Radiometry in Remote Sensing

Ivana Čapková, ivana.capkova@fsv.cvut.cz

August 29, 2004

1 Introduction

In remote sensing, two types of data are used: optical data, acquired with scanners or cameras, and radar data. The sensors can be placed on both aircraft and satellites. The disadvantage of aircraft is its unstable flight and undetectable position and orientation data. Satellite orbits are much more stable although the computations are more complicated.

In the following text, we will talk about "imaging" radars, i.e. radars which output an image. In order to place the radar onto a satellite, it can't be too long so as not to threaten the satellite stability; thus a synthetic aperture radar must be used (the resolution of the real aperture radars would not be sufficient). Synthetic aperture radar must always be coherent, i.e. store the phase of the transmitted signal for later comparison with the phase of the received signal.

SAR aperture is usually small but the post-processing of the received signal makes it possible to improve the resolution by simulating a much longer antenna.

In contrast to the optical data (acquired with a scanner or a camera), radar is a distance measuring device. It has no information about the angle. Radar resolution is therefore given naturally for the slant range, i.e. in the direction from the satellite to the scatterer. The ground resolution then depends on the look angle of the radar (see figure 1). In order to achieve a reasonable ground-range resolution, the radar must be side-looking. That also results in worse scattering features because most of the incident energy is reflected either away from the radar (for good reflectors, such as water bodies) or scattered to all directions (for coarse scatterers). So, most of the time we are just able to observe the coarseness of the scatterer.

Let's add here that for a radar, the worst reflectors are water bodies, roads etc. that reflect all the energy away from the radar. On the contrary, the best reflectors are corner reflectors, e.g. bridges over water or trees in the water [7]. Comparing the radar and optical data is quite difficult because different features are imaged in a different way, and there is also a great difference in wavelength and the look angle.

Also, there are radars with various wavelengths in use. Microwave radiation with different frequencies has different properties as to atmospheric delay or scattering. For ERS-1/2 satellites, C-band is used with the wavelength of $\lambda = 5.6cm$. Other satellites for radar mapping also use X-band.

For intereferometry, which interests us at most, the magnitude, i.e. the reflectivity of the received signal, is not of much importance. It is only suitable for coregistration and visual check. The more important property of the received signal is the phase, showing the "precise distance" from the satellite to the scatterer. But, this distance is coded in the interval of $(0, \lambda)$. Also, we will only deal with data acquired by two or more subsequent passes, not data acquired by two antennas at the same time.

In order to obtain not only the reflectivity but also the phase of the reflected signal, the radar must be coherent. That means, the phase of the transmitted signal must be stored for further processing, as well as the phase of the received signal.

2 SAR description

Let's start with ESA ERS-1/2 parameters. The velocity for the reference orbit is $v_s = 7456m/s$, the altitude above the WGS-84 ellipsoid is approx. $h = 785km$. Typical footprint (ground) velocity is $v_{gr} = 6650m/s$, nominal look angle of $\theta_{nom} = 23^o$. A typical distance between the sensor and scatterer at the mid-swath is $850km$, the microwaves are vertically polarized, both transmitted and received.

While describing SAR, we will talk about two main directions: the azimuth, i.e. along track direction, parallel to the vector $\vec{v}$ of the satellite movement; and the (slant) range direction, approximately perpendicular to this track, from the satellite to a given scatterer. Ground range is the slant range converted to the ground

with respect to the ellipsoid (or other body) used; the conversion depends on the look angle of the satellite (see figure 1).

Antenna footprint is the spot on the ground illuminated by a single pulse; it is approximately 100 km wide (ground range) and 4.5 km long (azimuth direction). *Swath* is the mapped strip, illuminated by the pulse sequence.

The principle of radar image acquisition is emission of pulses of a predefined length and shape and measuring the echo. The time of receiving the backscattered signal is given by the distance of the resolution cell, its magnitude is given by the backscattering properties of the resolution cell.

