3.12 Interferometric Synthetic Aperture Radar Geodesy

M. Simons and P. A. Rosen, California Institute of Technology, Pasadena, CA, USA

## 3.12.1 Introduction

3.12.1.1 Motivation

In 1993, Goldstein *et al.* (1993) presented the first satellite-based interferometric synthetic aperture radar (InSAR) map showing large strains of the Earth's solid surface – in this case, the deforming surface was an ice stream in Antarctica. The same year, Massonnet *et al.* (1993) showed exquisitely detailed and spatially continuous maps of surface deformation associated with the 1992 M_w7.3 Landers earthquake in the Mojave Desert in southern California. These papers heralded a new era in geodetic science, whereby we can potentially measure three-dimensional (3-D) surface displacements with nearly complete spatial continuity, from a plethora of natural and human-induced phenomena. An incomplete list of targets to date includes all forms of deformation on or around faults (interseismic, aseismic, coseismic, and postseismic) aimed at constraining the rheological properties of the fault and surrounding crust, detection and quantification of changes in active magma chambers aimed at understanding a volcano's plumbing system, the mechanics of glaciers and temporal changes in glacier flow with obvious impacts on assessments of climate change, and the impact of seasonal and anthropogenic changes in aquifers. Beyond detection of coherent surface deformation, InSAR can also provide unique views of surface disruption, through measurements of interferometric decorrelation, which could potentially aid the ability of emergency responders to respond efficiently to many natural disasters.

That InSAR can take advantage of a satellite's perspective of the world permits one to view large areas of Earth's surface quickly and efficiently. In solid Earth geophysics, we are frequently interested in rare and extreme events (e.g., earthquakes, volcanic eruptions, and glacier surges). Therefore, if

we want to capture these events and their natural variability, we cannot simply rely on dense instrumentation of a few select areas; instead, we must embrace approaches that allow global access. Given easy access to data (which is not always the case), this inherently global perspective provided by satellite-based InSAR also allows one the luxury of going on geodetic fishing trips, whereby one essentially asks "I wonder if….?", in search of the unexpected (e.g., **Figure 1**). In essence we must not limit
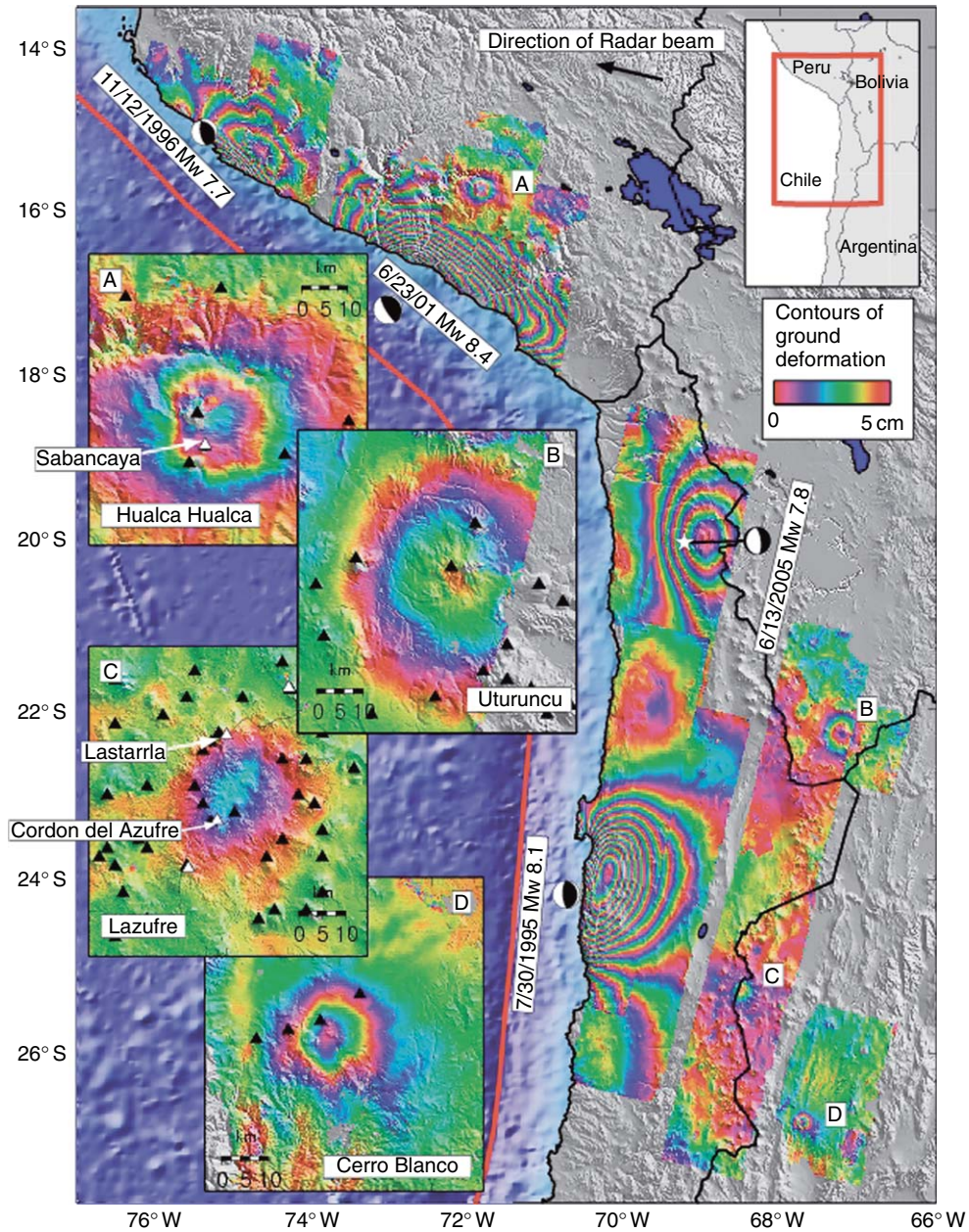
Figure 1 Mosaic of interferograms showing surface deformation associated with subsurface magma migration (inset panels A–D and associated tags in main image) and from three megathrust earthquakes and one intraslab earthquake all with M_w between 7.7 and 8.5. These interferograms are all from right-looking descending orbit C-band radars (*ERS*-1, *ERS*-2, and *ENVISAT*). The phase has been unwrapped from the natural fringe rate and rewrapped at 5 cm per fringe. The subduction trench is indicated by the red line and the centroid moment tensors by beachballs. Note the original survey for volcano deformation is a clear example of the 'fishing trip' approach enabled by InSAR. Figure courtesy of Matt Pritchard.

ourselves to hypothesis testing, but rather we must also tap the inherently exploratory power of InSAR.

3.12.1.2   History and Overview

Operating at microwave frequencies, synthetic aperture radar (SAR) systems provide unique images representing the electrical and geometrical properties of a surface in nearly all weather conditions. Since they provide their own illumination, SARs can image in daylight or at night. SAR mapping systems typically operate on airborne or spaceborne platforms following a linear flight path, as illustrated in **Figure 2**. Raw image data are collected by transmitting a series of coded pulses from an antenna illuminating a swath offset from the flight track. The echo of each pulse is recorded during a period of reception between the transmission events. When a number of pulses are collected, it is possible to perform 2-D matched-filter compression on a collection of pulse echoes to focus the image. This technique is known as SAR because in the along-track, or azimuth,

direction, a large virtual aperture is formed by coherently combining the collection of radar pulses received as the radar antenna moves along in its flight path (Raney, 1971). Although the typical physical length of a SAR antenna is on the order of meters, the synthesized aperture length can be on the order of kilometers. Because the image is acquired from a side-looking vantage point (to avoid left-side/right-side ambiguities), the radar image is geometrically distorted relative to the ground coordinates (**Figure 2**).

Figure 3 illustrates the InSAR system concept. By coherently combining the signals from two antennas, the interferometric phase difference between the received signals can be formed for each imaged point. In this scenario, the phase difference is essentially related to the geometric path length difference to the image point, which depends on the topography. With knowledge of the interferometer geometry, the phase difference can be converted into an altitude for each image point. In essence, the phase difference provides a third measurement, in
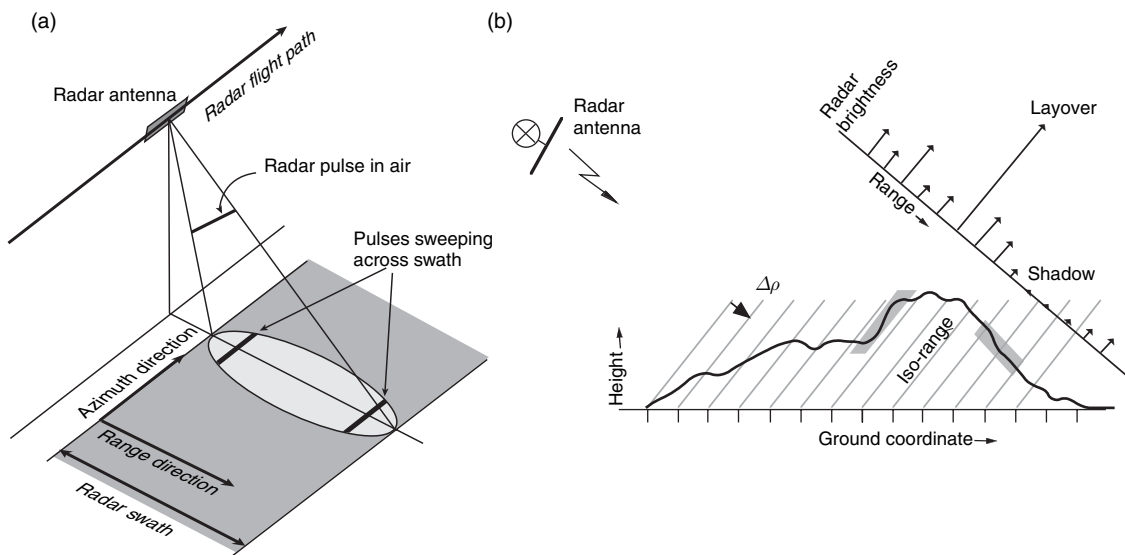
Figure 2   (a) Typical imaging scenario for a SAR system. The platform carrying the SAR instrument follows a curvilinear track known as the 'along-track' or 'azimuth' direction. The radar antenna points to the side, imaging the terrain below. The distance from the aperture to a target on the surface in the look direction is known as the the 'cross-track' or 'range direction' and is terrain dependent. The radar sends a pulse that sweeps through the antenna beam, effectively returning the integrated backscatter over the pulse and azimuth beam extent at any given instant. The azimuth extent can be many kilometers. Matched-filtering creates fine resolution in range. Synthetic aperture processing creates fine resolution in azimuth. (b) The 3-D world is collapsed into two dimensions in conventional SAR imaging. After image formation, the radar return is resolved into an image in range-azimuth coordinates. This panel shows a profile of the terrain at constant azimuth, with the radar flight track into the page. The profile is cut by curves of constant range, spaced by the range resolution of radar, defined as $\Delta\rho = c/2\Delta f_{BW}$, where c is the speed of light and Δf_{BW} is the range bandwidth of the radar. The backscattered energy from all surface scatterers within a range resolution element contribute to the radar return for that element.
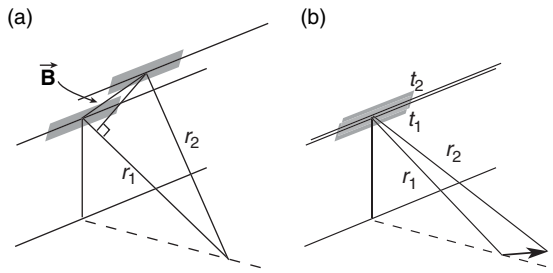
Figure 3 (a) InSAR for topographic mapping uses two apertures separated by a 'baseline', B, to image the surface. The phase difference between the apertures for each image point, along with the range and knowledge of the baseline, can be used to infer the precise shape of the imaging triangle to derive the topographic height of the image point. A range difference exists because the scene is viewed from two different vantage points. This is described by a shift in the point target response as presented in the text. (b) InSAR for deformation mapping uses the same aperture to image the surface at multiple times. A range difference is generated by a change in the position of the scene from one time to the next, imaged from the same vantage point. This range difference is described by a scene shift, not a point-target response shift, the mathematics is the same but for a sign change.

addition to the along- and cross-track location of the image point, or 'target', to allow a reconstruction of the 3-D location of the targets.

