InSAR Principles:

Guidelines for SAR Interferometry Processing and Interpretation

Acknowledgements

Table of Contents

Part C InSAR processing: a mathematical approach

Scope

This manual has been produced as a text book to introduce radar interferometry to remote sensing specialists. It consists of three parts.

Part A is meant for readers who already have a good knowledge of optical and microwave remote sensing, to acquaint them with interferometric SAR image processing and interpretation.

Part B provides a practical approach and the technical background for people who are starting up with InSAR processing.

In **Part C** a more mathematical approach can be found, for a deeper understanding of the interferometric process. There, the manual also includes an appreciation of themes such as super resolution and ERS/Envisat interferometry.

Part A

Interferometric SAR image processing and interpretation

1. Synthetic Aperture Radar basics

1.1 Introduction

Synthetic Aperture Radar (SAR) is a microwave imaging system. It has cloud-penetrating capabilities because it uses microwaves. It has day and night operational capabilities because it is an active system. Finally, its 'interferometric configuration', Interferometric SAR or InSAR, allows accurate measurements of the radiation travel path because it is coherent. Measurements of travel path variations as a function of the satellite position and time of acquisition allow generation of **Digital Elevation Models (DEM)** and measurement of centimetric surface deformations of the terrain.

This part of the InSAR Principles manual is dedicated to beginners who wish to gain a basic understanding of what SAR interferometry is. Real examples derived from ESA satellites, ERS-1, ERS-2 and Envisat, will be exploited to give a first intuitive idea of the information that can be extracted from InSAR images, as well as an idea of the limits of the technique.

1.1.1 Introduction to ERS

The European Remote Sensing satellite, ERS-1, was ESA's first Earth Observation satellite; it carried a comprehensive payload including an imaging Synthetic Aperture Radar (SAR). With this launch in July 1991 and the validation of its interferometric capability in September of the same year, an ever-growing set of interferometric data became available to many research groups. ERS-2, which was identical to ERS-1 apart from having an extra instrument, was launched in 1995.

Figure 1-1: An artist's impression of ERS-2

Shortly after the launch of ERS-2, ESA decided to link the two spacecraft in the first ever 'tandem' mission, which lasted for nine months, from 16 August 1995 until mid-May 1996. During this time the orbits of the two spacecraft were phased to orbit the Earth only 24 hours apart, thus providing a 24-hour revisit interval.

The huge collection of image pairs from the ERS tandem mission remains uniquely useful even today, because the brief 24-hour revisit time between acquisitions results in much greater interferogram coherence. The increased frequency and level of data available to scientists offered a unique opportunity to generate detailed elevation maps (DEMs) and to observe changes over a very short space of time. Even after the tandem mission ended, the high orbital stability and careful operational control allowed acquisition of more SAR pairs for the remainder of the time that both spacecraft were in orbit, although without the same stringent mission constraints.

The near-polar orbit of ERS in combination with the Earth's rotation (E-W) enables two acquisitions of the same area to be made from two different look angles on each satellite cycle. If just one acquisition geometry is used, the accuracy of the final DEM in geographic coordinates strongly depends on the local terrain slope, and this may not be acceptable for the final user. Combining DEMs obtained from ascending (S-N) and descending (N-S) orbits can mitigate the problems due to the acquisition geometry and the uneven sampling of the area of interest, especially on areas of hilly terrain (this is illustrated in Figure 1-14 on page A-15). The ERS antenna looks to the right, so for example a slope that is mainly oriented to the West would be foreshortened on an ascending orbit, hence a descending orbit should be used instead.

In March 2000 the ERS-1 satellite finally ended its operations. ERS-2 is expected to continue operating for some time, although with a lower accuracy of attitude control since a gyro failure that occurred in January 2001.

1.1.2 Introduction to Envisat

Launched in 2002, Envisat is the largest Earth Observation spacecraft ever built. It carries ten sophisticated optical and radar instruments to provide continuous observation and monitoring of the Earth's land, atmosphere, oceans and ice caps. Envisat data collectively provide a wealth of information on the workings of the Earth system, including insights into factors contributing to climate change.
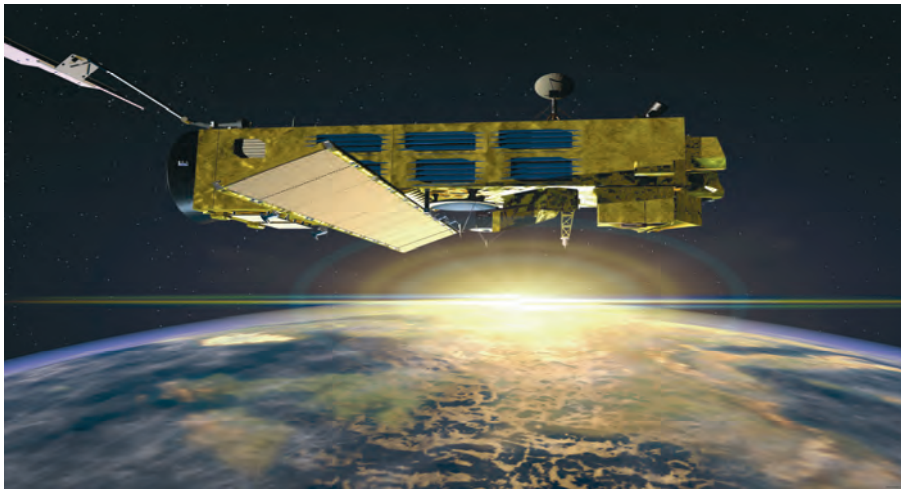
Figure 1-2: Artist's impression of Envisat

Furthermore, the data returned by its suite of instruments are also facilitating the development of a number of operational and commercial applications.

Envisat's largest single instrument is the Advanced Synthetic Aperture Radar (ASAR), operating at C-band. This ensures continuity of data after ERS-2, despite a small (31 MHz) central frequency shift. It features enhanced capability in terms of coverage, range of incidence angles, polarisation, and modes of operation. The improvements allow radar beam elevation steerage and the selection of different swaths, 100 or 400 km wide.

Envisat is in a 98.54° sun-synchronous circular orbit at 800 km altitude, with a 35-day repeat and the same ground track as ERS-2.

Its primary objectives are:

- to provide continuity of the observations started with the ERS satellites, including those obtained from radar-based observations;
- to enhance the ERS mission, notably the ocean and ice mission;
- to extend the range of parameters observed, to meet the need for increasing knowledge of the factors affecting the environment;
- to make a significant contribution to environmental studies, notably in the area of atmospheric chemistry and ocean studies (including marine biology).

1.2 SAR images of the Earth's surface

1.2.1 What is a strip-map SAR imaging system?

A SAR imaging system [Curlander91] from a satellite (such as ERS or Envisat) is sketched in Figure 1-3. A satellite carries a radar with the antenna pointed to the Earth's surface in the plane perpendicular to the orbit (in practice this is not strictly true, because it is necessary to compensate for the Earth's rotation). The inclination of the antenna with respect to the nadir is called the **off-nadir angle** and in contemporary systems is usually in the

range between 20° and 50° (it is 21° for ERS). Due to the curvature of the Earth's surface, the **incidence angle** of the radiation on a flat horizontal terrain is larger than the off-nadir (typically 23° for ERS). However, for the sake of simplicity we assume here that the Earth is flat, and hence that the incidence angle is equal to the off-nadir angle, as shown in the figure.
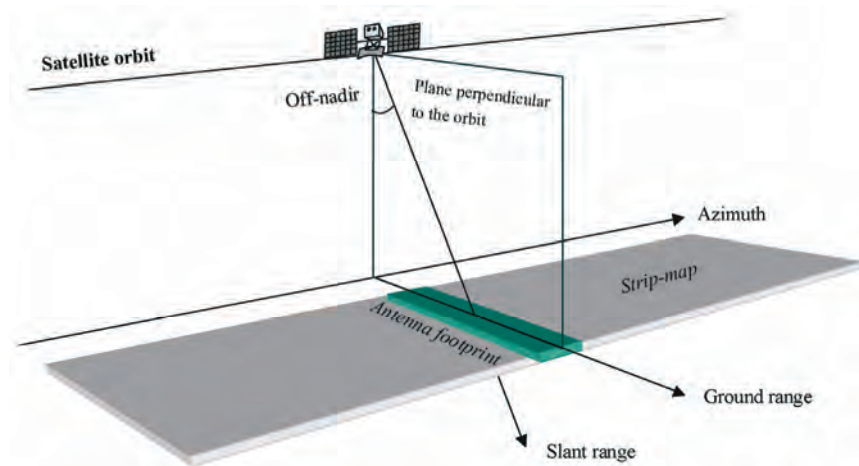
Figure 1-3: A SAR system from a satellite

Currently, operational satellite SAR systems work in one of the following microwave bands:

- C band – 5.3 GHz (ESA's ERS and Envisat, the Canadian Radarsat, and the US shuttle missions)
- L band – 1.2 GHz (the Japanese J-ERS and ALOS)
- X band – 10 GHz (the German-Italian X-SAR on the shuttle missions)

In the case of ERS, the illuminated area on the ground (the **antenna footprint**) is about 5 km in the **along-track direction** (also called the **azimuth direction**) and about 100 km in the **across-track direction** (also called the **ground range direction**).

The direction along the **Line of Sight (LOS)** is usually called the **slant-range** direction.

The antenna footprint moves at the satellite speed along its orbit. For ERS, the satellite speed is about 7430 m/s in a quasi-polar orbit that crosses the equator at an angle of 9° and an elevation of about 800 km. The footprint traces a **swath** 100 km wide in ground range on the Earth's surface, with the capability of imaging a strip 445 km long every minute (strip map mode).

1.2.2 What is a complex SAR image?

A digital SAR image can be seen as a mosaic (i.e. a two-dimensional array formed by columns and rows) of small picture elements (pixels). Each pixel is associated with a small area of the Earth's surface (called a **resolution cell**). Each pixel gives a complex number that carries amplitude and phase information about the microwave field backscattered by all the scatterers (rocks, vegetation, buildings etc.) within the corresponding resolution cell

projected on the ground (see section 1.2.3). Different rows of the image are associated with different azimuth locations, whereas different columns indicate different slant range locations.

The location and dimension of the resolution cell in azimuth and slant-range coordinates depend only on the SAR system characteristics.

In the ERS case, the SAR resolution cell dimension is about 5 metres in azimuth and about 9.5 metres in slant-range. The distance between adjacent cells is about 4 metres in azimuth and about 8 metres in slant range. The SAR resolution cells are thus slightly overlapped both in azimuth and in slant-range.

1.2.2.1 The detected SAR image

The detected SAR image contains a measurement of the amplitude of the radiation backscattered toward the radar by the objects (**scatterers**) contained in each SAR resolution cell. This amplitude depends more on the roughness than on the chemical composition of the scatterers on the terrain. Typically, exposed rocks and urban areas show strong amplitudes, whereas smooth flat surfaces (like quiet water basins) show low amplitudes, since the radiation is mainly mirrored away from the radar.

The detected SAR image is generally visualised by means of grey scale levels as shown in the example of Figure 1-4. Bright pixels correspond to areas of strong backscattered radiation (e.g. urban areas), whereas dark pixels correspond to low backscattered radiation (e.g. a quiet water basin).
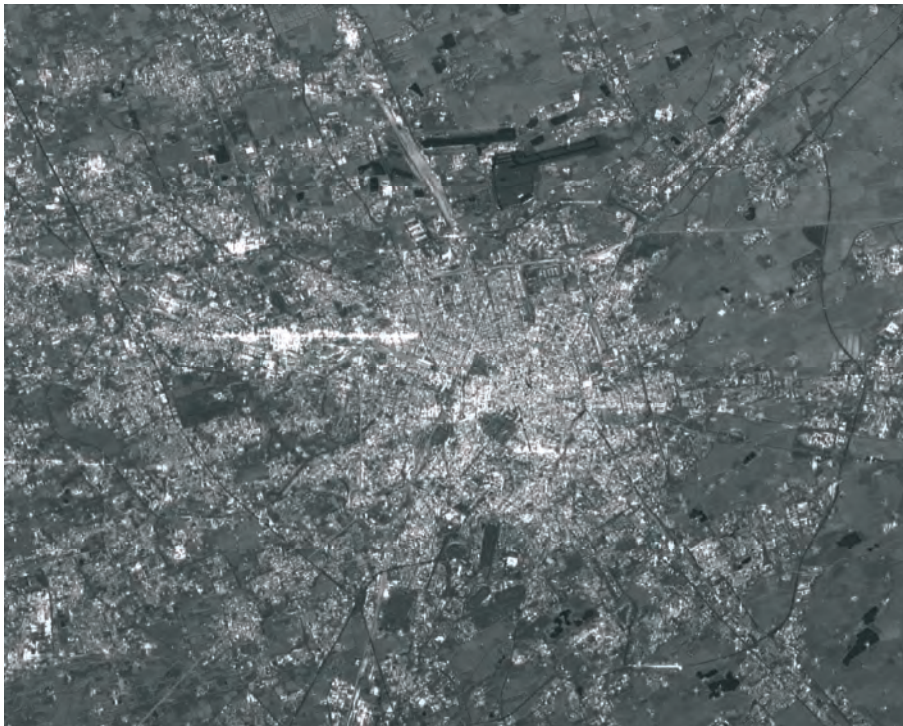
Figure 1-4: ERS SAR detected image of Milan (Italy). The image size is about 25 km in ground range (vertical) and 25 km in azimuth (horizontal).