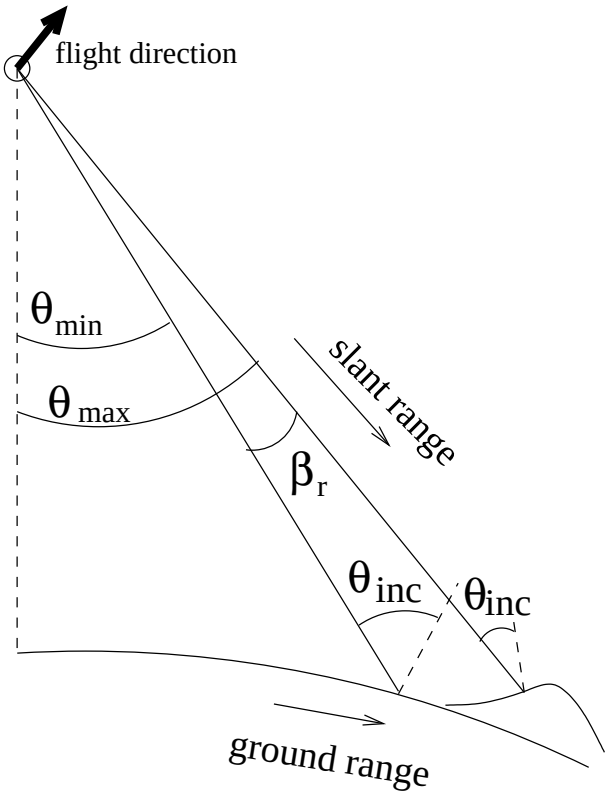

Figure 1: Look and incidence angles in the side-looking radar configuration

In remote sensing, we consider resolution in two directions:

- In the azimuth direction, the resolution is determined by the antenna length (aperture).
- In the slant range direction, the resolution is given by the effective pulse length.

In the ground range, the resolution is given by the slant range resolution and the incidence angle. In order to achieve good ground-range resolution, the look angle can't be too small. On the other side, it can't be too large either (too large distance, artifacts etc.). Considering a flat terrain, the ground range resolution is different in different parts of the image — in "near range", a pixel corresponds to a larger scatterer than in "far range" (for a more detailed description, see [2].

Let's emphasize here that the look angle is different from the incidence angle which is dependent on the local slope, see figure 1. For ERS-1/2, the look angle $\theta \in (21, 26)^o$, nominal $\theta_{nom} = 23^o$. The incidence angle is important for the reflectivity and is responsible for the image artifacts, the look angle is important for the ground-range resolution.

The synthetic aperture is (according to [6]) "equal to the distance the satellite travelled during the integration time" illuminating the given resolution cell; according to [10], it is the length of the radar footprint in the azimuth direction. Due to the relativity principle, it's the same.

Pulse repetition frequency (PRF) is limited by the antenna length and satellite velocity. According to [7], "in SAR the satellite must not cover more than half of the along-track antenna length between the emission of

successive pulses. For example, a 10-m antenna should advance only 5 m between pulses, to produce a 5-m-long final elementary resolution pixel. For a satellite traveling approx. 6 km s^{-1} over the ground, this implies a PRF of approx. 1 kHz.

In comparison to real aperture radars, where the resolution is given by the radar footprint, the important feature of SAR (see [3]) is that the azimuth and range resolution are independent. This is due to the fact that more distant areas are observed longer time.

Both the phase and amplitude, i.e. the entire complex signal, is constructed from many little reflectors within the given resolution cell (approximately corresponding to a pixel). That is, reflectors with a better reflectivity have more influence on the resulting phase (please note that the signal wavelength of 5.6 cm is much smaller than the resolution cell (approx. 20 m in the range direction)) than a less-reflecting object. Construction from many small object may sometimes be an obstacle for interferometry: the phase is naturally random and if the look angles of the two images differ too much (the parallel baseline is too long), i.e. also the incidence angles are different, the phases of the corresponding pixels are uncomparable and irregular with respect to the image. That's the reason why the upper limit for the baseline is about 1000 m for ERS-1/2 (derived e.g. in [7]). That's also the reason why data from different tracks or even different satellites cannot be combined to produce an interferogram. Even combining data of a descending and ascending passes of the same satellite is not easy; a solution is suggested in [9].

