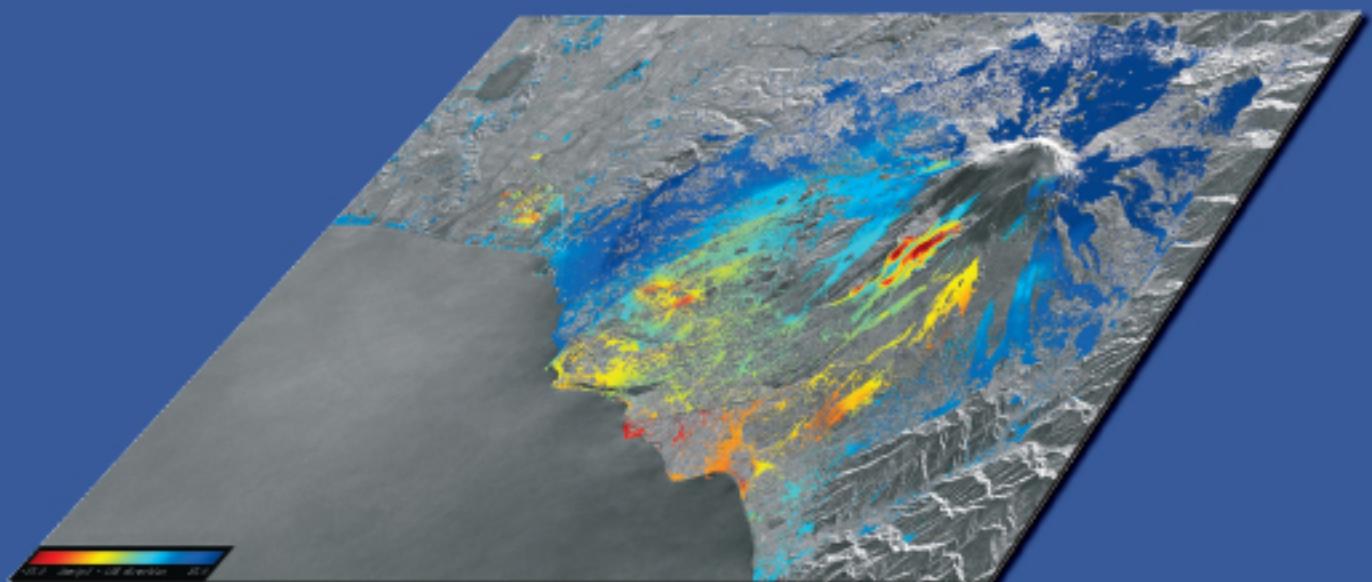


InSAR Principles



**Guidelines for SAR Interferometry Processing
and Interpretation**

InSAR Principles:

Guidelines for SAR Interferometry Processing and Interpretation

Acknowledgements

Authors:

Alessandro Ferretti, Andrea Monti-Guarnieri, Claudio Prati, Fabio Rocca
Dipartimento di Elettronica ed Informazione, Politecnico di Milano, Italy

Didier Massonnet
CNES, Toulouse, France

Technical coordination:

Juerg Lichtenegger
ESA/ESRIN (retired), Frascati, Italy

Publication:

InSAR Principles: Guidelines for SAR
Interferometry Processing and Interpretation
(TM-19, February 2007)

Editor:

Karen Fletcher

Published and distributed by:

ESA Publications
ESTEC
Postbus 299
2200 AG Noordwijk
The Netherlands
Tel: +31 71 565 3400
Fax: +31 71 565 5433

Printed in:

The Netherlands

Price:

€40

ISBN:

92-9092-233-8

ISSN:

1013-7076

Copyright:

© 2007 European Space Agency

Table of Contents

Scope	viii
--------------------	-------------

Part A Interferometric SAR image processing and interpretation

1. Synthetic Aperture Radar basics.....	A-3
1.1 Introduction.....	A-3
1.1.1 Introduction to ERS	A-3
1.1.2 Introduction to Envisat.....	A-4
1.2 SAR images of the Earth's surface	A-5
1.2.1 What is a strip-map SAR imaging system?.....	A-5
1.2.2 What is a complex SAR image?.....	A-6
1.2.3 SAR resolution cell projection on the ground.....	A-11
2. SAR interferometry: applications and limits	A-17
2.1 Introduction.....	A-17
2.2 Terrain altitude measurement through the interferometric phase	A-18
2.2.1 Interferogram flattening	A-19
2.2.2 Altitude of ambiguity	A-20
2.2.3 Phase unwrapping and DEM generation.....	A-20
2.3 Terrain motion measurement: Differential Interferometry	A-23
2.4 The atmospheric contribution to the interferometric phase	A-24
2.5 Other phase noise sources	A-25
2.6 Coherence maps	A-26
3. SAR Differential Interferometry basics and examples	A-31
3.1 Introduction.....	A-31
3.2 Landers co-seismic deformation	A-31
3.3 Small earthquake modelling	A-33
3.4 The quiet but complicated deformation after an earthquake.....	A-35
3.5 A case of coherence loss: India.....	A-37
3.6 A case of damaged raw data, studying a large earthquake in Chile.....	A-38

Part B InSAR processing: a practical approach

1. Selecting ERS images for InSAR processing.....	B-3
1.1 Introduction.....	B-3
1.2 Available information about ERS images.....	B-3
1.2.1 The ESA on-line multi-mission catalogue	B-3
1.2.2 DESCW.....	B-4
1.2.3 Expected coherence (prototype).....	B-6
1.3 Selecting images for InSAR DEM generation.....	B-8
1.4 Selecting images for Differential InSAR applications.....	B-9
2. Interferogram generation	B-11

2.1	Introduction.....	B-11
2.2	Generation of synthetic fringes.....	B-12
2.3	Co-registering	B-13
2.3.1	Co-registering coefficients	B-14
2.3.2	Co-registering parameter estimation	B-16
2.3.3	Implementation of resampling	B-17
2.4	Master and slave oversampling.....	B-17
2.5	Range spectral shift & azimuth common bandwidth filtering ...	B-18
2.5.1	Range spectral shift filtering	B-18
2.5.2	Azimuth common band filtering	B-20
2.6	Interferogram computation	B-22
2.6.1	Complex multi-looking	B-24
2.6.2	Generation of coherence maps.....	B-26
2.7	Applications of coherence	B-27
2.8	Interferogram geocoding & mosaicking	B-29
3.	InSAR DEM reconstruction	B-31
3.1	Introduction.....	B-31
3.2	Processing chain and data selection.....	B-31
3.3	Phase unwrapping techniques for InSAR DEM reconstruction.B-33	
3.3.1	What are we looking for?.....	B-34
3.3.2	Case $p=2$, Unweighted Least Mean Squares method ...B-37	
3.3.3	Case $p=2$, Weighted Least Mean Squares methodB-38	
3.3.4	Case $p=1$, Minimum Cost Flow method.....	B-38
3.3.5	Case $p=0$, Branch-Cut and other minimum L_0 methods	B-39
3.3.6	Outlook	B-41
3.4	From phase to elevation	B-42
3.4.1	Polynomial approximation of satellite orbits, point localisation and data geocoding	B-42
3.4.2	Data resampling	B-45
3.4.3	Impact of baseline errors on the estimated topography	B-45
3.4.4	Precise orbit determination	B-47
3.5	Error sources, multi-baseline strategies and data fusion	B-48
3.5.1	Multi-interferogram InSAR DEM reconstruction.....B-50	
3.6	Combination of ascending and descending passes	B-53
3.7	Conclusions.....	B-55
4.	Differential Interferometry (DInSAR).....	B-57
4.1	Examples of differential interferometry on land.....	B-57
4.1.1	Physical changes	B-57
4.1.2	Volcano: Okmok	B-57
4.1.3	Surface rupture: Superstition Hill	B-58
4.1.4	Subsidence: East Mesa.....	B-60
4.2	Example of differential interferometry on ice	B-62
4.3	Review of various criteria for data selection	B-63

4.4	Interferometric interpretation.....	B-63
4.4.1	Interferometry phase signal ruggedness.....	B-64
4.4.2	Fictitious example interferograms for analysis	B-65
4.4.3	Analysis of fictitious situations.....	B-67

Part C InSAR processing: a mathematical approach

1.	Statistics of SAR and InSAR images.....	C-3
1.1	The backscattering process	C-3
1.1.1	Introduction.....	C-3
1.1.2	Artificial backscatterers	C-3
1.1.3	Natural backscatterers: the spectral shift principle	C-4
1.1.4	Statistics of the return	C-7
1.2	Interferometric images: coherence.....	C-8
1.2.1	Statistics of coherence estimators	C-9
1.2.2	Impact of the baseline on coherence	C-12
1.3	Power spectrum of interferometric images	C-13
1.4	Causes of coherence loss	C-13
1.4.1	Noise, temporal change.....	C-13
1.4.2	Volumetric effects.....	C-13
2.	Focusing, interferometry and slope estimate	C-15
2.1	SAR model: acquisition and focusing.....	C-15
2.1.1	Phase preserving focusing	C-15
2.1.2	CEOS offset processing test.....	C-18
2.2	Interferometric SAR processing	C-18
2.2.1	Spectral shift and common band filtering (revisited)...	C-19
2.3	DEM generation: optimal slope estimate.....	C-21
2.4	Noise sources	C-24
2.5	Processing decorrelation artefacts.....	C-25
2.5.1	Examples of decorrelation sources.....	C-25
3.	Advances in phase unwrapping	C-29
3.1	Introduction.....	C-29
3.2	Residues and charges	C-31
3.2.1	Effects of noise: pairs of residues, undefined positions of the ‘ghost lines’	C-33
3.2.2	Effects of alias: unknown position of the ghost lines...	C-36
3.3	Optimal topographies under the L_p norm.....	C-37
3.3.1	L_2, L_1, L_0 optimal topographies	C-37
3.3.2	Slope estimates.....	C-40
3.3.3	Removal of low resolution estimates of the topography	C-41
3.3.4	Bias of the slope estimate.....	C-41
3.4	Analysis in the wave-number domain.....	C-42
3.4.1	L_2 optimisation in the wave-number domain	C-42
3.5	Weighting factors in the optimisation.....	C-43

4.	Multiple image combination for DEM generation and ground motion estimation	C-45
4.1	Multi-baseline phase unwrapping for InSAR topography estimation.....	C-45
4.2	Applications to repeat-pass interferometry.....	C-48
4.2.1	Example 1: the Vesuvius data set	C-50
4.2.2	Example 2: The Etna data set.....	C-53
4.3	The 'Permanent Scatterers' technique	C-56
4.3.1	Space-time estimation.....	C-58
4.3.2	Subsidence in Pomona	C-59
4.3.3	Ground slip along the Hayward fault.....	C-62
4.3.4	Seasonal deformation in the Santa Clara Valley.....	C-63
5.	Applications based on spectral shift	C-65
5.1	Introduction to spectral shift.....	C-65
5.2	Interferometric quick look (IQL).....	C-67
5.3	Super-resolution.....	C-69
6.	Differential interferometry	C-71
6.1	Introduction.....	C-71
6.2	Differential interferometry using an available DEM	C-72
6.3	Differential interferometry with three or more combined images	C-77
6.4	Techniques to avoid phase unwrapping	C-79
6.4.1	Integer combination	C-79
6.4.2	Interferogram stacking	C-82
6.5	Information contained in interferometric measurements	C-83
6.5.1	Residual orbital fringes	C-83
6.5.2	Uncorrected topography.....	C-86
6.5.3	Heterogeneous troposphere.....	C-86
6.5.4	Heterogeneous ionosphere	C-87
6.5.5	Static atmosphere	C-88
6.5.6	Radar clock drift	C-88
7.	Envisat-ASAR interferometric techniques and applications	C-91
7.1	Introduction.....	C-91
7.2	ScanSAR: an introduction	C-92
7.2.1	Acquisition.....	C-93
7.2.2	Focusing	C-94
7.3	ScanSAR interferometry.....	C-96
7.3.1	Common band (CB) filtering	C-97
7.4	Multi-mode SAR interferometry.....	C-98
7.4.1	Multi-mode interferometric combination.....	C-98
7.5	Applications	C-101
7.5.1	AP/AP/IM interferometry	C-101
7.5.2	WSM/WSM and WSM/IM interferometry	C-102
8.	ERS-Envisat interferometry	C-107
8.1	Introduction.....	C-107

8.2 ERS-Envisat interferometric combination	C-107
8.3 Frequency gap compensation.....	C-108
8.4 Vertical accuracy	C-108
8.5 Altitude of ambiguity	C-109
8.6 Effect of volume scattering	C-110
8.7 Experimental results.....	C-110
References	I

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.

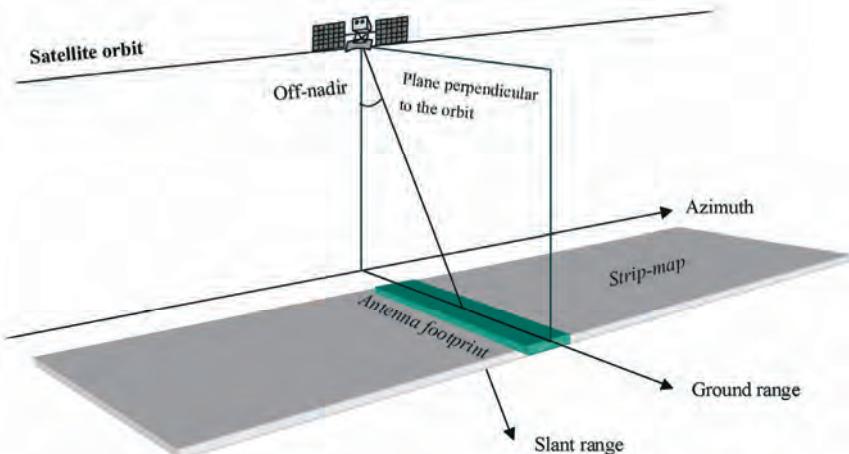


Figure 1-3: A SAR system from a satellite

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.

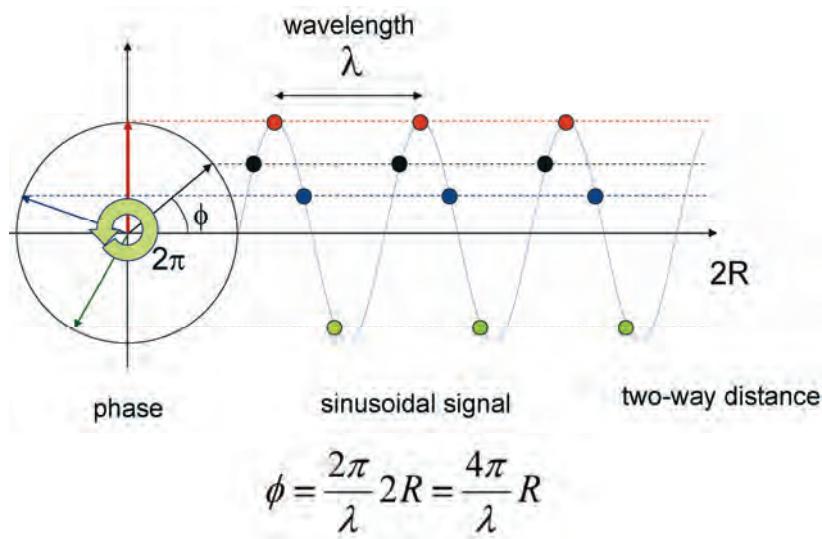


Figure 1-5: A sinusoidal function $\sin \phi$ is periodic with a 2π radian period. In the case of a relative narrow-band SAR (i.e. ERS and Envisat), the transmitted signal can be assimilated, as a first approximation, to a pure sinusoid whose angle or phase ϕ has the following linear dependence on the 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.

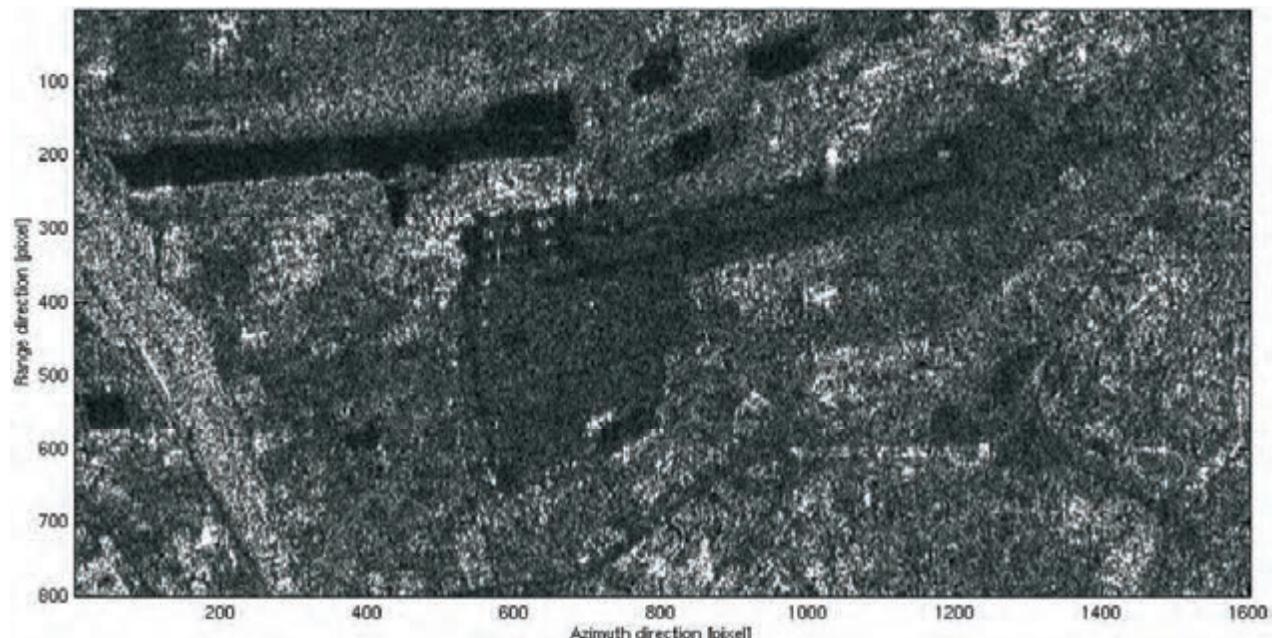


Figure 1-6: ERS-2 SAR detected image of the Linate Airport in the eastern part of Milan (Italy): the speckle effect on the homogeneous fields surrounding the airport is clearly visible.



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.

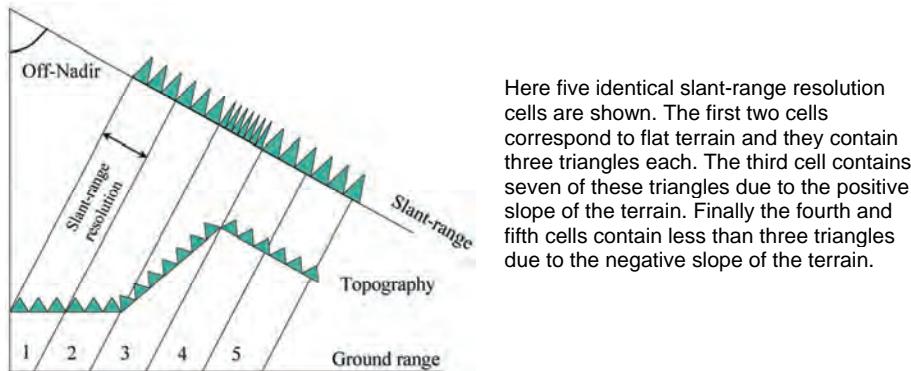


Figure 1-9: Effect of terrain on the SAR image. For SAR resolution cells in the plane perpendicular to the orbit, the part of the terrain imaged in each resolution cell clearly depends on the topography.

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 **loayover**, and is sketched in Figure 1-10.

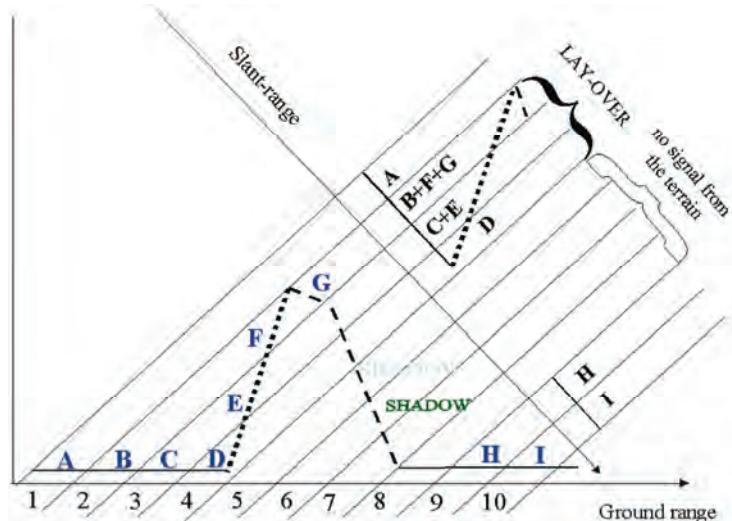


Figure 1-10: Layover and shadow effects. Depending on the terrain slope, scatterers that are located at increasing ground range positions can be imaged in reverse order by the SAR system (points D, E, F and G). Moreover they are imaged in the same SAR resolution cells as scatterers B and C, which belong 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.

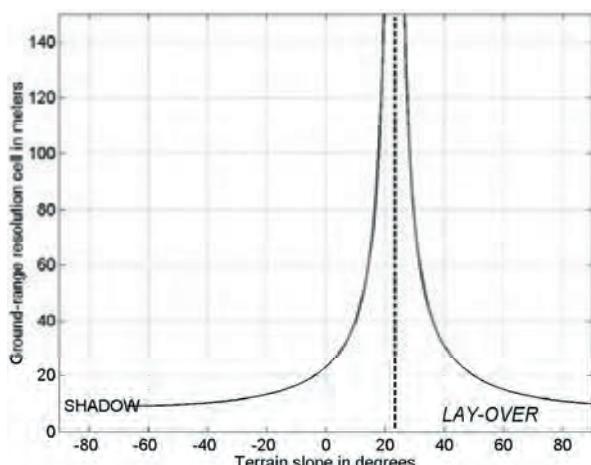


Figure 1-11: ERS resolution cell dimension in ground range as a function of the terrain slope. The vertical dotted line indicates the incidence angle relative to a flat horizontal terrain (23°).

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.

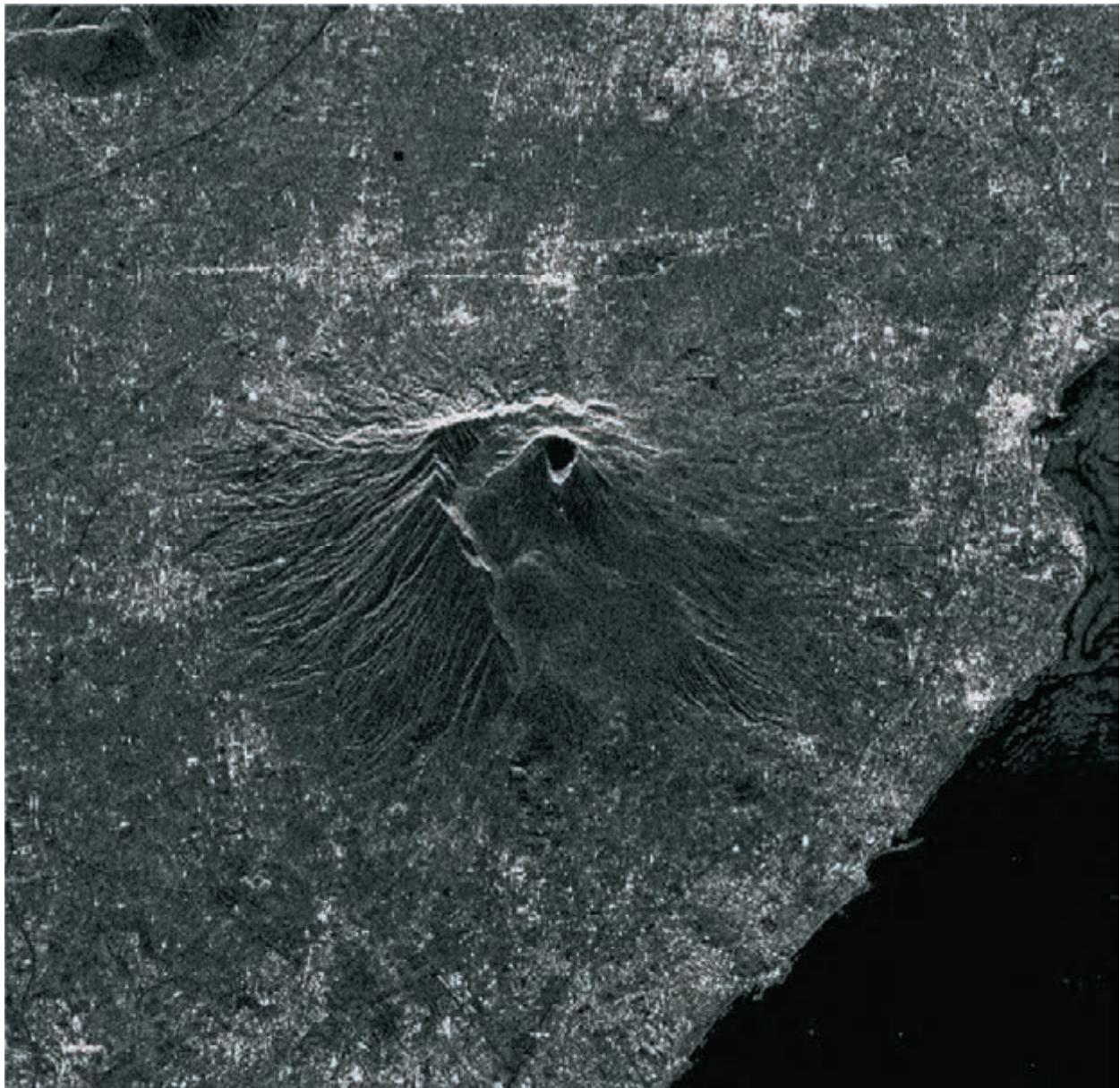


Figure 1-12: ERS-2 SAR image of Mount Vesuvius (Italy), as detected. The slant range direction is vertical on the image (near range is in the upper part of the image). Brighter levels correspond to stronger backscattered radiation. The coastline along the Tyrrhenian Sea is clearly visible (the sea is 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.

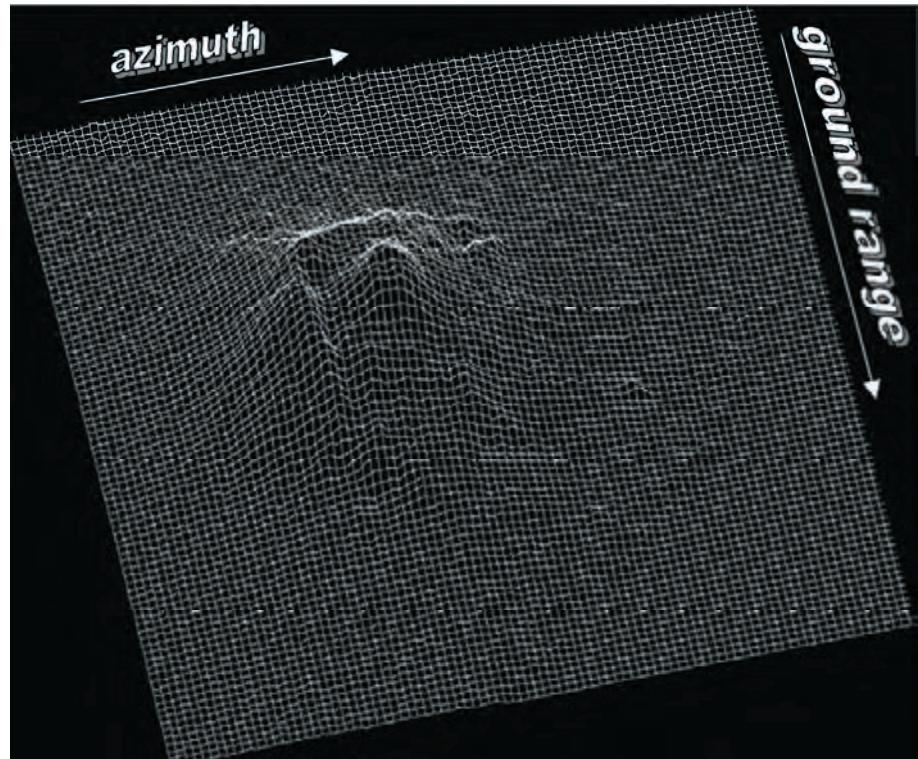


Figure 1-13: Deformation of the regular resolution grid on the ground when projected in SAR coordinates. The deformation is due to topography.

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.

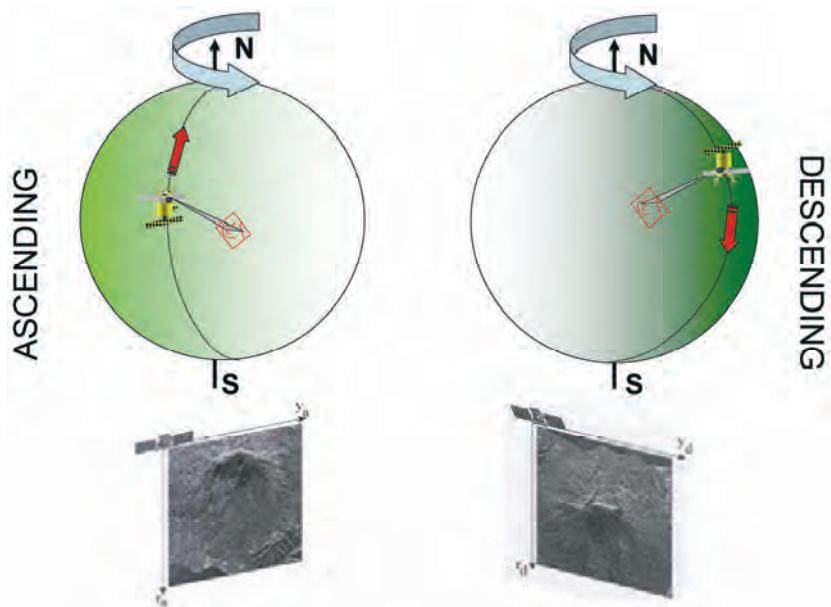


Figure 1-14: Ascending and descending passes

The ground range deformation of SAR images from ascending and descending passes are thus almost complementary. An example is shown in Figure 1-15.

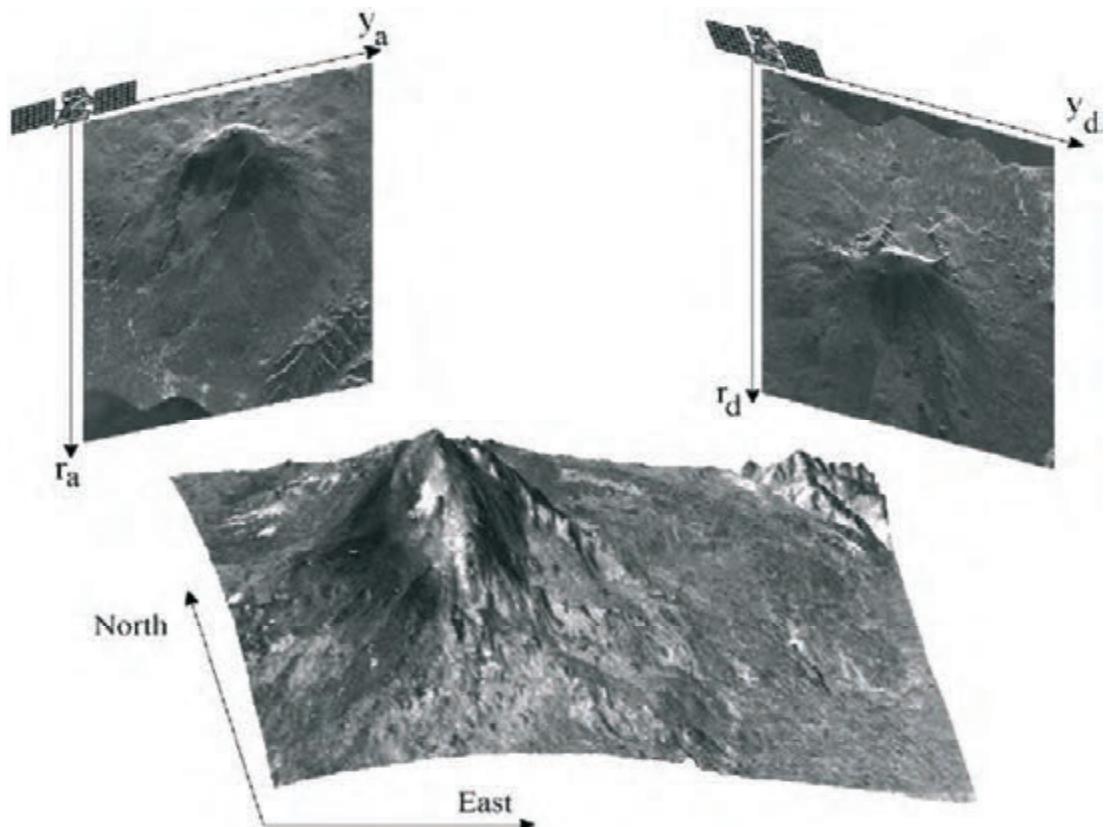


Figure 1-15: Ascending and descending ERS SAR images of Mount Etna (Italy), plus an elevation model

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**.

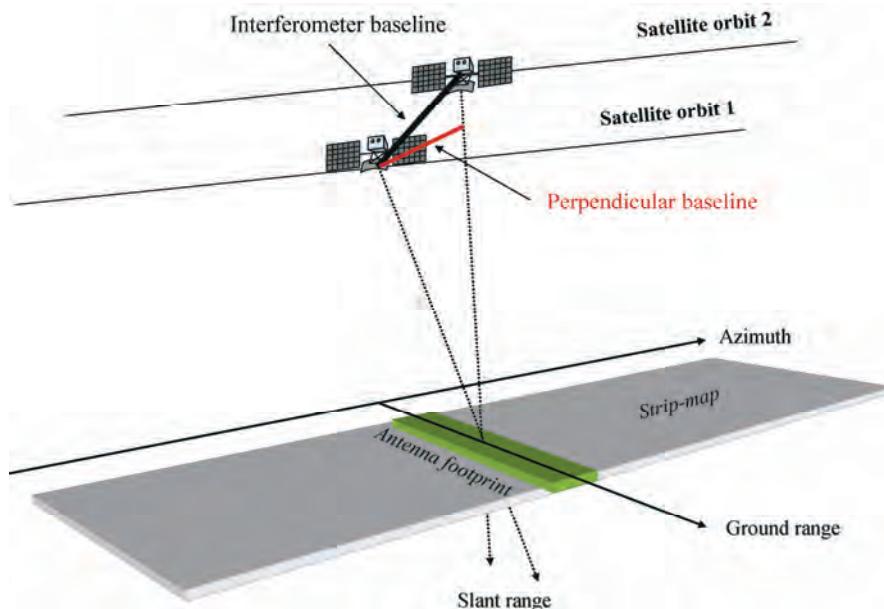


Figure 2-1: Geometry of a satellite interferometric SAR system. The orbit separation is called the interferometer baseline, and its projection perpendicular to the slant range direction is one of the key parameters of SAR interferometry.

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.

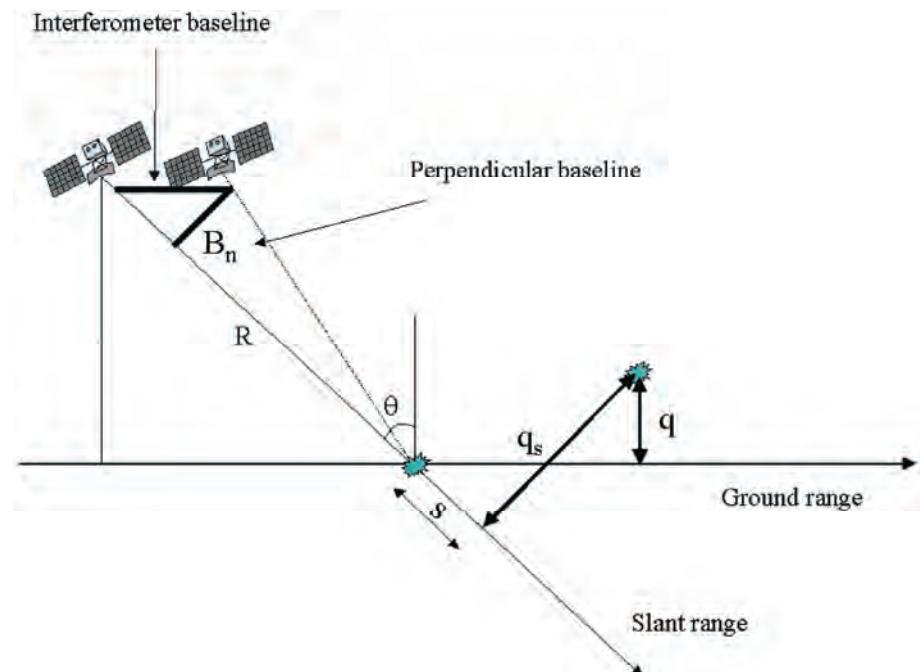


Figure 2-2: Geometric parameters of a satellite interferometric SAR system

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:

$$\Delta r = -2 \frac{B_n q_s}{R} \quad \text{Equation 2.1}$$

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

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

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

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

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).

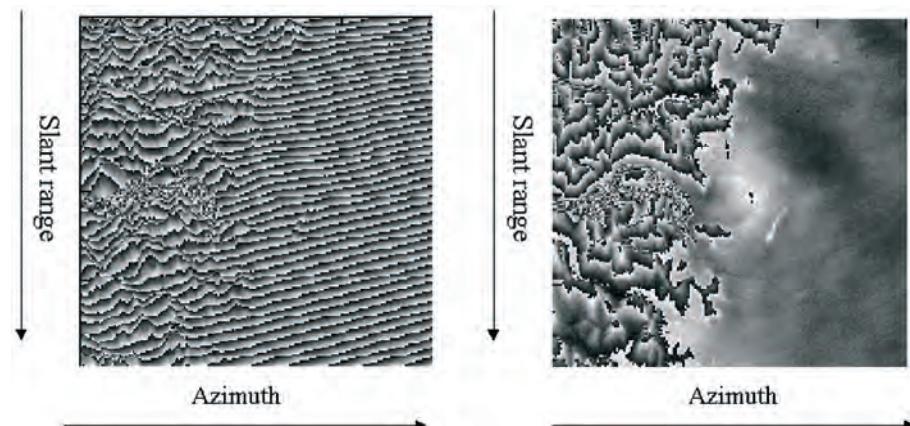


Figure 2-3: Left: interferogram of a portion of the Italian Alps and the Pianura Padana that has been obtained from ERS data. Right: flattened interferogram. Here the phase discontinuities resemble the contour lines.

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:

$$h_a = \frac{\lambda R \sin \theta}{2B_n} \quad \text{Equation 2.4}$$

In the ERS case with $\lambda = 5.6$ cm, $\theta = 23^\circ$, and $R = 850$ km, the following expression holds (in metres):

$$h_a \approx \frac{9300}{B_n} \quad \text{Equation 2.5}$$

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.

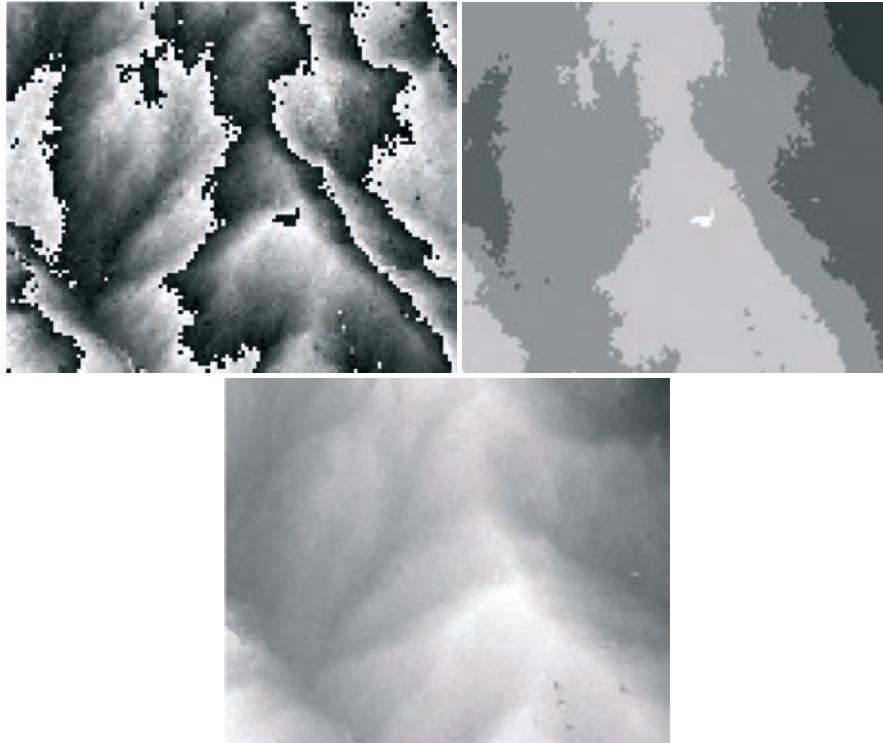


Figure 2-4: Left: SAR interferometric phase generated by means of two ERS images. The 2π phase discontinuities are clearly visible as black/white transitions.

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.

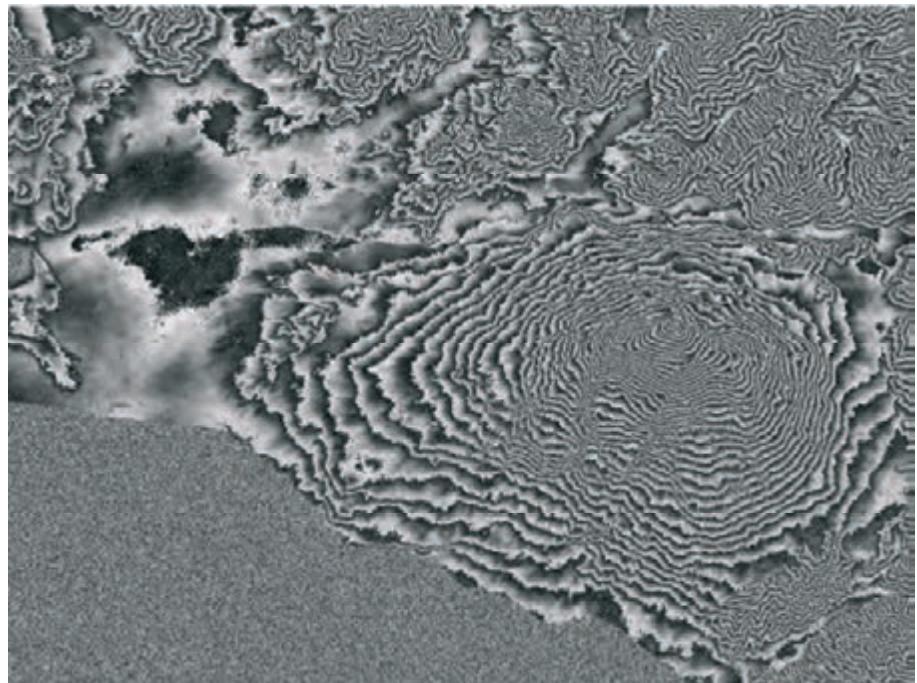


Figure 2-5: Flattened interferogram of Mount Etna generated from ERS tandem pairs. The perpendicular baseline of 115 metres generates an altitude of ambiguity of about 82 metres.

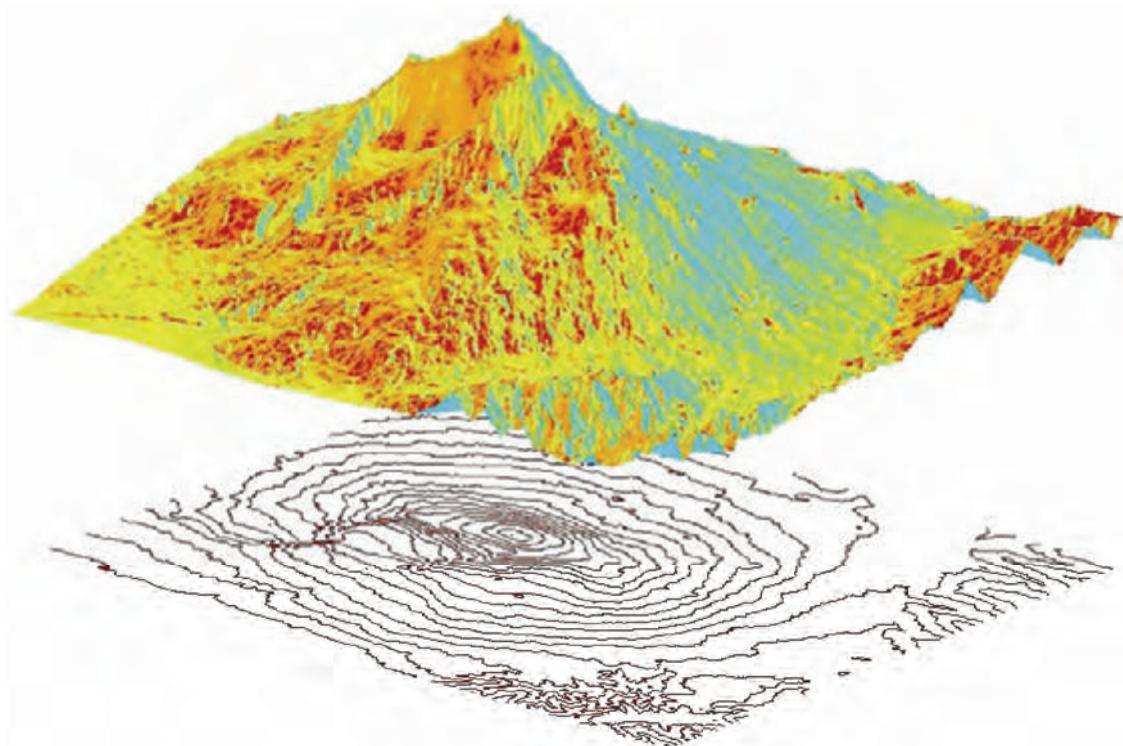


Figure 2-6: Perspective view of Mount Etna as seen from the Northeast. The DEM of Mount Etna has been generated by unwrapping and re-sampling the flattened interferogram of Figure 2-5: The estimated vertical accuracy is better than 10 metres. Contour lines are shown below the DEM.



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:

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

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:

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

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:

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

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.

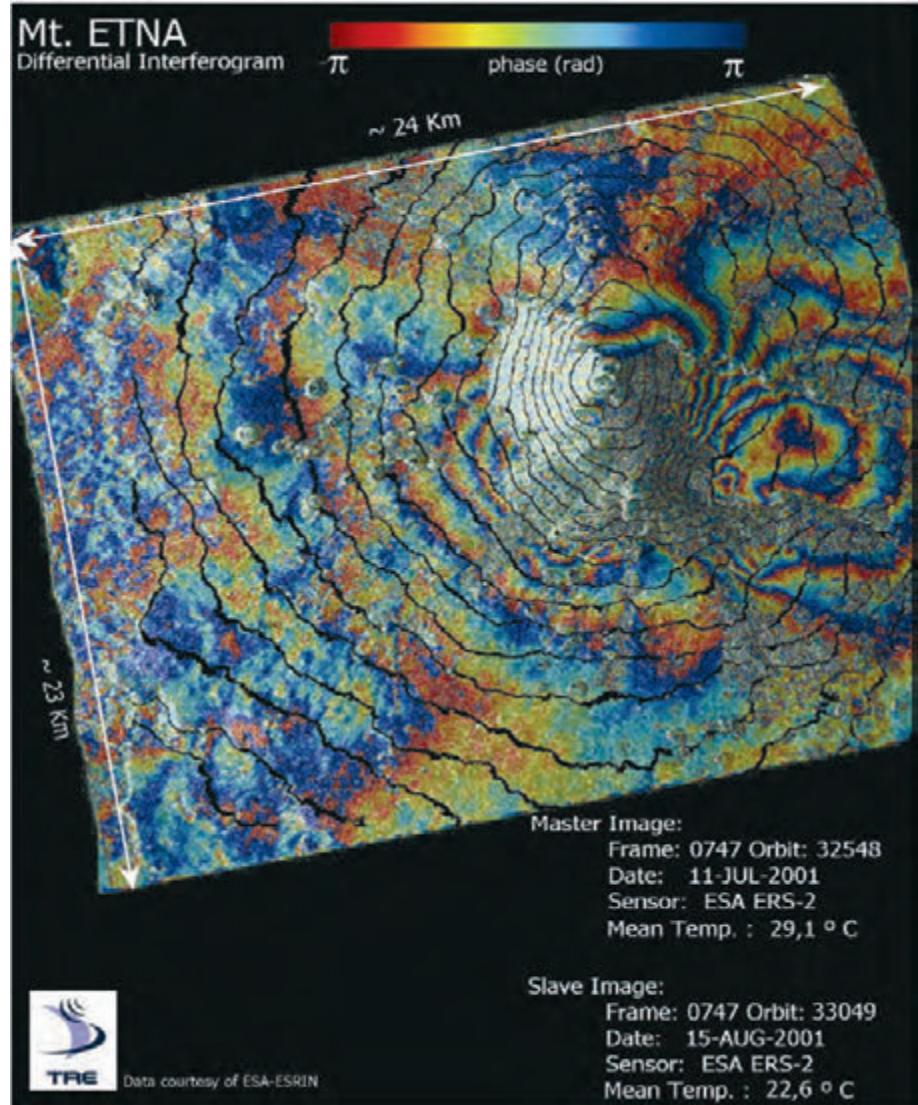


Figure 2-8: The differential interferogram of the Mt Etna eruption that occurred in July 2001. The interferogram has been generated by means of two ERS images taken before (11 July 2001) and after (15 August 2001) the eruption. The topography has been removed by means of an available DEM. Contour lines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

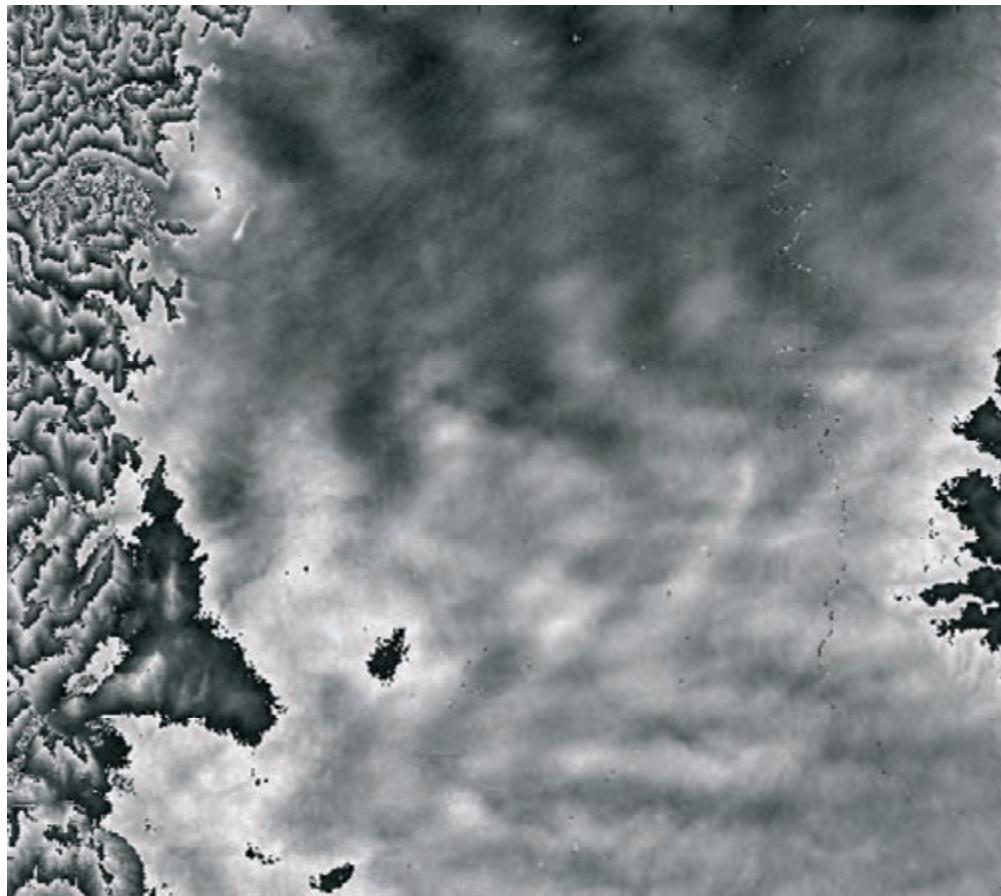


Figure 2-9: An example of atmospheric phase contribution to the ERS interferogram generated on the Pianura Padana. The perpendicular baseline is about 30 metres and the altitude of ambiguity (from black to white in the grey scale used) is about 300 metres.

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]:

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

From a mathematical point of view, this formula is a good approximation of the exact phase dispersion shown in Figure 2-11 when $\sigma_\phi < 12^\circ$. 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 $NL > 4$.

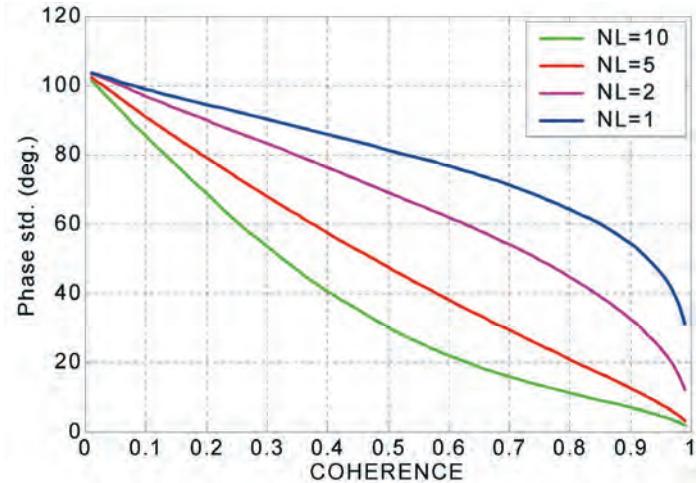


Figure 2-11: Interferometric phase dispersion (degrees) as a function of the coherence for varying numbers of looks (NL)

A comparison between the exact and approximated curves is shown in Figure 2-12.

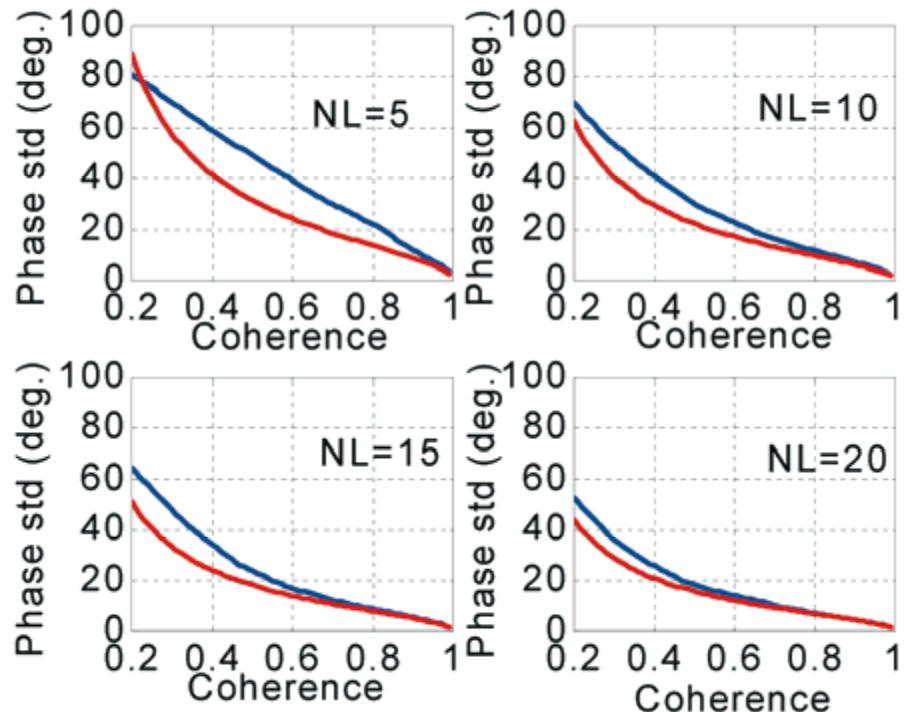


Figure 2-12: Interferometric phase dispersion exact values (blue curves) and approximated ones (red curves)

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:

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

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 whole 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.

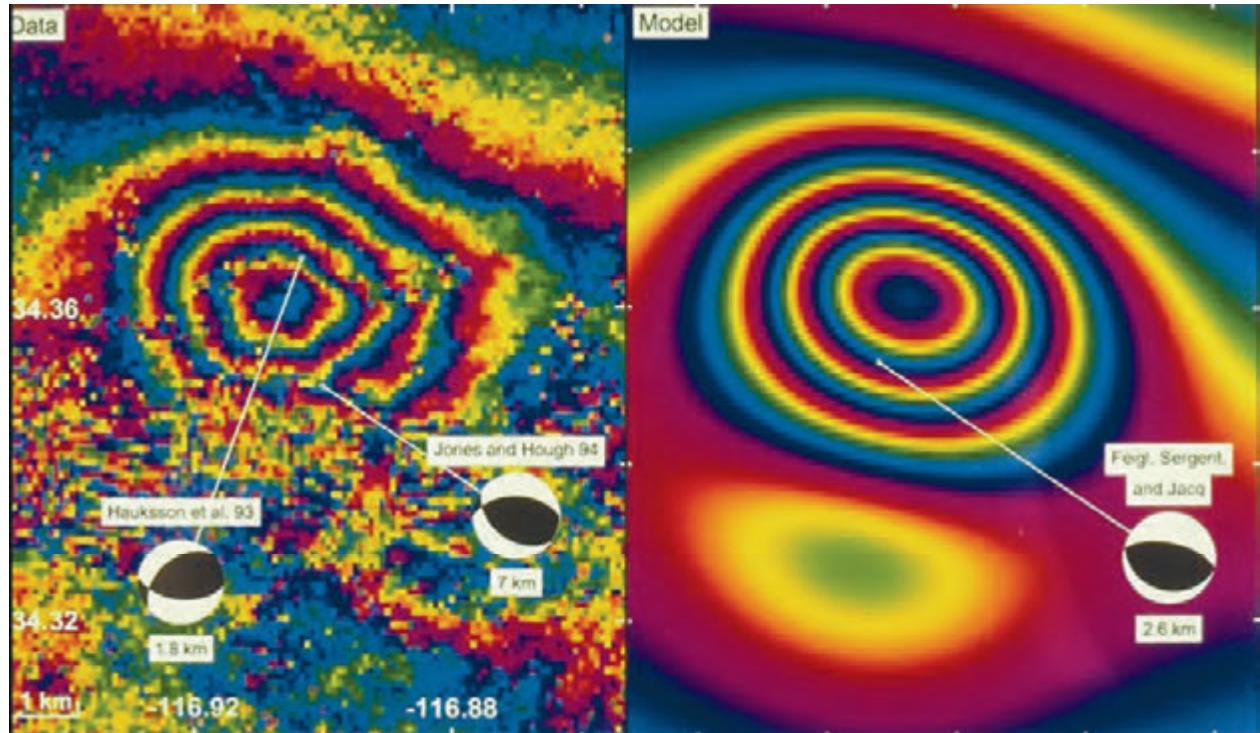


Figure 3-2: FawnSkin earthquake and its modelling

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.

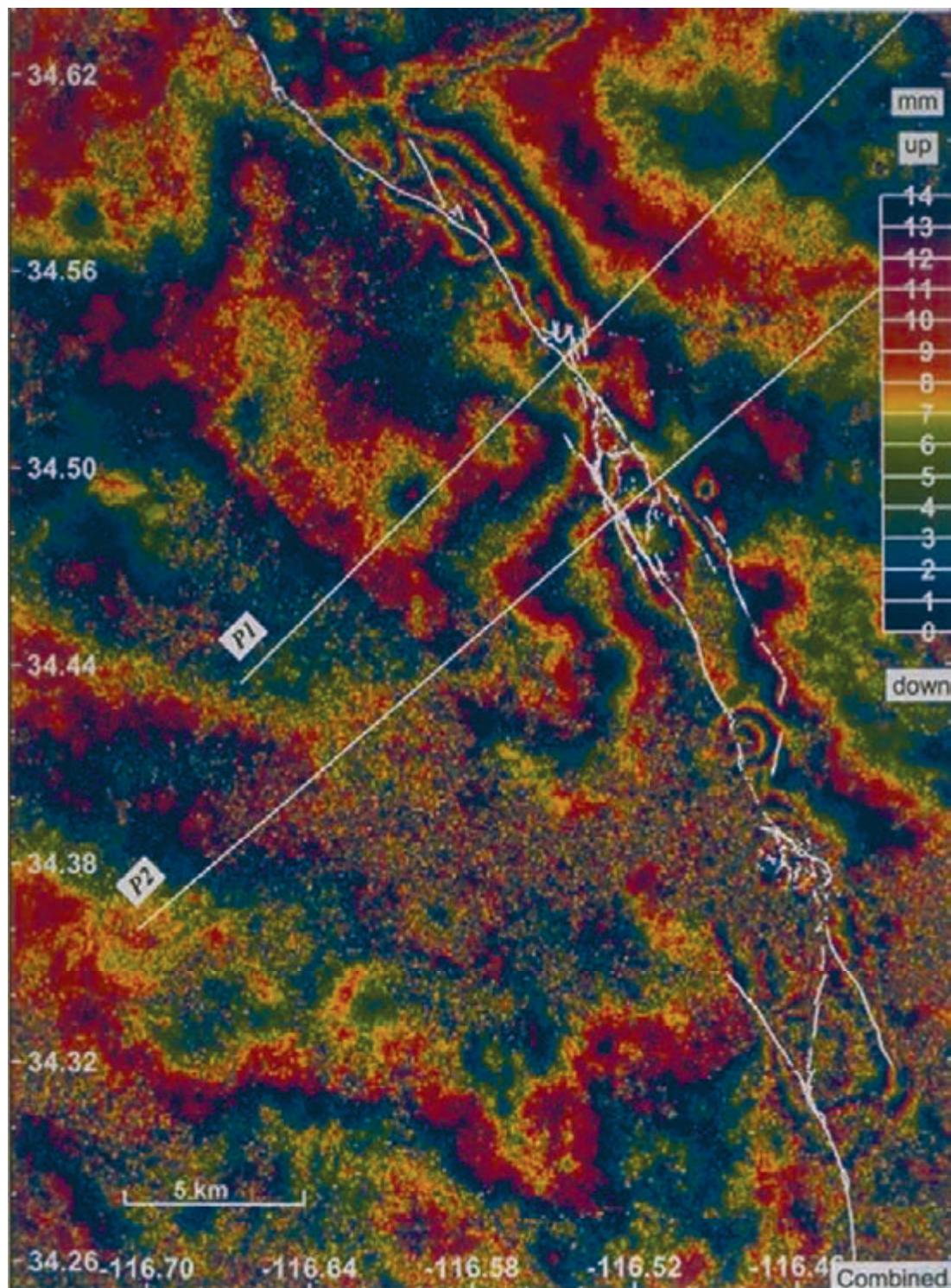


Figure 3-3: Illustration of post-seismic displacement

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.

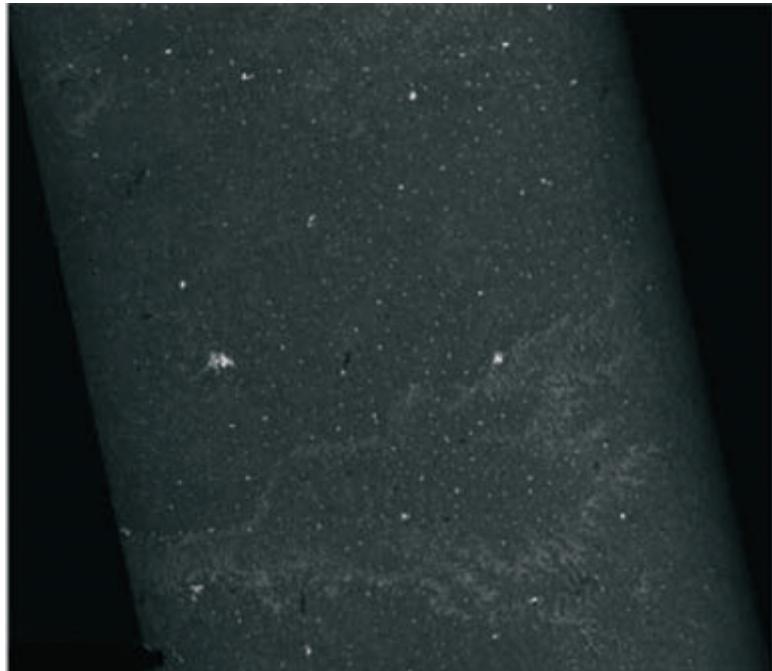


Figure 3-4a shows the typical landscape: a mostly agricultural landscape with many small cities and villages scattered from place to place. Some mild topography is detectable in the south of this amplitude image.

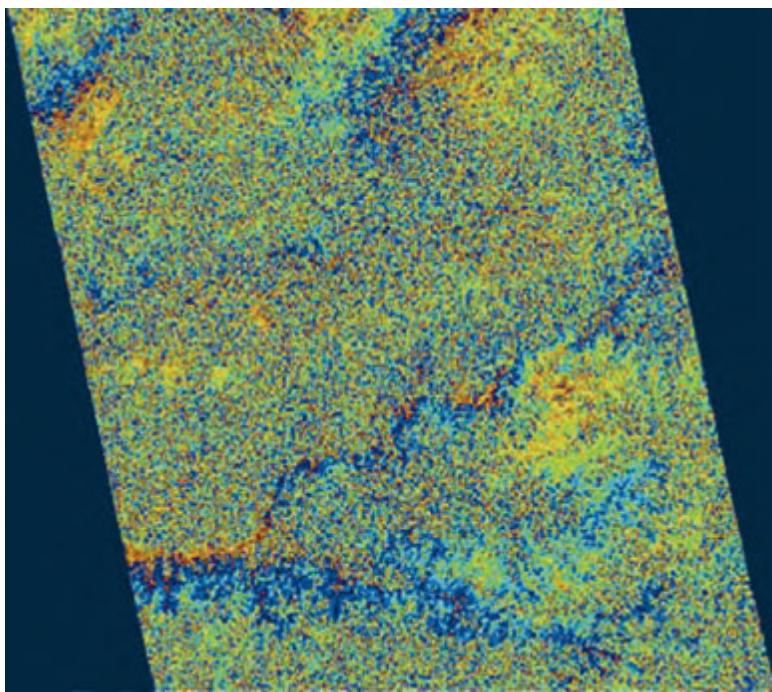


Figure 3-4b shows a co-seismic interferogram created from ERS-1 orbits 8409 and 11916, i.e. between 23 February 1993 and 26 October 1993. The quality of the interferogram, which includes a negligible topographic component because of a very small orbital separation, does not allow a clear recognition 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.

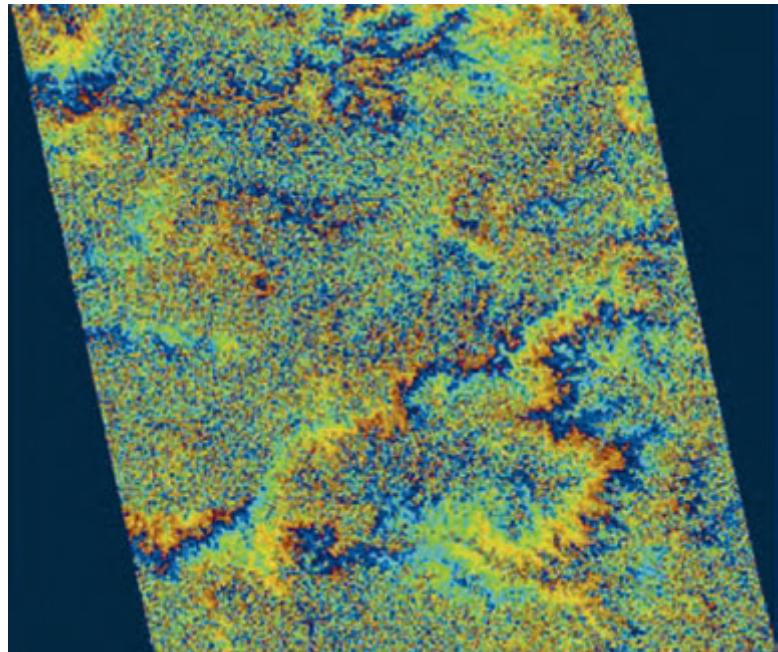


Figure 3-4: Failed example due to coherence loss: the Latur earthquake in India. The dramatic event of Latur, on 30 September 1993, was an opportunity to expand the demonstration of interferometry for the mapping of co-seismic deformation. Although some previous images from the site existed, and despite 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.

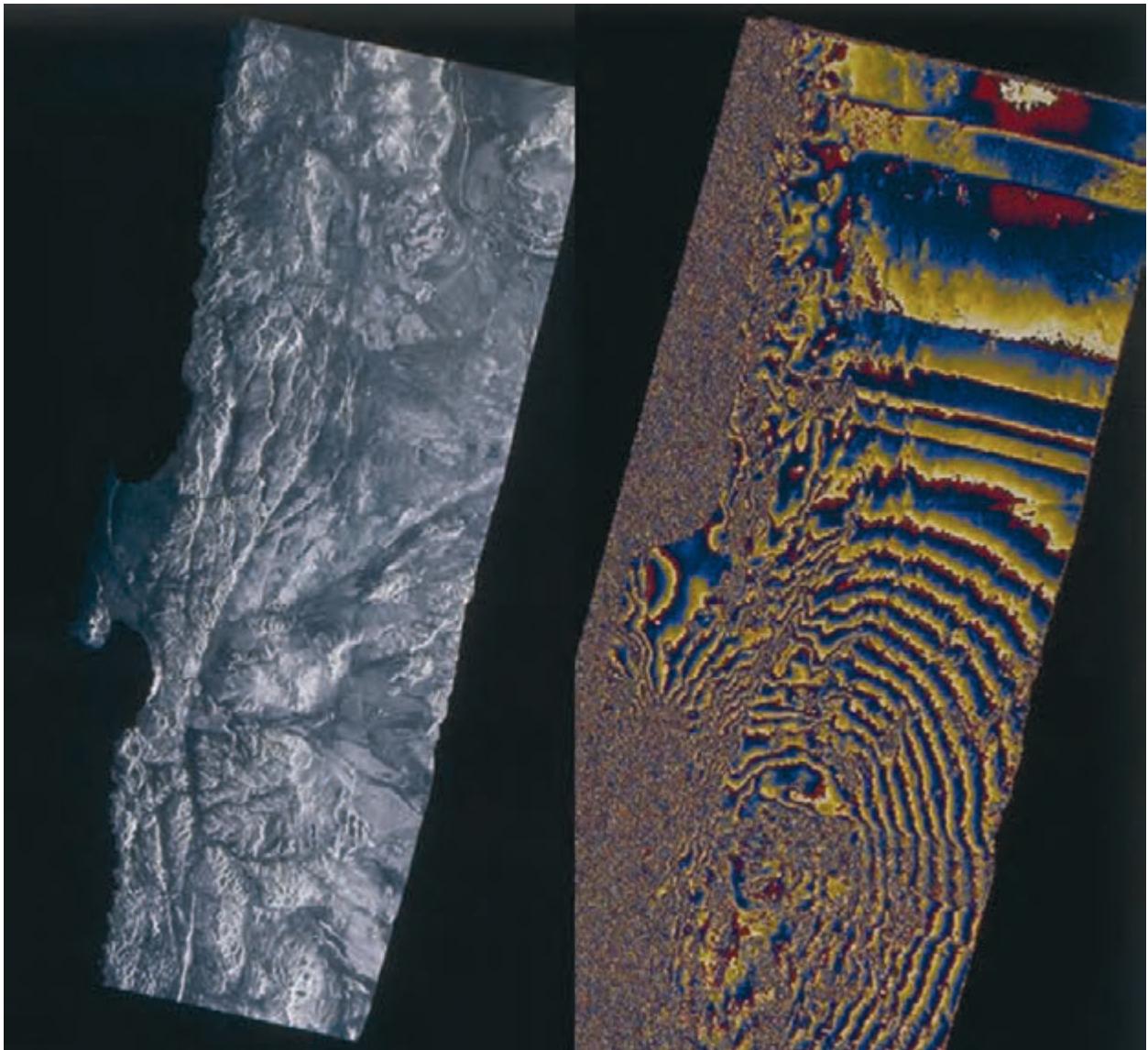


Figure 3-5: Large deformation field with errors: Chile earthquake

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:

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

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 :

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

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.

Part B

InSAR processing: a practical approach

1. Selecting ERS images for InSAR processing

1.1 Introduction

Selection of SAR images suitable for interferometry use is the first step to be carried out for any interferometric processing. It is a key step, since the criteria adopted for selection of the images have strong impact on the quality of the final results. These criteria depend upon the specific application for which the SAR interferometric images are required.

In this chapter a few selection criteria will be given concerning the two most important InSAR applications: Digital Elevation Model (DEM) generation and Differential Interferometry (DInSAR). In particular, we shall analyse how to select the following parameters in order to get the best results from the SAR interferometric analysis:

- View angle (ascending and descending passes)
- Geometrical baseline
- Temporal baseline
- Time of the acquisition
- Coherence
- Meteorological conditions

Before starting the analysis of the selection criteria, it is worthwhile spending a few words on the information available about ERS images.

1.2 Available information about ERS images

1.2.1 The ESA on-line multi-mission catalogue

A list of ERS images acquired over a certain area is easy to obtain, thanks to the EOLI software, available at the appropriate ESA/ESRIN site [EOLI].

This software allows the user to perform fast inventory searches on the major ESA-supported missions, by means of a user-friendly graphic interface. All images acquired over the area of interest can easily be identified. Moreover, if the ‘interferometry’ Query Mode is selected (see Figure 1-1), the relative baselines can be listed and the range set.