1.2.2.2 The phase SAR image

The radiation transmitted from the radar has to reach the scatterers on the ground and then come back to the radar in order to form the SAR image (two-way travel). Scatterers at different distances from the radar (different slant ranges) introduce different delays between transmission and reception of the radiation.

Due to the almost purely sinusoidal nature of the transmitted signal, this delay τ is equivalent to a phase change ϕ between transmitted and received signals. The phase change is thus proportional to the two-way travel distance $2R$ of the radiation divided by the transmitted wavelength λ. This concept is illustrated in Figure 1-5.

$$\phi = \frac{2\pi}{\lambda} 2R = \frac{4\pi}{\lambda} R$$
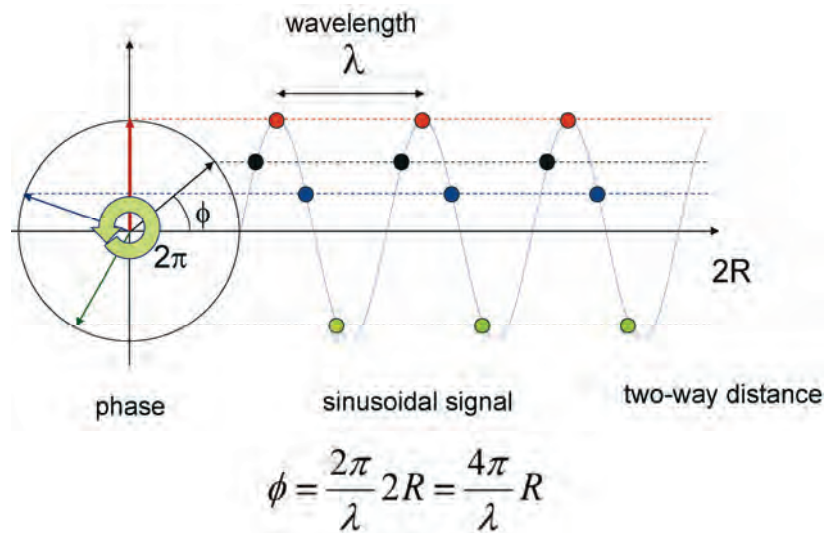
Figure 1-5: A sinusoidal function sin ϕ is periodic with a 2π radian period. In the case of a relative narrow-band SAR (i.e. ERS and Envisat), the transmitted signal can be assimilated, as a first approximation, to a pure sinusoid whose angle or phase ϕ has the following linear dependence on the slant range coordinate r: $\phi = 2\pi r/\lambda$ (where λ is the SAR wavelength). Thus, assuming that the phase of the transmitted signal is zero, the received signal that covers the distance $2R$ travelling from the satellite to the target and back, shows a phase $\phi = 4\pi R/\lambda$ radians.

However, due to the periodic nature of the signal, travel distances that differ by an integer multiple of the wavelength introduce exactly the same phase change. In other words the phase of the SAR signal is a measure of just the last fraction of the two-way travel distance that is smaller than the transmitted wavelength.

In practice, due to the huge ratio between the resolution cell dimension (of the order of a few metres) and wavelength (~5.6 cm for ERS), the phase change passing from one pixel to another within a single SAR image looks random and is of no practical utility.

1.2.2.3 Speckle

The presence of several scatterers within each SAR resolution cell generates the so-called 'speckle' effect that is common to all coherent imaging systems. Speckle is present in SAR, but not in optical images.

Homogeneous areas of terrain that extend across many SAR resolution cells (imagine, for example, a large agricultural field covered by one type of cultivation) are imaged with different amplitudes in different resolution cells. The visual effect is a sort of 'salt and pepper' screen superimposed on a uniform amplitude image.

This speckle effect is a direct consequence of the superposition of the signals reflected by many small **elementary** scatterers (those with a dimension comparable to the radar wavelength) within the resolution cell. These signals, which have random phase because of multiple reflections between scatterers, add to the directly reflected radiation. From an intuitive point of view, the resulting amplitude will depend on the imbalance between signals with positive and negative sign.

An example of speckle is shown in Figure 1-6. Here the 'salt and pepper' effect is clearly visible on the homogenous fields that surround the Linate Airport as seen by ERS-2.

The same area as seen from the SPOT optical system is shown in Figure 1-7. Here no speckle is present and the fields that surround the Linate Airport appear homogeneous.
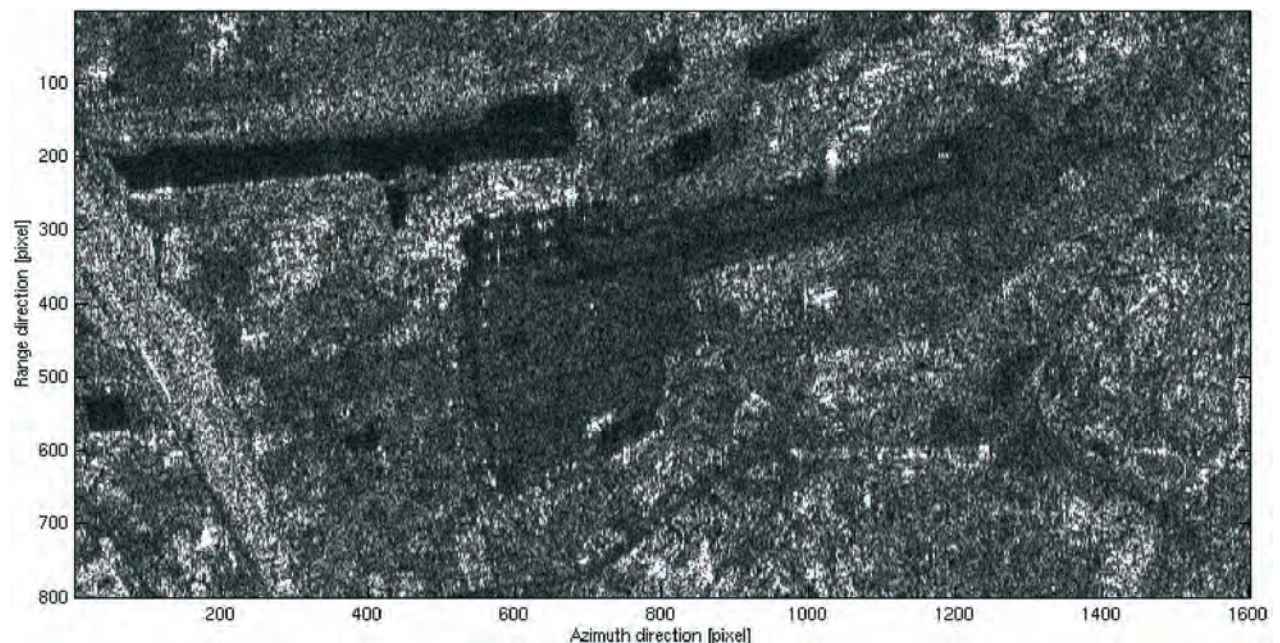
Figure 1-6: ERS-2 SAR detected image of the Linate Airport in the eastern part of Milan (Italy): the speckle effect on the homogeneous fields surrounding the airport is clearly visible.

Figure 1-7: Optical image of Linate Airport taken from the SPOT satellite. No speckle is visible and the fields that surround the airport look homogeneous.

Speckle has an impact on the quality and usefulness of detected SAR images. Typically, image segmentation suffers severely from speckle. However, by taking more images of the same area at different times or from slightly different look angles, speckle can be greatly reduced: averaging several images tends to cancel out the random amplitude variability and leave the uniform amplitude level unchanged.

An example of speckle reduction is shown in Figure 1-8. Here the average of 60 separate ERS-1 and ERS-2 SAR images of the area surrounding the Linate airport in Milan is shown. A comparison between this image and the single SAR image shown in Figure 1-6 gives an idea of the speckle reduction achieved and of the improved visibility of detail.
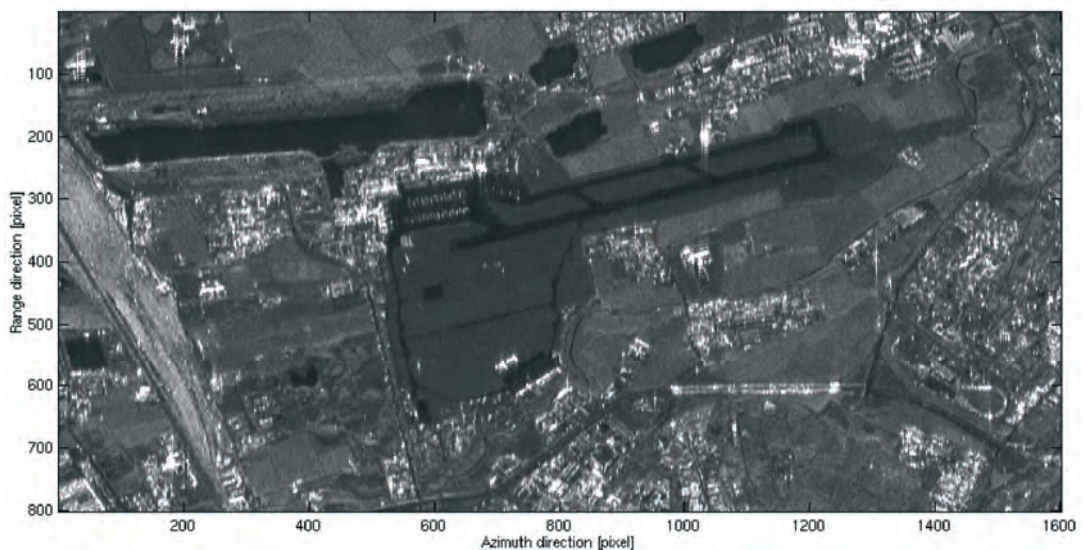
Figure 1-8: Average of multiple ERS SAR images of Linate airport: the speckle effect on the homogeneous fields around the airport has disappeared.

1.2.3 SAR resolution cell projection on the ground

The terrain area imaged in each SAR resolution cell (called the **ground resolution cell**) depends on the local topography. It strongly depends on the terrain slope in the plane perpendicular to the orbit (ground range direction), and on the terrain slope in the azimuth direction.

The dimension of the ground resolution cell in azimuth is related to that of the SAR resolution cell by the usual perspective deformation we experience every day looking at surfaces from different angles (e.g. a postcard seen at 90 degrees is a line).

The dimension of the ground resolution cell in range is related to that of the SAR resolution cell by an unusual perspective deformation. Figure 1-9 shows how slant-range is projected onto the ground.

Here five identical slant-range resolution cells are shown. The first two cells correspond to flat terrain and they contain three triangles each. The third cell contains seven of these triangles due to the positive slope of the terrain. Finally the fourth and fifth cells contain less than three triangles due to the negative slope of the terrain.
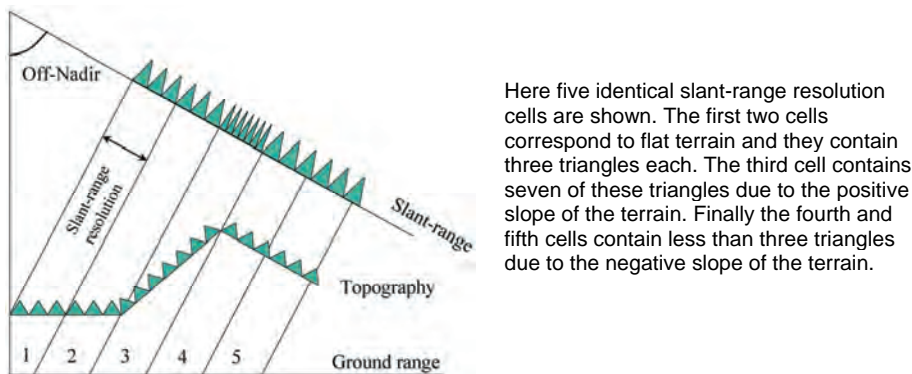
Figure 1-9: Effect of terrain on the SAR image. For SAR resolution cells in the plane perpendicular to the orbit, the part of the terrain imaged in each resolution cell clearly depends on the topography.