Also, the phase is "stable" and therefore useful only in "stable" areas. The "stability" is meant in relation to the wavelength — i.e. the artificial objects such as buildings, roads etc. are stable but trees (with moving leaves) are very unstable, except for the winter season when they have no leaves. The same applies to fields and meadows — these areas are useful for interferometry only in late autumn and early spring. Also, water or snow bodies are considered to be "unstable" except for areas of permanent snow or glaciers.

Let's note here that if we take the two images at the same time — as was the case of the SRTM mission [8], or airborne radars — this problem of stability arises only over water bodies whose surface changes very quickly. The "fundamental condition for interferometry" [7] is

$$2L(\sin\theta_1 - \sin\theta_2) < \lambda \tag{1}$$

where L is the ground-range pixel length (approx. 20 m at far range), θ_1 and θ_2 are the incidence angles of the two images, respectively, λ is the radar wavelength. The difference in round trip distance of both ends of a pixel is $2L\sin\theta$ and the condition (1) means that the phases of all the small targets within a pixel subtract in a similar way — if the condition weren't satisfied, the targets would add both constructively and destructively and the resulting phase difference would be random, without any meaning. Let's emphasize here that θ_1 and θ_2 are the incidence angles so the quality of the interferogram is also influenced by the local terrain slope.

Also, according to [7], the "direction of observation must be identical for the two images" in the along-track direction. "The degradation becomes total when the angle betwwen the two directions of observation exceeds the width of the antenna beam (0.3 ^{o} for ERS-1/2)." This corresponds to "excessively large difference" between the Doppler centroids of the two images. Doppler centroid will be explained below.

According to [10], let's consider two processes: SAR data acquisition, that is the conversion from the reality (with the infinitesimal resolution) to the raw data (blurred due to the SAR principle), and SAR processing, i.e. conversion from the raw data to the image (with resolutions cells about 20 by 5 meters large). SAR processing is therefore the inverse process to data acquisition.

3 Radar artifacts

There are three types of radar artifacts, all occuring in mountainous terrain. All of them are caused by the side-looking principle, and the amount is also dependent on the look angle.

- Foreshortening occurs when the radar beam reaches the base of a mountain first and after that it reaches its peak, i.e. reaches the terrain in the correct order but faster than it would for a flat terrain. The slope will appear compressed in the image and brighter (more energy is reflected). In comparison to other satellites, ERS-1/2 have quite small look angle, implying lots of these artifacts.

- Layover occurs when the radar beam first reaches the top of the mountaion and then the base, i.e. reaches the terrain in a bad order (the echo of the scatterers which are "closer" to the satellite (measured on the ground) is received later than the echo of the scatterer which are more distant). The effects look very similar to the foreshortening effects, but brighter.

- Shadow occurs when the radar is not able to illuminate a surface, i.e. these areas are very dark.

Figure 2 shows radar artifacts.