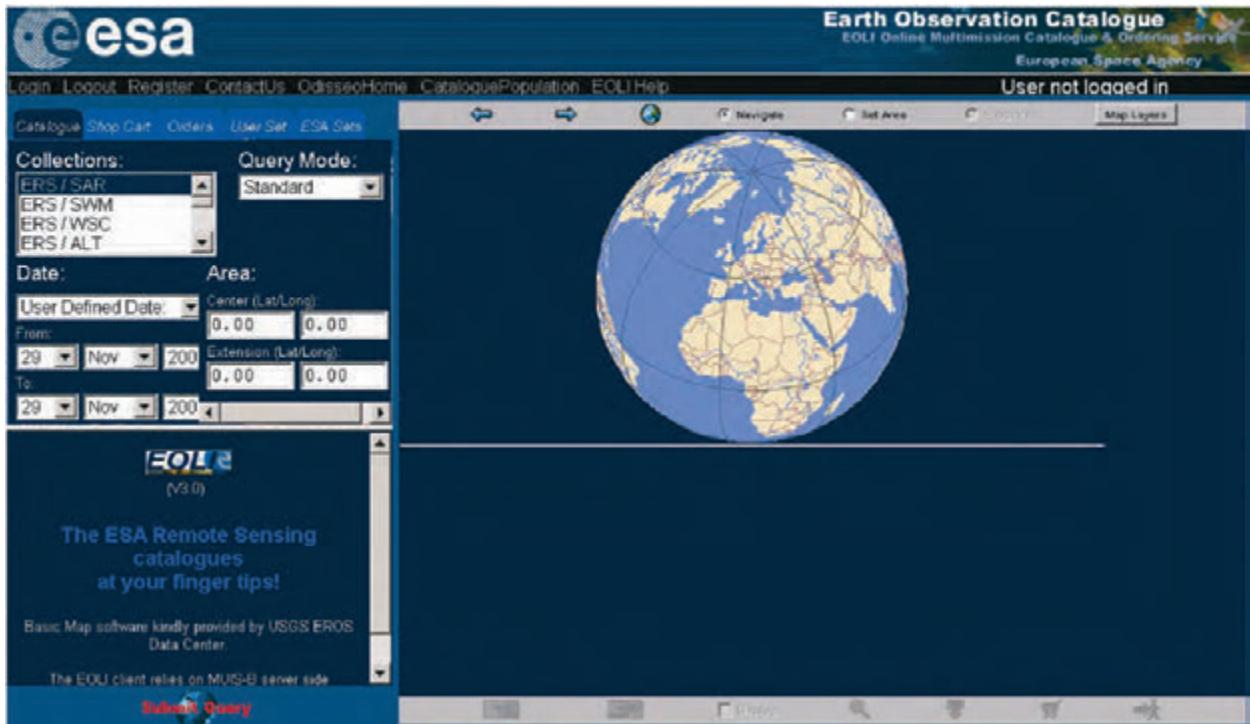


Figure 1-1: The EOLI query panel. The area of interest can be selected on the right panel by setting a window on the world map, or on the left panel by entering the geographical coordinates. On the left panel, users can select the mission type, the range of acquisition dates and the query mode. If the ‘interferometry’ query mode is selected, users can set the range of perpendicular baselines among ERS-SAR images. In this mode, the user can also select the satellite combination among ERS-1/ERS-2 (Tandem), ERS-1/ERS-1, ERS-2/ERS-2 and ERS-1/ERS-2.

1.2.2 DESCW

Another way to obtain the same information as supplied by EOLI is offered by an off-line application named DESCW [DESCW]. The main features offered by DESCW are shown in Figure 1-2 and Figure 1-3.

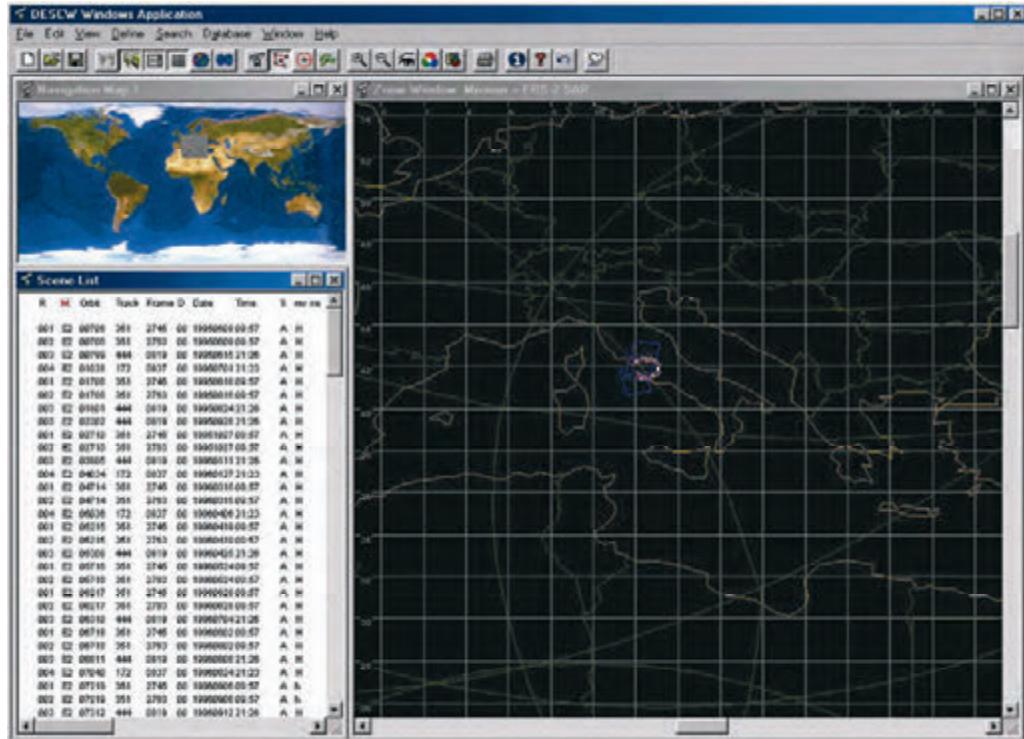


Figure 1-2: The DESCW query panel. The area of interest can be selected by providing the geographic coordinates and checking the result on the map on the right panel. Users can select the mission type, the range of acquisition dates and the range of baselines.

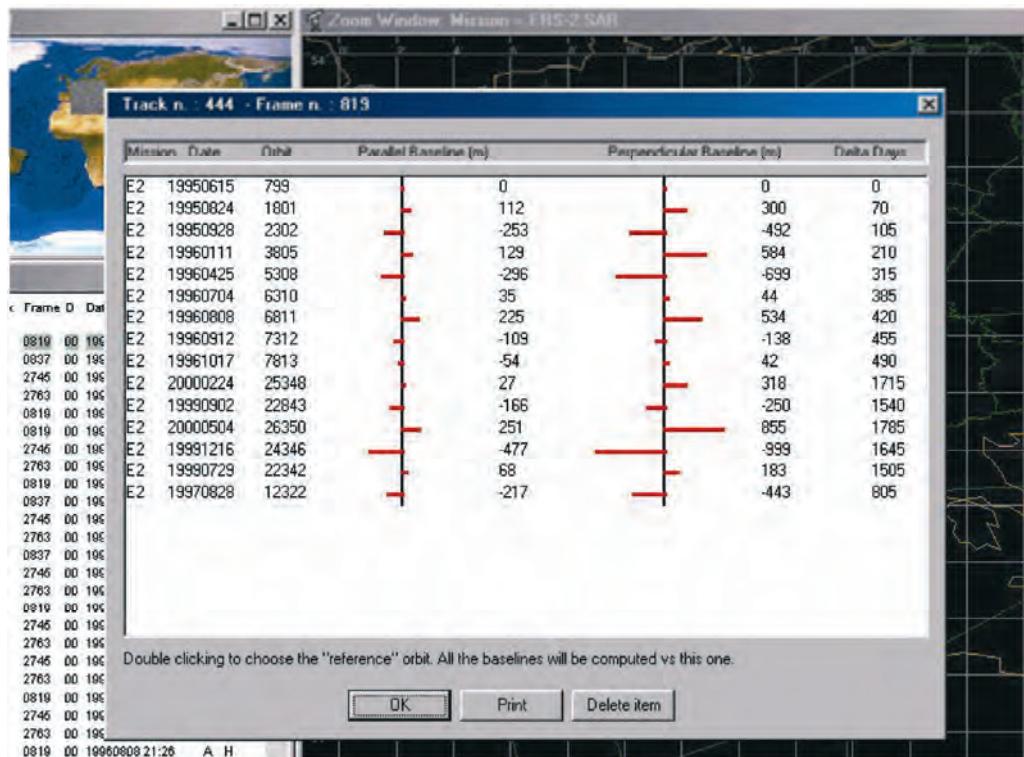


Figure 1-3: The DESCW software provides estimated perpendicular baselines of all the selected images with respect to a reference nominal orbit

1.2.3 Expected coherence (prototype)

View angle, geometrical and temporal baseline as well as acquisition time can be identified from the EOLI or DESCW catalogues. However, as far as coherence is concerned, a complete set of information is still not available. However, a prototype software application based on the Interferometric Quick Look (IQL) processor has been developed at ESRIN.

This software allows fast generation of SAR interferograms with reduced resolution, and coherence maps relative to long strips of ERS acquisitions (thousands of kilometres). Many examples, covering a wide range of land surfaces, have already been processed at ESRIN and are available on the web [INSI]. Users who are not familiar with SAR interferometry should take advantage of these examples that demonstrate the sensitivity of the coherence with respect to the land surface type. An example of the information supplied by the IQL software is shown in Figure 1-4 and Figure 1-5.

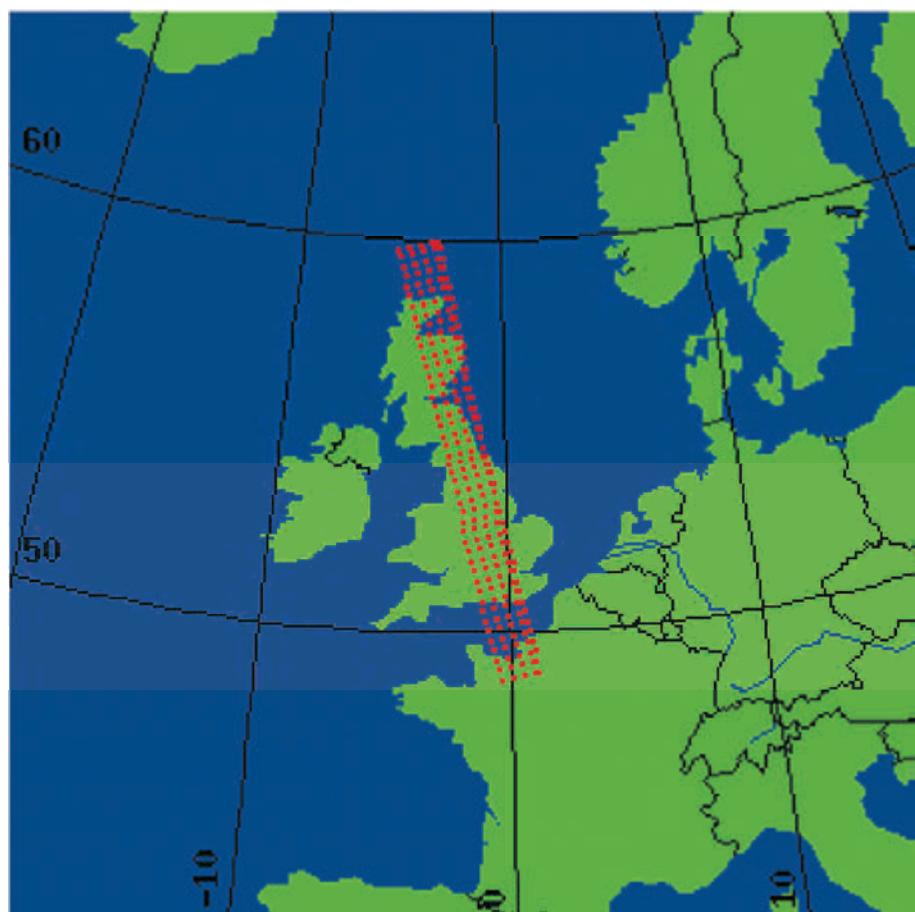


Figure 1-4: Map of the area processed by the IQL software

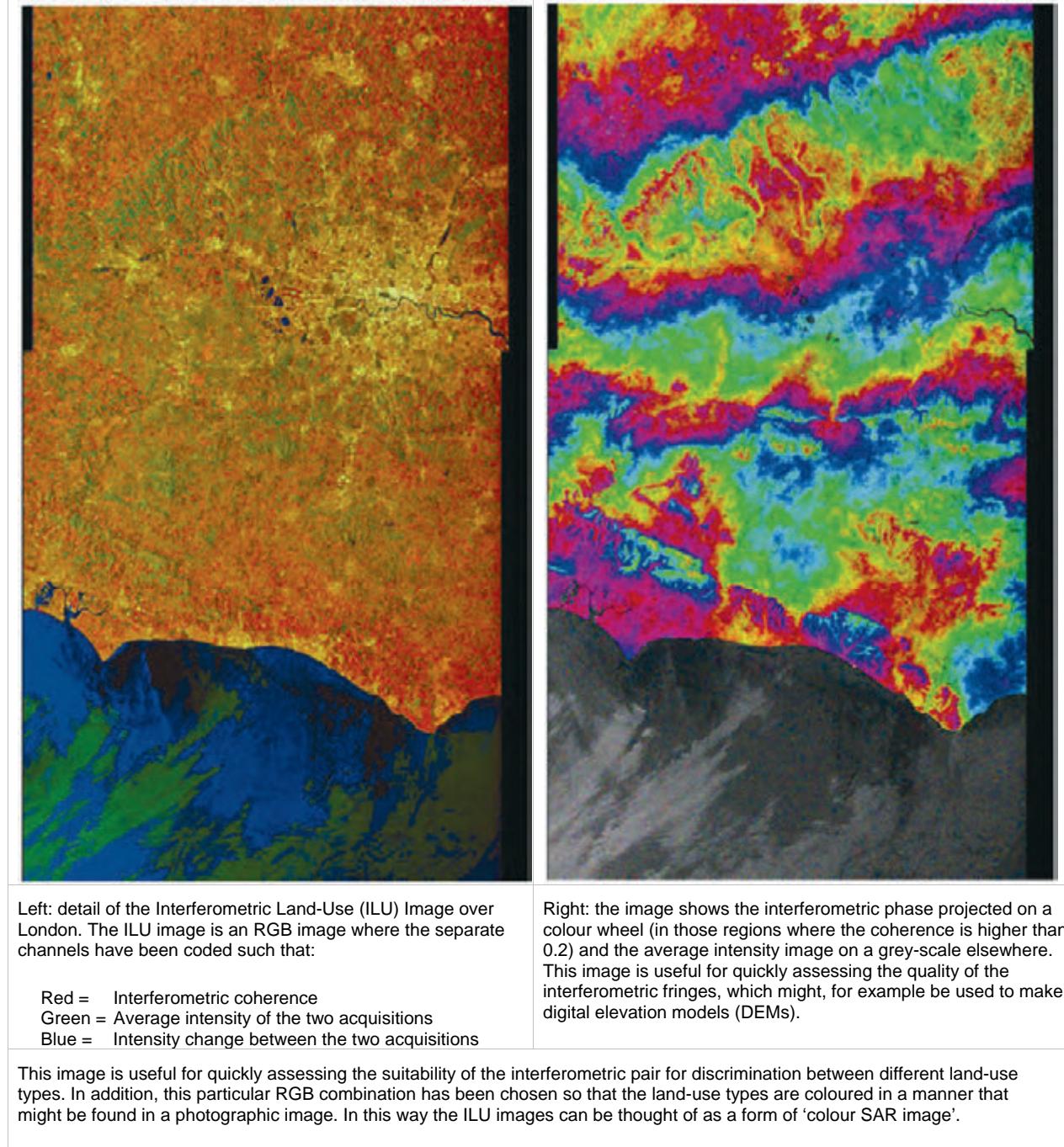


Figure 1-5: Interferometric Land-Use and phase images of London

From the analysis of thousands of InSAR images the following general conclusions on the expected coherence can be drawn:

- Urban areas and areas with exposed rocks maintain a high level of coherence even after several years.
- Sparsely vegetated areas and agricultural fields generally show high coherence on Tandem acquisitions (1-day interval) and much lower coherence after 35 days. Nonetheless, a suitable coherence value has

often been detected by comparing images acquired with a temporal interval of an integer number of years, i.e. at the same period of the year. Usually winter to winter data is best, when there is the least amount of vegetation on the ground.

- Forested areas (especially tropical forests) and water basins do not show a sufficient level of coherence even over a 1-day interval.

Finally, the geometrical deformation introduced by SAR should always be taken into consideration in order to maximise the coherence:

- Areas in foreshortening become non-coherent as soon as the perpendicular baseline is greater than a few metres.
- Areas with opposite slopes usually show the best coherence if not in shadow, since the spatial resolution is higher and the actual critical baseline is greater than that of flat terrain.

As an example, if the area of interest lies on a slope that is mainly oriented towards the West, it would be foreshortened on SAR ascending images (the ERS antenna looks to the right). Thus, descending ERS orbits should be selected.

1.3 Selecting images for InSAR DEM generation

In general the following recommendations should be followed when making digital elevation models from ERS interferometric data:

- Select Tandem acquisitions to reduce temporal decorrelation.
- Interferograms with very small perpendicular baseline values (< 30 m), though easy to unwrap, are almost useless due to their high sensitivity to phase noise and atmospheric effects.
- Interferograms with normal baseline (values higher than ~450 m) are usually almost impossible to unwrap if no *a priori* DEM is available and the topography of the area is not very smooth. Moreover the coherence is generally small, due to the high geometrical and volume scattering decorrelation [Gatelli94, Zebker92, Rodriguez92].
- The optimum perpendicular baseline is in the range between 150 and 300 metres. However, the best result is achieved by using more than one interferogram: interferograms with small baselines can be exploited to help unwrap interferograms with high baselines. Moreover, different interferograms can be combined in order to reduce the atmospheric artefacts.
- If no Tandem pair is available, consider using phase A, B and D ERS-1 acquisitions (3-day repeat cycle) instead of phase C (35-day repeat cycle).
- When the DEM will be used for differential interferometry applications, use the same track as that used to estimate possible ground deformations, in order to avoid the necessity of image interpolation.
- Coherence values are affected by local weather. Avoid acquisitions during rain, snow or strong wind. These phenomena usually cause loss of phase coherence. Weather information can be often recovered from historical databases available on the web.

- Nighttime acquisitions are usually less affected by atmospheric effects [Hanssen98].
- Discard images acquired during very hot days: hot air can hold much more water vapour than cold air (a major cause of atmospheric artefacts in SAR interferograms) [Hanssen98].
- Usually Tandem pairs acquired on vegetated areas during the dry season show higher coherence values than those acquired during a wet season.

1.4 Selecting images for Differential InSAR applications

In this section the criteria for selecting ERS images for measuring ground deformations are listed without detailed comments. Chapters B4 and C6 are dedicated to Differential InSAR (DInSAR) applications, with more detailed analysis.

Repeating Equation A.2.7, the interferometric phase is given by:

$$\Delta\phi = -\frac{4\pi}{\lambda} \frac{B_n q}{R \sin \theta} + \frac{4\pi}{\lambda} d \quad \text{Equation 1.1}$$

From this it can be seen that there are various different ways to produce a differential interferogram:

1. Single interferometric pair and near-zero baseline

With a single interferometric pair (two SAR images) and baseline B_n close to zero: the interferometric phase contains the motion contribution only (see Equation 1.1). No other processing steps are required.

2. Single interferometric pair and non-zero baseline

With a single interferometric pair (two SAR images) and non-zero baseline: the interferometric phase contains both altitude and motion contributions (see Equation 1.1). The following processing steps are required:

- 1) An available DEM must be re-sampled from geographic to SAR coordinates and the elevation must be converted into interferometric phases by inverting Equation 1.1. The same baseline should be used as for the interferometric pair.
- 2) These ‘synthetic’ fringes should be subtracted from those of the available interferometric pair. Notice that this operation can be conveniently done in the complex domain by multiplying the actual interferogram by the complex conjugate of the synthetic one.

3. Three interferometric images and no motion

With three interferometric SAR images and no terrain motion between two of them, one image should be selected as a common master. Two

interferograms are then formed: the two slave images are registered to the common master.

The shortest temporal difference (to gain coherence and avoid terrain motion) and a medium/high baseline (to gain elevation accuracy) should be selected for the first interferometric pair: typically one day and 100 – 300 metres in the ERS case. The second pair should have a larger temporal difference (it should contain the terrain motion) and a short baseline.

The following processing steps are required:

- 1) The first interferogram should be unwrapped and scaled by the ratio of the two baselines.
- 2) Its phase should be wrapped again and subtracted from that of the second interferogram (generally done in the complex domain as described in point 2 above).

However, if the baselines of the two pairs are in an integer ratio, no unwrapping is required. In this case the phases of one interferogram can be directly scaled by the integer ratio between baselines and subtracted from the phases of the other interferogram. The available collection of images should be analysed carefully to check if this very favourable condition can be met (phase unwrapping is still one of the most delicate points in SAR interferometry).

4. Two image pairs and no motion in one of them

With two interferometric pairs (four SAR images) and no terrain motion in one of them: there are two master images, each of them with a slave image. All the images should be registered to each other. We end up with two interferograms as in the case of three SAR images analysed in point 3, so the same steps are required.

1.4.1 Hints for image selection

- Select either ascending or descending passes, depending on which will avoid foreshortening in the area of interest.
- Select those image pairs with the smallest perpendicular baseline in the required range of dates. Bear in mind that the smaller the baseline, the smaller the topography contribution to the interferometric phase. As a consequence, a less precise DEM will be required for the topography subtraction. Moreover, the smaller the baseline, the higher the expected coherence.
- Check first the possibility of using only three images: a tandem pair (for DEM generation) and a third image, acquired after the desired time interval, that shows a small perpendicular baseline with either the first or the second image of the selected tandem pair (to make a second interferometric pair).

2. Interferogram generation

2.1 Introduction

This section discusses generation of full resolution interferograms and coherence maps. The algorithm described here is illustrated in the block diagram of Figure 2-1. This algorithm can be used to generate both interferograms and differential interferograms, i.e. interferograms corrected by the ‘known’ topography, provided as a Digital Elevation Model. It has been proposed and implemented in a prototype processor under ESA contract [MontiGuarnieri01B].

The mandatory inputs for interferogram generation are two Single Look Complex (SLC) images that are focussed and that preserve the phase. These are referred to as ‘master’ and ‘slave’; the meaning will be clarified in a following section. These images should have a suitable baseline, according to the image selection criteria just defined.

The scheme discussed here applies to two full-resolution images, such as Envisat IM, but will be generalised in the third part of this manual for a combination of different SAR modes (ScanSAR, AP, etc.).

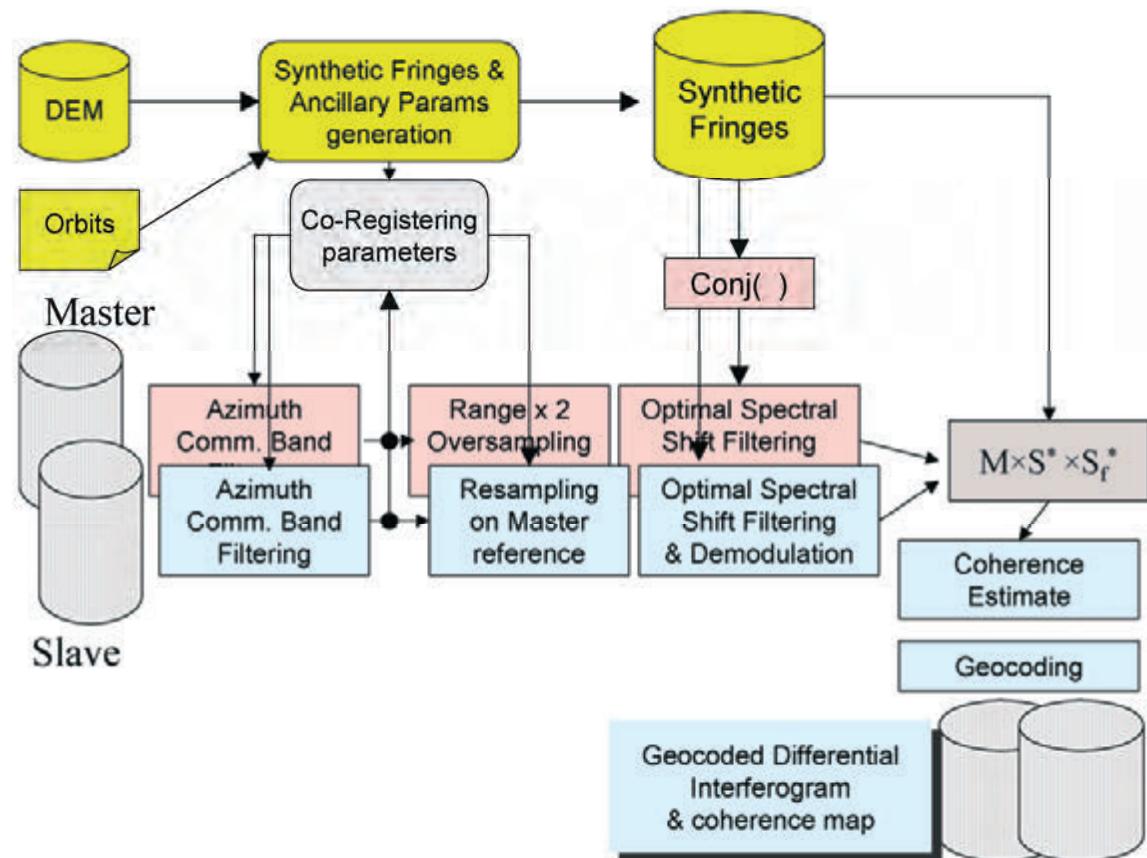


Figure 2-1: Schematic block diagram of an interferometric processor

According to the block diagram of Figure 2-1, besides the two interferometric SLC images, a DEM of the area imaged is assumed also to be available. The role of the DEM is twofold.

Firstly it is used, in conjunction with the precise knowledge of the satellite orbits, to estimate and compensate for topography in the final interferogram. Eventually, this produces a ‘differential’ interferogram, suitable for monitoring and detecting changes.

Secondly, the DEM is used to provide an optimal removal of ‘baseline decorrelation’, as will be discussed later.

In many cases a DEM is not available and a flat altitude profile is instead assumed. As a result, the final interferogram will be simply compensated for ellipsoidal Earth, or ‘flattened’: the scheme of Figure 2-1 still applies, but with some simplifications.

2.2 Generation of synthetic fringes

A synthetic interferogram is generated based on the precise sensor orbits [Scharroo94], timing information, and scene topography, e.g. from a DEM. This step assumes that a DEM is available: if an accurate DEM, sampled at the SAR resolution, is given, a complete removal of topography is possible, together with efficient spectral shift filtering. However, the use of low resolution DEMs with global coverage (like GTOPO30 [GTOPO30] or ACE [Berry2000]) could be sufficient in many applications or scenes. At a minimum, a proper ellipsoid is sufficient for interferogram flattening.

Generation of the synthetic interferogram is represented in Figure 2-2.

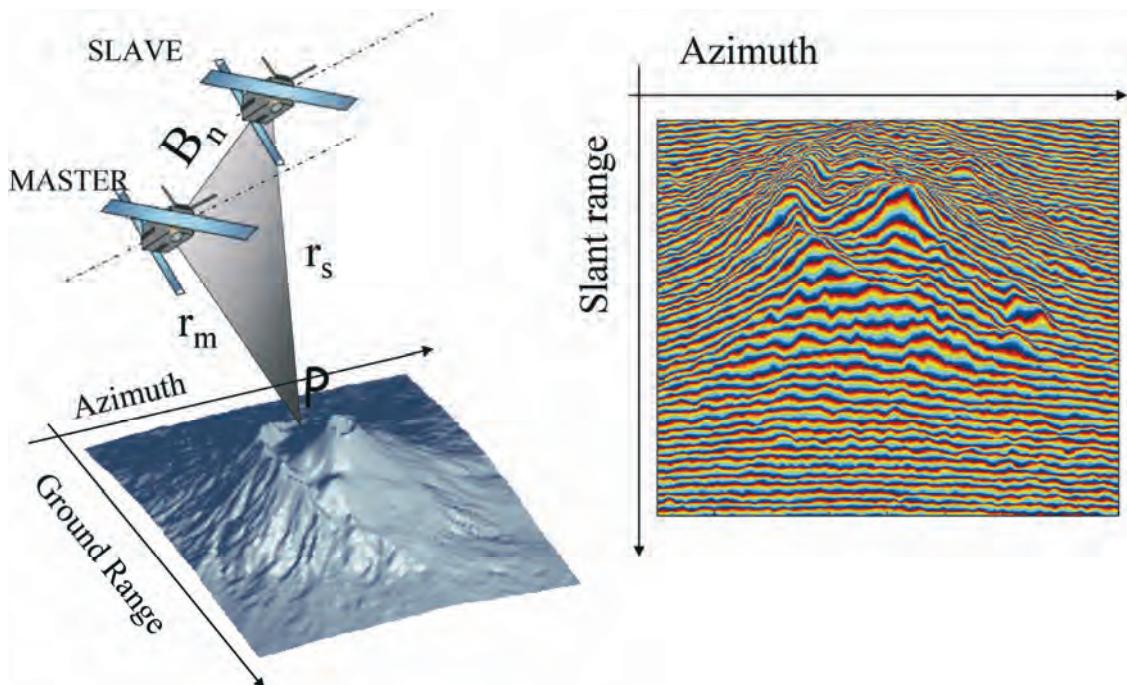


Figure 2-2: Generation of a synthetic interferogram (right) based on a DEM (left) and knowledge of the sensor orbits and timing

The interferogram phase is estimated by computing, for each target P in the (azimuth, slant range) plane, the difference in the sensor-target travel path distance for the two satellites (see Figure 2-2 (left)):

$$\phi(P) = \frac{4\pi}{\lambda} [r_M(\vec{P}) - r_S(\vec{P})] \approx \frac{4\pi B_n}{\lambda r_0} \Delta r(P) \quad \text{Equation 2.1}$$

The actual implementation is tricky, since the DEM is regularly gridded in a ground reference according to some cartographic projection, whereas $\phi(P)$ is to be computed on a regular grid in the (slant range, azimuth) SAR referenceⁱ. A suitable algorithm is needed [MontiGuarnieri01B]. This problem does not exist if a flat terrain profile is assumed, e.g. to flatten the interferogram with respect to a reference ellipsoid.

The synthetic interferogram provides an unwrapped phase field that can be used for the following purposes:

- The phase can be subtracted from the final SAR interferogram to remove the known topography, hence providing a **differential interferogram**, for monitoring changes
- The following term provides a map of the pixel to pixel mapping from the master to the slave image, to be used for **image co-registering**:

$$\Delta r(P) = \frac{4\pi}{\lambda} (r_M(\vec{P}) - r_S(\vec{P})) \quad \text{Equation 2.2}$$

- The information on local slopes implied in $\phi(P)$ can be used to provide an optimal **spectral shift filtering**, as will be discussed in a later section

2.3 Co-registering

The co-registration step is fundamental in interferogram generation, as it ensures that each ground target contributes to the same (range, azimuth) pixel in both the master and the slave image.

In an ideal case of perfect parallel orbits and aligned acquisitions, co-registration would only need to compensate for the differing geometry due to the different view angle (the parallax effect, see Figure 2-2 (left)). This would be compensated by a proper cross-track stretching of one image.

In practice, the co-registration should also account for:

- orbit crossings/skewing
- different sensor attitudes
- different sampling rates (maybe due to different **pulse repetition frequency (PRF)**, sensor velocities, etc.)
- along- and across-track shifts

All these effects are summarised in Figure 2-3.

ⁱ The complexity of the problem is due to the fact that the ground to SAR reference transformation depends on the elevation.

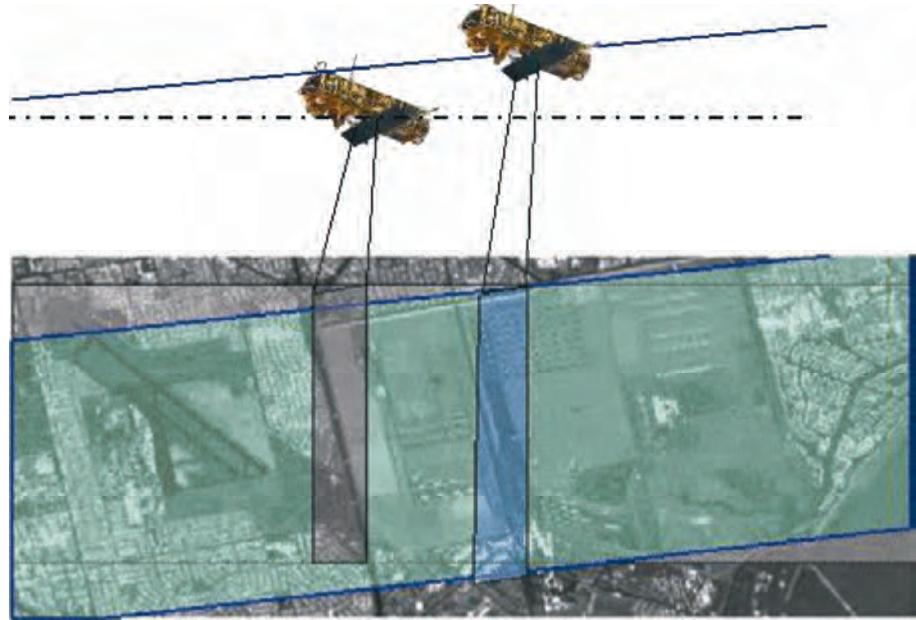


Figure 2-3: Co-registration should be performed to compensate for orbit skew, different sampling and baseline-induced deformations between the two acquisitions

In satellite sensors like those of ERS and Envisat the required transformation is mainly composed of:

- A small rotation of the two images
- A range and an azimuth ‘stretch’ (change in sampling rate)
- Further second-order effects

2.3.1 Co-registering coefficients

Proper space alignment between the two images should be performed on a pixel by pixel basis, with an accuracy of the order of one tenth of the resolution, or better (see chapter C.1 for details).

A map of pixel-to-pixel correspondence could be provided, within an accuracy of one centimetre, by the synthetic fringes (see Figure 2-2 (right)), scaled by the wavelength, i.e. the term Δr in Equation 2.1.

In theory, co-registration should depend on the (local) topography [Lin92]. However the impact of the elevation is almost negligible in most cases. As an example, in an ERS or Envisat geometry, with a baseline of 100 m, an elevation change of 2500 m would cause ~25 fringes, e.g. a shift of $25 \times 0.028 = 0.7$ m – that is, close to 1/10 of the slant range resolution.

Therefore, the co-registration map can be provided as a smooth polynomial that approximates the pixel-to-pixel shift with the assumption of targets lying on the ellipsoidal Earth surface.

In satellite-borne Synthetic Aperture Radars such as ERS and Envisat ASAR, the sensor velocity and attitudes are so stable that the master-slave deformation of an entire frame (100×100 km) can be well approximated by the following polynomial:

$$\begin{cases} r_s = a \cdot r_m^2 + b \cdot r_m + c \cdot a_m + d \\ a_s = e \cdot r_m^2 + f \cdot r_m + g \cdot a_m + h \end{cases} \quad \text{Equation 2.3}$$

where: (r_m, a_m) are the range and azimuth coordinates respectively of the master image
 (r_s, a_s) are the range and azimuth coordinates where the slave image should be evaluated

The convention here assumed implies that the ‘slave’ image is the one that is actually resampled, so that the final interferogram will be in the same (slant range, azimuth) reference of the master image.

The eight coefficients involved in Equation 2.3 represent the following transformations, illustrated in Figure 2-4:

- a fixed azimuth shift, coefficient (d), (due to different timing along orbit), and fixed range shift, (h) (mainly due to the perpendicular baseline component)
- a stretch in range, (b), due to the normal baseline variation with range and an azimuth stretch, (g) due to variation in PRF and or satellite velocity;
- a range and azimuth skew, (c,f) that approximate an image rotation, for small rotation angles;
- two second order terms (a,e) that are required for processing large range swaths.

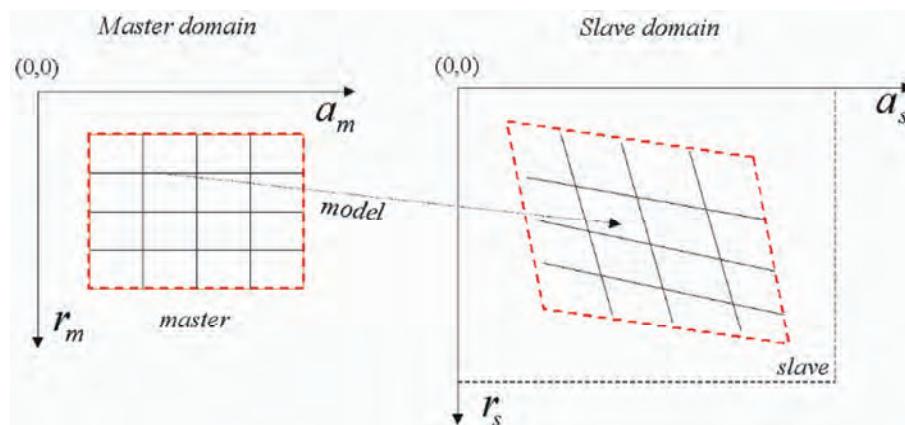


Figure 2-4: Deformation model to register the slave image on the reference grid of the master. Note the azimuth & range shift, stretch & rotation (actually implemented by two 1-D skews).

Note that, for ERS and Envisat InSAR, the major effects are due to the azimuth and range shift. The range and azimuth stretch are very small: a few pixels over an entire frame [Gatelli94, MontiGuarnier01B]. Note that in Envisat ASAR missions the azimuth stretch could increase within reason in order to accommodate the PRF variation in the two acquisitions.

The range and azimuth skew are rather limited, as the image rotation usually corresponds to an angle $\beta < 1/100^\circ$ (as measured for ERS, see [Solaas94]). Such values justify the approximation of a rotation as two skews, which can

be efficiently compensated by two 1-D operators (instead of a full 2-D resampling).

2.3.2 Co-registering parameter estimation

The simplest way to retrieve the proper values of the co-registering coefficients is by exploiting the known acquisition geometry, e.g. the $\Delta r(P)$ in Equation 2.1 that was already computed while generating the synthetic fringes. A Least Mean Square regression based on a regular grid of, say, 500 points displaced over the whole frame will be enough.

In practical cases of ERS interferometry this scheme will not work, due to the uncertainty in the acquisition timing (both for the fast and the slow times), leading to errors usually of a few pixels (but much larger, in some limited cases). This should not be the case for Envisat, whose predicted accuracy of timing will be much better.

It is thus useful to retrieve these co-registering parameters from the actual data. This estimate is usually accomplished by dividing each image into small patches and then finding the range and azimuth offset for each patch. Such processing is represented in Figure 2-5, which draws the sub-images together with an example of displacements in the form of a vector field.

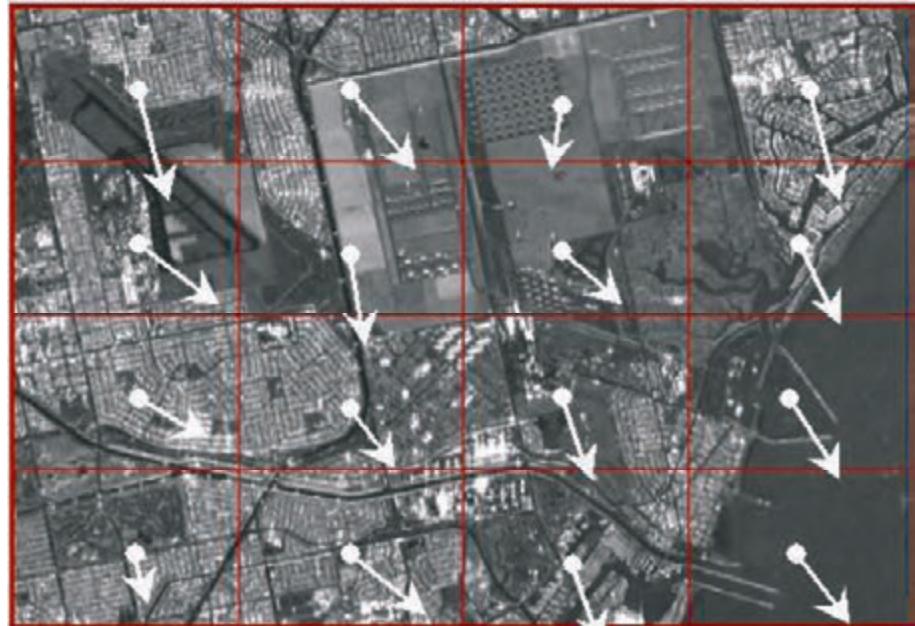


Figure 2-5: The estimate of co-registering parameters from data is here pictorially represented. A shift vector is drawn for each sub-image: it has been computed by maximising some statistical measure (e.g. the cross-correlation between master and slave amplitudes), and then a smooth polynomial is fitted by a weighed LMS technique.

These displacements are estimated by maximising some statistical measure, such as amplitude cross-correlation or fringe contrast [Carter87, Gabriel88, Lin92, Just94, Bamler98, Moreira 2000]. Finally, a smooth polynomial is fitted over the whole measure, for example by LMS, weighted according to

the local SNR estimates (such as the amplitude normalised correlation coefficient or the absolute value of coherence).

Note that both fringe-based and amplitude-based techniques have complementary features and drawbacks. Fringe-dependent techniques have superior performances when the topography is flat or the baseline is moderate so that fringes are slowly varying in the sub-image. However, they need to exploit small windows, hence they perform badly in the presence of image contrasts. Furthermore, they need to estimate the local fringe frequency, hence they have a demanding computational cost.

Amplitude-based techniques, on the other hand, could work very well with very wide baseline spans and image contrasts; the computational cost is moderate; but they have a coarser accuracy (at the same number of degrees of freedom and SNR), hence they need to exploit much larger sub-images.

Amplitude-based techniques are usually exploited for a first guess of the co-registering parameters, or in a multi-baseline environment where all the images must be co-registered with the same master (disregarding the baseline).

2.3.3 Implementation of resampling

Implementation of the slave image resampling according to the polynomial mapping in Equation 2.3 is quite efficient, since it can be approximated by two one-dimensional resampling steps: along range and then along azimuth. Each step can be efficiently performed in the space domain by means of small kernels (typically 6 points), that can be designed according to the General Least Square Filter approach (discussed in [Laakso96] or according to [Hanssen99]). Usually, a look-up-table (LUT) of kernels is pre-calculated for fractional pixel shifts in steps of 1/100. Note that, when co-registering parameters derived from orbits are sufficiently accurate, it is possible to include the co-registration in the focusing operator, as described in [Fornaro95].

2.4 Master and slave oversampling

Range oversampling by a factor of two is a mandatory step in high quality interferogram generation. The purpose of this step is to avoid the uncorrelated contributions that would arise in the spectral cross-correlation implied by the interferogram generation, e.g. Hermitian multiplication of the two focused images [Gatelli94]. The interferogram spectrum is then the cross-correlation of the spectra of the two images.

It can be shown that noise introduced without interpolating is marked where the range fringe patterns are dense (e.g. on the fore-slopes and for larger baselines), whereas it is quite small for lower baselines and in the azimuth direction. Azimuth oversampling can be usually be avoided since the effect of azimuth slopes is much more limited.

Range oversampling $\times 2$ is efficiently implemented by **Fast Fourier Transform (FFT)** techniques, both in the master and in the slave. Note that

the implementation of range oversampling prior to slave image resampling (for co-registering in the master reference), allows for the use of short, efficient space-domain kernels.

2.5 Range spectral shift & azimuth common bandwidth filtering

The purpose of these two different processing steps is to provide a sort of ‘phase co-registration’, such that the contributions that are mostly correlated in the two images are kept, but the uncorrelated contributions (that behave like noise) are removed prior to generation of the interferometric cross-product.

2.5.1 Range spectral shift filtering

A complete discussion of the range spectral shift is provided in chapters A2, B3. The basic idea is shown in Figure 2-6.

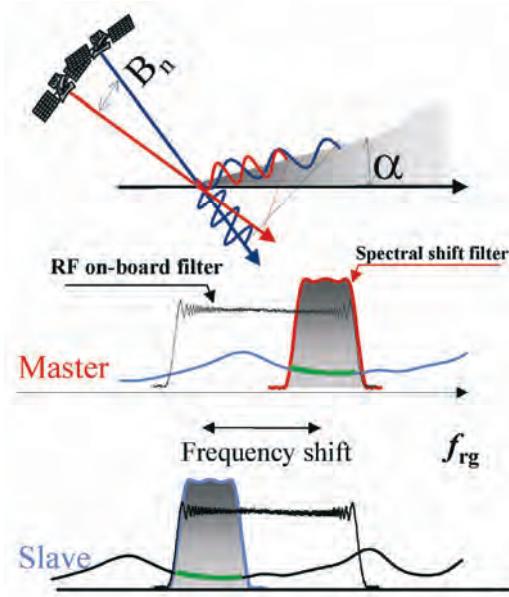


Figure 2-6: Range spectral shift. In range, the change in the view angle introduces a frequency shift in the reflectivity. The filters required to remove the uncorrelated spectral contributions are shown in the middle and bottom plots.

Note that two plane wave-fronts with the same **wave number** induce wave-fronts with slightly shifted wave numbers on the sloped terrain. This (spectral) shift has been computed in [Gatelli94] for constant sloped terrain:

$$\Delta f = -f_0 \frac{B_n}{r_0 \tan(\theta - \alpha)} \quad \text{Equation 2.4}$$

The capability of performing interferometry is subject to the fact that the same wave number is illuminated on the ground. In the example of Figure

2-6, the correlated spectral contributions are represented by the shaded areas: notice that here the spectral shift is approximately two thirds of the whole range bandwidth, and the common band one third. The non-correlated contributions should be removed by means of two complementary band-pass filters (represented in Figure 2-6) before computing the interferogram [Gatelli94]. Note that the central frequencies of the two filters are $\pm \Delta f/2$ and their bandwidth $B_r - \Delta f$.

The gain achieved by such filtering depends on the size of the spectral shift: in the case of flat Earth and for a baseline of 250 m, an ERS interferometric pair will have coherence ~ 0.75 in the absence of any other decorrelation source: an ideal unitary coherence value could be achieved by spectral shift filtering.

The implementation of such filtering is not trivial, as the spectral shift Δf of Equation 2.4 depends on the local incidence angle, θ , and hence on the terrain slope. Fixed filtering, tuned on the spectral shift for a flat Earth, is usually assumed. This however is not recommended in full resolution processing as, depending on the slope, it could even worsen the quality compared with no filtering at all [Fornaro01].

The proper range-varying implementation of such filtering has been proposed in [Davidson99, Fornaro01] and is detailed in Figure 2-7.

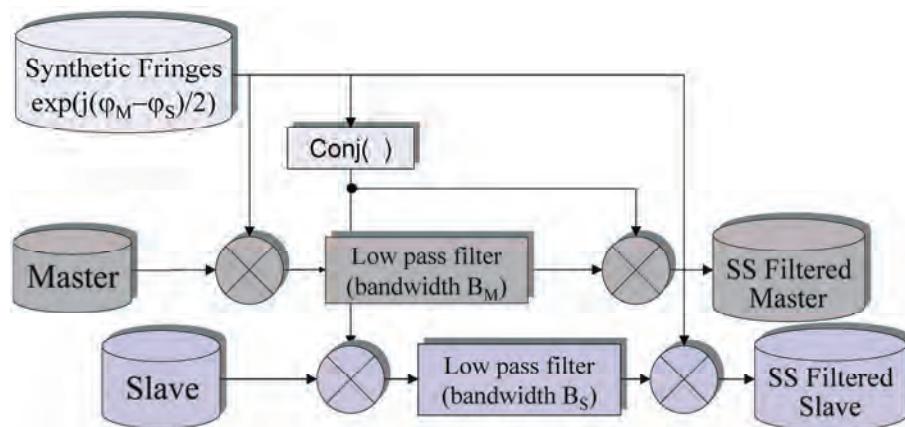


Figure 2-7: Optimal range-varying spectral shift filtering implementation

The scheme is fairly simple: one has to process each range line of the master image by:

- 1) demodulating by the synthetic fringe pattern (converted from phase to complex sinusoid),
- 2) low-pass filtering with the same bandwidth as the slave image acquisition,
- 3) reversing step (1) by modulating by the conjugate of the synthetic fringes.

Similar processing should be applied to the slave image. However, in this case the synthetic fringes should be conjugated, and the filter should have the same bandwidth as the master acquisition. Notice in Figure 2-7 that the final modulation by the synthetic fringe pattern is not applied to the slave

image, and this will lead to a differential interferogram (where topography has been removed).

The major limitation of this scheme is the need for a DEM in order to compute the synthetic fringes. If only an approximate DEM is available (such as global GTOPO30 or ACE), an alternative approach is suggested in [Fornaro01], where the synthetic fringe pattern is obtained by filtering and unwrapping the interferogram computed in a first iteration. This iterative approach allows a quality close to that achieved by using an accurate DEM.

2.5.2 Azimuth common band filtering

The azimuth common band filtering is somewhat complementary to the range one, the goal being once again to keep the mostly correlated contributions.

However, the case is quite different: the ‘azimuth spectral shift’ due to terrain slope is in fact quite small and can be completely ignored in full resolution SAR interferometry [MontiGuarnieri99A]. The ‘spectral shift’ is rather due to a possible variation in the antenna pointing, e.g. the Doppler Centroid (DC)ⁱⁱ between the two acquisitions.

The effect of a different Doppler Centroid on the azimuth spectra of the two acquisitions is shown in Figure 2-8. This concept is similar to the range spectral shift: there the same portions of two shifted reflectivity spectra were observed; here two different (shifted) portions of the same reflectivity spectrum are observed.

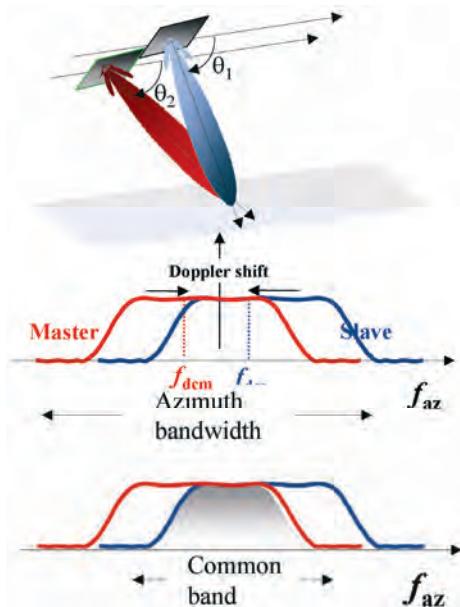


Figure 2-8: Azimuth common band filtering. In azimuth, the shift is due to a change in the squint angle. The filters required to remove the uncorrelated spectral contributions are shown in the middle and bottom plots.

ⁱⁱ The linear relation between angles and frequencies is due to the particular SAR’s time-frequency mapping.

The filters required to remove the decorrelated contribution are represented in the same figure. Each filter is band-pass, centered on the average Doppler Centroid:

$$f_c(r) = \frac{(f_{DC_M}(r) + f_{DC_S}(r))}{2} \quad \text{Equation 2.5}$$

where we have emphasised the (slight) dependence of the master and slave Doppler Centroid on rangeⁱⁱⁱ. The filter bandwidth should keep the ‘common bandwidth’, corresponding to the shaded area in the lower plot of Figure 2-8.

It is suggested, for computational efficiency, to perform Common Band (CB) filtering as the first step in the interferogram generation, e.g. before range oversampling both the master and the slave images, as shown in Figure 2-1. In this case, computing the filters is somewhat tricky, as the central frequency $f_c(r)$ in Figure 2-8 should be computed for ‘corresponding’ pixels in the two images, and this requires *a priori* knowledge, even approximate, of the co-registration coefficients (as master and slave are not pixel-to-pixel co-registered at that stage).

The proper implementation of the azimuth CB filtering should include compensation for the antenna pattern and the spectral ‘weighting window’ that are usually introduced during processing (see [Fornaro01, MontiGuarnieri01A] for details). This compensation is performed by an inverse filter that can be designed using the Remez approach. The maximum possible **spectral whitening** should be used, compatible with limits imposed by aliasing: usually up to 70–80% of the Doppler Bandwidth can be exploited.

Finally, a visual idea of the gain that can be achieved by performing azimuth CB filtering and range spectral shift filtering is shown in Figure 2-9, based on a realistic simulation of ERS tandem interferometry over Mount Vesuvius. In a real case, the gain would be reduced by the scene decorrelation noise.

ⁱⁱⁱ For space-borne SAR missions like ERS and Envisat, the variation of Doppler Centroid with azimuth can be ignored – for the purpose of CB filtering – within the extent of a frame.

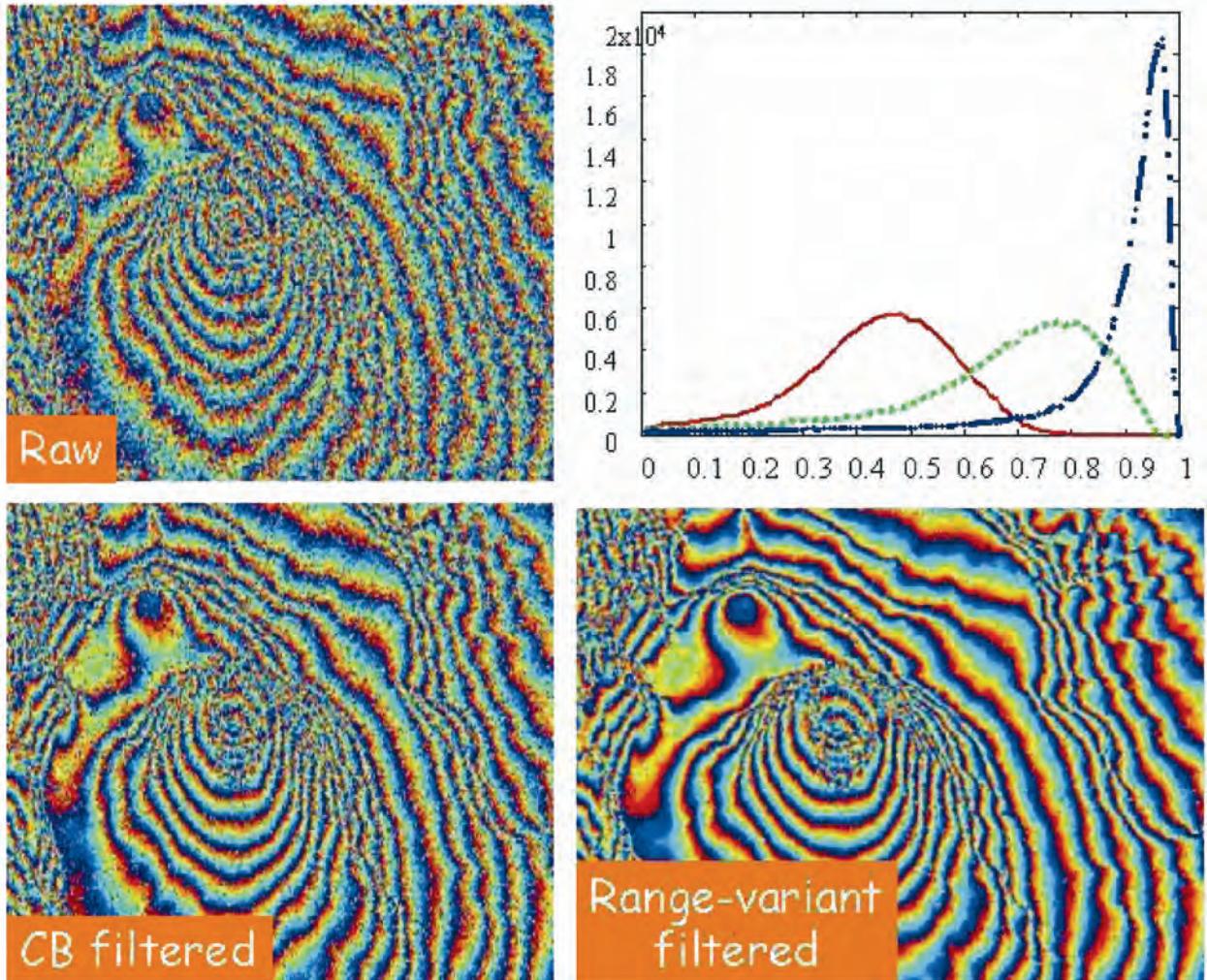


Figure 2-9: Effects of range and azimuth filtering on the final interferogram. A simulated interferogram of Mt. Vesuvius has been generated, (top left), by assuming no filtering. The decorrelation is entirely due to geometry and to the azimuth spectral shift (a Doppler centroid difference of 300 Hz was assumed). The other two pictures represent the interferogram achieved by performing azimuth CB filtering, and azimuth CB filtering + range spectral shift filtering. The histograms of coherence for the three cases are in the top right plot, where the higher coherence level plots correspond to the increasing levels of filtering.

2.6 Interferogram computation

Generation of the interferogram requires the pixel-to-pixel computation of the Hermitian product of two co-registered, spectral-shift-filtered images [Graham74, Gabriel88]:

$$v_i = u_M \times u_S^* \quad \text{Equation 2.6}$$

where u_M and u_S refer to the master and slave respectively.

The convention assumed here ensures that the interferogram is registered in the same (azimuth, slant range) reference as the master image, and its phase is the difference between the phase of the master and that of the slave, if necessary compensated for any further topography/flattening fringe pattern.

An example of a full-frame 100×100 km interferogram is provided in Figure 2-10.

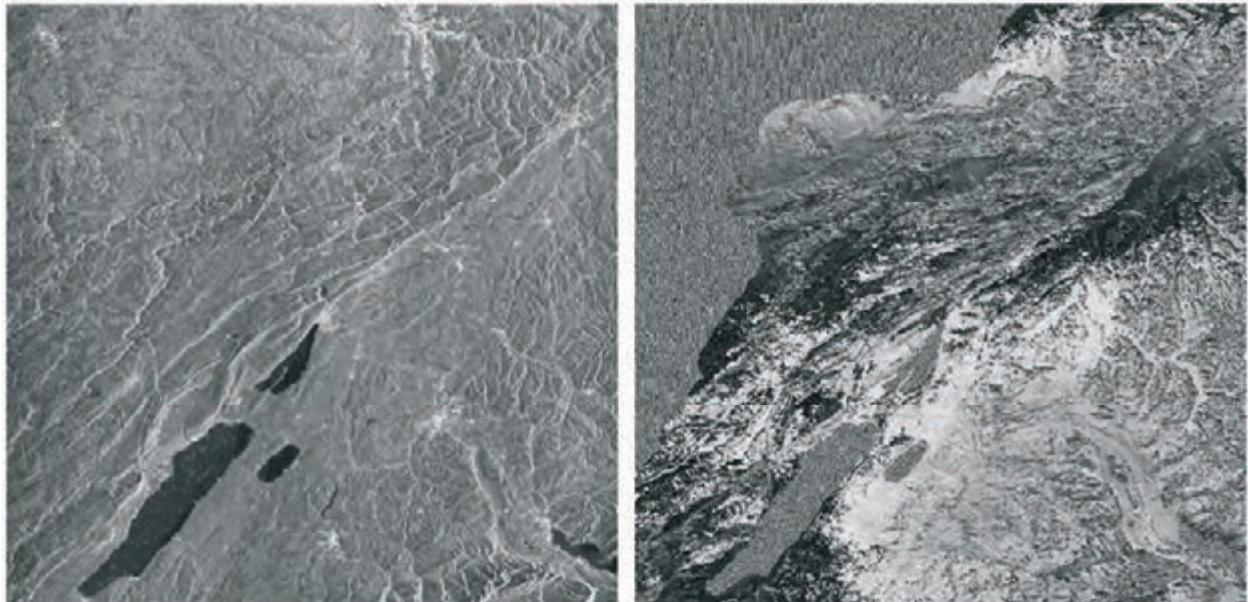


Figure 2-10: Full resolution, full frame Tandem ERS interferogram of Bern area. Baseline: 120 m. Left: absolute value. The interferogram phase (right) has been compensated by exploiting a 50 m DEM of Switzerland, according to the scheme in Figure 2-1. The residual phase shown in the figure is the result of DEM errors (uncompensated topography) and atmospheric artefacts. The fast fringe pattern in the top-left corner corresponds to an area of France that was not covered by the DEM.

Another example of a differential interferogram is provided in Figure 2-11; here residual fringes are mostly due to motion over the large revisit interval (two years). Further details are provided in the figure caption.

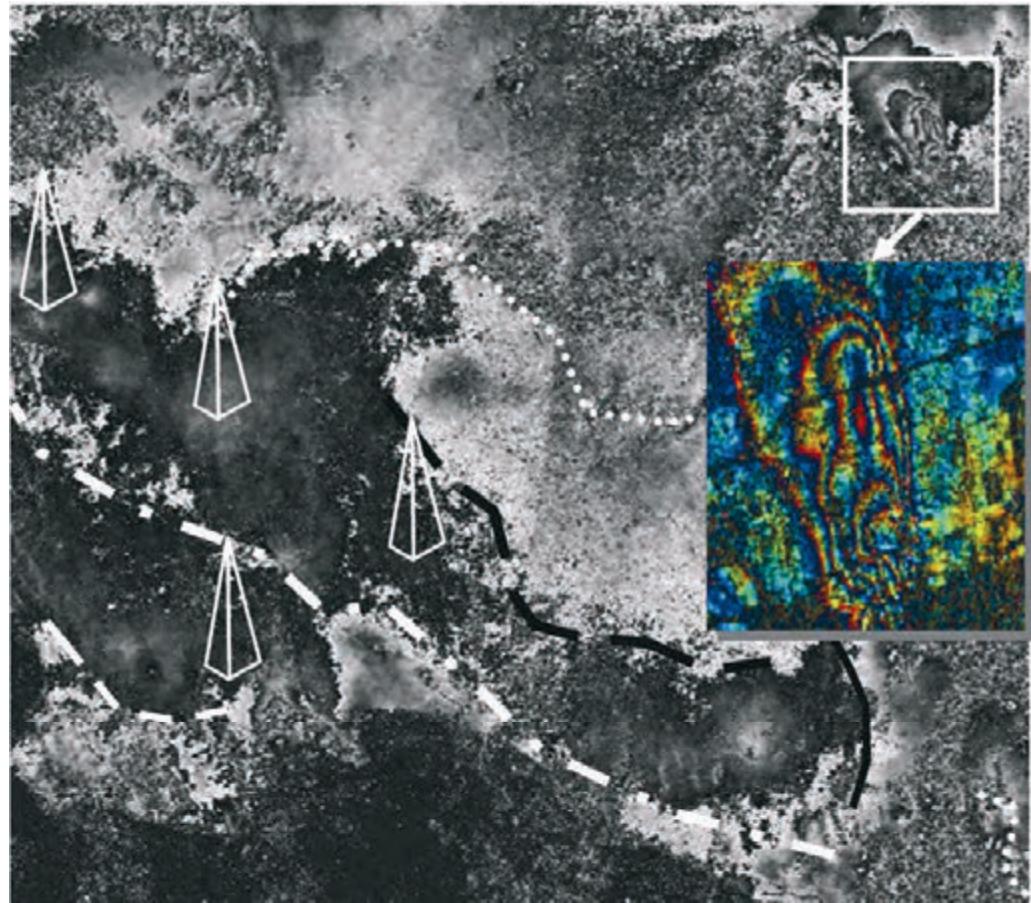


Figure 2-11: Full resolution, approximately half frame ERS interferogram of the LA area obtained by two acquisitions taken two years apart. Normal baseline: 120 m. The interferogram has been compensated by combining two different DEMs (the junction between the DEMs appears at the centre of the image). The residual phase, shown in the figure, is mostly due to motion and atmospheric effects. The principal faults and water pumping stations have been highlighted in the image. The image in the zoom-box is a close-up of the area of Pomona. Up to 5 fringe cycles (14 cm) were generated by subsidence due to water pumping.

2.6.1 Complex multi-looking

The interferogram described in section B.2.6 is usually referred to as a ‘raw interferogram’ as its phase is rather noisy, at least in the case of repeat-pass acquisitions that are strongly affected by temporal decorrelation. It is thus common practice to reduce the noise by averaging adjacent pixels in the complex interferogram. This processing, defined as ‘**complex multi-looking**’, [Rodriguez92, Goldstein98, Lee98] trades geometric resolution for phase accuracy (or altimetric resolution when the interferogram is exploited for DEM generation). Such averaging is quite effective with respect to any uncorrelated noise due to temporal, baseline, volume etc. sources. However it is not able to remove space-correlated artefacts, e.g. due to atmospheric turbulence, errors in flattening, or DEM removal etc. An example of the interferogram phases before and after space averaging is given in Figure 2-12.

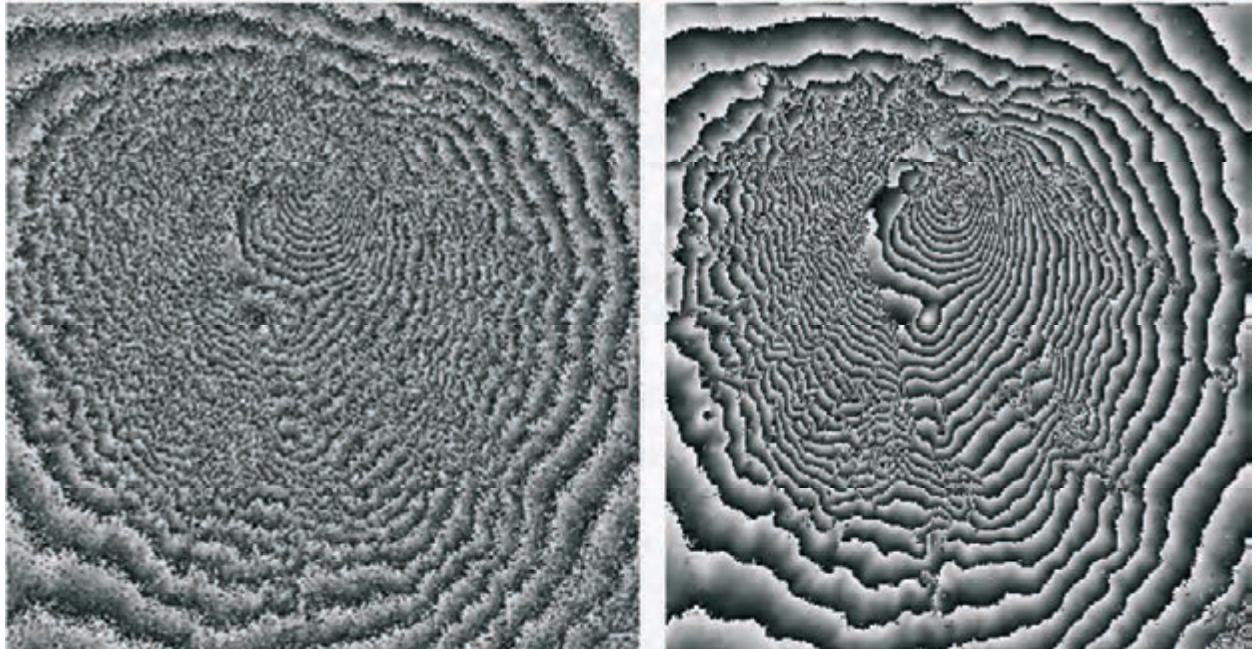


Figure 2-12: ERS tandem interferogram (compensated for flat Earth) before (left) and after (right) complex multi-looking. Mt. Vesuvius (Italy), normal baseline 250 m.

The implementation of interferogram complex multi-looking can be done by choosing from many different algorithms.

The simplest and most efficient is just averaging over a fixed mask, where masks are usually rectangular and designed to have the same size in azimuth and ground range. However, this technique is effective only if the interferogram phase is ‘constant’ (say within a small fraction of π) over the averaging window, and fails when phases vary. Therefore it can be applied to a differential interferogram, where most of the topography has been removed. In processing an ERS interferogram, such averaging is usually applied for flattened or non-flattened interferograms, by exploiting a box car window of 1×5 (slant range, azimuth) corresponding to $\sim 20 \times 20$ m on the ground.

The fixed mask filter can be improved by providing an estimate of the interferogram phase in the averaging window and compensating this phase before averaging (a sort of local flattening) [Prati92, Rocca94]. Usually linear phase is assumed and the interferogram is locally approximated by a complex sinusoid, whose frequency can be retrieved as discussed in section C2.3.

These techniques have still some drawbacks, as averaging makes sense only if performed on statistically homogeneous samples, i.e. within the same distributed target. The presence of different scattering mechanisms in the averaging window, and particularly of contrast variation, will introduce artefacts. A strong point target, for example, contributes to the multi-looked interferogram with a window of the same shape and size as that used for averaging. Therefore, a significantly better technique is suggested in [Ferretti96A], where areas of homogenous speckle have been previously

identified by a sort of speckle-Lee filter [Lee98], and then exploited for multi-looking. Such an approach results in averaging over non-uniform windows, whose shape and size is adapted according to the local statistics. The interferogram of Figure 2-12 has been filtered in this way.

2.6.2 Generation of coherence maps

Coherence, or better its absolute value (since it is a complex quantity), provides a useful measure of the interferogram quality (SNR) [Rodriguez92, Goodman63, Prati92, Rocca94]. Details and discussion on coherence and its applications are provided in third section of this manual. As an example, the coherence map related to the Bern interferogram in Figure 2-10 is shown in Figure 2-13.

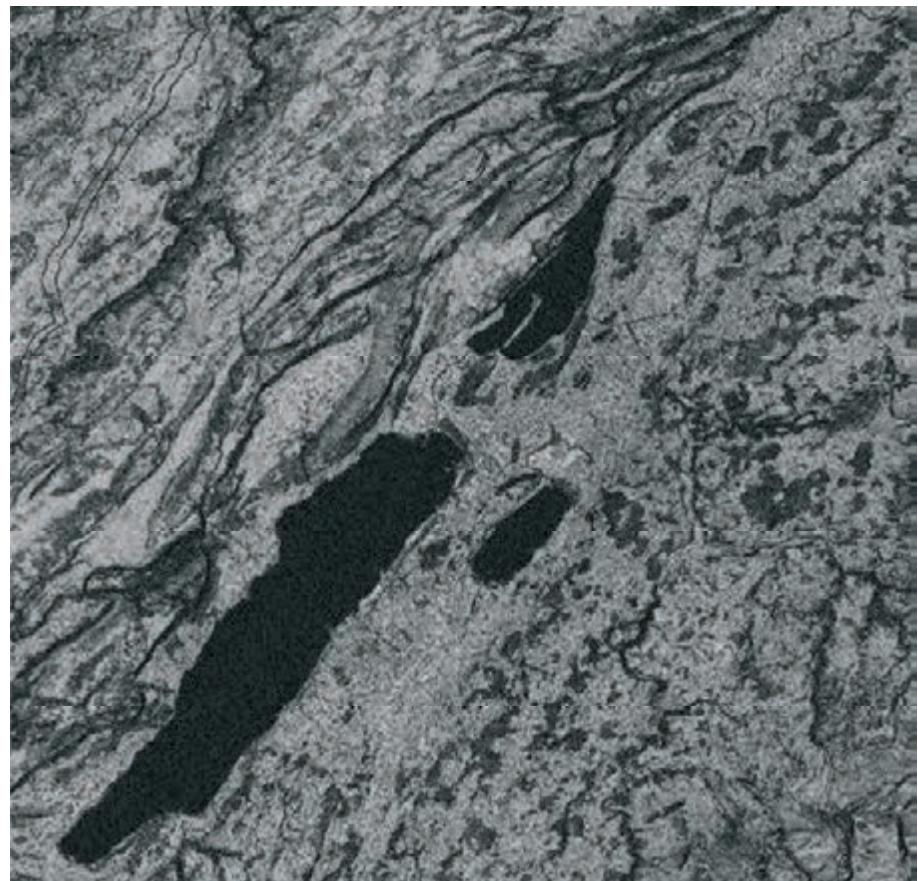


Figure 2-13: Map of the absolute value of coherence, referred to the ERS tandem interferogram of Bern in Figure 2-10 (a zoom on the central part of the frame)

The ‘sampled’ estimator is usually exploited [Prati92, Gatelli94, Rocca94]:

$$\hat{\gamma} = \frac{\sum_{n,m} u_1(n,m) \cdot u_2^*(n,m) \cdot e^{-j\phi(n,m)}}{\sqrt{\sum_{n,m} |u_1(n,m)|^2 \sum_{n,m} |u_2(n,m)|^2}} \quad \text{Equation 2.7}$$

where the summation is extended to the 2D patch of uniform speckle. The absolute value is then taken.

In Equation 2.7, $\phi(n, m)$ is the phase contribution due to topography, with the linear approximation just discussed. It is shown in [Seymour94, Touzi96] that this estimator provides the maximum likelihood (ML) estimate of coherence for homogeneous speckle. Using this estimate, pixels with weaker returns have less influence on the final estimate. The number of independent pixels generally used to estimate the coherence ranges from 16 to 40. Thus, the coherence is averaged over areas of thousands of square metres.

Note that the same expression used for the generation of coherence maps (Equation 2.7) also implements the multi-look filtering introduced in the previous section: the phase of the estimated coherence is in fact a filtered version of the original interferogram. Therefore, the same discussions on the averaging-window size and shape can be applied in this case, and an adaptive scheme for estimating coherence can be implemented by just ‘reshaping’ the rectangular box implied in the summation.

2.7 Applications of coherence

The coherence of an interferogram has an important diagnostic function (Figure 2-14, Figure 2-15 and Figure 2-16). Excluding random noise, the changes with time of the scattering properties of a target determine its coherence. For example, water bodies have low coherence because their surfaces are constantly moving; they therefore appear black in coherence images.