The InSAR approach for topographic mapping is similar in principle to the conventional stereoscopic approach. In stereoscopy, a pair of images of the terrain are obtained from two displaced imaging positions. The 'parallax' obtained from the displacement allows the retrieval of topography because targets at different heights are displaced relative to each other in the two images by an amount related to their altitudes (Rosen et al., 2000). The major difference between the InSAR technique and stereoscopy is that, for InSAR, the 'parallax' measurements between the SAR images are obtained by measuring the phase difference between the signals received by two InSAR antennas. These phase differences can be used to determine the angle of the target relative to the baseline of the interferometric SAR directly. The accuracy of the InSAR parallax measurement is typically several millimeters to centimeters, being a fraction of the SAR wavelength, whereas the parallax measurement accuracy of the stereoscopic approach is usually on the order of the resolution of the imagery (several meters or more).

Typically, the postspacing of the InSAR topographic data is comparable to the fine spatial resolution of SAR imagery, while the altitude

measurement accuracy generally exceeds stereoscopic accuracy at comparable resolutions. The registration of the two SAR images for the interferometric measurement, the retrieval of the interferometric phase difference, and subsequent conversion of the results into digital elevation models (DEMs) of the terrain can be highly automated, representing an intrinsic advantage of the InSAR approach. As discussed in later sections, the performance of InSAR systems is largely understood both theoretically and experimentally. These developments have led to airborne and spaceborne InSAR systems for routine topographic mapping.

The InSAR technique just described, using two apertures on a single platform, is often called 'cross-track interferometry' (XTI) in the literature. Other terms are 'single-track' and 'single-pass' interferometry (**Figure 3(a)**).

Another interferometric SAR technique was advanced by Goldstein and Zebker (1987) for measurement of surface motion by imaging the surface at multiple times (**Figure 3(b)**). The time separation between the imaging can be a fraction of a second to years. The multiple images can be thought of as 'time-lapse' imagery. A target movement will be detected by comparing the images. Unlike conventional schemes in which motion is detected only when the targets move more than a significant fraction of the resolution of the imagery, this technique measures the phase differences of the pixels in each pair of the multiple SAR images. If the flight path and imaging geometries of all the SAR observations are identical, any interferometric phase difference is due to changes over time of the SAR system clock, variable propagation delay, or surface motion in the direction of the radar line of sight (LOS).

In the first application of this technique described in the open literature, Goldstein and Zebker (1987) augmented a conventional airborne SAR system with an additional aperture, separated along the length of the aircraft fuselage from the conventional SAR antenna. Given an antenna separation of roughly 20 m and an aircraft speed of about $200 \, \text{m s}^{-1}$, the time between target observations made by the two antennas was about 100 ms. Over this time interval, clock drift and propagation delay variations are negligible. This system measured tidal motions in the San Francisco Bay area with an accuracy of several cm s^{-1} (Goldstein and Zebker, 1987). This technique has been dubbed 'along-track interferometry' (ATI) because of the arrangement of two antennas along the flight track on a single platform.

In the ideal case, there is no cross-track separation of the apertures, and therefore no sensitivity to topography.

ATI is merely a special case of space 'repeat-track interferometry' (RTI), which can be used to generate topography and motion. The orbits of several space-borne SAR satellites have been controlled in such a way that they nearly retrace themselves after several days. Aircraft can also be controlled to accurately repeat flight paths. If the repeat flight paths result in a cross-track separation and the surface has not changed between observations, then the repeat-track observation pair can act as an interferometer for topography measurement. For spaceborne systems, RTI is usually termed 'repeat-pass interferometry' in the literature (**Figure 3**).

If the flight track is repeated perfectly such that there is no cross-track separation, then there is no sensitivity to topography, and radial motions can be measured directly as with an ATI system. However, since the temporal separation between the observations is typically days to many months or years, the ability to detect small radial velocities is substantially better than the ATI system described above. The first demonstration of RTI for velocity mapping was a study of the Rutford ice stream in Antarctica (Goldstein *et al.*, 1993). The radar aboard the ERS-1 satellite obtained several SAR images of the ice stream with near-perfect retracing so that there was no topographic signature in the interferometric phase, permitting measurements of the ice stream flow velocity of the order of $1\,\mathrm{m\,yr^{-1}}$ (or $3 \times 10^{-8}\,\mathrm{m\,s^{-1}}$) observed over a few days (Goldstein *et al.*, 1993).

Most commonly for repeat-track observations, the track of the sensor does not repeat itself exactly, so the interferometric time-separated measurements generally comprise the signature of topography and of radial motion or surface displacement. The approach for reducing these data into velocity or surface displacement by removing topography is generally referred to as 'differential interferometric SAR.'

Goldstein *et al.* (1988) conducted the first proof-of-concept experiment for spaceborne InSAR using imagery obtained by the SeaSAT mission. In the latter portion of that mission, the spacecraft was placed into a near-repeat orbit every 3 days. Gabriel *et al.* (1989) used data obtained in an agricultural region in California, USA, to detect surface elevation changes in some of the agricultural fields of the order of several cm over approximately a 1-month period. By comparing the areas with the detected surface elevation changes with irrigation records, they concluded that these areas were irrigated in between the observations, causing small elevation changes from increased soil moisture. Gabriel *et al.* (1989) were actually looking for the deformation signature of a small earthquake, but the surface motion was too small to detect. These early studies were then followed by the aforementioned seminal applications to glacier flow and earthquake-induced surface deformation (Goldstein *et al.*, 1993; Massonnet *et al.*, 1993).

All civilian InSAR-capable satellites to date have been right-looking in near-polar sun-synchronous orbits. This gives the opportunity to observe a particular location on the Earth on both ascending and descending orbit passes (**Figure 4**). With a single satellite, it is therefore possible to obtain geodetic measurements from two different directions, allowing vector measurements to be constructed. The variety of available viewing geometries can be increased if a satellite has both left- and right-looking capability. Similarly, neighboring orbital tracks with overlapping beams at different incidence angles can also provide diversity of viewing geometry.

In an ideal mission scenario (see Section 3.12.5), observations from a given viewing geometry will be acquired frequently and for a long period of time to provide a dense archive for InSAR analysis. The frequency of imaging is key in order to provide optimal time resolution of a given phenomena, as well as to provide the ability to combine multiple images to detect small signals. Of course, many processes of interest are not predictable in time, thus we must continuously image the Earth in a systematic fashion in order to provide recent 'before' images. For a given target, not all acquisitions are necessarily viable for InSAR purposes. The greatest nemesis for InSAR geodesy comes from incoherent phase returns between two image acquisitions. This incoherence can be driven by observing geometry (i.e., the baseline is too large) or by physical changes of the Earth's surface (e.g., snowfall). Thus any InSAR study begins with an assessment of the available image archive. **Figure 5** uses an example from the ERS-1 and ERS-2 image archive to illustrate how one would go about choosing images for InSAR processing, assuming you wanted to make all available pairs that were not decorrelated due to large baselines, snow, or temporal separation that was too large. In theory, a future mission would have sufficiently tight control on the satellite orbit such that baseline selection would not be an issue.
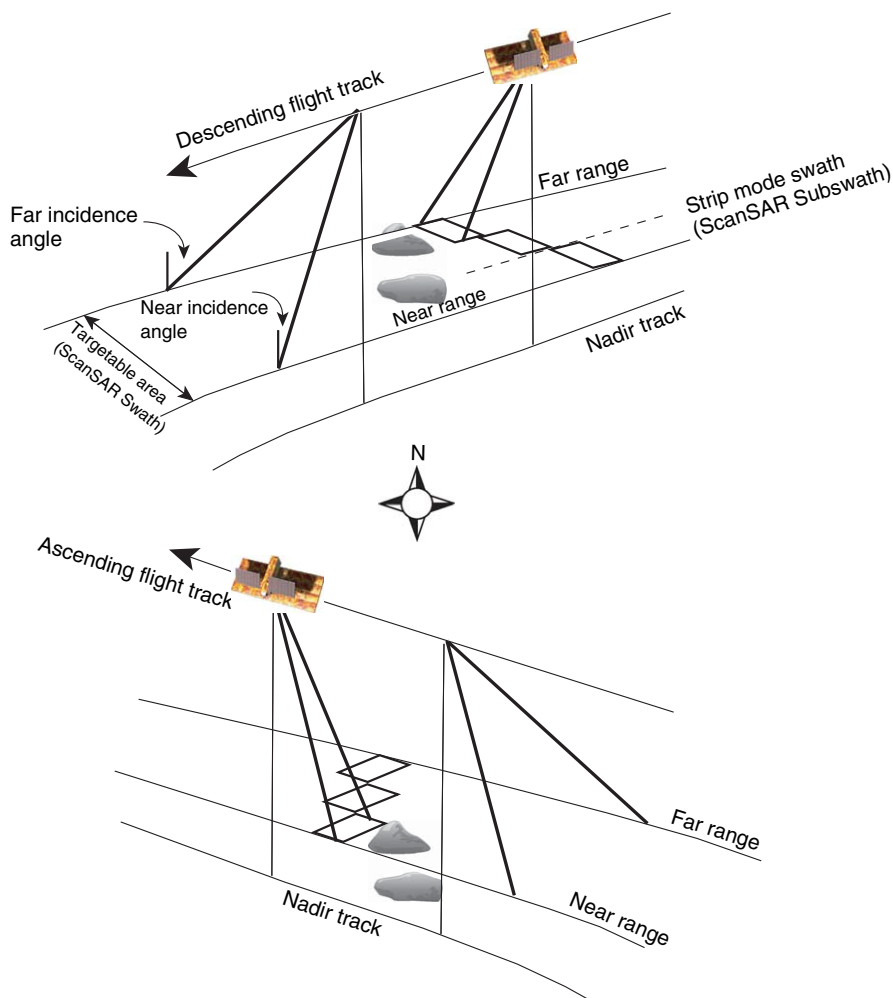
Figure 4 A rendition of ascending and descending tracks for a single right-looking radar system. On successive repeat tracks, geodetic measurements can be made from two different line-of-sight directions, giving possible vector measurements over time. Systems that are capable of imaging in a continuous strip can operate in strip mode with a swath width constrained by the details of the radar system. Systems that can electronically steer their antenna in the cross-track direction are usually capable of ScanSAR operation, and can image a much wider swath, as illustrated.