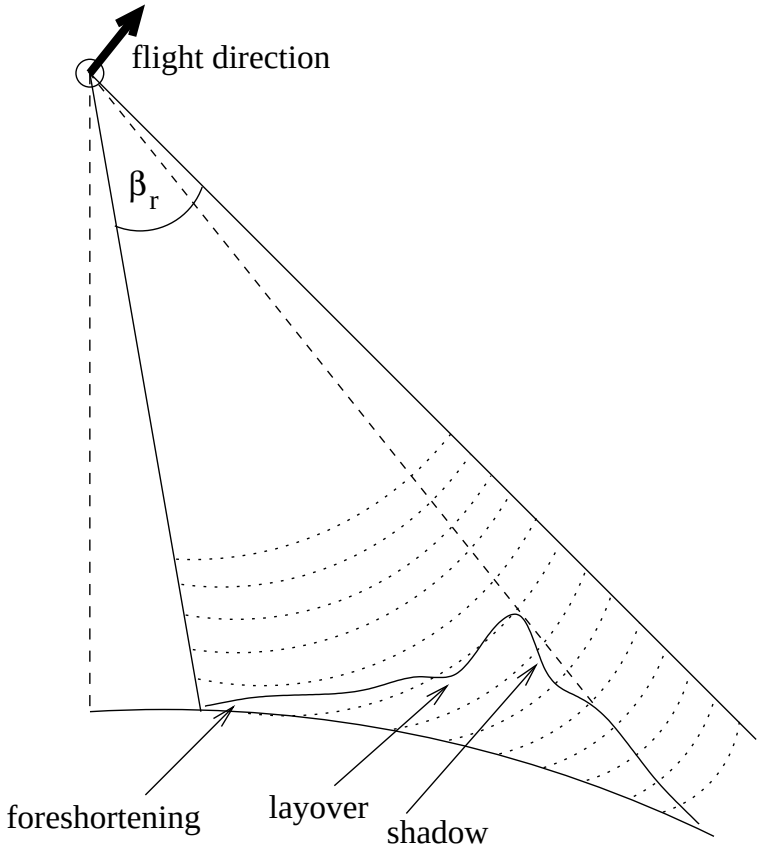

Figure 2: Illustration of radar artifacts

For a more detailed description and illustrations, see e.g. [2].

4 Antenna

This section is based on [1, 10].

Let's have a rectangular antenna with the (azimuth) length of L_a and (range) width of D_a. The normalized antenna pattern (i.e. the transmitted energy as a function of the off-center beam angles ϕ_r (range) and ϕ_a (azimuth)) can be written as (λ is the radar wavelength)

$$a(\phi_r, \phi_a) = \frac{\sin^2\left(\frac{D_a}{\lambda}\phi_r\pi\right)}{\left(\frac{D_a}{\lambda}\phi_r\pi\right)^2}\frac{\sin^2\left(\frac{L_a}{\lambda}\phi_a\pi\right)}{\left(\frac{L_a}{\lambda}\phi_a\pi\right)^2}$$

and a 2-dimensional simplification is shown in figure 3.

For ERS-1/2, the antenna length is $L_a = 10m$, the antenna width is $D_a = 1m$ (see [5]).

For simplification, let's consider (as usual) the antenna pattern to be a triangle, having the slopes where the "real" antenna pattern has a half power (-3 dB) (see the dashed line in figure 3). Now, the beamwidths in the azimuth and range direction are $\beta_r = 0.886\frac{\lambda}{D_a} = 2.870^o$, $\beta_a = 0.886\frac{\lambda}{L_a} = 0.287^o$.

The pattern within this angle limit is called the "main lobe", all the energy transmitted outside the limits is attributed to "sidelobes."

According to [1], the beamwidth of the main lobe is slightly broadened in range direction, "in order to get the power distributed more evenly across the full swath" [5]. In practice, the beamwidths in range and azimuth direction respectively, are [1] $\beta_r = 5.4^o$, $\beta_a = 0.228^o$. This is inconsistent with beamwidths presented in [10],