As the terrain slope increases with respect to a flat horizontal surface (i.e. as the normal to the terrain moves toward the **line of sight (LOS)**), the ground resolution cell dimension in range increases. This effect is called foreshortening. When the terrain slope is close to the radar off-nadir angle, the cell dimension becomes very large and all the details are lost. Moreover, when the terrain slope exceeds the radar off-nadir angle the scatterers are imaged in reverse order and superimposed on the contribution coming from other areas. This effect is called **layover**, and is sketched in Figure 1-10.
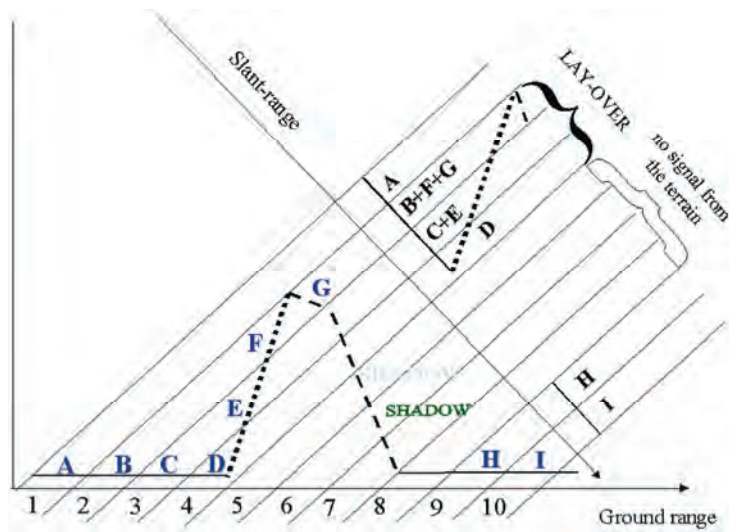
Figure 1-10: Layover and shadow effects. Depending on the terrain slope, scatterers that are located at increasing ground range positions can be imaged in reverse order by the SAR system (points D, E, F and G). Moreover they are imaged in the same SAR resolution cells as scatterers B and C, which belong to a different area on the ground (layover). On the other side of the elevation profile, scatterers located between points G and H cannot be illuminated by the radar since they are in shadow. As a consequence, SAR resolution cells from 5 to 8 do not contain any signal from the ground and they generate a dark gap on the detected image.

On the other hand, when the terrain slope decreases with respect to the flat horizontal reference surface, the resolution cell dimension decreases. The minimum resolution cell dimension (i.e. equal to the slant range resolution) is reached when the terrain is parallel to the LOS. This is also the lower slope limit that can be imaged at all by a SAR system, since beyond this angle the terrain is in shadow.

As an example, in the case of ERS systems, the resolution cell dimension as a function of the terrain slope is shown in Figure 1-11.
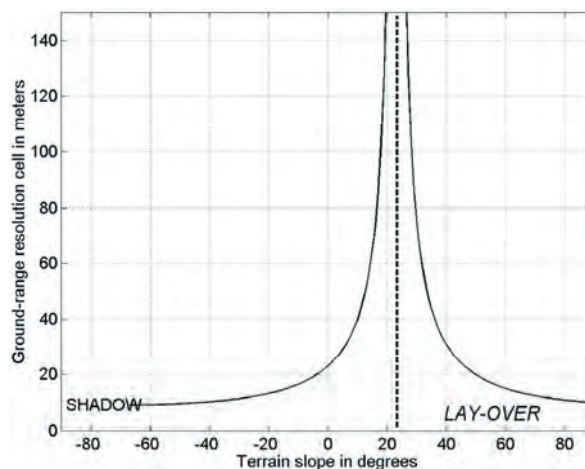
Figure 1-11: ERS resolution cell dimension in ground range as a function of the terrain slope. The vertical dotted line indicates the incidence angle relative to a flat horizontal terrain (23°).

It should be pointed out that foreshortening has a strong impact on the amplitude of the detected SAR image. Foreshortened areas are brighter on the image because the resolution cell is larger (hence more power is backscattered towards the satellite) and the incidence angle is steeper. An example that illustrates this effect is shown in Figure 1-12 with reference to the area of Mount Vesuvius (Italy) as seen by ERS-1.

Figure 1-12: ERS-2 SAR image of Mount Vesuvius (Italy), as detected. The slant range direction is vertical on the image (near range is in the upper part of the image). Brighter levels correspond to stronger backscattered radiation. The coastline along the Tyrrenian Sea is clearly visible (the sea is dark due to the almost specular reflection of electromagnetic waves). Urban areas can be identified as bright spots on the image (strong backscattering from buildings), as can the main crater of the volcano. It should be noted that positive slopes of the volcano (in the upper flanks on the image) are shortened with respect to the flanks descending to the sea. At the same time, the shortened flanks appear brighter on the image.

It should be pointed out that foreshortening has a strong impact on the amplitude of the detected SAR image. Foreshortened areas are brighter on the image because the resolution cell is larger (hence more power is backscattered towards the satellite) and the incidence angle is steeper. An example that illustrates this effect is shown in Figure 1-12 with reference to the area of Mount Vesuvius (Italy) as seen by ERS-1.

Figure 1-12: ERS-2 SAR image of Mount Vesuvius (Italy), as detected. The slant range direction is vertical on the image (near range is in the upper part of the image). Brighter levels correspond to stronger backscattered radiation. The coastline along the Tyrrenian Sea is clearly visible (the sea is dark due to the almost specular reflection of electromagnetic waves). Urban areas can be identified as bright spots on the image (strong backscattering from buildings), as can the main crater of the volcano. It should be noted that positive slopes of the volcano (in the upper flanks on the image) are shortened with respect to the flanks descending to the sea. At the same time, the shortened flanks appear brighter on the image.

Referring to the same area, Figure 1-13 shows how the regular resolution grid in SAR coordinates (azimuth and slant-range) is deformed by the topography when projected on the ground.
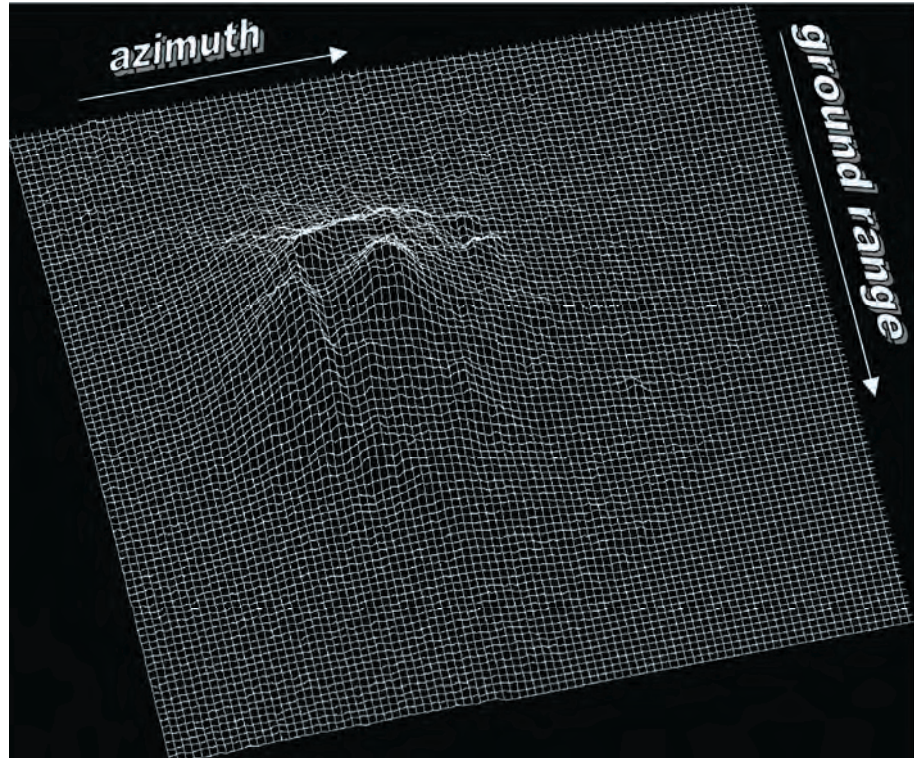
Figure 1-13: Deformation of the regular resolution grid on the ground when projected in SAR coordinates. The deformation is due to topography.

1.2.3.1 Geometric deformation from ascending and descending ERS passes

With ERS there is the possibility to observe the same scene with incidence angles of both plus and minus 23 degrees. Observation of the whole of the Earth's surface is achieved by combination of the orbital satellite motion along the meridians (almost polar orbits) and the Earth's rotation in the equatorial plane. This possibility comes from the fact that during orbits that go from South to North (**ascending passes**) and from North to South (**descending passes**), the SAR antenna pointing is usually fixed to the same side of the orbital plane with respect to the velocity vector (e.g. the radar antenna is always pointed to the right side of the track for ERS and Envisat), as shown in Figure 1-14.

Thus, the same scene on the ground is observed by the SAR antenna from the east during the descending passes and from the west during the ascending passes.

Here two detected ERS images of Mount Etna (Italy) taken from ascending and descending passes are shown together with an elevation model of the imaged area. A comparison of these two images clearly shows the effect of the different perspective: the summit is shifted away from the coastline in the ascending (left) ERS SAR image and towards it in the descending (right) image. From these images it is also evident that high resolution details of the western flank of the volcano are obtained from ERS ascending passes, whereas the eastern flank is 'squeezed' into a few pixels of the SAR image; the opposite happens with descending ERS passes. Thus, both ascending and descending passes should be exploited to get a high resolution SAR image of the whole area. It is necessary, however, to resample both images on a common reference grid in order to be able to make such a combination.

2. SAR interferometry: applications and limits

This section gives a brief overview of the subject. Details of the method may be found in parts B and C of this document.

2.1 Introduction

A satellite SAR can observe the same area from slightly different look angles. This can be done either simultaneously (with two radars mounted on the same platform) or at different times by exploiting repeated orbits of the same satellite. The latter is the case for ERS-1, ERS-2 and Envisat. For these satellites, time intervals between observations of 1, 35, or a multiple of 35 days are available.

The distance between the two satellites (or orbits) in the plane perpendicular to the orbit is called the **interferometer baseline** (see Figure 2-1) and its projection perpendicular to the slant range is the **perpendicular baseline**.
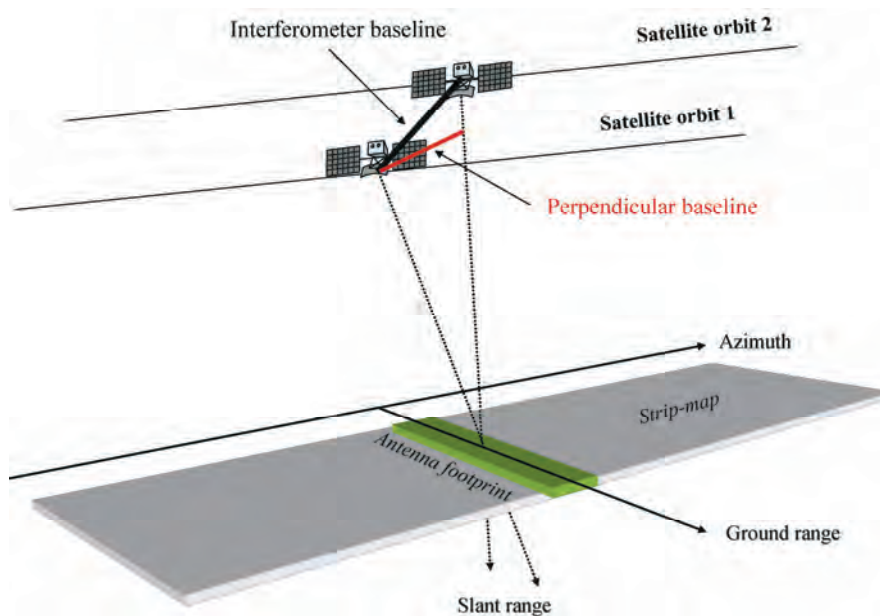
Figure 2-1: Geometry of a satellite interferometric SAR system. The orbit separation is called the interferometer baseline, and its projection perpendicular to the slant range direction is one of the key parameters of SAR interferometry.

The SAR interferogram is generated by cross-multiplying, pixel by pixel, the first SAR image with the complex conjugate of the second [Bamler98A, Massonnet98, Franceschetti99, Rosen00]. Thus, the interferogram amplitude is the amplitude of the first image multiplied by that of the second one, whereas its phase (the **interferometric phase**) is the phase difference between the images.

