperature.

At a specified wavelength \( \lambda \) or for a narrow channel \( \Delta \lambda \) around \( \lambda \), the Planck function (equation 2.7) may be easily inverted to retrieve the temperature, \( T \):

\[
T(\lambda) = \frac{hc}{\lambda k \ln \left(1 + \frac{2\pi hc^2}{\lambda^2 B(\lambda)}\right)}
\]

However, \( T \) will not be the body’s true temperature unless \( \varepsilon = 1 \). If the body’s emissivity is known, and the channel bandwidth \( \Delta \lambda \) is sufficiently narrow to enable \( \varepsilon \) to be considered as constant, an emissivity correction can be applied to the measured quantity (by inverting \( B(\lambda, T) = I_{IR}(\lambda, T) / \varepsilon(\lambda) \)), and the body’s temperature can be measured. Otherwise, by inverting the measured quantity \( B(\lambda, T) = I_{IR}(\lambda, T) \), a temperature \( T_{BB} \) (equivalent black-body temperature) is obtained that is lower than the body’s true temperature.

For a large variety of bodies where \( \varepsilon \) is close to 1, radiance observation in a TIR atmospheric window enables rather accurate temperature measurement. It is particularly accurate in the case of the sea, whose emissivity is close to 0.98. However, emissivity is not the only effect that needs to be corrected. As previously mentioned, the atmospheric windows are not perfectly transparent. For instance, the main window in TIR, 10–12 \( \mu \)m, is contaminated by the water vapour continuum, particularly on the long-wave side. One way to reduce this disturbance is to split the window into two channels, generally 10.3–11.3 \( \mu \)m and 11.5–12.5 \( \mu \)m. Differential absorption is then used to estimate a correction (total column water vapour can also be estimated as a by-product).

As a result of those effects, equation 2.15 indicates a dependence of the retrieved temperature on the wavelength (although as there is only one body temperature, that dependence should not exist). The spread of values with wavelength implies how certain information might be explained. For instance, by comparing \( T_{BB} \) measured at 3.7 \( \mu \)m and 11 \( \mu \)m it is possible to explain the difference in terms of different emissivity, or in terms of different contamination of the measurement from clouds.

It is important to note that the 3.7 \( \mu \)m window behaves very differently in daylight and at night. In daylight, it is strongly contaminated by reflected solar radiation, which needs to be subtracted before using the channel for quantitative thermal emission estimates. As previously noted, the 3.7 \( \mu \)m window is much more sensitive to high temperatures than the 11 \( \mu \)m window. However, the 3.7 \( \mu \)m window is of little use for low temperatures, such as those found in cloud tops in the upper troposphere. The differential response to the temperatures of the 3.7 and 11 \( \mu \)m windows can also be used for detecting fog at night.

As regards clouds, equation 2.14 is still approximately valid. However, with the exception of very thick clouds (such as cumulus or nimbus-stratus), emissivity is substantially lower than unity. Equivalent black-body temperature substantially underestimates true temperature, and a correction must be applied to account for low emissivity. The usual method is to couple the window channel with a channel that is strongly sensitive to water vapour. The difference between the two \( T_{BB} \) values indicates the cloud emissivity: the larger the difference, the lower the emissivity.

The penetration of infrared radiation in clouds is very low. Measured temperature refers to the top surface, and information about the interior of clouds is poor, especially for dense clouds. However, the temperature of cloud tops that is in equilibrium with air temperature at the same level is very important, because it indicates the altitude of the cloud in the troposphere, and therefore the cloud type.

Information derived from black-body temperature about the cloud-top level is often inaccurate. With thin clouds, such as thin cirrus, the background surface is much warmer than the air at
cloud level. As a result, surface radiation transmitted through the cloud adds to the cloud emission, the cloud appears warmer, and the assigned level is underestimated. Conversely, thick cirrus with high emissivity are observed as very cold, and may be confused with cumulonimbus.

In order to resolve such ambiguities, it is useful to plot VIS brightness and IR temperatures as bi-dimensional histograms (Figure 2.16). By using only one band (the projection of the 2D pattern on one axis), several clusters would be unresolved. By contrast, this example shows that 10 different objects can be identified through multi-band analysis. This is a simple example using an old instrument (a very high resolution radiometer (VHRR)) and only two channels in VIS and IR. Current multi-band analysis techniques can operate with many more channels.

2.2.3.4 Measurements in the microwave range

In the MW range, solar radiation is virtually zero, but there is a diffuse source of incoming radiation: the sky. Transmitted radiance is zero with regard to the Earth’s surface, including land and sea. There are only two contributions: thermal emission and reflected radiation. Those are controlled by emissivity and reflectivity coefficients which, since \( \tau = 0 \), are linked by the condition \( \varepsilon + \rho = 1 \) (i.e. \( \rho = 1 - \varepsilon \)). By expressing radiative power in units of temperature according to the Rayleigh-Jeans equation (2.10), observed brightness temperature, \( T_B(v) \), can be rendered:

\[
T_B(v) = \varepsilon(v) \cdot T + [1 - \varepsilon(v)] \cdot T_{\text{sky}}(v)
\]  

\( T_{\text{sky}}(\nu) \), sky brightness temperature, is composed of background cosmic radiation and the contribution of precipitating clouds; it changes with frequency. In the main window and under non-precipitating conditions, where \( \nu \sim 40 \text{ GHz} \), \( T_{\text{sky}}(\nu) \) may be \( \sim 140 \text{ K} \). Under heavy precipitation conditions, \( T_{\text{sky}} \) may reach values as large as \( \sim 250 \text{ K} \).

Figure 2.16. Scatterplot of VIS (0.65 \( \mu \)m) versus IR (11.5 \( \mu \)m), enabling the classification of 10 bodies. If projected onto one of either two axes, several clusters would be unresolved.

The impact of $T_{\text{sky}}(\nu)$ strongly depends on the value of the emissivity, $\varepsilon$. Sea and land have substantially different $\varepsilon$ values.

The emissivity of the sea in the MW range is very low: $\varepsilon \approx 0.5$. As a result, the two components in equation 2.16 have equal weight. In the absence of precipitating clouds, $T_{\text{sky}}$ is low and known: the measurement can therefore be associated with sea-surface temperature. The optimal frequency for sea-surface temperature is about 5 GHz (see Figure 2.17) where $T_{\text{sky}}$ is much smaller than that at 40 GHz. The measurement is rather accurate and is also applicable in all weather, since the wavelength ($\lambda = 6 \text{ cm}$) is much longer than any rain droplet. The 5 GHz sea-surface signal intensity is representative of the temperature of a few millimetres of the deep-water layer (sub-skin), that should be compared to a few tens of micrometres in the case of IR (skin temperature). At higher frequencies, $T_{\text{sky}}$ strongly increases, especially in the presence of heavy precipitation. The high reflectivity value ($1 - \varepsilon$) is such that the observation is mainly an indicator of precipitation.

Over land, emissivity is close to unity. The second component in equation 2.16 is therefore ineffective, and precipitation is poorly detected. At higher frequencies (~90 GHz ($\lambda = 3 \text{ mm}$)), radiation welling up from the surface is scattered by large droplets, and even more so by ice crystals in clouds. As a result, radiation reaching the satellite is decreased.

Polarization can also be used for measuring precipitation. Radiation reflected from the sea is strongly polarized: when crossing a precipitating cloud, it undergoes depolarization that can be measured to infer precipitation. The differential polarization can also be exploited over land, since the emerging radiation scattered from droplets and ice crystals is polarized.

Observation under several polarizations may also be useful, regardless of the objective of precipitation measurement. Differential polarization is sensitive to surface roughness – an effect which must be taken into account when measuring sea-surface temperature. This can also be used to infer wind speed over the sea, as indicated in Figure 2.17.

Figure 2.17 also indicates that MW radiation is sensitive to ocean salinity, but only at very low frequencies, typically about 1.4 GHz (L band). The figure also shows that in order to measure ocean salinity, it is necessary to account for sea-surface temperature and wind speed (or roughness). Similarly to salinity, there is a water vapour absorption band that contaminates the observation of temperature, wind and liquid cloud (precipitation). That band can also be used to infer total column water vapour (precipitable water) over the sea. In summary, different variables may have different signatures on various channels in the MW as well as the optical fields: multi-channel analysis is therefore needed.