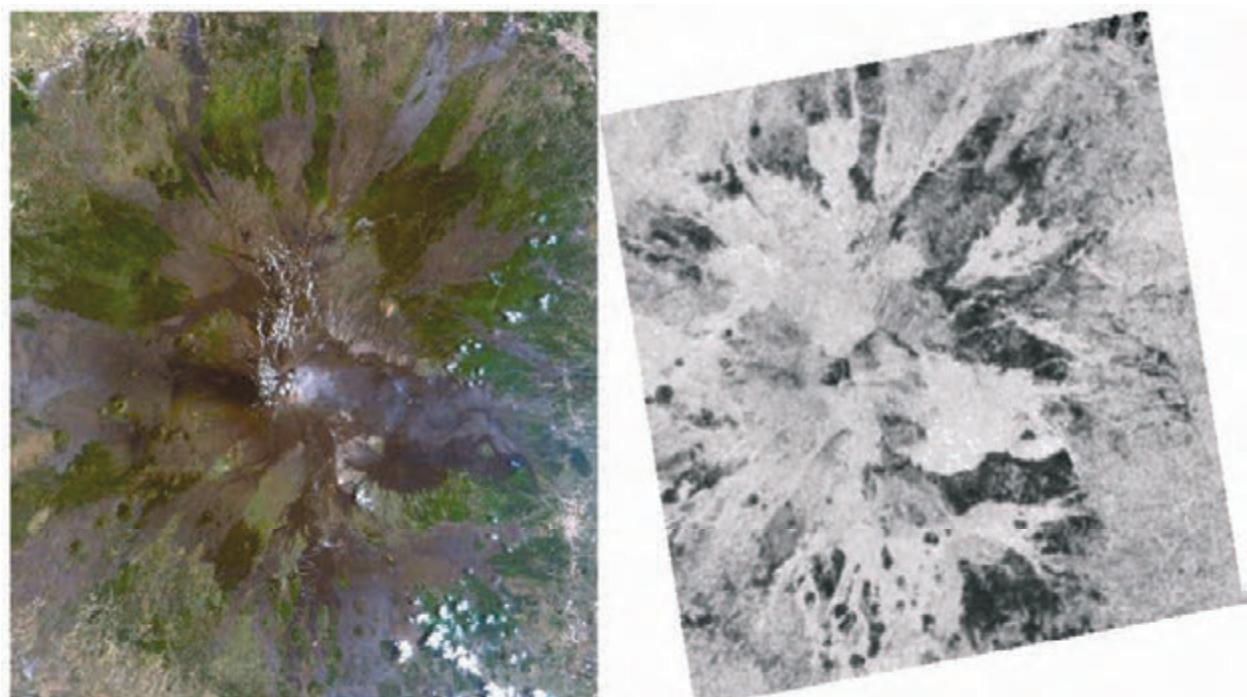


Figure 2-14: Volcano Mount Etna: LandsatTM image (left) and coherence (right)

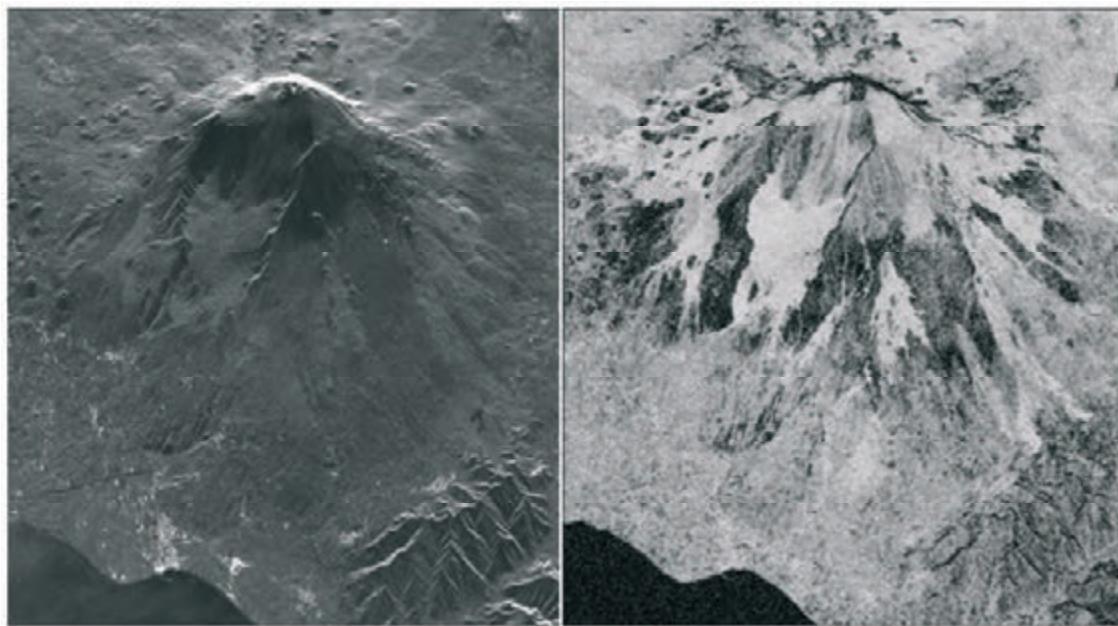


Figure 2-15: Coherence images (right) of Mount Etna together with the average amplitude image

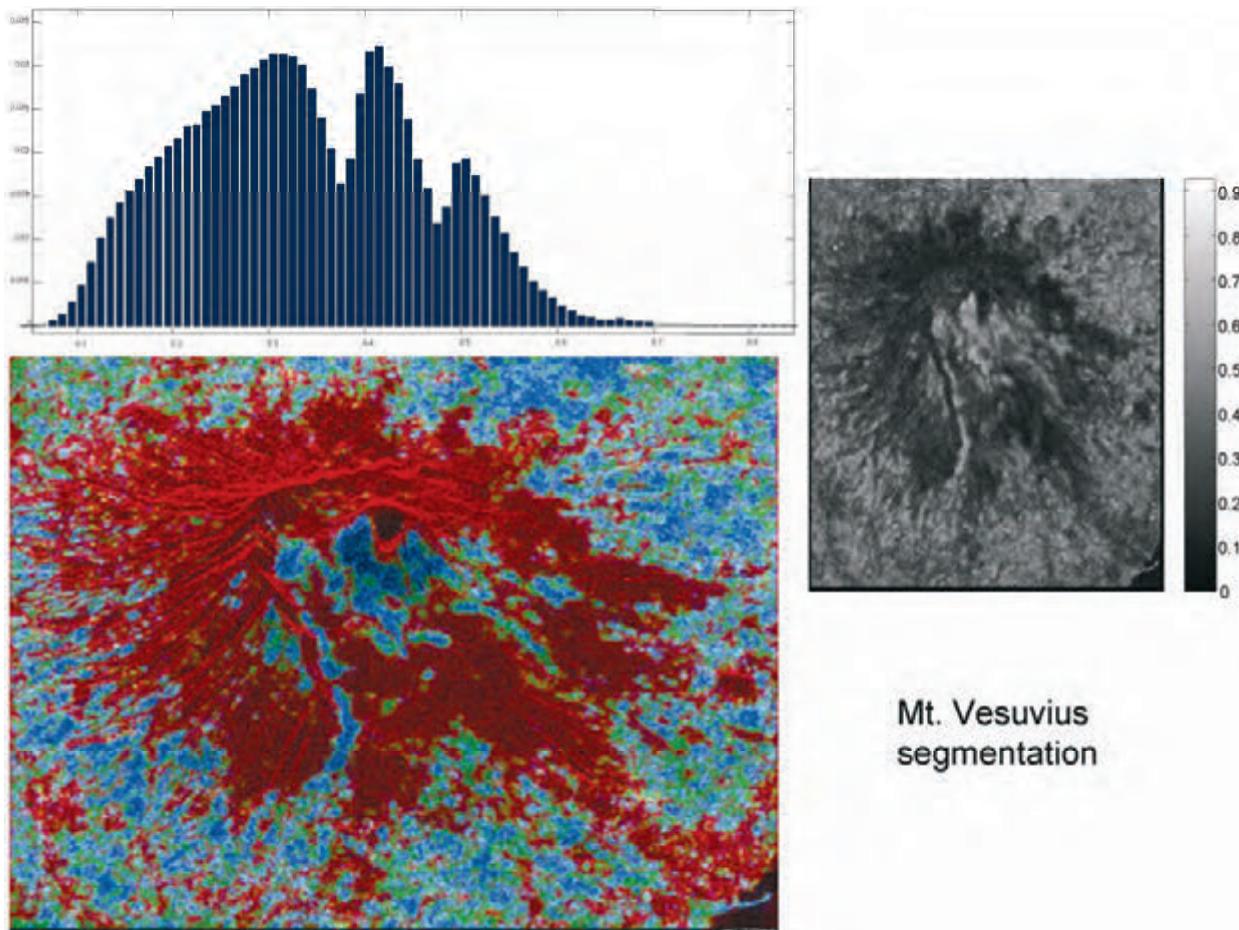


Figure 2-16: Coherence-based segmentation on Mount Vesuvius

Motion and change in vegetation also affect coherence. Leaf motion will usually cause a total loss of coherence, but this does not imply that areas of vegetation will always appear with zero coherence: radiation will often penetrate the foliage, at least partially, and can be backscattered by the terrain underneath or by the trunk and branches of the trees, which are mechanically much more stable and will therefore contribute to the coherence. In general, deciduous trees will show high coherence during winter when there are no leaves and less coherence in summer due to foliage effects. Similarly, different types of vegetation will show different one-day coherence values, depending on the height of the plant and on the lengths of the leaves: short leaves could be practically transparent to the C-band radiation of ERS satellites [Engdahl01]. Multitemporal interferometric analysis of the coherence and amplitude of the backscatter can therefore contribute to the detection and classification of forests and of vegetation in general [Koskinen01].

The joint use of coherence and the amplitude of the backscatter allows for better image segmentation. While the amplitude of the return depends on the electromagnetic structure of the target, the coherence is mostly related to its mechanical stability. For instance, in open vegetated fields, the level of the coherence is approximately linearly related to the biomass and the height of the crops [Moeremans99]. Other causes of coherence loss should be properly taken into account, for example, the force of the wind could be considered, using meteorological information [Ranson99]. Another application of coherence is forested/non-forested area segmentation, for example to find the extent of forest fires. In addition, areas of freeze and thaw in permafrost regions can be detected, and deciduous forests can be separated from coniferous ones [Schmullius99]. The penetration of radiation through dry ice can be evaluated using the volumetric effect and the change of coherence with baseline [Rott96, Weber00, Weydahl01]. In general, seasonal effects can be appreciated, using the regular series of ERS images available in selected locations [Wegmüller97]. Thus, multi-temporal techniques make it possible to identify the periodicities of the coherence that are connected to plant growth and to the visibility of the terrain in the background. They lead to segmentation techniques with results not so far from those obtainable with optical techniques in good weather [Askne97, Dammert99].

Finally, remember that if the baseline of the two acquisitions is equal to or greater than B_{cr} (the critical baseline), there is a complete loss of coherence in the case of extended scatterers. This effect has been discussed in the previous sections and intuitively shown as corresponding to the ‘celestial footprint’ of an antenna being as wide as the ground resolution cell. Moreover, unless the ‘non-cooperating’ wave number components (the useless parts of the spectrum of the signal) are filtered out, the coherence of the two acquisitions will decrease linearly with the baseline, becoming zero when it reaches B_{cr} .

2.8 Interferogram geocoding & mosaicking

Geocoding and mosaicking are the last steps in the interferogram generation chain shown in Figure 2-1.

Mosaicking is required when several interferograms (each, say, 30×100 km) are joined together to make a long strip. The need for block processing arises not only for computational efficiency, but to reduce the error due to the many approximations made so far (for example: the co-registering model, the DEM vs. SAR image alignment, the Doppler Centroid variation with azimuth etc.).

When overlapping adjacent blocks, a phase offset could arise due to small errors in image co-registering. This bias can be avoided if the image mapping is estimated over the whole strip. In some cases, the bias can be estimated, e.g. by cross-correlating the interferograms in the overlap area; however such techniques may lead to poor results in cases of low SNR.

Geocoding is performed on the mosaicked interferogram and consists of resampling it onto a uniform grid on the reference ellipsoid. In the geocoding step, one combines

- a) the range distance equation (a sphere centred in the sensor location),
- b) the Doppler equation (a plane orthogonal to sensor-target velocity, in the case of zero-Doppler focusing),

thus getting a circle in 3D space. The actual scatterer location is found by intersection with the hyperbola obtained by assuming $\Delta R(P)$ constant (e.g. the interferometric information) [Madsen93], or by exploiting the known DEM. In practice, this corresponds to the usual geocoding [Schreier93], where one has to substitute the flat Earth assumption (the ellipsoid or geoid model for Earth) by the actual interferometric information. Processing is then iterated for each point, as is done for normal geocoding.

3. InSAR DEM reconstruction

3.1 Introduction

This chapter discusses the study of topography estimation from SAR interferograms.

The first theoretical study on this subject dates back to the 1970s [Graham74], but the first applications related to a single-pass system mounted on an aircraft were published only twelve years later [Zebker86]. The feasibility of surface reconstruction by means of a repeat-pass satellite system was soon confirmed using SIR-B data [Gabriel88] and the first sensitivity analysis was then presented in 1990 [Li90]. With the launch of ERS-1 in July 1991, an ever-growing collection of interferometric data became available to many research groups. The advent of ERS-2 in April 1995 and the start of the so-called ‘Tandem Mission’ in August was regarded as a real breakthrough towards an extensive use of InSAR techniques for topography estimation, by the creation of a unique data set of high-coherence interferograms^{iv}.

While more and more InSAR DEMs were generated, the presence of atmospheric artefacts became more and more evident, and damped somewhat the enthusiasm [Massonnet95, Goldstein95, Zebker97, Hanssen98]. Research efforts were then devoted to different strategies for the combination of several Tandem pairs or the fusion of InSAR data with optical DEMs, in order to reduce the impact of the atmospheric disturbances. Results obtained using a multi-interferogram approach have recently shown how the vertical accuracy can be as high as that achievable by conventional optical satellite data (e.g. SPOT), although the amount of computational processing required is much greater than for conventional InSAR processing [Ferretti99]. Several large-scale DEM estimation projects have been set up using Tandem data [Muller99, Kooij99]. The subject has now gained popularity after the Shuttle Radar Topography Mission (SRTM), when the first single-pass radar interferometer in space flew on board the Space Shuttle in February 2000 [SRTM].

Here we focus on DEM reconstruction using ERS Tandem data (repeat-pass interferometry). However, most of the discussion holds for single-pass interferometry also.

3.2 Processing chain and data selection

The difficulties related to InSAR DEM reconstruction, and the computational burden (i.e. time and/or money) necessary to get the final result, strongly depend on the topography of the area of interest and the accuracy requirements. Although the generation of a low-resolution (and

^{iv} During the Tandem mission the orbits of the two sensors were phased to provide a 24 hour revisit interval, thus reducing temporal decorrelation.

low-quality) DEM of an area with smooth topography using a pair of SLC data sets^v can be done in a matter of minutes on a PC, high-quality topographic profile reconstruction of hilly terrain can be a very hard task. In fact, accurate DEM estimation usually requires a multi-interferogram framework (i.e. more than one interferogram related to the same area), and the processing chain can become rather complex.

Most of the processing steps for DEM generation are common to almost all interferometric applications (e.g. image focusing, re-sampling on the same acquisition geometry, common-band filtering, interferogram and coherence map generation, filtering of the fringes, etc.). Several review papers are now available on the subject and the interested reader should refer to [Gens96, Massonnet98, Bamler99, Rosen00, Franceschetti99] and references therein for a detailed analysis of the algorithms in use^{vi}. Here we will assume that images have already been focused and re-sampled on the same acquisition grid, and interferograms and coherence maps have been generated and properly filtered.

Figure 3-1 shows a simplified block diagram of the processing chain. Note that these processing steps do not need to be followed in a rigid sequence, so this is only one possible sequence.

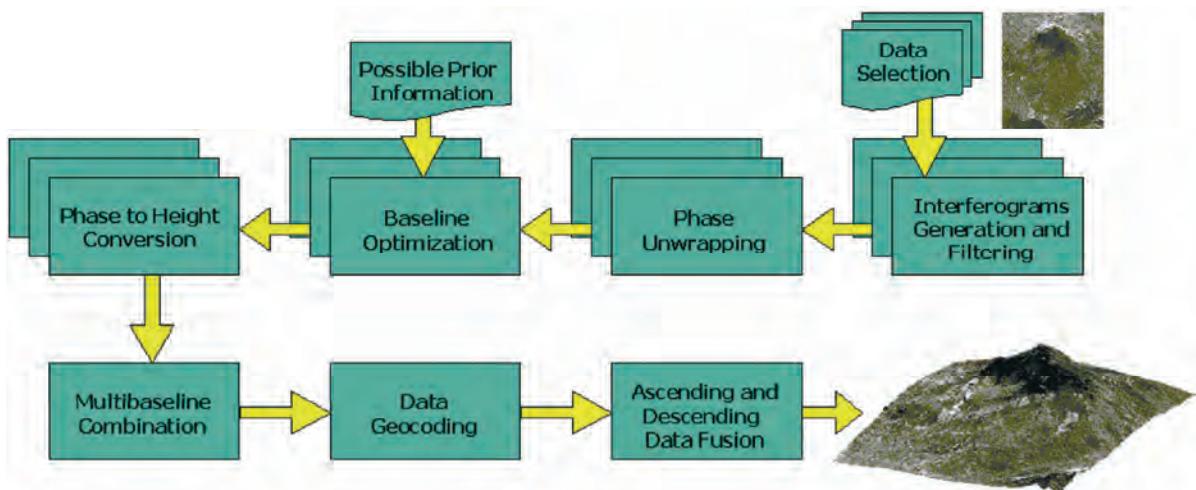


Figure 3-1: Block diagram of InSAR DEM reconstruction

Each block will be treated in some detail in the following sections. After phase unwrapping of the interferograms (section 3.3), phase values are converted to elevation (3.4) with respect to a reference ellipsoid. This step is usually preceded by a baseline optimisation procedure that uses possible Ground Control Points (GCP) to finely tune the phase-to-height conversion function. When a multi-interferogram approach is adopted, a data-fusion program (3.5) combines the DEMs obtained from the different interferograms (all relative to the same acquisition geometry, e.g. belonging

^v SLC stands for Single Look Complex. This product presents focused complex SAR data in full resolution.

^{vi} An excellent on-line searchable bibliography on SAR interferometry is available at the Dutch Interferometry Group web-site [DIG].

to the same satellite track), and possible prior DEMs, to obtain the best estimation of the local topography in SAR coordinates. After data resampling in geographic coordinates, a last processing step combining DEMs obtained from ascending and descending orbits (3.6) can be used to mitigate the problems due to the acquisition geometry and the uneven sampling of the area of interest (especially on areas of hilly terrain). In fact, the combination of Earth rotation (E–W) and satellite orbit (near polar) enables two acquisitions of the same area on each satellite cycle from two different look angles, ascending or descending^{vii}. 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. In general, for ERS data selection, the criteria listed in chapter 1 of this part of the manual should be followed.

3.3 Phase unwrapping techniques for InSAR DEM reconstruction

As already mentioned in the previous section, we will assume that interferograms have already been generated and properly filtered. The examples in this section are based on Figure 3-2.

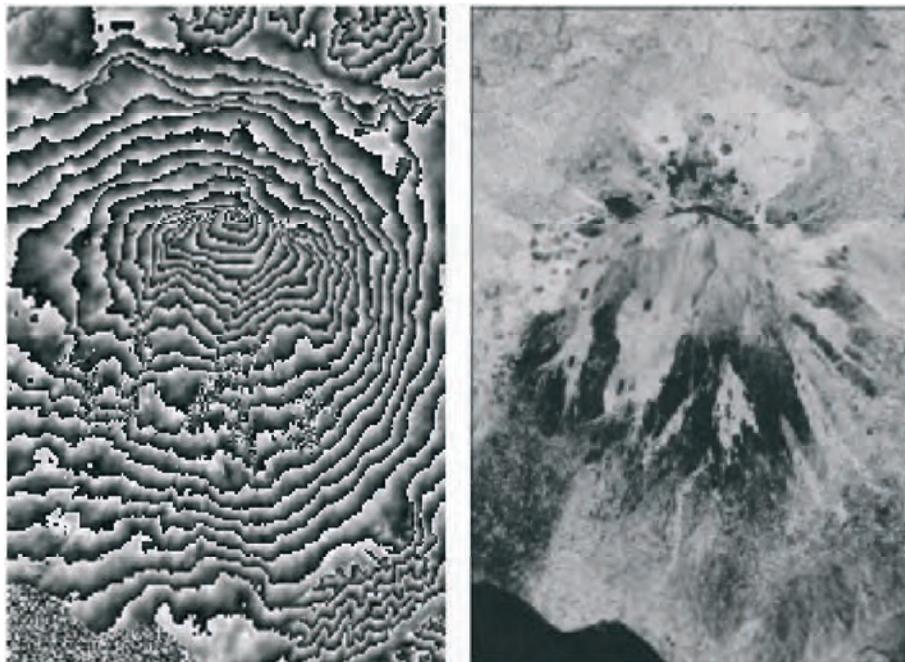


Figure 3-2: ERS Tandem data acquired over the Etna volcano (Sicily) on 1–2 August 1995 (the normal baseline was about 60 m). Left: flattened interferogram. Right: coherence map. Data have been filtered and down-sampled.

^{vii} Especially at high latitudes, satellite scenes acquired from different satellite tracks overlap, so an area of interest can be imaged by the sensor from two or more ascending tracks and two or more descending tracks. To limit the number of images to be processed, select just one track for ascending mode and one for descending mode, based on the available data-sets and their baseline distribution.

We start our analysis directly with the most important of the processing steps involved in InSAR DEM reconstruction. This a problem that will come back to our attention several times in the next chapters: phase unwrapping.

Since the interferometric phase is known only modulo- 2π and the maximum height variation in the area of interest can give rise to hundreds of cycles, an unwrapping procedure is necessary in order to estimate the local topography. If the phase contribution due to an ideally flat Earth has been properly estimated and compensated for (i.e. the interferogram has been flattened), phase unwrapping allows one to pass from the fringe pattern (similar to a set of contour lines) to a phase field proportional to the local topography. In most cases this is the major obstacle to be overcome in the processing chain for InSAR DEM reconstruction, and often cannot be performed in a totally automatic way [Ghiglia98]. The reasons for this become evident once we state the problem more precisely and analyse it from a mathematical point of view.

3.3.1 What are we looking for?

The aim of phase unwrapping (PU) is to recover the integer number of cycles n to be added to the wrapped phase ϕ so that the unambiguous phase value ψ can be finally obtained for each image pixel:

$$\psi = \phi + 2\pi \cdot n \quad \text{Equation 3.1}$$

In general, if no *a priori* information about ϕ is available, i.e. no constraint is given to the solution (e.g. maximum frequency band and signal power), phase unwrapping is an ill-posed inverse problem and therefore an infinite number of different solutions can be found, all honouring the data.

The most straightforward PU procedure would be a simple integration of the phase differences, starting from a reference point. However, because of phase discontinuities, it is not always accurate.

Almost all PU algorithms are based on the assumption that the true unwrapped phase field is ‘smooth’ and varies ‘slowly’. More precisely, neighbouring phase values are assumed to be within one-half cycle (π radians) of one another. Though this hypothesis is often valid for most of the image pixels, the presence of some phase discontinuities (i.e. absolute phase variations between neighbouring pixels of greater than π radians) causes inconsistencies, since integration yields different results depending on the path followed (Figure 3-3).

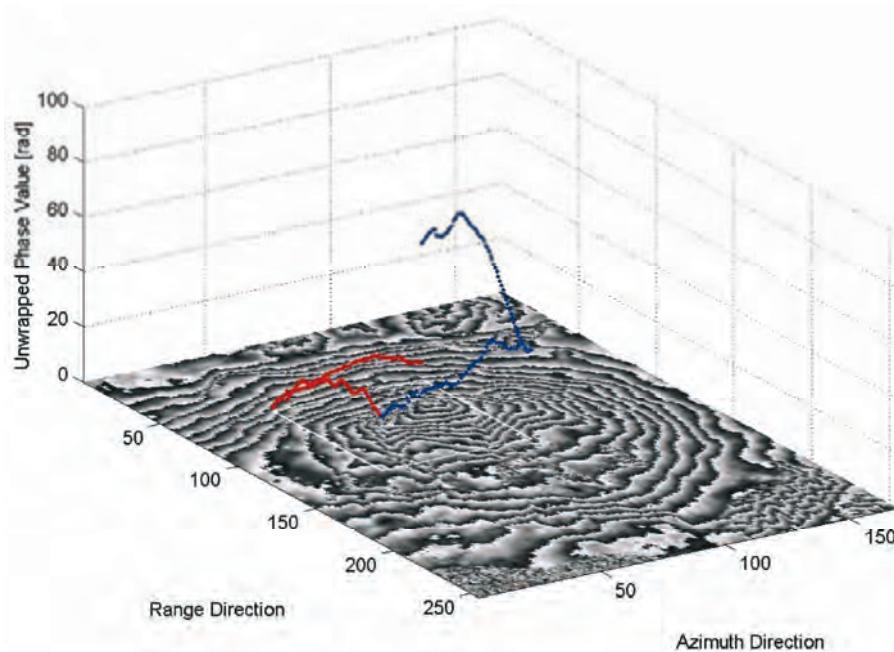


Figure 3-3: A visual example that shows the main problem related to phase inconsistencies: integration of the phase values yields different results depending on the path followed. In this example, the first and the last pixels of the integration paths are common, but one path crosses a layover area, characterised by very low coherence values, and propagates significant phase unwrapping errors.

This feature is evident whenever the sum of the wrapped phase differences (the integral of the estimated phase gradient) around a closed path differs from zero. To be consistent, a gradient field must be irrotational; i.e. the curl of $\nabla\phi$ should be zero everywhere [Spagnolini95] [Goldstein88, Ghiglia98]. Whenever this condition is verified over the whole interferogram, we have a ‘trivial PU problem’. Unfortunately, this is almost never the case in InSAR data processing.

The rotational component of the gradient field can easily be estimated by summing the wrapped phase differences around the closed paths formed by each mutually neighbouring set of four pixels. Whenever the sum is not zero, a residue is said to occur [Goldstein88]. Its value is usually normalised to one cycle and it can be either positive (+1) or negative (-1). The summation of the wrapped phase variations along an arbitrary closed path equals the algebraic sum of the residues enclosed in the path. An example of residue field (relative to the interferogram in Figure 3-2) is shown in Figure 3-4.

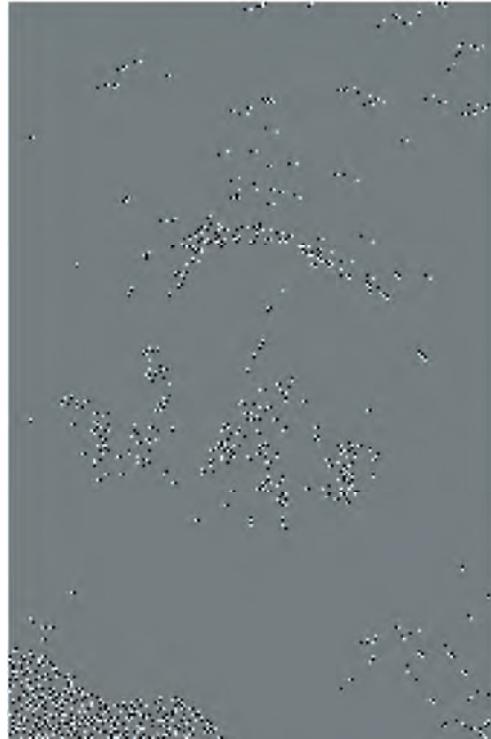


Figure 3-4: Map of the phase residues relative to the interferogram shown in Figure 3-2

Since phase residues merely mark the endpoints of the ‘discontinuity lines’, the true problem is their complete identification. Discontinuities are essentially due to two independent factors: (1) phase noise; (2) steep terrain slopes. In repeat-pass interferometry we usually deal with low SNR values (~ 0 dB) due to temporal or geometrical decorrelation, and the probability on flat terrain of a (noisy) phase variation greater than π is not at all negligible (for coherence $\gamma = 0.5$ and an effective number of looks

$$N_{\text{eff}} = 3 p(|\Delta\phi| > \pi) \approx 0.01 \quad \text{Equation 3.2}$$

On the other hand, the acquisition geometry of the sensor gives rise to an uneven sampling of the terrain in ground range (see Part A of this manual). The neighbouring pixels in the range direction can correspond to two scatterers very far apart in ground range with very different elevations. In general, the number of discontinuities will be a function of the local topography (characterised, for example, by a certain fractal dimension [Luca96]), the off-nadir angle, the normal baseline and the decorrelation noise.

In order to cope with phase discontinuities, different strategies have been followed and different algorithms have been developed. Following Ghiglia and Romero [Ghiglia98] and Chen and Zebker [Chen00], we will briefly describe them, using the ‘minimum L^p -norm’ framework. In fact almost all PU algorithms seek to minimise the following cost function:

$$C = \left\{ \sum_{i,j} w_{i,j}^{(r)} |\Delta^{(r)} \psi_{i,j} - \Delta_w^{(r)} \phi_{i,j}|^p + \sum_{i,j} w_{i,j}^{(a)} |\Delta^{(a)} \psi_{i,j} - \Delta_w^{(a)} \phi_{i,j}|^p \right\} \quad \text{Eq. 3.3}$$

where $0 \leq p \leq 2$;

Δ indicates discrete differentiation along range (r) and azimuth (a) directions respectively;

w are user-defined weights; and

the summations include all appropriate rows i and column j .

The suffix w to the differentiation operator Δ indicates that the phase differences are wrapped in the interval $-\pi$ to $+\pi$. We stress that this objective function has not been obtained from a theoretical analysis or a statistical description of a topographic phase signal in SAR coordinates. It is just a reasonable translation into mathematical terms of our basic assumption: $\Delta\psi = \Delta_w\phi$ almost everywhere. Nonetheless, it has been used for its simplicity and due to the fact that efficient algorithms are available for $p = 2$ and $p = 1$.

3.3.2 Case $p=2$, Unweighted Least Mean Squares method

Let us first analyse the unweighted least squares method (ULMS). In this case, $p = 2$ and no weight is present ($w_{ij} = 1 \forall i,j$). Equation 3.3 leads to a linear system of equations $\mathbf{A}\psi = \mathbf{b}$:

$$\begin{aligned} \psi_{i+1,j} - \psi_{i,j} &= \Delta_w^{(a)} \phi_{i,j} \\ \psi_{i,j+1} - \psi_{i,j} &= \Delta_w^{(r)} \phi_{i,j} \end{aligned} \quad \text{Equation 3.4}$$

to be solved with some boundary conditions. The equation system is patently over-constrained, we have roughly two gradient estimates for each phase, therefore it can be formulated in a normal equation form: $\mathbf{G}^T \mathbf{G} \psi = \mathbf{G}^T \mathbf{d}$.

Data vector \mathbf{d} is just the vectorised form of the wrapped phase differences estimated from the interferogram, while the model matrix \mathbf{G} is an incidence matrix whose non-zero elements assume the value $+1$ or -1 . This is the typical matrix encountered in geodesy for levelling networks [Strang97], and again efficient numerical solutions are well known. In this case the boundary conditions change and the network is solved with respect to (at least) one pixel of known elevation.

The drawbacks to this approach can be easily envisaged by examining the rationale behind it. *We don't care about the number and the position of the discontinuities.* We simply write the equations, we hope that only a few of them are wrong and we take advantage of ready-for-use solution packages to get the result. Unfortunately, simplicity and accuracy rarely go together. Unweighted LMS solutions are prone to severe errors caused by phase discontinuities. Each wrong equation gives rise to phase artefacts around it, so error propagation problems are by no means overcome. Moreover, the

solution is congruent with the data only in the trivial case ($\psi = \phi + 2\pi n$), where no discontinuities are present.^{viii}

3.3.3 Case $p=2$, Weighted Least Mean Squares method

The quality of the results is somewhat improved by weighting the equations. In fact, the coherence map associated with the interferogram and/or the amplitude images can be used successfully to identify areas where phase discontinuities are likely to occur [Ghiglia98], but in doing so we lose the regular structure of the matrix to be inverted. Efficient iterative numerical techniques can be adopted, but they lead to significant increases in computational time. Furthermore, for correct phase reconstruction, zero weights should be applied to phase discontinuities and unitary weights whenever the phase gradient is correct, but this would imply identification of the ‘cycle skips’, which is the problem that LMS methods would like to avoid.

Phase artefacts in the estimated unwrapped field due to noisy data can be partially avoided if local phase gradients are estimated on larger windows [Spagnolini95], using more reliable 2-D frequency estimation techniques (e.g. FFT analysis). The problem is basically a variational surface reconstruction problem from indirect measurements, well known in the field of computer vision. This approach is sometimes referred to as generalised LMS PU. Of course its use requires a not-easy trade-off between estimation accuracy and resolution. Interesting results have been obtained using a fast hierarchical implementation of a multi-resolution estimator [Davidson99].

Both LMS and WLMS results do not honour the original interferogram ($\psi \neq \phi + 2\pi n$), apart from the trivial case. Of course, the solution can be forced to be congruent with the data [Pritt97] (rounding the difference between the wrapped and unwrapped phase to the nearest integer number of cycles) but this is not equivalent to minimising Equation 3.3 with the ‘integer constraint’. Actually, this is an NP-hard problem^{ix} [Arora97] (like the famous travelling salesman problem) known as the ‘nearest lattice vector problem’ [Strang97], which complexity theory suggests is impossible for efficient algorithms to solve exactly.

3.3.4 Case $p=1$, Minimum Cost Flow method

An approach to PU based on network programming [Costantini98, Flynn97] has received a great deal of attention, since it provides an efficient tool for a global minimisation of Eq. 3.3 under the (weighted) L^1 -norm (minimum absolute deviation). PU is formulated as a constrained optimisation problem. The algorithm minimises the integer number of cycles to be added to the phase variations (i.e. the data $\Delta_w \phi$) to make them consistent. As already

^{viii} For an interesting relationship between unweighted LMS and 1-D integration the reader should refer to [Fornaro96] (see also Section C2 of this manual).

^{ix} The computation time of a problem is a function of its size (i.e. the number of variables). An NP-hard problem is one for which nobody has ever found an algorithm to solve it in a polynomial time.

mentioned, to be consistent a gradient field must be irrotational; only in this case is the unwrapped phase field independent of the integration path, up to an additive constant. The constraint to be satisfied is then $\nabla \times \nabla \phi = 0$, not that of an integer solution. This welcome property is just a consequence of the equations used [Costantini98].

In the Minimum Cost Flow (MCF) approach, the PU problem is equated to a general network flow problem [Costantini98]. This reformulation of the PU problem allows the use of powerful techniques developed for network optimisation [Ahuja93]. Since graph theory and network programming is a mature subject of operational research, several fast optimisation routines can be employed to seek the minimum cost flow. Moreover, many source codes are available on the web [SNAPHU]. The details of the algorithms can be found in [Ahuja93].

As for WLMS estimation, the user can define proper weights marking areas where phase discontinuities are more likely to occur. Even in this case, since flow magnitudes are restricted to being integers, the final unwrapped phase field is congruent with the original interferogram ($W[\phi] = \psi$); MCF algorithms are truly PU programs, since they merely add an integer number of cycles to every wrapped phase value.

3.3.5 Case $p=0$, Branch-Cut and other minimum L^0 methods

When $p = 0$, Equation 3.3 equals the number of samples for which the solution gradients do not exactly equal the measured gradients. Thus the gradients of the estimated unwrapped phase field exactly equal the data in as many places as possible (Figure 3-5). This is one of the frequently suggested goals for PU [Goldstein88, Ghiglia98, Chen00] and historically the first one used in InSAR data processing [Goldstein88].

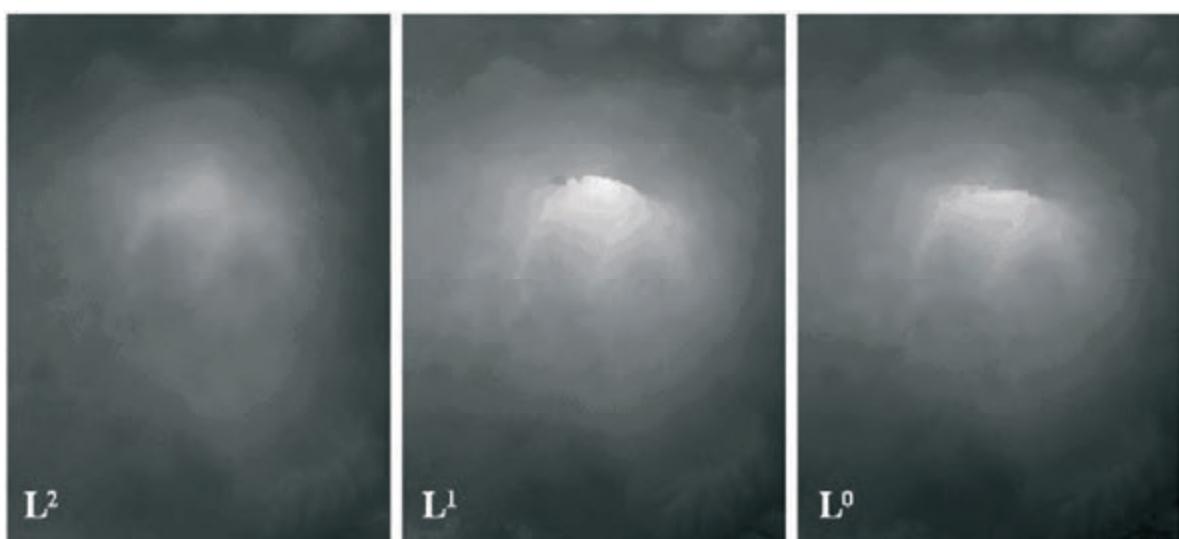


Figure 3-5: Three unwrapped phase values obtained by unwrapping the Etna interferogram shown in Figure 3-2 using the three different cost functions described in the text. The ‘smoothing effect’ typical of the ULMS approach can be seen.

The basic idea of the branch-cut method [Goldstein88] is to unwrap the interferogram selecting only paths of integration that lead to self-consistent solutions. The rotational part of the estimated phase gradient (i.e. the residue field) is used as a map to drive the integration along a consistent path, limiting the error propagation. As already mentioned, the summation of the wrapped phase variations along an arbitrary closed path equals the algebraic sum of the residues enclosed in the path. Paths of integration that encircle a net ‘charge’ must be avoided. This is accomplished by connecting oppositely-charged residues with branch cuts the integration paths cannot cross. Usually cuts are straight lines and are grown in a treelike manner as in a region-growing algorithm. Residues mark the endpoints of the ‘discontinuity lines’ (often referred to as ‘ghost lines’ [Prati90A] or ‘aliasing lines’ [Spagnolini95]). The strategy adopted in order to identify the minimum number of branch cuts distinguishes the different members of this class of algorithms. One of the major drawbacks of the method is that cuts can close on themselves in areas of low coherence, so that holes can exist in the unwrapped phase field and, in general, poor results are obtained in areas of low SNR.

A different approach has been proposed by Ghiglia and Romero [Ghiglia96]. It is an iterative scheme to the general L^p -norm solution and allows complete coverage. The governing equations are equivalent to those that describe the WLMS PU algorithm, but weights are data dependent. Convergence to a global minimum is not guaranteed (this should be expected given the highly non-linear nature of the cost function) and the algorithm is computationally intensive, though the application of the method is easy once a WLMS program is available.

In [Chen00], Chen and Zebker have demonstrated that the minimum L^0 -norm problem is NP-hard [Garey79] (roughly speaking, it would take too long to find the global minimum of the cost function) and research efforts should be focused on designing approximate algorithms. On the other hand, we recall here that the L^0 -norm criteria has not been obtained by means of a rigorous statistical analysis of the PU problem for InSAR DEM reconstruction; instead it is just an attempt to minimise the number of discontinuities in the estimated surface (Figure 3-6).



Figure 3-6: Map of the ‘cuts’ identified by a ‘branch and cut’ algorithm (L^0 -norm)

3.3.6 Outlook

The preceding analysis is by no means complete. Region growing techniques [Xu99], Kalman filtering [Kramer96] and model-based PU algorithms [Friedlander96], just to mention three other interesting approaches to PU, have been proposed as possible solutions to this problem. The multi-baseline PU algorithm [Ferretti97] is also a valuable tool when more interferograms of the area of interest are available (it will be discussed in Section C4.1).

At this point in time it is not possible to give a classification or suggestion as to the best PU algorithms. While further analyses on PU will be presented in Section C3, this is still an active research field, firstly because of the conviction that there is still room for new algorithms and improvements, and secondly because PU is fascinating and has interesting links with many other image-processing problems. We partially agree with the common opinion that PU can now be overcome by a “well-stocked arsenal of PU methods” [Ghiglia98, Chen00]. Probably a topography-based, statistical analysis of the signal of interest, together with a proper combination of all the available information (e.g. low-resolution DEMs, amplitude images, multi-pass combinations), will reduce the number of algorithms effectively used to a few and there will finally be a ‘standard PU procedure’ for InSAR DEM reconstruction.

3.4 From phase to elevation

After phase unwrapping, it is possible to localise every pixel in the image with respect to a Cartesian reference system: from radar coordinates (range, azimuth, phase variation) we can pass to a standard description of the local topography. To this end, it is necessary to know the acquisition geometry for both master and slave images, i.e. the satellite orbit positions. After point localisation, it is easy to compute the elevation with respect to a reference ellipsoid and, in general, to transform the coordinates into whichever reference system is required. The last processing step (called **data geocoding**) is unavoidable in order to compare the results with possible reference DEMs of the area of interest (e.g. to assess the quality of the estimated topographic profile) and to generate a product in a ‘standard format’.

Apart from phase noise and possible unwrapping errors, the accuracy of this processing step depends on the precision of the satellite state vectors used to model the satellite trajectories. Lack of precise orbits can be compensated by the availability of Ground Control Points (GCP) at known coordinates, though their identification is an operator-dependent processing step. To date, the accuracy of ERS orbits does not allow large-scale DEM generation without any tie point and all methods for phase-to-height conversion require at least one GCP.

Phase-to-height conversion methods can be divided into two groups [Small96]: those that operate on the flattened phase, and those based on the unflattened interferogram. In fact, even though the most common unwrapping algorithms operate on flattened data, flattening phase terms are deterministic and known, and can easily be added back to the data once they have been unwrapped. In general, the most accurate methods belong to this second category and we will focus on them in the following.

This section is organised as follows: After a brief discussion on the common polynomial approximation of a satellite trajectory, we present the system of equations that must be solved to perform data geocoding and DEM reconstruction. We then present a simplified sensitivity analysis of the impact of baseline errors on the estimated topography, and a brief overview of the major sources of precise orbits for ERS satellites.

3.4.1 Polynomial approximation of satellite orbits, point localisation and data geocoding

SAR data are usually delivered to users in frames corresponding to an area of about $100 \times 100 \text{ km}^2$. Even considering full-frame processing, a satellite trajectory can be well approximated by a low-order polynomial function. More precisely, for a third-order fitting, satellite position is defined by the following equation set:

$$\mathbf{S} = \mathbf{a} \cdot t^3 + \mathbf{b} \cdot t^2 + \mathbf{c} \cdot t + \mathbf{d} \quad \text{Equation 3.5}$$

where $\mathbf{S} = \{S_x, S_y, S_z\}$ are the satellite Cartesian coordinates with respect to a reference frame

t is the azimuth time
vectors $\{\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}\}$ can be obtained from the available state vectors by means of an LMS estimation.

The reference frame is usually an Earth-Centred Rotating (ECR) frame [Small98], and the target is assumed motionless. In particular, state vectors are usually delivered in both inertial and ECR reference frames, and the following discussion is valid in both cases. We start from the problem of data geocoding when target elevation is known *a priori*. We then extend the procedure to interferometric data.

If the local topographic profile is known, target coordinates are determined by simultaneous solution of three equations [Curlander82, Schreier93]:

- 1) Range equation;
- 2) Doppler equation; and
- 3) Earth model equation.

In fact, each pixel $P = P\{r, t\}$ of the image can be identified by its azimuth time t and its range coordinate r . Considering a zero-Doppler focusing algorithm [Curlander91], the Cartesian coordinates $P = \{X, Y, Z\}$ of the target must satisfy the following equations:

$$|\mathbf{P} - \mathbf{S}(t)| = r = r_{ca} + R_s \cdot n_r \quad \text{Equation 3.6}$$

$$(\mathbf{P} - \mathbf{S}(t)) \cdot \mathbf{V}_{PS}(t) = 0 \quad \text{Equation 3.7}$$

where r_{ca} is the distance between the sensor and the first sample of the range line,
 R_s is the range step,
 n_r identifies the sample of the range line we are working on,
 V_{PS} is the sensor-target relative velocity.

Equation 3.6, the range equation, describes a sphere of radius r centred in $S(t)$. Equation 3.7, the Doppler equation, identifies a plane orthogonal to V_{PS} . In order to localise P , another equation is needed. If the height of the target with respect to a reference ellipsoid (h) is known, it is possible to solve the system for $P = \{X, Y, Z\}$, using an iterative numerical technique (the system is non-linear) [Curlander82, Schreier93].

The third equation, the Earth model equation, is the following:

$$\frac{X^2 + Y^2}{(R_e + h)^2} + \frac{Z^2}{R_p^2} = 1 \quad \text{Equation 3.8}$$

where R_e is the equatorial radius of the Earth, and
 R_p is given by:

$$R_p = (1 - f)(R_e + h) \quad \text{Equation 3.9}$$

where f is the flattening factor of the reference ellipsoid (e.g. for the WGS84 ellipsoid, $f = 1/298.257223563$)
 $R_e = 6378137.0$ m

As pointed out in [Curlander91], the accuracy of this location procedure does not require attitude sensor information (though it depends on the accuracy of the sensor position and velocity vectors). Thus a SAR pixel location is inherently more accurate than that of optical sensors, which are strongly dependent on attitude parameters at the time of the acquisition.

In InSAR topographic applications, h is unknown and Equation 3.8 cannot be used; nonetheless target coordinates can be estimated using the interferometric phase ϕ and the state vectors relative to the second radar acquisition. More precisely, for the so-called ‘slave’ sensor, the following equations hold:

$$\begin{aligned} (\mathbf{P} - \mathbf{S}_{\text{slave}}(t_{\text{slave}})) \cdot \mathbf{V}_{\text{PSslave}}(t_{\text{slave}}) &= 0 \\ |\mathbf{P} - \mathbf{S}_{\text{slave}}(t_{\text{slave}})| &= r + \frac{\lambda}{4\pi}\phi \end{aligned} \quad \text{Equation 3.10}$$

In general, $t_{\text{slave}} \neq t = t_{\text{master}}$. Thus a system of four equations and four unknowns ($X, Y, Z, t_{\text{slave}}$) must be solved. For ERS Tandem data, the satellite orbits are almost parallel (typical cross-angle value is ~ 1 millidegree [Small98]) and we can usually neglect the Doppler equation of the slave acquisition. The following approximations are then adopted:

$$(S_{\text{slave}}(t_{\text{slave}}) - S(t)) V_{\text{PS}}(t) = 0 \quad \text{Equation 3.11}$$

(i.e. the zero-Doppler plane is the same for both the acquisitions), and

$$S_{\text{slave}}(t_{\text{slave}}) \approx S(t) + \mathbf{A} \cdot t + \mathbf{B} \quad \text{Equation 3.12}$$

where \mathbf{A} and \mathbf{B} are suitable constant vectors.

In other words, the baseline between the two orbits is approximated by a linear function of the azimuth time. This strategy allows a faster solution of the system with no significant loss in location accuracy. Of course, orbit indetermination impacts on target coordinate estimation, since vectors \mathbf{A} and \mathbf{B} (and hence the baseline of the interferometer) depend on the state vectors of both the master and the slave acquisition.

Some observations are now in order:

- As already mentioned, the accuracy of the satellite state vectors does not allow precise data geocoding without any GCPs, and the processing is calibrated using at least one reference pixel of known coordinates.
- Even if the user is merely interested in the height of the targets with respect to a reference ellipsoid (e.g. to generate a DEM in SAR coordinates), the computational burden is basically the same, since, in any case, it is necessary to solve the quoted non-linear system of equations. In fact, only after the computation of the Cartesian coordinates of the target is it possible to compute h . Once the coordinates (X, Y, Z) have been identified, it is then possible to transform them into different reference frames (e.g. geographical or UTM coordinates with respect to a local ellipsoid [Schreier93]).
- It should be pointed out that possible phase unwrapping errors, phase noise and atmospheric effects (to be discussed in the next sections) impact not only on the estimated target elevation but on its geo-

referencing as well. This can make a comparison with prior information in rough topography more difficult.

3.4.2 Data resampling

In general, in order to compare the estimated topographic profile with a reference DEM of the area and to obtain a standard product that can be delivered to possible customers, a last step is required. The uniform two-dimensional image grid (range and azimuth coordinates) gives rise to a non-uniform sampling of the geographical coordinates (latitude and longitude or Northern and Eastern) since they depend on the local topography (remember that the *ground* range coordinate is not uniformly sampled for non-flat areas). An interpolation is then necessary to generate a standard raster file with a constant sampling step. The choice of the best interpolator is beyond the scope of this manual.

In general, four solutions can be considered [Wackernagel98]:

- 1) Nearest Neighbour (NN);
- 2) Delaunay Triangulation and linear interpolation (DT);
- 3) Inverse Distance Weighting (IDW);
- 4) Kriging Interpolation (KI).

Whilst the NN technique is by far the most simple computationally, KI strongly reduces DEM artefacts in areas of steep slopes, at the cost of a much more computation-intensive processing. In most applications, DT or IDW can be a convenient compromise between accuracy and computational time. In any case, it is important to realise that, in areas of rough topography, only a combination of both ascending and descending passes can produce a reliable DEM, as will be discussed in Section C4.

3.4.3 Impact of baseline errors on the estimated topography

For the sake of simplicity, the following sensitivity analysis will be based on expressions that are valid only locally. Nevertheless, it turns out to be enough for our purposes. More complete analyses are available in several papers [Dixon94, Rosen00].

Using the linear (far-field) approximation (Section A2), the phase variation between two neighbouring pixels of the (unflattened) interferogram is related to their elevation difference q by a very simple expression that can be easily solved for q :

$$q = -\cos \theta \cdot \Delta R + \frac{\lambda R \sin \theta}{4\pi} \frac{1}{B_n} \Delta \phi \quad \text{Equation 3.13}$$

where $\Delta \phi$ is the unwrapped phase difference. This linear expression may be sufficient when the area of interest is small (say 2×2 km wide), i.e. when the orbits relative to the master and the slave acquisitions can be considered parallel (the baseline is not dependent on the azimuth coordinate).

Differentiating with respect to B_n , we can evaluate the impact of a small baseline error (ε_B):

$$\begin{aligned}\Delta q &= -\frac{\lambda R \sin \theta}{4\pi} \frac{\varepsilon_B}{B_n^2} \Delta \phi \\ &= -\frac{\lambda R \sin \theta}{4\pi} \frac{\varepsilon_B}{B_n^2} (\Delta \phi_{flat} + \Delta \phi_q) = \Delta q_{flat} + \Delta q_q\end{aligned}\quad \text{Eq. 3.14}$$

where the phase variation has been divided, as usual, into two contributions: a flat Earth term ($\Delta \phi_{flat}$) and topographic phase ($\Delta \phi_q$). The first one can be well approximated by a linear phase term (or by a low order polynomial for larger areas), while the second one is proportional to the local topography.

This simple analysis allows us to highlight two different kinds of distortion due to baseline errors:

- 1) An additive, low-order polynomial superimposed on the result (Δq_{flat})
- 2) An error modulated by the local topographic profile (Δq_q)

The first contribution is usually large (at least for interferograms with baseline values greater than 50 m) and that is the reason why DEM errors due to orbit inaccuracies are often compensated for simply by adding a suitable low-order polynomial to the topography. This can be easily estimated if some *a priori* information is available (e.g. a low-resolution reference DEM of the area under study).

The second term in Eq. 3.14 is not always negligible. In fact from equations A.2.7 and B.3.14 we obtain:

$$\Delta q_q = -q \cdot \frac{\varepsilon_B}{B_n} \Rightarrow \left| \frac{\Delta q_q}{q} \right| = \left| \frac{\varepsilon_B}{B_n} \right| \quad \text{Equation 3.15}$$

Thus the *relative* topographic error equals the *relative* normal baseline error (at least to a first order approximation). For $B_n = 100$ m, $\varepsilon_B = 1$ m and a maximum height variation of 1000 m the maximum error turns out to be ~10 m. Therefore, low baseline interferograms, which are easier to unwrap, are more prone to distortions due to orbit inaccuracies.

The results of the previous analysis can easily be generalised, at least to get a first insight, to a large scale DEM reconstruction case considering the normal baseline value (and its possible error) as a function of the pixel coordinates (range, azimuth). In general, orbit inaccuracies give rise to low-order polynomial distortions as well as an error term that is dependent on the local topography: the impact of both contributions on the estimated topography is a function of relative baseline error that depends on the state vector accuracies. Whenever tie points of known coordinates can be identified in the area of interest, it is possible to correct the baseline parameters by means of a non-linear optimisation [Werner93]. In general, the higher the desired accuracy, the higher the number of GCPs needed.

As previously discussed, apart from DEM errors, orbit inaccuracies can compromise data geocoding too, since the geographic coordinates of each

scatterer depend on the acquisition geometry. In order to mitigate this kind of problem, accurate state vectors should be used. In the following section we will discuss different sources of orbital data for the ERS satellites.

3.4.4 Precise orbit determination

Precise orbital data form a crucial element in InSAR data processing [Closa98, Reigber96, Kohlhase99]. Orbital uncertainties impact not only in DEM reconstruction and data geocoding (the geographic coordinates of each scatterer depend on the acquisition geometry) but also in differential applications (DInSAR), where small surface displacements should be detected and monitored. In fact, compensation for the topographic phase contribution cannot be carried out correctly using low quality **ephemerides** (satellite position and velocity vectors), and spurious fringes can be misinterpreted as surface displacements, especially when the local topography presents considerable height variations.

For the ERS satellites, different sources of orbit data are available, with different quality levels. In general, state vector accuracy depends on the data available about sensor position and velocity at different epochs (e.g. Satellite Radar Ranging – SLR; Precise Range and Range-Rate Experiment – PRARE; Radar Altimeter – RA [Massmann97]), and the processing used to get the estimation (in particular the mathematical model describing its motion). Since the ERS-1 launch in July 1991, state vector accuracy has been strongly improved. Nevertheless, precise orbit products are usually not available until several weeks after the satellite pass over the area of interest, due to all the processing steps involved in the estimation. This can be an obstacle for routine monitoring of seismic or volcanic areas by means of DInSAR techniques [Reigber96].

In general, the ERS-1 and ERS-2 operational orbit determination is performed by the Flight Dynamics Division at the ESA European Space Operations Centre (ESOC) in Darmstadt [ESOC]. The purpose of these data is to provide the ERS ground segment with the latest orbit determination and prediction, for satellite data acquisition, mission planning and fast delivery data processing purposes. Precise ESOC orbit products are also available, with a delay of typically one week necessary to collect most of the laser tracking.

More accurate state vectors are made available several months after the satellite acquisition. They are generated by two different groups: the German Processing and Archiving Facility (D-PAF) and Delft Institute for Earth-Oriented Space Research (DEOS). The main difference is the gravity field model adopted for orbit propagation: PGM055 for D-PAF and DGM-E04 for DEOS. ERS precise orbits provided by DEOS are believed to have a radial precision of 5–6 cm [Scharroo97], while the D-PAF precise orbits have an accuracy (derived from internal quality checks) of about 7 cm [DPAF]. Across- and along-track accuracy is lower (about 20–30 cm) and, in the end, the impact of possible baseline errors cannot be considered negligible in interferometric applications. Of course, the higher the accuracy the lower the number of GCPs requested for baseline optimisation (the so-called ‘orbital tuning’).

Both D-PAF and DEOS provide several orbit products with different accuracy levels and different delays from the ERS acquisition. Usually precise orbital products are available only after several months and consist of the satellite ephemeris (position and velocity vectors) with a certain time resolution (every 30 s for D-PAF and 60 s for DEOS) and other ancillary information. Though satellite state vectors are delivered with a lower time resolution, the DEOS geodesy group also provides the software for orbit propagation, starting from the state vectors available [DEOS]. Files containing dates and duration of the satellite manoeuvres (useful to estimate the reliability of the estimated satellite positions) are also available at the DEOS website. Both inertial and ECR data on state vectors estimated with respect to an inertial reference frame can be found at D-PAF. For further information about precise orbit products the reader should refer to [DPAF] and [DEOS] where available data are very well documented.

Note that for Envisat, more accurate state vectors are available as a standard product with no time delay.

3.5 Error sources, multi-baseline strategies and data fusion

Apart from the phase unwrapping problem, InSAR DEM accuracy depends on several diverse factors, including:

- 1) phase noise ϕ_w ,
- 2) atmospheric effects ϕ_a ,
- 3) orbit indetermination (baseline errors)

As already discussed in the previous section, baseline errors are systematic and can be strongly reduced by means of an optimisation procedure using a few Ground Control Points (GCPs) in a given image scene [Werner93]. Here we focus on the first two error sources. Again, for the sake of simplicity, we use the linear approximation for phase-to-height conversion, we neglect possible baseline errors, and we assume that the interferometric phase has been compensated for the flat-Earth term (i.e. the interferogram has been flattened). The estimated topographic profile is then given by the following expression:

$$q = \frac{K}{B_n} (\phi_t + \phi_w + \phi_a) = t + w + a \quad \text{Equation 3.16}$$

where K is a constant (for small areas)
 ϕ is the topographic phase contribution
 t is the local topographic profile
 a is the elevation noise due to the atmosphere
 w is the elevation noise due to phase decorrelation

A brief description of these two error contributions follows.

The phase noise term results from various factors [Zebker92] including thermal noise, image misregistration, processing artefacts, temporal and baseline decorrelation. All these noise sources increase the dispersion of the

interferometric phase value ϕ and thus the DEM. The noise power can be estimated using the absolute value $|\gamma|$ of the local coherence. This is computed from the data using a space average around each pixel in the image, assuming the process to be ergodic and stationary inside a small estimation window.

Phase distortion due to atmospheric effects (i.e. refractive index variations in the propagation medium) has gained increasing attention [Massonnet95, Goldstein95, Zebker97, Hanssen98, Ferretti99], since it can seriously compromise InSAR DEM quality, especially for those pass pairs with low normal baseline values. These effects are mainly due to the time and space variations of atmospheric water vapour and exhibit power law energy spectra (Figure 3-7). The corresponding correlation length extends well beyond the window dimensions used for coherence estimation, so the final topography can show strong distortions in spite of high coherence values.

The effects of phase noise and atmospheric artefacts are reduced if *high baseline* interferograms are used for DEM reconstruction (the same phase dispersion can give rise to very different elevation dispersions if different baseline values are used: see Equation 3.16). Unfortunately, high baseline interferograms have many tightly packed fringes and are usually very noisy and difficult to unwrap: the smaller the altitude of ambiguity (i.e. the height variation corresponding to one cycle of phase variation) the greater the probability of phase aliasing and the more difficult the unwrapping.

However, when more than one interferogram is available, we can better estimate the local topography, combining the DEMs obtained from each image pair. A *weighted* combination of several topographic profiles is recommended whenever possible, since it can strongly reduce the impact of phase artefacts on the final DEM. The key issue is proper weight selection in order to give a positive bias to the most reliable interferograms. In fact, weighting factors should take into account:

- 1) the baseline value,
- 2) the local coherence (phase noise power), and
- 3) the atmospheric disturbance power.

Unfortunately, the last term cannot be easily estimated from a single interferogram, since its contribution cannot be separated by the (unknown) topographic phase signal^x.

The following section is concerned with this latter issue. When more than three independent DEMs are available, under easy-to-meet assumptions it is possible to estimate, directly from the data, both the atmospheric distortion power and the decorrelation noise power for each datum. It is then possible to properly combine the DEMs by means of a *weighted* average. The resulting DEM is more reliable, since the uncorrelated atmospheric and noise phase contributions coming from single interferograms are averaged, thus reducing the elevation error dispersion.

^x Atmospheric disturbances and topographic signals exhibit a very similar spectral behaviour. Both signals have a power law energy spectrum.

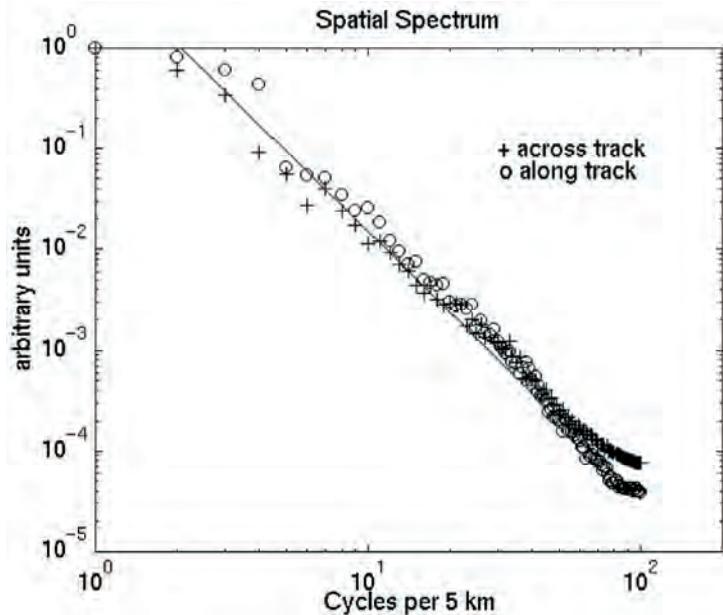


Figure 3-7: Power spectrum of the phase distortion due to atmospheric inhomogeneities (ERS Tandem – August 1995). The added line follows the slope to be expected from turbulence phenomena (from [Ferretti99]).

3.5.1 Multi-Interferogram InSAR DEM reconstruction

From 16 August 1995 until mid-May 1996, the ESA ERS-1 and ERS-2 satellites operated in Tandem mode (i.e. the orbits of the two sensors were phased to provide a 24-hour revisit interval). This configuration allowed the collection of more than 100 000 interferometric SAR pairs, acquired all over the world (Figure 3-8). While over South America and parts of South-East Asia just one tandem pass was made, more than three interferometric passes are often available for Europe and North America. These data can be used to increase the reliability of the estimated topography. The increased complexity of the algorithms involved in the processing chain required to process several data pairs is repaid by the quality of the results. Only the basic idea of the approach is presented here. For a thorough discussion, refer to [Ferretti99].



Figure 3-8: The ESA ERS-1 and ERS-2 Tandem mode

Let us suppose, temporarily, that atmospheric effects are negligible. If N ERS Tandem pairs are available, N independent DEMs q_i can be generated (see Equation 3.16):

$$q_i = t + w_i \quad (i = 1 \dots N) \quad \text{Equation 3.17}$$

As outlined in the previous section, the problem is the identification of the best linear estimator of the local topographic profile, given the available data. Under the hypothesis that w_i is a zero-mean additive white Gaussian noise (this may sound a rather strong hypothesis^{xi}, but it can help to introduce the basic idea and avoid cumbersome computations, the maximum likelihood (ML) estimation of q is given by the following expression:

$$\hat{q} = \frac{\sum_{i=1}^N \frac{q_i}{\sigma_{wi}^2}}{\sum_{i=1}^N \frac{1}{\sigma_{wi}^2}} \quad \text{Equation 3.18}$$

i.e. optimum weights are simply the inverse of the noise powers (up to a normalisation factor). Let us now suppose that three Tandem pairs are available, they have been successfully unwrapped and three DEMs in SAR coordinates have been generated. All images have been registered on the same grid, and we suppose that only *one* GCP of known elevation has been identified in the image scene. In order to reduce the impact of possible baseline errors, the DEM estimated from the Tandem with the largest

^{xi} Gaussian statistics are a good approximation for the elevation error whenever the unwrapping has been performed successfully and the signal-to-noise ratio (i.e. the coherence) is high enough.

baseline is assumed as a reference (it is less sensitive to orbital parameter errors), and a low order polynomial is subtracted from the remaining data, to fit the reference one.

The target is now the estimation of the unknown noise power superimposed on the data, in order to properly combine them. A possible solution can be the following. From three data sets we can compute three error maps r_{ij} , computing the difference between q_i and q_j :

$$r_{ij} = w_i - w_j \quad \text{Equation 3.19}$$

Since w_i and w_j are statistically independent processes, the error mean power (P_{ij}) then results from two independent contributions:

$$P_{ij} = \sigma_{wi}^2 + \sigma_{wj}^2 \quad \text{Equation 3.20}$$

We can then estimate the unknown noise powers by solving the following system of equations:

$$\begin{aligned} \begin{bmatrix} P_{12} \\ P_{23} \\ P_{13} \end{bmatrix} &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \sigma_{w1}^2 \\ \sigma_{w2}^2 \\ \sigma_{w3}^2 \end{bmatrix} \Rightarrow \\ \begin{bmatrix} \hat{\sigma}_{w1}^2 \\ \hat{\sigma}_{w2}^2 \\ \hat{\sigma}_{w3}^2 \end{bmatrix} &= \frac{1}{2} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} P_{12} \\ P_{23} \\ P_{13} \end{bmatrix} \end{aligned} \quad \text{Equation 3.21}$$

As long as the data have the same accuracy, the error powers (P_{ij}) show very similar values and so the estimated weights take values near $1/N$. On the contrary, suppose that one DEM (e.g. the first) is strongly corrupted by noise and it is significantly worse than the other two (e.g. a very low baseline interferogram was used to estimate the topography); in that case, P_{12} and P_{13} have values higher than P_{23} , and so the estimated noise power for the first DEM will be higher than the other two (and a lower weight will be assigned to the first DEM).

In general, if N data are available, we can generate $M = \binom{N}{2}$ error maps,

and get an estimation of the weights to be used solving an overdetermined system similar to Equation 3.21, using an LMS approach. Consideration of both noise and atmospheric effects makes the solution of the problem more complex, since the different spectral behaviours of the two components should be taken into account. By using a proper filter bank (e.g. the wavelet transform) it is possible to consider, in each sub-band, the signal samples as effectively uncorrelated [Wornell93] and to carry out a wave-number-dependent weighted average. In fact, we can use the previous approach to get an estimation of the optimum weights to be used in every sub-band. Moreover, the algorithm can cope with non-uniform noise powers inside each datum, and it takes advantage of the coherence maps associated with each interferogram to carry out a space-variant filtering.

We conclude this section with the following recommendations:

- The averaging must always be *weighted*, taking into account the different baselines: e.g. considering Gaussian statistics and the same phase noise power on all the interferograms, weights should be proportional to the *square* of the normal baseline of each acquisition (equations B3.16 and B3.18).
- Coherence weighting, i.e. weighting in proportion to the coherence values in the different interferograms (a combination strategy very often encountered in InSAR literature), has *no statistical basis*^{xii}.
- As shown in [Ferretti99], whenever the atmospheric distortion is significant, at least in some data, considering just the decorrelation noise powers (estimated from the coherence maps and the baseline values) is not enough for a proper DEM combination: *weights are no longer correct*.
- The accuracy of the final product is strongly dependent on different factors (e.g. baseline values, number of data available, kind of topography, etc.), but can be as high as that achievable with optical satellite data.
- In [Ferretti99], it has also been shown that the residual elevation error (after the weighted averaging) is still concentrated at very low spatial frequencies (due to the spectrum of the atmospheric disturbances) and that the fusion with coarse resolution DEMs obtained with other techniques can further improve the elevation accuracy.
- Finally, it should be pointed out that, using the wavelet approach, it is possible to obtain not only a DEM but a *quality map* of the final result. More precisely, we can get an estimation of the error variance *a posteriori*. This can be useful, for example, in ascending and descending data combination, as it will be discussed in the next section.

3.6 Combination of ascending and descending passes

Due to the low off-nadir angle of the ERS satellites (23° at mid-swath), significant layover effects are often observed in areas with rough topography. When the surface slope approaches the incidence angle, the ground-range sampling step becomes wider and wider, so that a single measurement (often unreliable due to geometric decorrelation [Gatelli94]) characterises many output pixels in geographical coordinates. For that reason, ascending and descending data fusion is essential for slope coverage in SAR interferometry [Pasquali94]: areas affected by foreshortening and layover in one mode are well covered (if not in shadow) in the other one. The quality of such a combination is strongly dependent on the accuracy of the ortho-rectification step. Moreover, since the accuracy of the estimated DEMs can be different, it is again necessary to carry out a weighted average.

^{xii} Coherence value is not proportional to noise variance. The relation is not linear, and depends on the number of looks (see, for example, [Zebker94]).

The multi-interferogram combination described in the previous section can be carried out on ascending and descending data, provided that more than two interferograms are available for both acquisition geometries. This processing step allows one to produce two combined DEMs in SAR coordinates (range, azimuth) as well as two ‘quality maps’ (the estimated error variance). In order to produce the final DEM, it is necessary:

- 1) to compensate the data for possible baseline errors;
- 2) to geocode the data; and
- 3) to properly combine the two DEMs, taking into account the estimated noise variance.

As already mentioned in the previous sections, the first step can be performed quite easily if GCPs of known coordinates have been identified in the area of interest, or a low-resolution DEM is available^{xiii}. If this is not the case, more sophisticated strategies must be adopted, involving non-linear optimisation algorithms. The cost function can be obtained by one of the following considerations:

- Both DEMs describe the same area: polynomial artefacts are minimised, thereby maximising the matching between the two estimated topographies [Ferretti98].
- The *amplitude* images describe the same area too. Once correctly geocoded, a good match between some image features (e.g. strong scatterers, bright features) should be visible [Stan00].

A thorough analysis of these algorithms is beyond the scope of this manual. Furthermore, where very few GCPs are available on large areas, no ‘standard’ procedure currently exists.

Once the systematic errors have been removed (or strongly reduced), it is possible to geocode the data on the UTM grid. As already mentioned, kriging interpolation [Wackernagel98], though computationally expensive, is a good technique for passing to a uniform image grid. In fact, it is possible to associate for each interpolated sample an estimation of its variance (dependent on the distances between the data and the position of the cell to be interpolated and the estimated error variance on each data sample) and this makes a more accurate combination possible. In fact, areas affected by foreshortening in one satellite acquisition mode (e.g. ascending) give rise to high variance samples; since only a few data are available, the average distance between them will be large and usually they are strongly affected by geometrical decorrelation. In these areas the final DEM will resemble the topography estimated from data relative to the opposite acquisition geometry (descending), where the spatial sampling (for areas not in shadow) will be good and the geometrical decorrelation lower.

^{xiii} The resolution depends on the local topography: e.g. GTOPO30 (global DEM, ~1 km posting, publicly available [USGS]) is not enough to remove systematic errors on a 20 m posting InSAR DEM, apart from in very smooth areas.

3.7 Conclusions

Despite the complex processing and the practical limitations of the technique, repeat-pass SAR interferometry data can provide a valuable tool for low-cost DEM generation on a wide range of land surfaces.

ERS Tandem data are still affected by temporal decorrelation and atmospheric disturbances, but, whenever a multi-baseline approach is feasible (i.e. a sufficient number of interferograms of the area can be generated), the final results can be strongly improved.

Phase unwrapping problems and tie-point identification for accurate geocoding are still time-consuming steps, since they usually require user interaction. With time, hardware constraints (in terms of computational power and memory requirements) are getting less and less severe and more sophisticated optimisation algorithms are becoming feasible, notwithstanding the huge numbers of image pixels usually involved. We expect, in the next few years, a sort of ‘standardisation’ of the processing chain for InSAR DEM reconstruction, where the best algorithms will be chosen as reference tools. Meanwhile, SRTM data are available and can be used (at least for 80% of Earth’s land mass [SRTM]) as a starting point for more accurate analyses and updates.

4. Differential Interferometry (DInSAR)

4.1 Examples of differential interferometry on land

4.1.1 Physical changes

A change in the phase of an electromagnetic signal can be caused by a variation of the length travelled by the wave, by a change in the refractive index of the medium or by a transition at an interface. Most of the geodetic applications of radar interferometry have dealt with straightforward interpretations as geometric differences, whether used to compute topography or to terrain displacements. Changes of phase can also occur if the electrical conductivity changes uniformly within the surface covered by the radar pixels.

The effect, demonstrated in the laboratory on various natural targets, can be partially responsible for the change of phase observed on irrigated fields, together with the mechanical swelling of soils [Gabriel89]. To separate these contributions, simultaneous observations with two different wavelengths (L and C band for instance) would be extremely useful.

A good knowledge of these phenomena would open new prospects of measuring our environment with radar interferometry.

4.1.2 Volcano: Okmok

The displacement of Mount Etna [Massonnet95C] was the first example of mapping a moving volcano, which deflated after its 1992–1995 eruption. It gave an assessment of the depth of the source of deformation (16 km below the surface). This was a remarkable result for interferometry since this volcano is one of the best surveyed in the world by conventional tools. Later studies showed that the volcano somewhat reinflated in the following years. Some controversy surrounded this result; some scientists wanted to explain it by a possible atmospheric effect, which reflected their reluctance to admit the existence of such a deep magmatic chamber. A recent result based on a non-geodetic method [Murru99], confirmed the hypothesis of a deep chamber.

More recently, a remarkable example of volcano monitoring was published [Lu2000], in which a complete deformation cycle (pre-eruptive inflation, co-eruptive deflation and post-eruptive inflation) was observed on Mount Okmok, a volcano in Alaska. The results are impressive, as the co-eruptive deflation (1997) amounts to 140 cm. This is more than half the radio-electric depth of the standard atmosphere. Figure 4-1 shows this deflation as observed in part (a), plus a model in part (b).

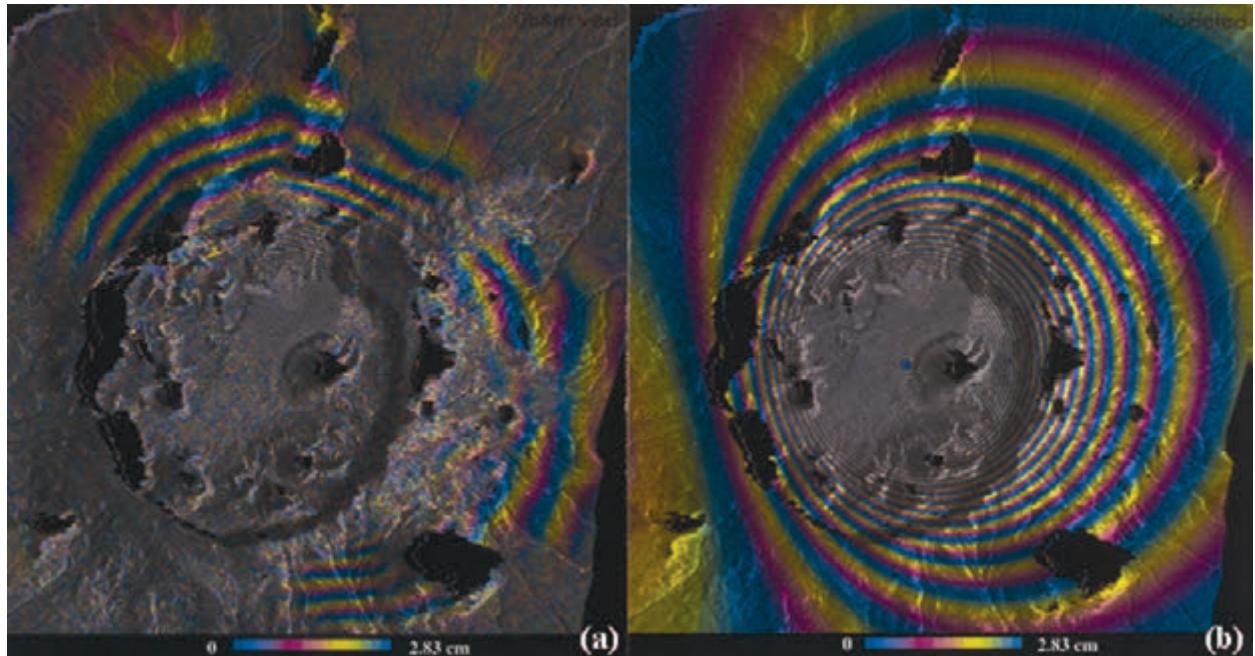


Figure 4-1: Full volcano cycle: inflation and deflation at Mount Okmok (from Z. Lu). This remarkable example of volcano monitoring includes a complete deformation cycle (pre-eruptive inflation, co-eruptive deflation and post-eruptive inflation). Mount Okmok, located in Alaska, is a good example for demonstrating the interest of remote monitoring in a ‘difficult’ environment in northern latitudes. As with other sub-arctic sites of high geophysical interest such as Iceland, data are useful only if acquired at certain periods of the year. The co-eruptive deflation (1997) amounts to 140 cm (part a) and is modelled in part b (using MOGI elastic modelling). The resemblance is very convincing, wherever surface coherence allows comparison.

The pre-eruptive inflation reached 18 cm in the 1992–1995 period. The uplift then resumed with 10 centimetres in 1997–1998. These results indicate how mature the use of radar interferometry is: the volcano is located in a ‘difficult’ environment at a northern latitude, which prevents all-year-long data takes. The effect of the atmosphere was permanently assessed and taken into account. Additional results of this study dealt with the behaviour of fresh lava flows with regard to coherence, due to the cooling-compaction processes.

4.1.3 Surface rupture: Superstition Hill

In the course of the study of the Landers earthquake [Massonnet93], several fault line slips were observed. They appeared as cuts through the otherwise homogenous fringe pattern generated by the continuous deformation caused by the earthquake. Some additional slips were detected farther from the epicentre, at the location of well-known faults [Massonnet94]. The result was all the more important because, albeit not recognised in the field, the section of the fault which ruptured corresponded to the area of maximum stress anticipated in earlier theoretical developments by geophysicists.

In Figure 4-2 we see another example of the power of interferometry for displacement mapping.

