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.

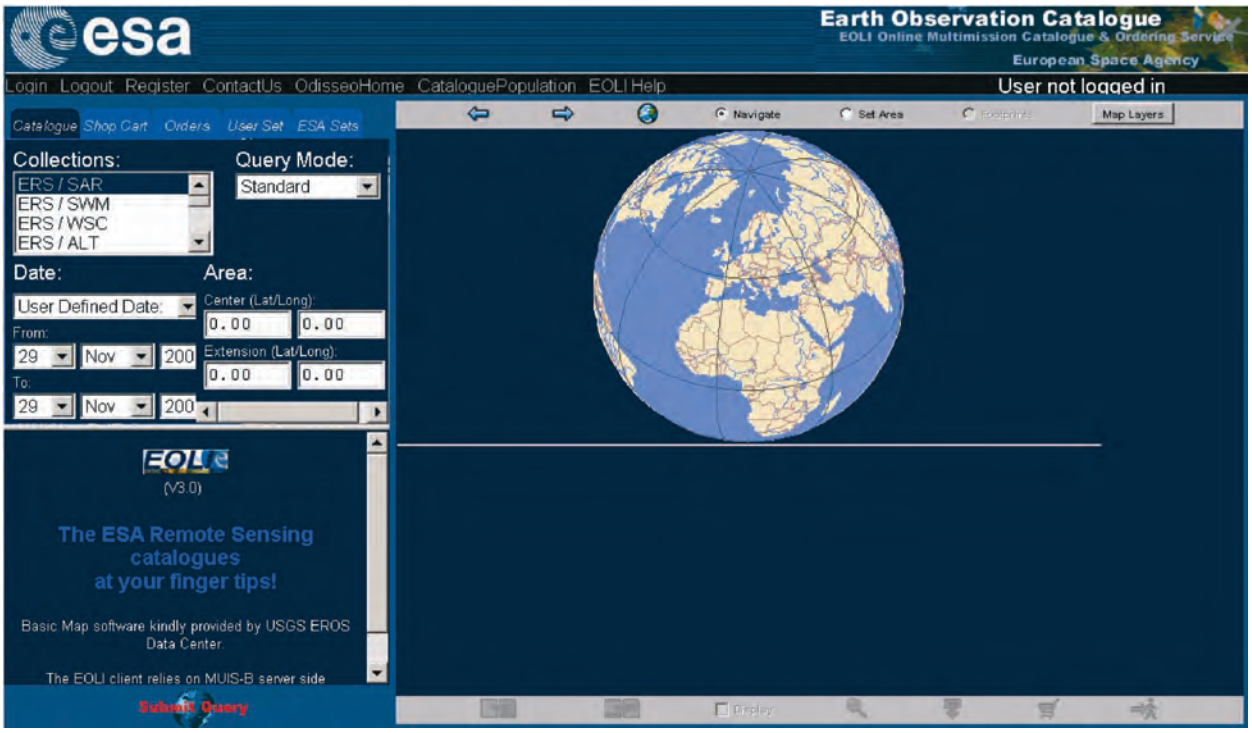


Figure 1-1: The EOLI query panel. The area of interest can be selected on the right panel by setting a window on the world map, or on the left panel by entering the geographical coordinates. On the left panel, users can select the mission type, the range of acquisition dates and the query mode. If the '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.

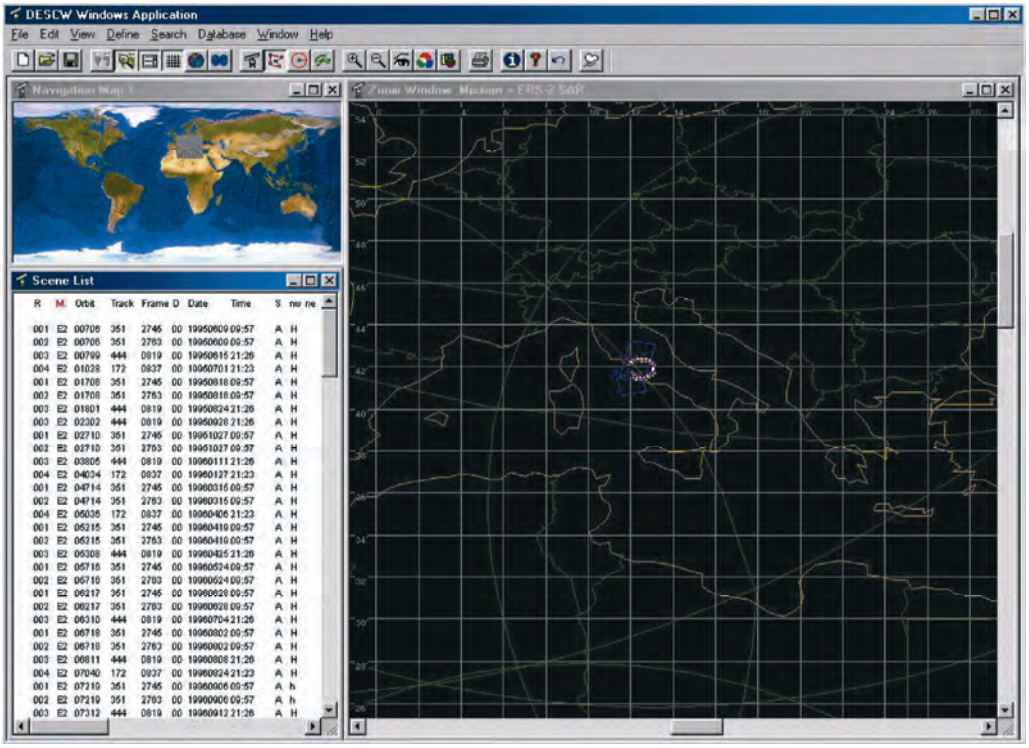


Figure 1-2: The DESCW query panel. The area of interest can be selected by providing the geographic coordinates and checking the result on the map on the right panel. Users can select the mission type, the range of acquisition dates and the range of baselines.

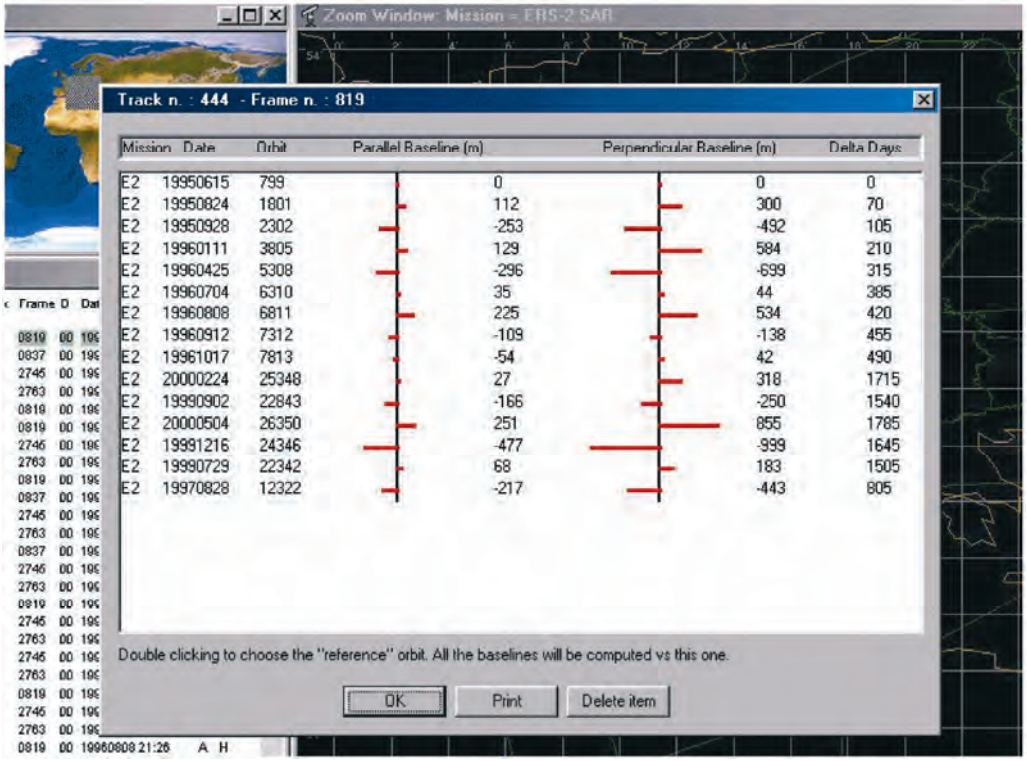


Figure 1-3: The DESCW software provides estimated perpendicular baselines of all the selected images with respect to a reference nominal orbit

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.