![Figure 2.17. Sensitivity (defined as $\Delta T_B/\Delta P_i$) of MW frequencies to several geophysical variables ($P_i$).](image-url)
Due to the totally different emissivity values of sea and land surfaces, the most obvious feature in a MW image is the land/sea boundary. Since the emissivity of ice is close to unity, sea ice is also an obvious observable in all weather conditions. MW images are particularly useful for geographical regions that are often overcast. In cases where emissivity is close to unity over land, a decrease in emissivity indicates the presence of water on the surface. This is because the emissivity of a body is controlled by its dielectric constant: water on land is a salt solution, which increases conductivity and thus decreases emissivity. This effect can be exploited to measure surface soil moisture and snow properties.

Soil moisture measurements can be rather accurate on bare soil, but can decrease in accuracy as vegetation increases. In order to penetrate vegetation, and to measure soil moisture at root level, very low frequencies must be used, either in L or P band. At higher frequencies (above 10 GHz), sensitivity to soil moisture is only significant if disturbance by vegetation is accounted for.

Two properties of snow are detectable in the MW range: surface melting conditions and, in the case of shallow snowpack, water equivalent. In the latter case, relatively high frequencies are preferred, as snow tends to be transparent at low frequencies. However, saturation can occur at very high frequency signals in the upper layers of the snowpack. Several frequencies with different penetration depths are therefore needed.

### 2.2.4 Observations in absorption bands

#### 2.2.4.1 The radiative transfer equation

In an atmospheric absorption band, each layer of thickness $dz$ absorbs radiation coming from below and re-emits it. Assuming zero reflectivity of the atmosphere in IR, the atmospheric transmittance from a height $z$ to a satellite altitude $H$ is given by:

$$\tau(\lambda, z) = e^{-\int_z^H \varepsilon(\lambda, z) N(z) \, dz} \quad (2.17)$$

where $N(z)$ is the concentration of the absorbing gas.

The radiative contribution of an atmospheric layer of thickness $dz$, at height $z$ and associated with a transmittance change of $d\tau$ $(I,z)$, is:

$$dI(\lambda, z) = B(\lambda, T(z)) \cdot d\tau(\lambda, z).$$

The radiation from the total atmospheric column to the satellite is:

$$I(\lambda) = \int_{\tau(\lambda, z_s)}^{\tau(\lambda, z_s)} B(\lambda, T(z)) \cdot d\tau(\lambda, z) \quad (2.18)$$

where $z_s$ is the height of the Earth’s surface.

A terrestrial contribution should be added, attenuated by the total atmospheric transmittance. In addition, the weighting function can be defined as:

$$K(\lambda, z) = \frac{d\tau(\lambda, z)}{dz} \quad (2.19)$$

The combined radiation reaching the satellite is given by the radiative transfer equation:

$$I(\lambda) = B(\lambda, T_s) \cdot \tau(\lambda, z_s) + \int_{z_s}^{H} B(\lambda, T(z)) \cdot K(\lambda, z) \cdot dz \quad (2.20)$$

Figure 2.18 shows that the transmittance (equation 2.17) tends to 1 as height $z$ increases (i.e. as the thickness of the atmospheric layer between height $z$ and satellite altitude $H$ decreases). This is accompanied by a decrease in emissivity $\varepsilon$ and a decrease in concentration $N$ of the absorbing gas. The weighting functions have peak values that correspond to the inflection point of the transmittance function. A simple way to read equation 2.20 is that each atmospheric layer of thickness $dz$ contributes to the radiation reaching the satellite according both to its temperature (through the Planck function), and also to its effectiveness to contribute, as quantified by the weighting function. The weighting function depends on the concentration of the absorbing gas and the strength of absorption/emission ($\varepsilon$). The shape is such that low atmospheric layers are
penalized because of absorption by the upper layers, and high layers are penalized because of their low concentration of the absorbing gas. The atmospheric layer that exhibits the greatest change in transmittance is usually the layer that contributes the most.

2.2.4.2 Profile retrieval

The inversion of equation 2.20 is not a trivial matter. It is a Fredholm equation of the second kind, for which the existence or the uniqueness of the solution are not mathematically guaranteed. In the present case, the existence is granted by nature. To ensure uniqueness of the solution it is necessary to add constraints to it, since the problem is ill-conditioned. Many methods have been developed since profile sounding from space began. Some are statistical and linear, others are physical and non-linear, while others are a combination of both.

The primary objective is to invert equation 2.20 in order to retrieve an atmospheric temperature profile. This is only possible if the transmittance function is known in advance, which implies working in the absorption bands of a gas that has a known and stable concentration profile. In the IR range, CO$_2$ has such a profile in the bands around 4.3 and 15 $\mu$m (see Figure 2.12). As previously noted, the 4.3 $\mu$m band is more sensitive to high temperature, and is thus representative of the lower troposphere. However, that band may be contaminated by radiation from other species, and in daytime, the tail of the solar black-body curve (> 4 $\mu$m) cannot be disregarded. The 15 $\mu$m band is spectrally more pure, but is somewhat contaminated by the water vapour continuum. The transmittances therefore need to be corrected, either a priori by using external information, or a posteriori, by iterating after the water vapour profile has been retrieved.

The next step is to retrieve the water vapour profile. Once CO$_2$ absorption band channels have been used to retrieve the temperature profile, H$_2$O absorption band channels are used. The main band is centred around 6.3 $\mu$m, and responds well to high temperature (in the low- to mid- troposphere). For climate monitoring, it is important to measure water vapour in the upper troposphere. But that requires using an 18 $\mu$m band, which is technologically difficult to construct due to a lack of efficient detectors in the FIR range.

It is not easy to retrieve a water vapour profile or, more generally, to retrieve the concentration profile of an absorbing gas. The weighting functions of the absorbing gas peak at varying altitudes in the atmosphere, depending on concentration and remote-sensing frequency.
In addition, retrieval is intrinsically inaccurate because the transmittance function to be inverted (equation 2.17) defines the content of a thin layer at altitude \( z \) as the difference between the respective content of two very thick layers: \( H \) to \( z \) and \( H \) to \( (z - \Delta z) \). In other words, a small number is derived by calculating the difference between two large numbers.

Further difficulties arise where clouds are present. If the instantaneous field of view (IFOV) is entirely filled by a cloud with uniform features, a profile may still be retrieved by the same method, although it will only cover the atmosphere above the cloud. If only a fraction \( \eta \) of the IFOV is filled by a cloud of emissivity \( \varepsilon_{\text{cloud}} \), the transfer equation becomes:

\[
\eta \cdot \varepsilon_{\text{cloud}} \cdot \eta = (1 - \varepsilon_{\text{cloud}} \cdot \eta) \cdot I_{\text{clear}}(\lambda) + \varepsilon_{\text{cloud}} \cdot \eta \cdot I_{\text{cloud}}(\lambda)
\]

(2.21)

where \( \varepsilon_{\text{cloud}} \cdot \eta \) is the effective cover.

Several methods deal with the effects of clouds. One starts the retrieval process by using the channels with weighting functions that peak above the cloud-top level to achieve a first-guess profile. The first guess is then iterated by changing effective cover values, until the measurements in all other channels fit best. Another method compares a number of nearby IFOVs on the assumption that the signals differ only because of the different fractional covers \( \eta \), and then extrapolates to zero \( \eta \).

In any event, it is acknowledged that, when cloud cover in the IFOV exceeds approximately 20%, no attempt should be made to retrieve profiles in IR. The IFOV of sounding instruments used to be several tens of kilometres. Fortunately, that has now been reduced to the order of 10 km, so that the probability of finding a substantial number of cloud-free IFOVs in a given area is much higher.

The problem of clouds is greatly alleviated in the MW range, where sounding is possible for all weather conditions except heavy rain. The species of well-known and constant concentration used for temperature profile retrieval is \( O_2 \), with absorption bands in the 50–70 GHz range and at about 118 GHz (not yet used from a satellite). For water vapour, the 183 GHz band can be used effectively. The 22 GHz band provides a weak signal that can provide total column-integrated amount over the sea. There are other absorption bands for temperature and water vapour at higher frequencies, but the radiative effect of the water vapour continuum makes it impossible to observe the troposphere using those spectral bands.

The transfer equation in the MW range is essentially a simpler version of the IR equation (2.20): instead of the Planck function (equation 2.7), it is possible to use the Rayleigh-Jeans approximation with linear temperature dependence (equation 2.10).