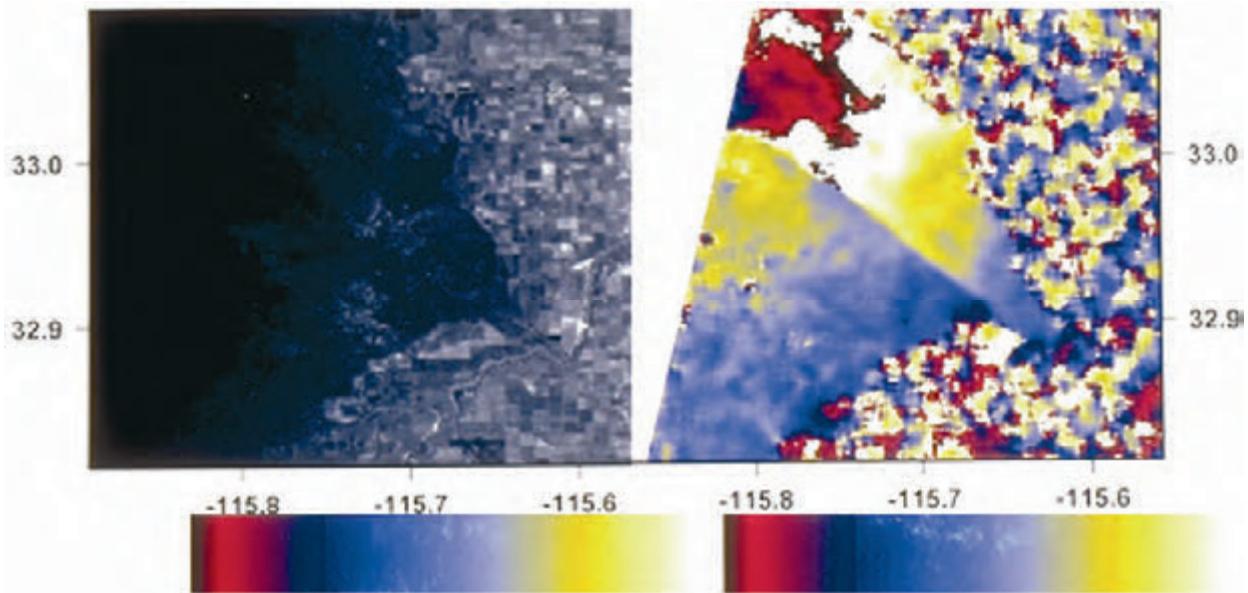


Figure 4-2: Example of fault slip detection: Superstition Hill. Two-year interferogram showing fault slip as a remote consequence of the Landers Earthquake. The slip is estimated as 10 mm. The sharp morphology of a fault crack as seen in an interferogram makes it very difficult to confuse with anything else. The accuracy of the measure is about one millimetre.

The area is located near Salton Sea, some 150 km south of the epicentre of the Landers earthquake. The image on the left is the amplitude of one of the two radar images from ERS-1, used to build the interferogram on the right. Coherence is lost on areas used for agriculture, as expected, since the interferogram is built with two radar images separated by two years. The two images were extracted from the ‘winter cycle’ of ERS-1, the first in 1992 and the second in 1994.

The interferogram covers the same area as the amplitude image, as shown by the latitude and longitude ticks. The fault slip of about 10 mm extends over more than 20 km on Superstition Hills, as a remote consequence of the Landers earthquake. As usual with ERS, one full colour cycle (i.e. red, white, yellow, blue) represents a deformation amounting to half a radar wavelength, or about 3 cm with the radar of ERS-1. The cut amounts to about one third of it, or 10 mm.

The co-seismic fault slip has been recognised in the field [Sharp92]. This example illustrates the ability of radar to see tiny phenomena, from very far away, without the help of any ground instrumentation. Such a fault slip is not as obvious as for example a crack in the tar of a road. Furthermore if, rather than being localised, the fault rupture area were distributed over, say, 20 metres, it would become extremely difficult to spot on the ground, but would appear with the same clarity in a radar interferogram.

An interesting feature of this kind of measurement is the easy and safe interpretation: an abrupt cut in an interferogram cannot be caused by any atmospheric phenomenon. The atmosphere cannot create a ‘step’ like this in the refraction index. The major cause for interferometric artefact is thus discarded. Similarly, such a feature can hardly be a topographic error: we

would need an unknown vertical cliff to create it; even an inaccurate DEM could smooth such a cliff across several pixels, but could not ignore it. We estimate the accuracy of the measurement to be about one millimetre, from the variations of the amplitude of the cut.

4.1.4 Subsidence: East Mesa

Interferometry can be used not only in purely scientific context, but also for helping understand industrial or legal problems. Figure 4-3 shows a long-term interferogram covering a location across the US-Mexico border. This example was observed by chance. We looked at an area used for agriculture, previously observed by Goldstein *et al.* in the eighties using SEASAT images, in the hope of observing changes linked to irrigation. We used ERS-1 and C-band, in order to compare results with different wavelengths. Short term interferograms acquired during the ‘ice-phase’, with time separations that were a multiple of the three-day orbital cycle, were generated and showed the expected effect: the intensity image on the left exhibits the expected loss of coherence that is linked to agricultural use.

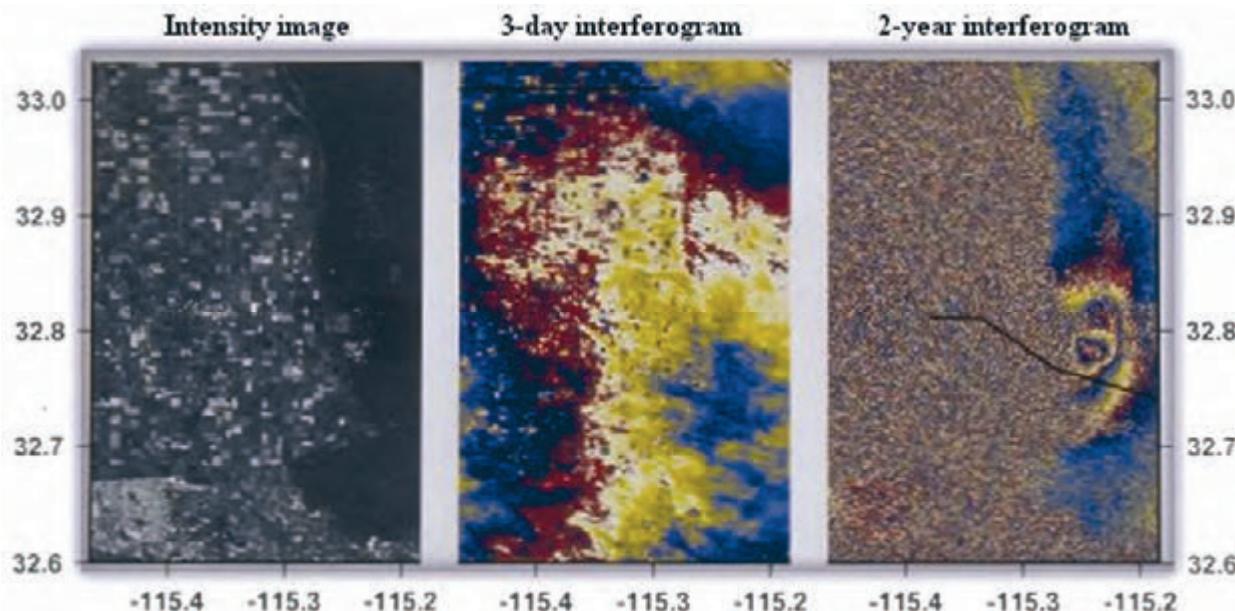


Figure 4-3: Example of industrial subsidence: the East Mesa Geothermal Plant

Out of mere curiosity, interferograms were also generated between the two ‘ice phases’, separated by two years. The surface used for agriculture does not show any fringes, as expected. Elsewhere, however, an ellipse-shaped subsidence bowl was observed. This area of deformation is clearly visible in the right-hand image, and it is centred on the East-Mesa geothermal plant. The black line indicates the profiles where levelling has been conducted to survey the extent of the ground deformation.

Ground deformation is a sensitive issue in this region as several irrigation channels cut across the area. This deformation is real and corresponds to an area of industrial subsidence in this extremely flat terrain. The interferogram

represented has a low sensitivity to topography (about 1000 m for each fringe). Furthermore, the deformation can be seen on several long-term interferograms, thus ruling out the possibility of any atmospheric propagation change which could be linked to a specific radar image. Assuming the displacement is vertical, and taking into account the local incidence angle, each fringe corresponds to 31 mm of vertical displacement, of which 28 mm is seen in range. The outer shell of the subsidence area, which corresponds to the first 31 mm of deformation, is 17 km by 8 km, or 105 km², or 2.9 million cubic metres of volume loss. The second 31 mm is more localised near the southern part of the field, and shows some ragged borders, possibly being more influenced by the actual geographic distribution of the extraction wells. It represents a much lower volume of about 900 000 cubic metres. Finally, the deformation reaches its highest amplitude of about 90 mm in the vicinity of one of the major production areas, representing a small additional volume loss of 200 000 cubic metres. The total loss of volume is therefore of the order of 4 million cubic metres. A direct integration of the volume loss conducted after local ‘phase unwrapping’ of the interferogram gives 3.8 million cubic metres, assuming that the deformation field has east-west symmetry.

For this site, the radar data are in excellent agreement with levelling data. Both the rate and extent of subsidence indicated by surveys conducted along the line shown in Figure 4-3 (right) are consistent with the subsidence indicated by the radar interferometry. Maximum rates of subsidence from the 1991–94 levelling are about 18 mm/yr, which compares to 18 mm/yr from 1992–94 interferograms. The radar data, however, provides a more detailed mapping of both the magnitude and area of surface deformation. The proximity of the subsidence area to two irrigation canals is of particular concern because the canals rely on gravity flow for their operation.

The radar interferogram also permits several observations about the relation between geothermal fluid production and the subsidence at East Mesa. The interferogram indicates that the maximum subsidence is over the southern end of the field, in a small area where about half of the production occurs. The map permits a direct comparison of the volume of the subsidence bowl to the volume of fluid removed from the geothermal reservoir. Although most of the extracted fluid is reinjected, a comparison of gross production to reinjected water indicates about 5 million cubic metres of water was removed from the reservoir from 1992 to 1994. This is comparable to the 4 million cubic metres volume of the subsidence bowl computed from the interferogram.

This example shows that radar interferometry can be used for problems in relation to legal issues and can monitor environmental damage to the environment. The study on East Mesa illustrates a whole class of problems to which interferometric data are particularly suited. Subsidence can be caused by natural gas storage, oil extraction, irrigation water pumping, or mining.

In contrast, landslides are difficult to monitor because they are always located on slopes, a difficulty even for radar with an angle of incidence less steep than the one of ERS.

4.2 Example of differential interferometry on ice

Although a few spectacular results were obtained on the Arctic and Antarctic ice caps in the early life of ERS-1, differential interferometry in these regions considerably improved with the availability of tandem data after the launch of ERS-2.

Ice surfaces lose their coherence quickly, often in a matter of days. In addition, flowing ice can very quickly create a displacement gradient that exceeds interferometric capabilities (i.e. they can create more than one fringe per pixel); ice motion can also lead to a general loss of coherence. The one-day time lag of the tandem passes elegantly solved both problems. The tandem mission generated large volumes of stunning results on the ice caps, causing a revolution in the field.

The main interest here is the stability of the ice caps, thought to be threatened by global warming. In this domain, ERS interferometry can really bring top level results into a very hot scientific debate.

Among the open questions is the problem of determining the line of buoyancy of a glacier (i.e. up to what point liquid water exists under a glacier). Interferometry can help to answer the question by detecting the flexion of the glacier with tide. Figure 4-4 shows an Arctic glacier in Greenland where this question has been studied and solved [Rignot98].

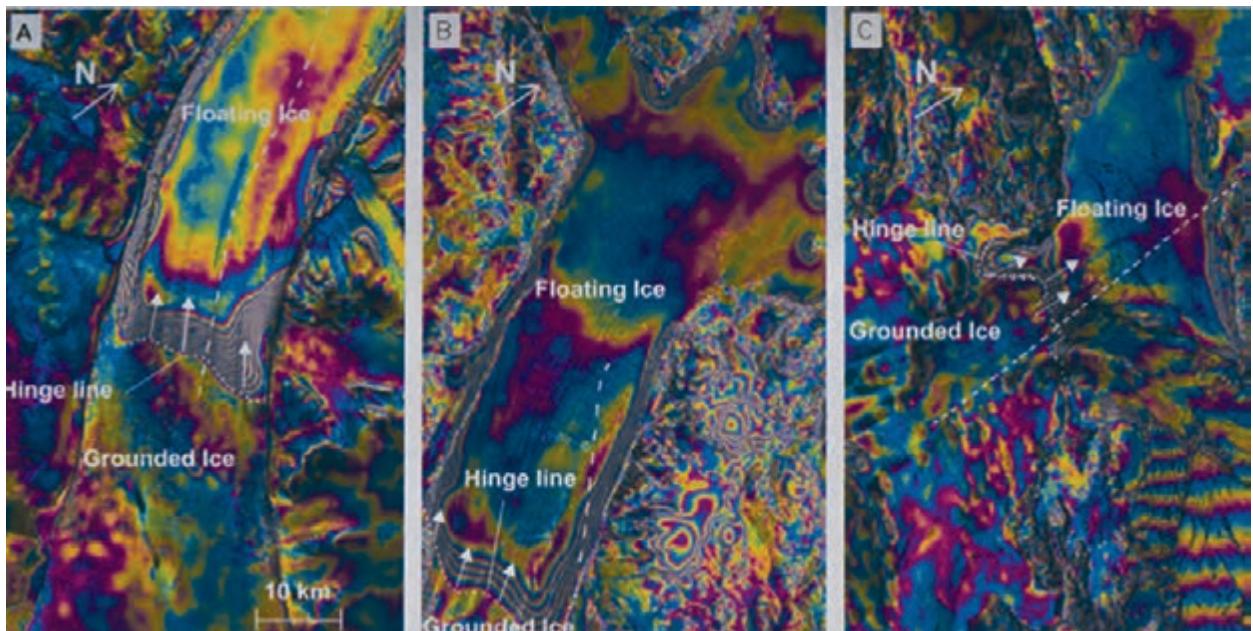


Figure 4-4: ‘Tandem’ ice flow mapping: Arctic glacier flowing [Rignot98]. (a) Petermann glacier, (b) Nioghalvfjerdsbrae glacier, and (c) Zacharia glacier.

In these three examples of glaciers in Greenland, the line of flexion of the glaciers has been determined by interferometry. The flexion of the glacier results from tides. The tongue of the glacier is lifted by the tide wherever it floats. The glacier flexes at the transition from ground lying to floatation. It is very important to know to where the liquid water can creep beneath the

glacier, because this determines the size of the interface where the melting can occur. Actually the displacement of a glacier can be split into two terms: a ‘secular’ term representing the generally constant flow of the glacier, and an oscillating term due to the action of tides. Subtracting two tandem interferograms can single out the tidal component, because the one-day secular flow is cancelled out by the subtraction, leaving a compound of the tidal effect at four different dates. Although the amplitude of the effect has been successfully modelled, the main output of the study is the location of the transition between the ground and the floating region.

Glaciologists have developed ad hoc methods to take the best from interferometry. They have sometimes had to make additional hypotheses, such as the assumption of the flow parallel to the slope, in order to use the intrinsically one-dimensional interferometric measurement.

4.3 Review of various criteria for data selection

In addition to what is stated in section B.1.4, data selection for displacement mapping is a complex mix of considerations, including, of course, the obvious prediction of the topographic sensitivities using the orbital files, but also taking into account some climatic aspects. For instance, in Iceland, only two months of each year are likely to be snow-free. If no reliable DEM is available, one might think of using a tandem pair for producing a DEM, or, alternatively, to select scenes that are likely to give good integer combinations. In the latter case, one must be very cautious to check the topographic sensitivity in several places on the test site, because the more ‘magic’ an integer combination is, the more likely it is to be unstable and to give much less useful values for the rest of the image. A good piece of advice is to check the predicted sensitivity for the four corners of the image, or, if it is much smaller than 100 km, to restrict it to the four corners of the test site.

Finally, since the worst artefacts in interferometry usually come from the atmospheric contribution, it is advisable to select scenes acquired under anti-cyclonic conditions, or at least to ensure that several independent interferograms contain the information about the required displacement signal. Note that the requirements for such ‘double-check’ interferograms are less stringent than for the ‘official’ one. They are useful only for checking that a geophysical measurement is not an artefact. In other words, that it can be detected on more interferograms, even those with low quality.

4.4 Interferometric interpretation

One of the main difficulties in interferometry is the mix of several different types of geometrical information in a given signal. Consequently, the measurement accuracy is not driven by the characteristics of the radar system (power, resolution, etc.) as much as by the terrain stability and the possibility of separating the various components in the signal. The main limiter of the basic accuracy of the measurement is change in the geometric and physical properties of the ground during the time separating the observations, for example if the moisture content of the soil changes, or

there is local motion. The standard deviation results directly from the ratio between the average amplitudes of the coherent and incoherent fractions of the signal. This basic accuracy will only be attained in the final measurement if the contributions of other types of geometrical measurements are eliminated or a sufficiently low upper limit for these effects is calculated.

4.4.1 Interferometry phase signal ruggedness

A predominant feature of the interferometry technique is the extraordinary resistance of the coherent part of a signal with regard to the incoherent part. Fundamentally, the phase is a sign change. Even when mixed with an incoherent signal of the same amplitude, the coherent signal will show a phase alteration of which the standard deviation is only $\pm 13\%$ of a cycle. When the amplitude of the incoherent fraction is equal to half the coherent fraction (i.e. simply a 6 dB protection margin), the standard deviation becomes $\pm 6\%$. When the amplitude of the incoherent fraction is 10%, the standard deviation on the error is $\pm 1.1\%$.

It is necessary to take into account the final size of the pixel or of the targeted geographic cell in interferometry. As an example, assuming an elementary 30 m^2 pixel used for creating a 30 m-sided grid as the final product (i.e. a surface area 30 times larger), the ratio between the coherent signal and the incoherent signal will show an improvement by a factor of the square root of 30. The four sketches in Figure 4-5 illustrate this effect.

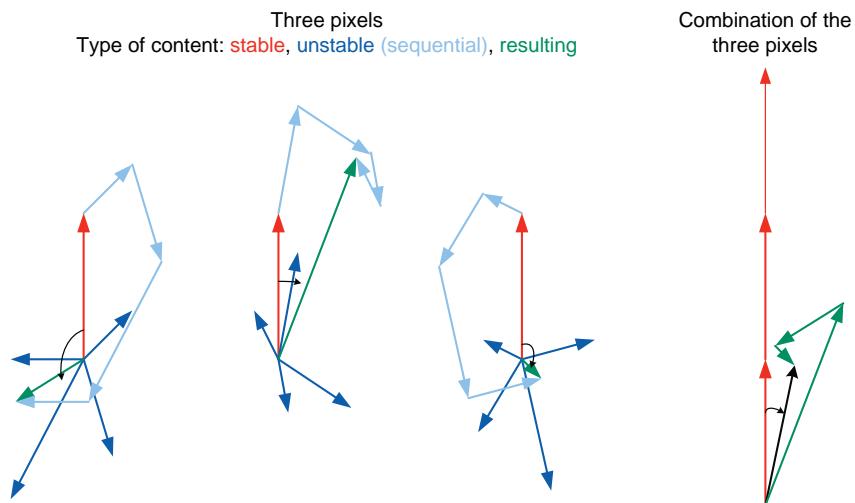


Figure 4-5: Combining coherent and incoherent signals

The first three ‘break down’ a radar pixel according to a ‘stable’ component (shown in red), of unit length, and four random (or ‘unstable’) components resulting from a statistical Gaussian draw on two components, x and y , so that the expected amplitude is equal to one for each random fraction shown in blue. The resulting vector is displayed in green, i.e. the value of the final pixel for which the phase differs from zero due to the random components. In the fourth sketch, the vectors of the three previous sketches have been

coherently summed. The amplitude of the ‘coherent’ component therefore reaches three. Due to the inefficiency of incoherent additions, the random fractions can no longer alter the final phase to the same extent as before summing.

Summing of the complex numbers from which phases are derived gives very different results depending on whether or not they are coherent. The amplitude of the sum of N coherent vectors is N times the amplitude of one of them. The power of the result (square of the amplitude) is then N^2 . On the other hand, the power of the sum of N random vectors is N times the average power of one of them, i.e. simply a \sqrt{N} gain on the amplitude, and hence an improvement of the final ratio of \sqrt{N} on the amplitudes.

4.4.2 Fictitious example interferograms for analysis

The improvement to the raw accuracy results from the filtering, if any, applied to the signal at the interferometric processing output. The main filtering trick is the complex summing described above, which gives an advantage to the coherent fraction of the signal equal to the square root of the ratio between the summed surface and the initial surface. Discrimination between different effects mixed in the measurement is obtained either from analysis of a parameter error depending on the conditions of the data take, or from logical reasoning based on the analysis of several interferograms of the same site. We are going to operate these mechanisms using several scenarios based on a series of images.

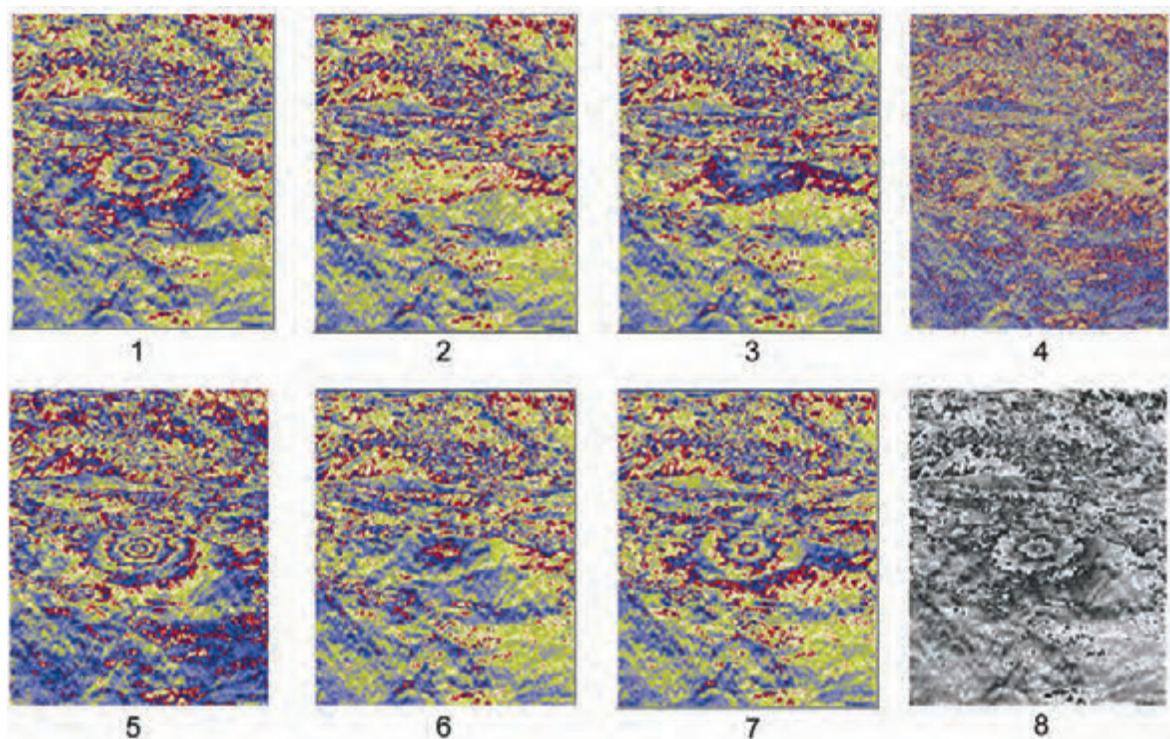


Figure 4-6: Series of images produced by superposing an elliptical signal (with a variable amplitude) and a background (extracted from a genuine interferogram)

The series of images (Figure 4-6) is produced by adding an ellipse-shaped feature of scalable amplitude and a background intended to represent the typical residues found in an interferogram, either from meteorological effects or from incorrectly compensated topographic effects. This background has been extracted from a genuine interferogram. The elliptical feature was added with a variable amplitude to the various members of the series. A phase cycle is represented by a coloured cycle. The advantage of this representation is that it is very easy to read and enables a much finer level readout than in black and white, as the eye is much more sensitive to colours than to levels of grey. On the other hand, while the sign of the black and white representation is not ambiguous, the sign of the coloured representation is arbitrary and depends on the table of colours used. In the last image of the series, identical to the first but in black and white, the growth of the phase from the outside to the inside of the circles is unambiguous. Crossing of an ambiguity is shown by the abrupt change from the maximum (in white) to the minimum (in black). The same is not true for colour representation as the table may be organised in a ‘Red-Yellow-Blue’ sequence or in a ‘Red-Blue-Yellow’ sequence. Therefore, it is necessary to compare the variation of the phase either to the colour table or to phenomena of which the sign is known and which have been processed using an identical procedure. The latter method is safer, as numerous phenomena occur which can change the sign of an interferogram. Inversion of real and imaginary parts of raw or processed data causes this type of change. Inversion of the pair made up of the reference image and the slave image during the creation of the interferogram also changes the sign of the result.

As can be seen in the series of colour images, it is not always easy to count the number of fringes characterising a structure even if it is an elementary shape such as these concentric fringes. The first image thus comprises three circular fringes, passing from the central yellow point to two other yellow circles, then to the background yellow colour. One can see that the background yellow is ‘the same’ everywhere, i.e. it is from the cycle, and that there is no phase transition. The second image does not include any fringes. The third image shows simply one fringe but it can be seen that the progression of colours is reversed with respect to the first image. If the first image comprises ‘+3’ fringes, the third therefore comprises ‘-1’. The fourth image comprises two fringes and the fifth image comprises -5 concentric fringes. The sixth image has only one fringe and the seventh ‘-3’. To illustrate the coherent summing effect on the readability of images, the fourth image is represented unfiltered whereas all the others result from a 3 by 3 summing. The phase of each pixel has been replaced by the phase of the complex number, which is formed by the sum of the nine complex numbers formed from the phase of this pixel and its eight neighbours.

These examples show to what extent it may be difficult to recognise a low amplitude structure in the presence of background noise. Remember that the artificial signal introduced in the interferograms is of a strict elliptical shape. Only its amplitude varies.

4.4.3 Analysis of fictitious situations

We are now going to propose some fictitious scenarios of interferograms. With regard to the interpretation, we will limit ourselves to the following phenomena that we will endeavour to characterise:

1. Atmospheric artefacts which will be characteristic as they are associated with a data take and therefore appear with the same amplitude, albeit possibly with a change of sign, in all interferometric combinations of a given image
2. Incorrectly compensated topographic contributions characterised in each interferogram by an amplitude inversely proportional to the altitude of ambiguity specific to this interferogram
3. Practically instantaneous deformation which will be found with the same amplitude, but possibly different sign, in each interferogram for which the dates of the images bracket the date of the event (for example, an earthquake)
4. Regular deformations over time, the amplitude of which in each interferogram is proportional to the difference in time between the data acquisition dates of the two images used

This list of phenomena is not exhaustive. Deformations of which the temporal behaviour is more complex could be added, or even partially reversible phenomena (swelling and sinking of volcanoes and water tables, etc.).

Scenario 1: three images

In the first scenario we have combined the images from orbit numbers 5222, 10232 and 11234 of satellite ERS-1. Remember that an Earth Observation satellite typically flies more than five thousand orbits per year. Notice that the differences in the orbit numbers are multiples of 501 which is the number of orbits flown by ERS-1 during its 35-day orbital cycle.

The combination of orbits 10232 and 11234 produces interferogram 7 of the series, which contains -3 fringes. The analysis of the orbits results in an ambiguity altitude h_a of 70 metres for this interferogram. Combining orbits 11234 and 5222 results in interferogram 1 of the series, which contains 3 fringes. Here, the ambiguity altitude is equal to -120 metres.

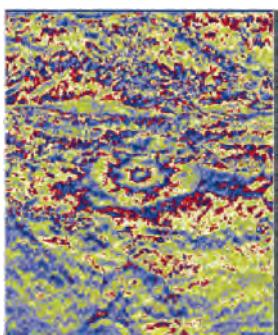


Image 7

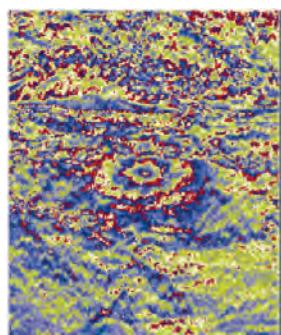


Image 1

Generally, interpretation by elimination is efficient. If the observed phenomenon were due to a faulty topographic correction, it would give a different number of fringes in each of the interferograms due to the significant difference (practically a factor of two) between the magnitudes of the ambiguity altitudes. A regular movement over time is also excluded as the elapsed times correspond respectively to 1,000 orbits (approximately two months) and 6,000 orbits (more than one year). However, the amplitudes of the ‘deformations’ are identical. In addition, the period covered by the first interferogram is totally included in the period of the second interferogram. The change to the sign for the number of fringes is explained by the ‘reversal of time’ in the second interferogram of which the first image is not the oldest.

Therefore, the scenario is compatible with a ‘–3 fringe’ deformation created between the dates of orbits 10232 and 11234, which naturally will also be observed by the pair 11234 and 5222, with reversal of the sign. However, we must ensure our explanation is the only one possible. In this case, the measurement is also compatible with an atmospheric effect on the image common to the two interferograms (image 11234): if this interferogram has taken a circular atmospheric phenomenon with an amplitude of three fringes, it will create 3 fringes in the interferogram 11234/5222 (phases of 11234 less phase of 5222) and –3 fringes in interferogram 10232/11234 (phases of 10232 less phase of 11234).

Therefore, despite the three orbits available, it is not possible to conclude on the nature of the phenomenon.

Scenario 2: four images

In a second scenario, the images from orbits number 5044 and 8050 of the satellite ERS-1 were combined into an interferogram represented by image 7 of the series (which comprises –3 fringes), with a topographic sensitivity h_a of 30 m. Two other orbits from the same satellite, 9052 and 7549, provide the interferogram represented by image 1 of the series (3 fringes), with a topographic sensitivity of –250 m.

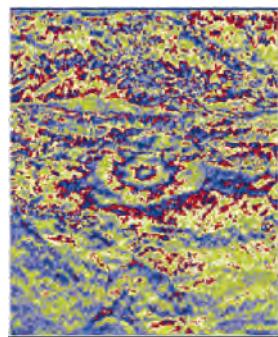


Image 7

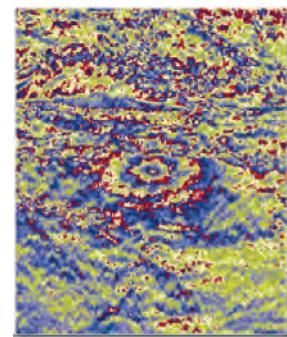


Image 1

Here, atmospheric effects cannot be invoked, as the interferograms do not share any images and it is highly improbable that a structure of atmospheric origin is identically repeated. The large difference in the ambiguity altitudes

excludes a topographic interpretation, as the two interferograms in absolute values have the same number of fringes. A deformation movement, regular over time, is unacceptable when the deformation amplitude is the same, whereas one of the time differences is twice the other.

All that remains is the assumption of an abrupt deformation, with a -3 fringe amplitude, which necessarily must have occurred in the common time interval of the two interferograms, i.e. between the dates of orbits 7549 and 8050, with the reversal of the sign being explained by the reversal of the time in the second interferogram.

The four orbits available now make it possible to conclude and confirm the nature of the phenomenon. The determination of the date of the event is better than that obtained from each of the interferograms considered separately.

Scenario 3: five orbits

In this scenario, the investigation is with five ERS-1 orbits numbered 5001, 5502, 6003, 7005 and 7506, combined as follows:

- 5001/7506 gives image 5 (-5 fringes) for which $h_a = 100$ m
- 7005/6003 gives image 4 (2 fringes) for which $h_a = 90$ m
- 5502/6003 provides image 3 ('-1' fringe) for which $h_a = 50$ m

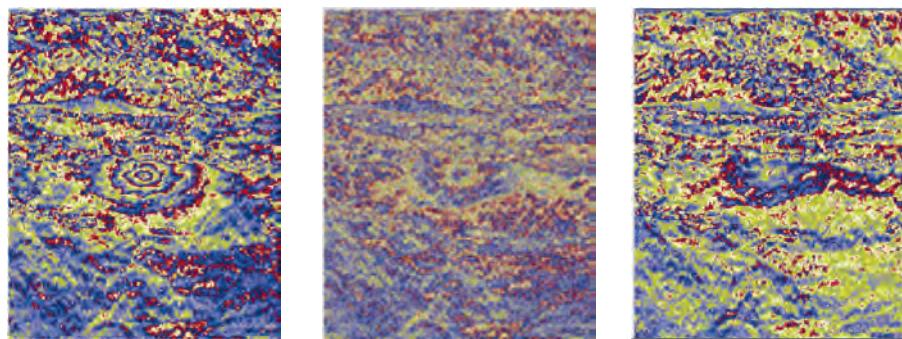


Image 5

Image 4

Image 3

Using a similar reasoning to the previous scenario, it is possible to deduce that the movement is regular over time, with a rate of one fringe per month, or more precisely a fringe every 35 days.

Scenario 4: seven orbits

The ERS-1 images from at least seven orbits are involved in this scenario (numbers 4530, 5031, 5532, 6033, 7035, 8037 and 12546). The first five images correspond to maximum temporal sampling in an interferometric series, i.e. an image every 501 orbits (or 35 days). Naturally, the scene could also be viewed by another interferometric series, either from a different direction (for example ascending instead of descending) or more generally from an image with a non-zero geographic intersection with ours. This type of image could be interleaved time-wise with our images, but they always give measurements over different multiples of 35 days.

The interferograms used are:

- image 2 (no fringe) for the pair 4530/12546 with $h_a = 1000$ m
- image 4 (2 fringes) for the pair 5031/5532 with $h_a = 60$ m
- image 7 (-3 fringes) for the pair 7035/8037 with $h_a = -40$ m
- image 6 (1 fringe) for the pair 6033/4530 with $h_a = 120$ m

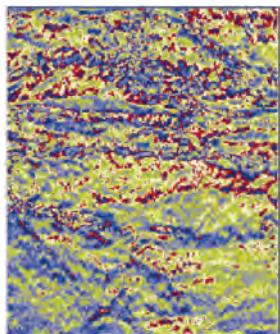


Image 2

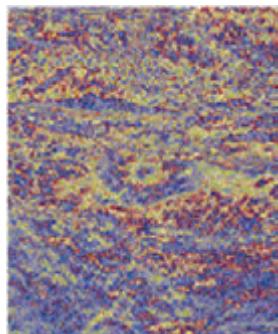


Image 4

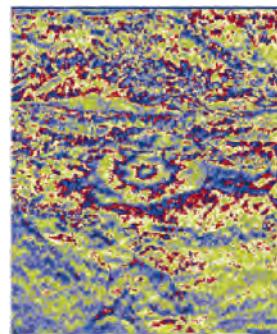


Image 7

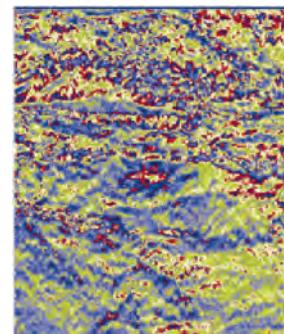


Image 6

In this scenario, the assumption of a geophysical deformation is immediately excluded as the first interferogram does not show it even though its dates bracket the entire series. For the same reason, it cannot be a regular deformation over time, and, as the series mainly comprises separate images, it cannot result from atmospheric effects. Notice that the number of fringes is inversely proportional to the ambiguity altitude, which indicates a 120-metre topographic error. Why is this not found on the first interferogram? Because its amplitude of only one tenth of a fringe is practically undetectable on a background quality as mediocre as the one found in our examples.

Scenario 5: using an L-band radar satellite

In this last scenario, we examine data from a satellite other than ERS-1, the Japanese radar J-ERS, of which orbits number 10032 and 10691 have made a unique interferogram represented by the fifth image of the series (-5 concentric fringes). Note that the difference in the orbit number is 659, which is the number of orbits flown by the J-ERS during its 44-day orbital cycle. The interferogram presents an ambiguity altitude of 250 metres.

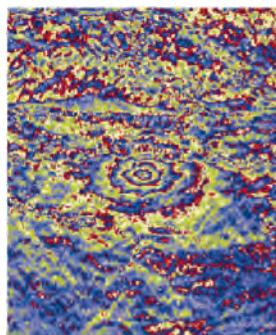


Image 5

Despite the uniqueness of the interferogram, it can be interpreted. In fact, a topographic error would reveal the existence of a 1,250 m deep hole. In this case, such a hole that was not filled with water would be known worldwide (as a giant open quarry, etc.). J-ERS operates in L-band. If it was caused by displacement, the five fringes would therefore represent approximately 60 cm. No tropospheric phenomenon can create a heterogeneity equal to one quarter of the atmospheric column propagation delaying effect. A phenomenon related to the ionosphere, if it can reach this amplitude range in L-band, could not be as localised as our circular fringes. A regular deformation over time also leads to extraordinary amplitudes (40 cm of swelling per month). Therefore, it can only be caused by an earthquake or a volcano in a highly active phase.

Conclusions

These five scenarios have enabled us to understand that there is no absolute rule or maximum or minimum number of images for correctly interpreting an interferometric sequence (nonetheless, at least *two* images per radar are always required!). Generally speaking, four images that can make up two completely independent interferograms make it possible to initiate a worthwhile discussion on atmospheric artefacts. The other combinations of these four images (there are six two-by-two combinations in all) make it possible, more often than not, to make sensible conclusions unless faced with very complicated terrain movements.

Part C

InSAR processing: a mathematical approach

1. Statistics of SAR and InSAR images

1.1 The backscattering process

1.1.1 Introduction

The dimension of the resolution cell of a SAR survey is much greater than the wavelength of the impingeing radiation. Roughly speaking, targets that have small reflectivity and are more distant than a wavelength, backscatter independently. The amplitude of the focused signal corresponds to the algebraic combination of all the reflections from independent scatterers within the cell, with their proper amplitudes and phases. This superposition of effects is only approximate as one should consider not only the primary reflections (i.e. satellite → target → satellite), but also multiple ones (say satellite → tree trunk → ground → satellite). Anyway, the focused signal is the combination of many independent reflections, with the possibility that some of them are much higher than the others. Therefore, statistics is the main tool to describe the backscattered signal: the probability density of the returns will be approximately Gaussian as the probability density of the sum of several independent complex numbers tends to be Gaussian, for the central limit theorem. The power of the reflections is additive, as usual. In this section we shall consider the amplitudes of the returns, first in the cases of artificial and then of natural back scatterers.

Obviously, as the artificial reflectors of interest are those highly visible from the satellite, they will correspond to scatterers made in such a way as to concentrate the incoming energy back towards the receiver; as the receiver is far away, the curvature of the surfaces will be small, and the scatterer will look like a mirror or a combination of mirrors. These scatterers are characterised by their **Radar Cross Section (RCS)**, i.e. the effective surface area that intercepts the incident wave and that scatters the energy isotropically into space.

The behaviour of natural scatterers, on the other hand, is better described by using the superposition principle and statistics.

1.1.2 Artificial backscatterers

Radiation coming from a satellite impinges on all sorts of backscatterers on the terrain under observation. If the surface of the scatterer is flat, the radiation is in general mirrored away from the satellite. For this reason, flat surfaces (lakes, quiet seawater, etc.) appear mostly black in the intensity images. However, it may also happen that a flat surface mirrors back towards the satellite (say, a flat roof with the right inclination and azimuth) and in this case the pixel appears very bright, as the RCS is high. Other man-made scatterers that appear bright are long cylinders (wires) aligned along the azimuth direction or vertically (the polarisation of the radiation in the case of ERS is VV). Dihedral scatterers (e.g. the corners formed by

buildings) parallel to the azimuth direction are also bright. The directivity of a dihedral filling a resolution cell is very high, since this cell is generally about 5 metres long. Any change in the azimuth direction that makes the ends of the dihedral move $\pm\lambda/4$ in the satellite direction would eliminate the backscatter, and this corresponds to a beam-width of about 0.33 degrees. Finally, metallic mirrors pointed precisely in the direction of the satellite ensure a bright reflection with a very high directivity if the radar cross section is high enough.

Taking into account the two-way path and the directivity, the radar cross section of a mirror pointed towards the satellite is:

$$S = \frac{A^2}{\lambda^2} 4\pi \rightarrow A = \frac{\lambda}{2} \sqrt{\frac{S}{\pi}} \quad \text{Equation 1.1}$$

where A is the area of the mirror.

Assuming a square mirror with side length l , in the case of ERS we have:

$$S = 3923 \times l^4 \quad \text{Equation 1.2}$$

If we wish a mirror to appear much brighter than the surroundings, its reflectivity should be, say, 100 times that of the neighbouring pixels. Now, the area of a pixel for Single Look Complex data is about 90 m^2 . If we suppose that the normalised local reflectivity of the terrain is $\sigma^0 = -6 \text{ dB}$, then:

$$S > 100 \times 90 \times 10^{-0.6} \quad \text{Equation 1.3}$$

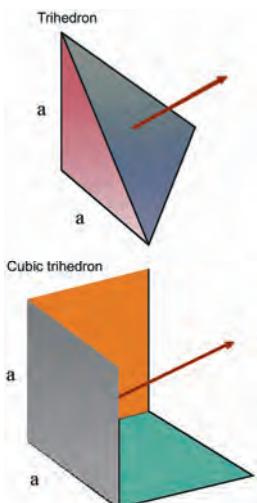


Figure 1-1: Trihedrons

The area of the mirror should be greater than 0.76 m^2 . The ensuing directivity is rather high (about 1.9 degrees between the zeros of the radiation pattern in azimuth and range directions, if the mirror is square). In order to decrease this directivity and simplify the positioning of the reflector, trihedral scatterers (corner reflectors) are often used. Three mirrors together form the corner of a trihedron. The multiple reflection of the incoming radiation is scattered exactly back, notwithstanding a low directivity (tens of degrees). The phase of the return is equivalent to that of a reflection from a point located at the corner of the trihedron. The effective area of the corner reflector is two thirds of that of the area of its triangular mouth, if the corner is pointed exactly; otherwise, it decreases gently with the angle. In order to have the same 0.76 m^2 area derived from the metallic mirror, the side of the trihedron corresponding to the corner reflector should be greater than 1.14 m.

Other corners have a cubic shape, rather than triangular. The RCS then increases by a factor of 9.

1.1.3 Natural backscatterers: the spectral shift principle

If we have natural (non-metallic) scatterers, then their reflectivity is usually rather small and we can neglect multiple reflections. We can use the principle of superposition of effects to calculate the total backscatter by combining the contributions of series of hypothetical idealised elementary

backscatterers. Two types of elementary backscatterers are of interest in interferometry:

- the grating (a spatial sinusoid)
- the point scatterer

a) The grating that we refer to lies in a horizontal plane, is uniform along the azimuth direction, has the length δ_g of the ground range resolution cell. We define $\Lambda/2$ as its period along the range direction.

Considering the two-way travel path, in order to have a reflection the returns from the elements of the grating should arrive in phase. If θ is the incidence angle, the spatial period of the grating should be in the interval:

$$\Lambda_{\min} = \frac{c}{2f_{\max} \sin \theta} \div \Lambda_{\max} = \frac{c}{2f_{\min} \sin \theta} \quad \text{Equation 1.4}$$

where:

$$f_{\max} = f_{\min} + B \quad \text{Equation 1.5}$$

and $B = 16 \times 10^6$ is the bandwidth (for the ERS case).

Any extended scatterer visible to the radar can be considered as superposition of gratings at these different wave numbers, or of point scatterers.

b) The point scatterer is wide band and has a very wide directivity function, i.e. the amplitude of its return changes very slowly with the wavelength and with the incidence angle.

In the case of the grating, the amplitude of the return of the extended scatterer changes rapidly both with the frequency and the direction of arrival of the radiation. In fact, as is the case for a mirror or a dihedral, the amplitude of the return is cancelled out if the contributions of the elementary scatterers that delimit the grating move $\pm \lambda/4$ in the direction of the satellite. Indicating the slant and ground range resolutions with δ_s and δ_g respectively, we get:

$$\delta_s = \frac{c}{2B}; \delta_g = \frac{\delta_s}{\sin \theta} \quad \text{Equation 1.6}$$

Let us consider now the return from a grating that occupies the entire range resolution cell and is tuned to the radar central frequency. If the radar frequency changes from f to $f \pm B/2$, i.e. at the edges of the band, the return will reduce to zero. Again the return is zero if the off-nadir direction changes from θ to $\theta + \Delta\theta$; to calculate $\Delta\theta$ we impose that the two-way travel path change between the nearest and the farthest elementary point scatterers (with the same amplitude) composing the grating is equal to λ . Then:

$$2\delta_g \sin(\theta + \Delta\theta) + 2\delta_g \sin \theta = \lambda \rightarrow \\ \Delta\theta = \frac{\lambda \sin \theta}{2\delta_g \cos \theta} = \frac{B}{f} \tan \theta \quad \text{Equation 1.7}$$

The change $\Delta\theta$ is obtained if the satellite orbit shifts in the sky by the so-called critical baseline B_{cr} :

$$B_{cr} = \frac{\Delta\theta \cdot h}{\cos\theta} = \frac{B \cdot h}{f} \frac{\sin\theta}{\cos^2\theta} \quad \text{Equation 1.8}$$

where h is the height of the satellite orbit.

Notice however, that if both the off-nadir angle increases by $\Delta\theta$ and the radar central frequency shifts (decreases) by B the grating is fully visible, since the perspective change compensates for the frequency change (the wave number shift concept). If incident angle and observation frequency change at the same time by the correct amount, the same ground grating is observed. If only one grating existed in the image, say at the wave number A_0 , it would generate a return only if illuminated with different temporal frequencies f_1 in one image and f_2 in the other image. The interferogram would show fringes at the frequency $f_2 - f_1 = \Delta f$.

$$\Delta f = -\frac{f \Delta\theta}{\tan\theta} = -\frac{f B_n \cos\theta}{h \tan\theta} \quad \text{Equation 1.9}$$

Now, the observation that the reflectivity of the terrain is the combination of the effects of different gratings is equivalent to the expansion of the terrain reflectivity in a Fourier series. Distributed scatterers have independent Fourier components and therefore the amplitudes of the gratings that compose the reflection will be independent random variables. In interferometric imaging, multiplying two images corresponds to multiplying the amplitudes of the component gratings. The average of the product of the amplitudes of two independent gratings is zero. If the reflecting grating is the same for the two images (because it was illuminated with different temporal frequencies in the two images) the amplitude of the product is non-zero, and the beat will have the frequency Δf (the fringe). The progressive increase of the baseline and therefore the progressive change in viewing angle implies that the fringe frequency $|\Delta f|$ increases to be finally greater than the signal bandwidth thus making the fringe not observable. Larger baselines thus produce fringes with increasing frequency. The gratings that do not find corresponding elements with which to interfere in the other image (the non-cooperating components) contribute to the noise. Finally, no grating at all will be visible in both images for a baseline greater than the critical one: in this case, distributed scatterers will be unable to generate interferograms. If the gratings that do not appear in both images are removed with proper filtering, the resolution of the interferogram is unchanged even if the single images have a reduced resolution. However, the level of the noise in the interferogram will be lower.

Isolated point scatterers interfere properly until the baseline exceeds the directivity of the scatterer itself. If the scatterer is small, this directivity is very low, i.e. it will scatter over a wide range of directions.

In conclusion, natural scatterers are the superposition of point scatterers located at random positions within the resolution cell with random amplitudes; using band limited radar, we see only their spatial frequency

components (the gratings) with wave numbers in the interval corresponding to $(\Lambda_{\min} \div \Lambda_{\max})$, depending on the incident angle and the illumination frequency. To have a good interferogram with distributed scatterers and high baselines, suitable range filtering should be applied.

1.1.4 Statistics of the return

For the **central limit theorem** to hold, the phase and quadrature components of the return, and superposition of many independent elementary scatterers, gratings or point scatterers, are independent Gaussian random variables, with variance dependent on the terrain reflectivity. The amplitude of a single pixel in a SAR image (the modulus of the return) then has a Rayleigh distribution. Its amplitude squared (or intensity) has a Laplacian distribution with mean \bar{I} [Bamler98A]. The value of \bar{I} in a pixel of horizontal coordinates r (range), a (azimuth) is dependent on the local reflectivity of the terrain characterised by a non-dimensional parameter, σ^0 , times the inverse of the sine of the slope of the terrain, to incorporate foreshortening effects that brighten any surfaces that verge towards the satellite.

The value of σ^0 decreases for increasing values of the off-nadir angle θ , and depends on the terrain cover [Laur98]. In order to correctly estimate σ^0 , it is necessary to average the value of the intensity (the amplitude squared) over several pixels that should have the same statistics.

$$\hat{\sigma}^0 = \frac{1}{N} \sum_{i=1}^N |u_i|^2 \quad \text{Equation 1.10}$$

The amplitude of each pixel being a random variable, the SAR images are covered in speckle. However, it has to be understood that for repeated acquisitions of a stationary object the speckle ‘noise’ remains the same, which is different from other kinds of random noise. To remove the randomness of the amplitudes, in order to have a good estimate of σ^0 , the amplitude squares of several neighbouring pixels should be averaged. The Parseval relation tells also that the same results of speckle reduction can be achieved by reducing the azimuth and/or the range resolution of each image, by a factor of say $N_r \times N_{az}$ (where N_r is the **number of range looks** and N_{az} the **number of azimuth looks**) with as many disjoint frequency and/or wave number filters and then averaging incoherently the amplitudes of the $N_r \times N_{az}$ different images thus obtained. The resulting average estimate of $\hat{\sigma}^0$, will have a reduced variance by a factor of $N_r \times N_{az}$ compared with σ^0 .

The formula that gives the dispersion of the estimate, however, depends also on the random noise superposed on the data that increases the dispersion of the estimate. With very noisy data (high signal-to-noise ratio, or *SNR*), the amplitude is due to noise more than to signal, and therefore the dispersion of the combined estimates increases. It is usually given in terms of the **Equivalent Number of Looks (ENL)**:

$$ENL = \frac{(\sigma^0)^2}{E\left[\left(\hat{\sigma}^0 - E\left[\hat{\sigma}^0\right]\right)^2\right]} = \frac{(\sigma^0)^2}{E\left[\left(\hat{\sigma}^0\right)^2\right] - E\left[\hat{\sigma}^0\right]^2} = \frac{N_r \times N_{az}}{\left(1 + \frac{1}{SNR}\right)^2} \quad \text{Eq. 1.11}$$

Radiometric resolution is another parameter used to characterise the image quality and therefore the amount of speckle on the data:

$$\begin{aligned} R_{dB} &= 10 \log_{10} \left(1 + \left(1 + \frac{1}{SNR} \right) \frac{1}{\sqrt{N_r \times N_{az}}} \right) \\ &= 10 \log_{10} \left(1 + \sqrt{\frac{1}{ENL}} \right) \approx \frac{4.3}{\sqrt{ENL}} \end{aligned} \quad \text{Equation 1.12}$$

1.2 Interferometric images: coherence

In the case of interferometric images, the amplitude $|v|$ of each pixel of the interferogram is proportional to the product of the amplitudes $|u_1|, |u_2|$ of the two initial images, and its phase ϕ is equal to their phase difference. We have seen that SAR image pixels are the realisation of random processes and therefore we can expect the amplitudes of the interferogram to fluctuate severely even in the most favourable case of no temporal decorrelation (i.e. the target did not change between the two takes) and zero baseline (i.e. there is absolutely no change in the off-nadir angle). Therefore, the phase noise changes from pixel to pixel due to the different impact of the random noise superposed on the random amplitudes of the pixels. Pixels with weak returns will show more dispersed interferometric phases; strong and stable scatterers will yield more reliable phases. In addition, there are important changes between the two acquisitions: temporal, due to the change in the off-nadir angle, and due to random noise. We now define the measure of this change γ , the **coherence** of the two SAR images (also called the **complex correlation coefficient**):

$$\gamma = \frac{E[u_1 u_2^*]}{\sqrt{E[|u_1|^2] \sqrt{E[|u_2|^2]}}} = \frac{E[v]}{I} ; |\gamma| < 1 ; \angle \gamma = \phi_0 \quad \text{Eq. 1.13}$$

where E is the expected value of the random variable x

The argument of γ, ϕ_0 , is equal to the expected value of the interferometric phase. Obviously γ is a function of the expected values that we have to estimate from the data. Using N independent image samples, the usual estimator is:

$$\hat{\gamma} = \frac{\sum_{i=1}^N u_{1i} u_{2i}^*}{\sqrt{\sum_{i=1}^N |u_{1i}|^2} \sqrt{\sum_{i=1}^N |u_{2i}|^2}} \quad \text{Equation 1.14}$$

Examples of tandem coherence images of the Etna volcano are shown in Figure 1-2.

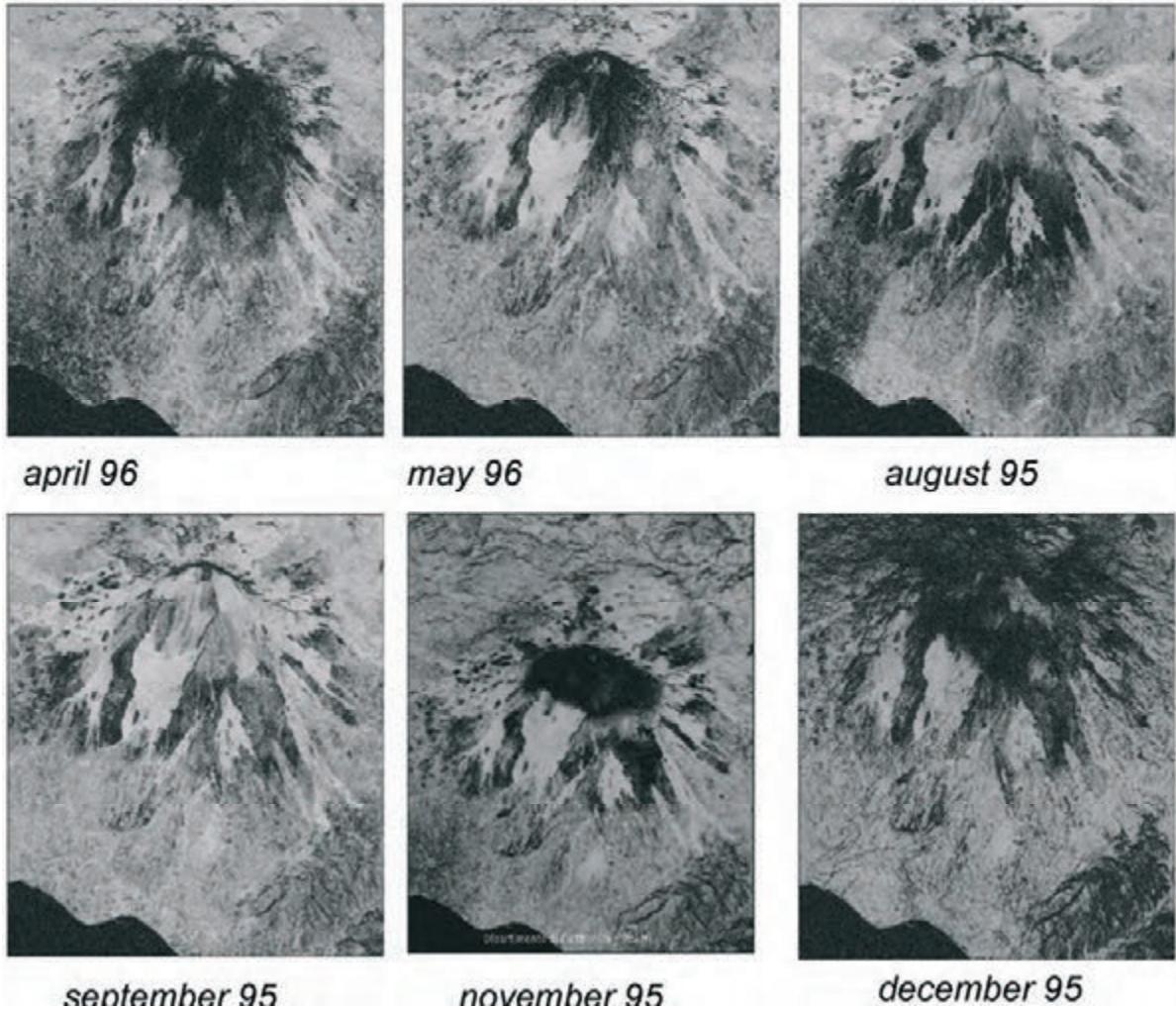


Figure 1-2: Tandem coherence of Mount Etna in different seasons

1.2.1 Statistics of coherence estimators

Estimation of the coherence is particularly difficult when its value is low; in fact, the estimator in Equation 1.15 is biased away from 0, as it is clear that if $N=1$, the estimate of the coherence is 1. For $N > 4$, and if $\gamma = 0$:

$$E[\hat{\gamma}] \approx \sqrt{\frac{\pi}{4N}} > 0 \quad \text{Equation 1.15}$$

The probability density of this estimate has been found [Touzi99]; the Cramer Rao Lower Bound (CRLB) for the dispersion is:

$$\text{var}(|\hat{\gamma}|)_{CR} = \frac{(1 - |\gamma|^2)^2}{2N} \quad \text{Equation 1.16}$$

The interferometric phase can be estimated as:

$$\hat{\phi} = \angle \hat{\gamma} = \angle \sum_{i=1}^N u_{1i} u_{2i}^* \quad \text{Equation 1.17}$$

We notice that using this estimate, weaker, noisier pixels have less influence on the final estimate. This estimate also shows that the interferometric phase is usually based on local averages. The value of N , i.e. the number of independent pixels generally used to estimate the coherence, ranges from 16 to 40. This limits the bias, but has the disadvantage of making local estimates of the coherence impossible since it is averaged on areas of thousands of metres square. This implies that the spatial resolution of an interferogram is intrinsically lower than that of a single SAR image, since the phases corresponding to single pixels are not necessarily reliable. In cases of low coherence (say 0.1), the number of looks to be averaged should increase up to 100 [Eineder99].

The Cramer Rao bound on the phase variance is:

$$\sigma_{\hat{\phi}} = \frac{1}{|\gamma|} \frac{\sqrt{1 - |\gamma|^2}}{\sqrt{2N}} \quad \text{Equation 1.18}$$

This bound can be assumed as a reasonable value of the phase estimator variance for N big enough; say, greater than 4 [Rosen00].

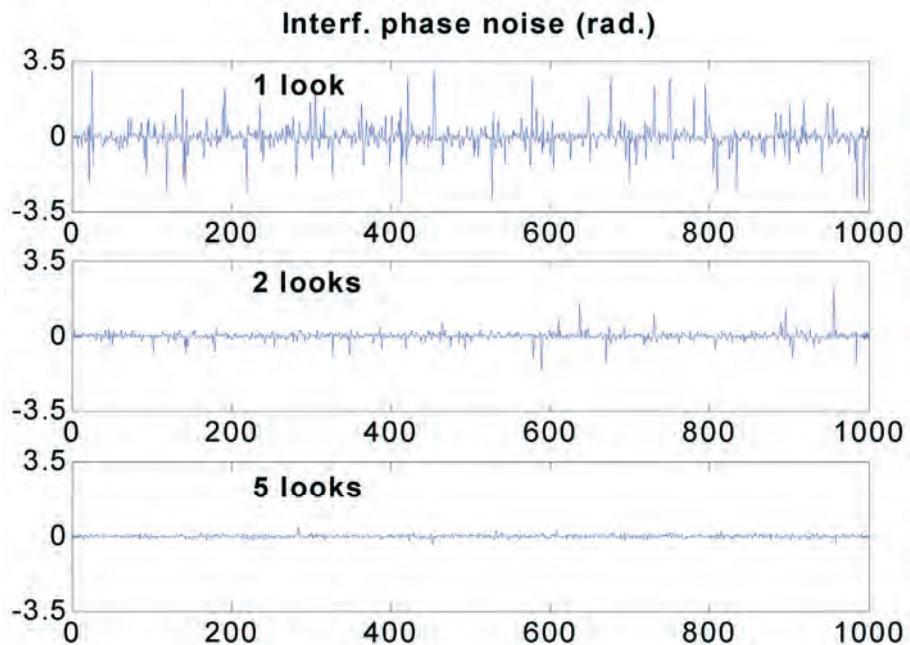


Figure 1-3: Argument of the coherence (estimated interferometric phase) as a function of the number of looks ($N = 1, 2$, and 5) for a sequence of 1000 pixels with random amplitude. The actual phase to be estimated is 0; the image SNR is 20 dB. Due to the randomness of the amplitude of the pixel, the dispersion of the phase estimates (for $N=1, 2$), is unacceptable even if the image SNR is very high. For $N > 4$, the phase dispersion decreases approximately with the square root of N .

For small N , as seen in Figure 1-3, it may happen that the amplitude of the pixel is low and thus small amounts of noise can create wide phase deviations. This effect decreases sharply by increasing the number of looks [Lee94]. Notice also the presence in Equation 1.18 of the term $1/|\gamma|$; this term accounts for the interaction of the noise in one image with the noise in the other and becomes important as soon as the coherence gets low. Then, to reduce the phase variance to acceptable values, N should increase and then the resolution will drop.

It is important to stress that, when the dimensions of the window used to estimate the coherence increase, a phase factor could be considered to compensate for volumetric effects within the window, that are unaccounted for using the current topographical model. This phase correction, if unknown, could be determined in such a way as to maximise the resulting coherence; hence, this correction would increase the coherence bias ([Dammert99]). To avoid this problem we can use another estimator of the coherence, which is phase independent but also biased and with higher variance. This estimate is based on the intensities of the two images, rather than on their complex values [MontiGuarnieri97]:

$$|\hat{\gamma}|_q = \sqrt{\frac{\sum_i I_{1i} I_{2i}}{\sqrt{\sum_i I_{1i}^2} \sqrt{\sum_i I_{2i}^2}}} - 1 \quad \text{Equation 1.19}$$

The advantage of the use of amplitudes is that the argument of the coherence, i.e. the interferometric phase, is spatially varying and therefore it has to be removed before averaging. This entails that in rugged areas the coherence will be systematically underestimated if the above formula is used without corrections or optimisation. On the other hand, we have seen before that if we insert a phase correction in that formula in order to optimise the coherence value, we get a bias.

The joint probability density of amplitude and phase of an interferogram is [Lee94]:

$$\begin{aligned} pdf(|v|, \phi) &= \frac{2|v|}{\pi I(1-|\gamma|^2)} \exp\left\{-\frac{2|\gamma||v|\cos(\phi-\phi_0)}{I(1-|\gamma|^2)}\right\} \\ &\times K_0\left(\frac{2|v|}{I(1-|\gamma|^2)}\right) \end{aligned} \quad \text{Equation 1.20}$$

where $|\gamma|$ is the coherence
 K_0 is the modified Bessel function

The marginal distributions of the phase and of the amplitudes are:

$$pdf((\phi) = \frac{1 - |\gamma|^2}{2\pi} \frac{1}{1 - |\gamma|^2 \cos^2(\phi - \phi_0)} \times \\ \left(1 + \frac{|\gamma| \cos(\phi - \phi_0) ar \cos(-|\gamma| \cos(\phi - \phi_0))}{\sqrt{1 - |\gamma|^2 \cos^2(\phi - \phi_0)}} \right) \quad \text{Eq. 1.21}$$

$$pdf(|v|) = \frac{4|v|}{I^2(1 - |\gamma|^2)} I_0\left(\frac{2|v||\gamma|}{I(1 - |\gamma|^2)}\right) K_0\left(\frac{2|v|}{I(1 - |\gamma|^2)}\right) \quad \text{Eq. 1.22}$$

1.2.2 Impact of the baseline on coherence

The coherence has an important diagnostic power. Remember that two acquisitions at baselines equal to or greater than B_{cr} will entail a complete loss of coherence in the case of extended scatterers. This effect has been discussed in the previous sections and was intuitively shown in Part A as corresponding to the **celestial footprint** of an antenna as wide as the ground resolution cell. Then, unless the ‘non-cooperating’ wave number components (the useless parts of the spectrum of the signal) are filtered out, the coherence of the two acquisitions will decrease linearly, becoming zero when the baseline reaches B_{cr} . In other words, if we do not eliminate the non-cooperating components we have, on flat terrain:

$$\gamma_b = \left(1 - \frac{B_n}{B_{cr}} \right) \quad \text{Equation 1.23}$$

The components at each wave number of the illuminated spectrum will be imaged at frequencies f_1 in one take and $f_1 + \Delta f$ in the other take.

$$\Delta f = -\frac{f \Delta \theta}{\tan(\theta - \alpha)} = -\frac{f B_n \cos \theta}{h \tan(\theta - \alpha)} \quad \text{Equation 1.24}$$

where: B_n is the component of the baseline in the direction orthogonal to the range

h is the satellite height

$$\Delta \theta = \frac{B_n \cos \theta}{h} \quad \text{Equation 1.25}$$

These formulas show that the situation becomes more complex when the slope greatly changes with range and/or azimuth, since then the filter necessary to remove the ‘useless’ components becomes space varying. Finally, when the change of slope is so severe as to become significant within a resolution cell, then coherence loss is unavoidable, due to the ‘volume’ effect as will be discussed in section 1.4.2 [Fornaro01].

1.3 Power spectrum of interferometric images

The power spectrum of an interferometric image is composed of two parts: the coherent part generating the fringe, and the noise. In order to evaluate it, first we oversample both images in range by a factor 2:1; the spectrum of each image is then a rectangle, occupying half of the available frequency range. Then, we cross-multiply the two images and thus we convolve the two spectra along the range axis. If the two images are totally incoherent, the resulting spectrum convolution of the two rectangles is triangular and centred in the origin. If there is a coherent component in the two images after cross multiplication, it generates a peak at zero frequency, if the baseline is zero. If the baseline is non-zero, the peak is located at the frequency Δf seen in a previous paragraph Equation 1.24. If we carry out the so called ‘spectral shift filtering’, removing the signal components corresponding to wave numbers visible in one image but not in the other (i.e. the components that do not contribute to the spectral peak at frequency Δf), the spectrum is still triangular, but centred at Δf .

1.4 Causes of coherence loss

1.4.1 Noise, temporal change

There are several causes that limit coherence in interferometric images: the most immediate to appreciate is the effect due to the random noise added to the radar measurement. From the definition of coherence, we get:

$$\gamma_{SNR} = \frac{1}{1 + \frac{1}{SNR}} \quad \text{Equation 1.26}$$

The changes with time of the scattering properties of the target should also be taken into account, which is done in the corresponding section in level 2 of this manual.

1.4.2 Volumetric effects

Another important effect that reduces the coherence with increasing baselines is the volumetric effect [Gatelli94, Rott96]. If we have stable scatterers having uniform probability density within a box of length δ_g along ground and height Δz , then the radar returns will combine with different phases as the off-nadir angle changes. The height of the box that entails total loss of coherence at baseline B_n is equal to the altitude of ambiguity h_a :

$$\Delta z_0 = h_a \frac{\lambda \cdot h \tan \theta}{2B_n} \approx \frac{9400}{B_n} \quad \text{Equation 1.27}$$

where $B_n = 250$ m, $\Delta z_0 = 38$ m (for ERS/Envisat)

For small values of Δz , the coherence loss due to volume is approximately [Gatelli94] :

$$\gamma_z = 1 - \frac{\pi^2}{6} \left(\frac{\Delta z}{h_a} \right)^2 \quad \text{Equation 1.28}$$

Similarly, for a Gaussian distribution of the terrain heights over a plane, with dispersion σ_z within the resolution cell, we get a phase dispersion of the returns equal to $\sigma\phi = 2\pi\sigma_z/h_a$. The coherence is:

$$\gamma_z = \exp\left(\frac{-\sigma_\phi^2}{2}\right) \cong 1 - \frac{\sigma_\phi^2}{2} = 1 - 2\pi^2 \left(\frac{\sigma_z}{h_a} \right)^2 \quad \text{Equation 1.29}$$

showing that the loss of coherence depends on the vertical dispersion of the scatterers.

Finally, as we have pointed out previously, the coherence γ is dependent on the baseline, unless proper filtering actions are taken. In the case of extended scatterers the coherence is always zero, when the baseline is greater than B_{cr} .

$$\gamma_b = \left(1 - \frac{B_n}{B_{cr}} \right); \quad B_n < B_{cr} \quad \text{Equation 1.30}$$

In conclusion, taking into account all the causes of coherence loss (noise, temporal change, volume, baseline) we get:

$$\gamma = \gamma_{SNR} \gamma_t \gamma_z \gamma_b = \frac{1}{1 + 1/SNR} \times \gamma_t \times \exp(-\sigma_\phi^2/2) \times \left(1 - \frac{B_n}{B_{cr}} \right) \quad \text{Eq. 1.31}$$

Further coherence reductions could be due to the imperfect alignment of the pointing of the antenna (Doppler centroid differences), that would cause a mismatch of the two azimuth spectra, similar to the effect of baseline decorrelation. The consequences could be reduced with a proper azimuth filtering of the two images.

2. Focusing, interferometry and slope estimate

In this chapter we study the effect of range and azimuth focusing on the final InSAR quality. We then derive an optimal technique for estimating the terrain topography, namely the slope. Finally, we give hints for optimising the processor, in order to get the best quality image, and at the same time we evaluate the impact of non-ideal processing.

2.1 SAR model: acquisition and focusing

The design of an optimal SAR focusing kernel for interferometric applications is based on the linear model for SAR acquisition and processing drawn in Figure 2-1.

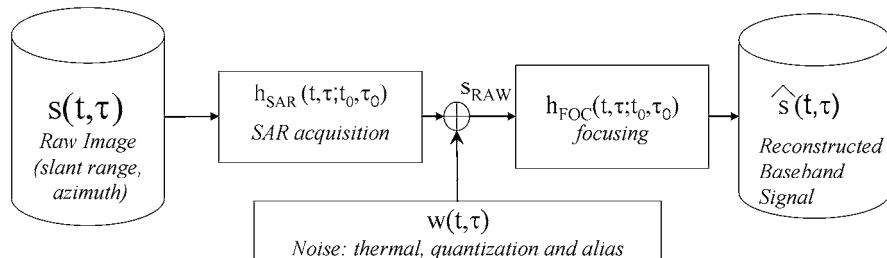


Figure 2-1: Equivalent simplified model of SAR acquisition and focusing

The **source** $s(t, \tau)$, is the scene reflectivity in the range, azimuth SAR reference (defined here by the fast-time, slow-time coordinates t, τ) of the Master image.

The **SAR acquisition** is modelled by the linear, space-varying impulsive-target response, $h_{SAR}(t, \tau; t_0, \tau_0)$, that depends on the target location t_0, τ_0 .

The **noise term**, $w(t, \tau)$, accounts for all the noise sources, mainly thermal, quantisation, alias and ambiguity.

The **SAR focusing** is represented by the space-varying impulse response $h_{FOC}(t, \tau; t_0, \tau_0)$.

The **output, Single Look Focused (SLC)** SAR image, is represented by the signal $\hat{s}(t, \tau)$.

2.1.1 Phase preserving focusing

The role of the SAR processor is to retrieve an estimate of the complex source reflectivity: $\hat{s}(t, \tau)$ that is the closest as possible to the true one: $\hat{s}(t, \tau) \approx s(t, \tau)$.

The optimisation would depend greatly upon the characterisation of the sources, of the noise, and the output. As an example, a processor optimised for measuring backscatter on a detected image would not necessarily be optimal for interferometry; likewise a processor optimised for interferometry with a distributed scene would not necessarily be optimal for applications on permanent scatterers.

Here we assume that the processor is intended for ‘classical’ interferometry, where the input is a distributed target modelled as a Gaussian circular complex process.

We make here two further assumptions, which are quite accepted in most SAR systems:

- 1) We approximate the SAR acquisition impulse response as stationary within the 2D support of the focusing operator
- 2) We assume that noise is also stationary, at least in the same signal support

These assumptions allows us to optimise the processor on a local base, and thus to approach the problem in the continuous-time domain or, equivalently in the frequency domain. In practice we ignore border effects. The estimate of the reflectivity, $s(t, \tau)$, can be achieved by **minimisation of the mean squared error (MMSE)** in Equation 2.1.

$$h_{FOC}(t, \tau; t_0, \tau_0) = \arg \min E[\hat{s}(t, \tau; t_0, \tau_0) - s(t, \tau; t_0, \tau_0)]^2 \quad \text{Equation 2.1}$$

The corresponding frequency domain expression is in Eq. 2.2, where we have defined f_a the azimuth frequency and we have used capital letters H and S for the Fourier Transforms of $h()$ and $s()$ respectively.

$$\begin{aligned} H_{FOC}(f, f_a; t_0, \tau_0) \\ = \arg \min E[\hat{S}(f, f_a; t_0, \tau_0) - S(f, f_a; t_0, \tau_0)]^2 \end{aligned} \quad \text{Eq. 2.2}$$

The minimisation involved in Equation Eq. 2.2 leads to the incorrelation between error and data, hence the optimal focusing in MMSE sense is the solution of Equation 2.3 (we have dropped the arguments for simplicity).

$$\begin{aligned} E[(H_{FOC} \cdot S_{RAW} - S) S_{RAW}^*] &= 0 \\ H_{FOC} &= \frac{E[S \cdot S_{RAW}^*]}{E[S_{RAW} \cdot S_{RAW}^*]} \end{aligned} \quad \text{Equation 2.3}$$

The optimisation would require knowledge of the power spectrum of the sources. According to our assumptions, source reflectivity is white (over a very large bandwidth) with **Power Spectrum Density (PSD)**

$$E[S(f, f_a) S^*(f, f_a)] = N_s \quad \text{Equation 2.4}$$

We then compute the PSD of the raw data by exploiting the model in Figure 2-1. The final expression of the optimal MMSE focusing kernel estimate is in Equation 2.5, where $N_w(f, f_a)$ is the PSD of the noise.

$$H_{FOC} = \frac{H_{SAR}^*}{|H_{SAR}|^2 + \frac{N_w(f, f_a)}{N_s}} \quad \text{Equation 2.5}$$

The result in Equation 2.5 is the well known MMSE Wiener inversion, that leads either to the matched filter, when $N_w \ll N_s$, or to the inverse filter, when $N_s \gg N_w$. In any case, *the focusing reference is phase matched to the SAR acquisition*, so that the overall end-to-end transfer function has zero phase (linear, if we add the proper delay): that is the reason why SAR focusing for interferometry is usually defined as ‘Phase Preserving’.

In a conventional SAR, we expect a constant PSD for both thermal and quantisation noise, whereas the azimuth ambiguity is weighted by the antenna pattern giving a coloured noise. A conventional rectangular antenna of length L causes the spectral weighting:

$$G(f_a) = \text{sinc}^2\left(\frac{L}{2v}(f_a - f_{DC})\right) \quad \text{Equation 2.6}$$

The azimuth ambiguity is due to the folding of the spectra due to PRF: the PSD of SAR data, the first left and right ambiguities, and the expected ambiguity-to-signal PSD are plotted in Figure 2-2.

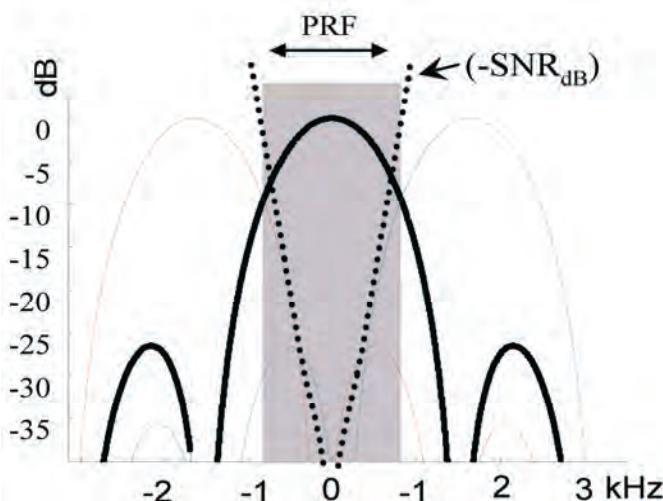


Figure 2-2: Azimuth Antenna Pattern in the frequency domain (dB) (continuous line) together with its left and right replica, folded by azimuth sampling (PRF). The ambiguity-to-signal ratio is plotted as the dotted line.