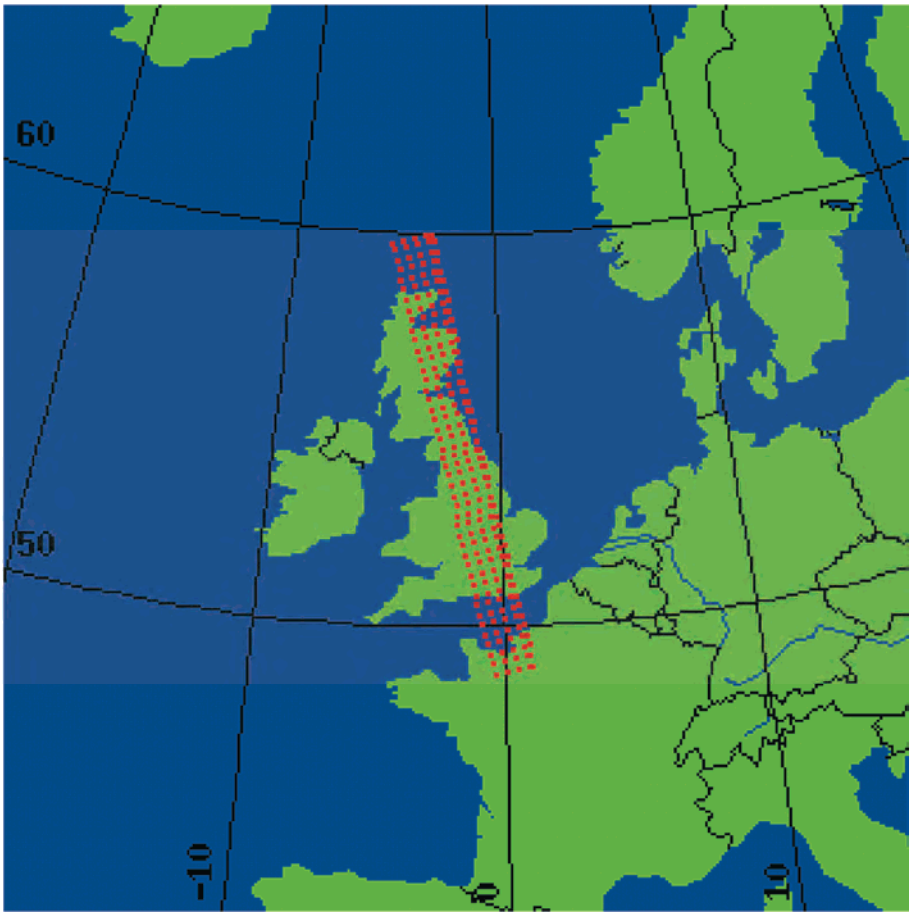
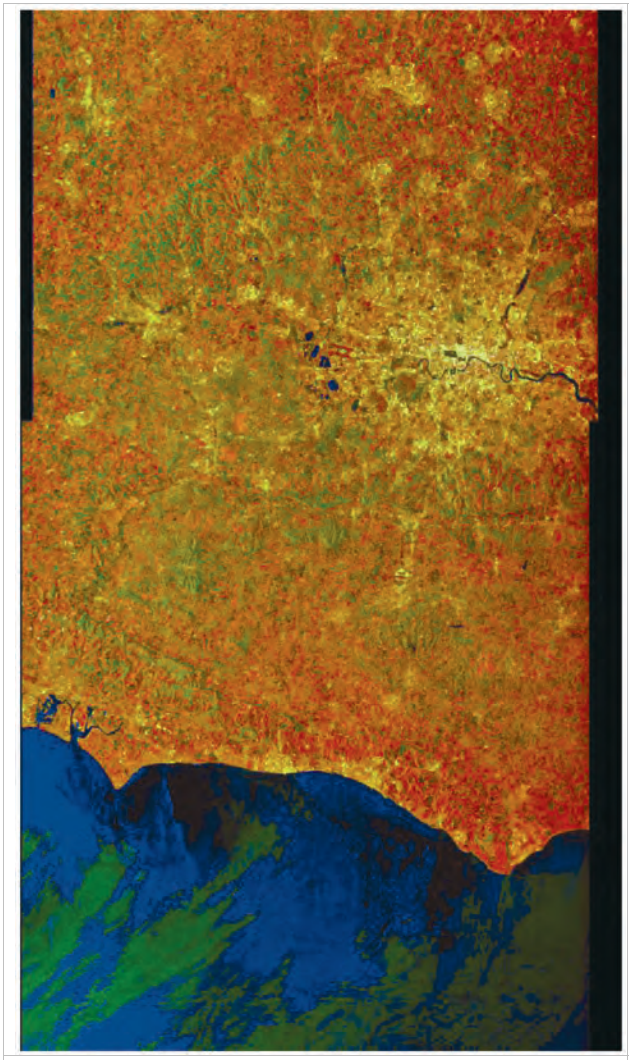


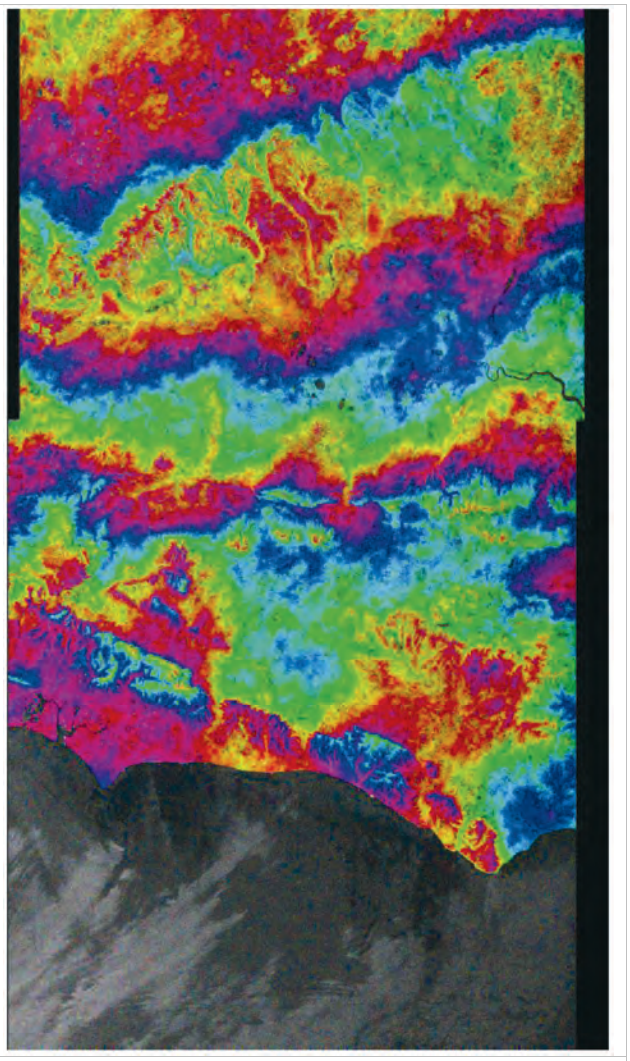
Figure 1-4: Map of the area processed by the IQL software



Left: detail of the Interferometric Land-Use (ILU) Image over London. The ILU image is an RGB image where the separate channels have been coded such that:

Red = Interferometric coherence
Green = Average intensity of the two acquisitions
Blue = Intensity change between the two acquisitions

This image is useful for quickly assessing the suitability of the interferometric pair for discrimination between different land-use types. In addition, this particular RGB combination has been chosen so that the land-use types are coloured in a manner that might be found in a photographic image. In this way the ILU images can be thought of as a form of 'colour SAR image'.



Right: the image shows the interferometric phase projected on a colour wheel (in those regions where the coherence is higher than 0.2) and the average intensity image on a grey-scale elsewhere. This image is useful for quickly assessing the quality of the interferometric fringes, which might, for example be used to make digital elevation models (DEMs).

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:

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

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.).

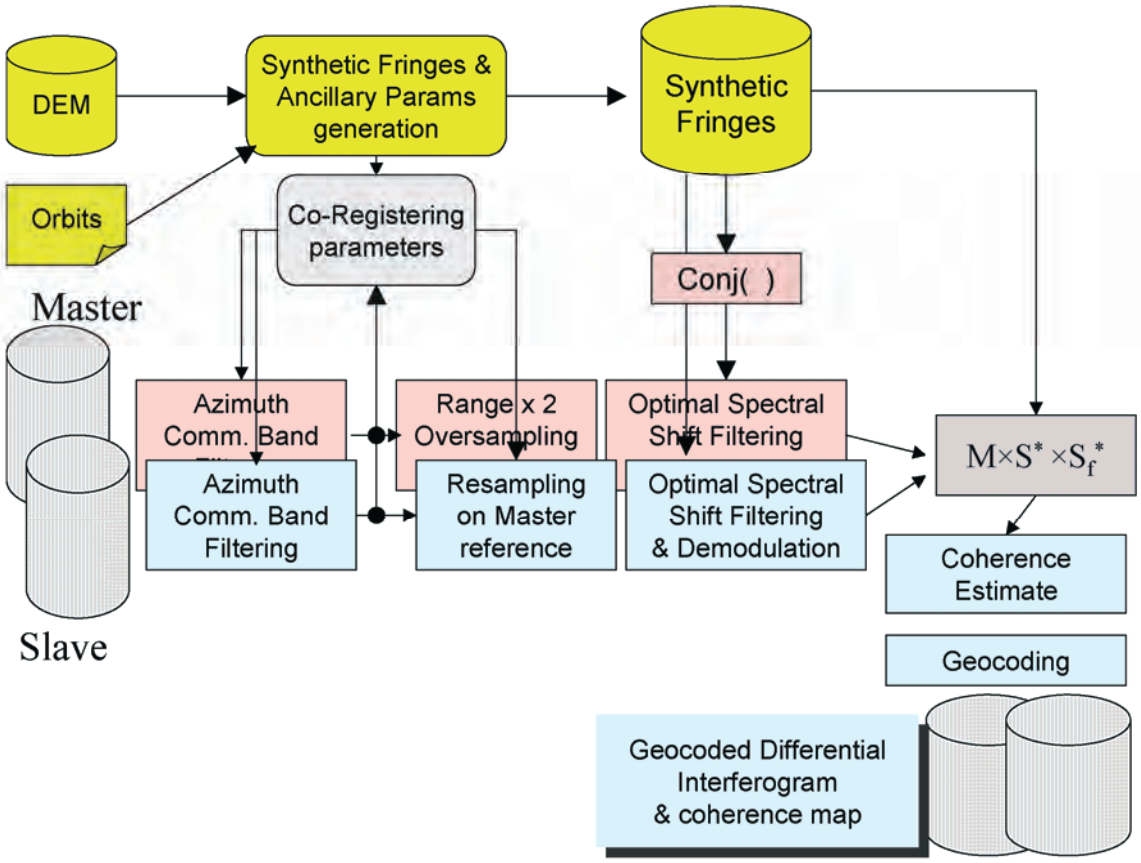


Figure 2-1: Schematic block diagram of an interferometric processor

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.

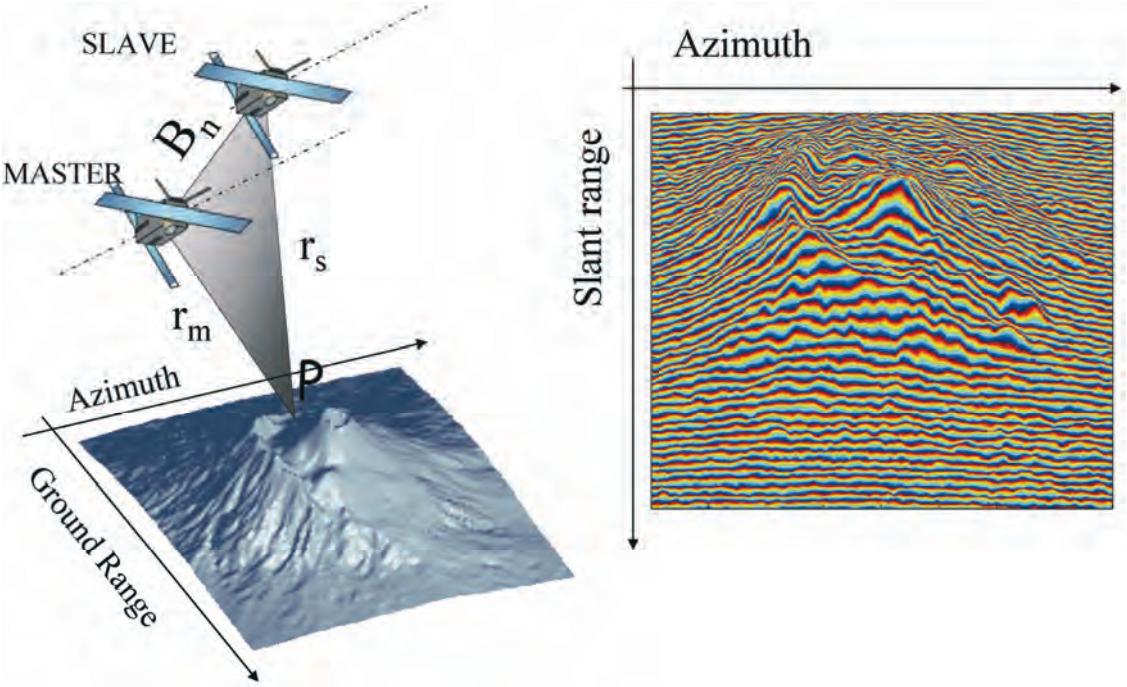


Figure 2-2: Generation of a synthetic interferogram (right) based on a DEM (left) and knowledge of the sensor orbits and timing

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)):

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

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 reference[i]. 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**:

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

- 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.