2.2 Terrain altitude measurement through the interferometric phase

Let us suppose we have only one dominant point scatterer in each ground resolution cell that does not change over time. These point scatterers are observed by two SARs from slightly different look angles as shown in Figure 2-1. In this case the interferometric phase of each SAR image pixel would depend only on the difference in the travel paths from each of the two SARs to the considered resolution cell. Any possible phase contribution introduced by the point scatterers does not affect the interferometric phase since it is cancelled out by the difference.

Once a ground reference point has been identified, the **variation** of the travel path difference Δr that results in passing from the reference resolution cell to another can be given by a simple expression (an approximation that holds for small baselines and resolution cells that are not too far apart) that depends on a few geometric parameters shown in Figure 2-2.
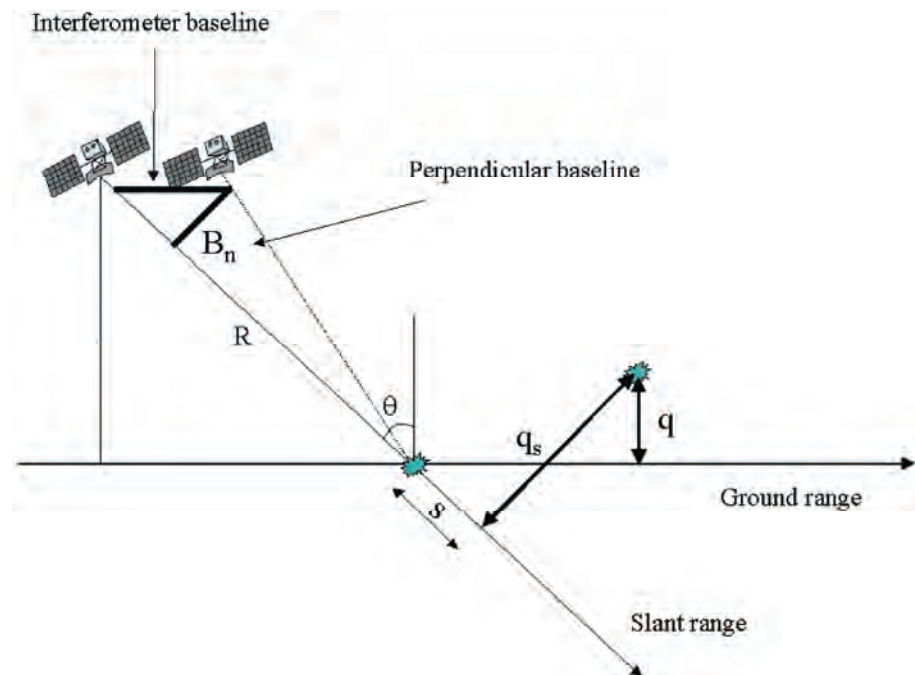
Figure 2-2: Geometric parameters of a satellite interferometric SAR system

The parameters are:

1. The perpendicular baseline B_n
2. The radar-target distance R
3. The displacement between the resolution cells along the perpendicular to the slant range, q_s

The following approximated expression of Δr holds:

$$\Delta r = -2 \frac{B_n q_s}{R}$$

Equation 2.1

The interferometric phase variation $\Delta\phi$ is then proportional to Δr divided by the transmitted wavelength λ:

$$\Delta\phi = \frac{2\pi\Delta r}{\lambda} = \frac{4\pi}{\lambda}\frac{B_n q_s}{R} \qquad\qquad \text{Equation 2.2}$$

2.2.1 Interferogram flattening

The interferometric phase variation can be split into two contributions:

1. A phase variation proportional to the altitude difference q between the point targets, referred to a horizontal reference plane

2. A phase variation proportional to the slant range displacement s of the point targets

$$\Delta\phi = -\frac{4\pi}{\lambda}\frac{B_n q}{R\sin\theta} - \frac{4\pi}{\lambda}\frac{B_n s}{R\tan\theta} \qquad\qquad \text{Equation 2.3}$$

where θ is the radiation incidence angle with respect to the reference

It should be noted that the perpendicular baseline is known from precise orbital data, and the second phase term can be computed and subtracted from the interferometric phase. This operation is called **interferogram flattening** and, as a result, it generates a phase map proportional to the relative terrain altitude.

An example of interferogram flattening is shown in Figure 2-3. An interferogram of a portion of the Italian Alps and the Pianura Padana that has been obtained from ERS-1 and ERS-2 data (taken one day apart with a normal baseline of about 30 metres) is shown on the left. The flattened interferogram is shown on the right side. Here the phase discontinuities resemble the contour lines. The altitude between two adjacent discontinuities is called the **altitude of ambiguity** (symbol h_a) and can be computed from the interferometer parameters (see section 2.2.2).
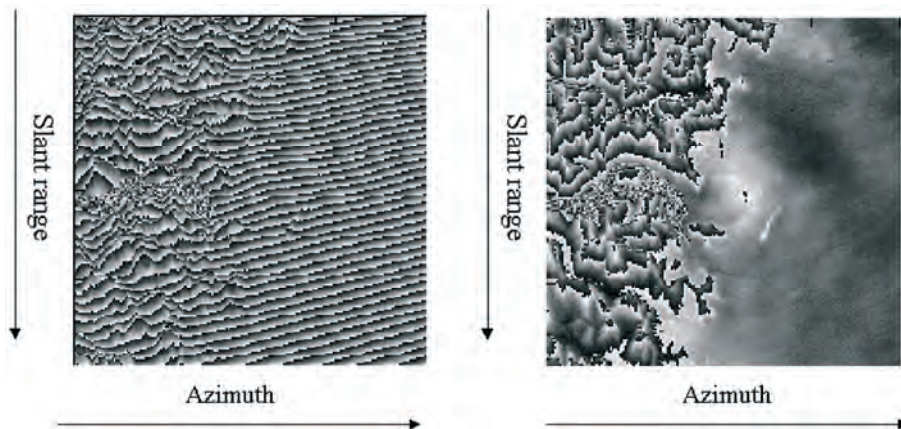
Figure 2-3: Left: interferogram of a portion of the Italian Alps and the Pianura Padana that has been obtained from ERS data. Right: flattened interferogram. Here the phase discontinuities resemble the contour lines.

2.2.2 Altitude of ambiguity

The altitude of ambiguity h_a is defined as the altitude difference that generates an interferometric phase change of 2π after interferogram flattening. The altitude of ambiguity is inversely proportional to the perpendicular baseline:

$$h_a = \frac{\lambda R \sin \theta}{2 B_n}$$

Equation 2.4

In the ERS case with λ = 5.6 cm, θ = 23º, and R = 850 km, the following expression holds (in metres):

$$h_a \approx \frac{9300}{B_n}$$

Equation 2.5

As an example, if a 100 metre perpendicular baseline is used, a 2π change of the interferometric phase corresponds to an altitude difference of about 93 metres. In principle, the higher the baseline the more accurate the altitude measurement, since the phase noise (see next section) is equivalent to a smaller altitude noise. However, it will be shown later that there is an upper limit to the perpendicular baseline, over which the interferometric signals decorrelate and no fringes can be generated. In conclusion there is an optimum perpendicular baseline that maximises the signal to noise power ratio (where the signal is terrain altitude). In the ERS case, such an optimum baseline is about 300–400 metres.

2.2.3 Phase unwrapping and DEM generation

The flattened interferogram provides an ambiguous measurement of the relative terrain altitude due to the 2π cyclic nature of the interferometric phase. The phase variation between two points on the flattened interferogram provides a measurement of the actual altitude variation, after deleting any integer number of altitudes of ambiguity (equivalent to an integer number of 2π phase cycles). The process of adding the correct integer multiple of 2π to the interferometric fringes is called **phase unwrapping**.

An example of phase unwrapping is shown in the following figure, in which the SAR interferometric phase, its unwrapped version and a map with the correct integer multiple of 2π added to the original phase are shown together.
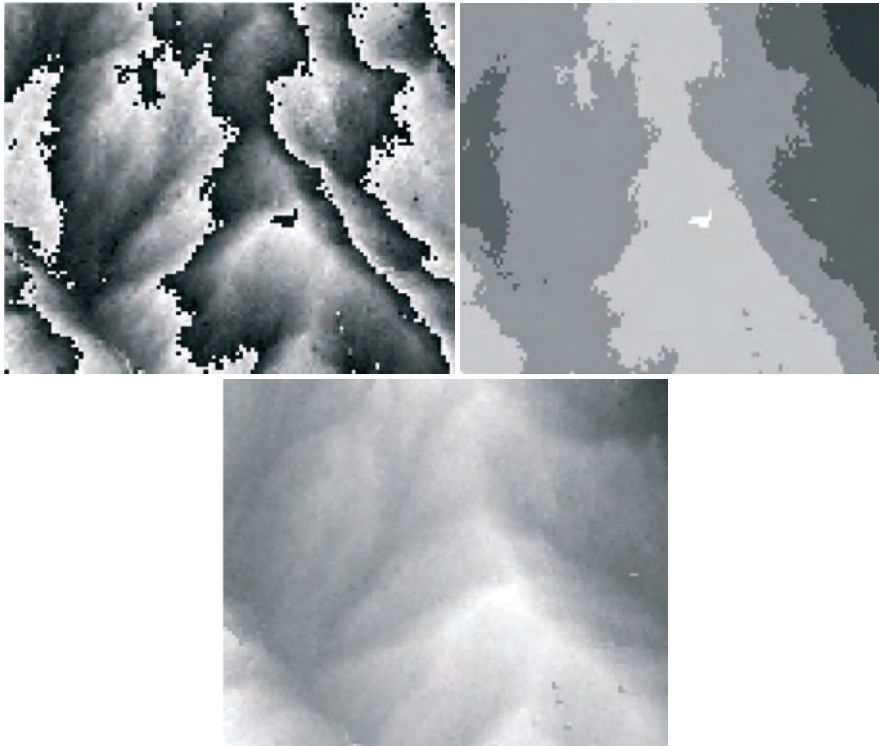
*Figure 2-4: Left: SAR interferometric phase generated by means of two ERS
images. The 2π phase discontinuities are clearly visible as black/white transitions.
Right: Unwrapped phase. Below: The 2π phase discontinuities have been
eliminated by adding or subtracting an integer multiple of 2π to each pixel of the
original interferometric phase image.*

There are several well-known phase unwrapping techniques that will be
described in the advanced section of this manual. However it should be
noted here that usually phase unwrapping does not have a unique solution,
and *a priori* information should be exploited to get the right solution.

Once the interferometric phases are unwrapped, an elevation map in SAR
coordinates is obtained. This is the first step towards getting a DEM. The
SAR elevation map should then be referred to a conventional ellipsoid (e.g.
WGS84) and re-sampled on a different grid (for example UTM).