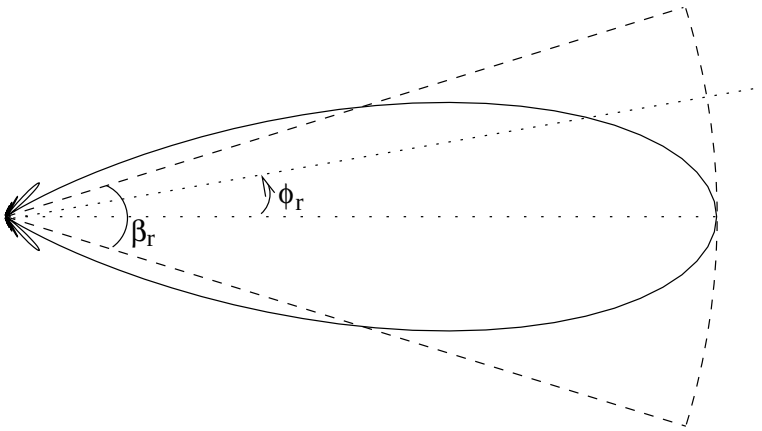

Figure 3: 2-dimensional antenna pattern, with β_r as the beamwidth

where $\beta_a = 0.208^o$. The inconsistency is probably caused by different consideration of limits (i.e. "first zero" versus "half power").

5 Sensor description — range processing

This section is based on [1, 5, 10].

Let's define two time scales: the first one, let's call it "fast time" [10] and denote τ, concerns the range processing. The transmitted pulse is just $37.1\mu s$ long and the echo, although a little larger, contains the information to be separated to pixels in the range direction. The other time scale, let's call it "slow time" and denote t, concerns the pulses; the pulses are transmitted with the frequency of $PRF = 1680$ Hz and the echo to each pulse generates a distinct line in the azimuth direction.

Because there is a several orders of magnitude difference between these two time scales, we can consider them to be independent. After range decompression, azimuth decompression is performed after range migration correction (will be discussed later).

The radar emits pulses of the length $\tau_p = 37.1\mu s$ with the PRF of 1680 Hz. After transmitting nine such pulses, first echo is received. According to [10], let's ignore that the spacecraft moved between the transmission and reception. The swath width (in ground range) is about 100 km, short enough to enable reception of the whole echo before transmitting another pulse. Thus, one antenna is used for both transmission and reception.

The transmitted frequency-modulated pulse can be written as

$$p_t(\tau) = g(\tau) \cdot \exp(-j2\pi f_0 \tau),$$

where $g(\tau)$ is the (complex) envelope, τ is the "fast time" and $f_0 = 5.3$ GHz is the carrier frequency (corresponding to the radar wavelength $\lambda = 5.6$ cm). The envelope could be a regular rectangle but, in this case, it would be a problem to transmit a great power within a short pulse; a longer pulse would cause worse resolution. That's why longer phase-coded pulses are used. Thus the envelope to be modulated is

$$g(\tau) = \exp(j\pi k \tau^2) \cdot \mathrm{rect}(\tau k / B_\nu),$$

where k is the frequency modulation rate (for ERS-1/2, $k \approx 0.42 MHz/\mu s$) and B_ν is the range bandwidth (i.e. maximal frequency of the chirp before modulation). The $\mathrm{rect}(x)$ function is defined in the following way [10]: for $|x| < \frac{1}{2}$, $\mathrm{rect}(x) = 1$; for $|x| = \frac{1}{2}$, $\mathrm{rect}(x) = \frac{1}{2}$; and for $|x| > \frac{1}{2}$, $\mathrm{rect}(x) = 0$.

That means that the frequency of the chirp is linearly increasing, reaching the maximum of $\nu = 15.5$ MHz (for ERS-1/2) for $\tau = \tau_p$.

The transmitted signal therefore looks like (omitting the magnitude)

$$p_t(\tau) = \exp\left(-j(2\pi f_0 \tau - k\tau^2)\right),$$

its phase is

$$\phi(\tau) = 2\pi f_0 \tau - k\tau^2$$

i.e. in the phase image and further processing, the radar wavelength of 5.6 cm (corresponding to the carrier frequency of 5.3 GHz) corresponds to the 4π phase cycle (because the phase is influenced by the round-trip distance).

The received signal (after demodulation and comparison with the transmitted one) has the form (omitting the magnitude)

$$p_r(\tau) = g(\tau - 2R/c) \cdot \exp(-j4\pi R/\lambda),$$

where R is the range and c is the speed of light. The phase $4\pi R/\lambda$ must be stored before further processing.