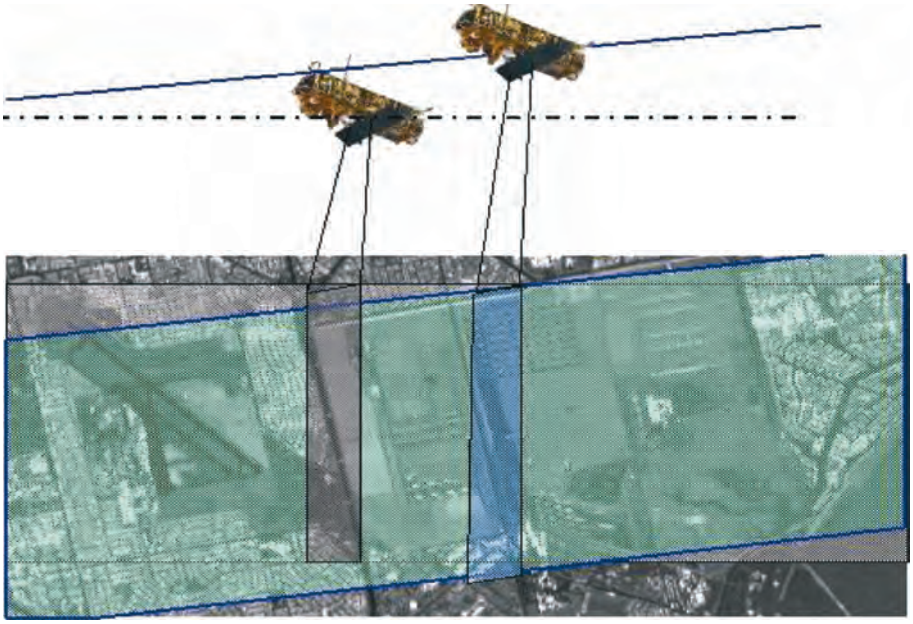


Figure 2-3: Co-registration should be performed to compensate for orbit skew, different sampling and baseline-induced deformations between the two acquisitions

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 × 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:

$$\begin{cases} r_S = a \cdot r_M^2 + b \cdot r_M + c \cdot a_M + d \\ a_S = e \cdot r_M^2 + f \cdot r_M + g \cdot a_M + h \end{cases} \qquad \text{Equation 2.3}$$

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.

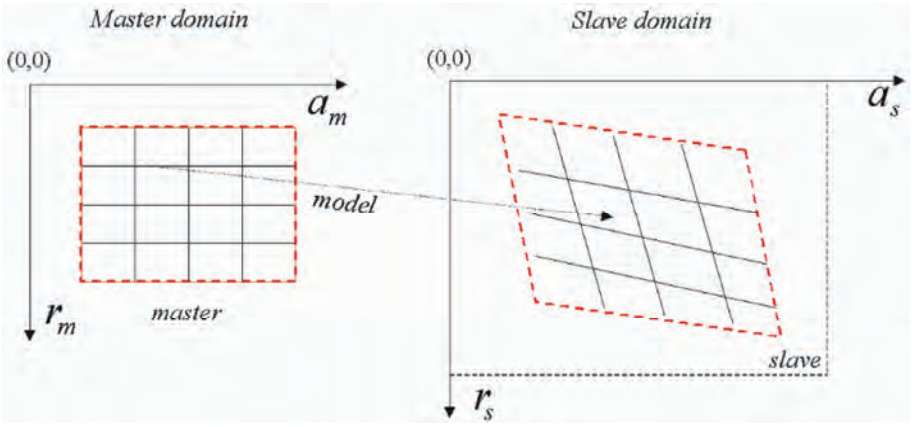


Figure 2-4: Deformation model to register the slave image on the reference grid of the master. Note the azimuth & range shift, stretch & rotation (actually implemented by two 1-D skews).

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 $\beta < 1/100^{o}$ (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.

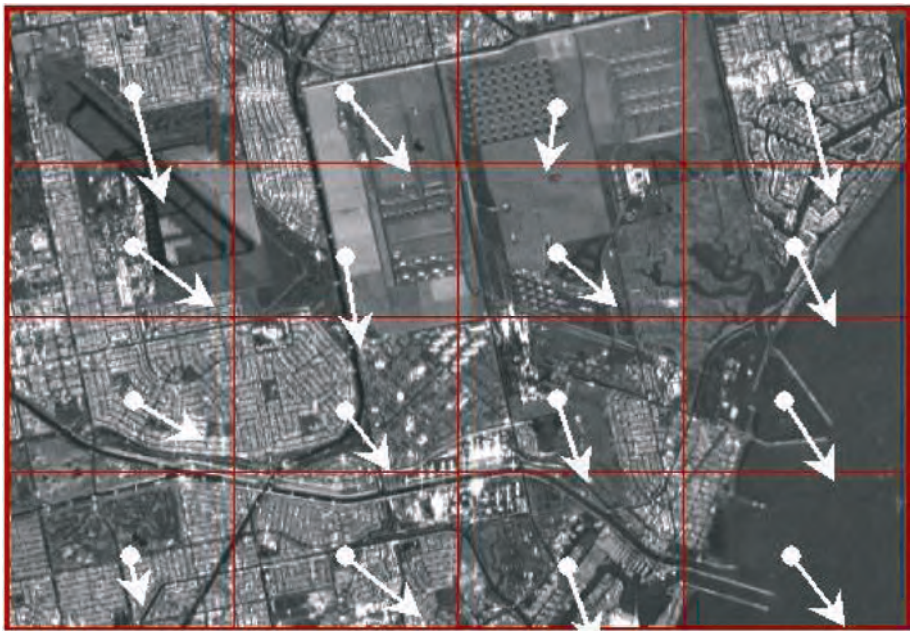


Figure 2-5: The estimate of co-registering parameters from data is here pictorially represented. A shift vector is drawn for each sub-image: it has been computed by maximising some statistical measure (e.g. the cross-correlation between master and slave amplitudes), and then a smooth polynomial is fitted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

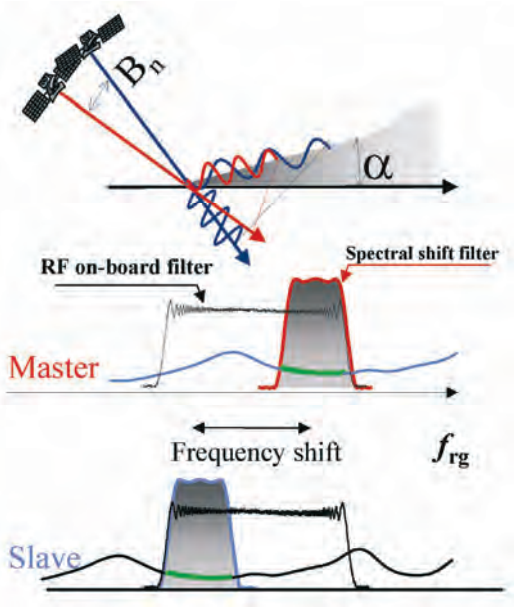


Figure 2-6: Range spectral shift. In range, the change in the view angle introduces a frequency shift in the reflectivity. The filters required to remove the uncorrelated spectral contributions are shown in the middle and bottom plots.

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:

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

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 $Br - \Delta f$.

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.

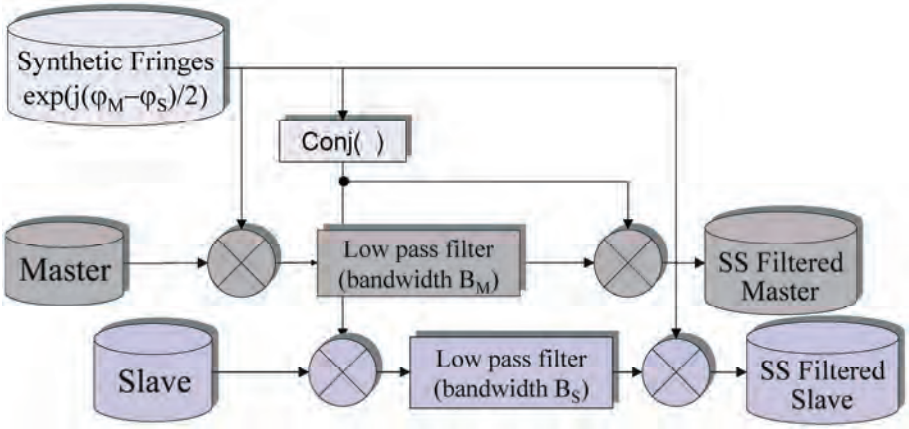


Figure 2-7: Optimal range-varying spectral shift filtering implementation

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.

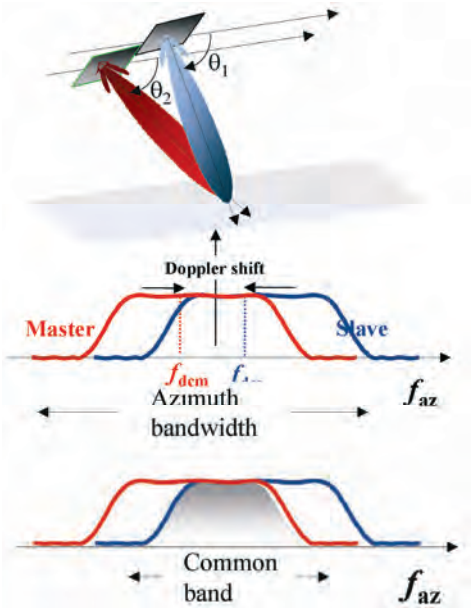


Figure 2-8: Azimuth common band filtering. In azimuth, the shift is due to a change in the squint angle. The filters required to remove the uncorrelated spectral contributions are shown in the middle and bottom plots.

[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:

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

where we have emphasised the (slight) dependence of the master and slave Doppler Centroid on range[iii]. 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.