The question may arise as to why the MW band is not exclusively used for temperature and humidity sounding, as it performs in nearly all weather conditions. This is because vertical resolution requires a high sensitivity to temperature variations (with height). Vertical resolution is best in the 4.3 \( \mu \)m band, where the Planck function varies roughly in relation to \( T^{12} \). In the 15 \( \mu \)m band, there is less sensitivity because the Planck function varies in relation to \( T^5 \). In the MW range, since \( B \) is a linear function of \( T \) (see equation 2.10), sensitivity \( (\partial B/\partial T)/B \) varies in relation to \( T^{-1} \) (it decreases as temperature increases). One interesting feature of the various bands is that, whereas the 4.3 \( \mu \)m band is well suited to the lower troposphere, and the 15 \( \mu \)m band is well suited to the middle and high troposphere, the MW band at 57 GHz is better suited to the stratosphere.

Vertical resolution is crucial to temperature and humidity sounding. Figure 2.18 shows an example in which the weighting functions are rather broad. That implies that the degrees of freedom (the number of independent pieces of information) are limited. Weighting functions become narrower when the spectral resolution of the instrument improves. Figure 2.18 relates to a radiometer with only seven channels and a low resolving power (\( \lambda/\Delta \lambda \approx 100 \)); its vertical resolution is \( -1.5–2 \) km in the mid-troposphere. Current sounding instruments (spectrometers) have thousands of channels and a higher resolving power (\( \lambda/\Delta \lambda \approx 1 \ 000 \)); their vertical resolution is less than 1 km in the mid-troposphere. Further increasing the resolving power to \( \lambda/\Delta \lambda \approx 10 \ 000 \) would not improve the vertical resolution of temperature and humidity profiles, but would enable single lines of trace gases to be observed for atmospheric chemistry purposes.
Current instruments for the MW range have already reached maximum vertical resolution performance: it cannot improve beyond ~1.5 km in the mid-troposphere, and will be worse in the lower troposphere because of strong contamination from the ground.

In the short-wave range (UV, VIS, NIR, SWIR) absorption band measurements are mostly used for atmospheric chemistry through spectroscopic methods. The radiative transfer equation is more complicated than equation 2.20. Instead of thermal radiation as described by Planck’s law, the more complex process of scattering is used. Retrieving geophysical variables relies on modelling rather than explicit equations. As well as being used in atmospheric chemistry, the analysis of absorption bands is used for other purposes (see the spectrum in Figure 2.11), such as deriving:

(a) Atmospheric pressure at the Earth’s surface: this is derived from estimates of the total column of oxygen in the band around 0.77 µm compared with nearby windows. It is one of the very few approaches available to measure surface pressure from space. Accuracy is limited by the scattering effect of aerosols, implying that the measurement also provides information on aerosols.

(b) Cloud-top height: this is derived from a deficit that arises when the total column of oxygen is measured; the deficit itself is the result of cloud masking the lower part of the column. In principle, this is more accurate than calculating the cloud-top height from the equivalent black-body temperature in IR, correcting for cloud emissivity, and transforming temperature into height using a temperature profile.

(c) Lightning: a very narrow bandwidth channel at 0.774 µm is used. Strong absorption from oxygen obscures the Earth’s surface and enables flashes to be detected even in daylight. The intensity and number of flashes in a given period over a given area are representative of convection, and so serve as a proxy for precipitation. In addition, lightning activity causes NOx to be generated in the atmosphere, and reflects the Earth’s electric field.

(d) Total column water vapour: the signal in one or more of the water vapour bands (about 0.94 or 1.37 µm) is compared with a signal from nearby windows. This can be more accurate than using IR or MW profiling.

2.2.4.3 **Limb sounding**

Figure 2.18 shows how weighting functions become broader as height increases. This indicates that the vertical resolution of temperature and humidity profiles using passive IR or MW radiometry degrades with increasing altitude. The resolution obtained from using spectrometers is currently considered adequate (~1 km) in the mid-troposphere; but it becomes marginal (~2 km) at tropopause level, where much better resolution is required. In the stratosphere, the vertical resolution degrades further and rapidly becomes unusable. Two techniques offer help: limb sounding (including through occultation of the Sun, moon or stars) and radio occultation.

In cross-nadir sensing mode, vertical resolution is determined by the sharpness of the weighting functions, which is in turn controlled by spectral resolution. In limb mode, vertical resolution is determined by mechanical scanning, i.e. by the instrument IFOV across the atmosphere when viewed transversally in the Earth’s limb region (Figure 2.19). Vertical resolution depends on the step change rate, which is tuned to the instrument viewing aperture and to the intensity of available radiation. It is generally set to between one and three kilometres. Horizontal resolution is relatively inaccurate, since the measurement is integrated over a large optical path, as shown in Figure 2.19. The total optical path may be thousands of kilometres long, but the effective path, once weighted by atmospheric density, extends to some 300–500 km around the tangent point.

The sources of radiation are solar radiation reflected from the atmosphere, or atmospheric thermal radiation in the IR or MW ranges. In general, limb observations address not only temperature and humidity profiles, but also trace gases for atmospheric chemistry purposes.

In the short-wave range (UV, VIS, NIR, SWIR), the atmosphere can be scanned by directly pointing to the Sun while the Sun is setting or rising (occultation). Observation is conducted
by measuring the damping of spectral lines in the solar spectrum. Sun occultation has the great advantage of avoiding any mechanical movement of the instrument telescope, and any calibration. That is because the spectra measured during occultation are compared to the solar spectrum measured shortly before (or after) occultation under the same conditions. One disadvantage is that, at least for polar-orbiting satellites, coverage is limited to high latitudes, where a satellite can observe sunrise or sunset as it enters or leaves the night arc of its orbit. More extended coverage is possible by using moon occultation, while all latitudes can be covered through occultation of the stars. However, less radiation is available in those cases.

2.2.5 **Active sensing**

The sensing methods described above assume that the sources for remote-sensing are reflected solar radiation and the Earth’s emitted thermal radiation (plus other minor sources, such as background sky radiation in MW and the moon or stars in occultation). These natural sources enable passive sensing, for which observation wavelengths are largely determined by natural targets. In active sensing, the source is artificial and the sensing wavelength is not entirely driven by the physical properties of the target. Instead, the wavelength can be chosen, while also taking account of signal generation and propagation constraints. The following active sensing principles are considered:

(a) Radio occultation (for high vertical resolution profiles of temperature and humidity);

(b) Radar (for altimetry, scatterometry, cloud and precipitation, and imagery);

(c) Lidar (for clouds and aerosols, air motion, altimetry, and atmospheric chemistry).

2.2.5.1 **Radio occultation**

Radio occultation is one of a number of limb-sounding techniques. It has a totally different approach compared to passive radiometric techniques. An artificial source (in this case, the signal from a navigation satellite: GPS, GLONASS, Galileo or Compass) is tracked by a receiver on a satellite in LEO (Figure 2.20).

The change of propagation direction due to refraction by the crossed atmosphere (bending angle \( \alpha \)) is converted into a phase shift. The shift is accurately measured and then converted into a refractivity profile during the occultation process, which lasts approximately 90 s.
The refractivity is linked to the atmospheric variables as follows:

\[ N = (n - 1) \cdot 10^6 = 77.6 \cdot \frac{p}{T} + 3.75 \cdot 10^5 \cdot \frac{p_w}{T^2} \]  

(2.22)

where \( N \) is refractivity, \( n \) is refractive index, \( p \) is dry-air pressure, \( p_w \) is water vapour partial pressure, and \( T \) is temperature. The coefficients for \( p \) and \( p_w \) are given in hPa, and in kelvin for \( T \).

The phase shift is ultimately a time measurement, one of the most accurate measurements in physics. The other measurement is the distance between satellites. Since time and distance are fundamental metric quantities, radio occultation provides absolute measurements (not requiring calibration): this is a very attractive feature in terms of climate monitoring. In fact, long-term observations of radio occultation are considered a benchmark among methods of detecting climate change.

Radio occultation data are difficult to process for two reasons: first, the position of the tangent point (see Figure 2.20, right hand panel) moves during the profile measurements; and second, pressure, temperature and humidity are not measured independently. Therefore, 4D assimilation into a numerical weather prediction model is needed. It is less complex to retrieve temperatures from the upper troposphere and stratosphere, since water vapour content is very low. Temperature retrieval is similarly straightforward in the lower troposphere, where water vapour is responsible for most variance.

The vertical resolution of radio occultation profiles in the upper troposphere and stratosphere, (approximately 0.5–1.0 km) cannot be matched by cross-nadir viewing IR or MW measurements (1.5–2 km). In addition, the frequencies used (L band: see Table 2.9) are insensitive to clouds, even if precipitating. As a result, although essentially performed in limb mode, the measurement can be extended down to the Earth’s surface, in order to observe, for example, atmospheric discontinuities such as the top of the planetary boundary layer. Furthermore, radio occultation is one of the few methods that can infer surface pressure: this is done by correlating the height of the tropopause and the air pressure at ground level.