As an example, the flattened interferogram and the relative DEM of Mount
Etna obtained through phase unwrapping and re-sampling are shown in
Figure 2-5, Figure 2-6 and Figure 2-7.
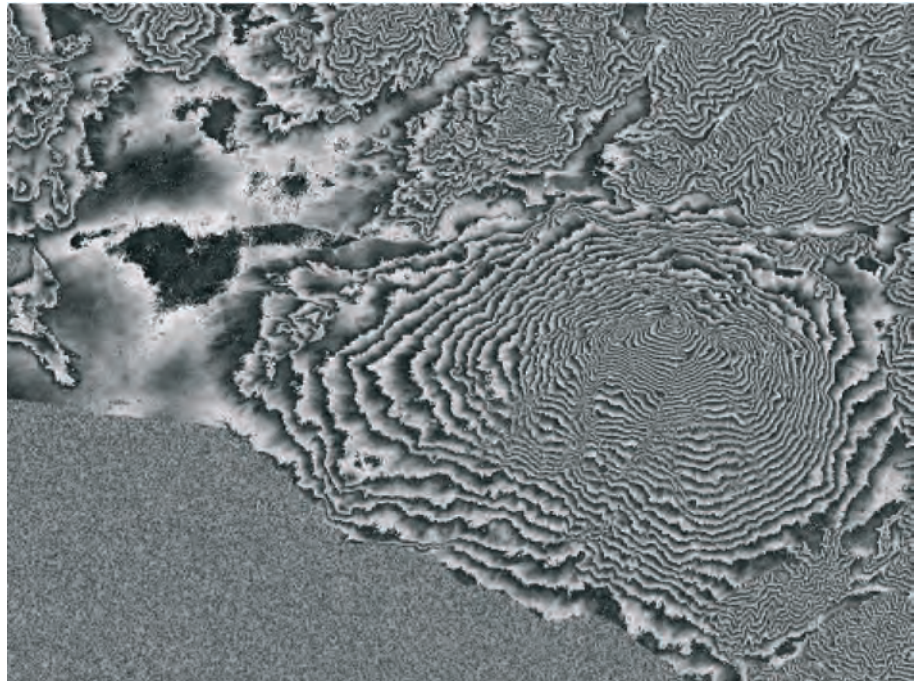
Figure 2-5: Flattened interferogram of Mount Etna generated from ERS tandem pairs. The perpendicular baseline of 115 metres generates an altitude of ambiguity of about 82 metres.
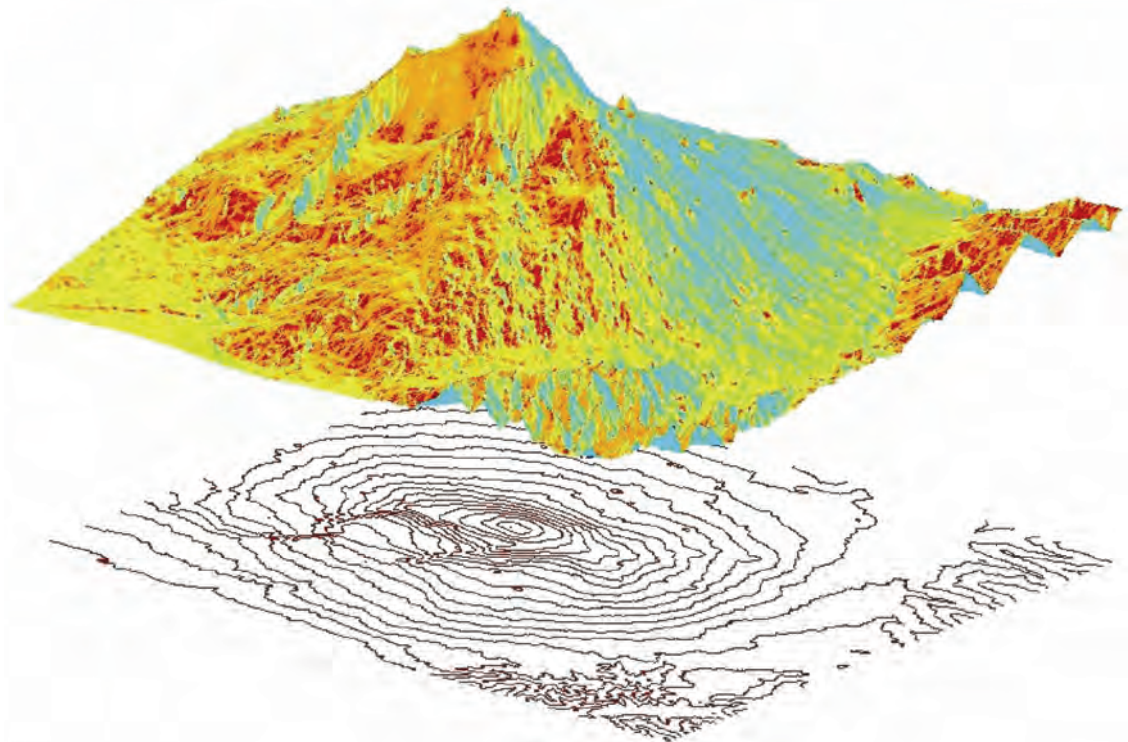
Figure 2-6: Perspective view of Mount Etna as seen from the Northeast. The DEM of Mount Etna has been generated by unwrapping and re-sampling the flattened interferogram of Figure 2-5: The estimated vertical accuracy is better than 10 metres. Contour lines are shown below the DEM.
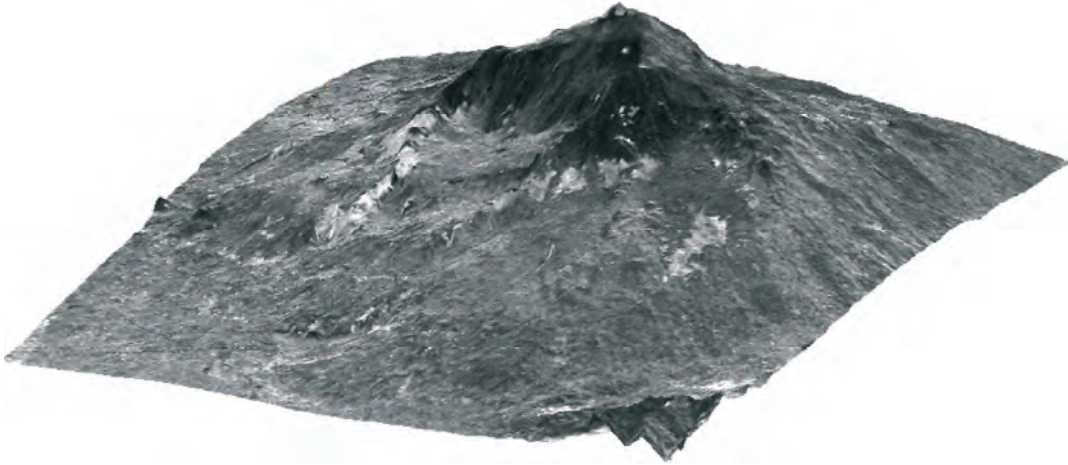
Figure 2-7: Perspective view of Mount Etna as seen from the Northeast. The average of many detected ERS SAR images has been draped on the DEM.

2.3 Terrain motion measurement: Differential Interferometry

Suppose that some of the point scatterers on the ground slightly change their relative position in the time interval between two SAR observations (as, for example, in the event of subsidence, landslide, earthquake, etc.). In such cases the following additive phase term, independent of the baseline, appears in the interferometric phase:

$$\Delta\phi_d = \frac{4\pi}{\lambda}d \qquad\qquad \text{Equation 2.6}$$

> where d is the relative scatterer displacement projected on the slant range direction.

This means that after interferogram flattening, the interferometric phase contains both altitude and motion contributions:

$$\Delta\phi = -\frac{4\pi}{\lambda}\frac{B_n q}{R\sin\theta} + \frac{4\pi}{\lambda}d \qquad\qquad \text{Equation 2.7}$$

Moreover, if a digital elevation model (DEM) is available, the altitude contribution can be subtracted from the interferometric phase (generating the so-called **differential interferogram**) and the terrain motion component can be measured. In the ERS case with $\lambda = 5.6$ cm and assuming a perpendicular baseline of 150 m (a rather common value), the following expression holds:

$$\Delta\phi = -\frac{q}{10} + 225\,d \qquad\qquad \text{Equation 2.8}$$

From this example it can be seen that the sensitivity of SAR interferometry to terrain motion is much larger than that to the altitude difference. A 2.8 cm motion component in the slant range direction would generate a 2π interferometric phase variation. As an example, the differential

interferogram showing the surface deformation that occurred during the Mount Etna eruption of July 2001 is shown in Figure 2-8.
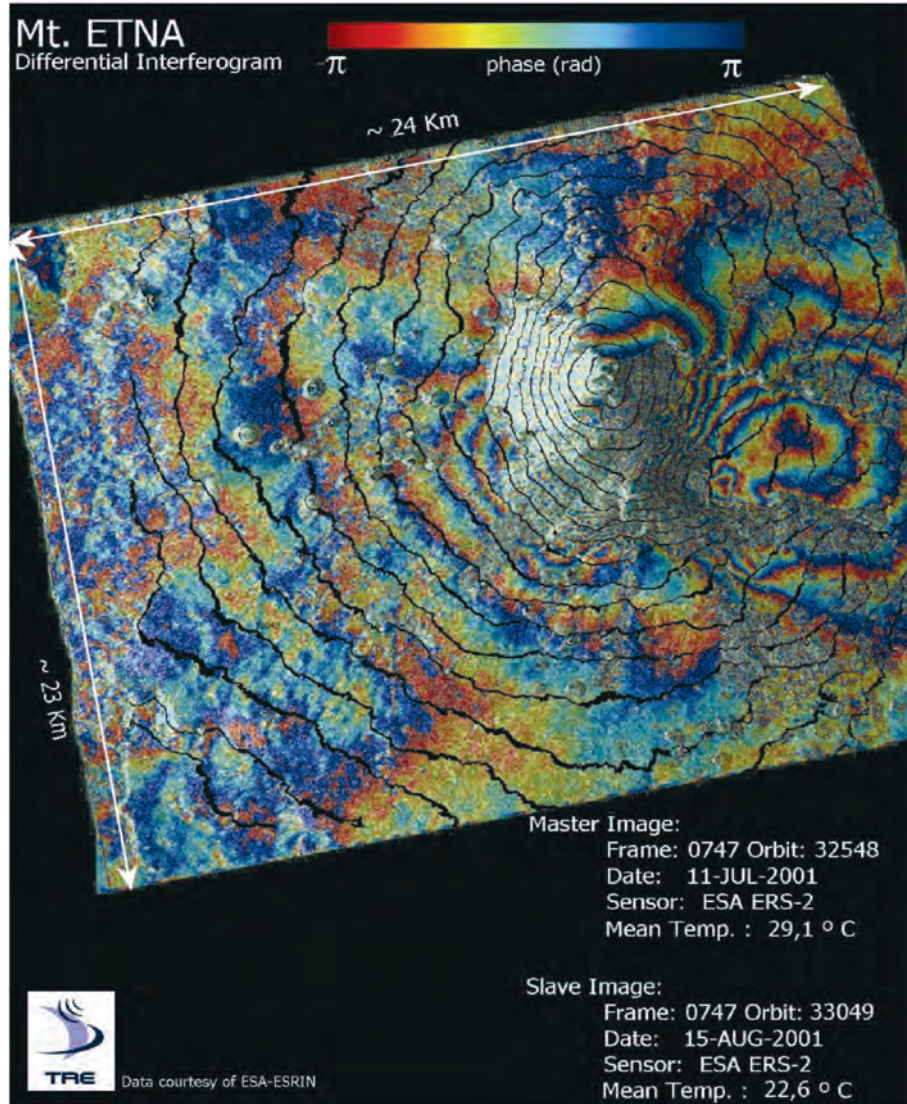
Figure 2-8: The differential interferogram of the Mt Etna eruption that occurred in July 2001. The interferogram has been generated by means of two ERS images taken before (11 July 2001) and after (15 August 2001) the eruption. The topography has been removed by means of an available DEM. Contour lines of the DEM are shown in black.

2.4 The atmospheric contribution to the interferometric phase

When two interferometric SAR images are not simultaneous, the radiation travel path for each can be affected differently by the atmosphere. In particular, different atmospheric humidity, temperature and pressure between the two takes will have a visible consequence on the interferometric phase. This effect is usually confined within a 2π peak-to-peak interferometric

phase change along the image with a smooth spatial variability (from a few hundred metres to a few kilometres). The effect of such a contribution impacts on both altitude (especially in the case of small baselines) and terrain deformation measurements.

As an example, the atmospheric phase contribution to the ERS interferogram generated on the Pianura Padana valley (North Italy) is shown in Figure 2-9. Here the perpendicular baseline is quite small (30 metres) and the differential turbulence effect is clearly visible on the interferogram where an almost flat phase contribution is expected from the known topography.
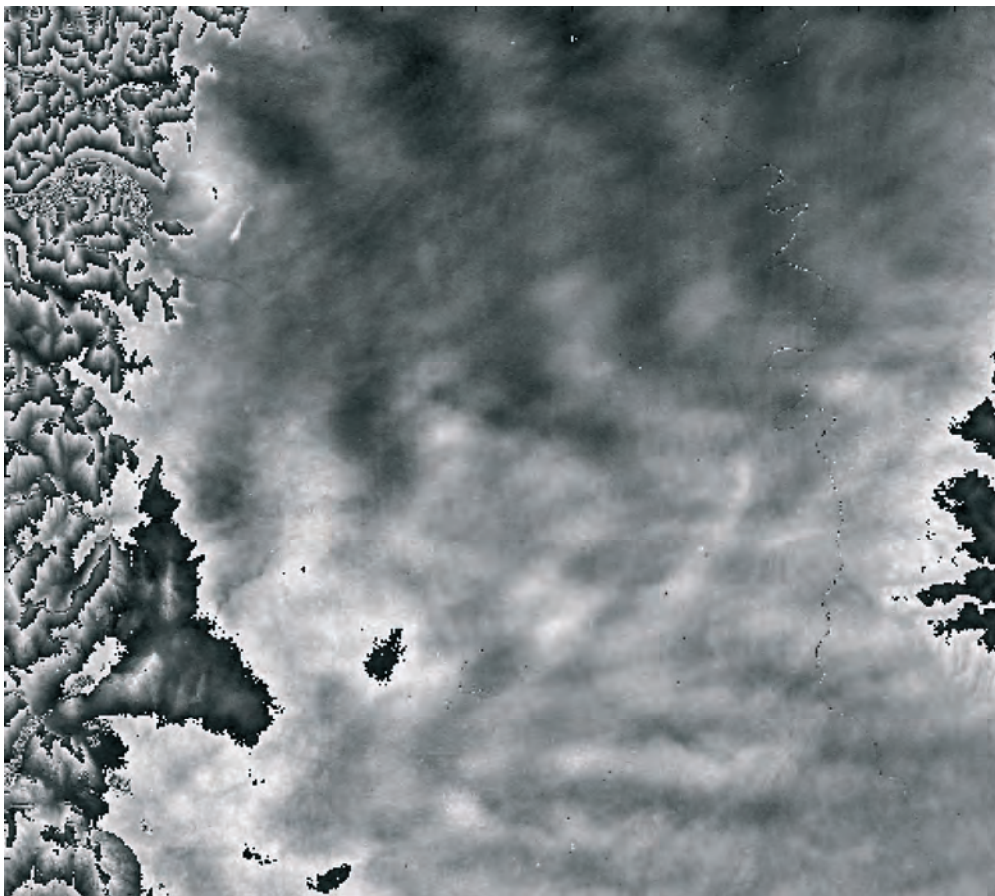
Figure 2-9: An example of atmospheric phase contribution to the ERS interferogram generated on the Pianura Padana. The perpendicular baseline is about 30 metres and the altitude of ambiguity (from black to white in the grey scale used) is about 300 metres.