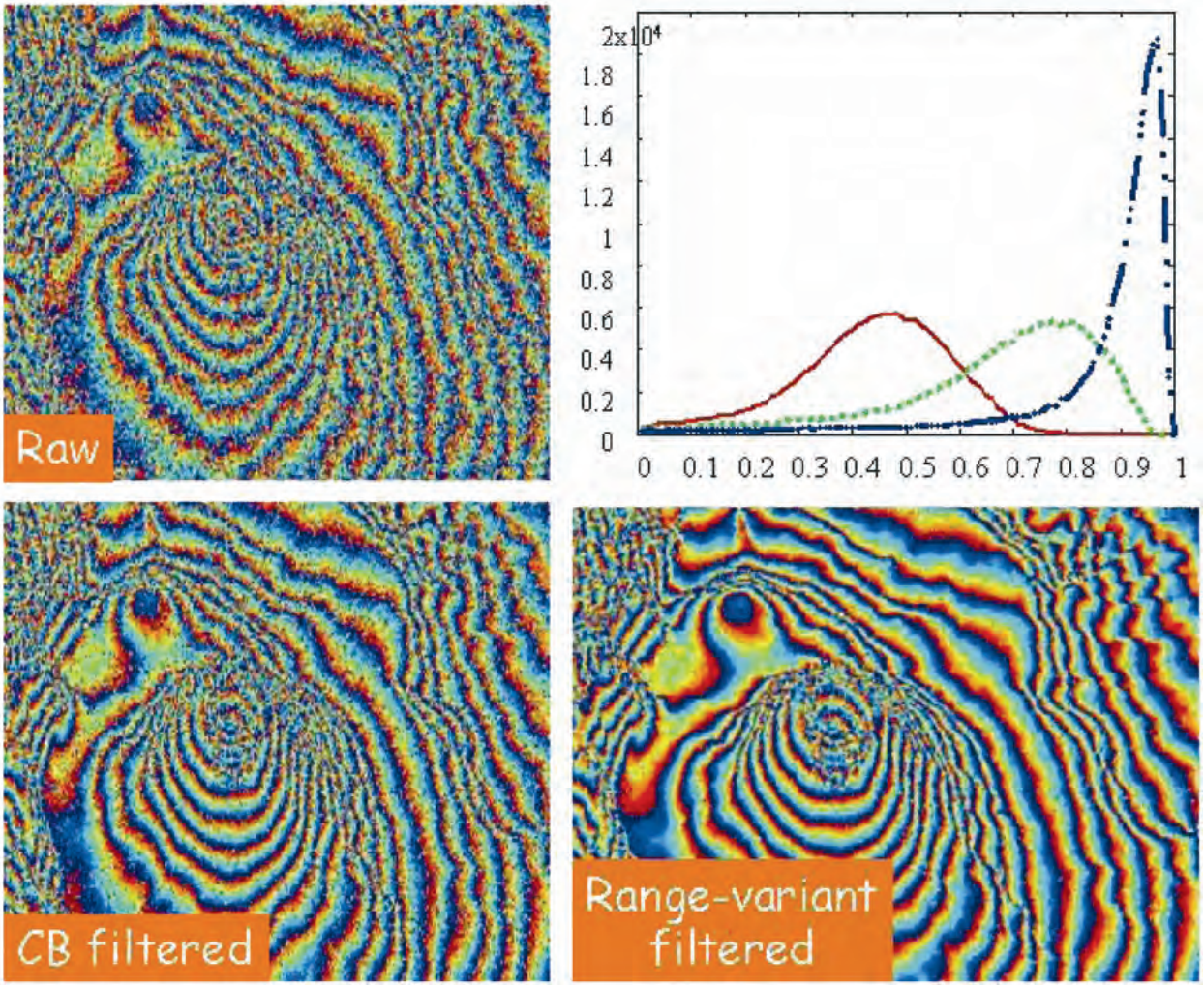


Figure 2-9: Effects of range and azimuth filtering on the final interferogram. A simulated interferogram of Mt. Vesuvius has been generated, (top left), by assuming no filtering. The decorrelation is entirely due to geometry and to the azimuth spectral shift (a Doppler centroid difference of 300 Hz was 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]:

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

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.

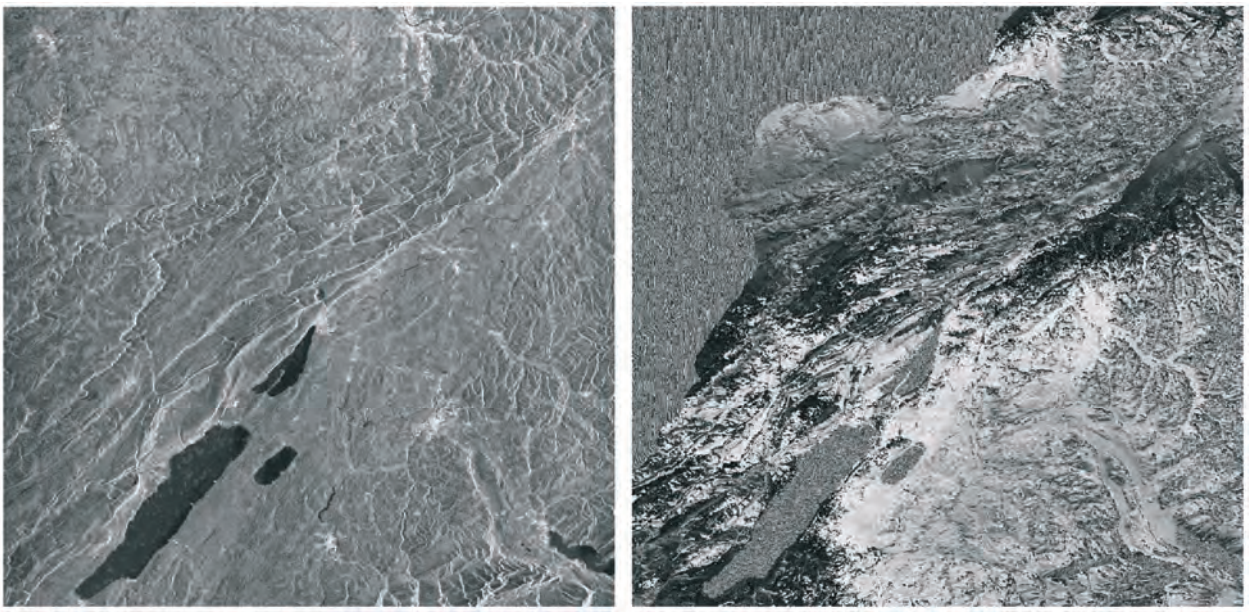


Figure 2-10: Full resolution, full frame Tandem ERS interferogram of Bern area. Baseline: 120 m. Left: absolute value. The interferogram phase (right) has been compensated by exploiting a 50 m DEM of Switzerland, according to the scheme in Figure 2-1. The residual phase shown in the figure is the result 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.

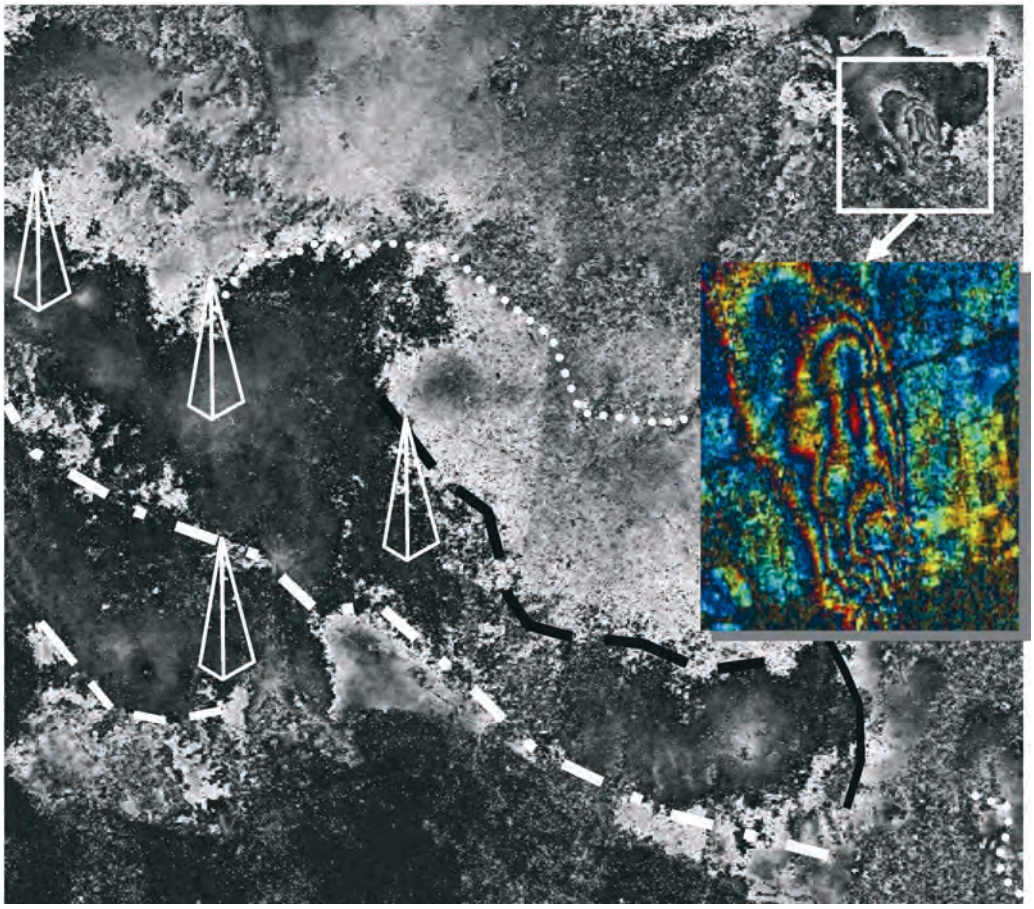


Figure 2-11: Full resolution, approximately half frame ERS interferogram of the LA area obtained by two acquisitions taken two years apart. Normal baseline: 120 m. The interferogram has been compensated by combining two different DEMs (the junction between the DEMs appears at the centre of the image). 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.