In order to account for the signal rotation induced by the ionosphere, transmissions from navigation satellites exploit at least two nearby frequencies. As by-products of the correction process, information relevant to space weather is obtained, such as total electron content and electron density profile.
2.2.5.2 Radar

Radars (radio detection and ranging) transmit pulsed signals to the object to be observed, and collect the backscattered signal. In essence, radars measure distance (or range) and the backscattered power resulting from the radar reflectivity or radar cross-section of the body.

The radar equation may be written in different forms. The easiest to understand is:

\[ P_s = \frac{P_t \cdot G}{4\pi \cdot r^2} \cdot \frac{\lambda^2}{4\pi} \cdot \sigma \cdot A_{\text{eff}} \quad (2.23) \]

where \( P_s \) is the power backscattered to the antenna, \( P_t \) is the power transmitted by the antenna, \( G \) is the antenna gain, \( A_{\text{eff}} \) is the effective area of the radar receiving antenna, \( r \) is the distance, and \( \sigma \) is the radar cross-section. Therefore:

- \( P_t \cdot G / (4\pi r^2) \) is the power hitting the target at distance \( r \);
- \( P_t \cdot G / (4\pi r^2) \cdot \sigma \) is the power reflected by the target;
- \( P_t \cdot G / (4\pi r^2) \cdot \sigma / (4\pi r^2) \) is the fraction of (isotropically) reflected power reaching back the antenna.

The antenna gain may be expressed as \( G = 4\pi \cdot A_{\text{eff}} / \lambda^2 \) (a relation that derives directly from the diffraction law). Expressing the effective area of the radar’s receiving antenna \( A_{\text{eff}} \) from this alternative expression and inserting it into equation 2.23 yields:

\[ P_s = \frac{P_t \cdot G^2 \cdot \lambda^2}{64\pi^3 \cdot r^4} \cdot \sigma \quad (2.24) \]

Different types of radar favour the measurement of either range accuracy (altimeters) or reflectivity/cross-section accuracy (scatterometers). Radar for clouds and precipitation focus on both range (for vertical profiling) and reflectivity. One feature that can be emphasized is image resolution by synthetic aperture radar.

Radar altimetry

The main purpose of altimetry is to measure the sea level and to map it so as to determine ocean dynamic topography. The radar characteristics are optimized to enhance range measurement as much as possible. Sea level is measured in terms of the time taken for a radar pulse to reach the sea surface and return to the satellite. Because sea level is computed as distance from the satellite, the satellite needs to be located with extreme accuracy. One or more of the following systems is used to ensure precise orbits:

(a) Laser tracking of the satellite by ground stations and laser-reflecting mirrors on the satellite;
(b) Radio positioning based on networks of ground transmitting-receiving stations and a transponder on the satellite;
(c) A GPS receiver aboard the satellite.

One drawback of radar altimetry is that viewing must be limited to nadir-only; otherwise echoes from surrounding areas interfere with time analysis. As a result, the observing cycle is very long. Corrections to the signal are requested to account for: ionospheric rotation (the two frequencies used are: ~13.6 GHz (main) and ~5.3 GHz (support)); and water vapour (a co-aligned MW radiometer is used at ~23 GHz (main) and ~35 GHz and/or ~19 GHz (support)).

In addition to measuring the range, an altimeter also records and analyses fluctuations and measures the intensity of the echo. The observations provided are:

(a) Significant wave height: derived from analysis of the spread in time of the collected echoes;
(b) Sea level: derived from filtering wave-related fluctuations and considering the instantaneous satellite altitude with respect to the geoid;
(c) Wind speed: derived from analysis of the fluctuation of the intensity of the echoes;
(d) Improved knowledge of the geoid: derived from long-term statistics of observed sea levels;
(e) Total electron content: derived as a by-product of the correction for ionospheric rotation.

Radar scatterometry

Unlike radar altimetry, where the focus is on range measurement, radar scatterometry optimizes the accuracy of the measured radar cross-section \( \sigma \) (see equation 2.24), which is often normalized and called \( \sigma^0 \) (sigma-naught). While the ranging subsystem may be absent from the instrument, calibration must be extremely accurate.

The radar cross-section is a function of the target’s dielectric property, the viewing geometry and the incident radiation (wavelength, polarization). Scatterometers are mainly used to derive sea-surface wind. The target is capillary waves, which are closely associated with wind stress. Sigma-naught changes with wind speed, relative wind direction and sight line. By measuring \( \sigma^0 \) under several azimuth angles, both speed and direction can be determined.

The relationship between \( \sigma^0 \) and wind is complicated: the practical solution is empirical or semi-empirical. Furthermore, it is not a unique relationship in terms of direction: with two viewing angles, several ambiguities remain (fewer ambiguities remain with three angles). When \( \sigma^0 \) values are directly assimilated into a numerical weather prediction model that accounts for wave–atmosphere interaction, ambiguities are solved by the model.

The differences between wind measurements taken with scatterometers and those taken with passive MW radiometers can be summarized as follows: (i) passive MW generally provides information on wind speed only; information on direction can only be acquired if several radiometric channels are equipped with full polarization capability; (ii) information from scatterometry is generally better quality, especially for low wind speeds (less than ~3 m/s); however, for high speeds (more than ~20 m/s) passive MW may perform better.

Primarily designed for sea-surface wind, scatterometers provide several kinds of observation:

(a) Sea-surface wind in all weather conditions (C band) or nearly all weather conditions (Ku band);
(b) Air pressure over the sea surface (achieved by applying geostrophic relations to wind maps);
(c) Soil moisture in scarcely vegetated areas (using C band, and occasionally Ku band);
(d) Leaf area index or total biomass in dense vegetation (forest);
(e) Ice type (age, roughness) at the polar caps;
(f) Snow water equivalent (for which Ku band is preferred).

Cloud and precipitation radar

While the radar altimeter focuses on ranging and the radar scatterometer focuses on the radar cross-section, cloud and precipitation radar emphasizes both. Ranging is necessary to measure the vertical profile of cloud particles, while \( \sigma \) is required to infer the concentration and size of reflecting particles. However, the accuracy required for ranging in order to obtain a precipitation profile is of the order of 100 m, instead of 1 cm for altimetry.

For rain droplets, provided that their diameter \( D \) is less than \( \lambda/10 \) (i.e. in Rayleigh-scattering conditions), the radar cross-section is:
The total backscattered radiation received by the radar is the sum of all reflectors of all diameters in the IFOV. Assuming a Marshall-Palmer distribution \( (N_0 \cdot e^{-\lambda D}) \) for the particle diameters, total reflectivity can be expressed as:

\[
Z = \pi \frac{5}{\lambda^4} \int_0^{D_{\text{max}}} K^2 \cdot N_0 \cdot e^{-\lambda D} \cdot D^6 \cdot dD \quad \text{(valid for } D \ll \lambda) \tag{2.26}
\]

Equations 2.25 and 2.26 are generally applicable to ground-based meteorological radar that use S band (~10 cm) or C band (~5 cm). They are to be compared with \( D \sim 0.5 \text{ cm} \), which is typical of precipitating clouds. Due to frequency regulations, the frequencies that can be exploited by a spaceborne radar are \( \sim 14 \text{ GHz} (~2 \text{ cm}) \), \( \sim 35 \text{ GHz} (~0.9 \text{ cm}) \) and \( \sim 94 \text{ GHz} (~0.3 \text{ cm}) \). Therefore, equations 2.25 and 2.26 are not fully applicable, but must be corrected in a complex way to account for Mie scattering conditions.

Once the reflectivity \( Z \) is measured, there are several ways to convert \( Z \) into precipitation rate \( R \). First, it is necessary to infer the precipitation rate at the surface. That cannot be directly measured from space, but must instead be derived from measured properties in the vertical column associated with the precipitation profile.

Cloud and precipitation radar is the only technique that can provide measurements of the cloud base, an important variable for aeronautical meteorology and for climate. The accuracy and reliability of the measurement depends on the radar frequency; the radar must also penetrate the full cloud thickness. From an operational viewpoint, cloud and precipitation radars have several disadvantages, particularly their limited swath, which prevents frequent observing cycles. And so, although passive MW imagery continues to be used as the basis for frequent precipitation observation, the accuracy of precipitation data from passive MW radiometry still needs to be improved. The continuing availability of at least one radar in space is necessary for “calibrating” the system of passive MW radiometers, along the lines of the concept for the Global Precipitation Measurement mission.