2.5 Other phase noise sources

In the previous sections it has been hypothesised that only one dominant stable scatterer was present in each resolution cell. This is seldom the case in reality. We should analyse the situation where many elementary scatterers are present in each resolution cell (distributed scatterers), each of which may change in the time interval between two SAR acquisitions. The main effect of the presence of many scatterers per resolution cell and their changes in time is the introduction of phase noise.

Three main contributions to the phase noise should be taken into consideration:

1. **Phase noise due to temporal change of the scatterers**

   In the case of a water basin or densely vegetated areas, the scatterers change totally after a few milliseconds, whereas exposed rocks or urban areas remain stable even after years. Of course, there are also the intermediate situations where the interferometric phase is still useful even if corrupted by change noise.

2. **Phase noise due to different look angle**

   Speckle will change due to the different combination of elementary echoes even if the scatterers do not change in time. The most important consequence of this effect is that there exists a **critical baseline** over which the interferometric phase is pure noise. The critical baseline depends on the dimension of the ground range resolution cell (and thus also on the terrain slope), on the radar frequency, and on the sensor-target distance. In the ERS case, the critical baseline for horizontal terrain is about 1150 metres. It decreases for positive terrain slopes and increases for negative ones. This phase noise term, however, can be removed from the interferogram by means of a pre-processing step of the two SAR images known as **spectral shift** or **common band** filtering. This will be described in detail in the advanced sections of this manual (part B section 2.5 and part C section 2.2.1).

3. **Phase noise due to volume scattering**

   The critical baseline reduces in the case of volume scattering when the elementary scatterers are not disposed on a plane surface but occupy a volume (e.g. the branches of a tree). In this case the speckle change depends also on the depth of the volume occupied by the elementary scatterers.

2.6 Coherence maps

The phase noise can be estimated from the interferometric SAR pair by means of the local **coherence** γ. The local coherence is the cross-correlation coefficient of the SAR image pair estimated over a small window (a few pixels in range and azimuth), once all the deterministic phase components (mainly due to the terrain elevation) are compensated for.

The deterministic phase components in such a small window are, as a first approximation, linear both in azimuth and slant-range. Thus, they can be estimated from the interferogram itself by means of well-known methods of frequency detection of complex sinusoids in noise (e.g. 2-D **Fast Fourier Transform (FFT)**).

The coherence map of the scene is then formed by computing the absolute value of γ on a moving window that covers the whole SAR image.

The coherence value ranges from 0 (the interferometric phase is just noise) to 1 (complete absence of phase noise).

As an example, a coherence map of the North East part of Sicily is shown in Figure 2-10. Here the exposed lava on Mount Etna shows a very high coherence value, whereas vegetated areas appear dark, showing lower coherence values. Note the very low coherence value of the sea (dark in the image), which changes completely in the one day interval between the two ERS observations.

Figure 2-10: Coherence map of the North East part of Sicily

The exact relation between the interferometric phase dispersion and coherence can be found through complicated mathematical computation [Lee94]. However, if the number of looks (*NL*) is greater than four, then independent pixels with the same coherence are averaged after topography compensation (multi-look interferogram) and the following simple approximation holds [Rosen00]:

$$\sigma_\phi = \frac{1}{\sqrt{2NL}} \frac{\sqrt{1-\gamma^2}}{\gamma} \qquad \text{Equation 2.9}$$

From a mathematical point of view, this formula is a good approximation of the exact phase dispersion shown in Figure 2-11 when $\sigma_\phi < 12°$ That is, when *NL* is large and γ close to one. However, for most practical applications of SAR interferometry, the approximated formula can be suitably exploited for coherence values higher than 0.2 and *NL* > 4.
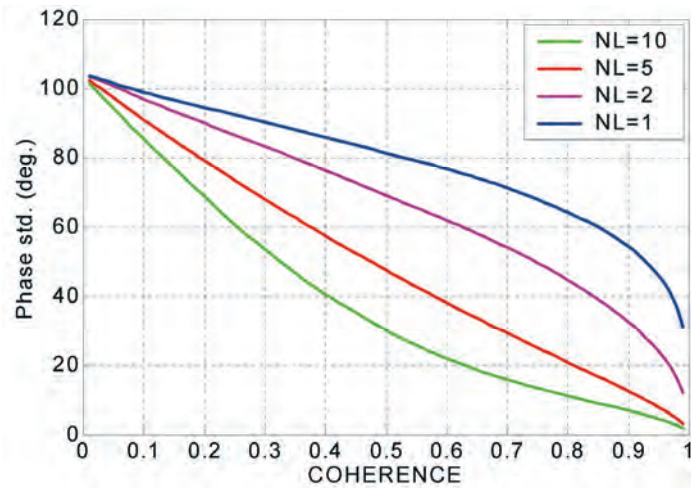
Figure 2-11: Interferometric phase dispersion (degrees) as a function of the coherence for varying numbers of looks (NL)

A comparison between the exact and approximated curves is shown in Figure 2-12.
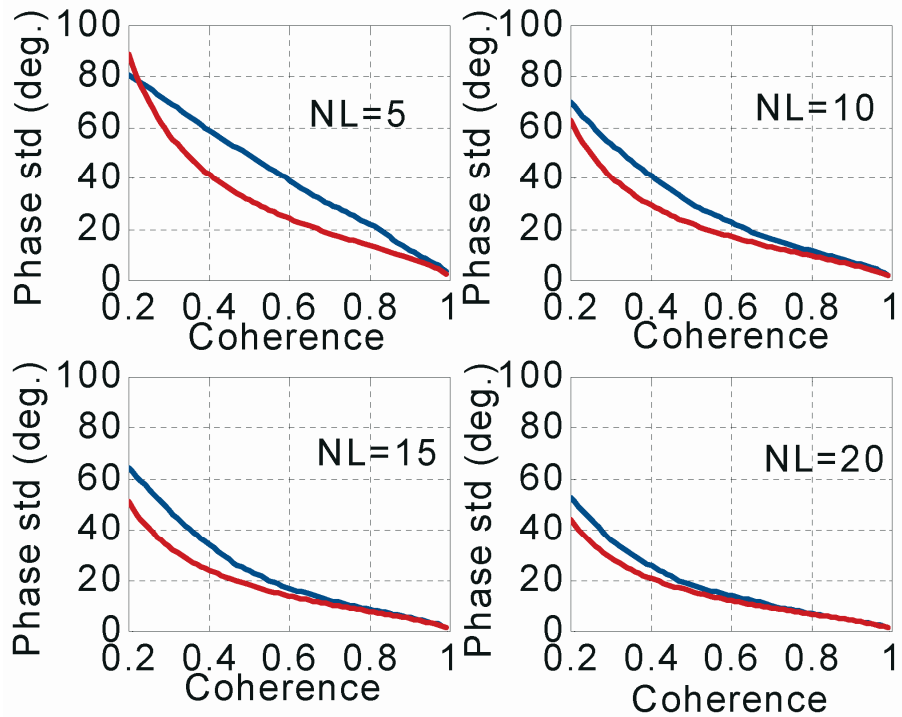
Figure 2-12: Interferometric phase dispersion exact values (blue curves) and approximated ones (red curves)

The phase dispersion can be exploited to estimate the theoretical elevation dispersion (limited to the high spatial frequencies) of a DEM generated from SAR interferometry:

$$\sigma_h = \sigma_\phi \frac{R\lambda \sin\theta}{4\pi B} \qquad \text{Equation 2.10}$$

On the other hand, low spatial frequencies of the DEM error cannot be predicted from the coherence map since the coherence estimation is carried out on small windows. The information carried by the coherence map can be usefully exploited to help image segmentation, as will be seen in part B.

3. SAR Differential Interferometry basics and examples

3.1 Introduction

'Differential interferometry' is the commonly used term for the production of interferograms from which the topographic contribution has been removed. However, the term may occasionally be misleading, because on the one hand interferometry is a differential technique right from the beginning, and on the other hand, the subtraction process can be pushed further as well as in other directions (e.g. subtraction of an expected geophysical contribution through earthquake or volcano dynamic modelling).

3.2 Landers co-seismic deformation

On 18 June 1992, a very large earthquake occurred in the desert northeast of the city of Los Angeles. It was named after the small city of Landers, which is nearby this largely unpopulated area. Its magnitude of 7.3 on the Richter scale made it one of the largest of the century in California. The earthquake was strongly felt in the w1hole area, including Los Angeles, but it caused few casualties and little damage because of its remote location. To geophysicists, the Landers Earthquake was an excellent opportunity to study the mechanisms of a large earthquake using the most recent geodetic and seismological instrumentation, which had previously been put in the field in the area, making it (along with Japan) one of the most densely instrumented areas in the world.

In a much less publicised way, the radar imaging community was eager to demonstrate the power of radar interferometry applied to displacement mapping. The Landers Earthquake was also an excellent opportunity for radar scientists. It took place after ERS-1 had been placed on its 35-day orbit, which was to be maintained for most of the useful life of the satellite and which guaranteed a regular flow of high quality data. The desert environment raised some hopes that the interferometric comparison of radar images acquired a long time apart could work, since the degradation of the soil during the time elapsed might be minimal, despite some previous pessimistic estimations that predicted a decay in a matter of days. Another positive was the availability of a topographic model of the area, of reasonable accuracy. Such a model would allow removal of the effect of topography in the interferograms (section C6.2), so that just two radar images would suffice to catch the displacements. Finally, that the area was so heavily instrumented and being studied was a great benefit, because other geodetic measurements could at the same time confirm the radar measurement and provide the highest level of 'geodetic competition' against interferometry.

The study of the Landers Earthquake actually exceeded all expectations. In the first study [Massonnet93] two images acquired before the earthquake

(24 April 1992) and after it (7 August 1992) combined into a nice interferogram despite the 105 days elapsed. A third image (3 July 1992) was used in combination with the 7 August image. This combination did not include the earthquake. It demonstrated the quality of the topographic model used in conjunction with the first interferogram and produced the error bars. The result of the study went beyond the mere demonstration of interferometry. It was a big surprise for geophysicists, who did not expect such a revolutionary way of looking at the Earth, but it also sets a new aesthetic standard in the geosciences. The cover of the magazine Nature popularised the 'fringes' as a new way to look at ground deformation with coloured, and sometimes shaky, contour lines, each amounting to 3 cm or so of additional deformation.

The interferometric image of Figure 3-1 provides a striking collision of scales: it shows the central part of the 100 km by 300 km area under study, where displacements are recorded with millimetre accuracy from 800 km away in space. The ratio of the width of the scene and the potential accuracy is 10^8. The ratio of the distance of observation and the maximum amplitude of the displacement in the image is 10^6.

Figure 3-1: The Landers Earthquake of 18 June 1992

One year later, another study [Massonnet94] demonstrated that the fringes were even more robust than anticipated. The interferogram in Figure 3-1 was actually part of this second study, and was made from two images separated by 18 months. Landers was also a good test site for demonstrating the method using three radar images (section C.6.3), which does not need a topographic model [Zebker94B] and for testing various mixes of geodetic and seismological data [Hernandez96] to refine earthquake modelling.