3.12.1.3   Scope

Today, spaceborne InSAR enjoys widespread application, in large part because of the availability of suitable globally acquired SAR data from the ERS-1, ERS-2, and ENVISAT satellites operated by the European Space Agency, JERS-1 and ALOS satellites operated by the National Space Development Agency of Japan, RADARSAT-1 operated by the Canadian Space Agency, and SIR-C/X-SAR operated by the United States, German, and Italian space agencies. As more and more radar data become available from international civilian radar satellites, and as scientific demands become greater on the use of these data, including extraction of ever more subtle and well-

calibrated geophysical signals, it is essential to understand the characteristics of the image, how they are processed, and how that processing can affect the interpretation of the image data.

There exist both commercial and freely available software for conventional InSAR processing. While they may differ in detail, they must all follow a basic processing flow. Figure 6 presents such a flow, derived from the authors' experience in developing the Repeat Orbit Interferometry Package (ROI_PAC) software suite (Rosen *et al.*, 2004). This flow diagram explicitly calls out potential iterative cycles and use of external data and intermediary models. Major phases of this processing are described
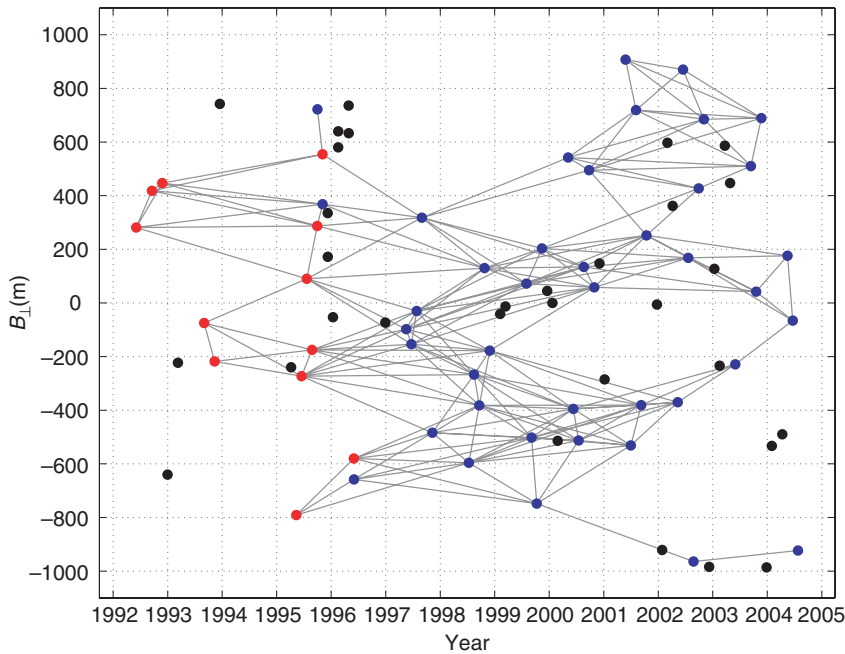
Figure 5 A typical perpendicular component of baseline, $B_\perp$, vs time plot used to select data for InSAR purposes. Here, each dot represents a different acquisition of the same scene by the ERS-1 or ERS-2 satellites. This example is for track 485 and frame 2853 corresponding to the Long Valley Caldera region of California. Lines connecting the dots represents interferograms that are likely to work. We have chosen to indicate only pairs that satisfy the constraints: $B_\perp < 250\,$m, neither scene occurs in winter when snowfall and rain cause decorrelation, and $\Delta T < 3.5\,$years. For viable scenes, red and blue dots indicate ERS-1 and ERS-2, respectively.

in the text. Given quality data and metadata, an initial complete processing from raw data to a georeferenced deformation image can now be done automatically and quickly (in under a few hours) on an average laptop computer. Subsequent iterative refinements to maximize the quality and quantity of the observations can require significantly more effort.

In this chapter, we aim to provide a review of the basic theory of InSAR for geodetic applications. Numerous review articles and books on the topic of InSAR already exist (e.g., Massonnet and Feigl, 1995; Rosen *et al.*, 2000; Burgmann *et al.*, 2000; Hanssen, 2001), and our goal is not to repeat these works more than necessary. Instead, we attempt to provide an overview of what data, processing, and analysis schemes are currently used and a glimpse of what the future may hold. As part of this discussion, we present our biased view of what constitutes best practices for use of InSAR observations in geodetic modeling. Finally, we provide a basic primer on the ties between different mission design parameters and their relationship to the character of the resulting observations. In general, this review borrows heavily

from our previous work with many colleagues, and where appropriate, we point the reader to the original sources for a more complete discussion. Much of the SAR processing discussion is derived and simplified from Rosen *et al.* (2000), although here, this discussion is augmented to include a variety of more recent techniques including persistent scatterers, ScanSAR interferometry, and pixel tracking.

3.12.2 InSAR

Interferometry relies on the constructive and destructive interference of electromagnetic waves from sources at two or more vantage points to infer something about the sources or the relative path length of the interferometer. For InSAR, the interference pattern is constructed from two complex-valued synthetic aperture radar images, and interferometry is the study of the phase difference between two images – acquired from different vantage points, different times, or both.
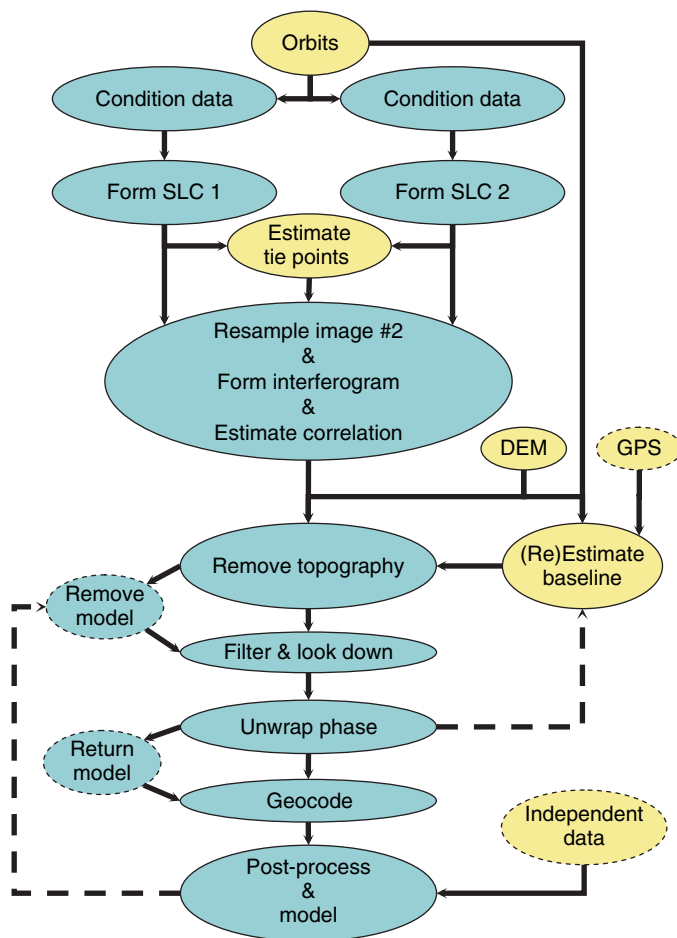
Figure 6 Representative differential InSAR processing flow diagram. Blue bubbles represent image output, yellow ellipses represent nonimage data. Flow is generally down the solid paths, with optional dashed paths indicating potential iteration steps. DEM, digital elevation model; SLC, single look complex image.

Appendix 1 describes the SAR technique, developing a model for the image one would obtain from an idealized surface that consists entirely of a single reflective point, known as a point target, then further considering the image effects of natural surfaces. Fine image resolution is achieved in the cross-track, or range, direction by transmitting a coded waveform with sufficient bandwidth. Matched-filter compression of each received signal pulse then recovers the range resolution. In the along-track, or azimuth, direction, a SAR forms a large synthetic aperture by coherently combining an ensemble of the radar pulses received as the SAR moves along in its flight path. Matched filtering then focuses the image in azimuth.

In Appendix 1, it is shown that there is a phase term $\exp\{-j2kr\}$, where $k = 2\pi/\lambda$ is the wave number and λ is the radar wavelength, that characterizes the two-way propagation distance, $2r$, from the radar sensor to the point target and back again. For a general surface, there is an additional phase term contributed by each surface scatterer. The net phase of each image point is the sum of these two terms: the intrinsic phase of the surface, which tends to be random, and the propagation phase term.

A resolution element can be represented as a complex phasor of the coherent backscatter from the scattering elements on the ground and the propagation phase delay, as illustrated in Figure 7. The backscatter phase delay is the net phase of the coherent sum of the contributions from all elemental scatterers in the resolution element, each with their individual backscatter phases and their differential path delays relative to a reference surface normal to the radar look direction.