### Synthetic aperture radar

In the MW range, spatial resolution is limited by diffraction. For a side-looking radar that views \( \theta^\circ \) off nadir, at orbital height \( H \), with an antenna diameter \( L \), and assuming a flat surface, the IFOV is:

\[
\text{IFOV} = 1.24 \cdot \frac{H \cdot c}{\cos \theta \cdot L \cdot v} \tag{2.27}
\]

Table 2.10 shows the IFOV and \( L \) relationship for radar in several bands, assuming that \( \theta = 23^\circ \) and \( H = 700 \text{ km} \) (parameters of the SAR on SeaSat, Figure 2.21). It illustrates that the requirements for IFOV = 1 km would be very difficult to meet, and requirements for IFOV = 100 m or below would be impossible with a real-aperture antenna concept.

<table>
<thead>
<tr>
<th></th>
<th><strong>L band</strong> ((-1.3 \text{ GHz}))</th>
<th><strong>C band</strong> ((-5.4 \text{ GHz}))</th>
<th><strong>X band</strong> ((-9.6 \text{ GHz}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFOV for ( L = 1 \text{ m} )</td>
<td>220 km</td>
<td>60 km</td>
<td>30 km</td>
</tr>
<tr>
<td>Required ( L ) for IFOV = 1 km</td>
<td>220 m</td>
<td>52 m</td>
<td>30 m</td>
</tr>
<tr>
<td>Required ( L ) for IFOV = 100 m</td>
<td>2 200 m</td>
<td>520 m</td>
<td>300 m</td>
</tr>
<tr>
<td>Required ( L ) for IFOV = 10 m</td>
<td>22 000 m</td>
<td>5 200 m</td>
<td>3 000 m</td>
</tr>
</tbody>
</table>
In the SAR concept (Figure 2.21) the antenna is elongated along the satellite motion. Its narrow dimensions determine the swath and are parallel to the sub-satellite track. The longer side determines the area where signals are to be analysed. The radar footprint corresponds to that of a real-aperture antenna, while the situation is varied for the resolution elements in the field (pixels). The pixels across the swath are at different distances from the satellite. The satellite can locate them due to its ability to distinguish small-range changes. The echoes from the pixels ahead of the sub-satellite cross-track line are affected by a positive Doppler shift (where a frequency is higher than the one transmitted); the echoes from the pixels behind that line undergo a negative Doppler shift. By capturing the moment of shift inversion, the pixels can be assigned to a location along the satellite track.

SAR may be used in different operating modes, depending on the trade-off between resolution and swath, and on the combination of transmitted and received polarizations. One operating mode is designed for wave spectra. A small part of the image (a vignette) is sampled at intervals. With Envisat, for example, the vignette is 5 km x 5 km at the best spatial resolution of 30 m, and sampled at 100 km intervals along the track. The echoes in the vignette are analysed to plot the power spectra, from which it is possible to determine the dominant wave direction, the dominant wave length/period and power associated with significant wave height.

The list of SAR applications is long, although not all bands are suitable for all applications:

(a) Ocean circulation features (eddies) and waves (preferably L band);
(b) Ocean pollution and oil spills (preferably C band);
(c) Sea-ice cover and type (age) (any band);
(d) Land-ice cover (glaciers) (any band);
(e) Snow melting conditions and snow water equivalent (preferably X band);
(f) Surface soil moisture (preferably L band, especially for the areas around plant roots);
(g) Vegetation type (preferably C band) and total biomass (preferably P band);
(h) Land use and urbanization (preferably C band);
(i) Geological structure detection (preferably X band);
(j) Disaster monitoring and damage inventories (preferably X band);
(k) Ship traffic surveillance and military surveillance (preferably X band).

Other important applications are possible through interferometry because the phase control of the SAR signal is extremely accurate. Signals from different orbits of the same satellite or different satellites can be accurately co-registered in order to implement interferometry. This enables, for instance, land surface topography to be obtained for an improved digital elevation model, and iceberg heights to be measured.

By using interferometry between passes of the same satellite in an orbit with a repeat cycle, it is possible to measure changes such as iceberg drift, variations of glacier cover and lake extent, volcanic surface topography changes and bradyseisms, coastal erosion and urbanization.

2.2.5.3 Lidar

While the principle of lidar (light detection and ranging) is the same as that of radar, the electromagnetic range is different: MW is used for radar, while SW is used for lidar. Most lidars make use of wavelengths in the UV (e.g. 355 nm), VIS (e.g. 532 nm), NIR (e.g. 1 064 nm) or SWIR (e.g. 1 550 nm) ranges; longer wavelengths (e.g. 10.6 µm) may also be used. The source is a laser (light amplification by stimulated emission of radiation). It is extremely directional, but the large distance between the satellite and the Earth means that significant electric power is needed and that large telescopes are required to collect the backscattered signal. The use of lidar in space therefore presupposes considerable resources. Although in principle, lidar could be used to scan an area for imaging purposes, lidar systems in space have so far only been nadir-pointing or monodirectional. The following missions are lidar-based:

(a) Backscatter lidar for aerosol and cloud-top height;
(b) Doppler lidar for wind profile in clear air;
(c) Lidar altimeter specifically for sea ice;
(d) Differential absorption lidar (DIAL) for atmospheric chemistry.

Backscatter lidar

Backscatter lidar primarily deals with the observation of aerosols. This implies the use of wavelengths that are similar in size to very small aerosols (~1 µm). In order to capture more aerosol properties (size, phase, absorption/scattering ratio, and ultimately, type), more concepts have been developed. Some are based on the use of two wavelengths, while other are based on high spectral resolution laser in order to distinguish Mie and Rayleigh scattering components. In any case, a backscatter lidar is a very large instrument (the telescope may have ~1 m aperture). The footprint may be as small as a few tens of metres, but more decorrelated echoes are included
to increase signal-to-noise ratio. Such echoes will help to ensure that the final resolution is in the range of a few hundred metres. The vertical resolution determined by the lidar ranging system is set to a few hundred metres.

Designed primarily for aerosols (the most demanding mission), the backscatter lidar enables different kinds of observations to be made, including:

(a) Aerosol profile and aerosol properties (size, phase, absorption/scattering ratio, type);
(b) Cloud-top height (much more accurately than with passive NIR and IR techniques);
(c) Optical thickness of thin clouds (cirrus) and cloud base of semi-transparent clouds;
(d) Polar stratospheric clouds;
(e) Atmospheric discontinuities such as tropopause and top of the planetary boundary layer, as revealed by the change of refractivity index.

**Doppler lidar**

Doppler lidar deals primarily with the observation of wind profiles, a key variable for weather forecasting.

The current operational technique available for observing wind profile consists of tracking the movement of clouds or water vapour patches in frequent images, either from geostationary satellites, or from polar orbiting satellites in polar areas with frequent satellite overpasses. This limits the opportunity to take measurements only at altitudes where tracers are present (generally one layer, sometimes two). Hyperspectral IR sounding in GEO is expected soon: when it becomes available, frequent water vapour profiles will be available, and water vapour patterns will be tracked at several heights. However, vertical resolution and accuracy are expected to be limited, and coverage will not include high latitudes.

Cloud-motion tracking is the only technique available for cloudy areas, while the Doppler lidar enables tracking to be conducted in clear air. The tracer consists of eddies of the turbulent atmosphere, of aerosols and of molecular scattering.

Exploitation of the Doppler shift due to the wind implies oblique viewing. The laser pulse repetition frequency is such that the corresponding IFOV may be less than 100 m. However, as with any radar or lidar, a number of decorrelated echoes need to be averaged to improve signal-to-noise ratio: the final resolution is an IFOV as large as several tens of kilometres. Because of the limited availability of electric power, the instrument has a limited duty cycle; for instance it may sample at 200 km intervals along a line parallel to the sub-satellite track. The vertical resolution of the wind data collected depends on the time sampling rate of the return echo. This is adjusted for a vertical resolution of about 1 km in the mid-troposphere, less than 1 km in the planetary boundary layer, and more than 1 km in the stratosphere. No technique based on cloud and water vapour tracking from GEO can compete with this performance. The anticipated accuracy of the wind horizontal component is less than 2 m/s.

A Doppler lidar is a very large instrument (the telescope may have a 1.5 m aperture). It requires a dedicated satellite in an orbit lower than that usually used for meteorological satellites.