Because it created a large surface rupture, the Landers Earthquake could be modelled rather accurately by elastic modelling (see an example of the latter in Figure 3-2) based on the rupture parameters, which are easier to determine when they are evidenced by fault shifts that reach the surface. The striking resemblance between the artificial fringes inferred from the geophysical elastic modelling and the actual interferogram was crucial to making people simply believe the result. The Big Bear Earthquake that took place three hours after Landers did not create a surface rupture and was much more difficult to model. On the interferogram, it is the set of six or seven large, circular fringes south of Landers. The same interferogram thus proved both the validity of the method (with Landers) and its unique capabilities (with Big Bear).

3.3 Small earthquake modelling

Unlike Landers, the earthquake that struck the northern side of the San Bernardino mountain range had nothing spectacular to draw attention. It had a small magnitude of 5.1 and was located far from populated centres. It took place on 4 December 1992, more than five months after the Landers Earthquake. It was, however, well recorded as a small concentric deformation in the southwest part of the previous illustration, which is considerably zoomed on the left panel of Figure 3-2. We dubbed this nameless small earthquake 'Fawnskin', after the name of the corresponding USGS topographic map.
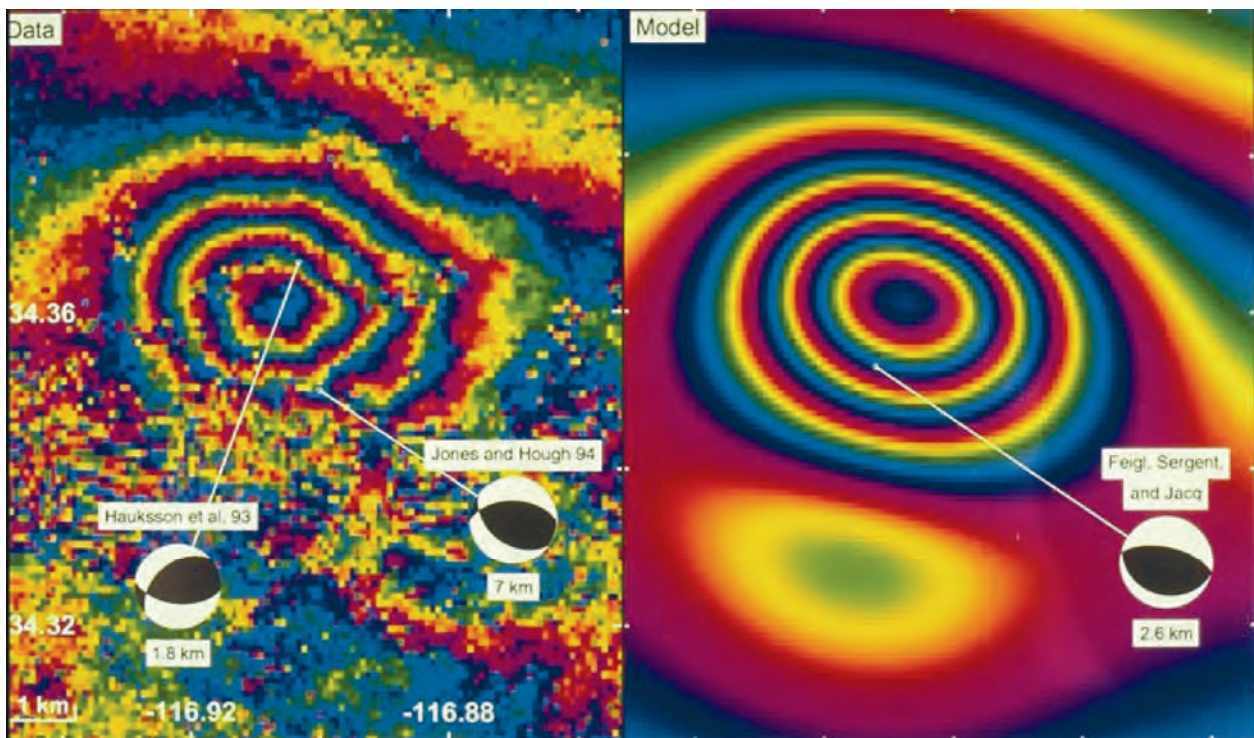
Figure 3-2: Fawnskin earthquake and its modelling

In interferometry, all the deformations occurring in the time elapsed between the images are stacked together, regardless of their date. The event is therefore superposed on a network of fringes created by the other, more powerful, earthquake in the area. We can see two of the fringes that cross our zoom obliquely. The smaller earthquake created four fringes of its own. The sign of these fringes, which is somewhat obscured because they are represented by an arbitrary colour table, indicates that the deformation brought the terrain closer to the radar during the second pass.

An anticipated limitation of interferometric geodesy that was expected to be serious before actual experiments took place, is that only the line-of-sight displacement is measured. Therefore, deformations that are basically 3-D are projected on 1-D. This limitation was not such a nuisance in practice because, for most events, geophysicists can recognise the nature of the displacement and propose a likely model. The radar data is then used to find the few free parameters of the model. The model starts from hypotheses of what the fault rupture mechanism at depth might be. This is well illustrated by the specific study of Fawnskin earthquake [Feigl95]. Using an approach which has become routine since then, ten parameters are required to characterise a rupture on a single planar patch:

- the position of the lower corner of the patch (3 coordinates)
- the two angles for the orientation of the plane (2 coordinates)
- the size of the ruptured rectangle (2)
- the vector indicating the amount of slip in direction and amplitude (3)

In the case of this study, Kurt Feigl and co-workers found that the patch was 2.9 by 3.1 km with an average 2.6 km depth. To infer this information from the radar data, which deals only with the surface deformation, geophysicists consider the Earth as being made of an elastic material like rubber. The rupture at depth, or 'focal mechanism' as it is called, is then equivalent to a cut in the rubber, followed by a relative displacement of the two lips of the cut. Mechanical equations are then used to convert the at-depth displacement into a 3-D surface displacement, using an assumption about the elastic modulus of the crust material. This displacement is itself converted to a line-of-sight displacement and scaled as fringes. The process lasts until the agreement between the model and the result is satisfactory. The best fit obtained by Feigl *et al.* is represented on the right of Figure 3-2.

Amusingly, when these elastic models were refined during the eighties, some approximations were made. The general feeling was that elastic modelling might have some flaws but that no geodetic method would ever provide measurements with sufficient density to reveal these flaws. This opinion was doubly pessimistic, as radar interferometry provided the required 100 or more measurements per square kilometre, and... proved that this modelling is basically sound and flawless!

3.4 The quiet but complicated deformation after an earthquake

The excitement over Landers as an ideal test site did not fade after the initial studies. Crucial questions were still unanswered. Is such a large earthquake preceded by geodetic precursors? What precisely happens to the ground in the months or years following the earthquake? For these studies a wealth of data became available as time went by after the earthquake and as ERS-1 and then ERS-2 continued to gather compatible data over the site. Unfortunately the amount of radar data before the event remained small, and the first of these questions, perhaps the most important in terms of hazard mitigation, remains unanswered.

During the course of these experiments, significant new facts concerning the interferometric technique were uncovered. In particular, the importance of atmospheric artefacts was suspected [Massonnet94] and fully characterised [Feigl95] mainly on the Landers site. Two studies [Massonnet96 and Peltzer96] aimed to model the post-seismic displacement that continued to take place on the site after the earthquake. Different mechanisms of post-seismic fault slip were proposed. Radar interferometry is all the more important in this activity since most of these displacements were aseismic, and therefore went unnoticed by conventional seismological records.

The scale on the post-seismic interferogram, Figure 3-3, speaks for itself. It is clear that, to catch the smallest structures of displacement with ground instrumentation such as GPS receivers, it would be necessary to literally cover the ground with instruments. Such a density is not realistic. On adequate surfaces, such as those that can be found in Iceland, the western United States, Chile and many more places, having a radar archive suffices for study of any upcoming event in the most remote regions. In this regard,

an archive such as the global coverage of the land surface gathered by ERS-1 and ERS-2 is a genuine 'memory of the Earth' that can be compared with new acquisitions taking place years later, possibly by another satellite. In a sense, radar interferometry can turn every pebble into a GPS receiver.
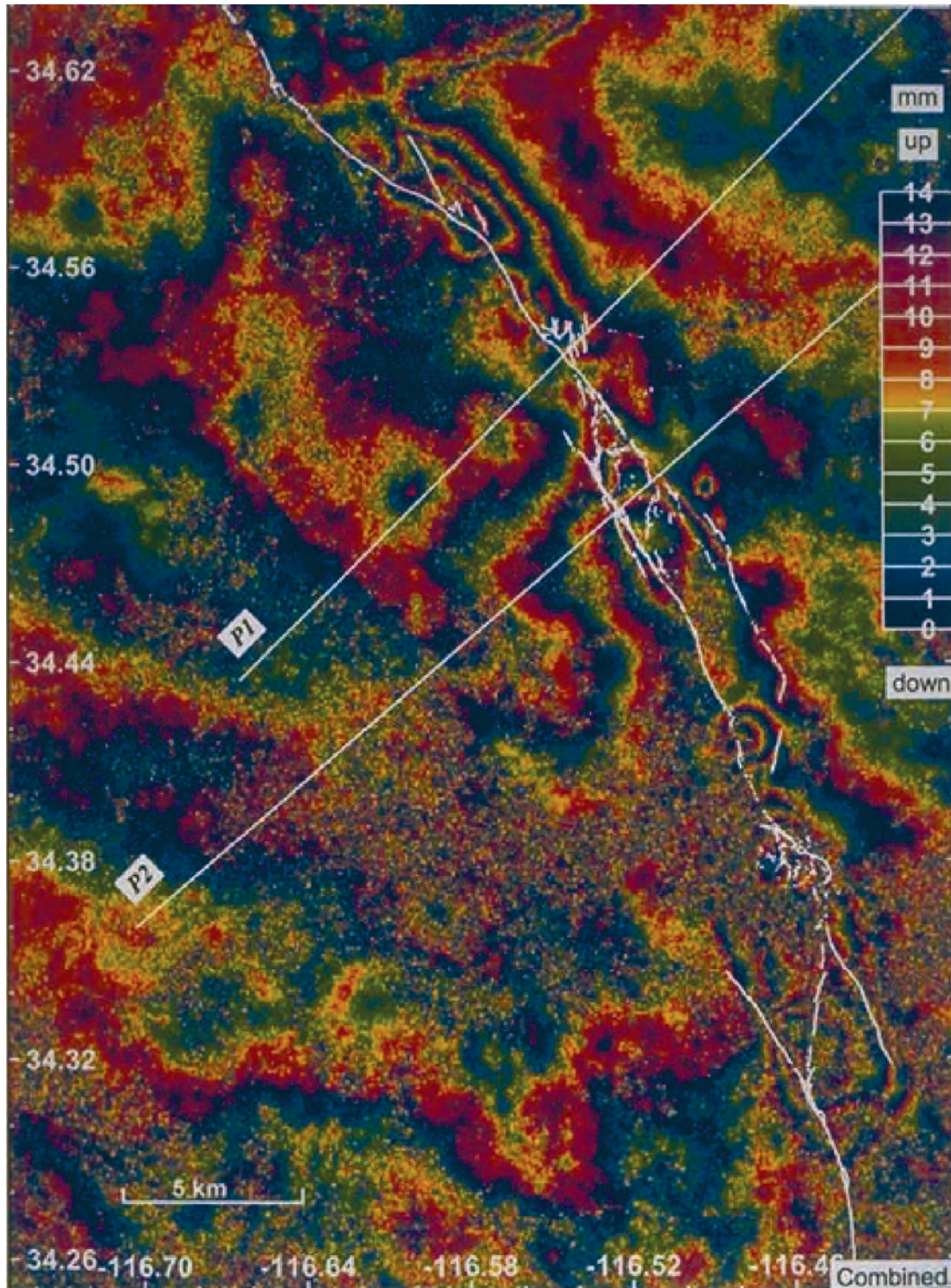

Figure 3-3: Illustration of post-seismic displacement

3.5 A case of coherence loss: India

One of the drawbacks of the academic way of communicating is that failures are never published. Here we describe one of these failures.