Radar images observed from two nearby antenna locations have resolution elements with
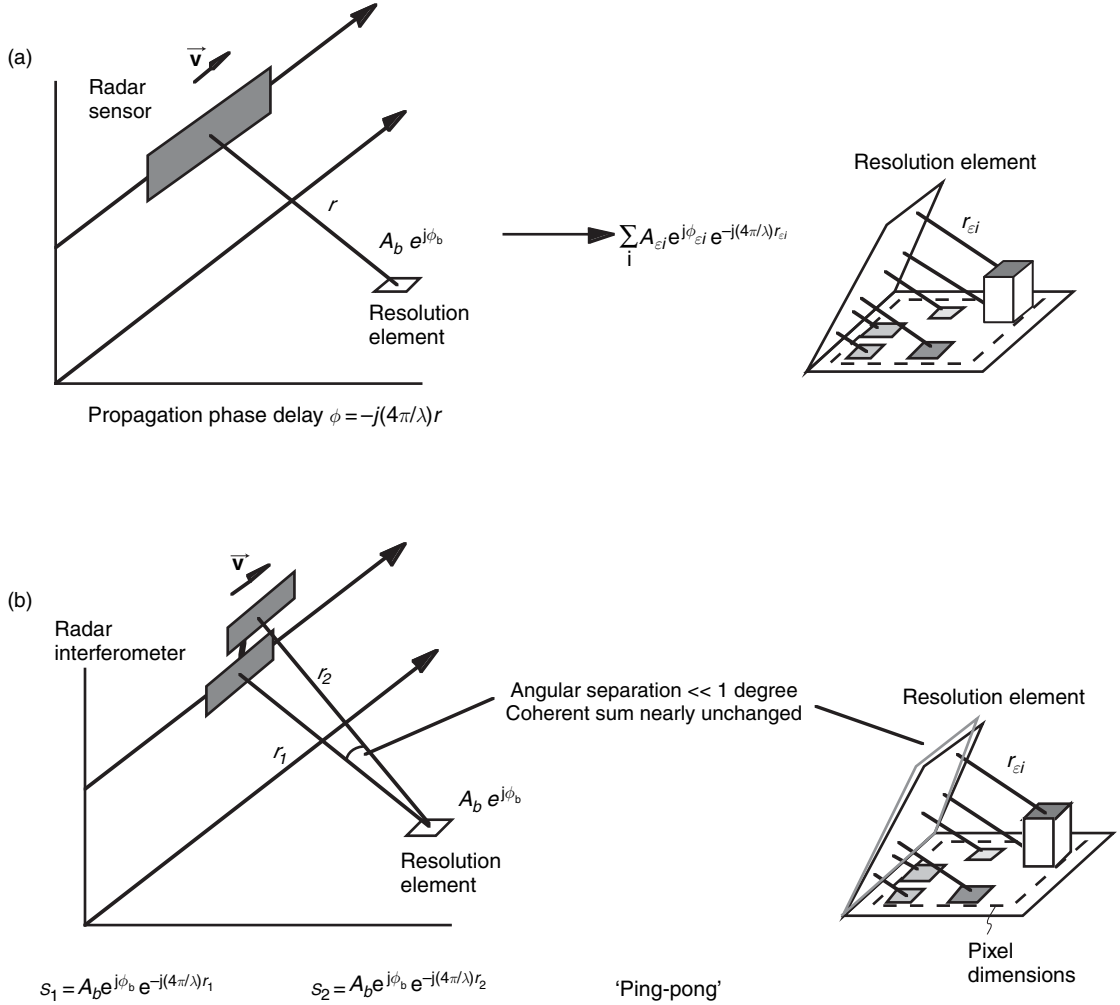
Figure 7 Illustration of the elements of phase in SAR and InSAR observations. (a) Each resolution element in a radar image has a complex value. Its amplitude, A_b, is related to the electrical reflectivity of the surface and its roughness. Its phase is the sum of the propagation phase delay, $-j(4\pi/\lambda)r$, and the surface backscatter phase, ϕ_b. The surface contribution, $A_b e^{j\phi_b}$, can be modeled as the coherent sum of contributions from elemental complex scatterers, $A_{\varepsilon i}e^{j\phi_{\varepsilon i}}$, each with their individual path delays, $r_{\varepsilon i}$, relative to a common reference. Since the arrangement of scatterers within a resolution element is generally random, the phase ϕ_b is also random from element to element. (b) For InSAR, when two observations are made from nearly the same viewing geometry, the backscatter values, $A_b e^{j\phi_b}$, are nearly the same in each, and the phase difference, $\phi_1 - \phi_2$, is essentially the difference in the path delay $-j(4\pi/\lambda)(r_1 - r_2)$.

nearly the same complex phasor return, but with a different propagation phase delay. In interferometry, the complex phasor information of one image is multiplied by the complex conjugate phasor information of the second image to form an interferogram, effectively canceling the common backscatter phase in each resolution element, but leaving a phase term proportional to the differential path delay. Ignoring the slight difference in backscatter phase of the surface observed from two different vantage points treats each resolution element as a point scatterer.

3.12.2.1 The Interferogram

As shown in Appendix 1, for a fixed point target and a platform moving to synthesize an aperture in azimuth, the range and azimuth-compressed point-target signal, $r_{zcc\delta}$ is

$$r_{zcc\delta}(x', r'; x_0, R_0) = e^{-j4\pi R_0/\lambda}\,\mathrm{sinc}\left(\frac{\pi}{\Delta R}(r' - R_0)\right)$$
$$\times \mathrm{sinc}\left(\frac{\pi}{\Delta X}(x' - x_0)\right) \quad [1]$$

where we have explicitly called out the dependence on the location of the point target in the definition of

the point-target response. As described in Appendix 1, x_0 represents the location of the fixed point target in the along-track coordinate direction, R_0 represents the 'closest approach' range from the platform to the target at that x_0, ΔR is the range resolution after range matched-filtering, and ΔX is the along-track resolution after synthetic aperture processing. The subscript z indicates that the function is complex valued, cc indicates both range and azimuth compression have been applied, and δ indicates this is a point-target, or delta-function response, for example a small bright reflecting object surrounded by a surface that reflects on energy back to the radar. For a general complex scene $\Gamma(x, r)$, the SAR image after compression is given by convolution of Γ with the point-target response

$$\Gamma_{zcc}(x', r') = \iint \Gamma(x, r) r_{zcc\delta}(x', r'; x, r) dx dr \qquad [2]$$

$$= \int dx \int dr\, \Gamma(x, r) e^{-j4\pi r/\lambda}$$
$$\times \operatorname{sinc}\left(\frac{\pi}{\Delta R}(r' - r)\right) \operatorname{sinc}\left(\frac{\pi}{\Delta X}(x' - x)\right) \quad [3]$$

which can be verified by substituting $\Gamma(x, r) = \delta(x - x_0, r - R_0)$ to recover the impulse response $r_{zcc\delta}$.

The SAR image estimate Γ_{zcc} is the convolution of the actual reflectivity with a 2-D function similar to a delta function but with finite width (the sinc function). This convolution smears out the intrinsic reflectivity as a point-spread function in optics. If the system bandwidths were to become infinite, the sinc functions would become delta functions

$$\operatorname{sinc}\left(\frac{\pi}{\Delta R}(r' - r)\right) \to \delta(r' - r) \quad \text{as} \quad \Delta R \to 0 \quad [4]$$

$$\operatorname{sinc}\left(\frac{\pi}{\Delta X}(x' - x)\right) \to \delta(x' - r) \quad \text{as} \quad \Delta X \to 0 \quad [5]$$

such that

$$\Gamma_{zcc}(x', r') \to \Gamma(x', r') e^{-j4\pi r'/\lambda} \qquad [6]$$
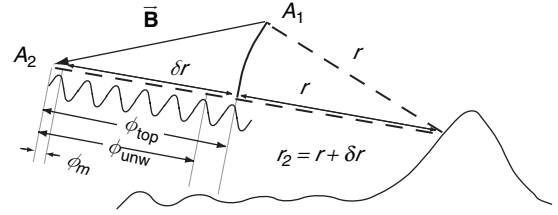
Now consider two observations of the reflectivity acquired from slightly different ranges (**Figures 7 and 8**).

For the original observation at r

$$\Gamma_{zcc,1}(x', r') = \int dx \int dr\, \Gamma(x, r) e^{-j4\pi r/\lambda}$$
$$\times \operatorname{sinc}\left(\frac{\pi}{\Delta R}(r' - r)\right) \operatorname{sinc}\left(\frac{\pi}{\Delta X}(x' - x)\right) \quad [7]$$

For an observation at $r + \delta r$, the point-target response is now shifted, but the scene must still be referenced to the original range r.

$$r_{zcc\delta}(x', r'x, r + \delta r) = e^{-j4\pi(r + \delta r)/\lambda} \operatorname{sinc}\left(\frac{\pi}{\Delta R}(r' - (r + \delta r))\right)$$
$$\times \operatorname{sinc}\left(\frac{\pi}{\Delta X}(x' - x_0)\right) \qquad [8]$$

Figure 8 Phase in interferogram depicted as cycles of electromagnetic wave propagating a differential distance δr. Phase in the interferogram is initially known modulo 2π: $\phi_m = W(\phi_{top})$, where ϕ_{top} is the topographically induced phase and $W()$ is an operator that wraps phase values into the range $-\pi < \phi \geq \pi$. After unwrapping, relative phase measurements between all pixels in the interferogram are determined up to a constant multiple of 2π: $\phi_{unw} = \phi_m + 2\pi k_{unw}$, where k_{unw} is a spatially variable integer dependent on the pixel coordinates of the interferogram. Absolute phase determination is the process to determine the overall multiple of $2\pi k_{abs}$ that must be added to the phase measurements so that it is proportional to the range difference. The reconstructed phase is then $\phi_{top} = \phi_m + 2\pi k_{unw} + 2\pi k_{abs}$.

$$\Gamma_{zcc,2}(x', r') = \int dx \int dr\, \Gamma(x, r) e^{-j4\pi(r + \delta r)/\lambda}$$
$$\times \operatorname{sinc}\left(\frac{\pi}{\Delta R}(r' - (r + \delta r))\right)$$
$$\times \operatorname{sinc}\left(\frac{\pi}{\Delta X}(x' - x)\right) \qquad [9]$$

Again letting the bandwidths tend to infinity, we get

$$\Gamma_{zcc,1}(x', r') \to \Gamma(x', r') e^{-j4\pi r'/\lambda} \qquad [10]$$

$$\Gamma_{zcc,2}(x', r') \to \Gamma(x', r' - \delta r) e^{-j4\pi r'/\lambda} \qquad [11]$$

Note a shift in the registration of the true scene reflectivity.

$$\Gamma_{zcc,2}(x', r' + \delta r) \to \Gamma(x', r') e^{-j4\pi(r' + \delta r)/\lambda} \qquad [12]$$

When the true scene reflectivity is aligned, there is a phase difference between the reconstructed phase proportional to δr, a geometric term. The true scene reflectivity is a complex number with random phase. However, after we have aligned the scene reflectivity by shifting $\Gamma_{zcc,2}(x', r')$ by δr, or equivalently by looking up the value of $\Gamma_{zcc,2}$ at location $r' + \delta r$, then the scene phase is common to both observations. We can form the product

$$I(x', r') = \Gamma_{zcc,1}(x', r') \Gamma_{zcc,2}^*(x', r' + \delta r) \qquad [13]$$

$$= |\Gamma(x', r')|^2 e^{-j4\pi(\delta r)/\lambda} \qquad [14]$$

The function $I(x', r')$ is the 'interferogram', a complex quantity, the phase of which is just a geometric term related to the range difference δr (when bandwidths are infinite) between the two images. The

range difference can be caused by a vantage point difference as described here, and illustrated in **Figure 3(a)**, or by a shift in the scene location (**Figure 3(b)**) or a combination of the two. If the scene shifts rather than the point target, then the sign of the range difference changes, but the form of the interferogram is the same.