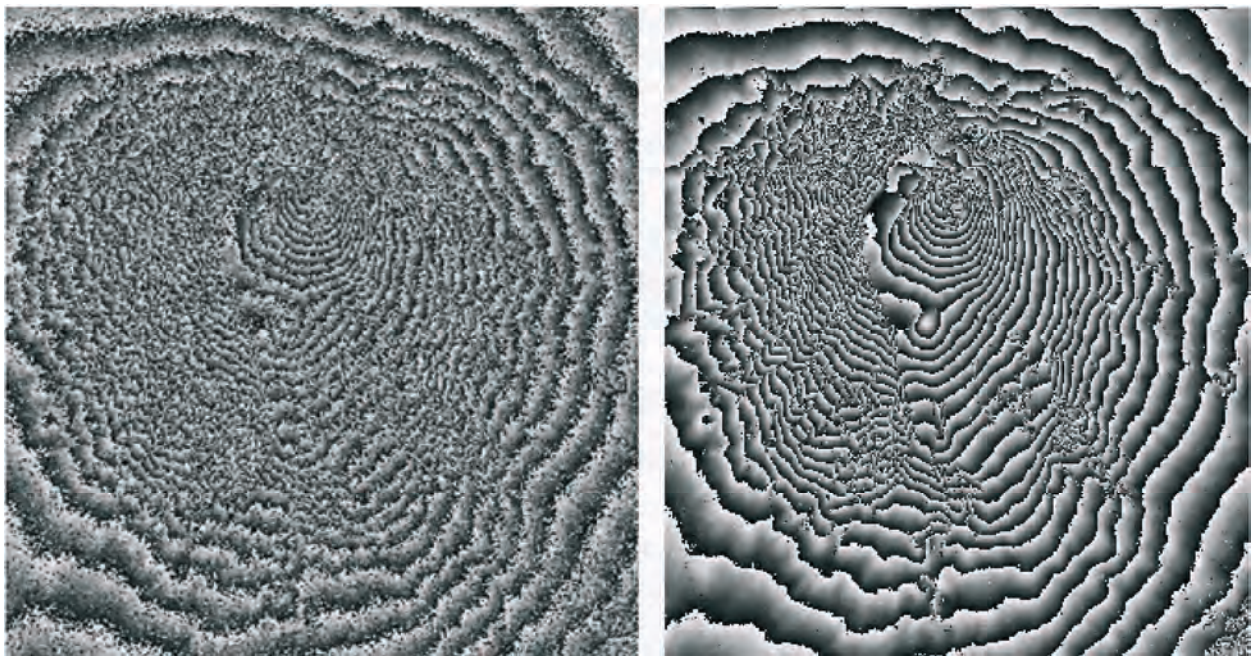


Figure 2-12: ERS tandem interferogram (compensated for flat Earth) before (left) and after (right) complex multi-looking. Mt. Vesuvius (Italy), normal baseline 250 m.

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 ~20×20 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.

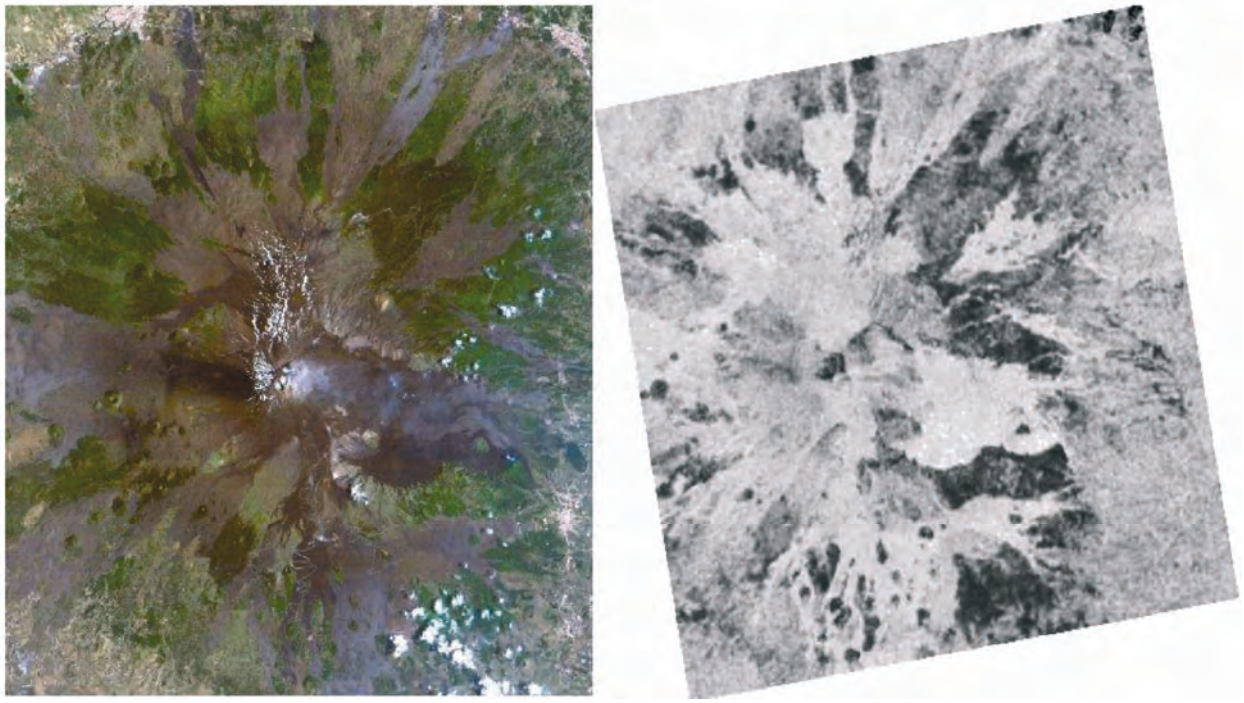


Figure 2-14: Volcano Mount Etna: Landsat^{TM} image (left) and coherence (right)

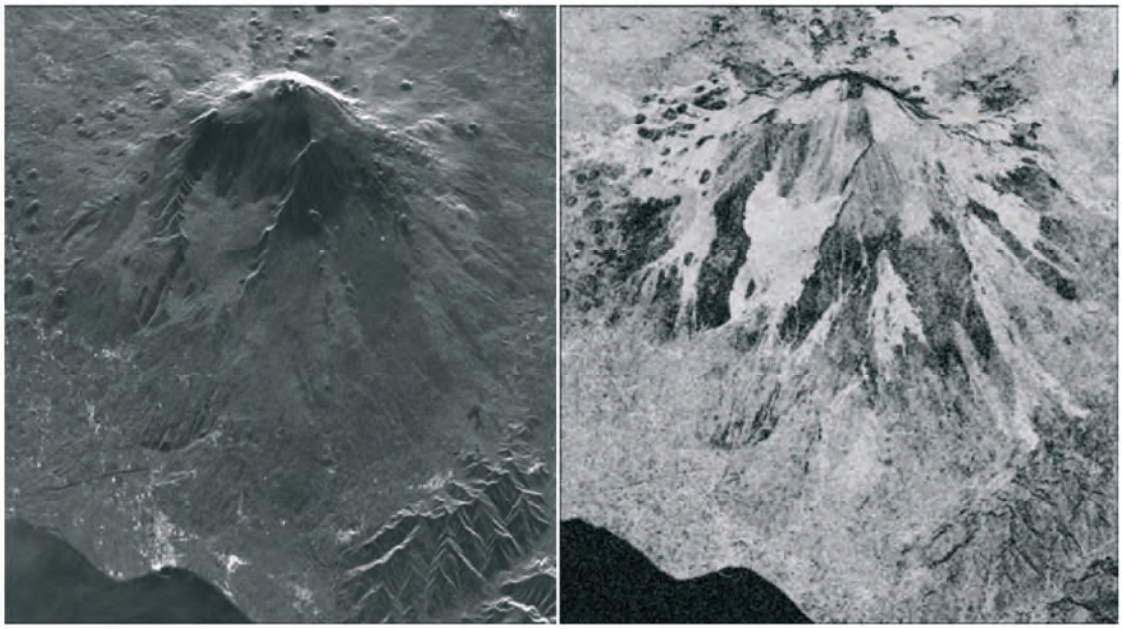


Figure 2-15: Coherence images (right) of Mount Etna together with the average amplitude image

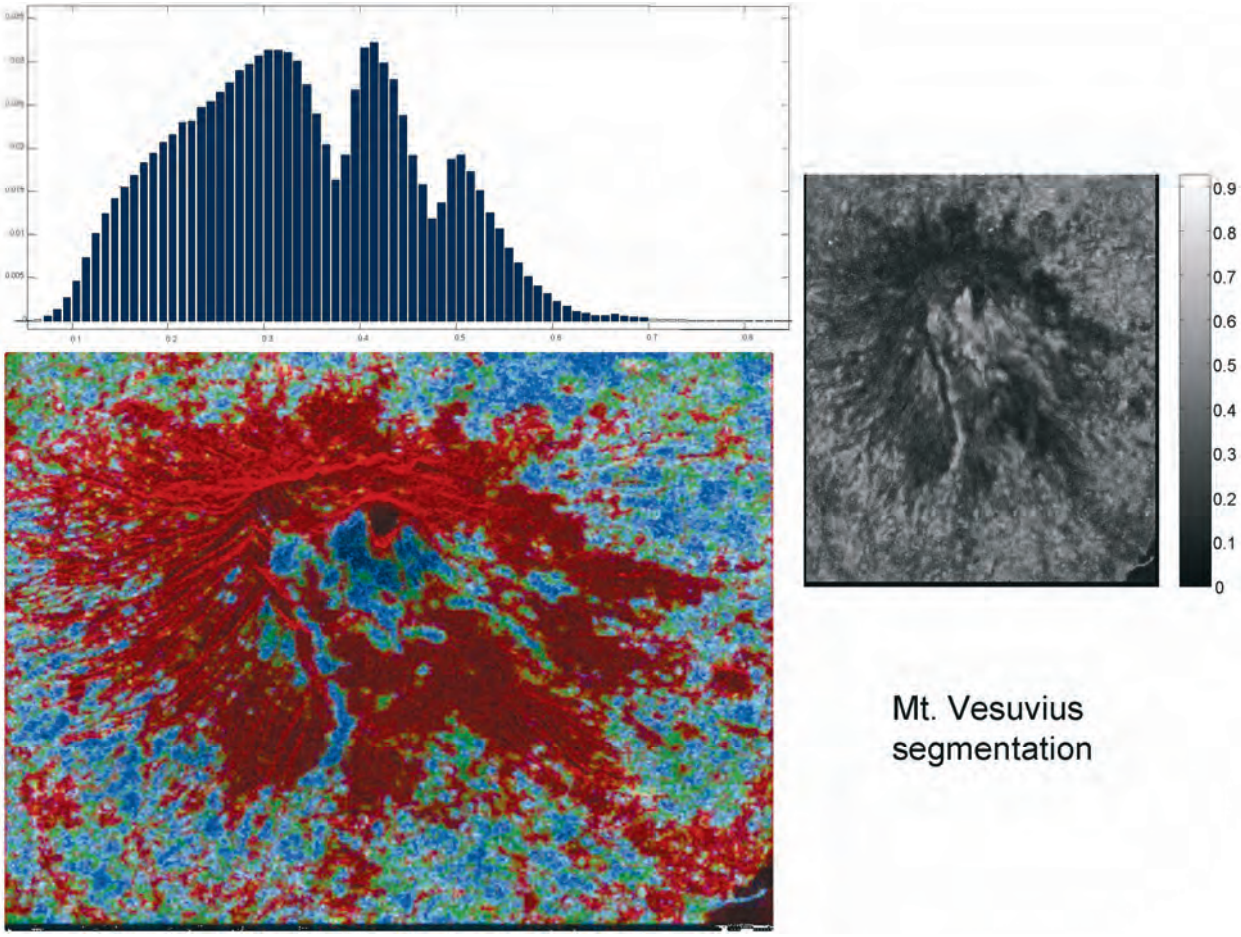


Figure 2-16: Coherence-based segmentation on Mount Vesuvius

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 $\Delta R(\mathrm{P})$ 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 dampened 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.