For wind lidar, coverage is a major limitation. Since one instrument only covers a line parallel to the sub-satellite track, a constellation of satellites would be required to provide frequent coverage. As long as the sustainability of such a constellation cannot be achieved, the baseline system for wind profile will continue to be imagery and hyperspectral sounding from GEO, with Doppler lidar providing support for calibration of the overall system.
Another difficulty with wind lidar is that measurement can only be along-sight (1D). That means that, in order to retrieve the required 2D horizontal component, assimilation into a numerical weather prediction model is necessary.

The Doppler lidar is primarily designed for wind, which is the most demanding mission. It provides several kinds of observations, including wind profile in clear air or in the presence of thin cirrus, aerosol profile (from echo intensity) and cloud-top height.

**Lidar altimeter**

Radar altimeters can provide measurements as accurate as a few centimetres. However, at above 20 km, their horizontal resolution is quite coarse. By implementing SAR processing of the along-track signal, the resolution can be brought to ~300 m, which is still insufficient for accurately detecting boundaries. Another limitation of radar altimeters is their unsuitability for observing surfaces with high emissivity (and thus low reflectivity) in the MW range, such as land and ice.

Lidar does not have those limitations. The horizontal resolution can be a few tens of metres and the vertical resolution less than 10 cm. That fine resolution makes it possible to capture the border between sea water and polar ice and (after successive passes) to profile the height along-track in order to map ice thickness.

In order to operate over the sea, even though reflectance in the VNIR range is very low, a lidar altimeter must have a large telescope (aperture ~1 m). For improved cost-effectiveness, a sensor using a second wavelength is usually added to observe the atmosphere (for example, 1 064 nm is used for surface and 532 nm is used for atmosphere). In this case, a lidar altimeter operates as an ordinary backscatter lidar.

As with any other altimeter, the lidar altimeter requires extremely accurate orbit determination, since the basic ranging measurement provides the distance of the object from the satellite in orbit. Precise orbit determination is achieved by a GPS receiver, and laser ranging from a network of ground stations with an array of retroreflectors on the satellite.

Lidar altimeter applications include:

(a) Sea-ice thickness and polar cap topography;

(b) Sea level and ocean-dynamic topography;

(c) Contribution to improved knowledge of the geoid;

(d) Land-surface topography, including glaciers and lakes;

(e) Aerosol profile and aerosol properties;

(f) Cloud-top height, optical thickness of thin clouds and cloud base of semi-transparent clouds;

(g) Polar stratospheric clouds;

(h) Atmospheric discontinuities such as tropopause and top of the planetary boundary layer.

**Differential absorption lidar**

The principle of differential absorption is to perform complementary observations of a target species at an absorption peak and in a nearby atmospheric window. Using lidar for this purpose is the only way to achieve high vertical resolution profiles in the planetary boundary layer, where such resolution is most important. (The profile is also observed in higher layers, including the stratosphere).
The conditions for DIAL to be applicable are: that the species is relatively abundant; that the absorption line is strong and lies in a spectroscopically “clean” region; and that there is a nearby “clean” atmospheric window. Examples of possible applications are:

(a) CO\textsubscript{2} by exploiting lines at 1.57 \(\mu\text{m}\) in the 1.6–1.7 \(\mu\text{m}\) window or 2.05 \(\mu\text{m}\) in the 2.0–2.3 \(\mu\text{m}\) window;

(b) H\textsubscript{2}O by exploiting lines at 935 nm, with windows on both sides;

(c) O\textsubscript{3} by exploiting lines and windows in the 305–320 nm region.

Missions exploiting DIAL are only at the stage of proposal or feasibility study. Since the spectral bandwidth (and thus the energy available for the observation) is very narrow and the reflector (a gas) is very weak, the instrument has to be very large and overall measurement is technologically challenging.

2.3 \textit{SPACE AND GROUND SEGMENTS}

Earth observation from space implies a complex system composed of (i) a space segment to perform observations, and (ii) a ground segment to manage the space segment and process observation data.

2.3.1 \textit{Space segment}

The space segment of a satellite system includes:

(a) The platform (also referred to as the bus);

(b) The instruments installed on board;

(c) The communication tools to receive commands and convey the instrument output to the ground.

The size and/or mass of satellites for Earth observation can range over two orders of magnitude:

(a) Nanosat: < 10 kg (actually unlikely to be used for operational Earth observation);

(b) Microsat: 10–100 kg;

(c) Minisat: 100–500 kg;

(d) Smallsat: 500–1 000 kg;

(e) Mediumsat: 1–2 tons;

(f) Large facility: > 2 tons.

2.3.1.1 \textit{Platform services}

The satellite platform hosts the instruments and provides several services:

(a) Power supply for instruments, telecommunications, and all other satellite subsystems;

(b) Navigation facilities for geographical referencing of observations;

(c) Attitude control for correct pointing of instruments and stabilization;
(d) Thermal control to keep the instruments within specified operating conditions;
(e) Housekeeping devices to monitor and control the status of all satellite subsystems;
(f) Propulsion for orbit keeping and, if needed, orbit change;
(g) Processing capability to administrate the various platform subsystems;
(h) Processing capability to handle instrument data and format data streams to be transmitted;
(i) Storage device for on-board global data recording;
(j) Communication facilities to receive commands from the ground;
(k) Communication facilities to transmit observational and housekeeping data to the ground;
(l) Other communication services, where the platform has a data relay function only.

2.3.1.2 Navigation and positioning systems

Navigation and positioning systems are necessary for geolocation of observed data, both during viewing, and afterwards for ground processing. The following systems are used:

(a) Laser retroreflectors;
(b) GNSS receivers;
(c) Radio-positioning systems;
(d) Star trackers.

Laser retroreflectors

These are mirrors which tend to be corner cubes. They reflect laser beams sent to the satellite by laser-equipped ground sites during positioning sessions. Laser retroreflectors are used on many satellites for a posteriori precise orbit determination. This is achieved by post-processing a number of measurements in night-time and clear sky only. The analysis involves a full network of coordinated ground stations. The results are sparse and only available after a certain delay. However, they are so accurate that they can be used for space geodesy applications.

Radio-positioning systems

These systems are specifically designed to support altimetry missions. They comprise radio links between the satellite on the one hand, and ground transmitting and/or receiving stations on the other. Positioning is performed in near real time and, with improved accuracy, after post-processing. Two examples of such systems are:

(a) Doppler Orbitography and Radiopositioning Integrated by Satellite (DORIS), which measures the Doppler shift of signals from ground stations;
(b) Precise Range and Range-rate Equipment (PRARE), which measures differential signals from a network of ground stations.
GNSS receivers

These systems make use of the phase difference of signals from several satellites in the GNSS. The GNSS includes the navigation satellite constellations of the United States of America (GPS), the Russian Federation (GLONASS), the European Union (Galileo), and China (Compass, known as Beidou in Chinese). A large number of satellites currently use GNSS receivers to support their navigation. Positioning is performed in real time.

Star trackers

These are charge-coupled device imagers that track bright stars, recognize their pattern, and send information to a satellite’s attitude control system. Star trackers provide continuous monitoring of satellite attitude much more accurately than systems based on horizon-sensing. This is necessary for instruments that require accurate pointing information (such as limb sounders), both for active attitude control during flight and for subsequent instrument data processing. An increasing number of satellites are now being equipped with star trackers.

2.3.1.3 Orientation and stabilization

The orientation and stabilization systems are primary platform features that determine instrument pointing capability.

The side of the platform on which the sensors are placed should ideally be kept facing the Earth’s surface, unless the satellite mission has a different purpose (such as monitoring the Sun). Since the platform tends to keep a steady orientation in relation to the stars during its orbital motion, a stabilization mechanism is required.

The stabilization mechanism known as spinning is most straightforward, as it is passive and inertial. The spin axis tends to have a constant orientation in relation to the stars, and therefore does not fulfill the Earth orientation requirement. For GEO satellites, if the spin axis is set parallel to the Earth’s rotational axis, the Earth’s surface is scanned for a small amount of time (about 5% of the satellite’s orbiting time) during each satellite rotation. For low orbiters, the orientation of the spin axis may be set specifically to enable instruments to be pointed towards the Earth’s surface for a fraction of time. In any event, spin stabilization is only suitable when an instrument’s radiometric budget is sufficient to carry out the measurements for which the instrument was designed, in spite of the small fraction of useful observing time. In addition, spin stabilization can only be implemented for one instrument, or very few instruments, on one platform.