Consider the case where the range difference δr arises from a cross-track separation of two observation points, as illustrated in **Figure 3(a)**.

The phase of the interferogram eqn [14] is the difference in the geometric path length phases of the two images

$$\phi_I = \phi_1 - \phi_2 = \frac{4\pi}{\lambda}(r_2 - r_1) = \frac{4\pi}{\lambda}\delta r \qquad [15]$$

There is a clear dependence on the relative lengths of the two sides of the triangle on the height of the surface, which in general is not known *a priori*. Thus δr is not known exactly to align the reflectivities to form the interferogram. However, in practical systems, one can match the reflectivity estimates in the two SAR observations to within sufficient accuracy (generally much better than the image resolution) to derive a sufficient estimate of δr for alignment. Once formed, the interferogram for the cross-track interferometer then contains a record of the variability of the height of the surface. It is possible to invert the phase to reconstruct the height. It turns out it can be done quickly and efficiently, and is a powerful tool for topographic mapping. Note that the sign of the propagation phase delay is set by the desire for consistency between the Doppler frequency, f_D, and the phase history, $\varphi(t)$ (Rosen *et al.*, 2000).

Only the principal values of the phase, modulo 2π, can be measured from the complex-valued resolution element. The total range difference between the two observation points that the phase represents in general can be many multiples of the radar wavelength, or, expressed in terms of phase, many multiples of 2π. The typical approach for determining the unique phase that is directly proportional to the range difference is to first determine the relative phase between pixels via the so-called 'phase-unwrapping' process. This connected phase field will then be adjusted by an overall constant multiple of 2π. The second step determines this required multiple of 2π, and is referred to as 'absolute phase determination.' **Figure 8** shows the principal value of the phase, the unwrapped phase, and absolute phase for a pixel.

### 3.12.2.2   Interferometric Baseline and Height Reconstruction

In order to generate topographic maps or data for other geophysical applications using radar interferometry, we must relate the interferometric phase and other known or measurable parameters to the topographic height. It is also desirable to derive the sensitivity of the interferometrically determined topographic measurements to the interferometric phase and other known parameters. In addition, interferometric observations have certain geometric constraints that preclude valid observations for all possible image geometries.

The interferometric phase as previously defined is proportional to the range difference from two antenna locations to a point on the surface. This range difference can be expressed in terms of the vector separating the two antenna locations, called the interferometric baseline. The range and azimuth position of the sensor associated with imaging a given scatterer depends on the portion of the synthetic aperture used to process the image (see Appendix 1). Therefore the interferometric baseline depends on the processing parameters, and is defined as the difference between the location of the two antenna phase center vectors at the time when a given scatterer is imaged.

The equation relating the scatterer position vector, $\mathbf{T}$, a reference position for the platform $\mathbf{P}$, and the look vector, $\mathbf{l}$, is

$$\mathbf{T} = \mathbf{P} + \mathbf{l} = \mathbf{P} + r\hat{l} \qquad [16]$$

where r is the range to the scatterer and $\hat{l}$ is the unit vector in the direction of $\mathbf{l}$ (**Figure 9**). The position $\mathbf{P}$ can be chosen arbitrarily, but is usually taken as the position of one of the interferometer antennas. Interferometric height reconstruction is the
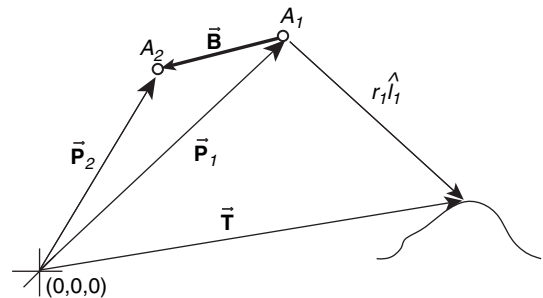
Figure 9   Vectors describing the relationship between the phase centers of the radar antennas defining the interferometer and the surface location, as described in the text.

determination of a target's position vector from known platform ephemeris information, baseline information and the interferometric phase. Assuming $\mathbf{P}$ and r are known, interferometric height reconstruction amounts to the determination of the unit vector $\hat{l}$ from the interferometric phase. Letting $\mathbf{B}$ denote the baseline vector from antenna 1 to antenna 2, setting $\mathbf{P} = \mathbf{P}_1$ and defining

$$\mathbf{B} = \mathbf{P}_2 - \mathbf{P}_1 \quad B = |\mathbf{B}| \equiv \langle \mathbf{B}, \mathbf{B} \rangle^{1/2} \qquad [17]$$

we have the following expression for the interferometric phase

$$\phi = \frac{2\pi p}{\lambda}(r_2 - r_1) = \frac{2\pi p}{\lambda}(|\mathbf{l}_2| - |\mathbf{l}_1|) \qquad [18]$$

$$= \frac{2\pi p}{\lambda} r_1 \left[\left(1 - \frac{2\langle \hat{l}_1, \mathbf{B} \rangle}{r_1} + \left(\frac{B}{r_1} \right)^2 \right)^{1/2} - 1 \right] \qquad [19]$$

where $p = 2$ for repeat-track systems and $p = 1$ for two-aperture systems with a single transmitter and two receivers (Rosen *et al.*, 2000), and the subscripts refer to the antenna number. This expression can be simplified assuming $B \ll r$ by Taylor-expanding eqn [19] to first order to give

$$\phi \approx -\frac{2\pi p}{\lambda} \langle \hat{l}_1, \mathbf{B} \rangle \qquad [20]$$

illustrating that the phase is approximately proportional to the projection of the baseline vector on the look direction (Zebker and Goldstein, 1986).

When the baseline lies entirely in the plane of the look vector and the nadir direction, we have $\mathbf{B} = (B \cos(\alpha), B \sin(\alpha))$, where α is the angle the baseline makes with respect to a reference horizontal plane. Then, eqn [20] can be rewritten as

$$\phi = -\frac{2\pi p}{\lambda} B \sin(\theta - \alpha) \qquad [21]$$

where θ is the look angle, the angle the LOS vector makes with respect to nadir, shown in **Figure 10**.

The intrinsic fringe frequency in the slant plane interferogram is given by

$$\frac{\partial \phi}{\partial r} = \frac{2\pi p}{\lambda} B \cos(\theta - \alpha)$$
$$\times \frac{1}{r \sin \theta} \left[-\frac{r}{b_p + r_e} + \cos \theta + \frac{b_0 + r_e \sin \tau_c}{b_p + r_e \sin(i - \tau_c)} \right]$$
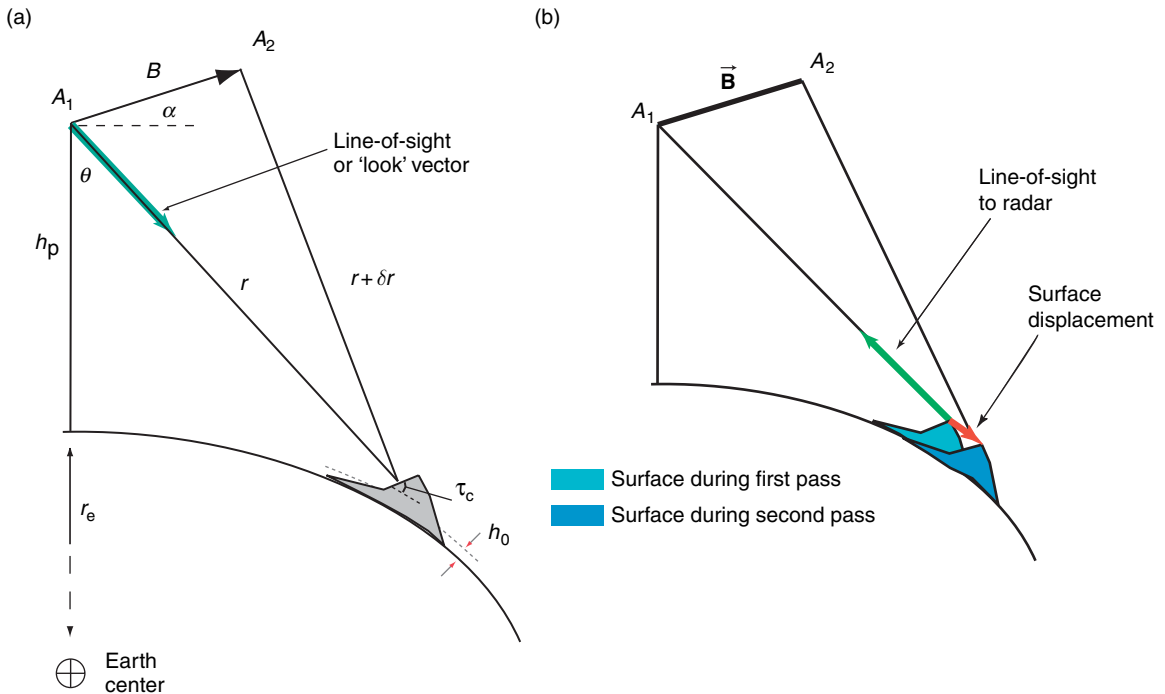$$[22]$$

Figure 10 SAR interferometry imaging geometry in the plane normal to the flight direction for topography (a) and deformation (b) mapping.

where

$$\sin i = \frac{h_{\mathrm{p}} + r_{\mathrm{e}}}{h_0 + r_{\mathrm{e}}} \sin \theta \qquad [23]$$

and i is the local incidence angle relative to a spherical surface, h_{p} is the height of the platform, and τ_{c} is the surface slope angle in the cross-track direction as defined in **Figure 10** at left. From eqn [22], the fringe frequency is proportional to the perpendicular component of the baseline: $B_{\perp} = B \cos(\theta - \alpha)$. As $B_{\perp}$ increases or as the local terrain slope approaches the look angle, the fringe frequency increases. Also from eqn [22], the fringe frequency is inversely proportional to λ, thus longer wavelengths result in lower fringe frequencies. If the phase changes by 2π or more across the range resolution element, Δr, the different contributions within the resolution cell do not add to a well-defined phase, resulting in what is commonly referred to as decorrelation of the interferometric signal. Thus, in interferometry, an important parameter is the critical baseline, defined as the perpendicular baseline at which the phase rate reaches 2π per range resolution element. From eqn [22], the critical baseline satisfies the proportionality relationship

$$B_{\perp, \mathrm{crit}} \propto \frac{\lambda}{\Delta r} \qquad [24]$$

This is a fundamental constraint for interferometric radar applied to natural (distributed scattering) surfaces. Point targets, sometimes called permanent scatterers, can maintain phase coherence beyond this critical baseline, however. Difficulty in phase unwrapping increases (see Section 3.12.2.4) as the fringe frequency approaches this critical value.