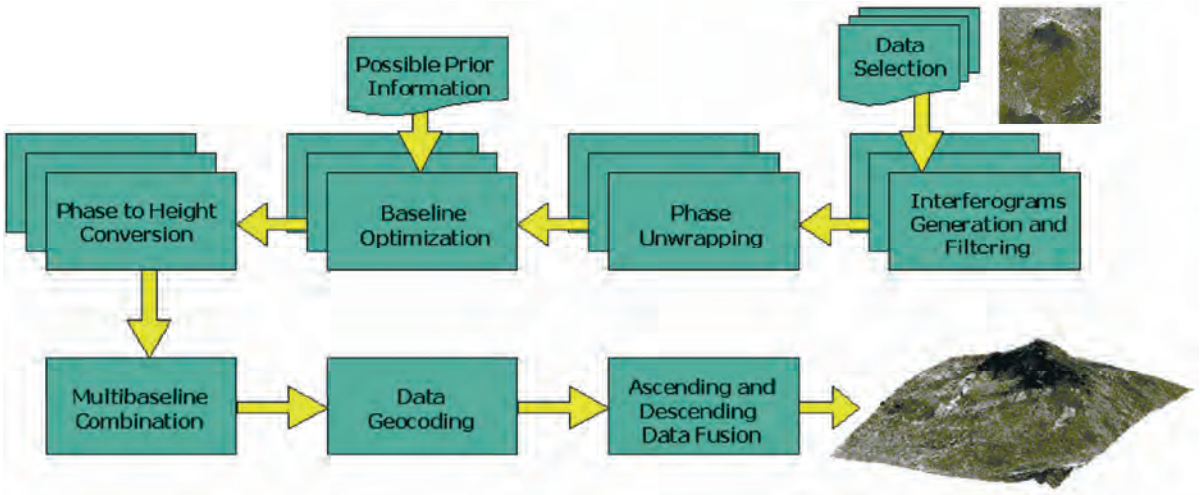


Figure 3-1: Block diagram of InSAR DEM reconstruction

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.

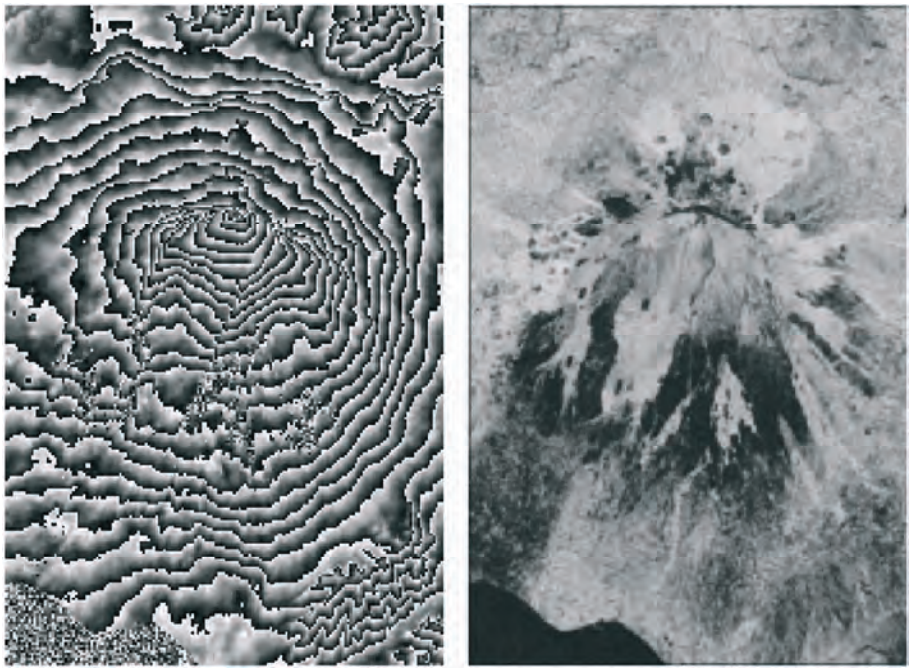


Figure 3-2: ERS Tandem data acquired over the Etna volcano (Sicily) on 1–2 August 1995 (the normal baseline was about 60 m). Left: flattened interferogram. Right: coherence map. Data have been filtered and down-sampled.

[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:

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

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).

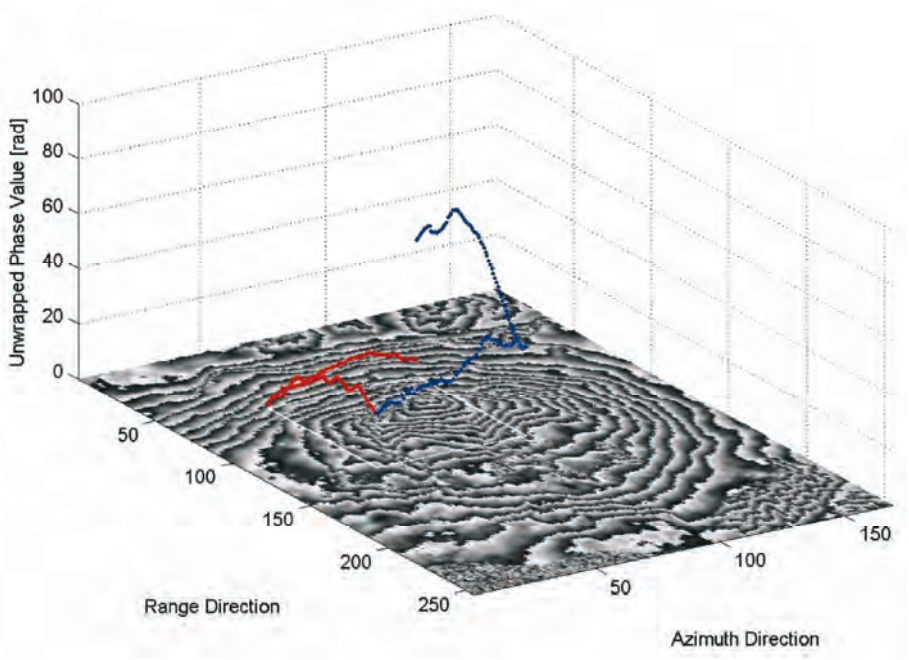


Figure 3-3: A visual example that shows the main problem related to phase inconsistencies: integration of the phase values yields different results depending on the path followed. In this example, the first and the last pixels of the integration paths are common, but one path crosses a layover area,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

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

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:

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

where $0 \le p \le 2$;
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 $\boldsymbol{A}\psi = \boldsymbol{b}$:

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

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: $\boldsymbol{G}^T\boldsymbol{G}\psi = \boldsymbol{G}^T\boldsymbol{d}$.

Data vector $\boldsymbol{d}$ is just the vectorised form of the wrapped phase differences estimated from the interferogram, while the model matrix $\boldsymbol{G}$ 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].

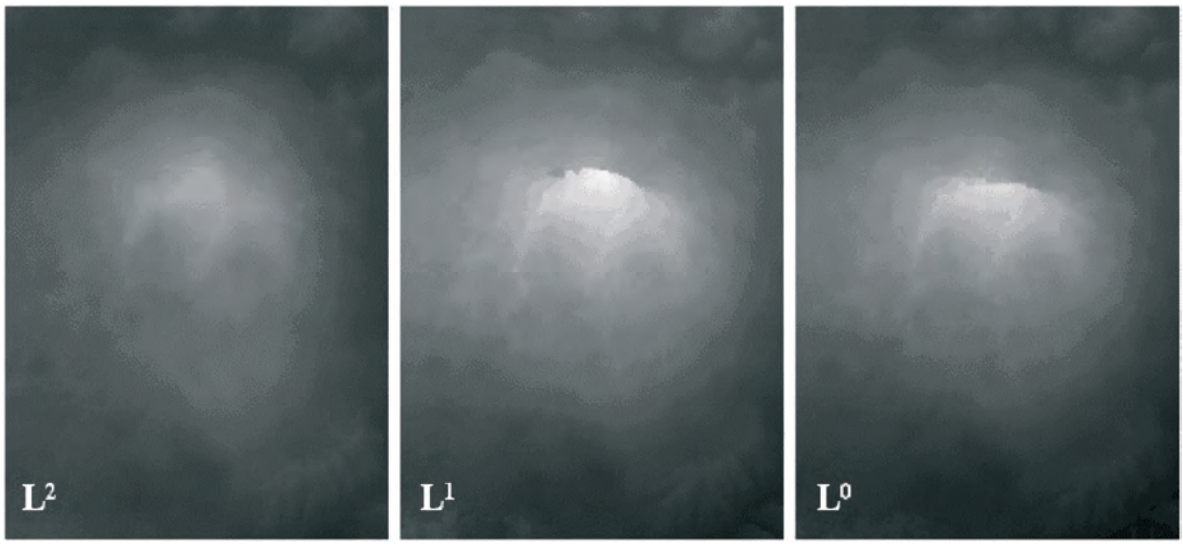


Figure 3-5: Three unwrapped phase values obtained by unwrapping the Etna interferogram shown in Figure 3-2 using the three different cost functions described in the text. The 'smoothing effect' typical of the ULMS approach can be seen.

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 x 100 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:

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

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

t is the azimuth time
vectors $\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}\}$ 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:

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

$$\left(\mathbf{P} - \mathbf{S}(t)\right) \cdot \mathbf{V_{PS}}(t) = 0 \qquad \text{Equation 3.7}$$

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:

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

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

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

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:

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

In general, $t_{slave} \neq t = t_{master}$. Thus a system of four equations and four unknowns (X, Y, Z, t_{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:

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

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

$$\boldsymbol{S}_{slave}(t_{slave}) \approx \boldsymbol{S}(t) + \boldsymbol{A} \cdot t + \boldsymbol{B} \qquad \text{Equation 3.12}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ 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 $\boldsymbol{A}$ and $\boldsymbol{B}$ (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:

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

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):

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

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:

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

Thus the *relative* topographic error equals the *relative* normal baseline error (at least to a first order approximation). For B_n = 100 m, ε_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 **ephemeredes** (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:

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

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.