We notice that for most of the azimuth bandwidth the signal is much larger than the ambiguity, so that the optimal focusing in Equation 2.5 is very close to **the inverse filter**. This is quite different from the matched filter used in RADAR detection and communications; in fact the matched filter comes out from the maximum likelihood optimisation for isolated targets, and not for a continuous distribution of scatterers. Finally, we observe that in most commercial Phase Preserving processors, proper **spectral shaping windows** are applied in order to reduce the side lobes in presence of non-stationary

targets (like strong point scatterers or man made artefacts). These shapes should then be equalised for interferometric processing on a nearly homogenous scene.

2.1.2 CEOS offset processing test

The CEOS phase preserving test has been introduced to check the quality of the space-domain varying implementation of the processor. SAR processors usually approach a space-varying transfer function by block-processing, and adapting the processing parameters in each block. The ‘CEOS offset’ tests the self consistency of such approach, in particular checking that the same transfer function is implemented at a specific range and azimuth, disregarding the relative position with respect to the block [Bamler95A, Rosich96]:

“Process two SLCs from the same raw data set and with the same orbit, but offset by 100 lines in azimuth and 100 sample in range. The interferogram formed from these two properly coregistered SLCs should ideally have a constant phase of zero and thus reveals processor induced artefacts.”

The test must be passed with the following results:

- Mean of interferogram phase $\leq 0.1^\circ$
- Standard deviation of interferogram phase $\leq 5.5^\circ$ (corresponding to a reduction in coherence $\leq 0.1\%$)
- No obvious phase discontinuity at the boundaries of processing blocks

The test does not provide a reference Doppler, however it is wise to perform it at the largest Doppler foreseen for the mission. Moreover, the azimuth shift should not be given when testing ScanSAR modes.

Clearly the test provides only a necessary condition, for example any range invariant operator would pass it, and the same for a one-dimensional operator. Hence, it is necessary to combine it with a measure of the actual focused quality.

We also notice that further self-consistency tests could be implemented, such as comparing the processing of blocks of different sizes.

2.2 Interferometric SAR processing

So far we have considered the design of an optimal focusing processor as the one that provides the most consistent reconstruction of the source reflectivity. However, in processing a specific interferometric pair, some further processing specifically matched to that image pair can be performed to provide the best quality final interferogram.

The simplified design of the system and the processing in the interferometric case is shown in Figure 2-3. Here again we assume a distributed target, as we refer to conventional interferometry. We observe that the source reflectivity is affected by a noise contribution, namely ‘scene decorrelation noise’ that will not be noticed in processing a single channel (see Figure 2-1). With this warning in mind, we represent this contribution as an additive

noise source. Scene decorrelation may arise from change in the scattering properties between the two acquisitions, from volumetric effects or baseline decorrelation [Zebker92]: for repeat-pass interferometry this may be indeed the strongest noise.

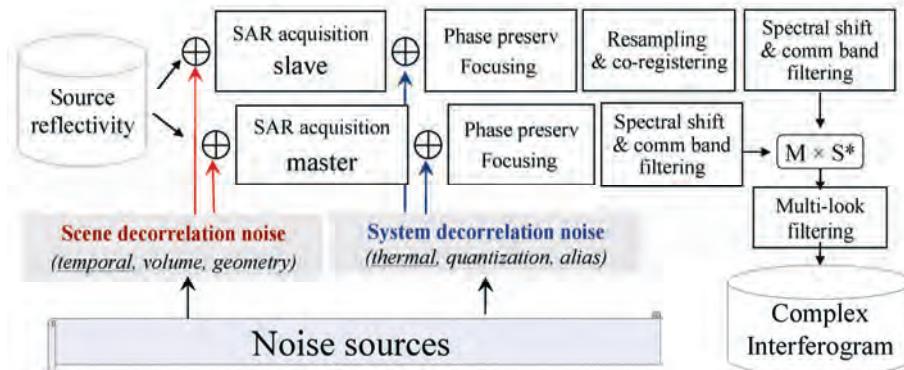


Figure 2-3: Schematic block diagram for the generation of a complex interferogram in the single-baseline case. The scene decorrelation noise is not noticeable on the single image.

The SAR acquisition channels for master and slave are not identical to the single channel shown in Figure 2-1, as here we must account for the different acquisition geometry that interferometry looks for. Following the discussion in section 1.1.3 we need to replace the reflectivity $s(t, \tau)$ by its modulated version:

$$\begin{aligned} s_M(t, \tau) &= s(t, \tau) \cdot \exp\left(+ j2\pi \frac{\Delta f}{2} t\right) \\ s_S(t, \tau) &= s(t, \tau) \cdot \exp\left(- j2\pi \frac{\Delta f}{2} t\right) \end{aligned} \quad \text{Equation 2.7}$$

where M and S refer to the master and slave respectively.

We have approximated the case of a constant sloped terrain, leading to the total frequency shift Δf expressed in Equation 1.19 (notice that the equal repartition of the total spectral shift into two contributions of $\Delta f/2$ in each channel is purely conventional). The case of space-varying slopes will be approached in a later section.

2.2.1 Spectral shift and common band filtering (revisited)

The object of the processing block named: ‘spectral shift and common band filtering’ in the diagram of Figure 2-3 is to extract only those contributions correlated in both channels, and likewise to remove all the decorrelated contributions. Let us refer to the master and slave phase-preserving complex focused images as $s_M(t, \tau)$ and $s_S(t, \tau)$ respectively.

We again have to face the problem of an optimal estimate in the MMSE sense, but now the task is to filter the master instead of focusing, thus getting an image as close as possible to the slave:

$$\hat{S}_{S|M}(t, \tau) = h_M * s_M \approx s_S(t, \tau) \quad \text{Equation 2.8}$$

where $\hat{S}_{S|M}$ is the reconstruction of the slave image given the master, and h_M is the reconstruction filter to be designed (the ‘spectral shift and common band’ block in Figure 2-3).

In the same way, a slave reflectivity would then be reconstructed from the master.

The design of filter h_M is accomplished by minimising the MMSE in the bidimensional frequency domain:

$$\begin{aligned} \hat{S}_{S|M} &= H_M \cdot S_M \\ H_M &= \arg \min_H E \left[\left| \hat{S}_{S|M} - S_M \right|^2 \right] \end{aligned} \quad \text{Equation 2.9}$$

This minimisation leads to the incorrelation between the error and data:

$$E[(\hat{S}_{S|M} - S_M) \cdot S_M^*] = 0 \quad \text{Equation 2.10}$$

Eventually we derive the formulation of the reconstruction filter:

$$H = \frac{E[S_S S_M^*]}{E[S_M S_M^*]} \quad \text{Equation 2.11}$$

This filter keeps the most correlated information. However, Equation 2.11 does not hold for non-zero spectral shifts ($\Delta f \neq 0$), i.e. it does not hold for different modulation of the sources that come out of the interferometric acquisition geometry (see section 1.1.3). In this case we have to shift the spectra of the two signals prior to correlating them, and this leads to the following result:

$$\begin{aligned} H &= \frac{E[S_M(f + \Delta f / 2, f_a) S_S^*(f - \Delta f / 2, f_a)]}{E[S_S S_S^*]} \\ &= \frac{H_{EM}^*(f - \Delta f / 2, f_a) H_{ES}(f + \Delta f / 2, f_a)}{|H_{EM}(f, f_a)|^2 \left(1 + \frac{|H_{FOCM}(f, f_a)|^2 N_{wM}(f, f_a)}{N_s} \right)} \end{aligned} \quad \text{Eq. 2.12}$$

where: H_{EM} and H_{ES} represent the end-to-end transfer functions (acquisition + focusing, as shown in Figure 2-1) of the master and slave channels respectively
 $N_{wM}(f, f_a)$ is the equivalent noise PSD on the raw master image
 H_{FOCM} is the focusing transfer function of the master

Be warned that the inversion involved in Equation 2.16 is meaningful only in the bandwidth of the end-to-end transfer function (out of this band there is

indeed no signal), or better in a smaller bandwidth to account for non-linearity of the on-board filter.

In a usual SAR system, the end-to-end transfer function is close to a band-pass in both range and azimuth, with the central frequency equal to the transmitted frequency f_0 in range and to the Doppler centroid in azimuth. Therefore, the numerator in Equation Eq. 2.12 has the fundamental role of keeping only the contributions in the common bandwidth (after providing spectral alignment), whereas the role of denominator is just to filter out-band noise.

2.3 DEM generation: optimal slope estimate

Let us derive here the estimate of the spectral shift Δf , by assuming the 1-D case (along range). We notice that Δf is related to the ground range slope, hence this estimate is what interferometry looks for in DEM-generation oriented applications.

The Maximum Likelihood estimator of the interferometric phase has been derived in [Tebaldini05] for the case of a distributed target, and properly accounts for both the non-stationary local topography and the discrete (sampled) case. However, in the particular case of a constant slope we can get the same result by a minimising the squared error in the frequency domain.

Let us assume the MMSE estimate of the slave SLC, obtained by processing the master with the filter computed in Equation Eq. 2.12. The error between the estimate and the given master image is:

$$E_{S|M}(f) = (\hat{S}_{S|M}(f) - S_S(f + \Delta f)) \quad \text{Equation 2.13}$$

We estimate the local spectral shift by imposing the minimisation of the integrated error, in the frequency domain:

$$\hat{\Delta}f = \arg \min_{\Delta f, H} \int |E_{S|M}(f)|^2 W_{S|M}(f) df \quad \text{Equation 2.14}$$

where the whitening term $W_{S|M}(f)$ provides the optimal weight, that should be inversely proportional to the variance of the error itself, $\sigma_{ESM}^2(f)$.

The schematic block diagram of the optimal estimate of the frequency shift is drawn in Figure 2-4.

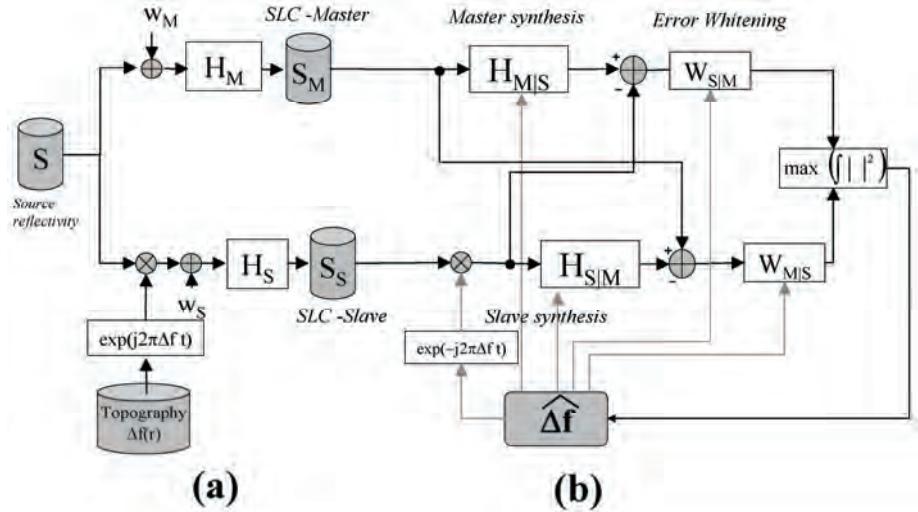


Figure 2-4: Schematic block diagram for (a) the interferometric model and (b) the optimal (MMSE) estimate of the local slope (or frequency shift, Δf)

The optimal expression of the weights is then computed as follows:

$$W_{S|M}(f) = \frac{A}{E\left[\left|\hat{S}_{S|M}(f) - S_S(f - \Delta f)\right|^2\right]} = \frac{A}{\sigma_{ESM}^2(f)} \quad \text{Eq. 2.15}$$

$$A = \int \sigma_{ESM}^{-2}(f) df$$

Note that the scheme assumes all the spectral shift is in the slave image. Moreover we are considering the 1-D case (along range). The master and slave reconstruction filters, from Equation Eq. 2.12, become:

$$H_{S|M} = \frac{H_{EM}^*(f, f_a) H_{ES}(f + \Delta f, f_a)}{H_{EM} H_{EM}^* \left(1 + \frac{\sigma_{wM}}{\sigma_s}\right)} \quad \text{Equation 2.16}$$

$$H_{M|S} = \frac{H_{ES}^*(f - \Delta f, f_a) H_{ES}(f, f_a)}{H_{ES} H_{ES}^* \left(1 + \frac{\sigma_{wS}}{\sigma_s}\right)}$$

where σ_{wM} , σ_{wS} , and σ_s represent respectively the standard deviation of the noise on the master image, the noise on the slave image, and the source reflectivity, and we assume that the inversion is applied only in the master and slave's bandwidths.

The computation of the error variance, $\sigma_{ESM}^2(f)$, can be performed by exploiting the interferometric model, the error definition in Equation 2.13, and the master synthesis expressed by Equation 2.9 and Equation 2.16.

Eventually we achieve the result expressed in Eq. 2.17, where we have assumed the same transfer function for the master and the slave end-to-end channel, $H_e(f)$.

$$\sigma_{ESM}^2(f) = (\sigma_{wS}^2 + \sigma_s^2) \left(|H_E|^2 - \sigma_s^4 \frac{|H_E^*(f - \Delta f) H_E(f)|^2}{(\sigma_{wM}^2 + \sigma_s^2)(\sigma_{wS}^2 + \sigma_s^2) |H_E|^2} \right) \quad \text{Eq. 2.17}$$

This equation can be further rearranged to evidence the weighting factor γ , defined in Equation 2.18, which – not surprisingly – corresponds exactly to the coherence of the interferometric pair.

$$\gamma^2 = \frac{\sigma_s^4}{(\sigma_{wS}^2 + \sigma_s^2)(\sigma_{wM}^2 + \sigma_s^2)} \quad \text{Equation 2.18}$$

We eventually compute the whitening factor by combining Eq. 2.17 and Equation 2.18 and using them in Eq. 2.15 to get the final frequency shift estimator:

$$\begin{aligned} \hat{\Delta f} &= \arg \min_{\Delta f} \int \frac{A(\Delta f) |H_E(f)|^2 |\hat{S}_{S|M}(f) - S_S(f - \Delta f)|^2}{|H_E(f)|^4 - \gamma^2 |H_E^*(f - \Delta f) H_E(f)|^2} df \\ A &= \int \frac{|H_E(f)|^2}{|H_E(f)|^4 - \gamma^2 |H_E^*(f - \Delta f) H_E(f)|^2} df \end{aligned} \quad \text{Eq. 2.19}$$

We can furthermore simplify this expression, by assuming that $H_E(f) = H_0$, constant in the whole range bandwidth, getting the result:

$$\hat{\Delta f} = \arg \min_{\Delta f} \int \frac{A(\Delta f)}{1 - \gamma^2 |H_E^*(f - \Delta f)/H_0|^2} |\hat{S}_{S|M}(f) - S_M(f - \Delta f)|^2 df \quad \text{Eq. 2.20}$$

to be windowed in the range bandwidth. This expression needs to be completed with its symmetric contribution (the estimate of the slave from the master), as shown in Figure 2-4.

Let us focus on the whitening term (the first factor in the integral in Eq. 2.20). For each search frequency, $\hat{\Delta f}$, we assume that correlated contributions come from $f > \hat{\Delta f}$, therefore we weight these contributions with the factor $1/(1-\gamma^2)$ that is always > 1 , and gets larger the higher the coherence. On the other side, the contribution outside the supposed common bandwidth, has unit weight. It is interesting to note that the Cramér Rao Bound would depend upon the width of the transition between the two regions ($f < \hat{\Delta f}$ and $f > \hat{\Delta f}$): a sharp transition (that would result from a wide extent of the data), would provide high accuracy, whereas a smooth transition (small range extent) would provide a coarse accuracy. [Tebaldini05] gives a simplification for $\gamma^2 \ll 1$; in that case Eq. 2.20 leads to:

$$\hat{\Delta f} = \arg \min_{\Delta f} \left(\int |\hat{S}_{S|M}(f) - S_M(f - \Delta f)|^2 df + \int |\hat{S}_{M|S}(f + \Delta f) - S_M(f)|^2 df \right) \quad \text{Eq. 2.21}$$

where the two symmetric contributions have already been added.

This would be the ML estimate for low SNR, thus quite robust with respect to model errors.

2.4 Noise sources

The impact of noise in interferometry is quite different from the single channel case due to the added source decorrelation, shown in Figure 2-3. In order to evaluate its impact with respect to the other contributions, we have evaluated the various signal-to-noise ratios for a typical ERS-Envisat case. More precisely, we have assumed:

- a typical Noise Equivalent σ_0 of -24 dB [Laur98];
- a mean reflectivity $\sigma_0 = -10$ dB (the 50% reference curve adopted for Envisat ASAR); and
- two different levels of scene decorrelation, roughly corresponding to ‘good’ (e.g. stable in time) and ‘very good’ cases ($\gamma = 0.7$ and $\gamma = 0.95$), this last achievable only on desert areas.

Results are plotted in Figure 2-5.

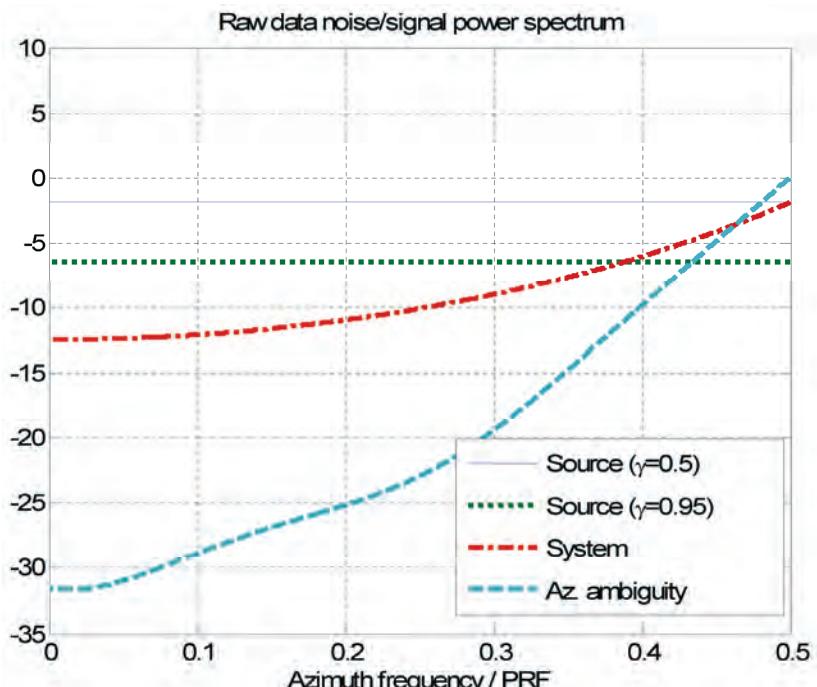


Figure 2-5: Ratio between Noise Power Spectrum and Signal PSD, estimated on the raw data, for the different kinds of decorrelation noise sources. Scene decorrelation appears to be the dominant effect of scene coherence up to $\gamma = 0.95$.

Note that we have assumed that data has been processed with the inverse filter, so that both system noise (thermal and quantisation) and azimuth ambiguity appear rather high-pass. Even though for most of the azimuth bandwidth the major contributor seems to be the scene decorrelation, the

best quality is simply achieved by maximising the ‘processed’ azimuth bandwidth, e.g. by performing antenna pattern deconvolution.

In conclusion, the **inverse focusing reference** (with some tapering at the bandwidth edges) is recommended for maximising interferometric quality in most cases (γ_{scene} up to 0.95), whereas a more complex scene-dependent optimisation is required for highly coherent scenes.

The advantage of performing antenna pattern deconvolution, compared for example with the use of a flat amplitude reference, can be approximated to a 3 dB improvement in SNR, as by compensating for the antenna pattern we increase the processed Doppler bandwidth from $\sim \text{PRF}/2$ (the antenna -3 dB bandwidth) to $\sim \text{PRF}$.

2.5 Processing decorrelation artefacts

So far, we have defined the optimal phase preserving focusing and interferometric processors by assuming perfect knowledge of the geometric and system-related parameters.

However, it may happen that, due to errors in these parameters or to the realisation of an approximate processor, the transfer function implemented in not exactly phase matched to the system. In these cases, the impact on the final quality can be expressed in terms of the coherence [Just94]. Let $H_{\text{SAR}1}$, $H_{\text{SAR}2}$ be the acquisition transfer function of the two interferometric channels (see Figure 2-1), and $H_{\text{FOC}1}$, $H_{\text{FOC}2}$ the two focusing operators, and $H_{\text{E}1} = H_{\text{IRF}1}H_{\text{FOC}1}$, the end-to-end transfer function of channel 1 (similarly for channel 2). Coherence can be expressed directly in the frequency domain by means of Parseval identity [MontiGuarnieri96B, Cattabeni94]. If we assume a totally correlated target and we ignore noise we get the processor coherence contribution:

$$\gamma_{\text{proc}} = \frac{\iint H_{\text{SAR}1}(f, f_a) H_{\text{SAR}2}^*(f, f_a) df df_a}{\sqrt{\left(\iint |H_{\text{SAR}1}(f, f_a)|^2 df df_a\right) \left(\iint |H_{\text{SAR}2}(f, f_a)|^2 df df_a\right)}} \quad \text{Eq. 2.22}$$

This contribution should then be multiplied by the other decorrelation sources (temporal, geometric etc.) to get the overall coherence [Rodriguez92, Villasenor92].

In theory, the integrals should be extended to the common support between the two datasets, both in range and in azimuth, to account for the filtering expressed in Eq. 2.12, and the result would be space-varying. In practice we will evaluate the impact of distortions by assuming the full range and azimuth bandwidth.

We discuss here the cases of an error in the Doppler rate, and an error in the estimate of co-registering parameters.

2.5.1 Examples of decorrelation sources

Let us assume that the phase artefact in SAR focusing is expressed by the (residual) pure phase transfer function $H_d(f) = \exp(j2\pi\phi(f))$ in both channels, one-dimensional in range. The coherence contribution can be estimated by exploiting Eq. 2.22, obtaining:

$$\begin{aligned}\gamma_{proc} &= \frac{1}{B} \int H_d(f - \Delta f) H_d^*(f + \Delta f) df \\ &= \frac{1}{B} \int \exp(j2\pi(\phi(f - \Delta f) - \phi(f + \Delta f))) df\end{aligned}\quad \text{Equation 2.23}$$

Note that no decorrelation results if the phase error is symmetric, or if the spectral shift is zero – whatever the distortion. As an example, an error in the replica quadratic phase coefficient leads to a defocusing, but not to a coherence loss, *provided that the same error is applied to both channels*.

In a general case, we can predict that the same phase distortion leads to no quality loss in azimuth, whereas it can give space-variant decorrelation (and phase bias) in range, according to the spectral shift.

However, if the phase distortion applies to one channel only, a decorrelation noise is introduced. This can be the case when a **fixed replica** is used to perform range focusing. Let us approximate the error as a pure quadratic phase, like in Equation 2.24, where k is the quadratic phase error.

$$H(f) = \text{rect}(f/B) \exp(jkf^2), \quad \text{Equation 2.24}$$

For ERS, we can expect a drift of $\pm 0.015\%$ with respect to the nominal value $k_o = 7.5 \cdot 10^{-12}$ rad/Hz². The coherence can be estimated by series expansion of $H(f)$, and assuming the worst case (no spectral shift), the correlation contribution of processing is expressed in Equation 2.25.

$$\begin{aligned}\gamma_{proc} &= \frac{1}{B} \int_{-B/2}^{+B/2} \exp(jkf^2) df = \frac{1}{B} \int_{-B/2}^{+B/2} \left(1 + jkf^2 - \frac{k^2}{2} f^4 \dots\right) df \\ \gamma_{proc} &\cong \left| \left(1 - \frac{k^2 B^4}{160}\right) + j \frac{k B^2}{12} \right|\end{aligned}\quad \text{Equation 2.25}$$

This gives a phase bias $\arg(\gamma_{proc}) \cong [(kB^2)/12] = 0.2$ rad and a decorrelation $1 - |\gamma_{proc}| = 0.3$. This decorrelation can be relevant for high quality processing.

Let now assume an **azimuth misalignment** Δ_t between the two images, e.g. a ‘residual’ T_F , expressed in Equation 2.26 on one channel (and zero phase on the other).

$$\begin{aligned}H_{T1}(f, f_x) &\cong \exp(j2\pi\Delta_t f_x) \\ f_{dc} - B/2 \leq f \leq f_{dc} + B/2\end{aligned}\quad \text{Equation 2.26}$$

In Equation 2.26 we have assumed a Doppler Centroid Frequency f_{dc} and a total bandwidth B (\leq PRF). This phase error leads to the coherence contribution shown in Equation 2.27.

$$\gamma_{proc} = \frac{1}{B} \int_{f_{dc}-B/2}^{f_{dc}+B/2} \exp(j2\pi\Delta_t f) df$$

$$= \text{sinc}(B\Delta_t) \exp(j2\pi f_{dc}\Delta_t)$$
Equation 2.27

The effect on the final interferogram is:

- a *loss of quality*, that depends on the **coherence amplitude**, i.e. a decorrelation: $1 - |\gamma_{proc}| = 1 - \text{sinc}(B\Delta_t) \equiv [1/6]\pi^2(B\Delta_t)^2$;
- a phase bias, equal to the coherence phase, $\angle\gamma_{proc} = 2\pi f_{dc}\Delta_t$.

The decorrelation imposes a limit on the maximum tolerable misalignment, for example if we accept $|\gamma_{proc}| < 0.95$ (the best case assumed in Figure 2-5), we get $\Delta_t \ll 0.15$ resolution cells. The phase bias is irrelevant *only* if it is constant over the processed data set (e.g. both misalignment and Doppler Centroid do not change), since the interferogram is usually known apart from a constant phase. Otherwise it can introduce significant phase artefacts: a misalignment of $\Delta_t = 0.15 / B$ would get a phase offset of 27° for a Doppler Centroid $f_{dc} = B/2$. This is quite important if, for example, two different looks should be coherently averaged [MontiGuarnieri99B].

Other examples in literature may be found in [Just94, Cattabeni94].

3. Advances in phase unwrapping

3.1 Introduction

One of the most difficult problems in interferometry is the extraction of the absolute phases from the available wrapped values. In fact, in order to estimate topography and motion of the terrain we need the absolute phases, whereas the data are phase differences between the two images, in their principal interval $\pm\pi$. Moreover, the wrapped phases are not available everywhere since we may have pixels without a significant radar return. And then, even if the phases are reliable, they may correspond to portions of the topography that are subject to alias and layover. This implies that, when using only one image pair, the operation of **phase unwrapping may be subject to uncertainties that cannot be solved**, unless we make hypotheses, usually statistical ones, on the structure of the underlying topography, or if we use additional information. In the section on multi-image phase unwrapping, we will see that this problem can have a unique solution if several image pairs with different baselines are combined. Obviously, this solution is available only where the target did not change during the entire span of the several surveys or it is seen simultaneously with multiple baselines and/or with multiple frequencies. Then, no hypothesis on the surface continuity is needed and the unwrapping can be carried out pixel by pixel. In this section we will approach the problem of phase unwrapping for image pairs, starting with simple cases (no or low noise, no aliasing) and then moving towards more difficult cases. The flat topography contribution is supposed to have been removed already.

In order to phase unwrap, we measure first the interferometric phase for all pixels and make an initial assumption that the measured values are reliable. Then, we add to each pixel phase value the integer multiple of 2π that is required to unwrap it. If we then scale the unwrapped phases to heights we get the topography in SAR coordinates. **Any assignment of integers is acceptable**, but then the topography may vary wildly from one pixel to the next. In other words, unlimited topographies exist, that honour the data.

Figure 3-1 shows an example of two-dimensional phase unwrapping on simulated data. The original topography is presented at left. The topography corresponding to the wrapped phase is presented at right.

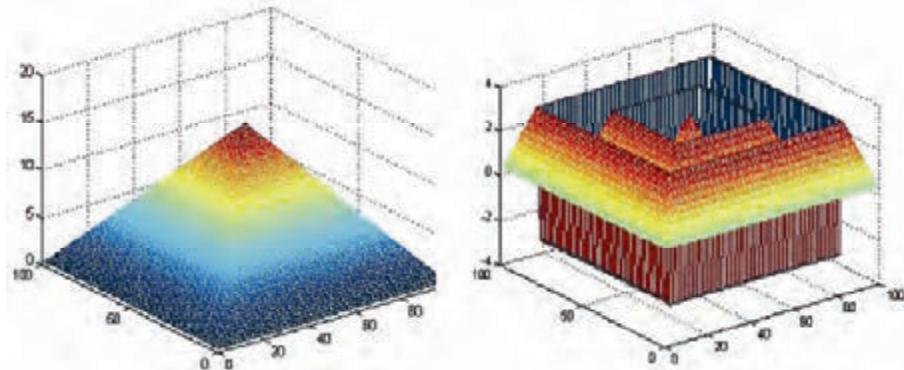


Figure 3-1: Phase unwrapping. Left: absolute phase, named ψ . Right: wrapped phase, named ϕ . Phase unwrapping regenerates the absolute phase, given the wrapped one (i.e. from right to left).

Problems of mono dimensional phase unwrapping are then exemplified in Figure 3-2: the unwrapping is carried out by supposing that the phase differences between neighbouring pixels are always under half cycle: whenever a phase difference greater than that is encountered, a cycle skip is surmised and compensated. In the case that the original phase difference is greater, this procedure generates an error.

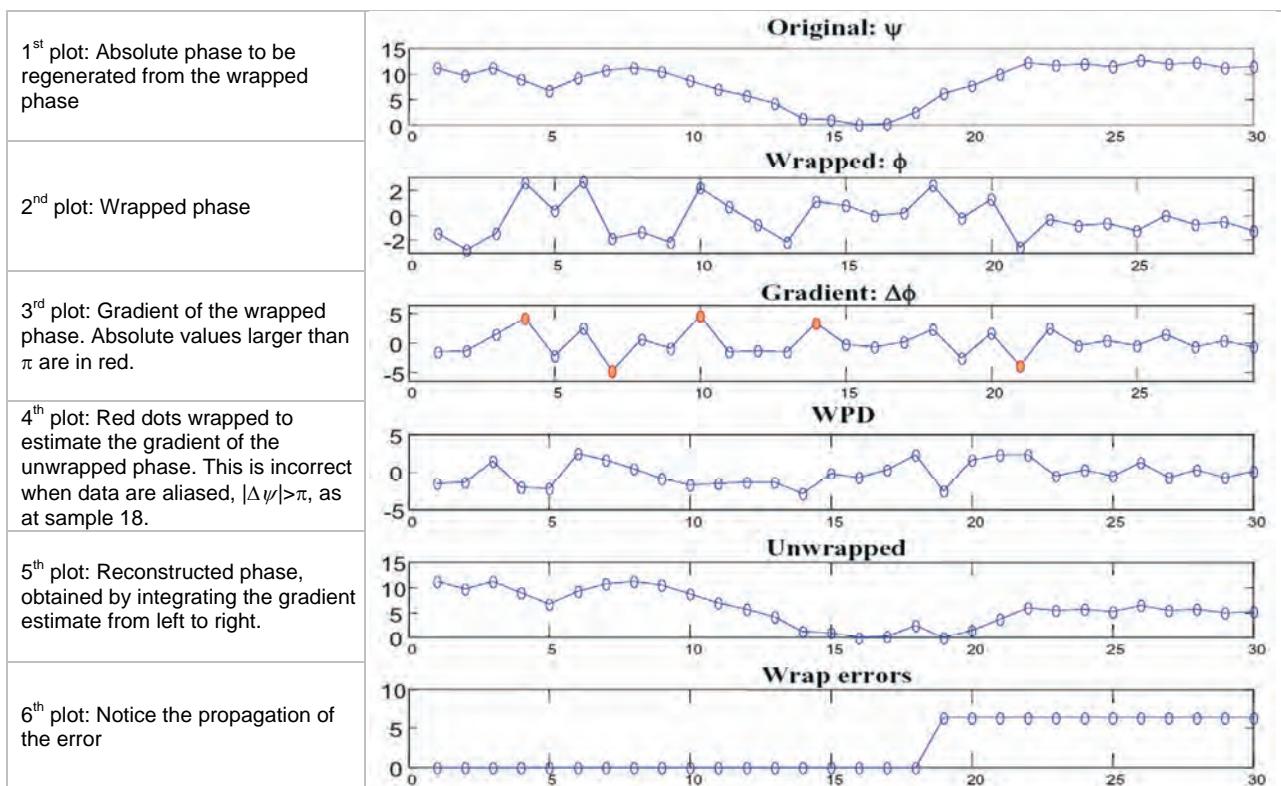


Figure 3-2: Mono-dimensional phase unwrapping carried out by ‘integration of the estimate’

3.2 Residues and charges

To get a unique solution from a single pair of images, besides assuming there is no noise, we have to further suppose that the initial surface was sampled without **alias**. The actual phase differences between neighbouring grid points are then less than π or, in terms of topography, the elevation difference between neighbouring grid points is smaller than $h_a/2$. This might be false in mountainous regions where layover is present in the resultant SAR data. We remember here that we can (and we should, as we shall see in the forthcoming sections) always remove the topography that is already known, to make the data as flat as possible. In other words, we should compensate for the local average slope of the terrain using *a priori* information of the topography, derived, for example, from lower resolution surveys. If there is no noise and no alias, then there is a true assignment of cycles to sum to the phases. The differences between the unwrapped phases of neighbouring pixels will then have a magnitude smaller than π and coincide with the **Wrapped Phase Differences (WPD)**. The WPD are the measured phase differences (Figure 3-3) of two neighbouring grid points (thus in the interval $\pm 2\pi$) re-wrapped by adding $\pm 2\pi$ (if necessary) so that they stay in the interval $\pm \pi$. From the WPD we can, for no noise and no aliasing, retrieve the unwrapped values.

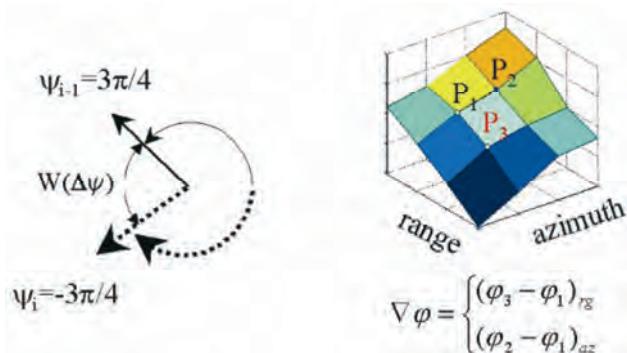


Figure 3-3: Left: The gradient is estimated by WPD. In the example, $\Psi_{i-1} = 3\pi/4$; $\Psi_i = -3\pi/4$; the gradient is then $\Delta\psi = -6/4 \pi$ wrapped to $WPD = -6/4 \pi + 2\pi = \pi/2$.

Right: In the 2-D case, the gradient is a vector with two components.

The problems with 2D phase unwrapping are exemplified in Figure 3-4 and Figure 3-5. The corruption of the WPD due to noise causes ambiguities.

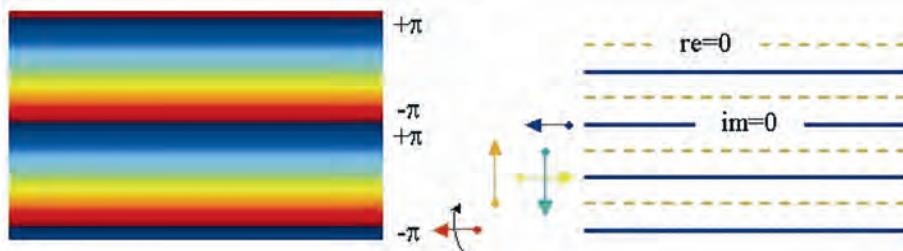


Figure 3-4: A constant phase slope for the noiseless case. Left: wrapped phase. Right: lines corresponding to zero crossings of the real and imaginary parts. These lines never cross ($modulus \neq 0$). However, when the slope increases, the distance between the lines decreases.

The problem is more easily understood in the continuous case. Suppose that we start from a discrete set of random complex numbers p_{ij} disposed on a square grid; then, we interpolate the real and imaginary parts of the p_{ij} 's, by convolving them, say, with a 2-D sinc function. We get a complex function $p(x,y)$ which is the sum of two continuous surfaces:

$$p(x,y) = \operatorname{Re}(x,y) + j \operatorname{Im}(x,y) \quad \text{Equation 3.1}$$

The corresponding sampled surface is:

$$p(x_i, y_j) = p_{ij} ; x_i = i\Delta ; y_j = j\Delta ; \quad \text{Equation 3.2}$$

The phase $\angle p(x,y)$ is well defined everywhere but at the points (say with coordinates (x_r, y_r)) where both $\operatorname{Re}(x,y)$ and $\operatorname{Im}(x,y)$ are zero, i.e. where the lines corresponding to the zeros of $\operatorname{Re}(x,y)$ ($\angle p(x,y) = \pm\pi/2$) cross the lines corresponding to the zeros of $\operatorname{Im}(x,y)$ ($\angle p(x,y) = \pi/2 \pm \pi/2$). These **phase contour lines cross** at the points (x_r, y_r) that are not necessarily on the sampling lattice [Schwartzman94]. Since the two continuous surfaces $\operatorname{Re}(x,y)$ and $\operatorname{Im}(x,y)$ are locally planar, only two possibilities exist, namely those circulating clockwise around (x_r, y_r) the phase either increasing or decreasing by 2π at each circuit (Figure 3-5). In the interferometric literature, following Goldstein who first studied this problem in the discrete case, we call these points **residues**, and **charge** refers to the sign of the curl of the phases. If we unwrap the phase by integrating the WPD following a closed path that encircles a residue, the result of the integration will be $\pm 2\pi$. Then, the *integration of the WPD is ambiguous if the phase contour lines cross*. Therefore, a random phase distribution does not allow a unique solution for phase unwrapping. Depending on the integration path, we will find infinite solutions, all potentially acceptable and none of them being any more significant than any other.

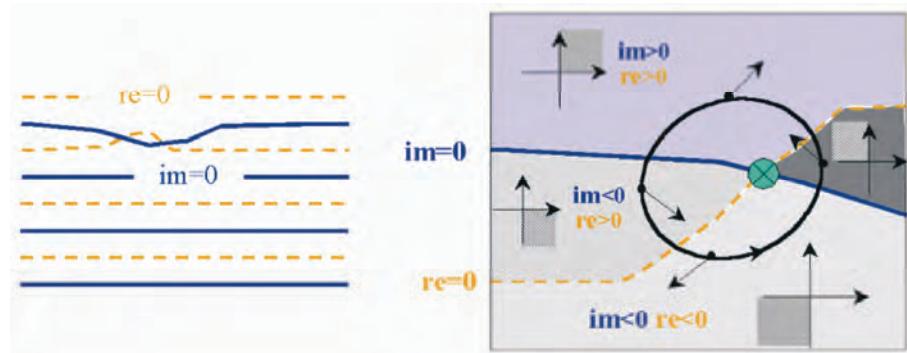


Figure 3-5: Left: In the presence of noise, amplitude may go to zero, and two neighbouring vortices with opposite charges are created. Right: If integration on a closed path embraces a vortex, a residual of $\pm 2\pi$ results at each circuit. In fact, whenever a zero line is crossed, the phase loses $\pi/2$.

If the phase contour lines cross, as will happen if the data are noisy enough, the vector field that has as components the WPD along azimuth and range is rotational. Then, the curl of this vector field is non-zero. If the continuous curl is non-zero and the total number of residues (both positive and negative) in a cell is non-zero, the **discrete curl** (the one measured on the discrete

lattice, if the sampling frequency is sufficient) will also be non-zero, as will be seen in the next section.

3.2.1 Effects of noise: pairs of residues, undefined positions of the ‘ghost lines’

We consider now the case of low noise and no alias: for instance when the complex values p_{ij} correspond to a constant slope and therefore the phases increase progressively along a direction parallel to the slope. The lines that correspond to zero imaginary part and those corresponding to zero real part are interleaved as seen in Figure 3-5. The spacing of the lines corresponding to the zeros decrease with increasing slope. However, if no crossings exist then no residues appear, and the unwrapping problem has a unique solution (if the sampling grid is fine enough that alias is absent).

Let us now consider an example containing some complex noise. The lines corresponding to the zeros of the real and imaginary parts now start to wiggle and, with increasing noise, may cross in two neighbouring points. We have a **pair of residues**, where both $\text{Re}(p)=\text{Im}(p)=0$; in these two points the total sum of signal and noise has modulus equal to zero (Figure 3-5).

The charges of the two residues have opposite signs, as visible in the figure, since in one case the phase around the residue increases clockwise and in the other case anticlockwise. If the noise level is increased, more and more pairs of residues appear, always with opposite charges (Figure 3-6 and Figure 3-7) [Rosen00].

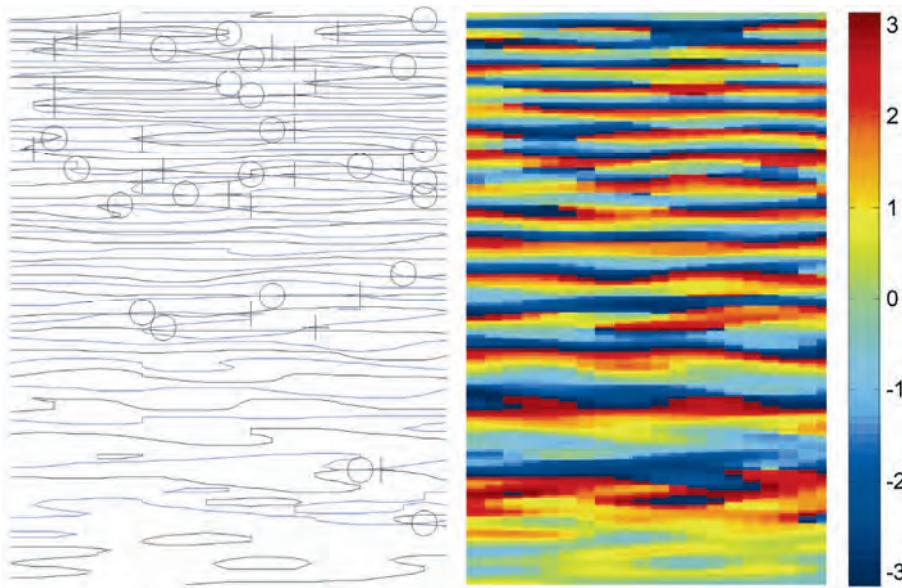


Figure 3-6: Residuals due to noise (SNR=8 dB) in the presence of a linear increasing slope. Left: residuals, marked with ‘O’ and ‘+’ and contour lines of zero-real, zero-imaginary parts. Right: phase field. The density of the residuals increases with the slope.

Now the unwrap problem is without a unique solution. If a phase integration path encircles only one of the two residues of any pair, 2π will be added (or subtracted) to the total phase and the final topography will depend on the integration path chosen. Any topography consistent with this noisy but continuous phase distribution cannot be continuous and should have an abrupt 2π discontinuity (e.g. a wall with height h_a) connecting two residues with opposite charge.

From what we have seen, the number of pairs of residues will increase both with the local slope (that makes the lines of zeros closer to each other) and with noise (that makes the lines of zeros wiggle more) (Figure 3-7).

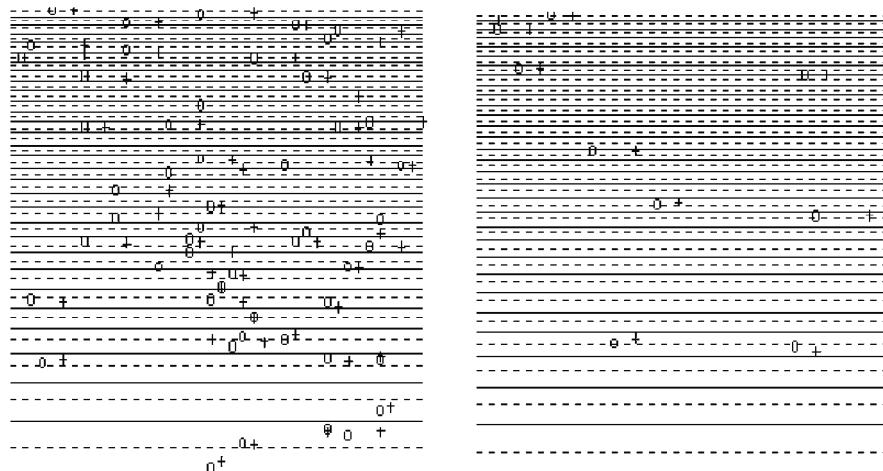


Figure 3-7: Residuals due to noise, as in Figure 3-6, but for different values of SNR: left: 8 dB, right: 4.5 dB.

It is interesting to remark that the gradient of the phase change induced by the couple of residues is in the direction opposite to that of the slope. In fact, if the noise is small and the lines of the zeros of real and imaginary parts barely cross (say are tangent), the gradient will keep the same direction, change sign, and point down slope. Therefore the effect of noise will always be that of reducing the apparent amplitude of the slope. This result has been noticed and explained by Spagnolini [Spagnolini95] and Bamler [Bamler98B].

Let us now consider what happens when we sample on a grid. If the local topography is smooth and thus there is no alias, no ambiguity occurs in the noiseless case. The effects of noise will be pairs of residues, which will be close to one another if the noise is small. The residues will be revealed by computing the sum of the WPD around a square of four neighbouring pixels of the grid (the **discrete curl**) (see Figure 3-8).

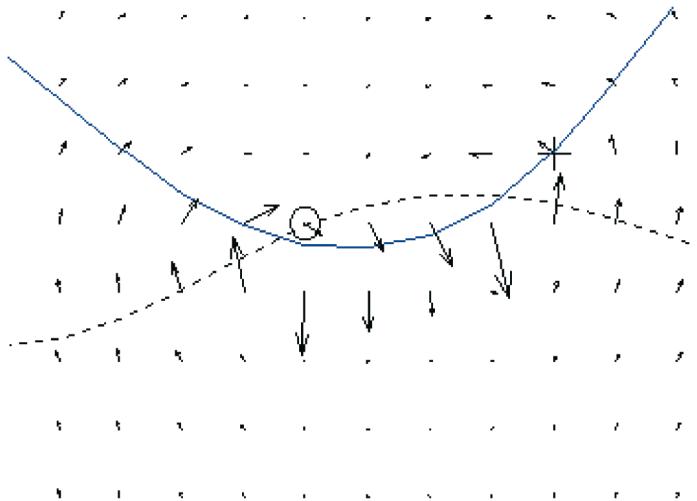


Figure 3-8: Gradient field in the neighbourhood of a dipole. Zero-real: continuous line. Zero-imaginary: dotted line. The gradient becomes very large approaching the vortices and its direction is opposite to the local slope.

The discrete curl will be zero in normal situations, and $\pm 2\pi$ if the four neighbouring pixels encircle a point where we have a crossing of the lines that correspond to zero real and imaginary parts. It will again be zero if two residues are encircled that have opposite charge.

If now we wish to find a topography consistent with this phase distribution, i.e. to unwrap it, any integration path should not encircle any single residue, but only pairs with opposite charge. The simple way to do that is to connect neighbouring residues with opposite charges to lines that are not to be crossed. These lines, which have no physical meaning and their location totally undefined but for their end points, are called '**branch cuts**' by Goldstein [Goldstein88] or '**ghost lines**' by Prati [Prati90A]. The first name corresponds mainly to the hypothesis that the residues are due to noise and the branch cuts correspond to lines that are totally undefined, apart from their end points. The second name is related to alias in the topography due to coarse sampling, and in this case the actual position of the ghost lines is well determined physically but is undetectable from the experiment. This will be better seen in the next section.

For increasing noise levels, the possible topographies become more and more complex, with meandering **walls** with height h_a that start from any positively charged residue and end in any negatively charged one. As said before, there is no way to 'detect' any h_a discontinuity from the wrapped phases because the phases are not affected. In principle, a good way to unwrap is not to cross any branch cut; but again, their locations are totally undefined and therefore *undeterminable from the phases*. However, if there is no alias and the residues are due only to noise, and if the noise is small, it is reasonable to couple together neighbouring residues with opposite charge.

Increasing the level of the noise without limit, we get a very high density of residues. They are asymptotically one in three of the lattice points for sampled independent Gaussian noise on a square lattice. The density is one in four on a triangular lattice. Thus, if we have a rather flat surface with a lot

of phase noise, the best solution to smooth the phases before unwrapping, averaging them and thus removing the effects of neighbouring residues with opposite charges.

3.2.2 Effects of alias: unknown position of the ghost lines

In SAR geometry, it may happen quite often that the slope of a hill facing the radar has, between two adjacent pixels, a variation of the elevation with range higher than $h_a/2$. Aliasing occurs since the WPDs are insensitive to any big elevation change and only measure the fractional part that is less than $h_a/2$. Thus, we are obliged to consider the case of aliased surfaces: as with noisy data, we have residues and thus we have again a failure of any simple technique to determine the unwrapped phase. If we initialise the process at any given point of the image and then integrate the WPD from there to everywhere else, crossing an elevation change in the topography that is greater than $h_a/2$ (a ghost line), we add 2π to the phases from this point on. Without alias and without noise, the sum of the WPD along any path would give us the same value as the unwrapped phase. In the case of alias, the discontinuity exists in the data but is visible only where it falls below $h_a/2$. So, suppose we start from a valley and ascend to the top of a hill, as in Figure 3-9. If we access this point following a path that does not cross any such discontinuity, we shall find the correct altitude. If, while climbing the hill, we cross a discontinuity without knowing it, the final altitude will be short by h_a .

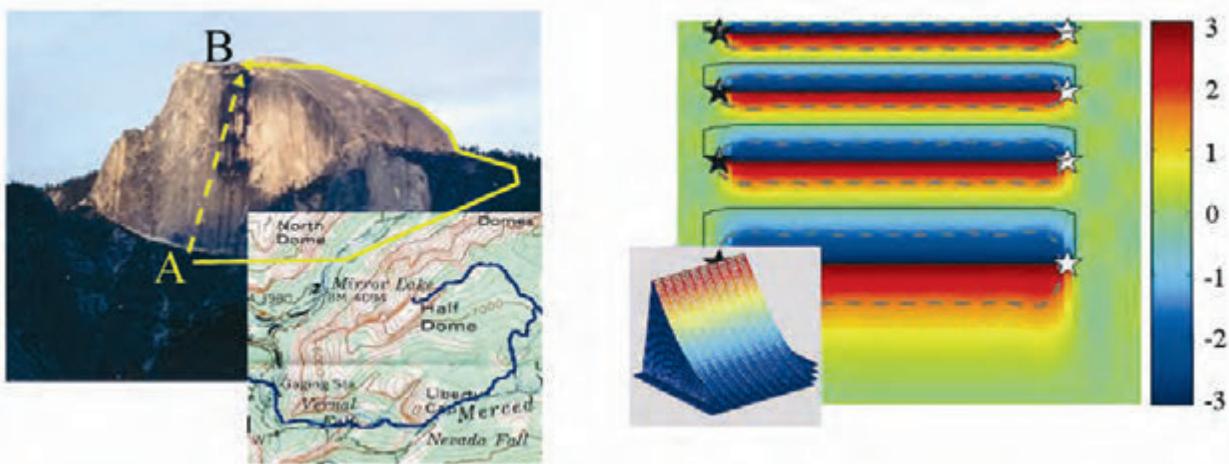


Figure 3-9: Phase due to topography. Left: A strong gradient causes a phase discontinuity that may lead to alias, depending on the baseline. Right: Wrapping a parabolic-like topography (shown in the shaded box). Pairs of residuals of opposite signs (marked with stars) are generated at the endpoints of the fringes.

How to avoid such a problem? We can detect the ends of the discontinuity since in these points the discrete curl of the WPD is non-zero. However, unless we use some *a priori* information of, say, geological nature, there is no way to tell where the discontinuity belongs, in between these two points (hence, the name ‘ghost lines’). In general, we may have many discontinuities in the image; we see only their end points, as in Figure 3-9

(right), and we do not know which residue to connect with which other (even if we know that we have to connect residues of opposite charges to begin and end any discontinuity). As long as we place discontinuities along ghost lines of height h_a beginning and ending at residues with the proper charge, we get topographies consistent with the phases. However, we have an unlimited number of solutions. In the case of alias, however, we can use *a priori* information. For example, alias is more likely in foreshortened areas, where the amplitude increases. The ghost line is likely to follow the direction of high amplitude lines [Prati90A].

3.3 Optimal topographies under the L_p norm

We can find now different unwrapped phase distributions (topographies) optimal in the sense that each minimises a norm (L_p) of the error. Just as a reminder, we recall that in the case of a random variable z that corresponds to noisy measurements of the variable a , i.e.

$$z_i = a + n_i \quad \text{Equation 3.3}$$

where z_i, n_i are the samples of the data and of the noise.

Then \hat{a}_p , the estimate of a optimal under norm p , is the number that minimises J :

$$\hat{a}_p = \arg \min_a [J_p] = \arg \min_a \left[\sum_i |a - z_i|^p \right] \quad \text{Equation 3.4}$$

For $p = 0$, \hat{a} is the value of the mode of the z_i 's; for $p = 1$, their median; and for $p = 2$, their mean.

In the case of phase unwrapping, we can consider the policy of minimising the differences, raised to the power p , between the unwrapped phase differences along azimuth and range $\Delta^{(a)}\psi_{i,j}, \Delta^{(r)}\psi_{i,j}$ corresponding to the chosen topography, and the WPD $\Delta_w^{(a)}\phi_{i,j}, \Delta_w^{(r)}\phi_{i,j}$ along the azimuth and range coordinates i, j . We minimise then [Ghiglia98]:

$$C_p = \sum_i \sum_j |\psi_{i+1,j} - \psi_{i,j} - \Delta_w^{(a)}\phi_{i,j}|^p + \sum_i \sum_j |\psi_{i,j+1} - \psi_{i,j} - \Delta_w^{(r)}\phi_{i,j}|^p \quad \text{Eq. 3.5}$$

Obviously the minimisation of such a figure of merit appears to be, for $p \neq 2$, a formidable problem, with no guarantees of being unique and, even more important, no guarantees that the optimal solution will be significant. However, there are important implications of the formula that show the usefulness of such an approach.

3.3.1 L_2, L_1, L_0 , optimal topographies

There are two apparently trivial phase unwrapping methodologies of a noisy and aliased phase field that are easy to describe: the first is to scan the image progressively, say left to right and then top to bottom, and to sum the WPDs

along each scan line, and along the vertical for the first pixel of each line. We get a topography with positive and negative ‘walls’, i.e. h_a elevation discontinuities, directed along the scan lines, that start at the residues, and most likely end at the boundaries of the image: a very unlikely topography indeed, but consistent with the data and not optimal under any norm. Then, we could scan the image in any another direction and we would get another topography, unlikely as well. It is very interesting to observe [Fornaro96, Franceschetti99] that the average of all these unlikely topographies yields the optimal solution under the norm L_2 . Apart from boundary effects, we would get the same result by minimising the mean square value of the differences between the WPD (range and azimuth) and the phase differences derived from the unwrapped phases (the topography). In other words, we get such a topography that the WPD are matched as far as possible in the least square.

Another solution (L_0) is that of minimising the total length of the mismatches (Figure 3-10). We suppose that the phases are reliable and aliases arrive seldom, so that we look for the topography which minimises the instances in which alias is supposed to happen and therefore the WPD and the differences of the unwrapped phase differ. The discontinuity heights would appear rarely, but then they could be multiple of h_a .

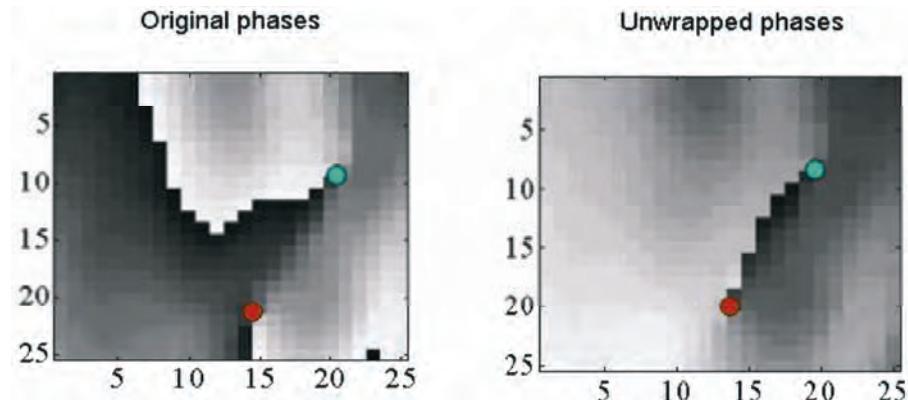


Figure 3-10: Connecting residuals under the L_0 norm. Left: a noisy phase field. Right: the L_0 unwrapped phase field (neglecting the boundaries close by).

A third solution (L_1) is to minimise the absolute total sum of the mismatch heights, controlling at the same time the instances and the heights of the discontinuities [Costantini98].

Notice that some of these solutions honour the values of the measured phases, adding to them $2\pi k$ (k is an integer). This is not the case of the solution of averaging the linear scans, since in that case the average of the discontinuity’s heights is an integer multiple of h_a at the residue, but goes to zero with $1/\text{distance from the residue}$. Then, the altitude correction is in general not an integer multiple of h_a . However, we might look for another (constrained) L_2 solution that allows only corrections of mh_a (m being an integer). We see that this solution, in the case that the corrections of more than $1 \times h_a$ are rare, is similar to that corresponding to (L_1). By the same token, the solution L_0 would then also be similar. Unfortunately, none of

these solutions needs to be ‘true’ and statistics is the only tool available to evaluate the quality of the result.

Examples of phase unwrapping are given in Figure 3-11, Figure 3-12 and Figure 3-13.

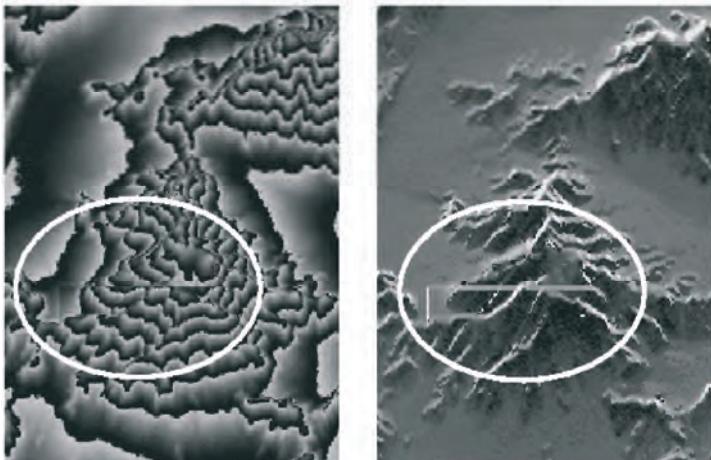


Figure 3-11: An example of errors in Phase Unwrapping. Left: unwrapped phase (error is encircled). Right: local slopes, estimated as the gradient of the unwrapped phase. Notice that the error is more evident in this image.

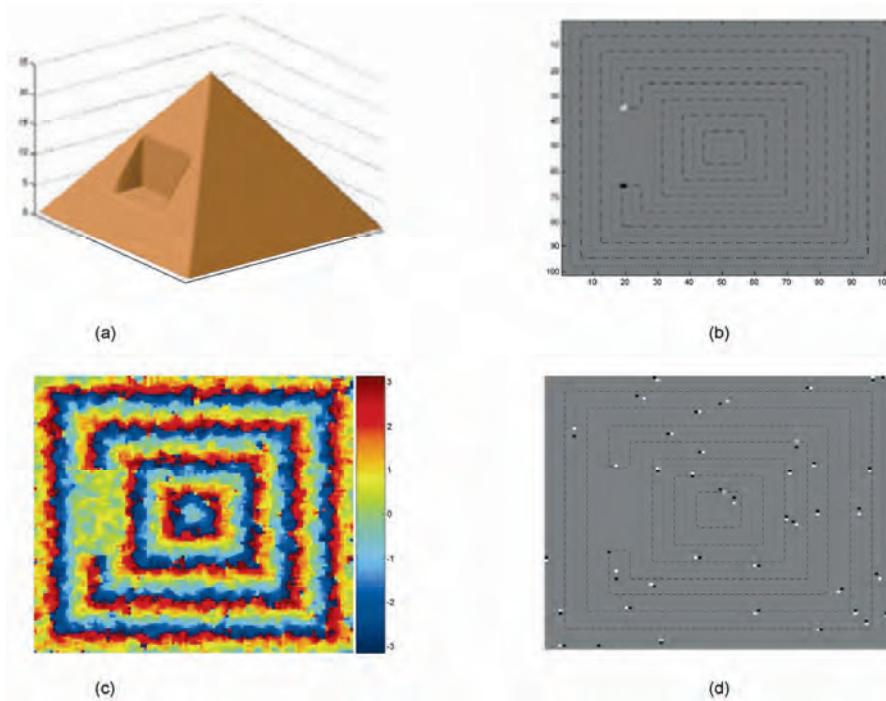


Figure 3-12: Example of 2-D phase unwrapping. (a) Noiseless unwrapped data; (b) Residuals, marked by white / black dots (depending on the sign) and zero-real, zero imaginary part contour lines. Only two residuals, due to topography, are found; (c) noisy wrapped phases; (d) distribution of residuals: of these, two are due to alias, the others to noise. Notice that the pair of opposite residuals tends to be aligned orthogonal to the slopes, and that ‘-’ anticipates ‘+’ in positive slopes.

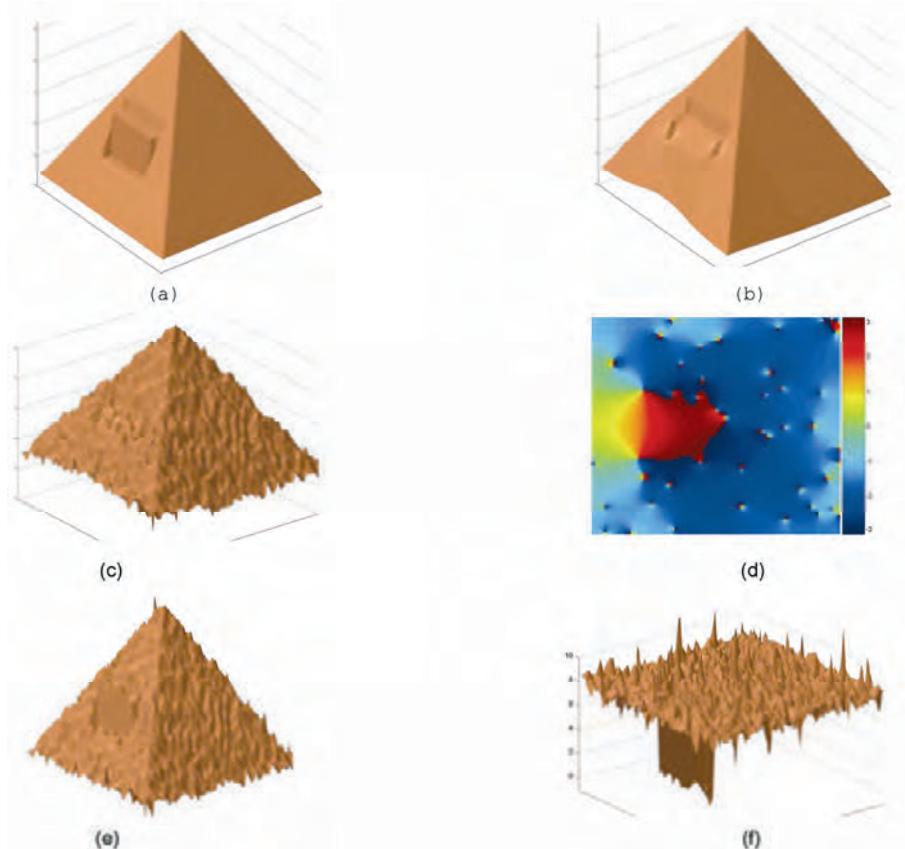


Figure 3-13: Examples of L_2 and L_0 norm phase unwrapping of the data in Equation 3.11. (a) L_2 solution in the noiseless case: inconsistent with the data; (b) L_0 solution in the noiseless case: this is consistent with the data, albeit not unique. (c) L_2 solution in the noisy case; (d) Phase error: it is a rotational field where the sources and sinks are in the residuals; (e) L_0 phase unwrapping, data consistent; (f) difference between the L_0 solution and the original, noiseless topography: the difference is due to the noise. Large errors (only multiples of the altitude of ambiguity h_a) may exist, depending on the structure of the original data.

3.3.2 Slope estimates

The previous techniques are designed to solve phase unwrapping problems in optics, Nuclear Magnetic Resonance, SAR interferometry etc. We will discuss now multi-pass SAR. Then, we have both alias and noise; in fact, the coherence on parts of the image might be quite low and the values of the interferometric phases very dispersed. Then, there is no other solution than making some sort of average and estimating the topography in a statistical sense. A reasonable approach is to hypothesise a locally planar surface and therefore the problem is reduced to that of slope estimation. Each slope corresponds to a sinusoidal behaviour of the fringes, uniform across the direction of the slope and with a frequency dependent on the slope. If the amplitudes of the interferometric signal were statistically uniform in the region of the slope, the maximum likelihood technique to estimate the slope in provided in section 2.3, a suboptimal but efficient technique could seek simply for the peak of the interferogram periodogram. In other words, we

make a local two-dimensional Fourier transform of the interferogram and look for the two spatial frequencies (along range and azimuth) of the maximum component of the spectrum. The wider the area where the slope is allowed to be uniform, the more precise is the estimate of its slope. While this technique to estimate the local slope is advisable, still there might be drawbacks when the amplitude is strongly non stationary (say a bright target surrounded by weaker ones). Then, the phase of the bright target would bias the result. Thus, it is at times reasonable to equalise the amplitudes before Fourier transforming, looking therefore only at the average phase gradient. Once we have the direction and the value of the local gradient of the topography, then we can merge all these data looking for the topography that, say in a mean square sense, best matches the estimated slopes. As said before, this solution matches that of the average of the scans; it does not honour the phases.

3.3.3 Removal of low resolution estimates of the topography

The periodogram technique coincides with the determination of the average amplitude of a 2D Fourier component after down-converting that component to zero frequency. Now, this operation can be carried out by analysing the Fourier components at all spatial frequencies (slopes) and then finding the component with maximum amplitude. Alternatively, instead of scanning all spatial frequencies, we might use *a priori* information on the average local slope. If this is known, say starting from a low resolution topographic map, then the local estimate can be made much simpler. We have thus the following method: start with a topography with low spatial resolution and determine the average slope, maybe using maps already available. Then, subtract the effect of that approximate topography from the data, and find locally the incremental slopes, that are on the average much smaller [Tarayre96]. This has two important outcomes:

- The alias is less likely, since we have reduced the absolute value of the slope; even if the noise level stays the same, the number of residues decreases because the distance between the lines of zeros increases;
- The residual topography in a given area may create less than a fringe: then a linear average of the complex numbers can be carried out to estimate their average phase. This coincides with estimating the phase as the argument of the complex coherence and allows for mutual compensation of the residues with different charges in the filter's pulse response.

3.3.4 Bias of the slope estimate

It is important to carry out this process with great care since the estimates of the gradient from the wrapped phases are biased [Davidson99]; in fact, if there is a positive local slope, the 2π skip, if any, will most likely counteract this slope. Therefore, the estimated gradient will be biased towards zero. Much less so, if the removal of the average local slope is previously carried out. Besides, it is very likely that the alias happens on the slopes facing the

satellite, where layover is likely. In these areas, the brightness is much higher due to the extent of the reflected energy that falls into a single resolution cell. Thus, it is reasonable to expect that most of the alias will occur in bright spots and, going from smaller to higher ranges, the aliased gradient will be short of 2π .

3.4 Analysis in the wave-number domain

An interesting technique derived from the paper [Costantini99] is the following: measure the noisy phase gradient and evaluate the rotational component. In the frequency domain we evaluate the noisy gradients $G_a(u,v)$, $G_r(u,v)$ of the unwrapped phase field $\psi(a,r)$ (the inverse Fourier transform of $P(u,v)$):

$$G_a(u,v) = juP(u,v) + n_a(u,v) ; E[|n_a|^2] = \mu_a |P|^2 \quad \text{Eq. 3.6}$$

$$G_r(u,v) = jvP(u,v) + n_r(u,v) ; E[|n_r|^2] = \mu_r |P|^2 \quad \text{Eq. 3.7}$$

where (a, r) are the two coordinates (azimuth, range)
 (u, v) are the corresponding spatial frequencies in the transformed domain.

In order to simplify the notation, we assume continuous instead of sampled data.

$P(u,v)$ is the transform of the unwrapped phase (and, apart from a constant factor, the topography in SAR coordinates).

G_a, G_r are the transforms of the unwrapped phase (topography) gradients and apart from the noise due to wrapping these are equal to ju, jv times $P(u,v)$.

n_a, n_r indicate the transforms of the noise mostly due to alias, i.e. of the walls with height equal to 2π or a multiple (proportional to multiples of h_a that extend along range and/or azimuth), and that are visible only at their ends.

Simplifying the notation, we first find the 2D transform of the curl of the data as:

$$C_R = jvG_a - juG_r = j(vn_a - un_r) \quad \text{Equation 3.8}$$

Its inverse transform corresponds to the residues (mostly negative and due to layover). Then, where the absolute value of the discrete curl is greater than a threshold ($\pi/4$ in [Costantini99]) we add 2π to reduce it and re-compute the phase, and then iterate.

3.4.1 L_2 optimisation in the wave-number domain

Using this notation, we can also give the results of the L_2 optimisation, which appeared in [Prati94B], since the unwrapped phase would be the optimal linear combination of the two noisy gradients:

$$\hat{P}(u, v) = A(u, v)G_a(u, v) + R(u, v)G_r(u, v) \quad \text{Equation 3.9}$$

Imposing orthogonality between data and reconstruction error we have:

$$\begin{aligned} E[(\hat{P} - P)G_a^*] &= 0 \\ E[(\hat{P} - P)G_r^*] &= 0 \end{aligned} \quad \text{Equation 3.10}$$

If we assume that the noise on the gradient is small and spectrally shaped as $|P|$, we find, consistently with [Ghiglia98], that if

$$\mu_r = \mu_a = \mu \quad \text{Equation 3.11}$$

Then:

$$A(u, v) = \frac{-ju}{\Omega^2 + \mu} ; R(u, v) = \frac{-jv}{\Omega^2 + \mu} ; \Omega^2 = u^2 + v^2 \quad \text{Eq. 3.12}$$

In other words, if the noise is small, the Laplacian of the unwrapped phases to be estimated is equal to the Laplacian obtained from the WPD. In fact we get:

$$(\Omega^2 + \mu)\hat{P}(u, v) = -juG_a(u, v) - jvG_r(u, v) \quad \text{Equation 3.13}$$

This is equivalent to the equations found in Equation B3.4, except that this is written in the frequency domain and Equation B3.4 is in the time domain.

Recalling that the noise on the gradient is shaped as walls in all directions, then:

$$\mu_r = \mu_a = \frac{\lambda_1}{\Omega^2} \quad \text{Equation 3.14}$$

the coefficients become:

$$A(u, v) = \frac{-ju\Omega^2}{\Omega^4 + \lambda_1} ; R(u, v) = \frac{-jv\Omega^2}{\Omega^4 + \lambda_1} \quad \text{Equation 3.15}$$

Further, if we admit that with unequal sampling in azimuth and range the noise would not have equal power (i.e. mostly across range) then the coefficients become:

$$\begin{aligned} A(u, v) &= \frac{-ju\mu_r}{\mu_r u^2 + \mu_a v^2 + \mu_r \mu_a} ; \\ R(u, v) &= \frac{-jv\mu_a}{\mu_r u^2 + \mu_a v^2 + \mu_r \mu_a} ; \end{aligned} \quad \text{Eq. 3.16}$$

$\mu_r \gg \mu_a$

and therefore the reconstructed ‘walls’ would be mostly along azimuth and across range, giving results similar to what could be obtained using the technique described in [Costantini99].

3.5 Weighting factors in the optimisation

The previous analysis shows that alias noise is limited if the average slope to be estimated is as close to zero as possible; in other words, in all cases it is useful to compensate the topography as much as possible, using *a priori* information. Then, the phase field to be estimated will be as close as possible to a flat horizontal surface and thus alias will be less likely, even if sudden slope changes can never be ruled out. If we can expect limited alias, then the optimal solution should be smoothed more if the noise level is higher, in order to reach optimal noise immunity. Therefore, in conditions of low coherence we should smooth the compensated topography as much as needed to limit the phase dispersion. This effect can be easily obtained by locally changing the expected value of the noise variance, in accordance with the level of the coherence. These weightings will be space-variant and thus cannot be applied using frequency domain techniques. Weighted L_p norm optimisation can be used then, as proposed by several authors. Weighting techniques could also be used to direct the phase integration paths to avoid unknowingly crossing h_a discontinuities, following the local topography, i.e. limiting the possibility of the paths to cross a layover area, or joining two areas with widely different scattering characteristics.

4. Multiple image combination for DEM generation and ground motion estimation

4.1 Multi-baseline phase unwrapping for InSAR topography estimation

The theory of topographic mapping using SAR interferograms and the difficulties related to phase unwrapping of InSAR data have been presented in detail in recent review papers [Bamler98A, Massonnet98, Rosen00], and books [Franceschetti99, Ghiglia98]; here we address the multi-baseline technique, that removes the unwrapping problem in practice.

In the ideal case of no noise and no artefacts due to changes in the propagation medium, the interferometric phase ϕ of each pixel is proportional to the travel path difference and a function of the local topography. Since the quantity that can be measured from the interferometric image is not the phase ψ , but its wrapped (principal) value ϕ , the value of ψ has to be determined by means of a 2-D phase-unwrapping procedure. The quality of the final DEM is strongly dependent on this processing step. Conventional orbit determination is not precise enough for absolute phase retrieval and usually the final DEM is a map where all the values are computed with respect to a reference point P_0 of known elevation z_0 . Therefore the datum used for DEM generation is not the phase ϕ of the generic pixel, but the unwrapped *phase difference* between P and P_0 .

Once the phase unwrapping step has been carried out, it is possible to obtain the elevation of each pixel in the image. The phase to height conversion function $g(\Delta\phi)$ can be computed if the acquisition geometry is known. In general $g(\Delta\phi)$ is a non-linear relation involving, besides $\Delta\phi$, the acquisition positions, i.e. the satellite orbits, the coordinates of the reference point, the range and azimuth variations between P and P_0 , and the system wavelength.

In Figure 4-1 the area of interest is a small patch of ground around P_0 ; the satellite trajectories are considered linear; and the phase to height conversion function can be well approximated by a linear relation:

$$\Delta z = z - z_0 = g(\Delta\phi) = A \cdot \Delta r + B \cdot \Delta y + C \cdot \Delta\phi \quad \text{Equation 4.1}$$

where A, B, C are constant parameters,

$\Delta r = r(P) - r(P_0)$ and

$\Delta y = y(P) - y(P_0)$

(i.e. they correspond to the variations between the two pixels in range and azimuth direction respectively).

This linear approximation is useful because it highlights the main features of the acquisition geometry.

A depends on the off-nadir angle.

B is usually small but not negligible, since the orbits cannot be assumed parallel.

C is inversely proportional to the normal baseline B_n in $y = y_0 + y(\mathbf{P}_0)$. The height difference corresponding to a phase variation of one cycle is the so-called altitude of ambiguity h_a :

$$h_a = C \cdot 2\pi = \frac{\lambda}{2} \frac{r_{M0}}{B_n} \sin \theta \quad \text{Equation 4.2}$$

where θ is the antenna off-nadir angle

r_{M0} is the distance between \mathbf{P}_0 and the master trajectory (Figure 4-1).