The fringe variation in the interferogram is 'flattened' by subtracting the expected phase from a surface of constant elevation. The resulting fringes follow the natural topography more closely. Letting $\hat{l}_0$ be a unit vector pointing to a surface of constant elevation, h_0, the flattened phase, ϕ_{flat}, is given by

$$\phi_{\mathrm{flat}} = -\frac{2\pi p}{\lambda} \left(\left\langle \hat{l}, \mathbf{B} \right\rangle - \left\langle \hat{l}_0, \mathbf{B} \right\rangle \right) \qquad [25]$$

where

$$\hat{l}_0 = (\sin \theta_0, -\cos \theta_0) \qquad [26]$$

and $\cos \theta_0$ is given by the law of cosines

$$\cos \theta_0 = \frac{r_0^2 + (r_{\mathrm{e}} + h_{\mathrm{p}})^2 - (r_{\mathrm{e}} + h_0)^2}{2(r_{\mathrm{e}} + h_{\mathrm{p}}) r_0} \qquad [27]$$

assuming a spherical Earth with radius r_{e} and a slant range to the reference surface r_0.

Equation [25] can be simplified by expanding the look angle $\theta = \theta_0 + \delta\theta$, where $\delta\theta$ is the contribution to the look angle at range r_0 from the topographic relief relative to a reference surface, and θ_0 is the look angle to the reference surface at range r_0. If the topographic relief is represented by δz, then $\delta\theta = \delta z / r_0 \sin \theta_0$, and

$$\phi_{\mathrm{flat}} = -\frac{2\pi p}{\lambda} \left(\left\langle \hat{l}, \mathbf{B} \right\rangle - \left\langle \hat{l}_0, \mathbf{B} \right\rangle \right) \approx -\frac{2\pi p}{\lambda} B_{\perp} \frac{\delta z}{r_0 \sin \theta_0} \qquad [28]$$

is the component of baseline perpendicular to the LOS. Equation [28] tells us several things about the fringes:

- The flattened fringes are proportional to the topographic height directly. A poorman's topographic map then can be generated by flattening the phase and examining the fringes.
- The flattened fringes are proportional to the perpendicular component of the baseline. For zero baseline, there are no fringes, even if there is a large parallel component of the baseline. For large baselines, there are many cycles of phase change for a given topographic change. From this equation, we can define h_{a}, the 'ambiguity height', as $h_{\mathrm{a}} = \partial h / \partial \phi = \lambda r_0 \sin \theta_0 / 2\pi p B_{\perp}$.
- In the absence of topographic variations, there is still an intrinsic variation of fringes across an interferogram given by the flattening phase.

3.12.2.3 Differential Interferometry

The theory just described assumes that the imaged surface is stationary over time, or that the surface is imaged by the interferometer at a single instant. When there is motion of the surface between radar observations there is an additional contribution to the interferometric phase variation. **Figure 10** at right shows the geometry when a surface displacement occurs between the observation at $\mathbf{P}_1$ (at time t_1) and the observation at $\mathbf{P}_2$ (at $t_2 > t_1$). In this case, $\mathbf{l}_2$ becomes

$$\mathbf{l}_2 = \mathbf{T} + \mathbf{D} - \mathbf{P}_2 = \mathbf{l}_1 + \mathbf{D} - \mathbf{B} \qquad [29]$$

where $\mathbf{D}$ is the displacement vector of the surface from t_1 to t_2. The interferometric phase expressed in terms of this new vector is

$$\phi = \frac{4\pi}{\lambda} \left(\left\langle \mathbf{l}_1 + \mathbf{D} - \mathbf{B}, \mathbf{l}_1 + \mathbf{D} - \mathbf{B} \right\rangle^{1/2} - r_1 \right) \qquad [30]$$

Assuming, as above, that $|\mathbf{B}|$, $|\mathbf{D}|$, and $|\langle \mathbf{B}, \mathbf{D}\rangle|$ are all much smaller than r_1, the phase reduces to

$$\phi = \frac{4\pi}{\lambda}\left(-\langle \mathbf{l}_1, \mathbf{B}\rangle + \langle \mathbf{l}_1, \mathbf{D}\rangle\right) \qquad [31]$$

Typically, for spaceborne geometries $B < 1$ km, and D is of order meters, while $r_1 \approx 600\text{--}800$ km. This justifies the usual formulation in the literature that

$$\phi_{\text{obs}} = \phi_{\text{topography}} + \phi_{\text{displacement}} \qquad [32]$$

In some applications, the displacement phase represents a nearly instantaneous translation of the surface resolution elements, for example, earthquake deformation. In other cases, such as glacier motion, the displacement phase represents a motion tracked over the time between observations. Intermediate cases include slow and/or variable surface motions, such as volcanic inflation or surging glaciers. Equations [31] and [32] highlight that the interferometer measures the projection of the displacement vector in the radar LOS direction. To reconstruct the vector displacement, observations must be made from different LOS directions (see Section 3.12.4.5).

The topographic phase term is not of interest for displacement mapping, and must be removed. Several techniques have been developed to do this. They all essentially derive the topographic phase from another data source, either a DEM or another set of interferometric data. The selection of a particular method for topography measurement depends heavily on the nature of the motion (steady or episodic), the imaging geometry (baselines and time separations), and the availability of data.

It is important to appreciate the increased precision of the interferometric displacement measurement relative to topographic mapping precision. Consider a discrete displacement event such as an earthquake where the surface moves by a fixed amount $\vec{\mathbf{D}}$ in a short time period. Neither a pair of observations acquired before the event (pair 'a') nor a pair after the event (pair 'b') would measure the displacement directly, but together would measure it through the change in topography. According to eqn [30], and assuming the same imaging geometry for 'a' and 'b' without loss of generality, the phase difference between these two interferograms (i.e., the difference of phase differences) is

$$\phi = \phi_a - \phi_b \qquad [33]$$

$$= \frac{4\pi}{\lambda}\left[\left(\langle \mathbf{l}_1 - \mathbf{B}, \mathbf{l}_1 - \mathbf{B}\rangle^{1/2} - r_1\right)\right. \qquad [34]$$

$$\left. - \left(\langle \mathbf{l}_1 + \mathbf{D} - \mathbf{B}, \mathbf{l}_1 + \mathbf{D} - \mathbf{B}\rangle^{1/2}\right.\right.$$

$$\left.\left. - \langle \mathbf{l}_1 + \mathbf{D}, \mathbf{l}_1 + \mathbf{D}\rangle^{1/2}\right)\right] = 0 \qquad [35]$$

to first order, because $\vec{\mathbf{D}}$ appears in both the expression for $\mathbf{l}_2$ and $\mathbf{l}_1$. The nature of the sensitivity difference inherent between eqns [31] and [35] can be seen in the 'flattened' phase (see eqn [28]) of an interferogram, often written (Rosen *et al.*, 1996)

$$\phi = -\frac{4\pi}{\lambda}B\cos(\theta_0 - \alpha)\frac{z}{r_0 \sin\theta_0} + \frac{4\pi}{\lambda}\delta r_{\text{disp}} \qquad [36]$$

where δr is the surface displacement between imaging times in the LOS direction, and z is the topographic height above the reference surface. In this formulation, the phase difference is far more sensitive to changes in topography (surface displacement) than to the topography itself. From eqn [36] $\delta r = \lambda/2$ gives one cycle of phase difference, while z must change by a substantial amount, essentially r_0/B, to affect the same phase change. For example, for ERS, $\lambda = 5.6$ cm, $r_1 \approx 830$ km, and typically $B \le 200$ m, implying $\delta r_{\text{disp}} = 2.8$ cm to generate one cycle of phase, $z \ge 450$ cm to have the same effect.

The time interval over which the displacement is measured must be matched to the geophysical signal of interest. For ocean currents, the temporal baseline must be of the order of a fraction of a second because the surface changes quickly and the assumption that the backscatter phase is common to the two images could be violated. At the other extreme, temporal baselines of several years may be required to make accurate measurements of slow deformation processes such as interseismic strain.

Reconstruction of the scatterer position vector depends on knowledge of the platform location, the interferometric baseline length, orientation angle, and the interferometric phase. To generate accurate topographic or displacement maps, radar interferometry places stringent requirements on knowledge of the platform and baseline vectors, as well as the intrinsic accuracy of the phase measurements and, in the case of differential interferometry, supporting topographic data sets. One source of phase noise is the refractivity of the atmosphere, which varies along the radar propagation path in space and time. Refractivity fluctuation due to turbulence in the atmosphere is a minor effect for two-aperture cross-track interferometers, but is much more important for repeat-track systems (Rosen *et al.*, 1996). Sensitivities to these parameters are discussed in detail in Rosen *et al.* (2000) and Zebker *et al.* (1994).

3.12.2.4 Phase Unwrapping

The phase of the interferogram must be unwrapped to remove the modulo-2π ambiguity before estimating topography or surface displacement. The literature describing approaches to phase unwrapping is quite large (Rosen *et al.*, 2000), with initial development of so-called branch-cut techniques for InSAR applications by Goldstein *et al.* (1988), followed by over a decade of exploration of other techniques.

A simple approach to phase unwrapping would be to form the first differences of the phase at each image point in either image dimension as an approximation to the derivative, and then integrate the result. Direct application of this approach, however, allows local errors due to phase noise to propagate, causing errors across the full SAR scene. The unwrapped solution should, to within a constant of integration, be independent of the path of integration. This implies that in the error-free case, the integral of the differenced phase about a closed path is zero. Phase inconsistencies are therefore indicated by nonzero results when the phase difference is summed around the closed paths formed by each mutually neighboring set of four pixels. These points have either a positive or negative integral (by convention performed in clockwise paths). Integration of the differenced phase about any closed path yields a value equal to the sum of the enclosed points of inconsistency. As a result, paths of integration that encircle a non zero sum must be avoided. In branch-cut methods, this is accomplished by connecting oppositely signed points of phase inconsistency with lines that the path of integration cannot cross. Once these barriers have been selected, phase unwrapping is completed by integrating the differenced phase subject to the rule that paths of integration do not cross the barriers.

The phase unwrapping problem becomes particularly difficult when the phase in the interferogram is intrinsically discontinuous, due to layover problems or true shear topography. Most algorithms are based on the assumption that the phases are continuous, and often natural phase discontinuities, often corrupted with inherent phase noise, are difficult to interpret.

A full treatment of phase unwrapping for geodetic imaging applications is beyond the scope of this chapter. There are a number of algorithms available for use, including branch-cut algorithms (Goldstein *et al.*, 1988) and statistical cost network flow techniques (Chen and Zebker, 2001). These techniques yield unwrapped phase images that are multiples of 2π of the original wrapped phase image. In the case of branch-cut algorithms, there are often regions that are blocked off from unwrapping by barriers that form a complete circuit. For network flow, the entire image is unwrapped. In all cases, there will be individual pixels or areas that are placed on the wrong multiple of 2π, and it is often quite difficult to identify these points without additional information.