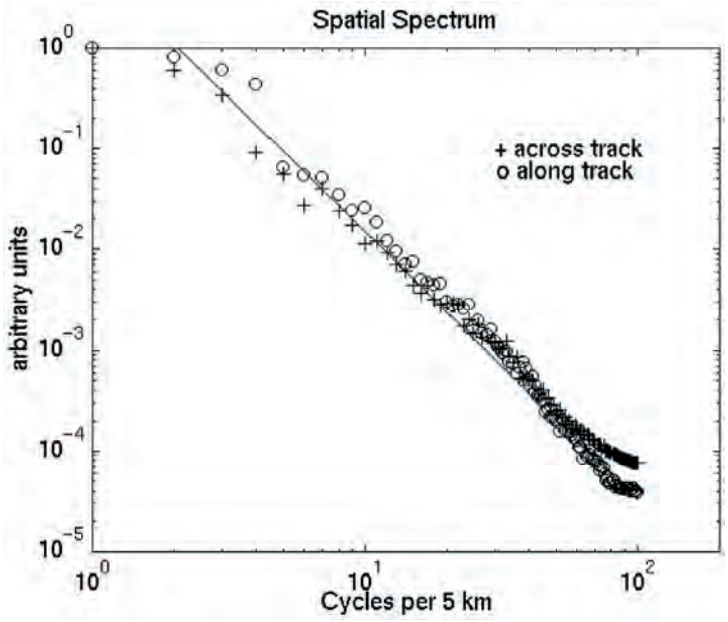


Figure 3-7: Power spectrum of the phase distortion due to atmospheric inhomogeneities (ERS Tandem – August 1995). The added line follows the slope to be expected from turbulence phenomena (from [Ferretti99]).

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):

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

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:

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

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:

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

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

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

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

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

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 byv 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).

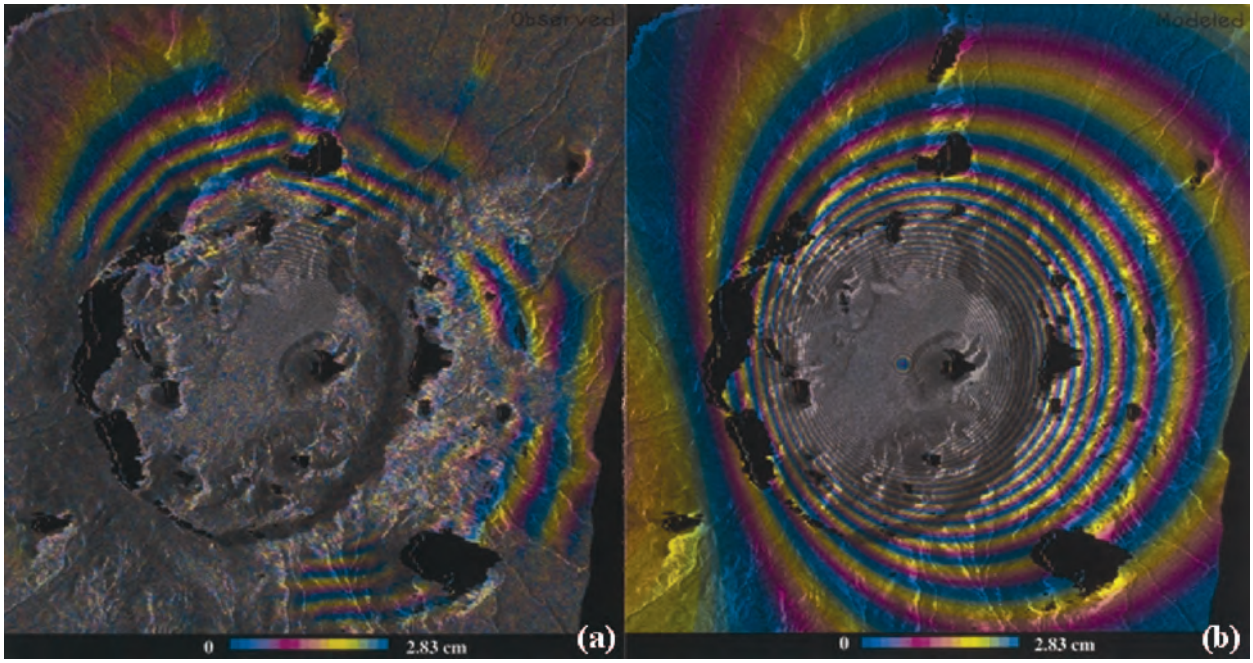


Figure 4-1: Full volcano cycle: inflation and deflation at Mount Okmok (from Z. Lu). This remarkable example of volcano monitoring includes a complete deformation cycle (pre-eruptive inflation, co-eruptive deflation and post-eruptive inflation). Mount Okmok, located in Alaska, is a good example for demonstrating 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.

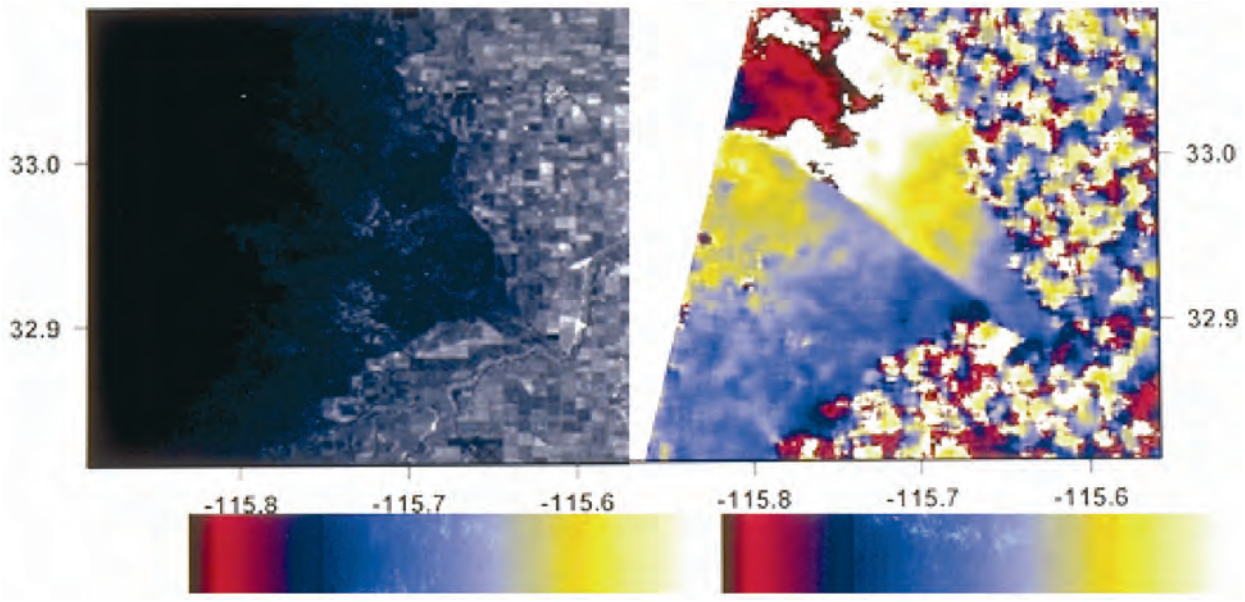


Figure 4-2: Example of fault slip detection: Superstition Hill. Two-year interferogram showing fault slip as a remote consequence of the Landers Earthquake. The slip is estimated as 10 mm. The sharp morphology of a fault crack as seen in an interferogram makes it very difficult to confuse with anything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

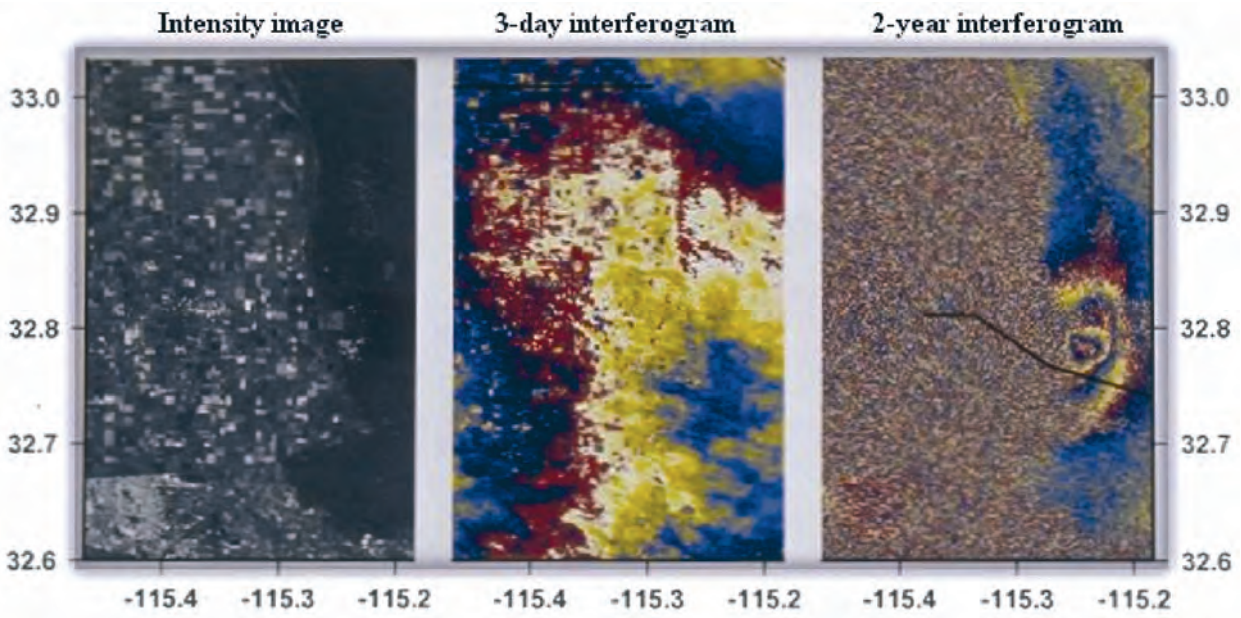


Figure 4-3: Example of industrial subsidence: the East Mesa Geothermal Plant

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 km^2, 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 warning. 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].