According to [5], "processing the returned signal involves stripping off the carrier frequency and performing a correlation with a copy of the transmitted signal". According to [1], the stripping off means down-conversion of the signal to the intermediate frequency of 123 MHz and sampling with the frequency of 18.96 MHz (higher than the bandwidth, although lower than the double bandwidth (Nyquist frequency) — I think that this is possible thanks to the knowledge of the pulse shape).

The received pulse after matched filtering has the form of [5]

$$h_r(\tau) = (\tau_p - |\tau|)\frac{\sin\left(k\tau\left(\tau_p - |\tau|\right)\right)}{k\tau\left(\tau_p - |\tau|\right)}\mathrm{rect}\left(\frac{\tau}{\tau_p}\right).$$

For the case of a single scatterer, this function has a sharp maximum at τ_d corresponding to the real time delay (phase). According to [5], the first zero of this signal is often taken as a measure of the time resolution

$$r_\tau = \frac{\tau_p}{2}\left(1 - \sqrt{1 - \frac{4}{B_\nu \tau_p}}\right).$$

Now the compression ratio is $\frac{\tau_p}{r_\tau}$ and r_τ is selected to be $r_\tau \approx \frac{1}{B} \approx 64$ ns. This is the effective pulse length, i.e. the length of a rectangular pulse to achieve the same range resolution.

The slant range resolution is then 9.68 m (according to [5, 1]) or 8.56 (according to [10] where half-power values are considered), corresponding to the ground range (with regard to ellipsoid) of 21.8 m (at far range) to 29.3 m (at near range) [5].

For radar data, it is usual for the pixel to be slightly larger than the resolution (i.e. the pixels overlap a little). This is given by the non-sharp antenna pattern and spectra both in the azimuth and range directions.

The resolution is then degraded by weighting the correlation function due to the sidelobes [5].

6 SAR principle — the azimuth direction

The azimuth beamwidth is $\beta_a \approx 0.287^o C$, corresponding to approx 4.5km pixels in the ground range. This is quite a large area, useless for further applications. But, if SAR processing/focusing is done, the resolution improves to approx. 5m. This derivation comes from [5].

The azimuth direction is the direction of the vector $\vec{v}$ of movement of the satellite. The received echos are influenced by the Doppler effect, i.e. the frequency of the received signal is different than the frequency of the transmitted signal. Also, the actual received frequency depends on the position of the scatterer in the beam. There is always a direction in which the Doppler shift is zero, called zero-Doppler plane. Unfortunately, due to small instabilities of the satellite movement, Earth rotation and other effects, this zero-Doppler plane is not in the center of the beam, i.e. the beam is not exactly perpendicular to the flight direction. The frequency which is received from the beam center is called the Doppler centroid and is always smaller than the PRF.

The value of Doppler centroid is computed during SAR processing and can be also changed during the process.

According to [7], both images should be focused by the same procedure; also it is better if they have the same Doppler centroid though not optimal for each image.

Because the echo of a given scatterer is received within many pulses (all the time while staying within the radar beamwidth), the SAR processing (focusing) procedure needs to be done first, in order to improve the resolution. Although improving range resolution is quite simple (it is a one-dimensional problem), improving azimuth resolution is more complex, mainly due to range migration.

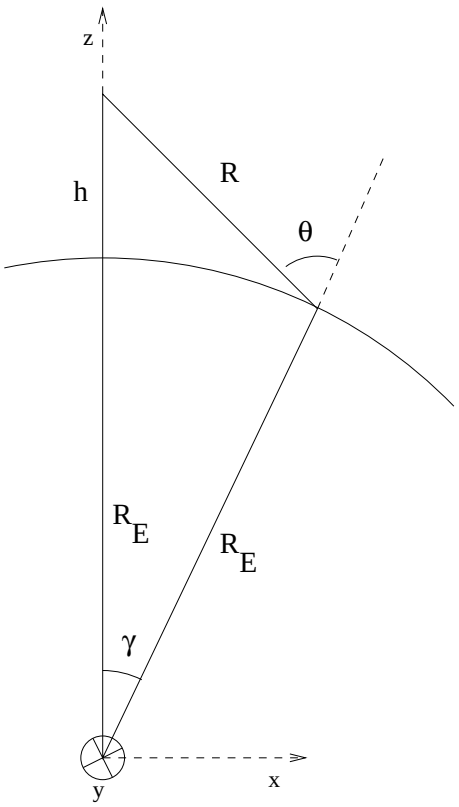

Figure 4: Coordinate system for azimuth resolution derivation. This figure is taken from [5].