After the disastrous earthquake of Latur, in India, ESA made some ERS-1 images available. Unfortunately, the images show a mostly incoherent result.
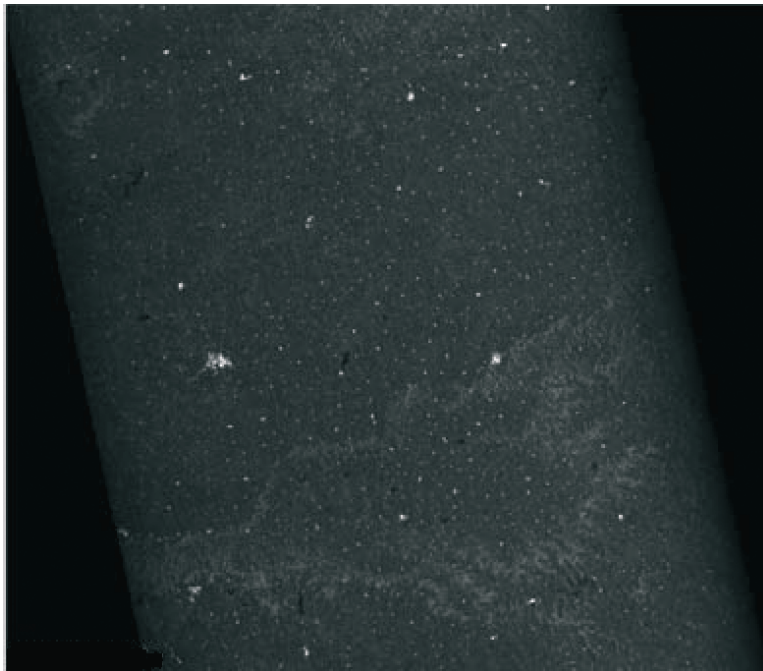

Figure 3-4a shows the typical landscape: a mostly agricultural landscape with many small cities and villages scattered from place to place. Some mild topography is detectable in the south of this amplitude image.

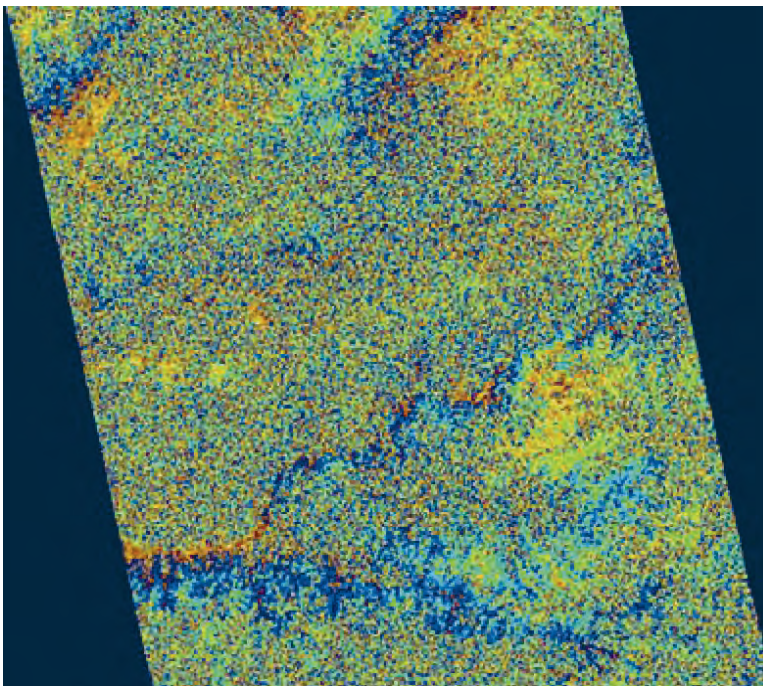Figure 3-4b shows a co-seismic interferogram created from ERS-1 orbits 8409 and 11916, i.e. between 23 February 1993 and 26 October 1993. The quality of the interferogram, which includes a negligible topographic component because of a very small orbital separation, does not allow a clear recognition of any ground displacement. The task is further aggravated because no reliable modelling of the expected displacement exists.

Figure 3-4c shows another interferogram created from orbits 8409 and 5403, i.e. between 23 February 1993 and 28 July 1992. The quality of the interferogram is similar to the previous one, and includes a topographic signature especially visible in the south of the image.
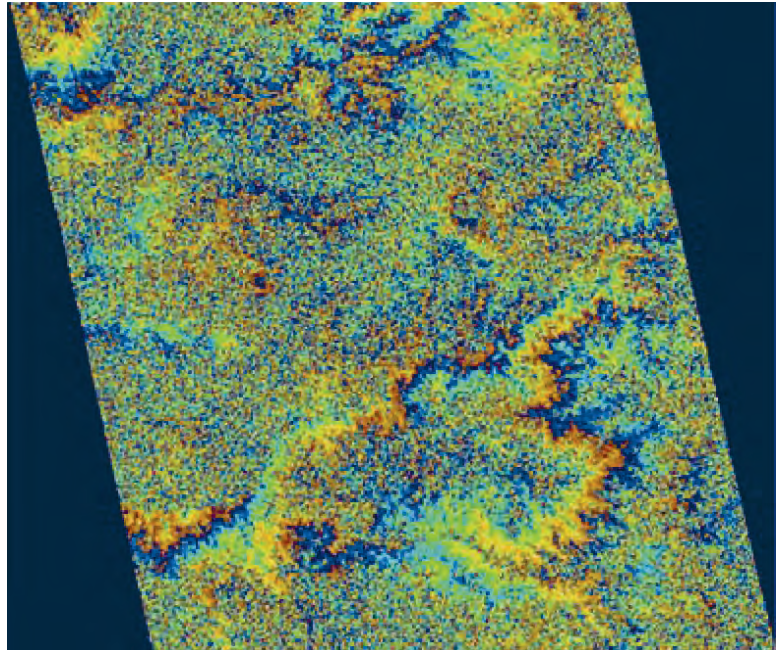
Figure 3-4: Failed example due to coherence loss: the Latur earthquake in India. The dramatic event of Latur, on 30 September 1993, was an opportunity to expand the demonstration of interferometry for the mapping of co-seismic deformation. Although some previous images from the site existed, and despite a systematic acquisition by ESA after the event, difficulties arose because it was monsoon time, which drastically changes the surface conditions. In addition, the region is used for agriculture.

The area is used mainly for agriculture and located in a monsoon region with heavy rainfalls. Surprisingly, the only areas that retained coherence are the cities and villages, despite the large-scale destruction they experienced. This is not a unique case; often cities heavily damaged by earthquake remain quite coherent. This might be an indication that the main contributors in the radar signal of a city are not much damaged by earthquakes. In any case, one should not restrict interferometric studies of damaged cities for this reason.

3.6 A case of damaged raw data, studying a large earthquake in Chile

Very large events emphasise the usefulness of the wide-area surveying that is possible with radar interferometry. Sometimes, however, the deformation field is so wide that even the wide 100 km swath of ERS cannot catch it entirely. A large earthquake took place in Chile in July 1995, and was studied, among others, by the IPGP (Figure 3-5). This example illustrates an unusual error in data management.
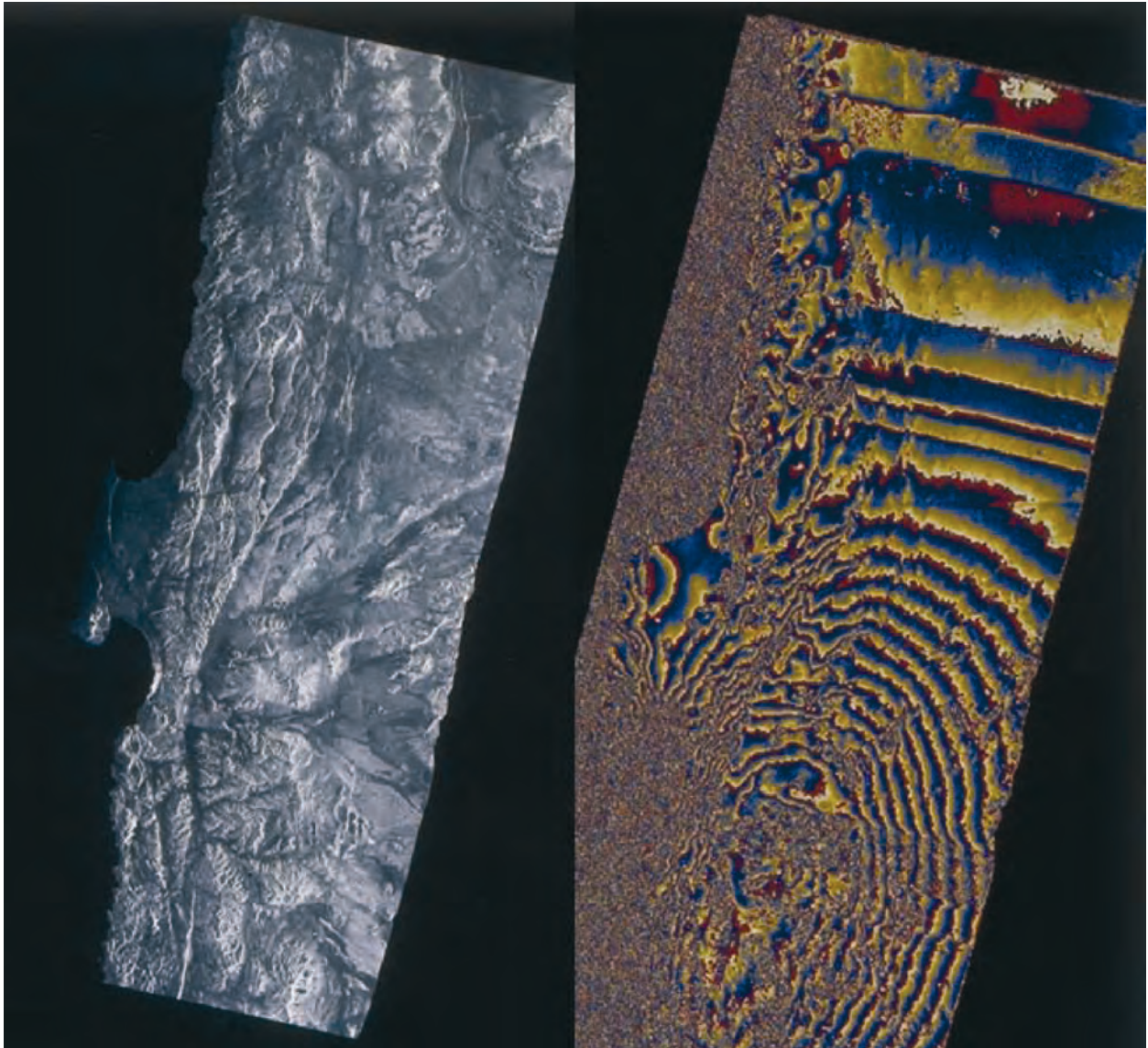
Figure 3-5: Large deformation field with errors: Chile earthquake

The large earthquake that struck the Atacama region in Chile on 30 July 1995 was an opportunity for interferometry very similar to the Landers example: a large earthquake in a mostly desert environment. We see in Figure 3-5 the typical behaviour of an interferometric signal (left is amplitude, right is co-seismic phase). The area covered by ocean, around the Mejillones peninsula, is not coherent. The rest of the landscape is exceptionally coherent (except for the interruption of fringe continuity which is explained below). This example illustrates how very big events can actually outspan the size of ERS images, in spite of their being the widest radar images available in standard mode (i.e. not SCANSAR). Another striking aspect of this example is the very smooth deformation pattern, without any visible surface rupture (again discounting the processing artefacts mentioned above). The data from this site that were processed,

analysed in cooperation with the IPG of Paris, were generally made of strips of four or five ERS images in length.

The interruption in the fringe continuity is due to missing lines in the raw data. The interferometric technique relies heavily on the strong self-consistency of the geometry of radar images. Missing lines can break this consistency. There are ways, however, to detect missing lines in raw data. Denoting the complex samples of a data take as $A(i,j)$, where i represents the range pixels (from 1 to N) and j represents the pulse lines, we can form the complex number:

$$\rho(j) = \frac{\sum_{i=1}^{i=N} A(i,j)A^*(i,j+1)}{\sqrt{\sum_{i=1}^{i=N} A^2(i,j)} \sqrt{\sum_{i=1}^{i=N} A^2(i,j+1)}} \qquad \text{Equation 3.1}$$

This quantity is similar to the one formed on interferograms to compute the coherence. Here the result is a complex number. The phase $\phi(\rho(j))$ of this complex number is an estimator for the mean Doppler of the scene D, once it has been multiplied by the pulse repetition frequency PRF:

$$D(j) = \frac{f(\rho(j))}{2\pi} * PRF \qquad \text{Equation 3.2}$$

The amplitude of $\rho(j)$ is also very interesting because it gives the correlation between adjacent lines of raw data. For ERS, this amplitude should be about 0.3, and if two lines exhibit a lower value (for instance 0.1 or less), it means that the two lines of raw data are not really adjacent, and so there must be at least one missing line between them!

The test is really easy to implement and can be used to check the raw data. Unfortunately, it cannot tell whether there is one or several missing lines between two lines which are found to be not adjacent.