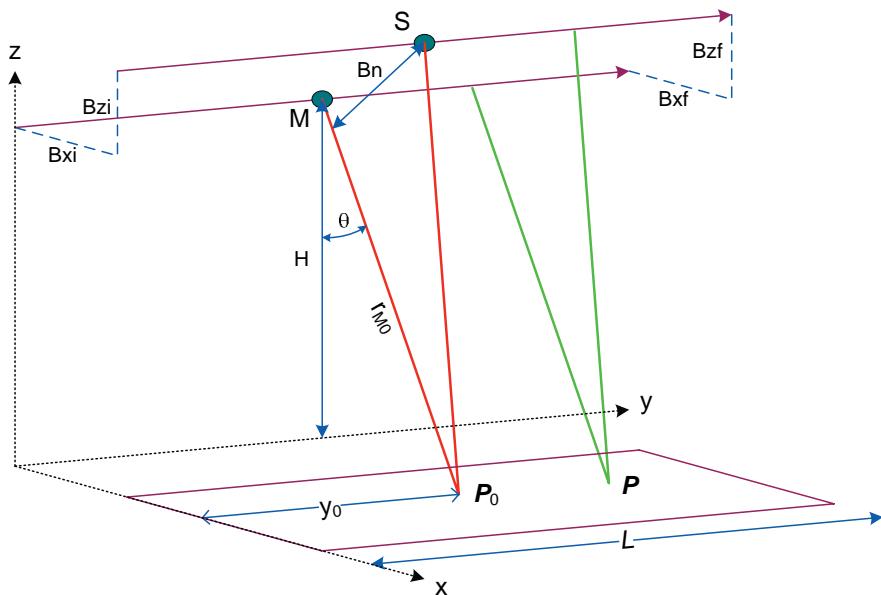


Figure 4-1: InSAR geometry and symbols used

From the previous equation it follows that the greater the baseline, the greater is the interferometric phase difference between two points and the more difficult will be the unwrapping process due to the higher probability of phase aliasing. Moreover, geometric decorrelation strongly limits the SNR in high baseline interferograms [Gatelli94]. On the contrary, low baseline interferograms show good fringes, usually easy to unwrap, but with poor topographic accuracy [Ferretti99]. If more than one interferogram of the same region is available, phase unwrapping should be carried out *simultaneously* on the entire data set. The underlying idea of the multi-baseline approach is simple:

Interferograms are different measures of the same physical variable: the topography [Ferretti96B, Ferretti97]

If no *a priori* information about the topography is available, phase unwrapping of a single SAR interferogram is an ill-posed inverse problem. Even in high coherence areas, the presence of phase aliasing can produce serious artefacts on the estimated topography. If NI interferograms of the

same region are available, with different baseline values, it is possible to combine the data to eliminate phase aliasing.

Theoretically it should be enough to have two interferograms with baseline ratios that are prime with respect to each other (their ratio is not a rational number) to be able to remove phase ambiguities (the Chinese remainder theorem). In practice, where data are noisy and baselines random, the use of multiple interferograms significantly increases the elevation ambiguity level. The use of Bayesian inference gives a rigorous framework for the estimation, taking into account possible *a priori* information about the DEM.

Let P_0 be a point of known elevation (**GCP: Ground Control Point**). The target is then to compute the elevation of each point P in the area of interest with respect to P_0 . Let $z(P)$ be the height variation between P and P_0 . This quantity can be considered as a random variable, its value can be estimated from the NI measures available. If $f(z/\Delta\phi^i)$ is the **probability density function (PDF)** of the variable z conditioned to the wrapped phase value $\Delta\phi^i$ (the apex specifies the interferogram used for the estimation), from the Bayes rule, if the measures are statistically independent, we have:

$$f(z/\Delta\phi_w^1, \dots, \Delta\phi_w^{NI}) = \frac{1}{C} \frac{f(\Delta\phi_w^1, \dots, \Delta\phi_w^{NI} / z)}{f(\Delta\phi_w^1, \dots, \Delta\phi_w^{NI})} f_{ap}(z) \quad \text{Eq. 4.3}$$

where C is a normalisation constant. The **Maximum A Posteriori (MAP)** estimate z of the variable z maximises this PDF. When the value of z is determined, it is easy to choose the multiple of 2π to be added to each interferogram: the correct value of the phase will correspond, in fact, to the height value nearest to the estimated one. In order to compute the conditional density we use the coherence maps associated with each interferogram. From the absolute value of the local coherence γ and the number of looks in the interferogram (the number of degrees of freedom in the estimate of $\Delta\phi$), it is possible to compute the expression of the PDF of the interferometric phase and thus of the elevation. The conditional density function of the elevation for each interferogram is periodic with a different period (the altitude of ambiguity) dependent on the baseline. In each period the higher is the quality of the fringes (the coherence) the sharper is the histogram.

The product of the conditional densities shows a neat peak whenever the coherence is not close to zero and the baseline errors are not too high. The sharper the global peak, the higher the reliability of the results, i.e. the probability that the correct value of the height variation lies inside a given interval. In Figure 4-2 an example is shown of this kind of computation, where three interferograms with three different baseline values are considered.

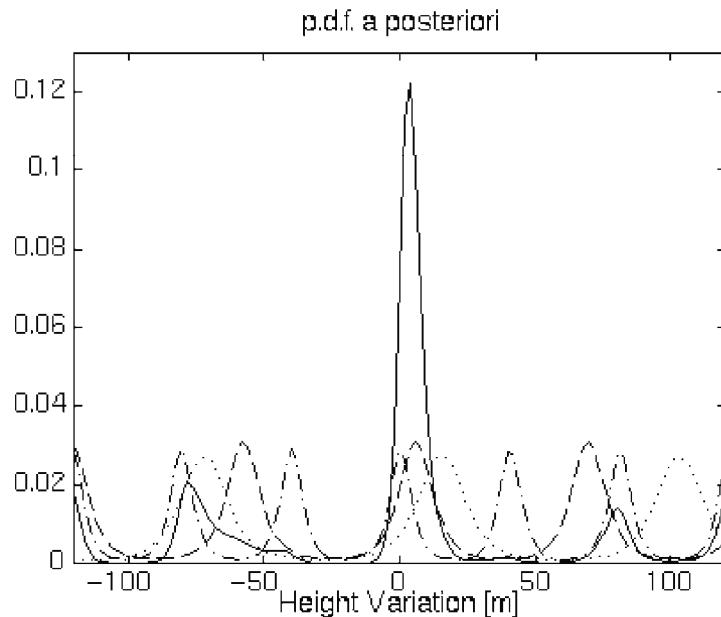


Figure 4-2: A posteriori probability density function (PDF) of the relative terrain elevation obtained with three interferograms

The benefits of the multi-baseline approach are twofold. First, combining all the information it is possible to limit the impact of the noise. Besides, there is minor risk of aliasing with conventional single interferogram phase unwrapping: *working simultaneously on more interferograms, phase unwrapping is possible even if the phase is undersampled*. Of course, the higher the noise, the worse the reliability and the more likely the phase unwrapping errors.

4.2 Applications to repeat-pass interferometry

If we knew exactly the phase-to-height conversion function for each interferogram, the implementation of the algorithm would be straightforward. All we would need to do is compute the PDF *a posteriori* conditioned to the data for each pixel in the image. Unfortunately, this is usually not the case. First of all, the satellite trajectory is not exactly known: only an estimate of the phase-to-height conversion function is available. Of course, attitude errors are systematic and usually introduce a tilt, or more generally a polynomial distortion on the map [Ferretti99]. Nevertheless, they cannot be neglected. Moreover, even if the orbits were known exactly, another low frequency contribution should be taken into account in repeat-pass interferometry: a phase distortion α due to random refractive index variations in the propagation medium [Zebker97, Hanssen98, Ferretti99]. The intensity of this distortion cannot be estimated from the local coherence, since the correlation length of these phenomena (usually more than 1 km) is far longer than the estimation window used for γ , and large ground regions will exhibit a common error. So the final topography can show strong distortions in spite of high coherence values.

Atmospheric effects and baseline errors thus strongly reduce the effectiveness of the multi-baseline approach outlined above. In order to correctly unwrap the phase values, a good match between the topographic profiles coming from each interferogram is necessary. These low-frequency phenomena make a straightforward implementation of the multi-baseline approach unfeasible. The ‘resonance’ of the PDFs relative to the interferograms would become more and more unreliable for pixels more and more distant from the reference point P_0 .

The process has to be localised in order to remove these low wave number effects before expanding the procedure to wider and wider areas. In order to compensate for possible baseline errors and atmospheric distortion, the phase-to-height conversion functions are iteratively optimised as more and more points are unwrapped. A Root Least Mean Square (RLMS) optimisation is used to minimise the error between the elevation values obtained from each datum. Since the locally optimised baseline values compensate not only for orbit indeterminations but for atmospheric distortions too, they are called **effective baselines**. The need for optimisation procedures would make the implementation of the algorithm cumbersome and very time consuming. For that reason the processing is carried out at two different resolution levels. The image is divided into small blocks (of about 1×1 km) such that:

- 1) orbits can be considered linear;
- 2) phase distortion due to possible atmospheric effects can be considered linear;
- 3) the phase to height conversion function can be well approximated, for each interferogram, by a linear function:

$$\Delta z^i = A^i \cdot \Delta r + B^i \cdot \Delta y + C^i \cdot \Delta \phi + D^i \quad \text{Equation 4.4}$$

where the apex specifies one of the NI interferograms. As already mentioned, all these variations are defined with respect to a reference point chosen inside each block. The reference point is a pixel having high coherence value in all the interferograms. The D^i parameter compensates for phase noise on $\phi^i(P_0)$. Inside each block it is then possible to carry out a *linear* recursive optimisation, fast and effectively.

Results concerning two difficult test areas (Vesuvius and Etna) are presented below. The unwrapped phase maps were obtained by averaging the interferograms by a factor 5 in the azimuth direction (three effective looks were considered for PDF computation), maintaining full resolution data in slant range. The final resolution cell is about 20×20 m for flat terrain. Remember that the reconstructed unwrapped phase is identical to the original wrapped phase when unwrapped. In order to assess the reliability of the results, an *a priori* DEM of both test sites was used.

4.2.1 Example 1: the Vesuvius data set

The multi-image reflectivity map (incoherent average of the co-registered SAR data) of the region around Mount Vesuvius is shown in Figure 4-3. The maximum height variation is 1281 m (from sea level to the top of the volcano). Seven Tandem interferograms were used, the baseline values ranging from 39 to 253 m. No *a priori* information was exploited during the processing.

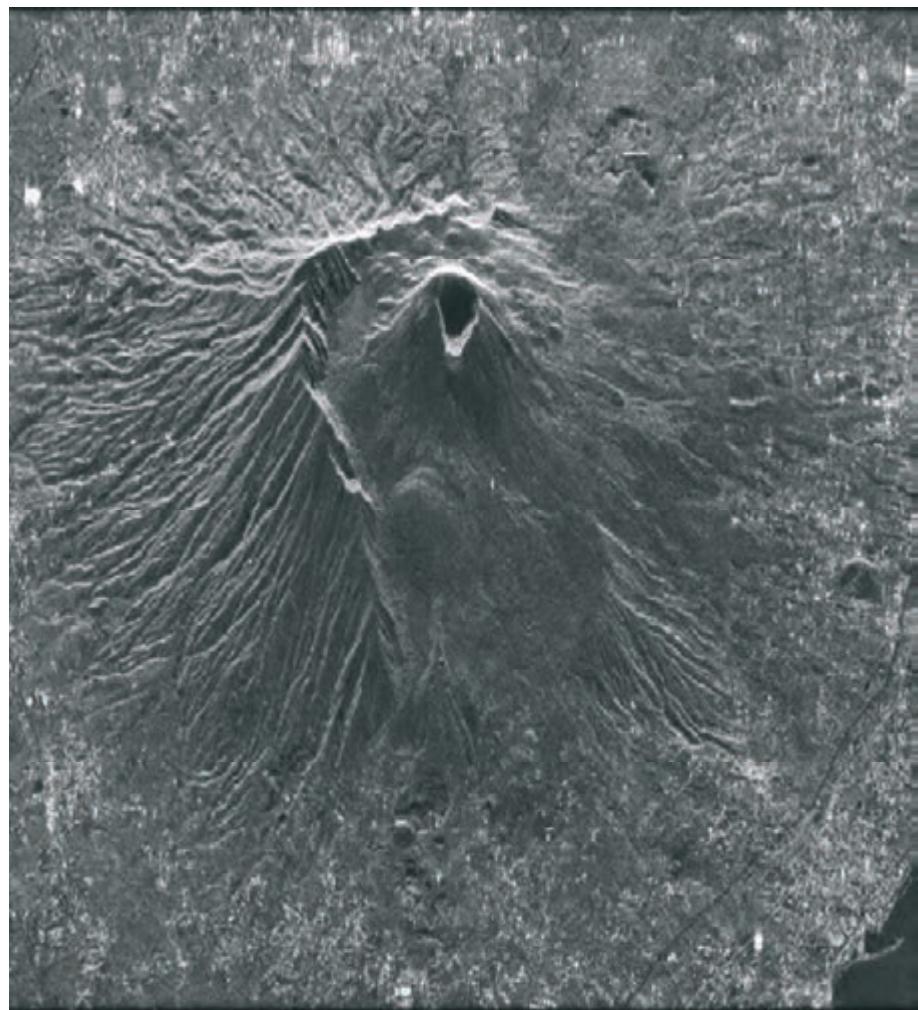


Figure 4-3: Mount Vesuvius: multi-image reflectivity map

Figure 4-4 shows the estimated topography after the first processing step. The blocks are still not unwrapped with respect to each other.

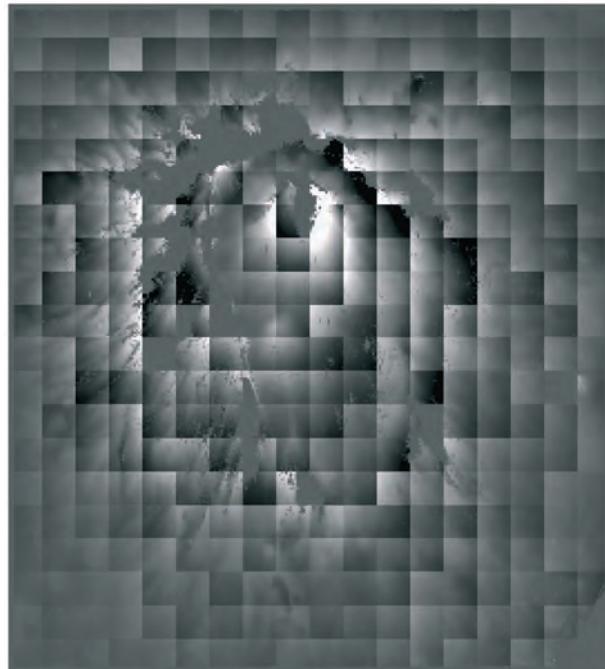


Figure 4-4: Estimated topography after the first processing step

The final DEM was computed using the seven unwrapped phase maps and the technique described in [Ferretti99]. In order to reduce baseline errors, the ESA precise orbits products (processed at GFZ/D-PAF, Oberpfaffenhofen) have been used to estimate the satellite trajectory relative to each image. A 3-D perspective of the final result is shown in Figure 4-5.

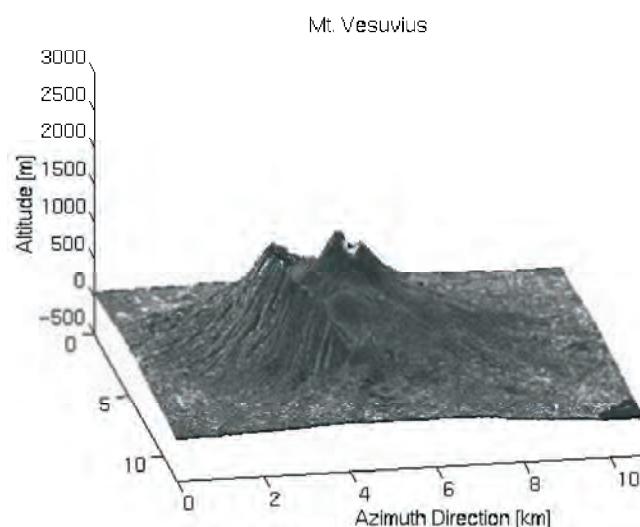


Figure 4-5: 3-D perspective view of the Mt. Vesuvius DEM

The error between the combined DEM and a reference (SPOT) topography (in SAR coordinates) is shown in Figure 4-6. White pixels correspond to areas not unwrapped (reliability under threshold). The error standard deviation is about 8 m. It should be noted that the available SPOT DEM has an estimated accuracy of 7 – 8 m, so multi-baseline InSAR data allow a precise estimation of the topographic profile.

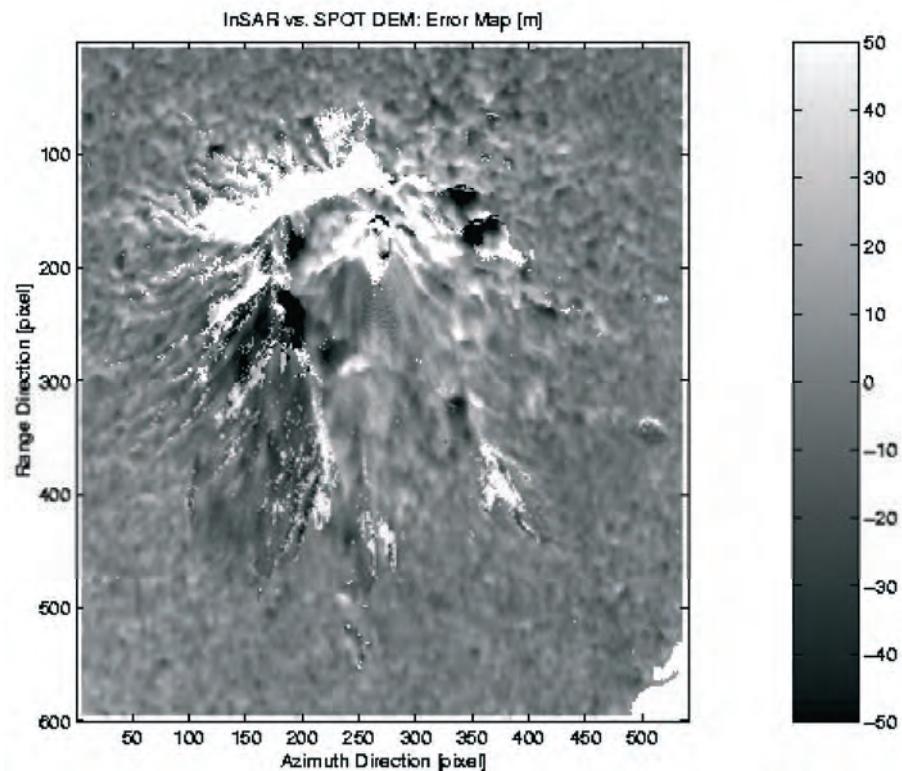


Figure 4-6: Difference between the combined DEM and a reference (SPOT) topography (in SAR coordinates)

4.2.2 Example 2: The Etna data set

The second example shown is the region around Etna. Eight interferograms were used. The area selected is about 30×30 km. As usual, the images were averaged only by a factor of five in the azimuth direction (Figure 4-7).



Figure 4-7: Mt. Etna: multi-image reflectivity map

An example of a reliability map generated by the multi-baseline phase unwrapping software is shown in Figure 4-8, while the best coherence map of the Etna data set is shown in Figure 4-9 for comparison.

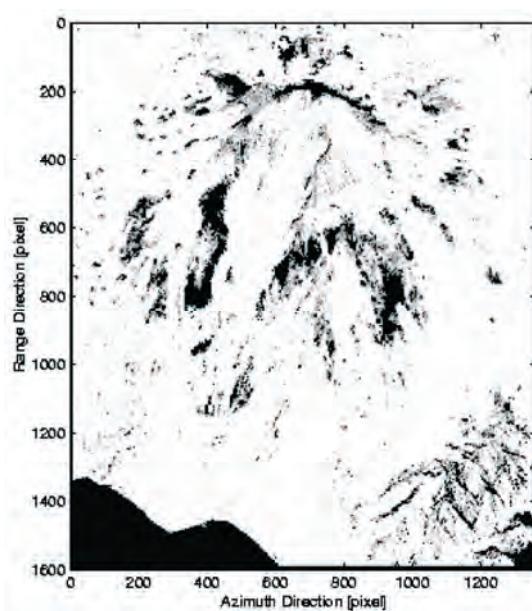


Figure 4-8: Reliability map generated by the multi-baseline phase unwrapping software. White=reliable; Black=unreliable.



Figure 4-9: Best coherence map

It should be noted that, using the multi-baseline approach, it was possible to unwrap the October 1995 Tandem pass, with a normal baseline of almost 400 m (Figure 4-10).

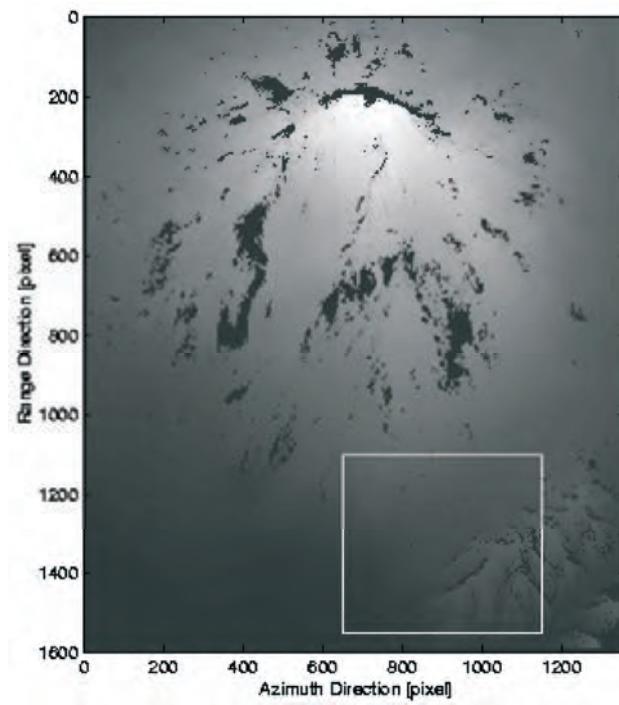


Figure 4-10: Unwrapped tandem interferogram with a 400m perpendicular baseline

The difficulties related to the unwrapping of this high baseline interferogram can be appreciated by analysing Figure 4-11, which shows a close up of the wrapped interferogram.

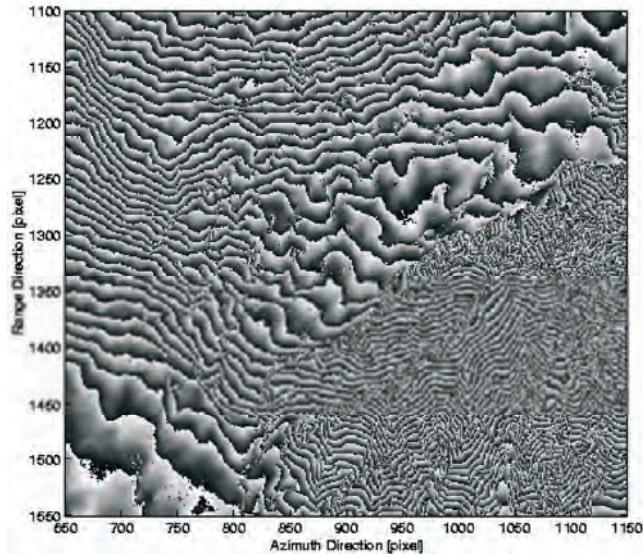


Figure 4-11: Close up of the wrapped interferogram with a 400m perpendicular baseline

A perspective view of the final DEM, obtained from the combination of all the unwrapped phase data [Ferretti99], is shown in Figure 4-12.

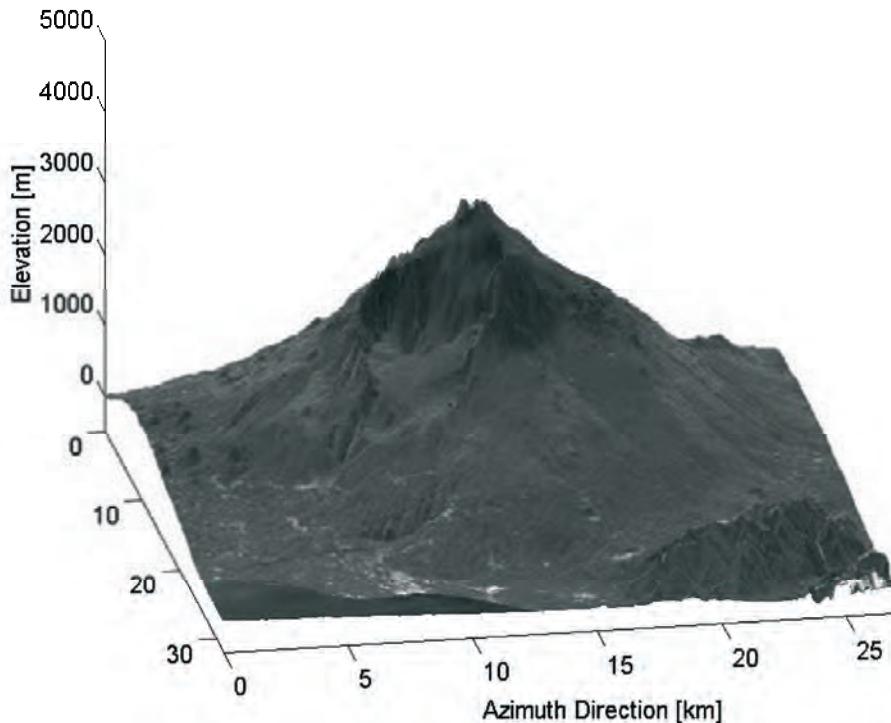


Figure 4-12: Perspective view of the reconstructed Etna DEM

In order to assess the reliability of the final estimated topography, Figure 4-13 shows the error histogram with respect to a reference DEM provided by IPGP (estimated accuracy 4 m). The error standard deviation (computed for unwrapped pixels) is 7.5 m.

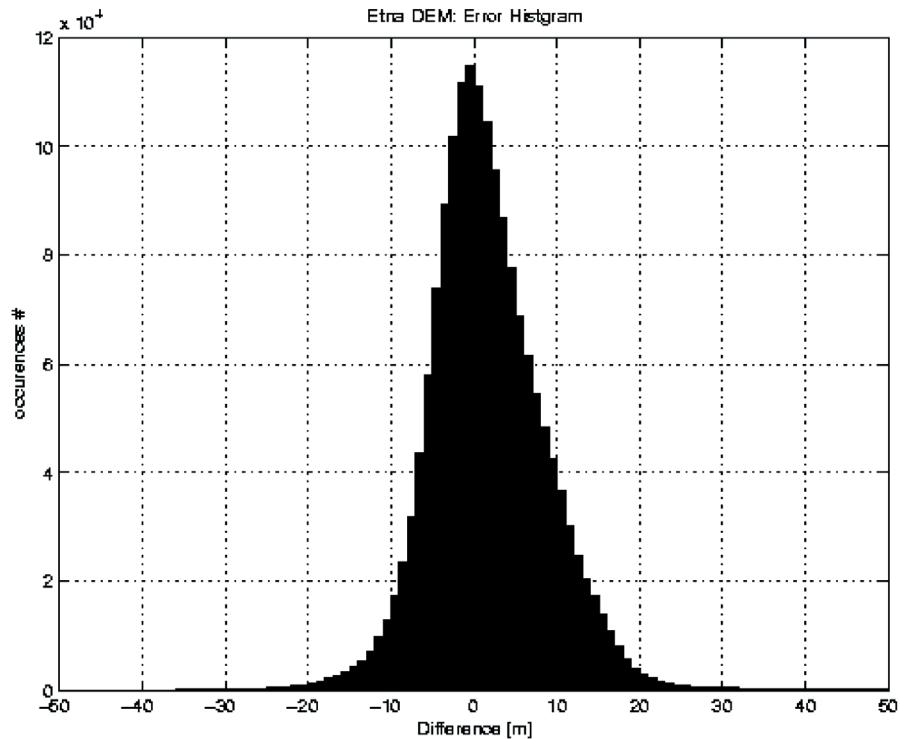


Figure 4-13: Elevation error histogram with respect to a reference DEM provided by IPGPⁱ (estimated accuracy 4 m)

4.3 The ‘Permanent Scatterers’ technique

One of the main difficulties encountered in differential InSAR (DInSAR) applications is temporal and geometric decorrelation. The main goal of this section is the identification of single pixels (called **Permanent Scatterers**, **PS**) that are coherent over long time intervals and for wide look-angle variations [Ferretti00, Ferretti01]. This allows one to use many or all ERS acquisitions relative to an area of interest. In fact when the dimension of the PS is smaller than the resolution cell, the coherence is good (the speckle is the same) even for image pairs taken with baselines greater than the decorrelation length. On those pixels, sub-metre DEM accuracy and millimetric terrain motion detection can be achieved, even if the coherence is low in the surrounding areas. Reliable elevation and deformation measurements can then be obtained on this subset of image pixels that can be used as a ‘natural’ GPS network.

The mathematical framework for this kind of estimation is relatively simple. Let us suppose that $N+1$ ERS SAR images of the area of interest are

ⁱ Institut de Physique du Globe de Paris

available. Data are first co-registered on a unique master, and a DEM of the area is estimated starting from low temporal baseline pairs [Ferretti99]. Next, N differential interferograms between all SAR images and the master are computed. After DEM compensation, the residual phase of interferogram i is:

$$\phi_i = \frac{4\pi}{\lambda} r_{Ti} + \alpha_i + n_i + \varepsilon_{topo_i} \quad \text{Equation 4.5}$$

where λ is the system wavelength,

α_i the atmospheric phase contribution,

n_i the decorrelation noise,

ε_{topo_i} the phase contribution due to possible errors in the DEM (proportional to the normal baseline of each image B_{n_i} , and r_{Ti} is the possible target motion in the direction of the satellite line-of-sight.

The first term in Equation 4.5 can then be written as follows:

$$\frac{4\pi}{\lambda} r_{Ti} = \frac{4\pi}{\lambda} v_r T_i + \mu_{NLi} = C_{vi} v + \mu_{NLi} \quad \text{Equation 4.6}$$

where v is the unknown component of the mean target velocity in the direction of the line of sight,

μ_{NL} is the phase term due to a possible non-linear target motion

T_i is the temporal baseline between the master acquisition and the generic i -th slave image.

Since we have N differential interferograms of the same area with different temporal and geometric baselines, we finally write, for each pixel, a linear system of N equations and two unknowns:

$$\phi_i = C_{zi} \varepsilon_z + C_{vi} v \quad i = 1, \dots, N \quad \text{Equation 4.7}$$

Where: ε_z is the DEM error

C_{zi} is proportional to B_{n_i}

The problem would be linear if we knew the unwrapped phase values. However, using a simple periodogram (albeit with an irregular sampling of the two dimensions: baselines and time) we can estimate both the residual elevation and the mean line of sight velocity, provided that the signal-to-noise ratio is high enough.

Actually the **Linear Phase Residues (LPR)** w_i (i.e. phase data after linear detrending in temporal and spatial baseline) are the sum of three contributions: atmospheric components (**Atmospheric Phase Screen – APS**) of the master and the slave acquisitions, noise, and non-linear motion:

$$w_i = \phi_i - C_{zi} \varepsilon_z - C_{vi} v_r = \mu_{NLi} + \alpha_i + n_i \quad \text{Equation 4.8}$$

Since μ_{NL} , α and n are independent random variables, the residual phase variance σ_w^2 is the sum of three contributions: $\sigma_{\mu_{NL}}^2$, σ_α^2 , σ_n^2 . As a matter of fact, a proper estimation of target motion and elevation is possible only if

σ_w^2 is low. To this end, the analysis is carried out in two steps. First a subset of image pixels is selected using the coherence maps of the area of interest: only locations of highly coherent targets are retained (low σ_n^2). If now we consider a pair of targets not too far apart (distance less than, say, 1 km), it is usually possible to estimate Δv (relative mean velocity) and $\Delta \varepsilon_z$ (relative elevation error) with a high degree of accuracy. In fact, low distance implies low variance of the atmospheric component: for points less than 1 km apart, values of σ_α^2 less than 0.1 rad² are common. Moreover, the motion of neighbouring pixels is usually correlated. If this hypothesis is verified, σ_μ^2 should be low as well. Estimation of (v, ε_z) is then obtained by integrating the values of $(\Delta v, \Delta \varepsilon_z)$ previously estimated from pairs of neighbouring pixels. Following this approach, *we can actually unwrap the differential interferograms*. In fact, if the condition $|w| < \pi$ is satisfied, we can easily recover the unwrapped phase differences $\Delta \varphi_i$ between pairs of neighbouring pixels and then integrate them all over the sparse grid of PS candidates.

We can summarise the results as follows. Phase unwrapping of differential interferograms characterised by high geometrical and temporal baseline is possible only on a sparse grid of targets, previously selected as **PS Candidates (PSC)**, and using a multi-image approach. The mathematical framework is easier if the constant velocity model can be applied. However, this hypothesis is not a real constraint and the technique can cope with non-uniform target motion, provided that the PS density and coherence is high enough (low σ_w^2). Phase differences are unwrapped taking advantage of the estimated values of relative velocity Δv and relative elevation $\Delta \varepsilon_z$ of each pair of scatterers. Phase data are then integrated using one of the techniques used for unwrapping regularly sampled data.

4.3.1 Space-time estimation

Once phase data are unwrapped, it is possible to estimate the signal of interest by using an optimum filter, and properly weighting the data vector. In general, we may be interested in estimates at unmeasured points and instants too (interpolation and/or prediction). It should be noted that the optimum filter is space and time variant.

The key point for optimum filtering is the estimation of the correlation values. While expression of the correlation matrix of the atmospheric component can be obtained from Kolmogorov theory, a statistical description of the local displacement field can be very difficult to achieve.

Basically, this choice should reflect our knowledge about the physical phenomenon we are dealing with. If a model is available, parameters and expression of the correlation matrix assume a specific physical meaning. If no *a priori* information is available about the displacement field of the area of interest, a possible approach that can be adopted is the whitening of the time series of the Linear Phase Residues (LPR) samples. Once the residual phases of the PS Candidates have been obtained (data compensated for the estimated DEM errors and the mean velocity field), we can start looking at their temporal evolution, to check for subsidence leakage. The time-smoothed component should be considered as an estimation of the non-

uniform motion contribution. Atmospheric Phase Screen (APS) estimation is then carried out to *spatially smooth the time high-pass filtered LPR data*.

To make it simple, we can for each pixel carry out a temporal smoothing using a triangular filter and remove the low pass component. However, seasonal effects can be removed from the data by using a band-pass filter. Phase residuals are then spatially filtered. After estimation and removal of all the APS superimposed on the data, we can finally estimate the motion of each pixel in the image and identify more PS's. In fact the PSC set is just a rough estimation of the PS set since isolated coherent targets are not detected due to the space average used for coherence estimation.

4.3.2 Subsidence in Pomona

An interesting case of subsidence, studied using differential interferometry and other techniques, is found in Pomona, California [Peltzereos]. 41 ERS images were obtained with the same track and frame (Track: 170. Frame: 2925). All were resampled on the same master acquisition (ERS-2, 6 April 1996) and 40 interferograms were obtained. Figure 4-14 shows the incoherent average of all the data (close up of Los Serranos area): due to the high number of looks, the radiometric quality of this image is comparable to that of an optical one. Using four Tandem pairs, we also determined a reference DEM (approximately 10 m accuracy, also limited by building effects) of the area of interest using the wavelet approach, and the corresponding phase was removed from each single image.

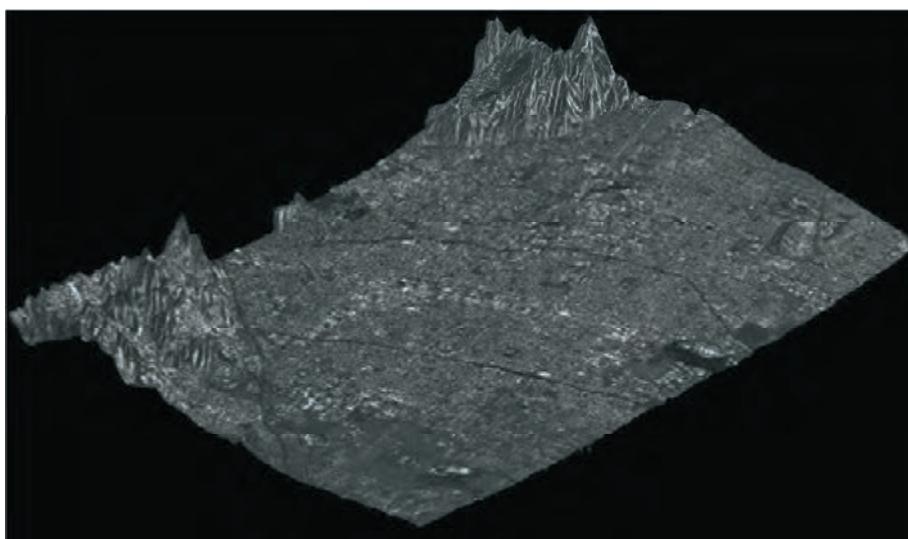


Figure 4-14: Geocoded multi-image reflectivity map of the test site (close up of Los Serranos area). The high radiometric quality is due to the incoherent data average

After the initial selection of the Permanent Scatterers set (about 3 PS/km² were identified), phase increments between each PS and all the others less than 1 km away were estimated using the periodogram technique. To illustrate this approach, Figure 4-15 shows the temporal distribution of the takes together with their normal baselines, referred to the master image.

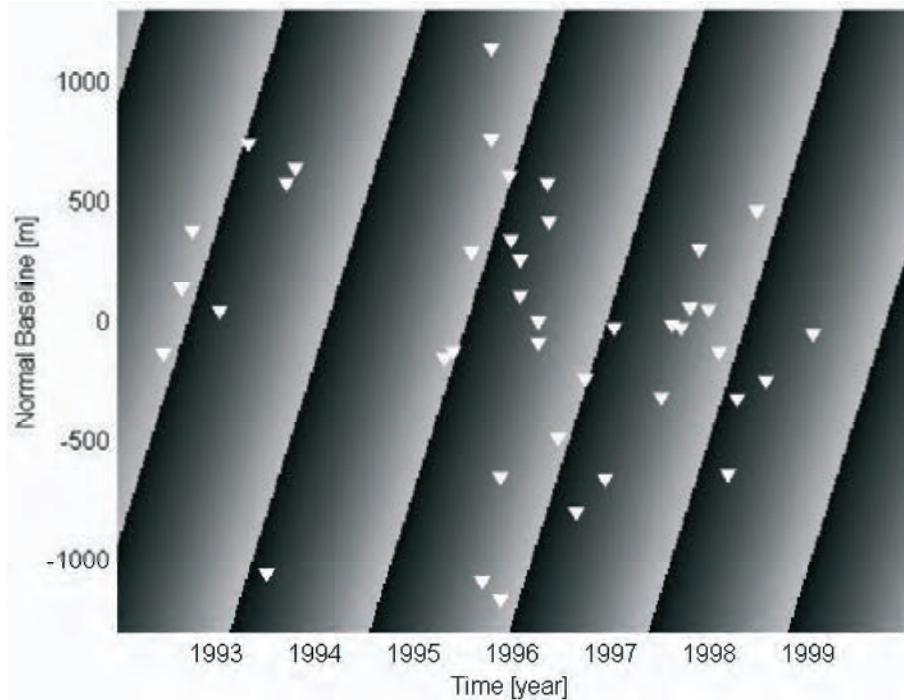


Figure 4-15: Space-time distribution of the available data. The bidimensional sinusoid represents the phase contribution for a LOS velocity of 2 cm/y and a DEM error of 5 m.

The range of normal baselines is about ± 1100 m, while the maximum temporal baseline is more than 6 years. If a PS had a **Line Of Sight (LOS)** velocity of, say, 2 cm/y and a residual elevation difference of 5 m with respect to a neighbouring scatterer, considered as a reference, its phase variations as a function of time and baseline would be a 2D sinusoid, also represented in Figure 4-15. If now we accept, temporarily, the hypothesis of constant LOS velocity of each pixel, then using a periodogram we can estimate both the residual elevation and the LOS velocity difference of the pixels. As already mentioned, this operation was carried out for all PS pairs less than 1 km apart, thus removing the effects of the residual elevation with respect to the average DEM and of the LOS velocity, and estimating the unwrapped phase values.

After estimation of both elevation and mean velocity of the targets, time series analysis of the phase residues in correspondence with each PS was carried out. The target is to identify possible non-linear motion contributions. APS's were then interpolated on the original regular grid and removed from each datum. It should be noted that each APS is actually the difference of the atmospheric component of the slave image and the APS of the master acquisition. Averaging the 40 APS's it was possible to get an estimation of the master contribution and then of each single contribution. An example of estimated APS is shown in Figure 4-16.



Figure 4-16: Example of APS estimated for 6 April 1996. The APS has been superimposed on the multi-image reflectivity map of the area. APS standard deviation is 0.86 rad.

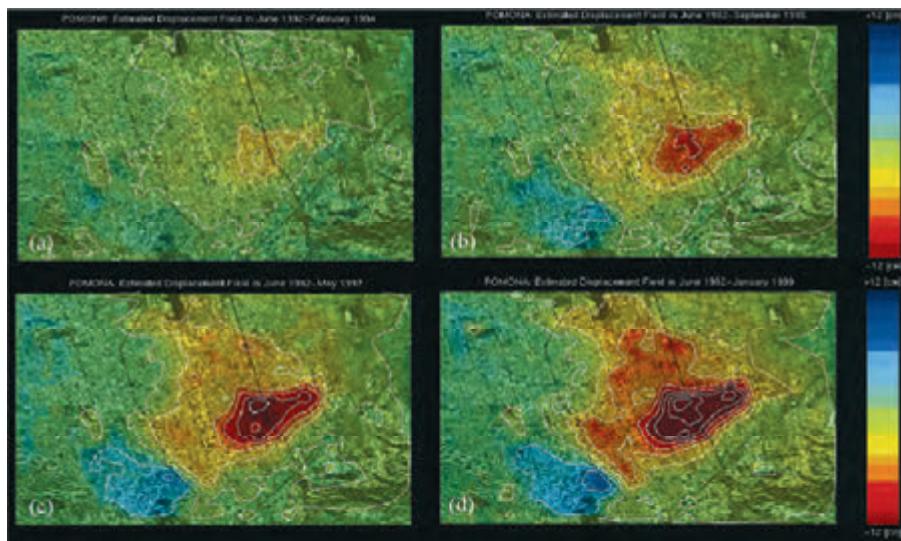


Figure 4-17: Pomona: estimated displacement field in cm relative to the time interval a) June 1992: February 1994; b) interval June 1992: September 1995; c) interval June 1992: May 1997; d) June 1992: January 1999. Colour scale has been saturated for visualisation purposes (± 12 cm). Contour lines step: 2.5 cm

After APS removal it is possible to estimate not only the mean velocity field of the area but a displacement field as a function of time, possibly interpolating the displacement maps on a regular temporal grid. Here we

report a sequence of four spatial maps corresponding to four equally-spaced time intervals (Figure 4-17).

The non-linear motion behaviour is reflected in the change of shape of the contour lines. A perspective view of the displacement field of Figure 4-17 is also shown in Figure 4-18. Maximum positive and negative displacements are +6 and -20 cm, respectively.

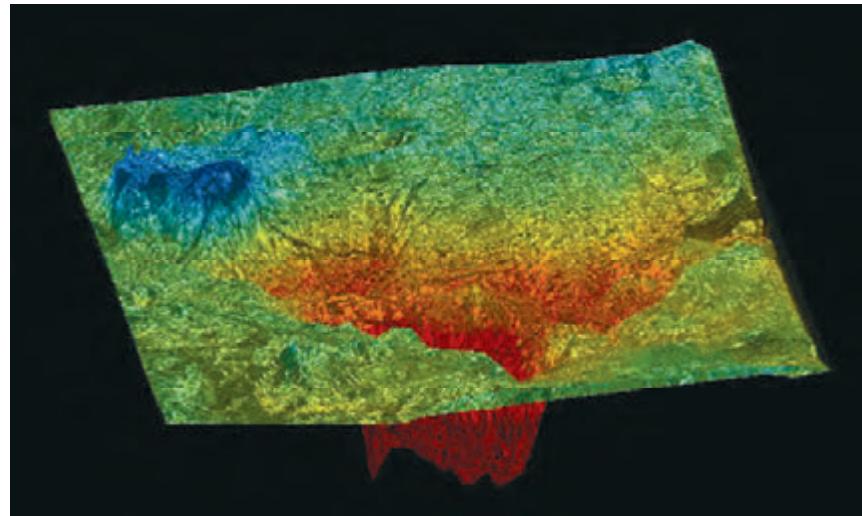


Figure 4-18: Perspective view of the displacement field relative to June 1992: January 1999. Minimum negative value: -20 cm. Maximum positive displacement: +6 cm.

4.3.3 Ground slip along the Hayward fault

The velocity field across the Hayward fault in California has been estimated by exploiting 55 ERS images acquired between June 1992 and November 1999 and processed with the Permanent Scatterers technique.

The PS density in the area is very high (over 200 PS/km²), so that the estimated LOS velocity field looks continuous in Figure 4-19. The relative dispersion of the velocity values in the two areas separated by the fault is lower than 0.4 mm/y. The stepwise discontinuity of about 2 mm/y in the average deformation rate can be identified easily and the hanging wall of the fault can be located with an accuracy of a few tens of metres.

In order to minimise interpolation artefacts, data are reported in SAR coordinates (range, azimuth) rather than in geographical coordinates. The sampling step is about 4 metres both in slant range and azimuth (ERS images have been interpolated by a factor of two in range direction). The radiometric quality of the image has been strongly improved by computing the incoherent average of the data.

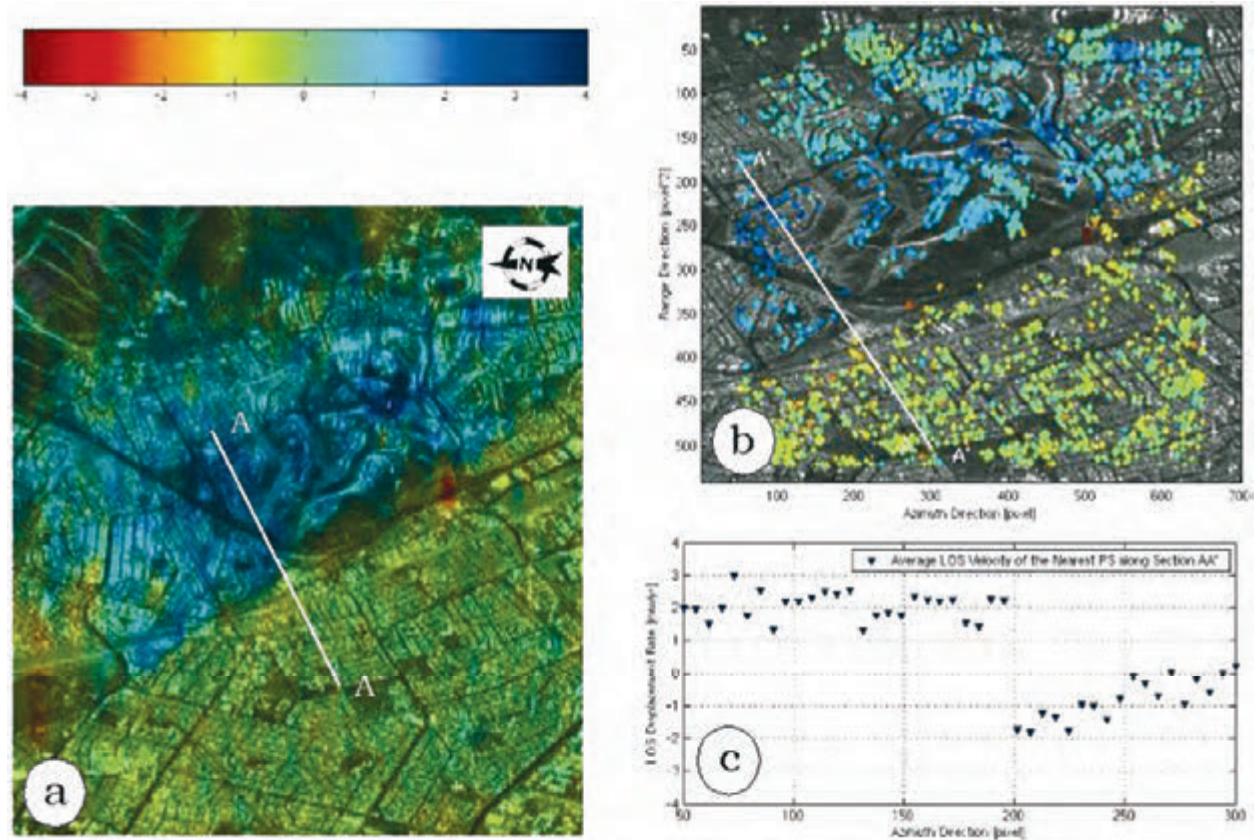


Figure 4-19: (a) Estimated LOS velocity field across the Hayward fault. (b) Close-up on cross section AA'. (c) LOS displacement rates relative to the PS along section AA'.

4.3.4 Seasonal deformation in the Santa Clara Valley

The PS technique can be exploited to identify seasonal deformation phenomena. As an example, the temporal analysis of the PS identified in the Santa Clara Valley (California) showed seasonal behaviour of the deformation due to the periodic water level variation, affecting an area limited by two known seismic faults.

The amplitude of this almost sinusoidal ground deformation is shown in Figure 4-20. Here, the border lines of the area affected by the seasonal deformation correspond to known seismic faults and are clearly visible on the colour image.

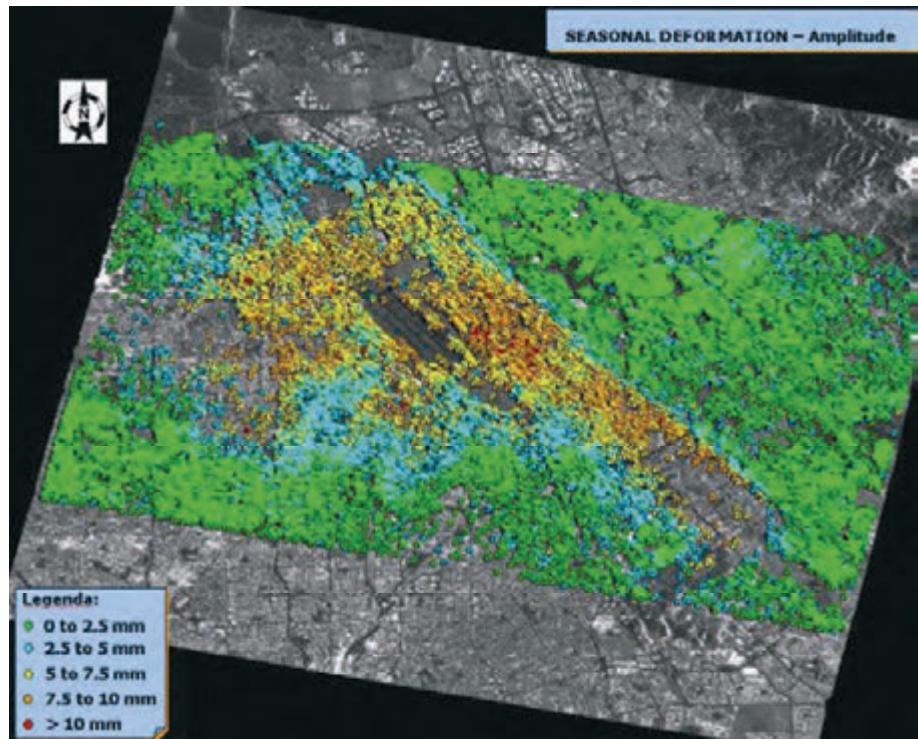


Figure 4-20: Map showing the amplitude of the seasonal ground deformation in the Santa Clara Valley (CA)

The typical behaviour of a PS in the area affected by seasonal motion is shown in Figure 4-21. In this analysis 46 ERS images, from May 1992 to September 2000, have been used.

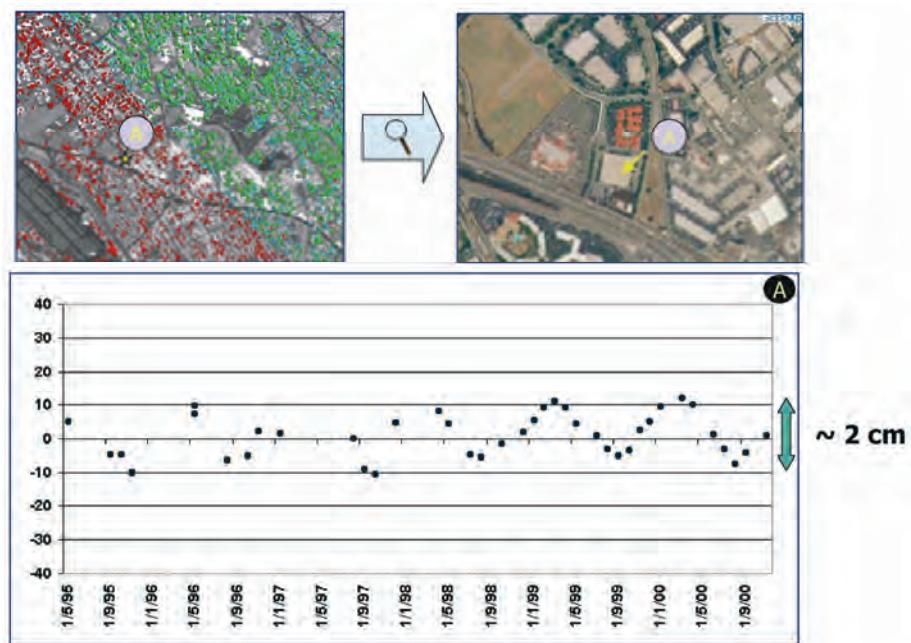


Figure 4-21: Typical behaviour of a PS in the area affected by seasonal motion

5. Applications based on spectral shift

It has been shown that spectral shift is an unavoidable consequence of the change in view angle (the baseline) that is always, to some degree, experienced in SAR interferometry. If not properly accounted for, the spectral shift causes an unwanted decorrelation, whose extent depends on the baseline and the local slope. In this chapter, we will discuss two different applications based on the spectral shift and its filtering.

- The **Interferometric Quick Look (IQL)** is based on an efficient implementation of the spectral shift filtering, discussed in section 2.2.1. In this specific case, a particular approximation of that technique is introduced, which sacrifices image quality with strong improvements of computational efficiency, in order to achieve real-time interferometry at coarse resolution.
- The **Super Resolution** technique takes advantage of the spectral shift to get a fine resolution image as a coherent combination of two or more interferometric images.

5.1 Introduction to spectral shift

Let us briefly review the principle of the spectral shift.

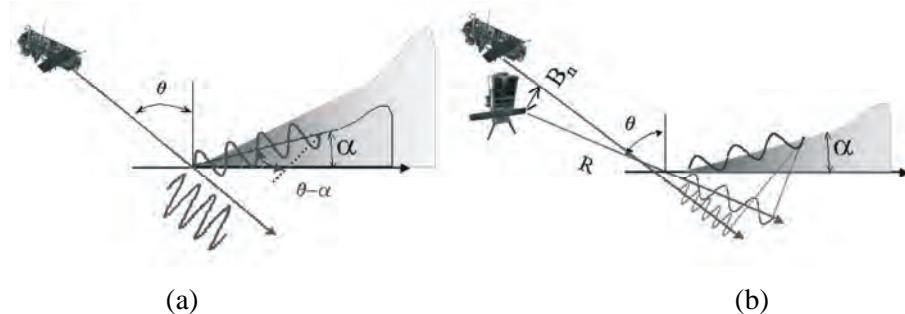


Figure 5-1: Spectral shift reference geometry. (a) The wave number on the ground changes with the slope. (b) Two sensors need to transmit with slightly shifted frequencies in order to observe the same wave number on the ground.

With reference to the SAR geometry in Figure 5-1 (a), we notice that a monochromatic plane wave, with wavelength λ , is projected onto the sloped terrain, as a ground wavelength:

$$\lambda_r = \frac{\lambda}{\sin(\theta - \alpha)} \quad \text{Equation 5.1}$$

i.e. a function of the original wavelength, and the local incidence angle ($\theta - \alpha$) (θ being the incidence angle for flat Earth and α the slope). A slight variation of the view angle, of an extent $\Delta\theta = B_n/R$, (B_n being the interferometer baseline), needs to be compensated by a shift in the transmitted frequency of Equation 1.9:

$$\Delta f = \frac{\partial f}{\partial \theta} \Delta \theta = -\frac{f \Delta \theta}{\tan(\theta - \alpha)} \quad \text{Equation 5.2}$$

This fact is shown in Figure 5-1(b). As already stated in chapter 1.1, interferometry is made possible by the actual finite bandwidth of the SAR system. As an example, the value of spectral shift has been plotted in Figure 5-2 as a function of the ground slope, for three different baselines. In the example, the baseline value of 1000 is defined as ‘critical’ since this would introduce a spectral shift, for a flat terrain, equal to the system bandwidth: fringes will vanish for larger baselines.

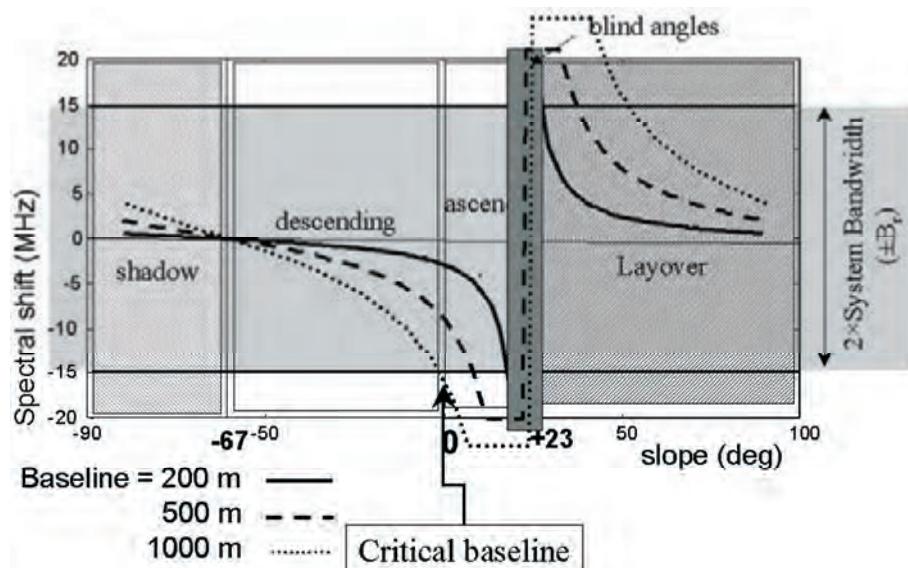


Figure 5-2: Spectral shift has been plotted as a function of the terrain slope, starting from back-slope (on the left) ending to the fore-slope, on the right. The ERS geometry was assumed, and three plots are drawn for different normal baselines. Dashed areas correspond to shadow and layover. The dark stripe in the middle marks the frequencies within \pm the system bandwidth: when the spectral shift exceeds that bandwidth, no correlated return is given.

The capability of getting interferometry over steep fore-slopes is thus linked to the amount of fractional bandwidth available. The fact is shown in Figure 5-3, where the common spectral components have been emphasised by the dark shading. If these components are selected by properly filtering the Master and Slave image, the spectra cross-correlations caused by the interferometric beat will result in an impulse, that is a linear phase in the time domain, e.g. the interferometric fringe represented in the plot of Figure 5-3 (on the right).

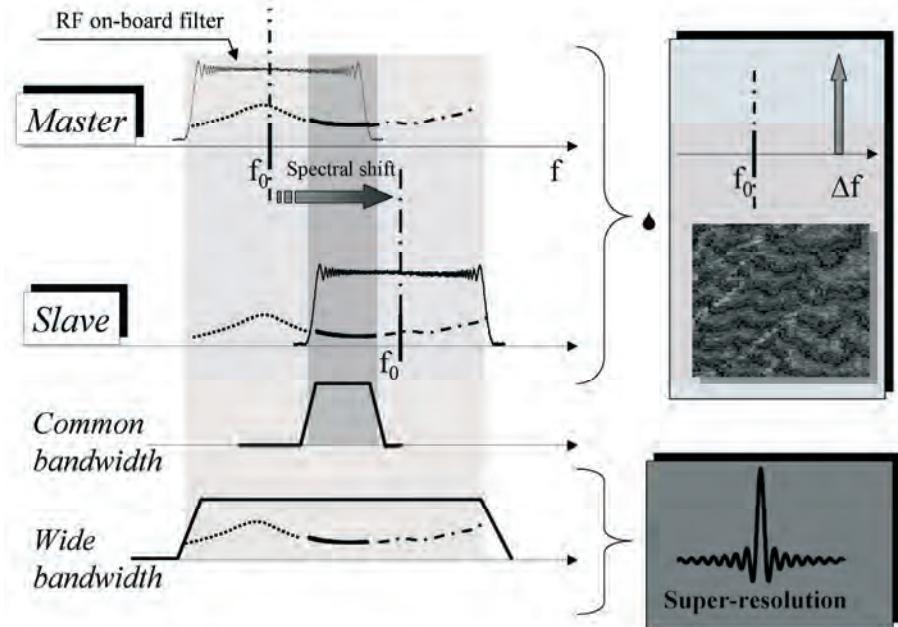


Figure 5-3: The master and slave image reflectivity are shifted by an amount that depends on the baseline. Their common spectral components (shaded in the first three plots on the left) are used for interferometry (top right). The other components can be combined to reconstruct a ‘wide bandwidth’ reflectivity, hence a super-resolution image (bottom left).

As a different and complementary application, the common and non-common spectral components can be added coherently to provide a large bandwidth, hence a super resolution image (see Figure 5-3 on the right).

5.2 Interferometric quick look (IQL)

IQL represents a complete interferometric processor intended to get medium resolution STRIPMAP interferograms with quite small computational complexity. The processor exploits the spectral shift principle in order to halve the data rate with only a moderate loss in the final quality. This idea, first expressed in literature in [MontiGuarnieri93, Gatelli94], was developed by Politecnico di Milano into a full processor, to be installed at ESA facilities for browsing purposes [Walker98].

The processor reduces the ‘computing’ costs (time, memory, disk-space) by moving the ‘spectral shift & common band filtering’ step, usually performed on the azimuth focused images (see chapter B2) *at the level of raw data*, i.e. as the *first step* of the interferometric processing chain. This is done by range half-band filtering and decimation. Usually decimation implies a loss of information; however, in this case, the range filters are tuned to the spectral shift (for flat Earth) so that most of the information removed would have contributed as noise in the final interferogram. According to Figure 5-2, a spectral shift of ~9 MHz (half of the sampling bandwidth) corresponds to a baseline of ~500 m (for flat terrain). When the interferometric baseline is 500 m, the quality available by the IQL will be close to that of the full

resolutionⁱⁱ. For smaller baselines, the half-bandwidth filtering results in some loss of quality with respect to a full resolution processing.

Besides range spectral shift, the IQL implements an efficient **azimuth common band filtering** (see chapter B1), once again applied directly to the raw data. This filtering is simply implemented by means of an $N = 8$ channel polyphase and DFT filter bank [Fliege94], that divides the raw data into eight azimuth looks, of which three are discarded and the other kept. Such processing is well suited to the ERS Tandem mission, which experiences a Doppler Shift of 300 Hz (~ 1/6 of the PRF).

The joint use of this unconventional scheme for range subsampling and azimuth look formation provides the following advantages:

- 1) the computing costs are reduced by a factor ~3 due to data decimation,
- 2) all the known advantages of SAR multi-look processors in terms of parallelism, scalability and efficiency have been achieved, and
- 3) the SNR is enhanced, thus improving performances of the required parameter estimates.

The IQL implementation also integrated several ‘tricks’ to gain efficiency, like the exploitation of small, optimised kernels for image co-registration, and an original, quick-and-dirty algorithm to compute coherence maps [MontiGuarnieri97].

An overview of the whole algorithm is provided in the schematic block diagram in Figure 5-4; more details will be found in the journal paper [MontiGuarnieri99B].

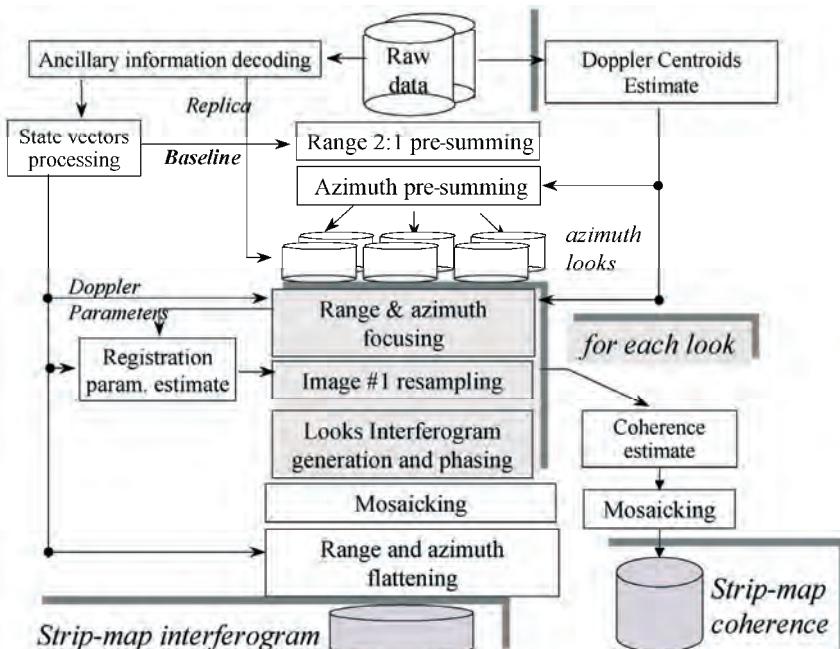


Figure 5-4: Schematic block diagram of the IQL processor

ⁱⁱ Note that the processor actually performs a spectral shift filtering tuned for flat Earth, which is not the best choice on a sloping terrain.

The processor was installed on a workstation at ESA/ESRIN in June 1996 to routinely produce quick-look tandem interferograms. The robustness of the algorithm, and specifically of the image alignment step, has been proven for non-contrasted areas (such as snow and ice), and also for low-coherence areas. The processor passed the ESA acceptance test at the beginning of 1997: the quality of resulting fringes was comparable with those achieved with a full bandwidth interferometric processor. The coherence maps were acceptable in order to discriminate the product quality. A wide variety of processed images is available on the ESA web server. An example of a strip interferogram and coherence map is in Figure 5-5: this specific example, on the area of Greenland, shows the potentials of SAR interferometry in monitoring ice detaching from the shelf.

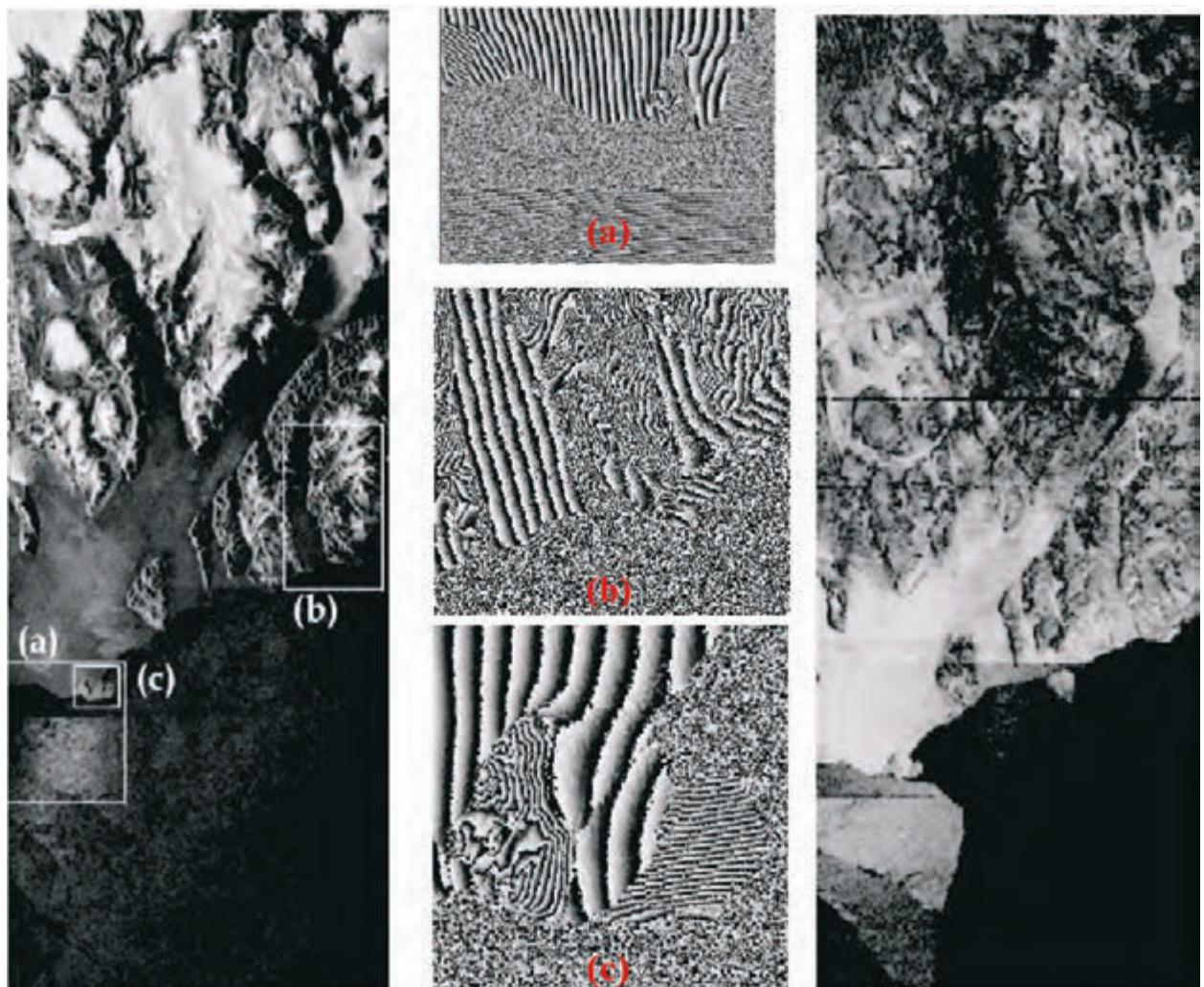


Figure 5-5: ERS tandem quick-look interferogram amplitude (left) and coherence map (right) of Greenland (100 × 385 km). Acquisition date 25 and 26 Jan. '96, baseline 130 m. Centre: interferogram phases of the three areas marked in the amplitude image (rescaled). Note the fringes in the sea and along the coast due to ice detaching from the shelf.

5.3 Super-resolution

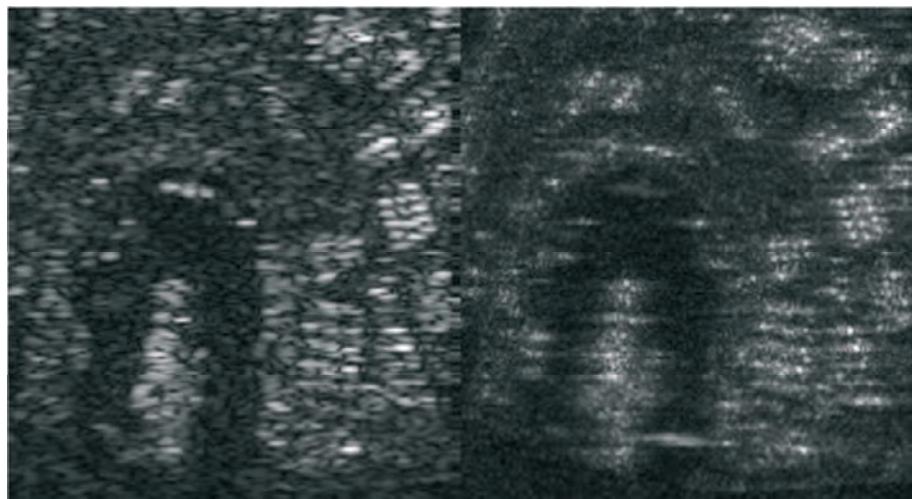
The range super-resolution technique is based on the observation that, when the same target is imaged from different view angles, all the views can be combined to get an image at an enhanced resolution. This principle is the very root of Synthetic Aperture Radar, and is furthermore exploited in SAR-SPOT systems that steer the antenna azimuth pattern in order to pinpoint a fixed target.

The extension to the range case, suggested for the first time in [Prati92], results in range super-resolution. The basic principle is the one shown in Figure 5-3 (lower part) for two images, but it can be extended to a large number of images. The proper in-phase summation of all the images is slightly complicated by the fact that the image should be corrected for a constant phase shift and also properly calibrated in amplitude. Notice that the in-phase summation of overlapped spectral contributions may cause spectral modulation effects, and these are quite complicated to compensate for as the cross-correlation of each spectral component is not known *a priori*.

The final result would be a single super-resolution complex image. The actual resolution enhancement would depend upon the baseline span and the local topography.

One of the major limitations of the technique in repeat pass SAR interferometry is represented by the temporal decorrelation that would arise in the long time interval between acquisitions (> 35 days for ERS). In fact, in processing with ERS datasets, this technique provided its best results in presence of point scatterers, which are most likely to correlate over the long term. For this reason, very good results are expected by applying super-resolution enhancements in simultaneous or in very short revisit time SAR interferometry, like for the **interferometric cart-wheel** combination proposed by D. Massonnet (see for example [Massonnet01]), or for the ERS-Envisat tandem discussed in chapter C7.

One of the better examples of the capability of this technique is provided in [Suess98]: a resolution enhancement of a factor 2.7 was obtained by exploiting the technique described in [Prati93, Gatelli94] and coherently averaging 9 images (see Figure 5-6).



*Figure 5-6: Super-resolution in urban areas: Harbour with oil storage, Amsterdam.
Left: incoherent average (9 images) Right: coherent average (range resolution
enhancement of a factor 2.7). Range is horizontal in the images.
[Images courtesy of M. Suess, Daimler-Benz Aerospace Dornier]*

6. Differential interferometry

6.1 Introduction

Distinguishing between ‘differential’ interferometry and ‘conventional’ interferometry is largely artificial, because interferometry is an intrinsically differential technique. Its ability to summarise all the geometric differences in range between two radar images is the result of several widely different phenomena. These phenomena can be regarded as different layers of information. Therefore, we must remove the various layers of geometric information one after another, in order to single out the one that is useful for a given application. As often in radar, the crucial word here is ambiguity. If there are n various causes of geometric difference involved in a final result (interferogram), one must know, in theory, the $n - 1$ first layers of information in order to acquire a good knowledge of the n^{th} layer.

In the most general case, four different information layers are stacked in an interferogram. In other words, $n = 4$. Of these layers, two arise from the difference of point of view between the two images used to form the interferogram. These are the **orbital residuals** and the **topographic residuals**. Two additional layers come from the difference between the acquisition times of the pair of images. These are the **displacements** and the **atmospheric artefact**.

Specially designed systems allow a reduction of the number of layers. For instance, the Shuttle Radar Topography Mission (SRTM), flown for the first time in 2000, eliminated the time difference by acquiring the two images simultaneously. Thus it reduced the number of layers to two (topography and orbital residuals). Furthermore, since the two antennae were linked by a physical beam, the degrees of freedom of the **orbital fringes** were themselves severely reduced. This system was therefore well suited for topographic mapping.

The ideal situation in which all the contributors to the geometric changes except one are perfectly known almost never occurs. The $n - 1$ information layers will usually leave residuals that will create unwanted artefacts. These unwanted layers will also govern the strategy of image selection, as the user will seek to select scenes whose conditions of acquisition will amplify the useful information and minimise the rest. In the next section, we will recall the orders of magnitude of the various layers.

The most basic step in interferometry consists of subtracting the phases between two radar images in order to get rid of the randomness built within each pixel by the layout of elementary targets. But this step is only the beginning of a long series of subtractions, indicating that interferometry is differential by nature.