3.12.2.5 Correlation

The relationship between the scattered electromagnetic fields seen at the interferometric receivers after image formation is characterized by the complex correlation function, γ, defined as

$$\gamma = \frac{\langle \Gamma_1 \Gamma_2^* \rangle}{\sqrt{\langle |\Gamma_1|^2 \rangle \langle |\Gamma_2|^2 \rangle}} \qquad [37]$$

where Γ_i represents the SAR reflectivity at the i antenna, and angular brackets denote averaging over the ensemble of speckle realizations. For completely coherent scatterers such as point scatterers, we have that $\gamma = 1$, while $\gamma = 0$ when the scattered fields at the antennas are independent. The magnitude of the correlation $|\gamma|$ is often referred to as the 'coherence.' (Several authors distinguish between the "coherence" properties of fields and the correlation functions that characterize them (e.g., Born and Wolf (1989)), whereas others do not make a distinction.)

In general, the correlation will comprise contributions from a number of effects:

$$\gamma = \gamma_N \gamma_G \gamma_Z \gamma_T \qquad [38]$$

where γ_N is the correlation influenced by noise in the radar system and processing approach, γ_G is that influenced by the different observing geometries, γ_Z describes the influence on correlation of the vertical extent of scatterers (e.g., due to vegetation), and γ_T describes the influence of repositioning of scatterers within a resolution element over time (Li and Goldstein, 1990; Zebker and Villasenor, 1992; Rodríguez and Martin, 1992; Bamler and Hartl, 1998; Rosen *et al.*, 2000). It is often more convenient to discuss decorrelation, defined as $\delta_X = 1 - \gamma_X$, where X is N, G, Z, or T.

Geometric decorrelation, δ_G, also called baseline or speckle decorrelation, is due to the fact that, after removing the phase contribution from the center of

the resolution cell, the phases from the scatterers located away from the center are slightly different at each antenna (see **Figure 7**). The degree of decorrelation can then be estimated from the differential phase of two points located at the edges of the area obtained by projecting the resolution cell phase from each scatterer within the resolution cell, as shown in **Figure 7**. Using this simple model, one can estimate that the null-to-null angular width of the correlation function, $\Delta\theta$, is given by

$$\Delta\theta \approx \frac{B_\perp}{r} \approx \frac{\lambda}{\Delta r_l} \qquad [39]$$

where $B_\perp$ is the projection of the interferometric baseline onto the direction perpendicular to the look direction, and Δr_l is the projection of the ground resolution cell along the same direction, as illustrated in **Figure 11**.

The geometric correlation term is present for all scattering situations, and depends on the system parameters and the observation geometry. A general expression for it is

$$\gamma_G = \frac{\int ds\, dr\, W_1(r, x) W_2^*(r + \delta_r, x + \delta_x)}{\times \exp[jr(p\kappa_r + 2\delta k)]\exp[j\tan\tau_x\kappa_z s]} \qquad [40]$$

where $k \equiv 2\pi/\lambda$ is the wave number; δk represents the shift in the wave number corresponding to any difference in the center frequencies between the two interferometric channels; δ_r and δ_x are the misregistration between the two interferometric channels in
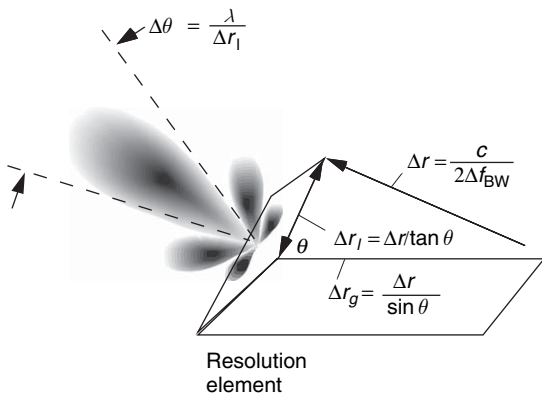
Figure 11 A view of geometric decorrelation showing the effective beam pattern of a ground resolution element 'radiating' to space. The mutual coherence field propagates with radiation beam width in elevation of $\Delta\theta \approx \lambda/\Delta r_l$. These quantities are defined in the figure.

the range (r) and azimuth (x) directions, respectively; $W_i(r, x)$ is the SAR point-target response in the range and azimuth directions; and τ_x is the surface slope angle in the azimuth direction. In eqn [40], κ_r and κ_z are the interferometric fringe wave numbers in the range and vertical directions, respectively. They are given by

$$\kappa_r = \frac{kB_\perp}{r\tan(\theta - \tau_c)} \qquad [41]$$

$$\kappa_z = \frac{kB_\perp \cos\tau_c}{r\sin(\theta - \tau_c)} = \kappa_r \frac{\cos\tau_c}{\cos(\theta - \tau_c)} \qquad [42]$$

The value of δk can be adjusted to recenter the range spectrum of each interferometric channel. This can be accomplished in principle by bandpass filtering the range spectrum differently in each channel. Under the right conditions, one can adjust the center frequencies to create the condition $2\delta k = -\kappa_r$, which leads to $\gamma_G = 1$ (Gatelli *et al.*, 1994). In other words, the geometric decorrelation term in principle can be eliminated by proper choice of center frequencies for two observations. In practice, as can be seen from the equation above, κ_r depends on the look angle and surface slope, so that adaptive iterative processing is required in order to implement the approach exactly.

The volumetric correlation term can be understood in terms of an effective increase in the size of the projected range cell Δr_l because the scattering elements in a given range cell are now extended not just on a surface but in a volume (Rosen *et al.*, 2000). If the range resolution is infinitesimally small, the volume decorrelation effect can be understood as being due to the geometric decorrelation from a plane cutting through the scattering volume perpendicular to the look direction. It was shown in Rodríguez and Martin (1992) that the volumetric correlation γ_Z can be written as

$$\gamma_Z(\kappa_z) = \int dz\, f(z)\exp[-j\kappa_z z] \qquad [43]$$

provided the scattering volume could be regarded as homogeneous in the range direction over a distance defined by the range resolution. The function $f(z)$, the 'effective scatterer probability density function (pdf)', is given by

$$f(z) = \frac{\sigma(z)}{\int dz\, \sigma(z)} \qquad [44]$$

where $\sigma(z)$ is the effective normalized backscatter cross-section per unit height. The term 'effective' is

used to indicate that $\sigma(z)$ is the intrinsic cross-section of the medium attenuated by all propagation losses through the medium. The specific form for $\sigma(z)$ depends on the scattering medium. Models for this term, and its use in the remote sensing of vegetation height, will be discussed later.

In repeat-pass systems, there is another source of decorrelation. Temporal decorrelation, $\delta_{\rm T}$, occurs when the surface changes between the times when the images forming an interferogram are acquired (Zebker and Villasenor, 1992). As scatterers become randomly rearranged over time, the detailed speckle patterns of the image resolution elements differ from one image to the other, so the images no longer correlate.

While it is difficult to describe these effects analytically, this can often be a strong limitation on the accuracy of repeat-pass data, so a few illustrative examples are in order. Open water, where the surface is roughened by wind or turbulence, is constantly changing over time, so two images will completely decorrelate ($\gamma = 0$). Similarly, an agricultural field, where the entire surface has been turned over due to tilling, will have $\gamma = 0$. Standing water with no vegetation present above water will also completely decorrelate because no signal is scattered back toward the radar. However, with vegetation present above the surface, the water serves as a mirror that permits signal return from the water as scattered off the vegetation (Alsdorf *et al.*, 2001; Wdowinski *et al.*, 2004). Rain can change the arrangement of vegetation on a surface (e.g., sagging branches or stalks), reducing the correlation by an amount dependent on the density of altered vegetation. In some cases, as the surface dries, the vegetation bounces back to its original position and correlation is at least partially restored. However, wind typically will alter the positions of scatterers in vegetation canopies over time, so correlation is generally degraded in vegetation. Snow can destroy correlation in the winter months, with correlation restored after the snow is gone. For interferometry applied to geophysical processes, we rely on the block motion of pixels without scatterer rearrangement to provide estimates of the geodetic motion. These effects, and changes due to ground shaking (building collapse, landslides, liquefaction, etc.), can impair the ability to measure displacements. On the other hand, it can also be a means for understanding the nature of the surface and the severity of the geophysical effects on the ground.

In addition to these field correlations, thermal noise in the interferometric channels also introduces phase noise in the interferometric measurement. The correlation due to thermal noise alone can be written as

$$\gamma_{\rm N} = \frac{1}{\sqrt{1 + {\rm SNR}_1^{-1}}\sqrt{1 + {\rm SNR}_2^{-1}}} \quad [45]$$

where ${\rm SNR}_i$ denotes the signal-to-noise ratio for the i channel (Zebker and Villasenor, 1992). In addition to thermal noise, which is additive, SAR returns also have other noise components, due to, for example, range and Doppler ambiguities. An expression for the decorrelation due to this source of error can only be obtained for homogeneous scenes, since, in general, the noise contribution is scene dependent. Typically, for simplicity these ambiguities are treated as additive noise as part of the overall system noise floor.

The effect of decorrelation is the apparent increase in noise of the estimated interferometric phase. Rodríguez and Martin (1992) presented the analytic expression for the Cramer–Rao bound (Sorenson, 1980) on the phase variance

$$\sigma_\phi^2 = \frac{1}{2N_{\rm L}} \frac{1 - \gamma^2}{\gamma^2} \quad [46]$$

where N_L is the number of independent samples used to derive the phase, and is usually referred to as the 'number of looks.' The actual phase variance approaches the limit eqn [46] as the number of looks increases, and is a reasonable approximation when the number of looks is greater than four. An exact expression for the phase variance can be obtained starting from the probability density function for the phase when $N_L = 1$, and then extended for an arbitrary number of looks (Goodman, 1985; Joughin *et al.*, 1994; Lee *et al.*, 1992; Touzi and Lopes, 1996). The expressions, however, are quite complicated and must be evaluated numerically in practice.

Note that the estimate of the correlation is usually accomplished by computing the expectation operations in eqn [37] as spatial averages over a number of pixels in an interferometric pair. This leads to biased estimates of the correlation, and care must be exercised in interpreting the estimate. For example, in open water, where actually the coherence of the fields is zero, the correlation estimate will produce decidedly nonzero estimates, in the range of 0.1–0.3, depending on the number of samples used in the estimate, N_L. The estimator is a random variable with a probability distribution shape that depends on the intrinsic coherence, and the number of samples used in the estimate. In the limit, where only one sample is used, the correlation estimate will be 1!