Three-axis stabilization is much more suitable for maintaining a constant orientation towards the Earth and also supports more instruments on one platform. This allows for active control of the satellite attitude with respect to rotations around: the axis perpendicular to the orbital plane (pitch), the axis tangent to instantaneous motion in orbit (roll), and the nadir direction (yaw).

Active control is critical, because it implies accurate attitude determination (by, for example, horizon detection, star trackers or GNSS receivers) and efficient actuators (such as micropropulsion devices, gyros, very fine-angle change detectors and efficient control electronics). A loss of active control is among the primary causes of mission failure. Active control may affect data quality because of limited accuracy, particularly in the case of high-resolution instruments and high orbits (GEO), and mechanical perturbation of instrument pointing associated with turning on the actuators.

In addition to the main orientation and stabilization systems (spinning and three-axis control), smart attitude control systems are also in use, especially with small satellites. For instance, the gravity-gradient system uses a long boom that tends to the nadir direction, and thus keeps one side of the platform pointing towards the Earth’s surface.
2.3.1.4 **Housekeeping system**

A basic trade-off for satellite design lies between the capabilities that are implemented on board and those that can be achieved on the ground if sufficient information regarding on-board features is available. Hardware implementation on board may be expensive, prone to irrecoverable failure, and provide limited performance. Therefore, it is advisable to reserve hardware implementation to cases where it is absolutely indispensable and reasonably safe. Moreover, the housekeeping system provides all the ancillary information necessary to accurately process data on the ground.

The housekeeping system manages both the platform (deformations, temperature of radiant surfaces, attitude, status of power generators and all other subsystems) and the instruments (status and temperatures of the various parts, control signals for electronics, etc.). In general, instrument housekeeping is at least partially implemented inside the instrument itself.

The amount and completeness of housekeeping constitutes a discriminating factor for the class of a satellite. Operational satellites are equipped with plenty of housekeeping devices for subsystem monitoring and the possible activation of recovery manoeuvres, such as by switching to redundant units. Housekeeping information is also a basic element of accurate instrument calibration and data georeferencing.

Notwithstanding the importance of a good housekeeping system, there are limitations to the accuracy of what can be achieved by such software processing. The residual errors of software corrections or reconstructions may exceed what is allowed by the application. Therefore, certain corrections need to rely on on-board hardware.

2.3.1.5 **Data transmission**

The platform must transmit to the ground the observation results from the various instruments. Whatever the satellite height over the Earth's surface, radio transmission to the Earth will have to cross the ionosphere and plasmasphere, which block the propagation of electromagnetic waves with frequencies lower than the critical plasma frequency (~25 MHz). Direct visibility is needed between the transmitters and the receivers both on the satellite and the ground station.

The simplest method of collecting observed data from a satellite is by direct broadcast in real time. For a LEO satellite, a ground station will acquire all the data that a satellite transmits when passing within the acquisition range. The size of the acquisition range is the same as a satellite’s FOV for zenith angle $\zeta = 90^\circ$ (in principle), or $\zeta = 85^\circ$ (with a reduced risk of interferences from ground sources or occlusion from orography). Table 2.1 shows that, for a satellite height of 800 km for instance, the acquisition range is a circle with a diameter of 5 000 km (for $\zeta = 85^\circ$ or elevation = 5°).

Where the satellite velocity is in the region of 400 km/min and the satellite pass is centred over the acquisition station, the acquisition session lasts 15 min at most, and is reduced to a few minutes for peripheral passes.

This type of acquisition is the most convenient because it provides the observed data to the user in real time for immediate processing. However, only data observed and transmitted during the satellite pass inside the acquisition range can be acquired by the local receiving station. For GEO, direct broadcast can be continuously received by a station located within the FOV.

An alternative way for a LEO satellite to receive data is to store the observed data on board and transmit them on command, when the satellite overflies a central acquisition station. The central station, also used to send commands to the satellite, has the same acquisition range as any local station; if it is placed at a high latitude, it can collect data from many orbits. Table 2.4 shows that most Sun-synchronous orbits pass less than $10^\circ$ from the pole, so that a central station placed at a latitude of, for instance, $80^\circ$ will intercept all orbits and provide global data acquisition. The store-and-dump acquisition method has the advantage of enabling data recovery for the whole globe, but also has several drawbacks:
(a) Access to data is slower, since the delay includes the time needed to: run the whole orbit (up to 100 min), receive at the central station (about 10 min), relay to the central processing facility (about 10 min) and redistribute to the users (about 10 min). The total delay in the availability of data is therefore 2–3 h.

(b) The satellite has to transmit the data accumulated in one orbit (about 100 min) during the time it spends within the acquisition range of the central station (about 10 min). Therefore data rate and bandwidth need to be one order of magnitude higher than for direct read-out, which heavily impacts on the cost, size and complexity of the station. This acquisition mode is suitable for a satellite operator but generally not suitable for an individual user.

(c) There are instruments with data rates so high that they cannot be fully stored on board: selection of data to be stored may be needed, either by reducing the resolution or by prior selection of frames (e.g. the local area coverage (LAC) mode of NOAA Polar-orbiting Operational Environmental Satellites (POES)).

If not acquired locally in real time, the data gathered in a central processing facility need to be retransmitted to the users, generally after preprocessing. In the case of GEO satellites, a transponder on the same satellite can be used for data retransmission to local user stations.

For LEO satellites, between the two extreme cases (direct read-out providing data over a limited area in real time, or store-and-dump of global data with 2–3 hours’ delay), there are alternative or complementary data recovery schemes, which can use:

(a) Several downlink stations spread over the globe, including, for instance, one near each polar region; this reduces the length of time that data need to be stored on board the spacecraft;

(b) A network of direct read-out stations spread over the globe; each one acquires data on limited areas and retransmits them to data centres; this reduces the delay in the availability of data to a few tens of minutes, but does not necessarily achieve global coverage; or

(c) A data relay satellite that receives data in real time from observing platforms and relays them to a central processing facility; this reduces the delay in the availability of data to a few minutes.

Data recovery timeliness is a critical issue for operational satellites; that is particularly the case for meteorology, because of the coexisting requirements for timeliness and global coverage. For research and development applications, the timeliness requirement is less stringent, and the store-and-dump method tends to be used in conjunction with an effective archiving and retrieving facility, which provides advanced data stewardship.

2.3.1.6 Data relay services

In addition to providing Earth observation data from a platform in orbit, satellites can support other services, and so act as a telecommunications relay. The most common forms of such relays are:

(a) Data collection from in situ platforms located on the ground, on aircraft, on balloons, on buoys, on ships and even on migrating animals. The data collection platform (DCP) may transmit all the time, at fixed intervals, or upon interrogation from the satellite. Mobile platforms may be located from the satellite if in LEO. GEO satellites serve either DCPs within their line of sight (regional) or DCPs carried on mobile platforms (ships, aircraft, etc.) that migrate among the views of different GEO satellites (international).

(b) Search and rescue distress signals are detected from small pieces of transmitting equipment carried by those in distress. The request for help is then relayed to one of the centres of a worldwide search and rescue network. Search and rescue missions are conducted
cooperatively by several operational meteorological satellites in LEO and GEO. The service from LEO is called the Search and Rescue Satellite-aided Tracking System (SARSAT), while the GEO-based service is known as Geostationary Search and Rescue (GEOSAR).

(c) Relay of meteorological information from meteorological centres to end users as a broadcast, or to selected centres within view of a GEO satellite. The central facility of the system may perform the uplink or delegate it to auxiliary stations close to the information production centre that are equipped to uplink the satellite.

2.3.2 **Ground segment**

The ground segment of a satellite system includes:

(a) The central station for satellite command and global data acquisition;

(b) Peripheral stations for data acquisition;

(c) Mission and operation control centres;

(d) Data processing and archiving centres;

(e) Data and product distribution systems.

2.3.2.1 **Central station for satellite command and global data acquisition**

This element of the ground segment may be generically termed the command and data acquisition station (CDA). Typical tasks of a CDA are:

(a) To collect command sequences from the mission and operation control centre and uplink commands to the satellite (for payload configuration, satellite configuration, orbit control, etc.);

(b) To acquire satellite telemetry data (for attitude and orbit determination, satellite and payload status, etc.) and immediately deliver it to the mission and operation control centre(s);

(c) To acquire geophysical and ancillary data (housekeeping, calibration, etc.) and deliver it to the data processing and archiving centres;

(d) To index the acquired data streams with accurate time and orbital elements.