In this section, in order to better understand differential interferometry, we will assume the ideal situation in which we know every detail of the geometry of the site and of the orbits during each data take. We will then

explore the ways in which the actual situation drifts away from this ideal and how we can characterise these drifts and learn from them.

6.2 Differential interferometry using an available DEM

In principle, building differential interferometry software is straightforward if reliable external sources of information exist. For instance, if the shape of the landscape is correctly known from a digital elevation model (DEM), we can screen each point in the DEM, once properly sampled to meet the user's needs, and define for each point the corresponding Cartesian vector \mathbf{P} in a given reference system. Assuming we accurately know the orbits of the two satellite passes involved in the formation of an interferogram in the same reference system, we can determine the closest distance \mathbf{r} to point vector \mathbf{P} and the corresponding satellite position vector \mathbf{S} and velocity vector \mathbf{V} .

It is clear that the above knowledge comes either from conventional geometry (such as computing \mathbf{P} for each point in the terrain grid), or from orbit propagation software (such as is required for determining \mathbf{r} , \mathbf{S} and \mathbf{V}). We label the two radar images involved in the interferogram 1 and 2, using the convention that 1 refers to the oldest image, or, if the two are acquired simultaneously, to the one used as a geometric reference (i.e. the phase in 2 is subtracted from the phase in 1).

For each point of our ground sampling grid we can obtain \mathbf{P} as well as \mathbf{S}_1 , \mathbf{V}_1 , \mathbf{r}_1 for the first image, and \mathbf{S}_2 , \mathbf{V}_2 , \mathbf{r}_2 for the second one. Writing software to do this can be simplified by reusing subroutines from geographic systems as well as from orbit computations. At the same time, we can easily define a unit vector \mathbf{U}_1 by:

$$\mathbf{U}_1 = \frac{\mathbf{S}_1 - \mathbf{P}}{\mathbf{r}_1} \quad \text{Equation 6.1}$$

We define a vector \mathbf{U}_2 similarly. These vectors point from the point of interest on the ground to the respective satellite's positions at closest range.

If we now use the expression $\text{mod}(\mathbf{r}, \frac{\lambda}{2})$ to mean the fractional part of \mathbf{r}

expressed in units of half the wavelength λ , then the interferogram between images 1 and 2 should in principle exhibit the following phase value:

$$\Delta\phi = \frac{4\pi}{\lambda} (\text{mod}(\mathbf{r}_1, \frac{\lambda}{2}) - \text{mod}(\mathbf{r}_2, \frac{\lambda}{2})) \quad \text{Equation 6.2}$$

We can remove $\Delta\phi$ from the interferogram, thus turning it into a *differential interferogram* as it is depicted in Figure 6-1. This operation involves comparing very different numbers, for \mathbf{r} amounts typically to 3×10^7 times half the wavelength for ERS-1. However using double precision numbers or similar precautions can take care of the potential rounding problems in a software program.

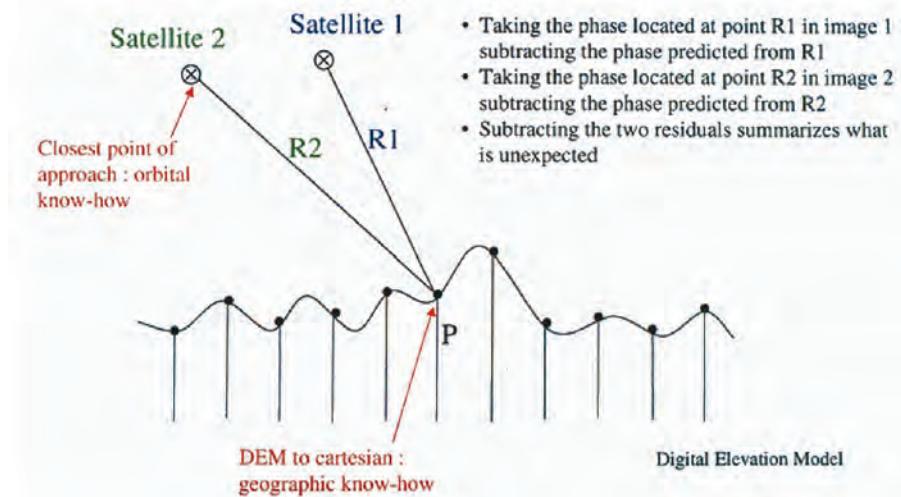


Figure 6-1: Drawing of 'perfect world' interferometry

Interferometry is a very straightforward technique. Assuming that the position of a satellite can be ascertained with arbitrary accuracy, so can the distance between it and any given point on the ground. We are then free to express this distance in units of wavelength. The expected 'geometric phase' can be computed for any point on the ground. However, this geometric phase cannot be extracted from a single image, where it is mixed with a seemingly arbitrary 'construction phase' which depends on the internal layout of elementary targets inside the pixel surface. Fortunately, if the interferometric conditions are met, the construction phase can be eliminated between two radar images, leaving the difference of the geometric phases. The difference of the geometric phases can also be simulated from the distances of the two satellite passes. The difference between the experimental phase difference and the simulated phase difference should leave a 'flat' interferogram as a result of a perfect knowledge of all the ground and orbital geometric features.

The sampling on the ground must obey two kinds of constraints: first, it is desirable to take advantage of the full content of information in the radar image. Therefore the ground sampling must be denser than the equivalent range and azimuth sampling of the radar, once it has been projected on the ground. Second, the signal represented by $\Delta\phi$ must be correctly sampled. In particular the change of $\Delta\phi$ from a point on the ground to any of its neighbours must stay below 2π . This is nothing more than the interferometric condition we have met before, in an earlier section (A.2.4).

For the sake of simplicity, we will assume that the generic differential processor consists, for each point on the ground, of the following four steps (this does not represent the most widely used procedure; however, an experimental processor [Massonnet00] implements it):

1. Selecting the position of the given point on the ground in range and time (i.e. across track and along track) in the two radar images to be combined,

2. Resampling the equivalent radar pixels in both images,
3. Subtracting the phases of these pixels and
4. Subtracting the geometric $\Delta\phi$ (the phase predicted by geometry).

The result of these steps is a map of the difference of phase in geographic coordinates, whose value should be zero if the assumptions for phase correction were fully accurate (i.e. the geometric prediction of the phase is exactly the observed phase difference between the images). In this case, the process would be useless because we would not have learnt anything from it!

The real world, however, is better represented by Figure 6-2, where all our initial assumptions are inaccurate to a certain extent.

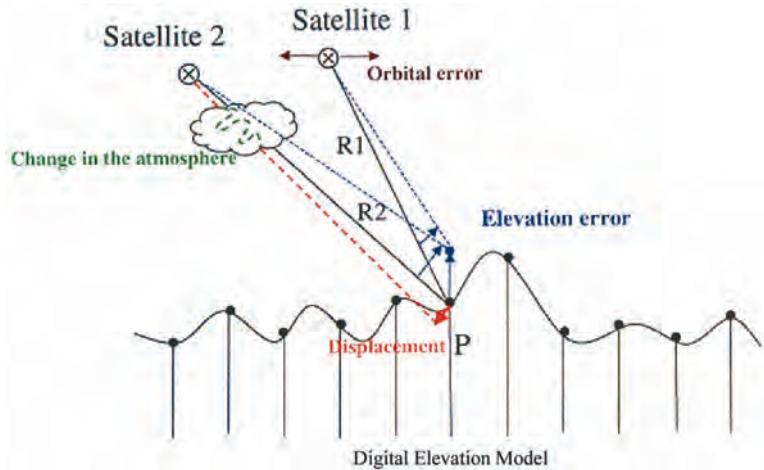


Figure 6-2: Categorising interferometric content

In the real world, four causes make the prediction of Figure 6-1 wrong; two of these are related to the different viewing geometries of the two satellites:

- 1) Inaccuracy of orbit knowledge (for both satellites)
- 2) Inaccuracy of elevation knowledge

Two of the causes are related to the time elapsed between the data acquisitions:

- 3) The change in the radio-electric ‘depth’ of the atmosphere, itself linked to changes in pressure and moisture or to turbulence
- 4) Displacements of the ground taking place between the imaging sessions

One can therefore say that, with respect to the point of view, interferometry must obey a condition (the similarity of the point of view of both images) and provides two information layers (the accurate orbital positioning and the topographic residuals); whereas, with respect to the elapsed time, interferometry must obey another condition (the surface conservation) and provides two additional information layers (the atmospheric depth change and the ground displacements). All these ‘layers’ exist together in any interferogram, with different scales and sensitivities, unless a specific design cuts some of these. This is the case for the SRTM mission: having no time

difference eliminates two layers, and the fixed-length beam reduces the degrees of freedom of the baseline (i.e. orbital) inaccuracy.

If our knowledge of the position of, say, the first satellite is wrong by $\delta\mathbf{S}_1$, this error causes a change of range of $\mathbf{U}_1 \cdot \delta\mathbf{S}_1$. The altitude of the point on the ground not being perfectly known also results in an error $\delta\mathbf{P}$. We assume that only the vertical component, δh , of this $\delta\mathbf{P}$ is not zero. If we denote the vertical component of \mathbf{U}_1 ' \mathbf{U}_{1u} ' (for 'up'), the error in altitude creates a change in range of the first image:

$$\delta\mathbf{r}_1 = \frac{\delta h}{(1 - U_{1u}^2)} \quad \text{Equation 6.3}$$

The topographic error *necessarily exists also in the second image*, creating an almost identical change $\delta\mathbf{r}_2$ obtained by changing \mathbf{U}_{1u} to \mathbf{U}_{2u} in the expression. Apart from this shift in positioning, the error in altitude creates a round-trip phase change in the interferogram, due to the misalignment of the two wave-fronts, equal to $(\delta\mathbf{r}_1 \mathbf{U}_{1u} - \delta\mathbf{r}_2 \mathbf{U}_{2u})$ or, in phase:

$$\Delta\phi = 4\pi \frac{\delta h (U_{1u} - U_{2u})}{\lambda (1 - U_{1u}^2)} \quad \text{Equation 6.4}$$

We denote the change of elevation capable of creating a full 2π phase cycle between the images as the 'altitude of ambiguity'. This is therefore equal to:

$$h_a = \frac{\lambda(1 - U_{1u}^2)}{2(U_{1u} - U_{2u})} \quad \text{Equation 6.5}$$

It is clear that since the vectors \mathbf{U} can be produced from procedures not related to any knowledge in the radar field, so can the altitude of ambiguity h_a too. For changes in the two distances \mathbf{U}_1 and \mathbf{U}_2 , it is only their difference that is accessible. This explains why interferometry is relatively insensitive to variations in topographic height. The angle $(\mathbf{S}_1, \mathbf{P}, \mathbf{S}_2)$ is so acute, constrained by the interferometric condition, that several tens or hundreds of metres are required to create a single topographic fringe, depending on the orbital separation. A good consequence of this, often surprising at first sight, is that a digital terrain model of only moderate accuracy is often sufficient to reduce the topographic content of an interferogram down to a negligible fraction of a phase cycle, thus leaving a 'clean' displacement map, accurate to millimetres.

The displacement in the terrain of point P occurring in the time elapsed between the acquisitions of the two images also creates a $\delta\mathbf{P}$ vector, which imposes changes *only on the second image*, creating a round-trip phase change in the interferogram equal to:

$$\Delta\phi = -4\pi \frac{\mathbf{U}_2 \cdot \delta\mathbf{P}}{\lambda} \quad \text{Equation 6.6}$$

As a consequence, the line-of-sight component of any displacement of the terrain is directly recorded as a phase shift, and unlike topography, displacement in the millimetre range can be observed and measured.

Finally, the refractive index of the atmosphere, for the entire path traversed by the radar waves during the acquisition of either image, creates a contribution to the difference of phase. In a subsequent section we will distinguish between the effect of static atmosphere over flat terrain, the effect of static atmosphere over mountainous terrain, and the effect of heterogeneous troposphere or ionosphere on the signal. These effects introduce phase shifts that generally remain within one ERS fringe (equivalent to a terrain height displacement of 28 mm), but might occasionally reach several fringes.

In the above discussion, we have seen how the pointing vector \mathbf{U} , a pure result of geographic and orbital knowledge, can be used to explain the behaviour of various geometric changes in an interferogram. The pointing vector also allows the definition of the limits of interferometric measurements to be examined. Let us define the ‘altitude of pixel’ h_p as the elevation change that creates a change of range equal to the size of a range pixel. Q is the ratio of the carrier frequency and the range sampling frequency of the radar (Q is 279.5 for ERS), and the one-way range pixel size is $\frac{Q\lambda}{2}$, hence:

$$h_p = \frac{Q\lambda}{2U_u} \quad \text{Equation 6.7}$$

The altitude of pixel can be related to the altitude of ambiguity:

$$h_a = \frac{h_{p1}h_{p2}}{Q(h_{p2} - h_{p1})} \left(1 - \left(\frac{Q\lambda}{2h_{p1}}\right)^2\right) \quad \text{Equation 6.8}$$

Considering the change of altitude corresponding to the range pixel of image 1, in the presence of a displacement *gradient* creating $\delta\mathbf{P}$ between each range pixel, and with a *slope* creating $\delta\mathbf{h}$ between each range pixel, the global condition for which interferometry is feasible is:

$$\frac{h_p}{h_a} + \frac{\delta\mathbf{h}}{h_a} - 2 \frac{\mathbf{U}_2 \bullet \delta\mathbf{P}}{\lambda} < 1 \quad \text{Equation 6.9}$$

It is theoretically possible that a pair of images outside this interferometric domain (for instance because $\frac{h_p}{h_a} > 1$) could nonetheless produce a local

interferogram because of a local displacement gradient that counteracts the first term. Alternatively, a strong deformation gradient may push an interferogram outside the interferometric domain, even if the satellite orbital baseline separation is small. This may happen in the vicinity of terrain faults or in glaciers.

6.3 Differential interferometry with three or more combined images

A ‘radar only’ approach to producing differential interferograms without using any external information is very appealing. It requires at least *three* radar images.

Figure 6-3 shows the assumed geometry of a landscape captured by any two radar images, labelled Master and Slave. The third image (and any additional images) would be similar to the Slave, but is not shown in order to simplify the drawing.

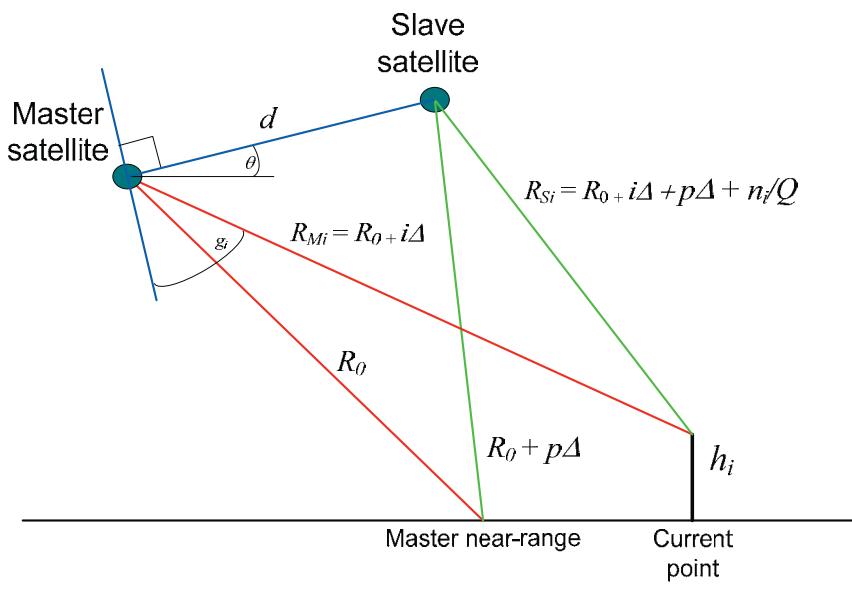


Figure 6-3: Description of interferometric geometry

In the master image, from the near-range to the i^{th} pixel, the distance changes from R_0 to $R_0+i\Delta$, Δ being the range pixel size. In the slave image, the point observed at near-range by the master is actually offset by p pixels (p can be positive or negative, and is generally not an integer, unlike i). In the current point with altitude h_i , the range to the slave track equals the one of the master, plus the near range offset, plus the number of fringes counted from the near-range to point i . The fringe count must be divided by the factor Q to be expressed in units of range pixels. The distance between the tracks is d , angles θ and g_i , used in the development, complete the geometric description.

Of course, in reality the Earth is not flat, but taking a ‘flat’ reference (for instance given by the plane tangential to the surface at mid-swath) is more convenient and does not affect the principles underlying the following discussion. As usual, we choose one of the images arbitrarily as a geometric reference (the *master* image). The others will be called, respectively, the *slave* image and the *complementary* image. Their values are denoted by m , s

and c , respectively. The range pixels for the master image are denoted i . We have, from a straightforward calculation:

$$\sin g_i = \frac{R_{si}^2 - \delta^2 - R_{mi}^2}{2\delta R_{mi}} \quad \text{Equation 6.10}$$

We express all lengths as multiples of the pixel size, Δ , such that $\Delta = \frac{Q\lambda}{2}$

(Q , which represents the number of fringes per slant-range pixel, has the value 279.5 in the case of ERS). Letting R_0 be the near-range of the master image, we have:

$$R_{mi} = R_0 + i\Delta \quad \text{Equation 6.11(a)}$$

$$R_{si} = R_0 + i\Delta + p\Delta + \frac{n_i}{Q} \quad \text{Equation 6.11(b)}$$

where:

p is the difference, expressed as a number of pixels, between the slave image near range and the master image near range. This quantity is obtained by prior correlation of the images.

n_i is the number of fringes (of course not necessarily an integer), counted from the near range of the interferogram to pixel i , so that $\frac{n_i}{Q}$ is the number of ‘equivalent’ pixels.

Thus, n and g are functions of i , but p is not. Continuing further:

$$\sin g_i = \frac{p\Delta + \frac{\lambda n_i}{2}}{\delta} + \frac{(p\Delta + \frac{\lambda n_i}{2})^2}{2\delta(R_0 + i\Delta)} - \frac{\delta}{2(R_0 + i\Delta)} \quad \text{Eq. 6.12}$$

The n_i fringes counted at point i determine the local elevation h_i at master range i . We can predict the fringes produced by the orbital configuration of a pair of images on a landscape by means of a second pair. This is an attempt to ‘simplify’ the altitude effects by making two interferometric pairs from three images. Supposing that the value of g_i is known, we may infer n_i :

$$\frac{n_i}{Q} = \sqrt{(R_0 + i\Delta)^2 + \delta^2 + 2\delta(R_0 + i\Delta)\sin(g_i)} - R_0 + i\Delta + p\Delta \quad \text{Eq. 6.13}$$

This formula allows the fringe pattern of one interferogram to be predicted from a second one, so long as the second is due only to relief (i.e. there is no effect from atmosphere, no displacement and no unwrapping error). This subtraction could be performed again with any additional complementary image after it has formed an interferogram with the same master image.

In practice, using three radar images, we construct two interferograms with the same master. We give the two interferograms the indexes s and c , for slave and complementary respectively.

First, we calculate $g_s(i)$ from p_s , $n_s(i)$ and d_s , which are the values for the first interferogram (master-slave).

Next, we define the look angle $g_s(i) - \theta_s$, which is equal to $g_c(i) - \theta_c$ because the master radar beam direction is common to the two interferograms. In practice, only the difference $\theta_c - \theta_s$ is needed for this operation. Finally, we recalculate $\sin(g_c)$ by means of δ_c and p_c to calculate $\frac{n_c(i)}{Q}$, from which $n_c(i)$ follows.

We may subtract these calculated fringes from the second interferogram, leaving only the phase shifts due to changes on the ground. The second interferogram does not need to be unwrapped at this stage. The advantage of the method is that it does not require external data other than radar images. The first drawback of the method is that it assumes the geometric differences between master and slave to be due only to topography, which is usually wrong to some extent due to atmospheric artefacts.

Another drawback is the difficulty in deciding whether movements in the terrain have occurred in the master-slave image pair or the master-complement pair. Both can influence the final result. This drawback is shared by the integer combination technique (see below).

Finally, it is necessary to resolve the phase ambiguity (see section “phase unwrapping”) prior to the application of the technique.

6.4 Techniques to avoid phase unwrapping

6.4.1 Integer combination

The ambiguous nature of the fringes within a wrapped interferogram make it of limited practical value. In particular, no linear combination nor any kind of scaling can be undertaken on a ‘wrapped’ interferogram. Despite the improvements of various methods for phase unwrapping, an active field of research that goes beyond radar imagery, unwrapping is still difficult, risky and damaging. It is risky because of the possible introduction of errors that will be incorporated in any further processing steps and which will become impossible to recognise easily after any scaling of the resulting interferogram. It might also be damaging because the software will be forced to black out parts of the image where the coherence is too low. The areas that cannot be unwrapped have poor coherence, but often contain some information that will be lost by this thresholding process. After correction for topographic and orbital contributions, the interferogram contains three kinds of information:

1. The fluctuation in the number of fringes due to the noise ν , either from thermal noise or partial incoherence of the terrain in the time elapsed
2. The number of fringes $\frac{\varepsilon}{h_a}$ due to the errors ε in the DEM, where h_a is the altitude of ambiguity.

3. The number of fringes μ caused by any differential effects, including local variations in the atmospheric thickness, as well as local ground motion between the data takes.

All three quantities (ν , ε and μ) vary greatly across the image. In contrast, h_a is rather stable, usually varying gently with range and not at all with azimuth.

The interferogram in which the effects of topography and the ‘orbital fringes’ have been retrieved gives us the fractional part of:

$$\nu + \mu + \frac{\varepsilon}{h_a} \quad \text{Equation 6.14}$$

It cannot be mathematically manipulated by the formulas of the double difference unless it has been unwrapped. However, we can manipulate the interferogram while preserving its integer and fractional parts. If we multiply the interferogram by an integer q_1 , we will likewise multiply the number of phase transitions and the initial integer values. Note that the integer values remain unknown unless unwrapping has been performed. We obtain the fractional part of:

$$q_1\nu + q_1\mu + \frac{q_1\varepsilon}{e_a} \quad \text{Equation 6.15}$$

If we combine the first interferogram with a second, characterised by another independent noise distribution, another altitude of ambiguity, other differential effects and another multiplication factor q_2 , but the same local DEM error ε , we obtain the fractional part of:

$$(q_1\nu_1 + q_2\nu_2) + (q_1\mu_1 + q_2\mu_2) + \varepsilon \left(\frac{q_1}{h_{a1}} + \frac{q_2}{h_{a2}} \right) \quad \text{Equation 6.16}$$

where the quantities related to the first and second interferograms are indexed by 1 and 2 respectively.

As a consequence of the multiplications, the RMS amplitude of the noise becomes:

$$\sqrt{q_1^2 n_1^2 + q_2^2 n_2^2} \quad \text{Equation 6.17}$$

This resulting increase of the noise in practice limits the choice of the absolute value of q_1 or q_2 to less than 3. Otherwise, once multiplied by a typical factor of $\sqrt{q_1^2 + q_2^2}$, the noise distribution could too easily exceed one cycle, and the result would be useless. Noise is usually responsible for a phase uncertainty equivalent to a few percent of a cycle.

The interferogram resulting from this kind of combination exhibits an equivalent altitude of ambiguity h_{ae} , written as:

$$\frac{1}{h_{ae}} = \frac{q_1}{h_{a1}} + \frac{q_2}{h_{a2}} \quad \text{Equation 6.18}$$

The consequences of the above formula are easy to understand. Phase unwrapping is only required for altitudes of ambiguity lower than the typical accuracy of globally available Digital Elevation Models and higher than the ones set by the interferometric limit. Formerly this ranged from 200 m (the accuracy of GLOBE dataset) to 10 m (the interferometric limit of ERS). Now the altitudes of ambiguities for which unwrapping might be required range from 30 m (accuracy of the publicly available global DEM produced by the Shuttle Radar Topography Mission) to a few metres. As a consequence, even if the choice of q_1 and q_2 is limited to ± 1 , ± 2 or ± 3 , there will be combinations which can emulate equivalent altitudes of ambiguity larger than 30 m most of the time. Even if the resulting interferograms are generated in a map-projected geometry, which means independently of the radar image geometry, it is possible to use the technique between pairs from ascending and descending orbits or between pairs from different satellites.

The technique of integer combination often contradicts intuitive expectations about fringe behaviour. After modelling the dense fringe pattern (Figure 6-4(a)) with a horizontal fringe pattern of the same density (Figure 6-4 (b)), we subtract the horizontal fringe model from the initial fringe pattern and obtain, surprisingly, a purely vertical pattern (Figure 6-4 (c)).

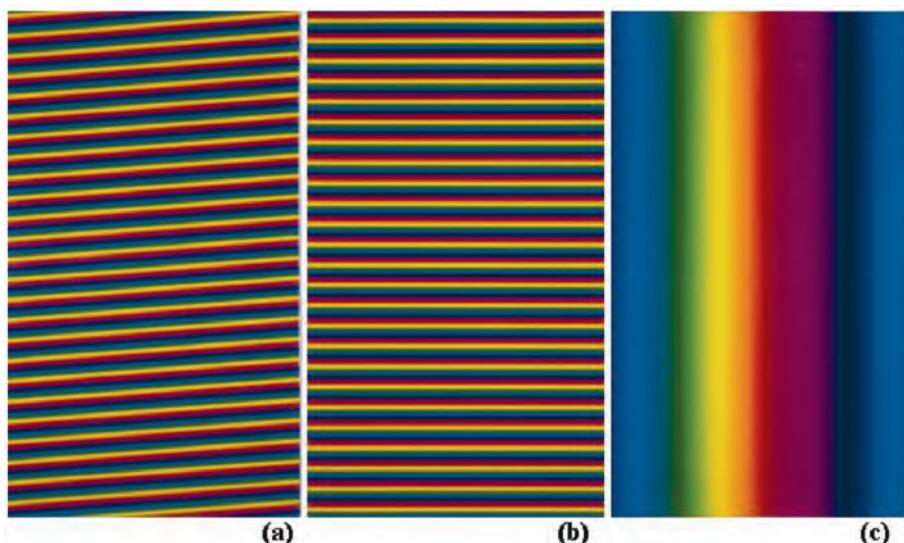


Figure 6-4: Illustration of interferometric combination. Set of fringes (a to c) that contradicts ‘natural’ opinion about fringe behaviour.

Similarly, even ambiguous (i.e. ‘wrapped’) interferograms can be manipulated by linear combination provided they are multiplied by whole numbers. In an ambiguous interferogram, we can basically ignore the integer part of the fringes. This part remains an integer only if it is multiplied by an integer. We simulate an interferogram on Mount Etna with a topographic sensitivity of 130 m per fringe (Figure 6-5(d)). We then simulate another interferogram with a sensitivity of -255 m per fringe (Figure 6-5(e)). Multiplying the first interferogram by two (which doubles the number of fringes) and adding the result to the first interferogram creates a new interferogram with a topographic dependence decreased to -6630 m per fringe (Figure 6-5(f)).

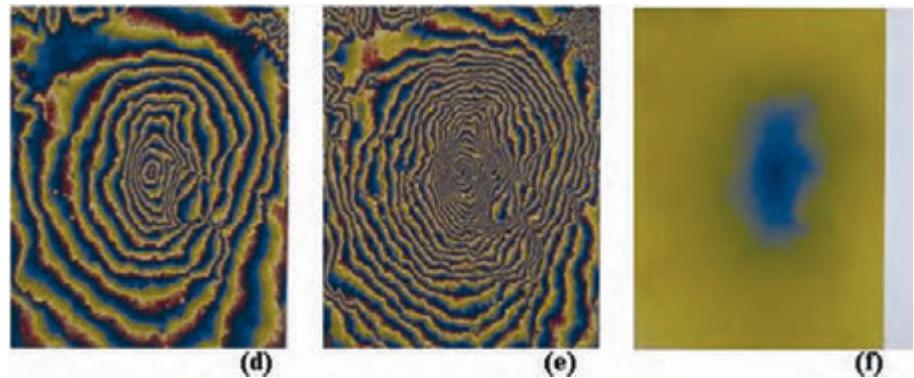


Figure 6-5: Illustration of interferometric combination (d to f). In d), we have a simulated mountain with a topographic sensitivity of 130 metre-per-fringe. In e), we have a sensitivity of -255 metre-per-fringe. Simply adding twice e) to d) provides an image of the simulated mountain featuring a topographic dependence decreased to -6630 metre-per-fringe seen in f).

In this last image, unwrapping becomes unnecessary because the mountain is not as high as the elevation represented by a single fringe. The combined interferogram (Figure 6-5(f)) can therefore be multiplied by a non-integer coefficient in order to automatically unwrap the image of (Figure 6-5(d)) (the most sensitive to topography) and to obtain an accurate elevation map.

6.4.2 Interferogram stacking

Another useful idea to bypass, at least partially, the need for phase unwrapping is **interferometric stacking** [Sandwell98]. According to this procedure, each interferogram of a series is first derived (i.e. the gradients of the phase are computed). The phase gradients are not ambiguous and can be scaled according to the orbital separation in order to reach a normalised topographic sensitivity. After being scaled, the gradients of the series are stacked and averaged. At this stage, this or that contributor (for instance the topographic contribution) to the interferometric information can be removed using *a priori* knowledge. During this operation, one may reasonably hope that atmospheric residuals are efficiently attenuated and that their averaged gradient is close to zero. The average interferogram is then reconstructed by integration. Global, FFT-based reconstruction methods are especially efficient for unwrapping gradients with values close to zero. Similarly to the integer combination method, the method can be used not only to cancel out topography, but also to eliminate deformation gradient proportional to time, by proper scaling, or any other specific contributor.

The drawback of the method is the difficulty in computing gradients on noisy phase images, which requires delicate pre-filtering. Sandwell et al. describe the necessary precautions to be taken.

6.5 Information contained in interferometric measurements

6.5.1 Residual orbital fringes

Usually the first term to be corrected right after the production of an interferogram corresponds to residual orbital fringes. These fringes are generally the dominant signal in the raw differential interferogram, and even a very good assessment of orbital position can lead to a significant number of residual fringes. Using the notations of previous sections, an error δS on the satellite's position creates a number of fringes f amounting to:

$$f = 2 \frac{\mathbf{U} \bullet \delta \mathbf{S}}{\lambda} \quad \text{Equation 6.19}$$

Let us assume that we can assess the number of residual fringes, both at near range f_n and at far range f_f , together with the corresponding pointing vectors \mathbf{U}_n and \mathbf{U}_f , themselves almost insensitive to the changes caused by δS . Since \mathbf{U}_n and \mathbf{U}_f are not collinear, we may assume:

$$\delta S = \alpha \mathbf{U}_n + \beta \mathbf{U}_f \quad \text{Equation 6.20}$$

which leads to:

$$\lambda f_n = 2\alpha + 2\beta \mathbf{U}_f \bullet \mathbf{U}_n \quad \text{Equation 6.21 (a)}$$

$$\lambda f_f = 2\alpha \mathbf{U}_n \bullet \mathbf{U}_f + 2\beta \quad \text{Equation 6.21(b)}$$

The system yields α and β and finally δS . A first problem is that we do not know f_n and f_f but only their difference $f_f - f_n$. A quick fix would be to assume $f_n = 0$, but what are the consequences? To answer the question, let us assume that δS results from $f_n = 0$ and $f_f = f$ and that $\delta S'$ results from $f_n = n$ and $f_f = f + n$. By definition, δS and $\delta S'$ are equally valid at far range and near range. Let us assume an intermediate range where the pointing vector is written:

$$\mathbf{U} = x \mathbf{U}_n + y \mathbf{U}_f \quad \text{Equation 6.22}$$

with the unit condition:

$$\mathbf{U}^2 = x^2 + y^2 + 2xy \mathbf{U}_f \bullet \mathbf{U}_n = 1 \quad \text{Equation 6.23}$$

The difference in correction for \mathbf{U} depends on whether it is computed from δS or $\delta S'$:

$$\frac{\lambda}{2} (\delta S - \delta S') \bullet \mathbf{U} = \frac{\lambda}{2} (\delta S - \delta S') \bullet (x \mathbf{U}_n + y \mathbf{U}_f) \quad \text{Equation 6.24}$$

Using the relations:

$$\begin{aligned} \frac{\lambda}{2} (\delta S \bullet \mathbf{U}_n) &= 0; \frac{\lambda}{2} (\delta S \bullet \mathbf{U}_f) = f; \\ \frac{\lambda}{2} (\delta S' \bullet \mathbf{U}_n) &= n; \frac{\lambda}{2} (\delta S' \bullet \mathbf{U}_f) = f + n \end{aligned} \quad \text{Equation 6.25}$$

we have:

$$\frac{\lambda}{2}(\delta S - \delta S') \bullet U = yf - xn - y(f+n) = -n(x+y) \quad \text{Equation 6.26}$$

We assumed a shift of n fringes between the two corrections. The unexpected difference between the corrections, which we later call the ‘error’, is therefore: $n(1-x-y)$ fringes. Let us now observe that $\mathbf{U}_f \bullet \mathbf{U}_n$ is the cosine of the antenna aperture of the radar in elevation (pointing difference between near range and far range). If W is the useful height of the antenna, that is about one metre for ERS, we have:

$$\mathbf{U}_f \bullet \mathbf{U}_n \approx 1 - \frac{\lambda^2}{2W^2} \quad \text{Equation 6.27}$$

hence:

$$(x+y)^2 \approx 1 + \frac{xy\lambda^2}{W^2} \quad \text{Equation 6.28(a)}$$

$$x+y \approx 1 + \frac{xy\lambda^2}{2W^2} \quad \text{Equation 6.28(b)}$$

leading to an error of magnitude: $n \frac{xy\lambda^2}{2W^2}$, $x+y$ being close to one. The maximum is obtained for $x=y=0.5$, that is for the intermediate point between far and near range. The worst case error becomes: $\frac{n\lambda^2}{8W^2}$; we want this ‘error’ to remain a small fraction of the wavelength. We choose for instance: $\frac{n\lambda^2}{8W^2} < \frac{\lambda}{32}$

Finally, we observe that n fringes is the error made on the estimation of the range of the second image when it corresponds to the near-range of the first image. The absolute number of fringes has no importance as long as it is

$$\text{such that: } n < \frac{W^2}{4\lambda}$$

For ERS-1, n fringes must be smaller than 4.5 m to obey the above condition. Since the interferometric processing requires the two images to be superposable to better than a fraction of the range (or azimuth pixel), we may consider that it is always met (ERS range pixel size is 7.9 m).

A second problem is that we do not know which orbit to move, that is, how the error is distributed between the orbits. In order to know more about this, we will use the following trick: we take two widely separated ERS orbits in order to be in a ‘worst case’ situation. We then generate a fake interferogram: a constant image filled with zeroes. We then use an ‘orbital tuning’ program to correct the fake interferogram by a number of residual fringes typical of the accurate assessment of ERS’s orbital paths. We correct the interferogram with a typical ‘orbital’ fringe pattern, amounting to zero,

eleven, six and a half and fifteen fringes at each corner. As a consequence, our interferogram, initially ‘flat’, exhibits this number of fringes (shown in Figure 6-6(a)). We ask the program to execute this first correction by moving only the first orbit.

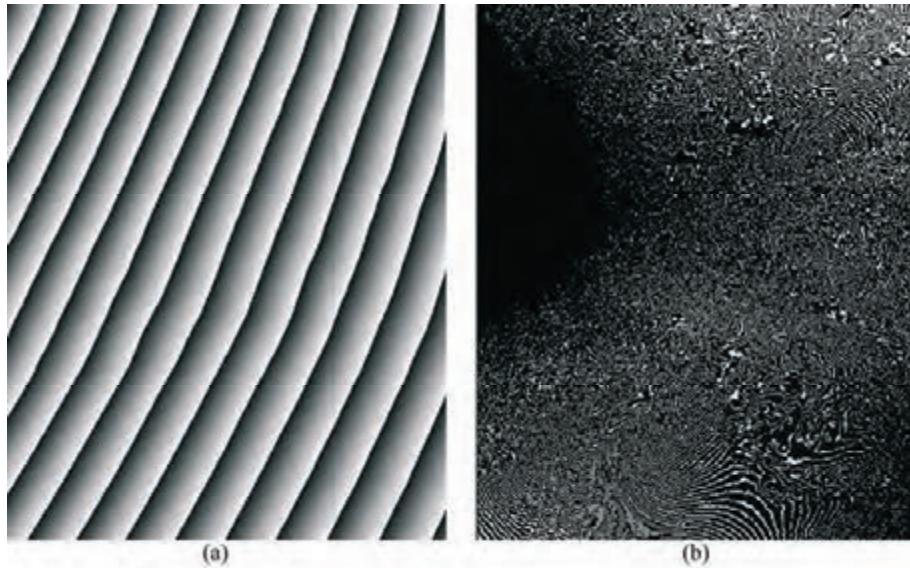


Figure 6-6: Illustration of the error in orbital tuning

Next, we run the program again in order to eliminate the fringes we just created, but we ask the program to correct only the second orbit. This time, we correct the four corners of the interferogram by the opposite quantities, respectively, by zero, minus eleven, minus six and a half and minus fifteen fringes. The result (Figure 6-6(b)) is flat to within the maximum error that can be committed by a wrong sharing of the responsibility between orbits. The figure clearly shows that this error is not a matter of concern (typically one grey level). Such a conclusion would not apply if the initial orbital error could have produced tens of residual fringes.

Orbital tuning provides a unique solution if one of the trajectories can be said to take all the error for itself. In reality of course both trajectories are faulty to some extent. What is the residual error we commit in correcting one trajectory for the full error? To assess this we take two orbits close to the maximum orbital separation permitted by the principles of interferometry, then we ‘correct’ an initially flat interferogram, introducing a number of fringes by moving one of the orbits (part a). We then correct the fringes, considered as an artefact, by modifying the second orbit. The result (part b) is representative of the maximum error that could be made using the ‘single orbit hypothesis’. In order to make the result visible, the error image (b) has been multiplied by 256, so that the maximum error (the white surfaces) corresponds to 1/256 cycle, well below any reasonable accuracy threshold of SAR interferometry. The structure of the error is mainly governed by the

topography of the area. A high relief area has been selected in this test, run by H. Vadon of CNESⁱⁱⁱ.

6.5.2 Uncorrected topography

Some topographic residuals arise from uncorrected topography or errors in the DEM used. The errors are scaled by the topographic sensitivities of the pairs involved, which is the basis of the pair-wise logic that can be used to discriminate these effects from others.

6.5.3 Heterogeneous troposphere

The contribution of heterogeneous atmosphere was first detected in 1994 [Massonnet94], and quickly confirmed as the major limitation of interferometric accuracy for displacement mapping as well as for topographic reconstruction. The signature of this artefact is linked to the date of the data takes. A given image will propagate the structures created by various delays through clouds, pressure waves, etc. in any interferogram in which the image is a part constituent. The structure would always have the same amplitude, although its sign would change depending on whether the faulty image is used as master or slave in the combination. This behaviour allows an easy screening of atmospheric effects if several interferograms are available on a given site.

Figure 6-7 shows the effect of cloud ‘rolls’ in eastern United States. To the experienced eye, the atmospheric effects can be recognised directly (i.e. even if a single interferogram is available), because they have typical shapes and ‘ragged’ borders that are not usually found in topographic or displacement signals. Figure 6-7 is an example of visualisation of meteorological phenomena through radar interferometry. These phenomena are characterised by their insensitivity to the time elapsed between the acquisition of the images used to form the interferogram, as well as their insensitivity to the orbital separation of the corresponding satellite tracks. However, a given radar image is responsible for the same structure observed in several interferograms with a constant magnitude.

ⁱⁱⁱ By personal communication

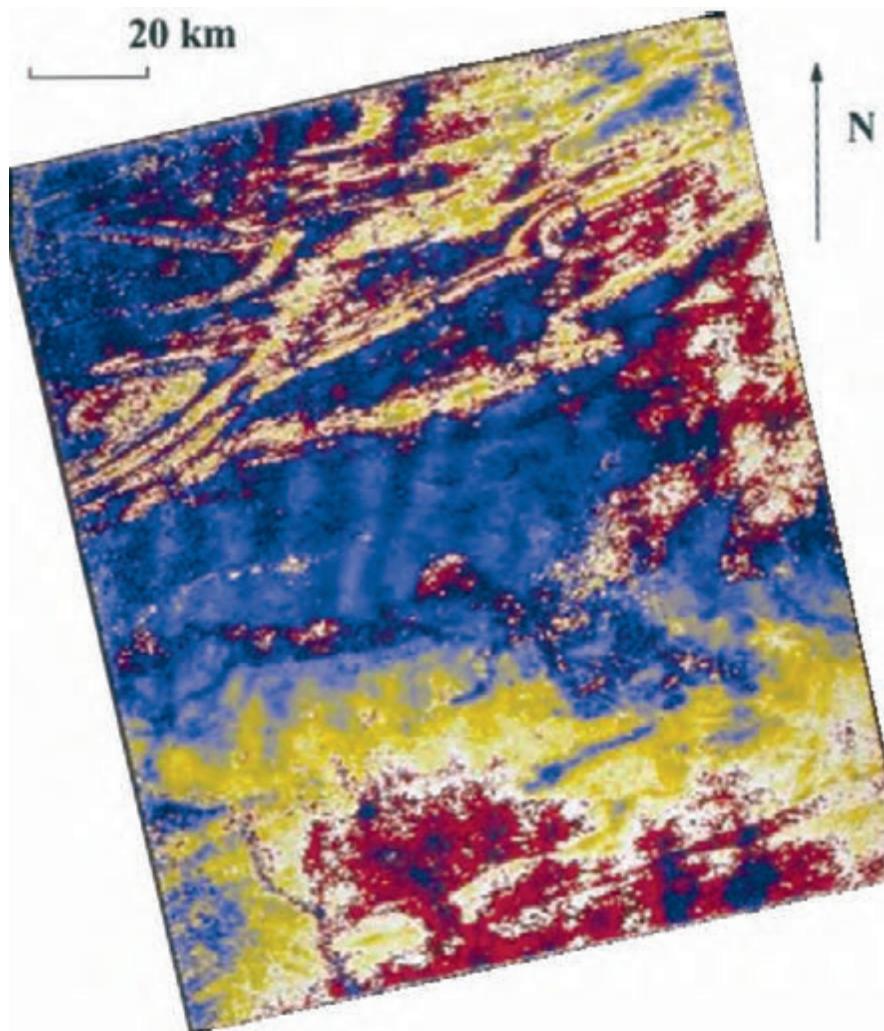


Figure 6-7: Atmospheric heterogeneity, cloud ‘rolls’ in eastern US. Here we see the effect of atmospheric turbulence generated by winds interacting with topography. The topographic contribution from the Blue Mountains, east of Washington, has not been corrected. The 1000 m or so high mountains can hardly create one fringe in this interferogram that is not very sensitive to topography. However, the wind ‘rolls’ as it crosses the topography, creating waves that are easy to see in dark blue. Their amplitude is very low – 3 mm crest to crest – but their structure makes them easy to spot. The images were separated by 3 days.

6.5.4 Heterogeneous ionosphere

The ionosphere is likely to alter the signal in a similar manner to the troposphere [Massonnet94], except that the ionosphere is a much thicker layer than the troposphere and actually extends up to the satellite’s orbit. As a consequence, it is more difficult to relate external data (such as GPS paths) to an artefact. However, the heterogeneity of the ionosphere is less than that of the troposphere and is not influenced by the topography. Similarly, the ‘static ionosphere’ does not create artefacts, because it is above any topography. The way to discriminate the effect of the ionosphere is the same as for detecting a heterogeneous troposphere.

6.5.5 Static atmosphere

Since the refractive index of the atmosphere is not exactly equal to one, all rays emitted by a radar undergo a curvature which results in a consistent apparent displacement of the satellite [Tarayre96]. However, if some topography is present in the scene, the depth of atmosphere effectively crossed by the wave varies. If the general features of the atmosphere (e.g. the moisture density) change between the two data takes used to form an interferogram, a difference in delay will appear, modulated by the topography. The signature of this effect is an interferogram in which structures linked to topography appear, but without any relationship to the orbital separation of the pair. This ‘topography’ would be seen even without any orbital separation.

6.5.6 Radar clock drift

In the course of interferometric experiments with ERS-1 over very long data strips, some parallel fringes were found to remain even after orbital fringes had been removed. Although it is difficult to prove firmly the origin and nature of these remaining fringes, a possible explanation is that this is due to a frequency drift of the local oscillator onboard the radar [Massonnet95B].

We assume the carrier frequency is:

$$f_c = f_0 + t f_b \quad \text{Equation 6.29}$$

where f_b is a time-dependent bias of the ‘official’ frequency f_0 (5.3 GHz for ERS, used in the experiment). The bias can create ‘clock fringes’ if the carrier frequency has drifted somewhat when the pulse returns. It is then mixed with a slightly different frequency, typically after a number k of interpulse periods (k equals 9 for ERS). If f_a is the pulse repetition frequency, the fringe rate, expressed in fringes per second, is:

$$\frac{k f_b}{f_a} \quad \text{Equation 6.30}$$

If data from ERS-1 and ERS-2 are mixed in an interferogram, only the difference of their frequency biases will contribute to clock fringes. A difference in their f_0 would be cancelled when remixed. So there is no real difference between mixing ERS-1 with ERS-2 or working with the same satellite. In either case, the fringe rate would be:

$$\frac{k(f_{b1} - f_{b2})}{f_a} \quad \text{Equation 6.31}$$

Several attempts to find more examples of the ‘clock artefact’, in particular for ERS-2, failed. The only example remains the one which was obtained by combining two very long data strips (about 2000 km) over eastern Europe (Figure 6-8). The data were acquired on 25 September and 1 October 1991 and showed very little sensitivity to topography.

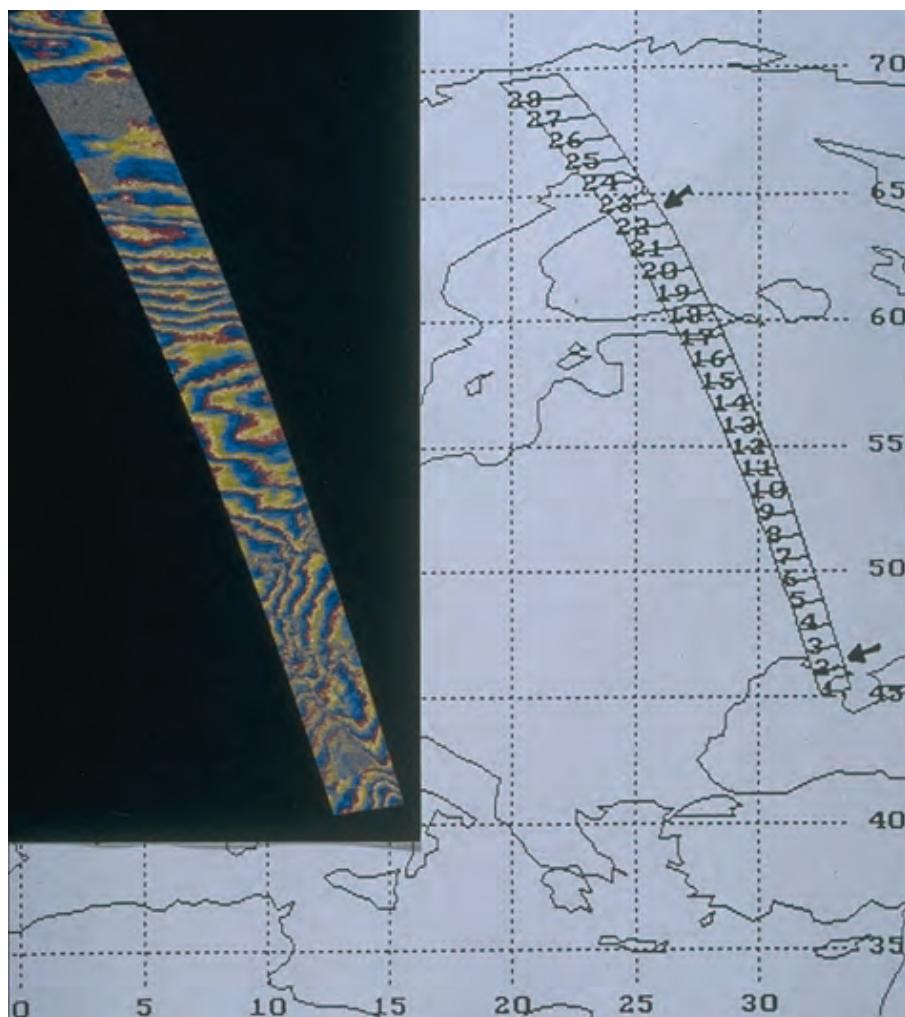


Figure 6-8: Processing a long strip of ERS-1 data from Ukraine to Finland that exhibits a series of fringes that cannot be explained by topography (the whole area is rather flat), displacement, or by atmospheric changes (the amplitude is too large)

The stability of the clock for ERS-1 was designed to be good over about one second, but by good fortune it was found to perform well in practice over a period of five minutes, very much better than specified. This allowed the radar to produce long segments of interferometric data without ‘interfering with itself’ too much. To allow this to be done again in future, a much higher clock stability should be planned at the design phase and preferably made a requirement.

7. Envisat-ASAR interferometric techniques and applications

In this chapter we will extend the interferometric techniques and applications discussed so far to the case when two acquisitions are made from different SAR modes, for example ScanSAR and Alternating Polarisation mode, and combined interferometrically. Such combinations have achieved progressively more importance since the first ScanSAR sensors (SIR-C, SRTM, RADARSAT) demonstrated the capabilities of ScanSAR interferometry.

7.1 Introduction

We will focus here mainly on the Envisat-ASAR sensor, which has three acquisition modes that can be combined interferometrically:

- The conventional full resolution SAR mode, named Image Mode (IM)
- An Alternating Polarisation Mode with multi-polarity capabilities (APM)
- A ScanSAR mode with Wide Swath coverage (WSM)

The geometry of these acquisition modes is sketched in Figure 7-1, and the main features have been summarised in Table 7-1.

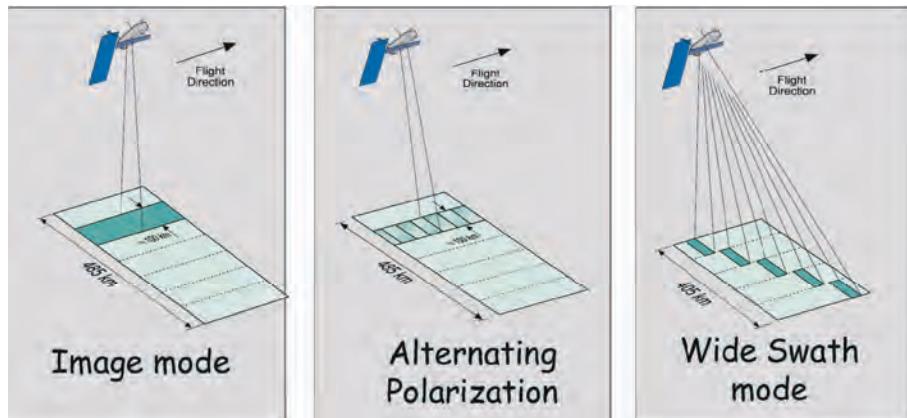


Figure 7-1: Interferometric modes of Envisat ASAR

Mode	Swath	Coverage (km)	Space resolution	Polarisations
IM	IS1-IS7 14°–45°	50–100	30 m	VV or HH
AP	IS1-IS7 14°–45°	50–100	30 m	VV+HH; VV+HV; HH+HV
WSM	SS1-SS5 16°–44°	400	150 m	VV or HH

Table 7-1: Envisat acquisition modes

Both AP and WSM are ScanSAR modes: the capability of performing interferometry in such modes will be discussed in the next section. We will generalise the interferogram generation technique, discussed in chapter 2 for the combination of two full resolution SAR images, to the combination of two images taken with different modes and resolution. Thereafter, we will detail the possible interferometric combinations of Envisat ASAR, and present some applications.

7.2 ScanSAR: an introduction

ScanSAR is Synthetic Aperture Radar with swath coverage, in slant range, that is wider than that of conventional SAR systems. This coverage is achieved by scanning different sub-swaths, i.e. by switching the antenna look angle into different positions [Moore81, Currie92].

The geometry of a five sub-swath ScanSAR (e.g. Envisat's ASAR WSM) is shown in Figure 7-2. The ScanSAR sensor acquires short 'bursts' of radar echoes for a time interval T_D (**dwell time**) by cyclically scanning all the N_s sub-swaths.

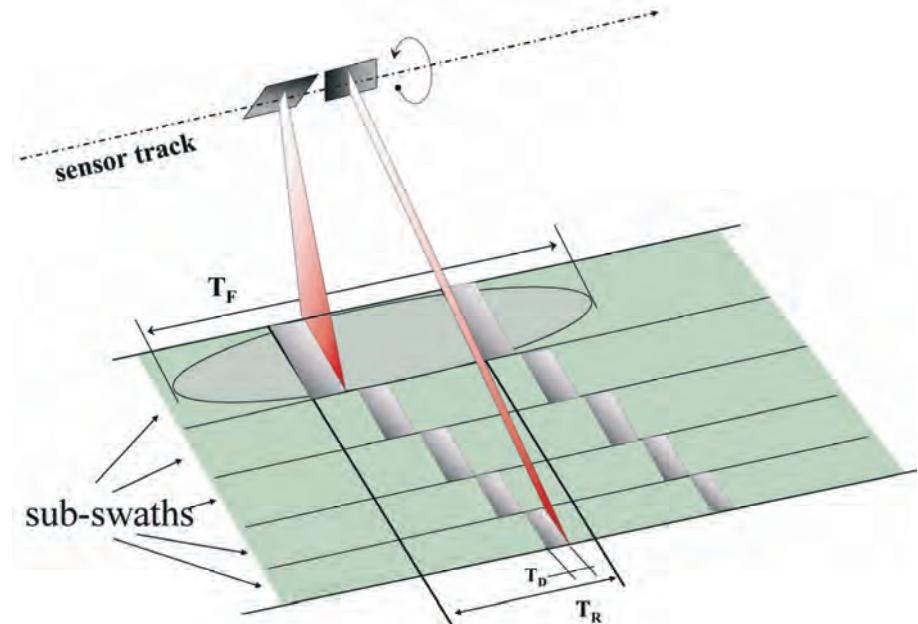


Figure 7-2: Geometry of the five beams (or sub-swaths) of ScanSAR

This specific acquisition scheme allows for the coverage of a ground scene that is N_s times wider than that of a conventional SAR mode (~400 km for Envisat and RADARSAT), at the same SNR balance (the transmitted power is always concentrated in one beam only), but with a proportional loss in the azimuth resolution [Currie92].

An undesired effect of the burst mode acquisition is a non-stationary Doppler history. This fact is evident in Figure 7-3: when an acquisition is being made, each target in the footprint is imaged from a different angle (three targets are represented here). Hence, both the amplitude and the phase

of an echo depend upon the displacement between the target and the burst. In particular, the target located at the centre of the antenna beam when the acquisition is on provides the strongest contribution and its spectrum will be centred on the Doppler centroid, f_{DC} .

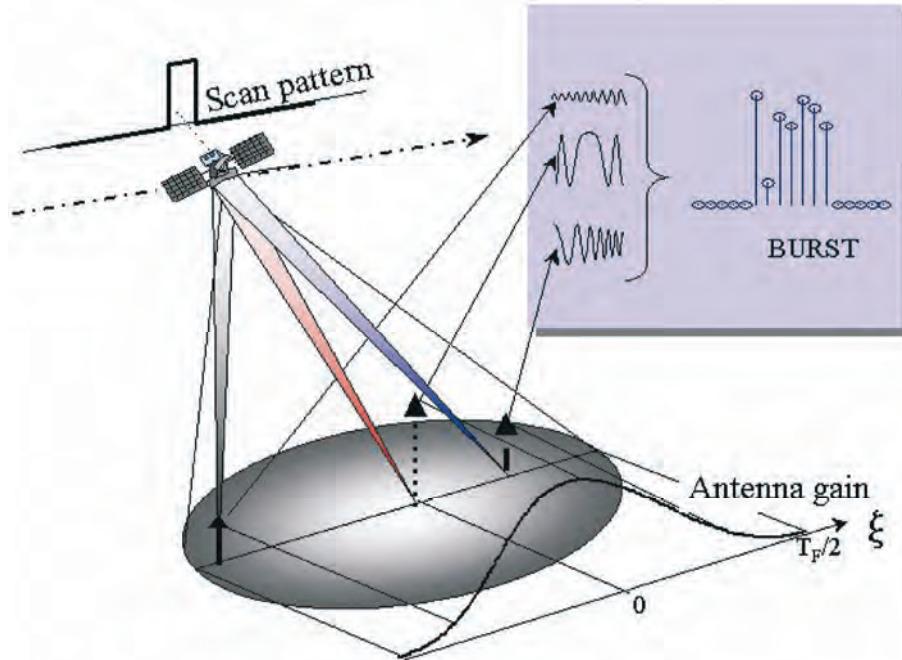


Figure 7-3: ScanSAR acquisition of three different targets. The Doppler phase history and the amplitude are differently weighted depending on the azimuth time of the target with respect to the scan pattern.

The resulting **azimuth-variant amplitude** is known as **scalloping**. Scalloping is the major drawback of ScanSAR, when fine amplitude calibration is required, as it introduces an azimuth variant **CNR (Clutter to Noise Ratio)**. The sole possibility for reducing scalloping is to have the same target imaged in several looks or bursts, N_b , from different viewing angles. In ASAR WSM, for example, three azimuth ‘looks’ have been acquired (i.e. $N_b = 3$) for each target. The **dwell time** is then limited to $T_D < T_F/(N_s N_b)$, $N_s N_b$ being the product of looks \times sub-swaths. This in turn worsens the geometric resolution, giving 150 m for the ASAR WSM. However, the radiometric accuracy is improved.

7.2.1 Acquisition

The impact of the burst acquisition of the focused ScanSAR data can be easily understood in the 1-D case (quite reasonable, as ScanSAR has a large resolution cell). We start from the SAR acquisition, in a simplified, straight geometry (see for example [Bamler92]):

$$hs(t, \xi) = a(v(t - \xi)/r_0 + \theta_s) \exp(j\pi f_R(t - \xi)^2) \quad \text{Equation 7.1}$$

where $hs(t)$ is the signal incident on a point scatterer
 ξ is azimuth time

r_o is the slant range

$a(\cdot)$ represents the antenna gain, as a function of the angle

$$\beta(t) = -v(t-\xi)/r_0,$$

f_R is the Doppler rate

θ_s the squint angle

The notation $hs(t, \xi)$ has been introduced to distinguish the signal domain, t , from the data domain, ξ . However, the SAR acquisition is stationary (azimuth invariant) as hs actually depends upon the difference $(t - \xi)$. But when we apply the burst window, $\text{rect}(t/T_D)$, the azimuth invariance is lost:

$$hs(t, \xi) = a(v(t-\xi)/r_0 + \theta_s) \exp(j\pi f_R(t-\xi)^2) \text{rect}(t/T_D) \\ -T_F/2 < t < T_F/2 \quad \text{Equation 7.2}$$

This is made clear in Figure 7-3, which shows in the upper-left corner the contributions of the three echoes that add up to the burst.

The effect of the burst on the spectrum of the acquired echo can be simply evaluated by transforming (Equation 7.2). We need to apply the **Method of Stationary Phase**^{iv} (MSP) (see, for example, [Curlander91]).

$$H_s(f, \xi) \cong a_\beta \left(\frac{f}{f_R} \right) \exp \left(-j\pi \frac{f^2}{f_R} \right) \exp(-j2\pi\xi f) \text{rect} \left(\frac{f + f_R \xi + f_{DC}}{f_R T_D} \right), \quad \text{Eq. 7.3}$$

$$|f + f_{DC}| < \frac{PRF}{2}$$

where $f_{DC} = -2v/c f_0 \sin(\theta_s)$ is the Doppler centroid.

Here again we have indicated the dependence of the spectrum of each echo on the actual location of the scatterer: $H_s(f, \xi)$. The burst mode acquisition accounts for the last term in Equation 7.3. The effect is twofold:

- the spectrum of each echo has a reduced bandwidth $B_s = f_R T_D$, hence also a coarse resolution;
- the spectrum is bandpass, and its central frequency depends on the target location:

$$f_c = f_{DC} + f_R \xi \quad \text{Equation 7.4}$$

7.2.2 Focusing

Focusing, in ScanSAR systems, means assigning to each target in the whole (large) footprint the proper contribution to the (short) burst. The simplest way to get phase-preserving ScanSAR focusing is to apply the same SAR reference, i.e., matched to (Equation 7.1), to the burst, after zero-padding it for the duration of the footprint. We assume here to compensate for the azimuth antenna pattern by the proper inverse reference^v, we get:

^{iv} We assume here that the azimuth time-bandwidth product (in the burst duration) is much greater than one. For very low resolution ScanSAR, different focusing schemes should be implemented, see reference [MontiGuarnieri2001].

^v In SAR processing, the antenna pattern compensation is not usually performed. In ScanSAR this is required if scalloping compensation is needed on the basis of a single burst, see also [Bamler95B].

$$h_r(t) = a^{-1} \left(vt/r_0 + \theta_s - t \right) \exp(j\pi f_R t^2) \cdot \text{rect}((t + t_{DC})/T_F) \quad \text{Eq. 7.5}$$

The reference transfer function can be derived, again by exploiting MSP:

$$H_r(f) \cong a_\beta^{-1} \left(\frac{f}{f_R} \right) \cdot \exp\left(j\pi \frac{f^2}{f_R}\right) \text{rect}\left(\frac{f + f_{DC}}{f_R T_F}\right) \quad \text{Eq. 7.6}$$

Eventually, we combine the ScanSAR acquisition, Eq. 7.3, and focusing, Equation 7.6, to get the overall ‘ScanSAR acquisition + focusing’ transfer function^{vi}:

$$H_{ss}(f; \xi) = H_r(f) H_s(f; \xi) = \\ \exp(-j2\pi\xi f) \text{rect}\left(\frac{f + f_R\xi + f_{DC}}{f_R T_D}\right) \text{rect}\left(\frac{f + f_{DC}}{f_R T_F}\right) \quad \text{Eq. 7.7}$$

The amplitude scalloping has disappeared, due to the use of the inverse filter. However, the spectrum of focused ScanSAR data is again narrow bandwidth, centred on the frequency $(f_{DC} + f_R\xi)$, that *sweeps the azimuth at the Doppler rate*, as appears in the spectrograms of Figure 7-4 on the right.

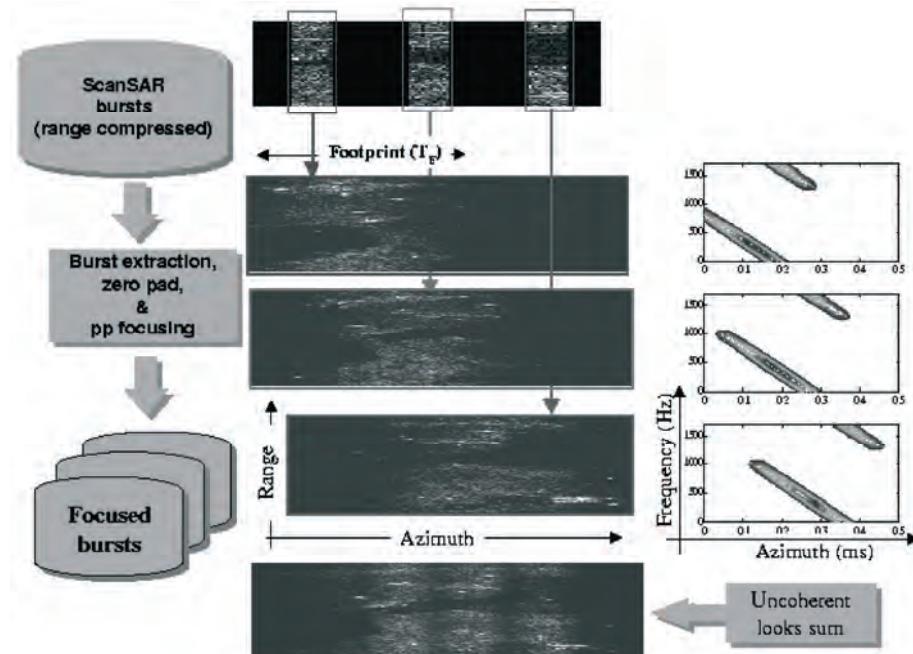


Figure 7-4: Azimuth focusing of three consecutive ScanSAR bursts. Left: schematic block diagram. Centre: amplitude images of the three focused bursts and the final multi-look average. Notice the scalloping due to the antenna pattern, which was intentionally not compensated during focusing. Right: The spectrograms of the three complex focused bursts.

It should be now clear that phase preserving ScanSAR focusing can be achieved by means of any standard SAR processor, just by processing burst-

^{vi} Note once again that Equation 7.4 is approximated for large azimuth time \times bandwidth product.

by-burst, and by (previously) zero-padding each burst to the duration of the whole footprint.

If the desired output is a multi-look image, it would be sufficient to detect and average all the focused bursts (in practice, only N_b bursts contribute at each azimuth), as shown in Figure 7-4.

The use of conventional SAR focusing to perform ScanSAR focusing is however not efficient at all, as the output image will be sampled at the SAR PRF rate, that is many times the Shannon limit (the inverse of the resolution). For example, in ASAR WSM, the ScanSAR resolution is 30 times coarser than the IM one.

To this end, several algorithms have been proposed in the literature for computing ScanSAR phase-preserving focused images directly at the low sampling rate: see [Bamler96, MontiGuarnieri96A, Moreira96, Cumming97, Lanari98]. These algorithms are different implementations of the same transfer function, with comparable computational efficiency.

7.3 ScanSAR interferometry

Let us suppose we have two (repeat-pass) ScanSAR acquisitions, by perfectly synchronised scans, parallel orbits and the same Doppler Centroid.

The generation of a multi-look ScanSAR-ScanSAR interferogram can be achieved by separately processing each burst [MontiGuarnieri96A]. First, phase-preserving focusing should be performed (see section C.6.2.). Next, all the steps described in chapter B2 for normal SAR interferometry should be applied, with the exception of the common band filtering (not required for synchronised acquisitions):

- 1) range oversampling
- 2) spectral shift filtering
- 3) slave co-registration
- 4) Hermitian multiplication

The interferograms achieved by processing different bursts should be coherently averaged: the mosaicking is similar to that shown for the amplitudes in Figure 7-6. A slightly different approach is found in [Bamler99].

In the actual implementation, one should ensure that the data were focused with the correct Doppler rate, and that the co-registering parameters are quite accurate (better than the ScanSAR resolution!). It can be shown that very small errors in these parameters will introduce phase artefacts and bias. A technique to provide the proper estimate of co-registering parameters for ScanSAR is described in [Moreira00].

7.3.1 Common band (CB) filtering

So far we have assumed perfectly time-aligned ScanSAR acquisition scan patterns, but in practical ScanSAR-ScanSAR interferometry we expect some mismatch.

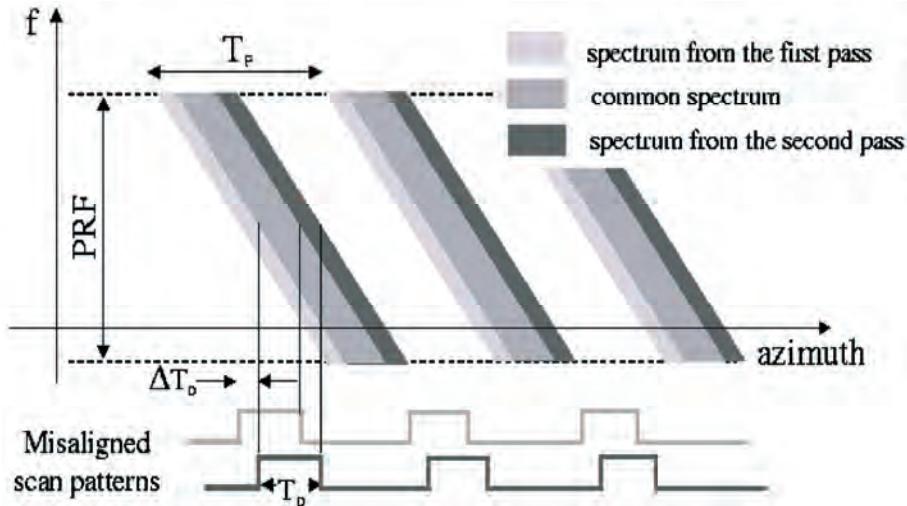


Figure 7-5: A shift of ΔT_D in the two ScanSAR acquisition patterns introduces a proportional shift in the acquired reflectivity azimuth spectra. The common band is reduced, causing a loss in coherence (avoidable by proper CB filtering) and an (unavoidable) loss of resolution.

Let us consider a time shift ΔT_D in the two scans. The azimuth linear sweep of the focused ScanSAR data spectrum of Equation 7.7 will convert that time shift into a frequency shift: $\Delta f_e = f_R \Delta T_D$. The consequence is twofold [MontiGuarnieri96A]:

- 1) The common band in azimuth is reduced to a value: $B_{az} = |f_R|(T_D - \Delta T_D)$, hence *the interferogram azimuth resolution will be worsened* by a ratio equal to the inverse of the bandwidth. Clearly, when $\Delta T_D > T_D$ there is no overlap at all in the two acquisitions, and no interferometric signal^{vii}.
- 2) A decorrelation is introduced, for non-aligned acquisitions, due to the non-common spectral contributions.

The fact is quite clear from Figure 7-5, which shows the azimuth varying spectral support of the two focused images. The same figure shows how this decorrelation can be avoided just by pre-filtering both images, keeping the common, azimuth-varying spectral contributions [MontiGuarnieri96A]. We can think of this filtering as an extension of the ‘common band filtering’ (see chapter B2) for ScanSAR interferometry. The same result can be achieved by superimposing time-aligned windows on the raw data bursts [Bamler96], however this would involve re-focusing the two datasets.

^{vii} We assume the source is a homogenous, indefinite target, so that its spectrum is uncorrelated at different frequencies. For different types of scatterer (such as point scatterers), correlation can be found even for a totally misaligned scan [MontiGuarnieri00].

7.4 Multi-mode SAR interferometry

The availability of repeat pass ScanSAR interferometry is subject to constraints on the synchronisation of the two acquisitions. For high resolution ScanSARs, these constraints are likely to be met: for example a 2-beams-1-burst ScanSAR (RADARSAT ‘ScanSAR Narrow A’ or Envisat AP modes) always has some overlap in the scans, even for random synchronisation. This is not the case for low-resolution ScanSARs: for the Envisat WSM, for example, an along-track synchronisation of the scan cycle better than 30 ms is required. The fact is made worse by second order effects, here neglected, like the azimuth spectral shift [MontiGuarnieri96A].

Another constraint for low resolution ScanSAR interferometry is the stringent baseline limit, due to the small system bandwidth. For the ASAR WSM mode the ‘critical’ baseline (for flat Earth) is $B_{nc} = 450$ m, at mid range, leading to ‘practical’ (or ‘useful’) baselines in the order of 100 m. The occurrence of such baselines is expected to be quite marginal (10% based on the interferometric acquisitions of the ERS mission [Solaas94]).

These limitations can be overcome by combining **a low resolution ScanSAR acquisition with a full-resolution SAR one** as both these modes are usually available on the latest sensors. This combination would be useless for mapping topography, due to its poor geometric resolution. However, it could be quite effective if a DEM is provided. In that case, an optimal combination is summarised in [MontiGuarnieri99A] to get a differential interferogram with the same quality of full resolution ones, and comparable baseline constraints. The geometric resolution would be the ScanSAR one.

7.4.1 Multi-mode interferometric combination

Performing ScanSAR – SAR interferometry requires a sort of spectral shift & common bandwidth filtering [MontiGuarnieri99A]. The concept is illustrated in Figure 7-6. We assume one very fine resolution focused image (‘master’ in the figure), and one very coarse resolution one, maybe ScanSAR (‘slave’ in the figure). We also assume knowledge of the topography, e.g. a DEM (like the SRTM ones). The steps to follow are explained below.

1. Synthesise the low-resolution complex reflectivity image

The slave image is synthesised by exploiting the master image (the fine resolution one) and the DEM. We do it as discussed in section 2.2, here extended to the 2-D case, with attention to the ScanSAR ‘sweeping’ spectrum. We end up with Equation Eq. 7.8, in which the exponential is the ‘synthetic fringe pattern’, and h_s is a filter with the same bandwidth as the slave acquisition.

$$\hat{v}_s(x, r) \cong v_M(x, r) \exp\left(j \frac{4\pi f_0}{c} \Delta R(x, r)\right) * h_s(x, r) \quad \text{Eq. 7.8}$$

This filter plays the fundamental role of *integrating all the contributions in the large resolution cell*, properly compensated for the travel path distance (see Figure 7-6).

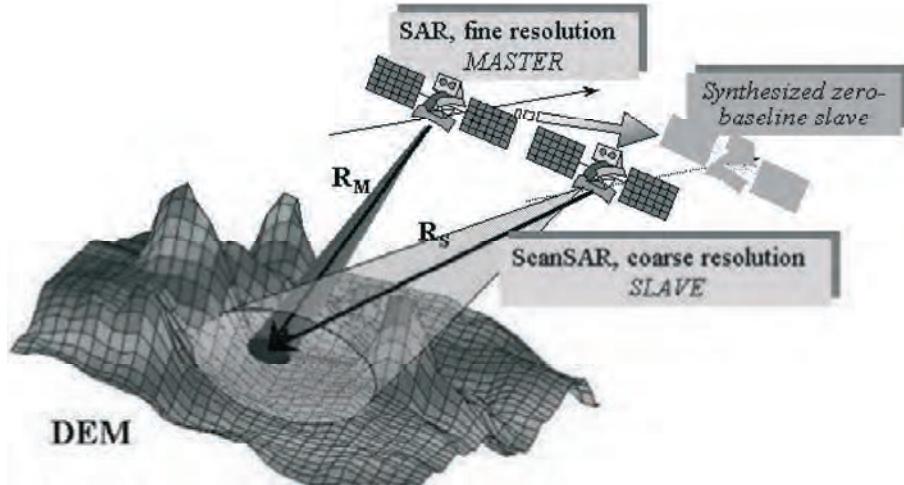


Figure 7-6: The low resolution (ScanSAR) slave image can be synthesised, given the fine resolution (SAR) one plus a DEM

2. Interferometrically combine the synthesised slave with the actual slave

This gives a (theoretical) **zero baseline interferogram**, e.g. an interferogram in which most of the topographic contributions have been removed [MontiGuarnieri99A].

This technique provides the proper removal of volume decorrelation, extending the interferometric capability *to the same baseline range of the conventional full resolution case*. Notice that the approach is consistent with the optimal slope estimate derived in section 2.3.

The generalisation of such techniques to the combination of any mode (SAR, ScanSAR, SPOT etc.) and resolution is provided in reference [Fornaro01]. The general case, in which the geometric resolutions of the two images *may* be comparable, requires pre-filtering to be applied to both images. The design of the pair of *bidimensional, space-variant filters* is approached by a statistical point of view in [Fornaro01], where a search for the maximisation of coherence (hence, an L_2) is made under the assumption of a stationary, indefinite homogenous target.

The solution is found in a three-step procedure, which is a generalisation of the one in Figure 2-3.

It consists of the following steps:

- 1) An estimate of the slave reflectivity \hat{v}_S is derived, given the master reflectivity, v_M and the synthetic fringes;
- 2) An estimate of the master reflectivity \hat{v}_M is derived, given the slave reflectivity, v_S , and the synthetic fringes;

- 3) The two estimates are combined to get the ‘cleaned’ interferogram:
 $i = \hat{v}_M \times \text{conj}(\hat{v}_S)$

These steps are summarised in Figure 7-7.