Range migration means that the range between the scatterer and the receiver is changing during acquisition. During acquisition, received data are stored in lines with respect to the "fast time" of their reception. That means that a scatterer lies on an approximate hyperbola (see [10] for more exact approach or derivation) in the raw data matrix. Also, the curvature of hyperbolas is dependent on the minimal range of the scatterer (see [10]). This makes the problem of SAR focusing "two-dimensional and non-separable" [10].

Now, let's derive the achievable azimuth resolution (taken from [5]):

A point X (located on the Earth's surface) has coordinates of $X = R_E(\sin\gamma, 0, \cos\gamma)$ (see 4), the satellite has coordinates of $P = (R_E + h)(0, \sin\Omega t, \cos\Omega t)$, where Ω is the Earth rotation, t is the time and h is the satellite height. Let's define R_0 as the minimal slant range at time 0, $R_0 = |X - P|_{t=0}$. Then

$$R_0 = \sqrt{(R_E \sin\gamma)^2 + (R_E \cos\gamma - (R_E + h))^2},$$

$$|X - P| = \sqrt{(R_E \sin\gamma)^2 + ((R_E + h)\sin\Omega t)^2 + (R_E \cos\gamma - (R_E + h)\cos\Omega t)^2} = \sqrt{R_0^2 + 2R_E(R_E + h)\cos\gamma\cos\Omega t},$$

$$|X - P| \approx R_0 + \frac{R_E(R_E + h)\cos\gamma\Omega^2 t^2}{2R_0}$$

for small values of Ωt, corresponding to the small beamwidth. Therefore, the corresponding (two-way) phase of the received signal is [5]

$$\phi(t) \approx \frac{-4\pi R_0}{\lambda} - \frac{2\pi}{\lambda}\frac{R_E(R_E + h)\cos\gamma\Omega^2 t^2}{R_0},$$

which is equivalent to linear frequency modulation. Frequency variation of

$$f_d = \frac{1}{2\pi}\frac{d\phi}{dt} \approx -\frac{2}{\lambda}\frac{R_E(R_E + h)}{R_0}\cos\gamma\Omega^2 t$$

gives the half-Doppler-bandwidth for maximum allowed t_{max} which is the half-time of illumination of a ground point.

The illuminated area on the ground (the synthetic aperture length) is approximately $R_0\beta_a$ and the ground velocity of the beam is $v = R_E\Omega\cos\gamma$, so the half-time of illumination is

$$t_{max} = \frac{1}{2}\frac{R_0\beta_a}{R_E\Omega\cos\gamma},$$

corresponding to the resulting (Doppler) bandwidth of

$$B = \frac{2}{\lambda}\beta_a(R_E + h)\Omega.$$

Time resolution (according to the above relations) is $\tau_a = 1/B$, and during the time, the beam moves a ground distance of $v\tau_a$ and the azimuth resolution therefore is

$$r_a = \frac{R_E}{R_E + h}\cos\gamma\frac{\lambda}{2\beta_a},$$

i.e. approximately 4.5m for ERS-1/2 SAR.

7 SAR focusing

This section is based on [10].

The SAR focusing procedure should "unblur" the raw data, i.e. separate the contributions from the "real" resolution cells. As already noted, it is approximately an inverse procedure to the data acquisition. Without SAR focusing, the data would be unusable because of very large resolution cells.