It is possible to have only one CDA for geostationary satellites. For near-polar satellite systems, it is possible to avoid blind orbits by placing the CDA at a very high latitude (such as Svalbard at 78°N). In order to improve the timeliness of the availability of observed data, auxiliary stations may be used (such as an Antarctic station). For low-inclination orbits, a network of CDAs is necessary, with one acting as the main CDA.

CDAs use S-band frequencies (about 2 100 MHz) to command the satellite. S band is nearly insensitive to weather and less critical to pointing accuracy. For geophysical data acquisition, the L band is used (about 1 700 MHz) if the data rate is below 10 Mbps; otherwise either the X band is used (about 8 GHz) for data rates of up to some 100 Mbps; or the K band (about 26 GHz) is used for data rates of several hundreds of Mbps.
2.3.2.2  **Mission and operation control centres**

These elements may be generically termed the operation control centre (OCC). Their tasks are:

(a)  To collect information on satellite, payload, and orbit status from the CDA and (in the case of the orbit only) from other ranging stations;

(b)  To collect requirements for elements including payload configuration and measurement sequence planning from data processing and archiving centres, and from other users entitled to input requirements into the mission plan;

(c)  To analyse the information on satellite, payload, and orbit status, and on payload or mission configuration requirements; to generate instrument performance monitoring reports; to elaborate the operations plan and deliver commands to the CDA for uplinking the satellite;

(d)  To provide data processing and archiving centres with ancillary information, which is relevant for data processing and results from activities relating to operations, payload or mission control (including accurate orbit determination, satellite attitude behaviour and payload status).

Mission control centres are closely connected with users, application centres and scientific teams. OCCs are closely connected with CDAs and the units responsible for satellite development. OCCs also have full knowledge of satellite design features. The two centres are often co-located, although this is not compulsory, and should preferably be secured by a back-up centre.

2.3.2.3  **Data processing and archiving centres**

These elements are responsible for:

(a) Acquiring geophysical, calibration and selected auxiliary data from CDAs;

(b) Acquiring auxiliary data on orbit, satellite and payload from OCCs;

(c) Monitoring instrument calibration and performing inter-calibration as appropriate;

(d) Generating and controlling the quality of various products;

(e) Archiving all products;

(f) Distributing a selection of products;

(g) Analysing mission status, payload status and requirements for mission planning;

(h) Delivering requirements for payload and mission control to OCCs.

Core products are normally generated by the satellite operator at a central facility. External specialized centres may supplement those by processing other specific products.

Satellite data archiving requires the maintenance of high levels of hardware availability for data ingest and storage, as well as discovery and retrieval services, with provision for long-term data preservation over decades. Data must be associated with metadata that contain all the information necessary to use and evaluate the data. Comprehensive, standardized metadata, and standardized globally interoperable catalogue systems enable data discovery to be extended to a worldwide scale within the WMO Information System.
2.3.2.4 **Data and products distribution**

Depending on data volumes and timeliness requirements, several methods of accessing data and products are available:

(a) Direct read-out from the satellite (when available; particularly prevalent among LEO satellites). This provides the best timeliness but presupposes the capability to receive raw data on an appropriate receiving station and to preprocess that data with an adequate software package.

(b) Near-real-time satellite retransmission of data after on-ground preprocessing or full processing. For GEO satellites, retransmission may be performed through the same satellite. Currently, retransmission is best achieved via commercial telecommunications satellite channels, such as the GEONETCast system, which consists of three coordinated services: EUMETCast (of the European Organization for the Exploitation of Meteorological Satellites), CMACast (of the China Meteorological Administration), and GEONETCast-Americas (of NOAA). Using this approach, onward dissemination services can be optimized within the ground segment, while taking into account the various missions and product sources available independently of the design constraints of any particular satellite.

(c) Near-real-time retransmission of data via specialized networks such as the WMO Global Telecommunication System.

(d) Active FTP retrieval from data centres for off-line data, specifically from archives.

Data and product distribution may be subject to conditions, depending on the status and data policy of the space agency running the programme (operational, research and development, commercial) and what the programme will be used for. Access to data retransmission services is generally controlled by encryption and subject to registration, even if no charges are levied.

2.3.2.5 **User receiving stations**

User stations are installed to make use of real-time or near-real-time data transmission from the satellite. Depending on the satellite access modality and the user requirement, there may be:

(a) High-data-rate acquisition stations for full reception of the data available either by open access or by agreement with the satellite owner;

(b) Low-data-rate acquisition stations for a data selection of either reduced volume or quality;

(c) Receiving terminals of commercial telecommunication satellites used for data dissemination after preprocessing or processing in the data-processing and archiving centre.

Frequencies used by the high-data-rate acquisition stations are in the L band (about 1 700 MHz) if the data rate is below approximately 5 Mbps; otherwise the X band (about 8 GHz) is used for data rates of up to approximately 100 Mbps. The low-data-rate acquisition stations make use of relatively low frequencies (L band: about 1 700 MHz for GEO, and very high frequency (VHF): about 137 MHz for LEO) that can be used on mobile stations, such as ships. Commercial telecommunication satellite terminals use Kᵤ band (about 11 GHz) or C band (about 3.8 GHz).

2.3.2.6 **Product processing levels**

Satellite observations are retrieved from the raw data acquired from the instruments through a processing chain. Different processing levels are usually referred to, the detailed definitions of which depend on the instrument in question. Table 2.11 provides a generic description of these processing levels.
Table 2.11. Generic description of processing levels (to be adapted to each instrument)

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Instrument and auxiliary data reconstructed from satellite raw data after removing communications artefacts.</td>
</tr>
</tbody>
</table>
| 1     | Instrument data extracted, at full original resolution, with geolocation and calibration information.  
        | 1a (for LEO) or 1.0 (for GEO): calibration and geolocation attached but not applied.  
        | 1b (for LEO) or 1.5 (for GEO): calibration and geolocation applied.  
        | 1c, 1d, etc: optional for specific instruments. |
| 2     | Geophysical product retrieved from a single instrument in the original projection. |
| 3     | Geophysical product retrieved from a single instrument, mapped on uniform space and time grid scales, possibly on a multi-orbital (for LEO) or multi-temporal (for GEO) basis. Irreversible process due to resampling. |
| 4     | Composite multi-sensor and/or multi-satellite products; or result of model analysis. |

Level 0 data are processed from the raw data stream by removing communication artefacts (such as synchronization frames and communication headers) and appending all necessary auxiliary data, including housekeeping and station-added information on timing and tracking. Level 0 data should be archived permanently to enable reprocessing with an updated instrument model (such as improved calibration or georeferencing).

Level 1a or 1.0 data consist of instrument files (counts) in the original instrument projection, with an appended (but not applied) deformation matrix or algorithm for georeferencing and calibration coefficients. The process from Level 0 to Level 1a/1.0 is fully reversible. Level 1a/1.0 data are normally permanently archived, although in principle they could be reproduced if Level 0 data have been archived.

Level 1b or 1.5 data consist of calibrated, co-registered and geolocated data in physical units (generally radiances), still in the original instrument projection. The process from Level 1a/1.0 to Level 1b/1.5 is not reversible because of truncation, discretization and resampling operations. Although Level 1b/1.5 may in any case be reprocessed from Level 1a/1.0 or Level 0, the processing effort is such that, in general, Level 1b/1.5 data are permanently archived.

Level 1c data are processed from the Level 1b data of certain instruments in order to enable end users to make use of that data. The process may be fully reversible (for example, spectra from interferograms by Fourier transform) but equally may not be (such as with apodized spectra). In general, these data are permanently archived. For certain instruments, further Level 1 steps (1d, 1e, etc.) may be defined (the addition of a cloud flag, for instance).

Level 0 and Level 1 processing is performed by the satellite operator. Where there is a direct readout, the satellite operator generally ensures the availability of Level 0 and Level 1 preprocessing software to local data users.

Level 2 products are generated from Level 1 data by applying algorithms that make limited use of external information. Data quality information is appended. These products are generated in the original instrument projection and tend to be permanently archived.

Level 3 products are generated by compositing a sequence of Level 2 products from successive orbits (with LEO) or at successive times (with GEO). Possible gaps in the sequence may be filled by interpolation. Due to the resampling operations implied by mapping on uniform space and time grids, Level 3 is an irreversible process. Products are generated offline by the satellite operators or by end users; they tend to be permanently archived.
Level 4 products are generated by blending data from different instruments on the same or different satellites, either with other data sources, or by assimilation in a model. In Level 4 products, the contribution of a specific satellite instrument may be hardly recognizable.