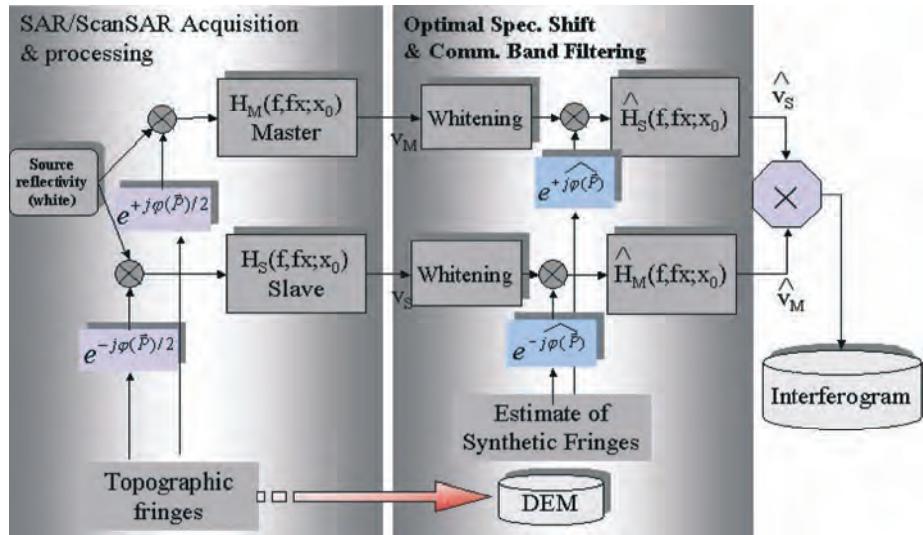


Figure 7-7: Generalised scheme for the spectral shift & common band filtering, allowing for any interferometric combination of SAR and/or ScanSAR modes

Note that these steps correspond to the MMSE estimators already discussed in section 2.2, here extended to the general case in which slopes vary locally. The proper design of the two filters is detailed in [Fornaro01]. A good approximation of these filters is provided in the scheme of Figure 7-7.

In the figure, a white reflectivity is assumed as the source (left part of the figure). Only two demodulated and band-pass filtered versions of this reflectivity are known: they represent the focused master and slave images (v_M and v_S in the figure). The 2D transfer functions H_M and H_S account for the overall acquisition and focusing. As an example, for ScanSAR, we should use:

$$H(f, f_x) = H_{ss}(f_x; \xi) \text{ rect}(f/B_R) \quad \text{Equation 7.9}$$

where $H_{ss}(f_x; \xi)$ is defined in Equation 7.7.

The generalised spectral shift and common band filtering, in the right-hand shaded block in Figure 7-7, is implemented in three steps for both the master and slave image:

- 1) Whitening of the 2D spectrum: this is mainly a compensation for the azimuth antenna pattern as discussed in section 2.1.1;
- 2) Multiplying by the synthetic fringe, $\exp(j\varphi(P))$, in order to get spectral alignment between the two images;
- 3) Filtering of the master image in the slave bandwidth and vice versa.

A further modulation by $\exp(j\varphi(P))$, (not shown in the scheme of Figure 7-7), should be applied to the master image if a differential interferogram is to be computed.

Finally, notice that interferometry is possible if there is some spectral overlap. If we translate the time modulation and filtering into convolution and multiplication in the frequency domain, the following should hold:

$$\begin{aligned} H_M(f, f_{az}) * F[\exp(j\varphi(P))] \times H_S(f, f_{az}) &\neq 0 \text{ and} \\ H_S(f, f_{az}) * F[\exp(j\varphi(P))] \times H_M(f, f_{az}) &\neq 0 \end{aligned} \quad \text{Eq. 7.10}$$

where convolution times the Fourier transform of the synthetic fringes, $F[\exp(j\varphi(P))]$, can be interpreted, for a flat terrain, as a spectral shift. In so doing, we derive the proper constraint for ScanSAR synchronisation discussed in the previous section.

7.5 Applications

The multi-mode, multi-resolution combination just derived is useful for multi-mode missions such as Envisat ASAR. With reference to the ASAR acquisition modes in Figure 7-1 (IM, AP & WSM), and the constraint of the previous equation, the following interferometric combinations are possible:

- a. Conventional full resolution interferometry: IM/IM
- b. ScanSAR/ScanSAR and ScanSAR/SAR interferometry: WSM/WSM and IM/WSM
- c. Alternating Polarisation interferometry: AP/AP (ScanSAR/ScanSAR),
- d. Alternating Polarisation/full resolution interferometry: AP/IM (ScanSAR/SAR)

7.5.1 AP/AP and AP/IM interferometry

The Alternating Polarisation (AP) mode of ASAR (see Figure 7-1) is capable of providing multi-polarimetric images by means of ScanSAR acquisitions, in which switching is made on polarisations instead of sub-swaths.

The purpose of AP interferometry is to estimate the **Polarimetric Phase Difference (PPD)** and the polarimetric correlation [Pasquali98, Pasquali98a]. This information is of great help in the processes of identification and classification of different scattering mechanisms, and where the penetration depth is different at different polarisations. As an example, the PPD is 0° for an odd number of reflections (point target) and 180° for an even number of reflections (dihedrals).

The ASAR sensor is able to simultaneously acquire two different AP (ScanSAR) images on the same swath: either HH/VV, or HH/HV, or VV/VH. These two images have different spectral contributions and cannot be combined with each other, but they can be combined with another AP mode or IM mode acquisition (either HH or VV) of the same area.

The IM data is split into two parts, each coherent with either the HH or the VV AP polarisation, through an azimuth space-varying filter [Pasquali98, Pasquali98a]. An interferogram is then formed for each of the two coherent AP-IM combinations. A differential interferogram is finally generated after an adaptive filtering process that improves the SNR of the two AP-IM

interferograms. Note that the temporal baseline would be a multiple of the orbit repeat time (35 days). Hence, long-term stability is required.

An example of the results that can be expected is shown in Figure 7-8, obtained by exploiting SIR-C multi-polarimetric data.

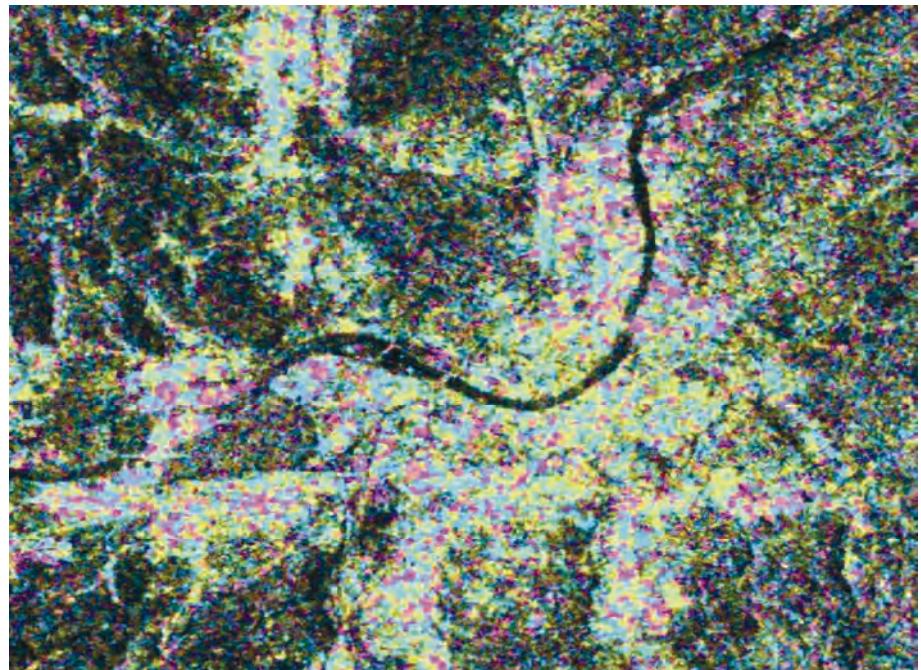


Figure 7-8: PPD estimation, Envisat-ASAR data, Basel, obtained by combining one AP image (polarisation VV, HH acquisition date 16/9/02) and one IM image (polarisation HH, acquisition date 21/10/02), baseline = 60 m. The image represents amplitude coloured with the PPD, so different colours are due to different polarimetric behaviours.

7.5.2 WSM/WSM and WSM/IM interferometry

The capabilities provided by the large swath imaged in one Envisat ScanSAR WSM acquisition open up the possibility of novel and interesting applications [MontiGuarnieri00]. This combination would be useless for topography mapping, due both to the poor geometric resolution and the long revisit time (≥ 35 days). However, it could be quite suited for monitoring, particular over large areas.

An example of the results that can be achieved by interferometrically combining two complex WSM images is provided by the co-seismic fringes of the Bam Earthquake in Figure 7-9, to be compared with a full resolution interferogram, obtained by IM surveys, in Figure 7-10. The WSM has much wider coverage, therefore it allows the detection and measurement of large-scale, low frequency fringes due to the earthquake that would not be possible in full-resolution mode.

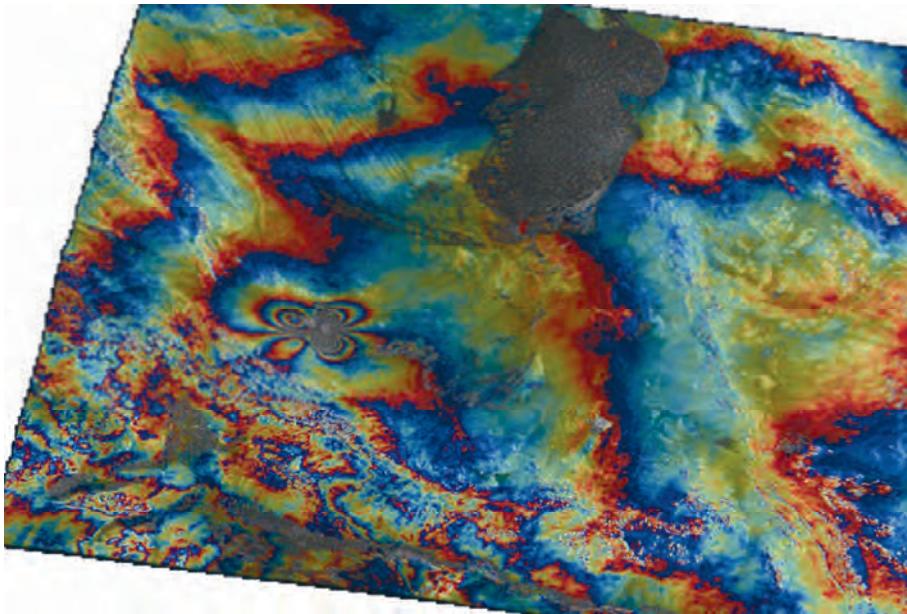


Figure 7-9: Large scale (200×400 km), geocoded interferogram (absolute value coloured with phases), of the area of Bam (Iran), achieved by combining two Envisat WS images acquired on 2 September 2003 and 8 June 2004, track 306. The normal baseline was 100 m and the scan patterns were synchronised up to 84%, leading to a resolution of approximately 40 m (range) \times 150 m (azimuth). The topographic contributions were removed by exploiting an SRTM DEM. The fringe contributions are mainly due to the earthquake of 26 December 2003 (some atmospheric artefacts are still visible in the lower left part of the image).

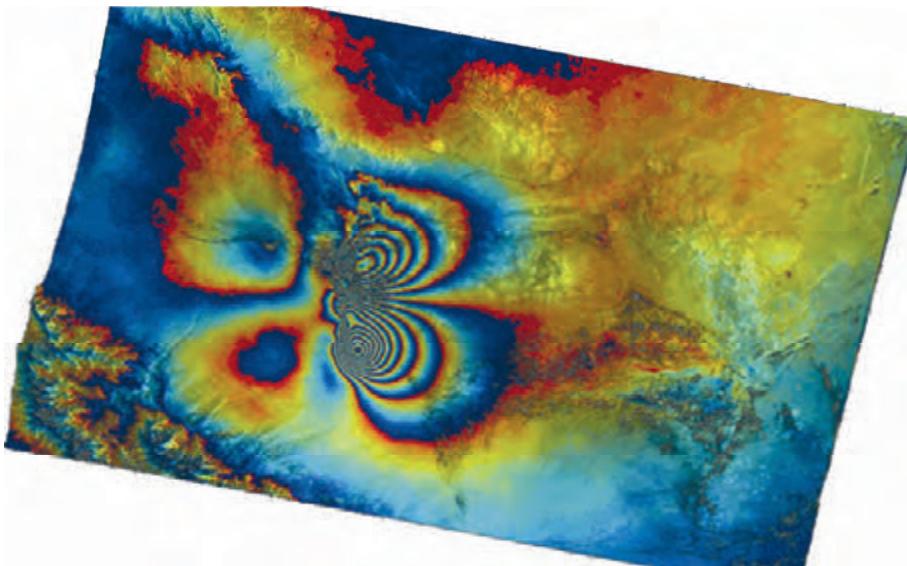


Figure 7-10: Interferogram of BAM area, as in Figure 7-9, but achieved by combining two IM images: 3 December 2003 and 11 February 2004, with a baseline close to zero. The interferogram resolution is pretty fine, approximately 20×10 m, but the ground coverage is much smaller than the previous image. The different fringe pattern is due to the difference in the look angle.

Furthermore, the large WSM swath allows for frequent revisit-time monitoring. Such a system could be based on the mixed use of ScanSAR and full resolution images for WSM/IM interferometry. An example of a WSM/STRIPMAP interferogram is shown in Figure 7-11. The advantages of WSM/STRIPMAP versus WSM/WSM are twofold: firstly there is no need for scan pattern synchronisation; secondly there is much less volumetric decorrelation, provided that the optimal combination here discussed is used. The conclusion is that we have a higher probability of getting valuable information in this case than in WSM/WSM interferometry. However, the interferogram coverage would be the same as the STRIPMAP, and its resolution the same as the WSM.

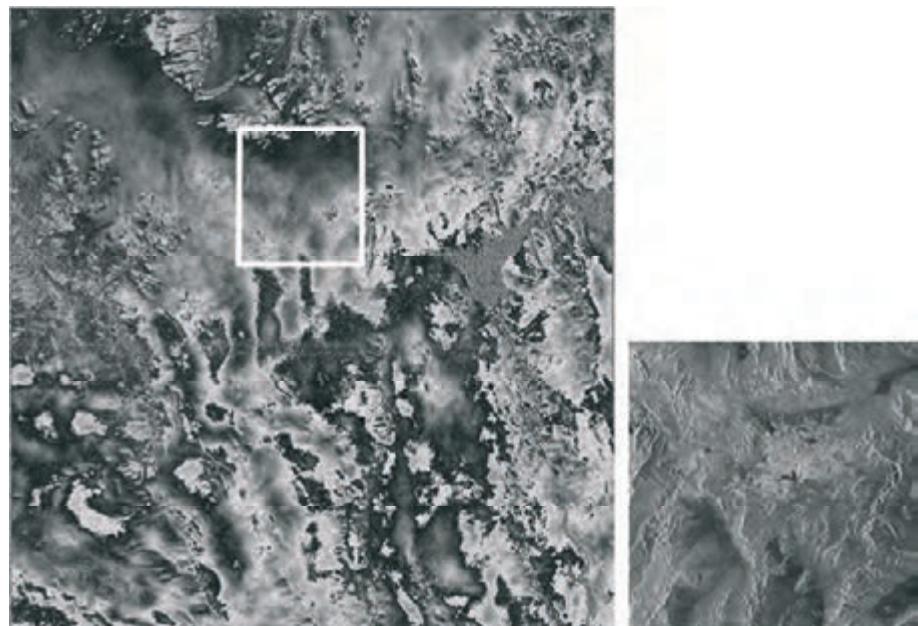


Figure 7-11: Combined ScanSAR/STRIPMAP interferogram of the area around Las Vegas. Left: fringes, right: detected amplitude, closeup of Las Vegas downtown. Topography has mostly been removed by exploiting an SRTM DEM: residual fringes are thus mainly due to atmospheric effects.

We could likewise take advantage of the wide ScanSAR swath for frequent monitoring, in two different ways. A first system could provide alerts over hazardous areas. The use of WSM would reduce the revisit time to a few days due to the very large ScanSAR swath. Such a system would require that, for any new WSM acquisition, an IM with a suitable baseline has already been acquired over the same area. The number of IM images that must be previously acquired over the area of interest has been plotted in Figure 7-12 as a function of the revisit time and the orbit span.

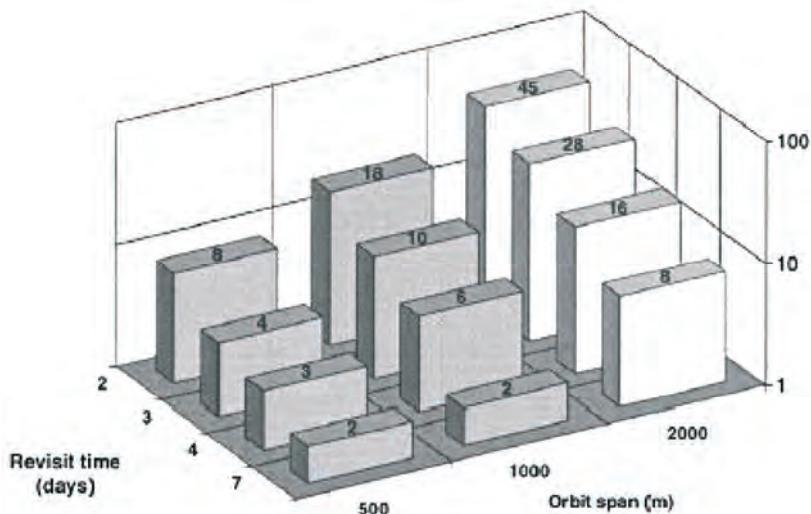


Figure 7-12: Number of images to be acquired to build the minimum database, computed here by assuming random orbit sampling and latitudes of 45°. The number must be doubled to include ascending and descending orbits.

A second system could be provided by reversing the order of acquisitions. A first set of WSM images (the actual number is the same as in Figure 7-13) could be acquired over the area of interest in a short time. Thereafter, every new full resolution IM image could be combined with an WSM image in the database to get a (differential) interferogram. Such a system would be suitable to monitor large-scale changes like surface deformation, meteorological effect, Earth tides, according to user demand. The scheme of such a system is illustrated in Figure 7-13.

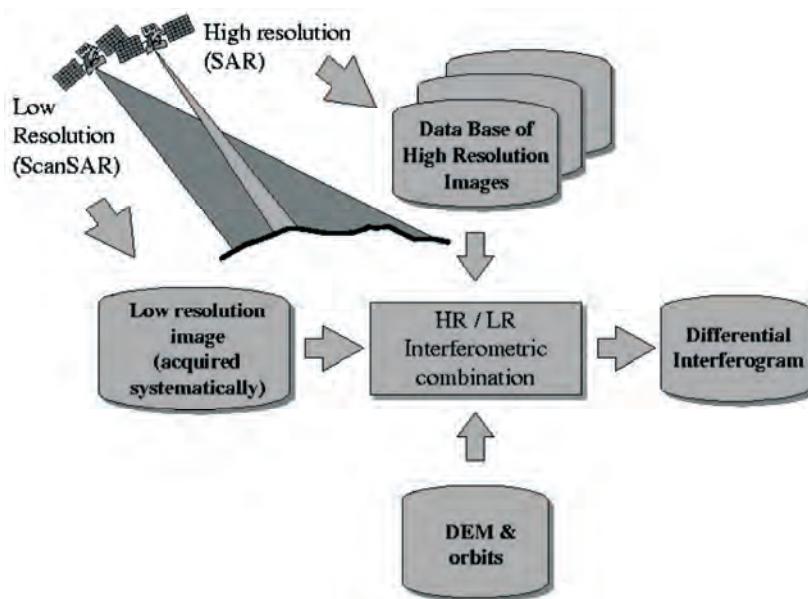


Figure 7-13: Monitoring system that exploits low and high-resolution acquisitions to get short revisiting time intervals. Low and high resolution images correspond to WSM and IM acquisitions respectively.

8. ERS-Envisat interferometry

8.1 Introduction

Let us assume a combination of two monostatic SAR systems with slightly different central frequencies f_0 and almost the same bandwidth W , such as ERS-2 ($f_0 = 5.3$ GHz, $W \approx 16$ MHz) and Envisat ($f_0 = 5.331$ GHz, $W \approx 16$ MHz). The two sensors could revisit the same areas, say after half an hour. This system would have very low temporal decorrelation, even better than the Tandem ERS mission, making the combination suitable for:

- High accuracy mapping on rolling topographies
- Upgrading accuracy of the existing Tandem ERS1/2 DEM to a few cm, and removing atmospheric artefacts

This frequency gap could be compensated by tuning the baseline, following the tuneable SAR idea [Prati94B, Gatelli94]. Hence, both super-resolution images and interferometry could be performed, as described in chapter C5. The trade-off between super-resolution and interferometry would be controlled by the baseline.

8.2 ERS-Envisat interferometric combination

Figure 8-1 shows the ERS-Envisat interferometric geometry. The ERS-2 sensor operates at a slightly lower frequency than Envisat.

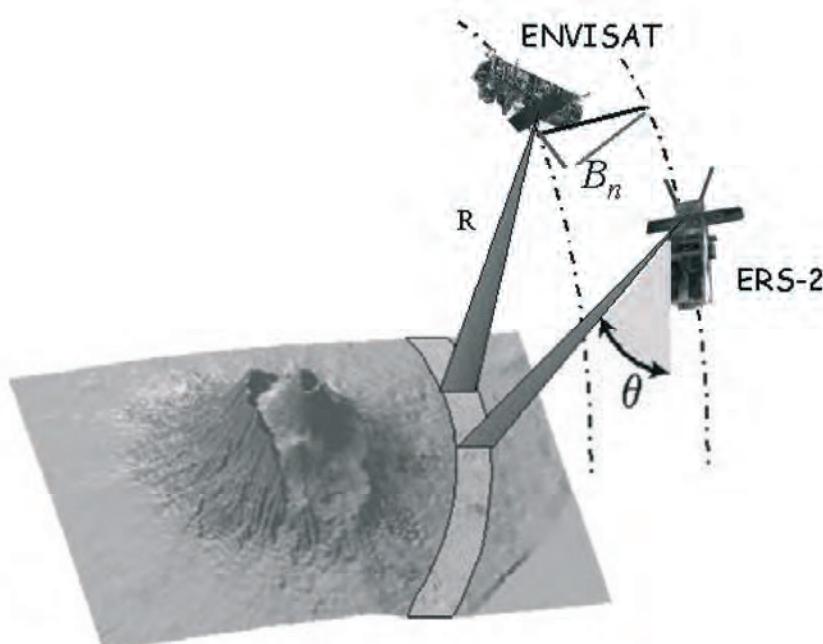


Figure 8-1: Geometry of the combined ERS-Envisat InSAR system. R is the slant range axis, B_n the normal baseline, θ the incidence angle and α the terrain slope in range.

8.3 Frequency gap compensation

The frequency gap can be compensated by exploiting the baseline. For interferometry, we require that the same wavenumber on the ground be illuminated by both sensors. According to Figure 8-2, Envisat can compensate for its intrinsic higher carrier frequency, $f_o = 5.33$, by exploiting a lower incidence angle.

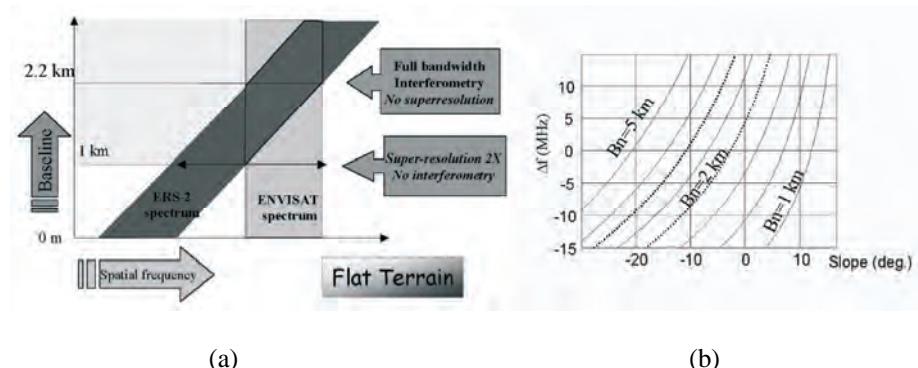


Figure 8-2: (a) Spectral shift in MHz as a function of the range slope expressed for baselines from 1 km to 5 km. (b) Overlap of ERS and Envisat wave-number spectrum at different baseline for flat terrain. Super-resolution $\times 2$ is achieved when the two spectra are marginally overlapped, whereas full resolution interferometry requires a complete overlap.

The baseline required to provide the proper wave-number shift on a ground terrain slanting at an angle α is:

$$B_\alpha = \frac{R\Delta f \tan(\theta - \alpha)}{f_o} \quad \text{Equation 8.1}$$

The baseline B_{α_0} would generate a total spectral overlap of the two images and thus a constant phase interferogram with coherence $\gamma = 1$ in the absence of temporal variations and volume effects. Terrain slopes at range $\alpha \neq \alpha_0$ would generate a relative spectral shift:

$$\Phi(\alpha, \alpha_0) = \Delta f \left[\frac{\tan(\theta - \alpha_0)}{\tan(\theta - \alpha)} - 1 \right] \quad \text{Equation 8.2}$$

As shown in Figure 8-2(a), if a 2000 metre baseline is adopted, a 5° slope in range would be perfectly compensated (no spectral shift) and slopes within the range $\pm 5^\circ$ would have a spectral shift smaller than 10 MHz (about $2/3$ of the system bandwidth W). Thus, if we assume rolling topography sites with gently varying slopes (i.e. almost constant within at least 3 consecutive pixels: about 75 m), and limited between $\pm 5^\circ$, an almost unitary coherence could be obtained on the entire image by means of the so called ‘common band filtering’ technique.

8.4 Vertical accuracy

The main advantage of such large-baseline interferometric combination lies in its intrinsic vertical accuracy. This accuracy can be computed by linking the standard deviation of the phase in Equation 1.18 with the phase standard deviation as a function of the coherence, γ , in Equation 1.16. We get an approximation for the standard deviation of the height:

$$\sigma_h = \sqrt{\frac{1-\gamma^2}{2\gamma N}} \frac{R\lambda \sin \theta}{4\pi B} \quad \text{Equation 8.3}$$

The coherence is expected to be good, due to the 1-day (or less) repeat pass interval. As an example, if we assume an overall coherence of $\gamma = 0.5$, and we average over $N = 5$ independent samples (i.e. a ground patch of $\sim 200 \text{ m}^2$ on flat terrain), we get $\sigma\phi \cong 30 \text{ cm}$. The same accuracy would be achieved with a conventional Tandem interferogram (assuming $B_n \cong 200 \text{ m}$) by exploiting a ground patch that is 100 times bigger, e.g. 20 km^2 . However, in that case we would lose spatial accuracy.

8.5 Altitude of ambiguity

Once the terrain slope α is compensated by means of the baseline B_α , the altitude of ambiguity h_a can be computed similarly to the section 2:

$$h_a = \frac{c R \sin(\theta - \alpha)}{2 f_o B_\alpha} \quad \text{Equation 8.4}$$

Then, by substituting the expression of the compensation baseline B_α , given by 8.1, we get the following simple expression:

$$h_a = \frac{c \cos(\theta - \alpha)}{2 \Delta f} \quad \text{Equation 8.5}$$

In the case of ERS-Envisat interferometry the following expression of the altitude of ambiguity (in metres) holds:

$$h_a = \frac{3 \cdot 10^8 \cos(\theta - \alpha)}{2 \cdot 31 \cdot 10^6} \cong 5 \cos(\theta - \alpha) \quad \text{Equation 8.6}$$

As an example, if we consider a terrain slope of 2.5% (about 1.5°), the altitude of ambiguity would be about 4.6 metres. It should be noted that with such a small altitude of ambiguity compared with the pixel size ($25 \times 5 \text{ m}$ range/azimuth) the resulting interferogram (even if strongly coherent) would be difficult to unwrap, especially in the azimuth direction where the terrain slope is not compensated. Nonetheless, the main application of this system would be the improvement of the accuracy of the low spatial frequencies of an existing DEM (e.g. generated from ERS Tandem images). Thus, the existing DEM could be used to demodulate the ERS-Envisat interferogram and the residue could be low-pass filtered, thus improving the coherence, which might be low due to temporal changes and volume scattering.

8.6 Effect of volume scattering

Due to terrain roughness around the slope α_0 compensated with the baseline B_{α_0} , the interferogram coherence decreases. We can refer to the volume scattering decorrelation computed in Equation 1.28:

$$\gamma_z \cong 1 - 2\pi^2 \left(\frac{\sigma_z}{h_a} \right)^2 \quad \text{Equation 8.7}$$

In the case of interest for us, i.e. $h_a = 4.6$ m, either $\sigma_z < 0.6$ m, or $\Delta z \ll 2$ m. These conditions are rather stringent, but are likely to be satisfied in flat areas or in rolling topography.

8.7 Experimental results

Notwithstanding the problems due to the unavailability of ERS-2 gyros, experimental evidence for the possibility of generating a SAR-ERS/ASAR-Envisat interferogram can be provided. The main goal of this experiment was to quantify the actual vertical accuracy of an ERS-Envisat interferometric DEM of a rural area along the rivers Réveillon and Yerres (tributaries of Seine) in the immediate neighbourhood of Paris. Second was to show that with a 30' temporal baseline and 1.5 km geometric baseline, neat fringes can be obtained on vegetated areas, whereas a random location phase screen is obtained on urban areas. Following the spectral shift theory, it should be pointed out that with a normal baseline of about 1.5 km the 30 MHz frequency shift can be practically compensated on cross-track terrain slopes ranging from -1 to 8 degrees. Moreover, in our data the DC difference was found to be about 800 Hz, thus leaving a common band in azimuth slightly greater than 20% of the ERS Pulse Repetition Frequency (see Figure 8-3).

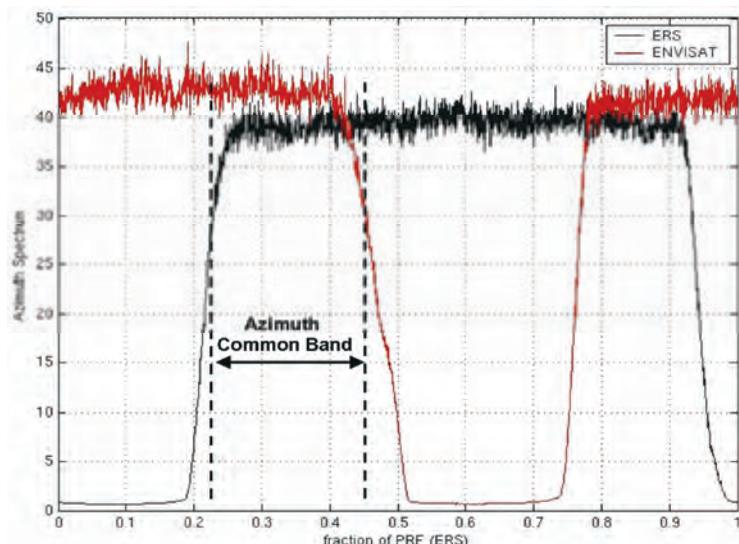


Figure 8-3: SAR-ERS and ASAR-Envisat azimuth spectra. The resulting azimuth common band is about 20% of the ERS pulse repetition frequency.

The unfiltered interferogram that was obtained after accurate image registration, and its azimuth common band filtered version, are shown in Figure 8-4.

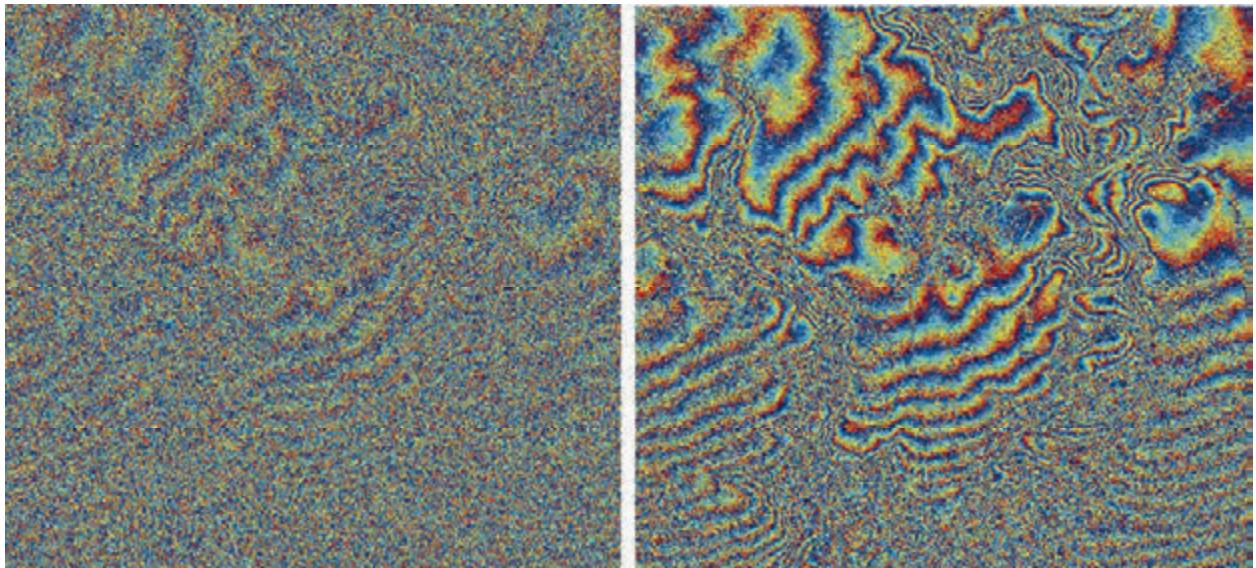


Figure 8-4: Left: unfiltered interferogram obtained after accurate ERS-Envisat image registration. Right: The interferometric fringes after the azimuth common band filtering.

The slant range common band filtered interferogram is shown in Figure 8-5.

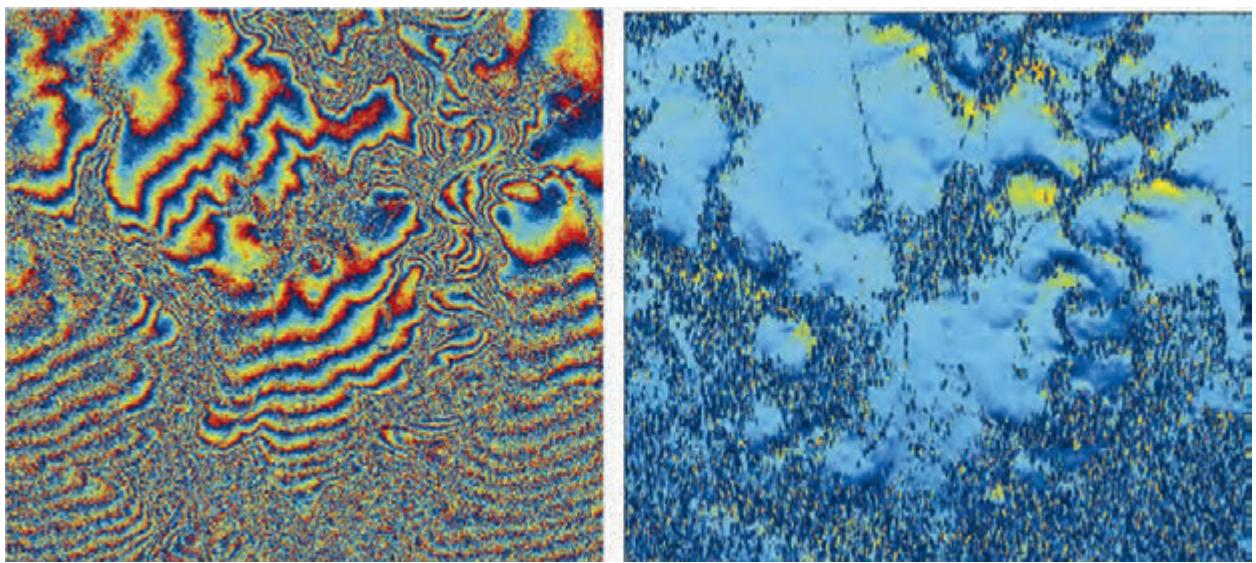


Figure 8-5: Left: The ERS-Envisat interferometric fringes after slant-range common band filtering. Right: Local ground range slope estimated from the local spectral shift that maximises the local coherence.

Due to the 1.5 km baseline, the altitude of ambiguity is about 6 m. The map of the ground range slope shown in Figure 8-5 has been derived from the value of *frequency* (changed by means of the common band filtering described above) that maximises the local complex coherence (an approximation of the optimal estimator in section 2.3).

Figure 8-6 shows a detail of the ERS-Envisat interferometric fringes of the rural area including the rivers Réveillon and Yerres. On this area the coherence and the relative elevation standard deviation histogram have been computed.

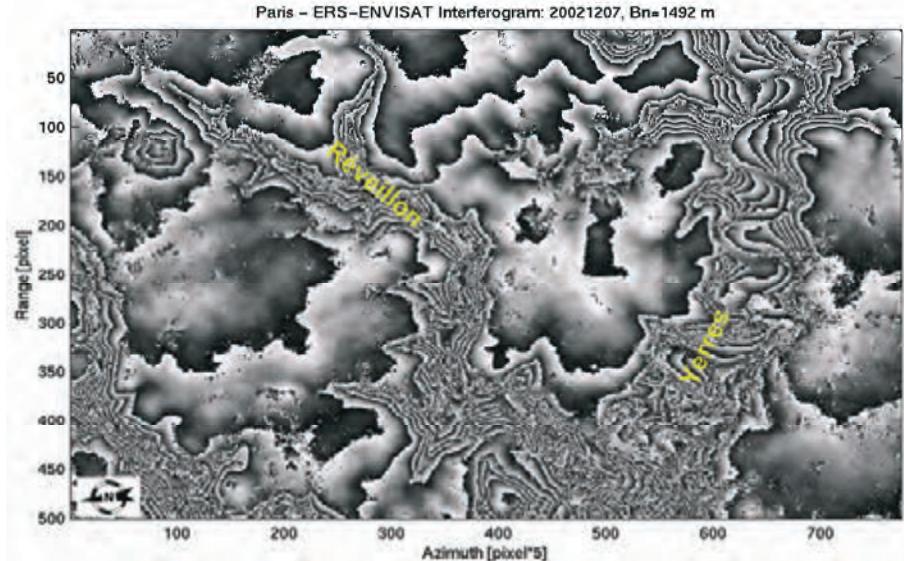


Figure 8-6: Detail of the ERS-Envisat interferometric fringes of the rural area including the rivers Réveillon and Yerres

The two histograms are shown in Figure 8-7. Here it is possible to appreciate the very high vertical accuracy that can be obtained by combining ERS-2 and Envisat SAR images.

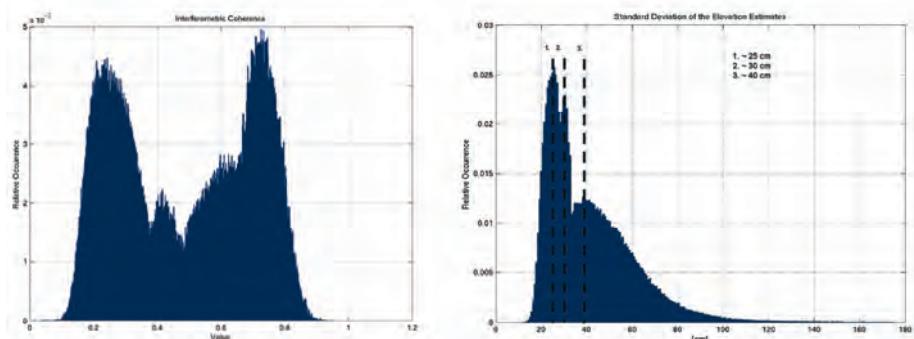


Figure 8-7: Left: Coherence histogram of the area shown in Figure 8-6. Right: Histogram of the elevation accuracy derived from the coherence map shown on the left.

Finally, the ERS-Envisat interferogram of Paris is shown in Figure 8-8. Here it is interesting to note that neat fringes are clearly visible on the city parks (the 30' temporal baseline avoids temporal decorrelation), whereas on the urbanised areas an almost random phase field (the location phase screen of point scatterers) is present. This is exactly the opposite of what we are familiar with from ERS/ERS interferograms (small geometric and large

temporal baselines), and it highlights the kind of applications the technique would lead to.

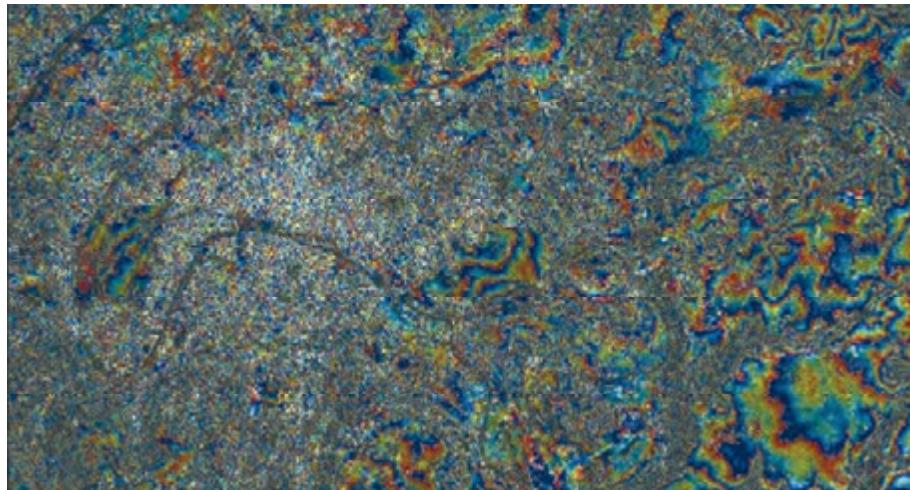


Figure 8-8: The left part of the image is centred on Paris. Here neat fringes are clearly visible on the city parks, whereas on the urbanised areas an almost random phase field is present.

Reference Documents

References

- [1] [Ahuja93] Ahuja, R. K., Magnanti, T. L., Orlin, J. B., Network Flows: Theory, Algorithms, and Applications, Prentice-Hall, Englewood Cliffs, New Jersey, USA -1993.
- [2] [Arora97] S. Arora, L. Babai, J. Stern and Z. Sweedyk, “The Hardness of Approximate Optima in Lattices, Codes, and Systems of Linear Equations,” *J. Comput. System Sci.* 54, pp. 317-331, 1997.
- [3] [Askne97] Askne, J.I.H.; Dammert, P.B.G.; Ulander, L.M.H.; Smith, G C-band repeat-pass interferometric SAR observations of the forest Geoscience and Remote Sensing, *IEEE Transactions Volume: 35*, Jan. 1997, Page(s): 25 -35
- [4] [Bamler92] Bamler, R. A comparison of range-Doppler and wave-number domain SAR focusing algorithms, *IEEE Trans. GARS*, 30(4):706-713, July 1992.
- [5] [Bamler95A] Bamler, R. and Schaettler B.. Phase-preservation in SAR processing: Definition, requirements and tests. *DLR Tech, Note Ver 1.0*, May 1995.
- [6] [Bamler95B] Richard Bamler. Optimum look weighting for burstmode and ScanSAR processing. *IEEE Transactions on Geoscience and Remote Sensing*, 33(3):722-725, May 1995.
- [7] [Bamler96] Bamler, R. and Einder. ScanSAR processing using standard high precision SAR algorithms, *IEEE Trans. GARS*, 34(1):212-218, 1996
- [8] [Bamler98A] Bamler, R., Hartl, P., Synthetic Aperture Radar Interferometry, *Inverse Problems*, Vol. 14, 1998, R1 – R54.
- [9] [Bamler98B] Bamler, R.; Adam, N.; Davidson, G.W.; Just, D. Noise-induced slope distortion in 2-D phase unwrapping by linear estimators with application to SAR interferometry, *IEEE Trans. GARS*, Vol. 36, N. 3, May 98, pp. 913 – 921
- [10] [Bamler99] Bamler, R., D. Geudtner, B. Schättler, P. Vachon, U. Steinbrecher, J. Holzner, J. Mittermayer, H. Breit, and A. Moreira, Radarsat ScanSAR interferometry, in Proc. IGARSS'99, Hamburg, Germany, June 1999.
- [11] [Berry2000] Berry, P. A. M., Pinnock, R. A., Johnson, C. P. D., Hilton, R. D., 2000c, ACE: A New Global Digital Elevation Model Incorporating Satellite Altimeter Derived Heights
- [12] [Carter87] G Clifford Carter. Coherence and time delay estimation. *Proceedings of the IEEE*, 75(2):236-255, February 1987.
- [13] [Cattabeni94] Cattabeni, M., Monti-Guarnieri A., and Rocca F. Estimation and improvement of coherence in SAR interferograms. In IGARSS94, Pasadena, CA, USA, 8-12 August 1994, pages 720-722, 1994.
- [14] [Chen00] Curtis W. Chen and Howard A. Zebker, “Network approaches to two-dimensional phase unwrapping: intractability and two new algorithms,” *Journal of the Optical Society of America A*, volume 17, pp. 401-414, March 2000.
- [15] [Closa98] J.Closa “The Influence of Orbit Precision in the Quality of ERS SAR Interferometric Data” Document No: ES-TN-APP-APM-JC01 – June 1st, 1998 http://earth.esa.int/rootcollection/sysutil/ORB_QL.html

- [16] [Costantini98] Costantini, M. A novel phase unwrapping method based on network programming, 1998, IEEE Trans. GARS, 36 (3), 813 – 821.
- [17] [Costantini99] Costantini, M.; Farina, A.; Zirilli, F., A fast phase unwrapping algorithm for SAR interferometry, IEEE Trans. GARS, Vol. 37, N. 1, Jan. 99, pp. 452 – 460
- [18] [Cumming97] Cumming I. G. A comparison of phase preserving algorithms for burst-mode SAR data processing. In IGARSS97, Singapore, 3–8 Aug 1997, volume 2, pages 731–733, 1997.
- [19] [Curlander82] J.C. Curlander “Location of Spaceborne SAR Imagery,” IEEE Trans. Geosci. and Remote Sensing, GE-20, pp. 359-364, 1982.
- [20] [Curlander91] Curlander J C and R N McDonough. Synthetic aperture radar: systems and signal processing. John Wiley & Sons, Inc, 1991.
- [21] [Currie92] Currie A and M A Brown. Wide-swath SAR. IEE Proceedings F, 139(2):123-135, 1992.
- [22] [Dammert99] Dammert, P.B.G.; Askne, J.I.H.; Kuhlmann, S. Unsupervised segmentation of multitemporal interferometric SAR images, IEEE Trans. GARS, Volume: 37 5 1 , Sept. 1999 , Page(s): 2259 -2271
- [23] [Davidson99] Davidson, G. W., and Bamler, R., “Multiresolution Phase Unwrapping for SAR Interferometry “, IEEE Trans. Geosci. Remote Sensing, 37, 1999, 163-174.
- [24] [DEOS] Delft Institute for Earth-Oriented Space Research (DEOS) - Orbit Determination. <http://www.deos.tudelft.nl/ers/precobs/>
- [25] [DESCW] <http://earth.esa.int/services/descw>
- [26] [Dixon94] Dixon, T., Editor, “SAR Interferometry and Surface Change Detection”, Report of the Workshop in Boulder, February 3 and 4, 1994, <http://southport.jpl.nasa.gov/scienceapps/dixon/index.html>
- [27] [DPAF] The German Processing and Archiving Facility: http://www.gfz-potsdam.de/pb1/ERS/GPM/GPM_DPAF.html
- [28] [Eineder99] Eineder, M., Holzner, J., “Phase unwrapping of low coherence differential interferograms”, in Proc. IGARSS’99, Hamburg, Germany, June 1999
- [29] [Engdahl01] Engdahl, M., Borgeaud, M., Rast, M. Use of ERS 1 / 2 tandem interferometric coherence in the estimation of agricultural crop heights, IEEE Trans GARS, 2001, Vol. 39, N. 8, pp 1799 –1806.
- [30] [EOLI] <http://www.eoli.esa.int>
- [31] [ESOC] “ESOC orbit products”, <http://nng.esoc.esa.de/ers/intro.html>
- [32] [Feigl95] Feigl, K.L., A. Sergent and D. Jacq, Estimation of an earthquake focal mechanism from a satellite radar interferogram: application to the Dec. 4 1992 Landers aftershock. Geoph. Res. Letter, 22, 1037-1048, 1995
- [33] [Ferretti00] Ferretti A., Prati C. and Rocca F.: Nonlinear subsidence rate estimation using Permanent Scatterers in differential SAR IEEE Trans. GARS Vol. 38, pp. 2202 – 2212, September 2000.
- [34] [Ferretti01] Ferretti A., Prati C. and Rocca F.: Permanent scatterers in SAR interferometry IEEE Trans. GARS - January 2001, Vol. 39, N.1. pp 8 –20.
- [35] [Ferretti96A] Alessandro Ferretti, “Generazione di mappe altimetriche da osservazioni SAR multiple”, Ph. D. Thesis, Politecnico di Milano, 1997.

- [36] [Ferretti96B] Ferretti A., Monti Guarnieri A., Prati C. and Rocca F.: “Multi Baseline Interferometric Techniques and Applications”; Proc.FRINGE 96 Workshop - Zurich, (1996), <http://www.geo.unizh.ch/rsl/fringe96/>. ESA SP-406 pp243-252
- [37] [Ferretti97] A Ferretti, C Prati, F Rocca, and A Monti Guarnieri. “Multibaseline SAR interferometry for automatic DEM reconstruction,” Proc. 3rd ERS Symposium, Florence, Italy, 17-21 March 1997, ESA SP-414, pp. 1809-1820, 1997.
- [38] [Ferretti98] Ferretti A., Monti-Guarnieri A., Prati C., Rocca F., “Multi-image DEM Reconstruction” - Proceedings IGARSS’98 International Geoscience and Remote Sensing Symposium, 6-10 July 1998, Seattle, USA.
- [39] [Ferretti99] Ferretti A., Prati C. and Rocca F.: Multibaseline INSAR DEM reconstruction, the wavelet approach IEEE Trans. GARS, 37, (1999) pp. 705-715.
- [40] [Fliege94] Norbert J Fliege, Multirate digital signal processing, John Wiley & Sons, Ltd. New York, 1994.
- [41] [Flynn97] T.J. Flynn, “Two-dimensional phase unwrapping with minimum weighted discontinuities,” J. Opt.Soc.Am.A, 14, pp. 2692-2701, 1997.
- [42] [Fornaro01] Fornaro G. and A. Monti Guarnieri. Minimum Mean Square Error Space-Varying Filtering Of Interferometric SAR Data, IEEE Trans. GARS, 39, 2001
- [43] [Fornaro95] Fornaro, G.; Franceschetti, G, Image registration in interferometric SAR processing Radar, Sonar and Navigation, IEE Proceedings - , Volume: 142 Issue: 6 , Dec. 1995 Page(s): 313 -320
- [44] [Fornaro96] Fornaro, G., Franceschetti, G., Lanari, R. Interferometric SAR Phase unwrapping using Green’s formulation, IEEE Trans. GARS, Vol. 34, N. 3, May 1996, pp. 720 – 727
- [45] [Franceschetti99] Franceschetti, G. and Lanari R.: Synthetic Aperture Radar Processing, CRC Press - 1999.
- [46] [Friedlander96] B. Friedlander and J. M. Francos, “Model Based Phase Unwrapping of 2-D Signals,” IEEE Trans. Signal Process., vol. 44, pp. 2999-3007, December 1996.
- [47] [Gabriel88] A. Gabriel, R. Goldstein, Crossed orbits interferometry: theory and experimental results from SIR-B, International Journal on Remote sensing, No.5, pp. 857-872, 1988.
- [48] [Gabriel89] Gabriel A.K., Goldstein, R.M. and Zebker H.A. Mapping small elevation changes over large areas: differential radar interferometry, J. of Geophy. Research, 94, 9183-9191, 1989.
- [49] [Garey79] M.R. Garey and D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
- [50] [Gatelli94] Gatelli F., Monti Guarnieri A., Parizzi F., Pasquali P., Prati C. and Rocca F.: IEEE Transactions on Geoscience and Remote Sensing , 32, (1994) pp. 855-865.
- [51] [Gens96] R Gens and J L van Genderen. “Review article: Sar interferometry--issues, techniques, applications,” International Journal for Remote Sensing, 17(10):1803-1835, 1996.

- [52] [Ghiglia96] D.C. Ghiglia and L.A. Romero “Minimum L_p-norm two-dimensional phase unwrapping,” Journal of the Optical Society of America A, Vol.13, No.10, pp. 1-15, October 1996.
- [53] [Ghiglia98] Ghiglia D.C. and Pritt M.D.: in Two-Dimensional Phase Unwrapping - Theory, Algorithms, and Software, John Wiley and Sons - 1998.
- [54] [Goldstein88] Goldstein, R., Zebker, H., Werner, C., Satellite radar interferometry: two dimensional phase unwrapping, Radio Science, Vol. 23, N. 4, pp. 713 – 720, Aug. 1988.
- [55] [Goldstein95] R. Goldstein, “Atmospheric Limitations to repeat-pass Interferometry”, Geophysical Research Letters, vol. 22, no. 18, pp. 2517-2520, September 1995.
- [56] [Goldstein98] Richard M Goldstein and Charles L Werner. Radar interferogram filtering for geophysical applications. Geophysical Research Letters, 25(21):4035-4038, November 1998.
- [57] [Goodman63] N R Goodman. Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). The annals of mathematical statistics, 34(1):152-177, March 1963.
- [58] [Graham74] L.C. Graham. Synthetic interferometer radar for topographic mapping. Proceedings of the IEEE, 62(6):763-768, June 1974.
- [59] [GTOPO] <http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html>
- [60] [Hanssen98] Hanssen R., Assessment of the role of Atmospheric heterogeneities in ERS Tandem SAR interferometry - Delft University Press, DEOS Report no 12698.1, 1998, Delft, the Netherlands.
- [61] [Hanssen99] Ramon Hanssen and Richard Bamler. Evaluation of interpolation kernels for SAR interferometry. IEEE Trans. on Geoscience and Remote Sensing, 37(1):318-321, January 1999.
- [62] [Hernandez96] Hernandez, B., Cotton, F., Campillo, M. and Massonnet, D., A comparison between short-term (co-seismic) and long term (one year) slip for the Landers earthquake: Measurements from strong motion and SAR interferometry, Geophys. Res. Lett., 1997. 24(13): p. 1579-1582
- [63] [INSI] <http://earth.esa.int/images/INSI/>
- [64] [Just94] Just D., and Bamler, R. Phase statistics of interferograms with applications to synthetic aperture radar. Applied Optics, 33(20):4361-4368, 1994.
- [65] [Kohlhase99] A. Kohlhase and K. Feigl “Measuring the quality of precise orbits by applying InSAR: an approach to avoid orbital tuning”, Proc. FRINGE’99 – Liege, Belgium -November 1999. ESA SP-478
- [66] [Kooij99] M. Van Der Kooij, “Production of DEMs from ERS tandem Data,” Proc. FRINGE’99 – Liege, Belgium -November 1999. ESA SP-478
- [67] [Koskinen01] Koskinen, J., Pulliainen, J., Hyppaa, M., Engdahl, M., Hallikainen, M., The seasonal behavior of interferometric coherence in boreal forest, IEEE Trans GARS, Vol. 39, Apr. 2001, pp. 820 – 829
- [68] [Kramer96] R. Kramer O. Loffeld, “Presentation of an improved Phase Unwrapping Algorithm based on Kalman filters combined with local slope estimation”, Proc. FRINGE’96 – Zurich – September 1996. – ESA SP-406 pp253-260 <http://www.geo.unizh.ch/rsl/fringe96/papers/kraemer-loffeld/>

- [69] [Laakso96] Timo I Laakso, Vesa Valimaki, Matti Karjalainen, and Unto K Laine. Splitting the unit delay. IEEE Signal processing magazine, pages 30-60, January 1996.
- [70] [Lanari98] Lanari R, S Hensley, and P A Rosen. Chirp z-transform based SPECAN approach for phase preserving ScanSAR image generation. IEE Proceedings Radar Sonar Navigation, 145(5):255-261, 1998
- [71] [Laur98] Laur, H., P Bally, P Meadows, J Sanchez, B Schättler, E Lopinto, and D Esteban. Derivation of the backscattering coefficient sigma nought in ESA ERS SAR PRI products. Technical Report ES-TN-RS-PM-HL09, ESA, September 1998. Issue 2, Rev. 5b.
- [72] [Lee94] Lee J.S., Hoppel K.W., Mango S.A., and Miller: Intensity and phase statistics of multilook polarimetric and interferometric sar imagery. IEEE Trans. GARS, 30:1017, 1994
- [73] [Lee98] Jong-Sen Lee, Konstantinos P Papathanassiou, Thomas L Ainsworth, and Mitchell R. Grunes and Andreas Reigber. A new technique for noise filtering of sar interferometric phase images. IEEE Trans. on Geoscience and Remote Sensing, 36(5):1456-1465, September 1998.
- [74] [Li90] F.K. Li and RM. Goldstein. Studies of multibaseline spaceborne interferometric synthetic aperture radars. IEEE Trans. on Geoscience and Remote Sensing, 28(1):88-97, January 1990.
- [75] [Lin92] Qian Lin, John F. Vesey, Howard A. Zebker}, New approaches in Interferometric SAR IEEE Trans. on Geoscience and Remote Sensing, 30(3):560-567, may 1992.
- [76] [Lu2000] Lu Z., D. Mann, J. T. Freymueller, and D. J Meyer. Synthetic aperture radar interferometry of Okmok volcano, Alaska: Radar observations. Journal of Geophysical Research, 105(B5):10791-10806, May 2000.
- [77] [Luca96] D. Luca, M. Datcu and K. Seidel, “Multiresolution Analysis of DEMs: Error and Artifact Characterization” Proc.FRINGE’96 – Zurich, Switzerland - October 1996. ESA SP-406 pp61-68 <http://www.geo.unizh.ch/rsl/fringe96/>
- [78] [Madsen93] S N Madsen, H A Zebker, and J Martin. Topographic mapping using radar interferometry: Processing techniques. IEEE Trans. on Geoscience and Remote Sensing, 31(1):246-255, January 1993.
- [79] [Massmann97] F.-H. Massmann, K.H. Neumayer, J.C Raimondo, K. Enninghorst and H. Li “Quality of the D-PAF ERS Orbits before and after the Inclusion of PRARE Data”, Proc. 3rd ERS Symposium, 1997, ESA SP-414 Vol. III, pp1655-1660.
- [80] [Massonnet00] “Phasemap, a simple software intended to give an efficient access to interferometry”, ERS-Envisat symposium, Göteborg, 16-20 October 2000, ESA SP-461 Filename 415masso.pdf
- [81] [Massonnet01] Massonnet D., The interferometric cartwheel, a constellation of low cost receiving satellites to produce radar images that can be coherently combined, International J. Rem. Sensing, 2001.
- [82] [Massonnet93] Massonnet, D., Rossi, M. , Carmona, C., Adragna, F. , Peltzer, G., Feigl, K. and Rabaute, T., The displacement field of the Landers earthquake mapped by radar interferometry. Nature, 1993. 364: p. 138-142. Cover paper
- [83] [Massonnet94] Massonnet, D., Feigl, K., Rossi, M. and Adragna, F., Radar interferometric mapping of deformation in the year after the Landers earthquake. Nature, 1994. 369: p. 227-230.

- [84] [Massonnet95A] Massonnet, D. and K.L. Feigl, Discrimination of geophysical phenomena in satellite radar interferograms. *Geophys. Res. Lett.*, 1995. 22(12): p. 1537-1540
- [85] [Massonnet95B] Massonnet, D. and H. Vadon, ERS-1 Internal clock drift measured by interferometry. *IEEE Trans. Geoscience & Rem. Sensing*, 1995. 33(2): p. 401-408
- [86] [Massonnet95C] Massonnet, D., P. Briole, and A. Arnaud, Deflation of Mount Etna monitored by spaceborne radar interferometry. *Nature*, 1995. 375: p. 567-570
- [87] [Massonnet96] Massonnet, D., W. Thatcher, and H. Vadon, Detection of Postseismic Fault Zone Collapse following the Landers Earthquake. *Nature*, 1996. 382: p. 612-616
- [88] [Massonnet98] D. Massonnet and K.L. Feigl, "Radar interferometry and its application to changes in the earth's surface. *Reviews of Geophysics*, 36(4):441-500, November 1998.
- [89] [Moeremans99] B. Moeremans and S. Dautrebande The Use of ERS SAR Interferometric Coherence and PRI Images to Evaluate Crop Height and Soil Moisture, Fringe 99, ESA SP-478
- [90] [MontiGuarnieri00] A. Monti Guarnieri, A. Ferretti, Visibility of Permanent Scatterers by ScanSAR, in proc EUSAR 2000 (Munich, Germany), May, 23-25, 2000, pp. 725-728, 2000.
- [91] [MontiGuarnieri01A] Monti Guarnieri A, Pietro Guccione. Optimal "focusing" for low resolution ScanSAR. In *IEEE Trans. GARS*, 39(3):479-491, Mar 2001
- [92] [MontiGuarnieri01B] A Monti Guarnieri, et al, ASAR ScanSAR Interferometrics Products, Software Requirement Document, ESA contract report, contract N. 14143/00/NL/GD.
- [93] [MontiGuarnieri93] Monti Guarnieri A, C Prati, and F Rocca. SAR interferometric quick-look. In IGARSS93, Tokyo, Japan, 18-21 August 1993, pages 988-990, 1993.
- [94] [MontiGuarnieri96A] Monti Guarnieri A. and Claudio Prati. ScanSAR focusing and interferometry. *IEEE Trans. GARS*, 34(4):1029-1038, July 1996.
- [95] [MontiGuarnieri96B] Monti Guarnieri, A.. Residual sar focusing: an application to coherence improvement. *IEEE Trans. GARS*, 34(1):201-211, January 1996.
- [96] [MontiGuarnieri97] Monti Guarnieri A and Claudio Prati. SAR interferometry: A "quick and dirty" coherence estimator for data browsing. *IEEE Trans. GARS*, 35(3):660-669, May 1997.
- [97] [MontiGuarnieri99A] A Monti Guarnieri and F Rocca. Combination of low- and high-resolution SAR images for differential interferometry. *IEEE Trans. on Geoscience and Remote Sensing*, 37(4): 2035-2049, July 1999.
- [98] [MontiGuarnieri99B] Monti Guarnieri A and C Prati. An interferometric quick-look processor *IEEE Trans. GARS*, 37(2):861-866, 1999.
- [99] [Moore81] Moore R K, J P Claassen, and Y H Lin. Scanning spaceborne Synthetic Aperture Radar with integrated radiometer. *IEEE Trans. AES*, 17(3):410-420, 1981.
- [100] [Moreira00] Moreira A. and R Schreiber. Coregistration of interferometric SAR images using spectral diversity *IEEE Trans. GARS*, 38(5):2179-2191, 2000.

- [101] [Moreira96] Moreira A., J Mittermayer, and R Schreiber. Extended chirp scaling algorithm for air and spaceborne SAR data processing in stripmap and ScanSAR imaging modes. *IEEE Trans. GARS*, 34(5):1123-1136, 1996.
- [102] [Muller99] J.P. Muller, "The LANDMAP project for the creation of multi-sensor geocoded and topographic map products for the British Isles based on ERS-tandem interferometry," *Proc. FRINGE'99*, November 1999. ESA SP-478
- [103] [Murru99] Murru M., Montuori C., Wyss M., Privitera E. (1999): The location of magma chambers at Mt. Etna, Italy, mapped by b-Values; *Geophys. Res. Letts.*, Vol. 26, N. 16, pp2553-2556.
- [104] [Pasquali94] P Pasquali, R Pellegrini, C Prati, and F Rocca. Combination of interferograms from ascending and descending orbits. In *International Geoscience and Remote Sensing Symposium*, Pasadena, CA, USA, 8-12 August 1994, pages 733-735, 1994.
- [105] [Pasquali98A] P. Pasquali, O. Stebler, D. Small, F. Holecz, and D. Nuesch. SAR Polarimetric Interferometry Experiments, PIERS Workshop on Advances in Radar Methods, Baveno, July 20-22, 1998.
- [106] [Pasquali98B] P. Pasquali, D. Small, F. Holecz, A. Schubert, E. Meier, D. Nuesch, Study of the ENVISAT ASAR alternate polarization mode: preliminary results, *Proceedings of EUSAR Symposium*, Friedrichshafen, May, 25-27 1998, pp. 335-338.
- [107] [Peltzer96] Peltzer, G, P. Rosen, F. Rogez and K. Hudnut, Post seismic in fault step overs caused by pore fluid flow, *Science*, 273, 1202-1204, 1996
- [108] [Peltzereos] G. Peltzer, "Monitoring of Ground Subsidence," <http://www.esrin.esa.it/esrin/eos/mon2.html>
- [109] [Prati90A] Prati, C., Giani, M., Leuratti, N., SAR Interferometry: a 2D Phase Unwrapping Technique Based on Phase and Absolute Value Informations, In *International Geoscience and Remote Sensing Symposium*, Washigton, pp. 2043-2046, 1990.
- [110] [Prati90B] Prati, C.; Rocca, F.; Guarnieri, A.M.; Damonti, E. Seismic Migration For Sar Focusing: Interferometrical Applications *IEEE Trans. GARS*, Volume: 28 Issue: 4 , July 1990 Page(s): 627 -640
- [111] [Prati92] C Prati and F Rocca. Range resolution enhancement with multiple SAR surveys combination. In *International Geoscience and Remote Sensing Symposium*, Houston, Texas, USA, May 26-29 1992, pages 1576-1578, 1992.
- [112] [Prati93] Prati C., and F Rocca, Improving Slant-Range Resolution With Multiple SAR Surveys, *IEEE Trans. AES* 1993 jan, vol 29(1), pp 135-144.
- [113] [Prati94A] Prati C., F. Rocca, 1994, Process for generating Synthetic Aperture Radar Interferograms, U.S. Patent N.5,332,999, July 26.
- [114] [Prati94B] Prati C., Rocca F., Monti Guarnieri A., Pasquali P.: Interferometric Techniques and Applications - ESA Study Contract Report. Contract N.3-7439/92/HGE-I.C.
- [115] [Pritt97] M.D. Pritt, "Congruence in Least-Squares Phase Unwrapping," In *International Geoscience and Remote Sensing Symposium*, Singapore, 3-8 August 1997, pp. 875-877, 1997.
- [116] [Ranson99] Ranson K. Jon, Guoqing, Sun; Forest Disturbance Characterization Using ERS Tandem Data, *Fringe 99*, ESA SP-478

- [117] [Reigber96] Reigber C., Xia Y., Kaufmann H., Massmann F., Timmen L., Bodechtel J. and Frei M.: "Impact of precise orbits on SAR interferometry", Proc.FRINGE 96 Workshop - Zurich, (1996), ESA SP-406 pp223-232.
- [118] [Rignot98] Rignot, E.J., Fast recession of a west Antarctic glacier, *Science*, 281, page 549-551, 1998
- [119] [Rocca94] Rocca F., Prati C., Pasquali P., Monti Guarnieri A., "ERS-1 SAR Interferometry techniques and applications", ESA contract report n.3-7439/92/HGE-I, 1994.
- [120] [Rodriguez92] E Rodriguez and J M Martin. Theory and design of interferometric synthetic aperture radars. *IEE Proceedings-F*, 139(2):147-159, April 1992.
- [121] [Rosen00] Rosen, P., Hensley, S., Joughin, I., Li, F., Madsen, S., Rodriguez, E., Goldstein, R., Synthetic Aperture Radar Interferometry, Proc. IEEE March 2000, pp. 333 – 379.
- [122] [Rosich96] Rosich Tell B. and H Laur. Phase preservation in sar processing: the interferometric offset test. In IGARSS96, Lincoln, Nebraska, USA, 27-31 May 1996.
- [123] [Rott96] Rott, H. Siegel, A., Glaciological Studies in the Alps and in Antarctica using ERS interferometric SAR, Proc. Fringe 1996, ESA SP-406 Vol. 2, pp. 149 – 159.
- [124] [Sandwell98] Sandwell, D.T. and E.J. Price, Phase gradient approach to stacking interferograms, *JGR*, vol.103, pages 30183-30204, 1998
- [125] [Scharroo94] Scharroo, R., K. F. Wakker, and G. J. Mets, The orbit determination accuracy of the ERS-1 mission, in Proceedings of the Second ERS-1 Symposium, Hamburg, Germany, 11-14 October 1993, ESA SP-361, vol. 2, edited by B. Kaldeich, pp. 735-740, 1994.
- [126] [Scharroo97] Remko Scharroo and Pieter N. A. M. Visser "ERS Tandem Mission orbits: is 5 cm still a challenge?", Proc ERS International Symposium – ESA SP-414 pp1643-1648, 1997
- [127] [Schmullius99] C. Schmullius, A. Holz, U. Marschalk, A. Roth, J. Vietmeier, W. Wagner Operational Readiness of ERS SAR Interferometry for Forest Mapping in Siberia, Fringe 99, ESA SP-478
- [128] [Schreier93] Gunter Schreier, editor. SAR Geocoding: data and systems. Wichmann Verlag, Karlsruhe, 1993.
- [129] [Schwartzman94] Schwartzman, N., Freund, I., Wave field phase singularities: near neighbor correlations and anticorrelations, *JOSA* Vol. 11, N. 10, October 1994, pp. 2710 – 2718
- [130] [Seymour94] M S Seymour and I G Cumming. Maximum likelihood estimation for SAR interferometry. In International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 8-12 August 1994, pages 2272-2275, 1994.
- [131] [Sharp92] Sharp, R. EOS, *Transactions AGU* vol 73, no. 43, p380, 1992
- [132] [Small96] D. Small, P.Pasquali, and S.Fuglistaler, "A Comparison of Phase to Height Conversion Methods for SAR Interferometry" – Proc. IGARSS'96 – Lincoln, Nebraska – USA – 27-31 May 1996.
- [133] [Small98] D. Small, "Generation of Digital Elevation Models through Spaceborne SAR Interferometry" – PhD Thesis – Remote Sensing Labs – University of Zurich – 1998.

- [134] [SNAPHU] http://www-star.stanford.edu/sar_group/snaphu/
- [135] [Solaas94] Gaute Aarbakke Solaas. Ers-1 interferometric baseline algorithm verification. Technical Report ES-TN-DPE-OM-GS02, ESA-ESRIN, March 1994.
- [136] [Spagnolini95] Umberto Spagnolini. 2-D phase unwrapping and instantaneous frequency estimation. IEEE Trans. on Geoscience and Remote Sensing, 33(3):579-589, May 1995.
- [137] [SRTM] "Shuttle Radar Topography Mission - Mapping the World in Three Dimensions", <http://www.jpl.nasa.gov/srtm/>
- [138] [Stan00] S. Stan, W. Knoepfle, D. Scales, "Registration of Ascending-Descending SAR Images for the SRTM Mission," Proc. EUSAR2000, 23-25 May 2000, Munich, Germany.
- [139] [Strang97] G. Strang and K. Borre, Linear Algebra, Geodesy, and GPS – Wellesley-Cambridge Press - 1997.
- [140] [Suess98] Suess M., M. Völker, J. J. W. Wilson and C. H. Buck, Superresolution: range resolution improvement by coherent combination of repeat pass SAR images, Proc. EUSAR98 (Friedrichshafen, Germany), 25-27 May, 1998, pp 565-569
- [141] [Tarayre96] Tarayre, H., and Massonnet, D., 1996, Atmospheric propagation inhomogeneities revealed by ERS-1 interferometry, Geophysical Research Letters, 23 (9) 989 – 992.
- [142] [Tebaldini05] A. Monti Guarnieri, S. Tebaldini, Channel phase estimate in time variant SIMO systems, in Proc. ICASSP 2006, (Toulouse, France), May 14-19, 2006, Page(s): IV-817 - IV-820
- [143] [Touzi99] Touzi, R.; Lopes, A.; Bruniquel, J.; Vachon, P.W.; Coherence estimation for SAR imagery; Geoscience and Remote Sensing, IEEE Transactions on Volume 37, Issue 1, Part 1, pp:135 – 149, Jan. 1999.
- [144] [Touzi96] Ridha Touzi and Armand Lopes. Statistics of the Stokes parameters and of the complex coherence parameters in one-look and multilook speckle fields. IEEE Trans. on Geoscience and Remote Sensing, 34(2):519-531, March 1996.
- [145] [USGS] GTOPO30 Documentation
<http://edcdaac.usgs.gov/gtopo30/README.html>
- [146] [Villasenor92] Villasenor J., and Howard A Zebker. Studies of temporal change using radar interferometry. In Proceedings of the SPIE, volume 1630, pages 187-198, 1992.
- [147] [Wackernagel98] H. Wackernagel, "Multivariate Geostatistics – 2nd Edition" – Springer-Verlang – 1998.
- [148] [Walker98] Walker N. P., Stefano Mica, and Henri Laur. The ERS interferometric quick look processor: a powerful system to assess and exploit the ERS tandem archive for land-use information retrieval. Publication off-print, 1998
- [149] [Weber00] Weber Hoen, E., Zebker, H., Penetration Depths inferred from interferometric volume decorrelation observed over the Greenland Ice Sheet, IEEE Trans. GARS, Vol. 38, N. 6, Nov. 2000, pp. 2571 - 2583

- [150] [Wegmüller97] Wegmüller U., "Soil Moisture Monitoring With ERS SAR Interferometry." In Proceedings of the third ERS Symposium, ESA SP-414 pp47-52, 1997.
- [151] [Werner93] C. L. Werner, S. Hensley, R. M. Goldstein, P. A. Rosen and H. A. Zebker, "Techniques and Applications of SAR Interferometry for ERS-1: Topographic Mapping, Change Detection and Slope Measurement," Proc. 1st ERS-1 Symposium - Cannes, France, 4-6 November 1992 - ESA SP-359 pp205-210 - March 1993.
- [152] [Weydahl01] Weydahl, D. J., Analysis of ERS Tandem Coherence from glaciers, valleys and fjord ice in Svalbard, IEEE Trans GARS, Vol. 39, Sept. 2001, pp. 2029 – 2039
- [153] [Wornell93] G. W. Wornell, "Wavelet-Based Representations for the 1/f Family of Fractal Processes" IEEE Proceedings, vol. 81, no. 10, pp. 1428-1450, Oct. 1993.
- [154] [Xu99] Wei Xu and Ian Cumming. "A region-growing algorithm for InSAR phase unwrapping," IEEE Trans. on Geoscience and Remote Sensing, 37(1):124-134, January 1999.
- [155] [Zebker86] H.A. Zebker and R.M. Goldstein "Topographic mapping from interferometric synthetic aperture radar observations" Journal of Geophysical Research, 91(B5):4993-4999, April-10, 1986.
- [156] [Zebker92] H.A. Zebker and J. Villasenor. Decorrelation in interferometric radar echoes. IEEE Trans. on Geoscience and Remote Sensing, 30(5), pp. 950-959, September 1992.
- [157] [Zebker94A] Zebker, H.A., C.L. Werner, P. Rosen, and S. Hensley, "Accuracy of topographic maps derived from ERS-1 radar interferometry", IEEE Trans. on Geoscience and Remote Sensing, Vol. 32, No. 4, pp 823-836, July 1994
- [158] [Zebker94B] Zebker, H. A., P.A. Rosen, R.M. Goldstein, A. Gabriel, and C.L. Werner, On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake, J. of Geophysical Research, 99, 19617-19634, 1994
- [159] [Zebker97] H.A. Zebker and P. Rosen, "Atmospheric Artifacts in Interferometric SAR Surface Deformation and Topographic Maps" J.Geophys.Res.- Solid Earth – vol. 102 (B4), pp. 7547-7563, 1997.