The first step of SAR focusing is range decompression, i.e. correlation of the received signal with the transmitted one (both demodulated), resulting in the delay τ, determining the pixel to where the phase and magnitude of the received signal are stored.

Second, the Doppler centroid is estimated. It is done by the spectral analysis in the azimuth direction. According to [10], "since the Doppler centroid frequency varies over range, estimation is performed at several range positions."

After that, range migration correction is performed. That means transforming the approximate hyperbola in which a given scatterer lies (the exact shape of the curve depends, among others, on the actual range) to a line in the azimuth direction. This means that the signal needs to be shifted back in the τ direction and, because a single pixel of the actual data corresponds to many scatterers with different range migration corrections, this is not a trivial problem and there are at least four approaches to its solution. One of them is range-Doppler (the others are named in [7]), which first computes the Fourier transformation of the data in the azimuth direction, separating the data with different Doppler frequency and therefore with different range. The data are then focused in this range-Doppler domain (i.e. the range direction is not transformed, the azimuth direction is). For more details, see e.g. [10].

After range migration correction, azimuth decompression is a one-dimensional problem. It is done by correlation with the azimuth chirp.

8 The altitude of ambiguity

For SAR interferometry, two images are neccessary. For data selection, some criteria are neccessary: good weather conditions, because during storms etc. the athmosphere properties change very "quickly" (spatially), and the altitude of ambiguity h_a, i.e. the height difference corresponding to one fringe in the future interferogram.

Of course, h_a is influenced by the radar wavelength λ, but what is more important, it is inversely proportional to the perpendicual baseline $B_\perp$. The higher the h_a, the less the interferogram is influenced by topography — the better for extracting Earth deformation, the worse for constructing the DEM. According to [1],

$$h_a = \frac{\lambda R_{1p}\sin\theta_p}{2B_{\perp p}} \tag{2}$$

where the p index applies to a given pixel and R_{1p} is the distance between the "first satellite" and the ground patch corresponding to the pixel. That means, that the height ambiguity moderately changes within the image. For data selection, these changes may be ignored.

With respect to (1), $h_a > 10m$. Having two sattelites (the radar phase centers) at the same position, $h_a \to \infty$ which is the ideal configuration for deformation mapping.

References

[1] Ramon F. Hanssen: Radar Interferometry (Data Interpretation and Error Analysis), Kluwer Academic Publishers, 2001

[2] Remote Sensing Tutorial by Canada Center for Remote Sensing, http://www.ccrs.nrcan.gc.ca/ccrs/learn/tutorials/fundam/fundam_e.html

[3] PHARUS: More on the SAR principle, http://www.tno.nl/instit/fel/os/fac/ra_fac_pha_more_sar.html

[4] ESA Radar Course: SAR principle, http://earth.esa.int/applications/data_util/SARDOCS/ spaceborne/Radar_Courses/Radar_Course_III/synthetic_aperture_radar_SAR.htm

[5] Chris Oliver, Shaun Quegan: Understanding Synthetic Aperture Radar Images, Artech House, 1998

[6] ESA Instrument Description, http://envisat.esa.int/cgi-bin/printer_friendly.cgi?/dataproducts/asar/CNTR3-1.htm?06.12.2002

[7] D. Massonnet, K. L. Feigl: Radar Interferometry and Its Application to Changes in the Earth's Surface, Reviews of Geophysics, 36(4):441-500

[8] B. Rabus, M. Eineder, A. Roth, R. Bamler: The shuttle radar topography mission — a new class of digital elevation models acquired by spaceborne radars, ISPRS Journal of Photogrammetry and Remote Sensing, 54(2003):241-262

[9] M. Crosetto: Calibration and Validation of SAR Interferometry for DEM generation, ISPRS Journal of Photogrammetry & Remote Sensing, 57 (2002):213-227

[10] R. Bammler, B. Schättler: SAR Data Acquisition and Image Formation, in G. Schreier: SAR Geocoding: data and systems, pp. 53 – 102, Wichman Verlag, Karlsruhe, 1993