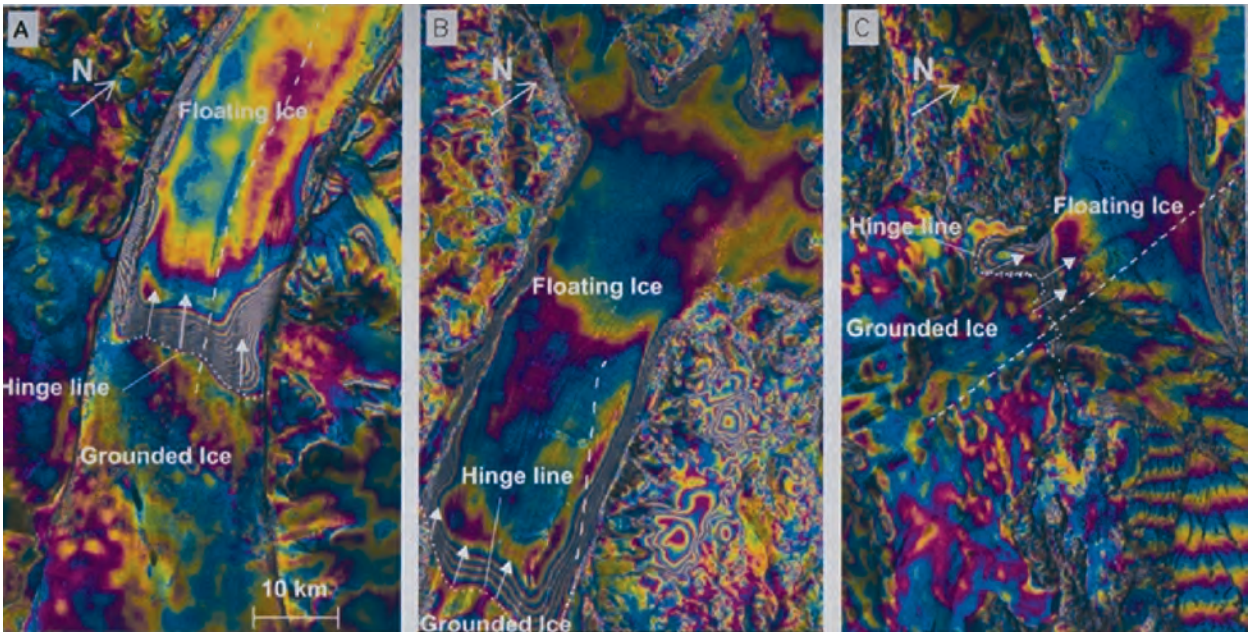


Figure 4-4: 'Tandem' ice flow mapping: Arctic glacier flowing [Rignot98]. (a) Petermann glacier, (b) Nioghalvfjerdsbrae glacier, and (c) Zacharia glacier.

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 ±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 ±6%. When the amplitude of the incoherent fraction is 10%, the standard deviation on the error is ±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.

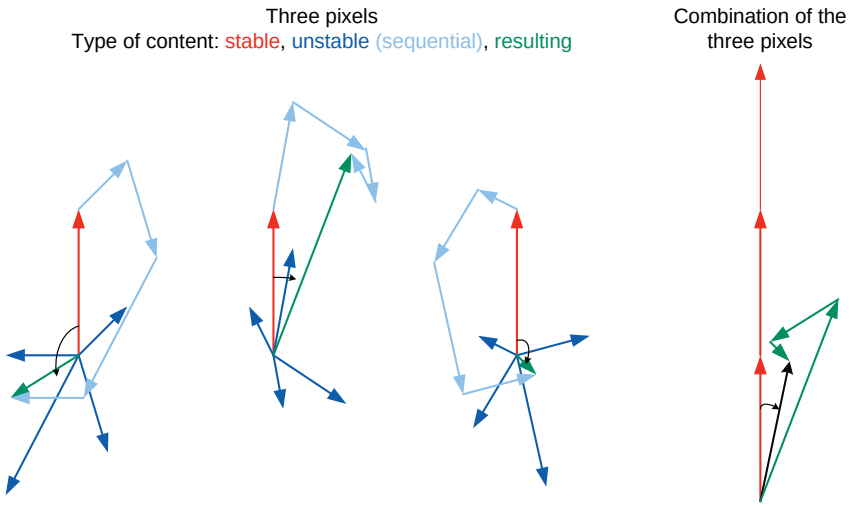


Figure 4-5: Combining coherent and incoherent signals

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.

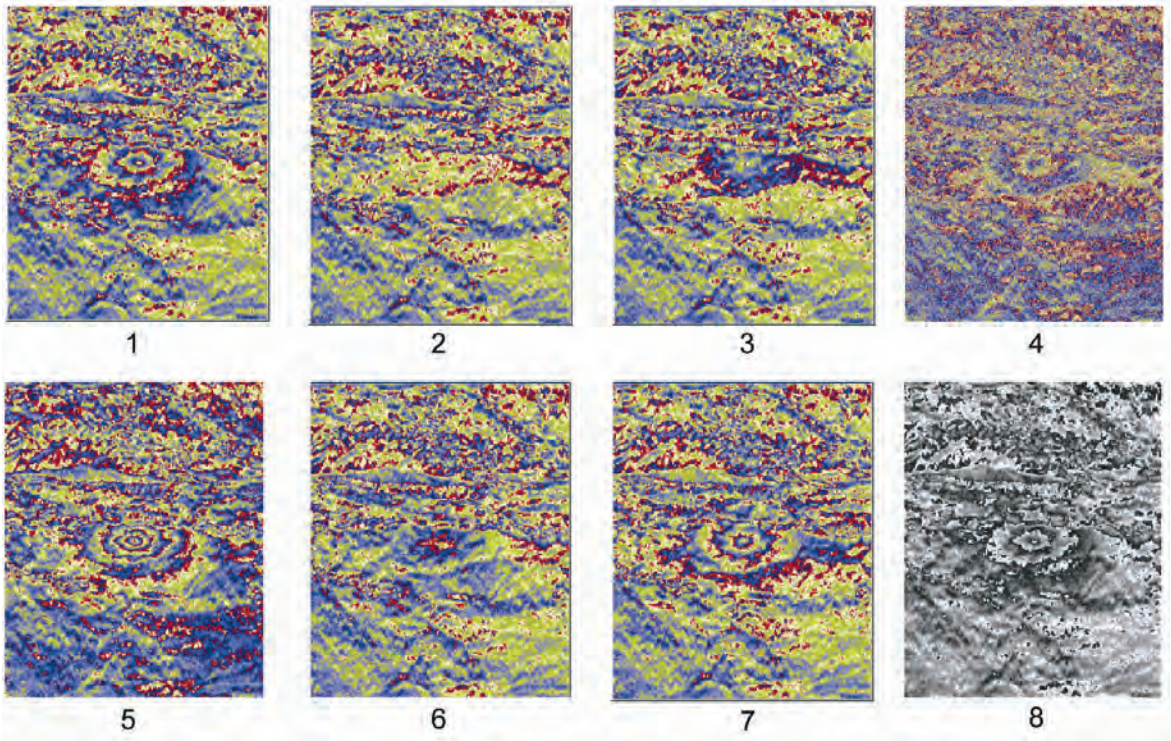


Figure 4-6: Series of images produced by superposing an elliptical signal (with a variable amplitude) and a background (extracted from a genuine interferogram)

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.

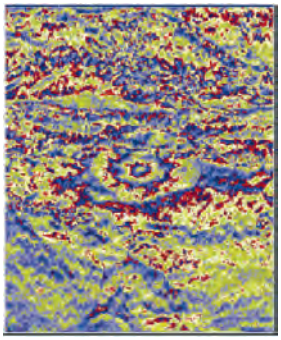


Image 7

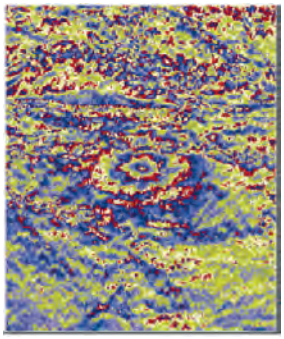


Image 1

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.

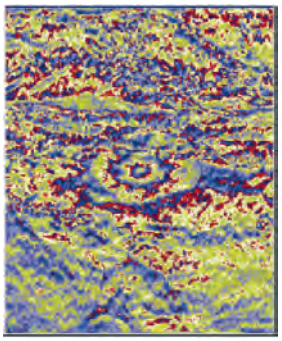


Image 7

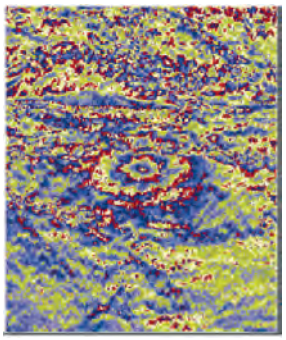


Image 1

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

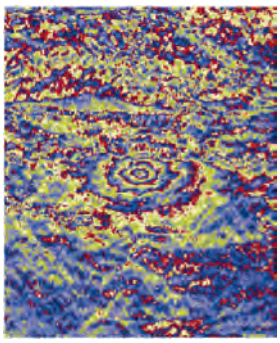


Image 5

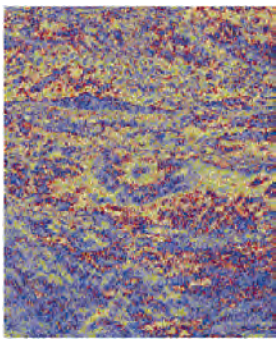


Image 4

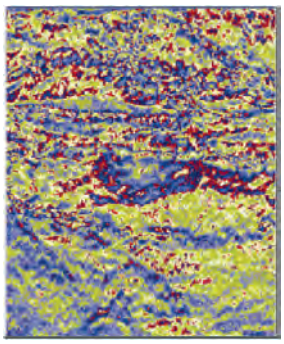


Image 3

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

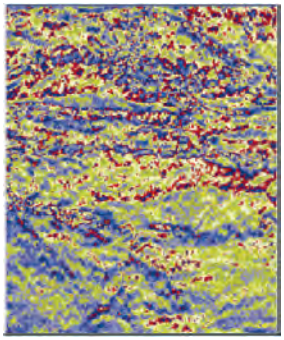


Image 2

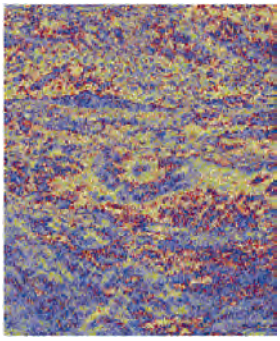


Image 4

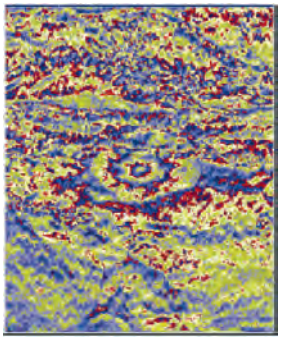


Image 7

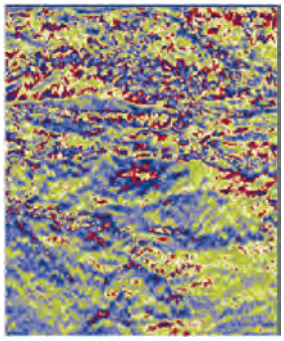


Image 6

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.

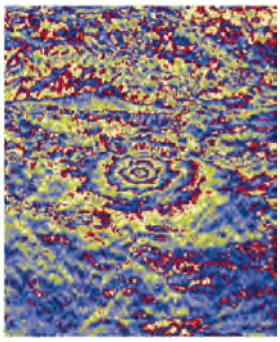


Image 5

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.