3.12.3 InSAR-Related Techniques

3.12.3.1 ScanSAR or Wide Swath Interferometry

For previously flown SAR systems, the width of the swath has been limited to somewhat less than 100 km. As discussed in Section 3.12.5, SAR antennas must satisfy particular minimum area criteria to ensure noise due to ambiguities below a required level. For wide swath, they must also be quite long, which can be difficult and costly to implement. For example, to achieve a 300 km swath in typical Earth orbits using the typical strip mapping method, the antenna would have to be over 40 m in length. To achieve wide swaths with an antenna sized for a swath smaller than 100 km, the ScanSAR technique has been developed (Tomiyasu, 1981; Moore *et al.*, 1981). ScanSAR requires an antenna with rapid electronic steering capability in the elevation direction. In ScanSAR,

the antenna is electronically steered to a particular elevation angle and radar pulses are transmitted and echoes received for a time period that is a fraction (say one-tenth) of the synthetic aperture time. After that 'dwell period,' also known as the 'burst period,' the antenna is electronically steered to another elevation angle (and other radar parameters such as the PRF, bandwidth, and antenna beam shape are changed), and observations are made for another short dwell period. This process is repeated until each of the elevation directions, needed to observe the entire wide swath is obtained at which point the entire cycle of elevation dwell periods is repeated (**Figure 12**).

For any given elevation direction, or subswath, there are large gaps in the receive echo timeline, yet after processing the data, a continuous strip mode image can be formed. This is true because the extent of the antenna beam in the along-track direction on the ground is equal to the synthetic aperture
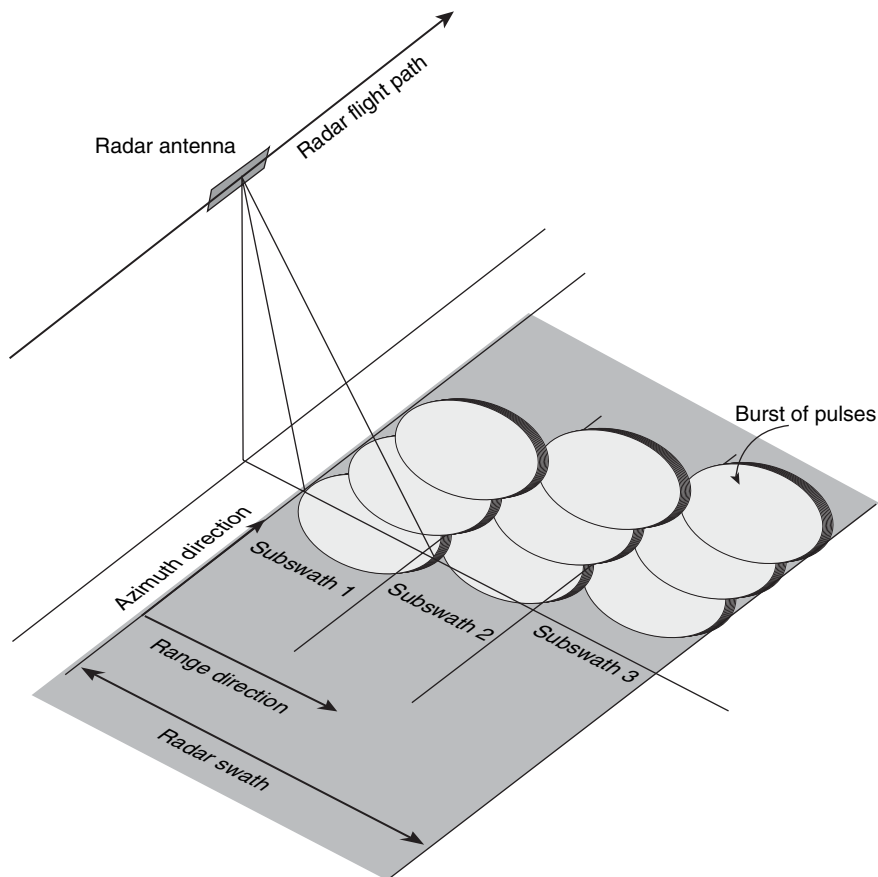
Figure 12 ScanSAR for a three-beam system. The radar transmits a burst of pulses to illuminate subswath 1, then electronically switches to point to subswath 2, then subswath 3. This cycle is repeated for the extent of the data take. The aspect ratio of the beams in this figure is highly stretched in azimuth to illustrate the pulsed behavior for the bursts and the significant beam overlap from burst to burst within a subswath. This overlap allows for construction of continuous maps.

length. As long as the dwell periods for any given subswath occur more than once in the synthetic aperture time, there is guaranteed continuous coverage of all points on the ground.

It is important to understand this method of generating radar data because it has strong implications for the quality of the geodetic data that are derived, and for the constraints that are imposed on the use of the data. First we note that the data contained in any given pulse include the full Doppler spectrum of information. We are transmitting over a broad range of angles (the beam width) and that defines the Doppler frequency content. So each burst of pulses contains the full Doppler spectrum of information. If one were to derive the spectrum in the along-track dimension, the full Doppler bandwidth would be represented. However, we note that any given scattering element within a burst period only contributes a portion of its full Doppler history because it is not observed over the full synthetic aperture time. Thus each scattering element is only resolved commensurate with the burst period relative to the synthetic aperture time: if the resolution in strip mode is $L/2$, then, the resolution in ScanSAR mode is $(L/2) * T_a/T_b$, where T_a is the synthetic aperture time, and T_b is the burst duration. If one were to attempt to achieve maximal resolution possible, one would divide the synthetic aperture time by the number of subswaths needed, and set the burst duration to this time. However, as described following, it is generally better to create several short bursts within one synthetic aperture. This degrades resolution further, but improves the radiometric characteristics of the data.

For interferometry, the most important aspect of ScanSAR is the fact that each scattering element provides in a burst only a small portion of its total Doppler history (Bamler and Eineder, 1996; Guarnieri and Prati, 1996; Lanari *et al.*, 1998). This is equivalent to the statement that each scattering element is observed over a small range of azimuth angles within the beam. In order for interferometry to work, a scattering element must be observed with the same range of Doppler frequencies from one pass to the other. Only then will the images be coherent, with similar speckle patterns. This implies that from pass to pass, each observation must observe from the same group of angles. In the case of strip mode SAR, this implies that the intrinsic pointing of the antenna beam be the same from pass to pass. In this case, the Doppler history of each scattering element will follow the same course. For ScanSAR, this condition also implies a timing constraint on the bursts. Each burst must occur at the same location in space relative to a scattering element from pass to pass. This constraint makes it much more challenging for mission operators, particularly with short bursts. Figure 13 illustrates the workflow associated with ScanSAR interferometry as well as required intermediary data objects.

For a satellite in a long repeat period orbit, for example 32 days, the ground swath will be sized to achieve global coverage, around 85 km for the 32-day repeat period orbit. So from an interferometric point of view, using ScanSAR with this period does not improve on the interferometric interval intrinsically. However, ScanSAR will increase the number of times a given scattering element will be observed by a factor of the number of ScanSAR beams. For a four-beam ScanSAR with 340 km swath in a 32-day repeat orbit, a scattering element will be observed roughly every 8 days, so one could make interleaved 32-day repeat interferograms. Alternatively, one could place the radar in a shorter repeat period orbit, for example 8 days, and observe always in ScanSAR mode with four beams. This would allow 8-day interferograms with no interleaving. The advantage of the latter is that decorrelation will be less of an issue in the basic repeat-pass measurement. The advantage of the former is that there will be greater angular diversity in the measurement, potentially resulting in better constrained models of deformation.

3.12.3.2   Permanent Scatterers and Time-Series Analysis

The two major error sources in InSAR are decorrelation due to temporal and geometric effects and phase errors introduced by the spatially and temporally random variations of the refractive index of the atmosphere and ionosphere. Decorrelation creates areas that are spatially disjoint, and irregularly so over time, leading to difficulty in interpreting the geodetic measurements. Because the repeat period of the orbital satellites is on the order of weeks, temporal sampling of dynamic phenomena is also poor. Continuous GPS measurements share similar characteristics – spatially disjoint measurements that also are subject to atmospheric effects – but have the advantage of dual-frequency measurements for ionospheric correction, and continuous temporal sampling while observing multiple satellites, which
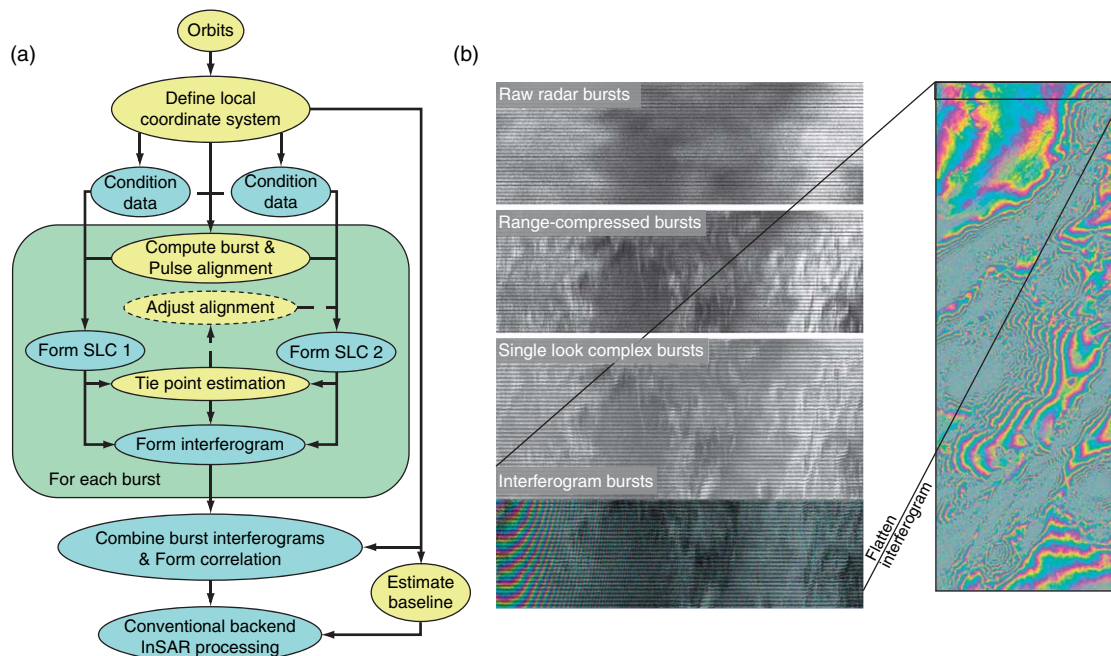
Figure 13 (a) Representative differential InSAR processing flow diagram for ScanSAR interferometry. Blue bubbles represent image output, yellow ellipses represent nonimage data. Flow is generally down the solid paths, with optional dashed paths indicating potential iteration steps. (b) Illustration of the workflow by explicity example. The top three black and white panels show a collection of bursts from one scene for one ScanSAR subswath. The top panel displays raw radar data, with a small artificial gap between bursts to delimit them. The next panel shows the bursts after range compression – features appear to be sharper in the range (across image) dimension. The next panel shows the bursts after azimuth compression – single look complex bursts – features are now sharp in both dimensions, and the existence of surface features in three or four successive bursts is apparent, showing the intrinsic overlap of the data. The bottom center panel shows burst interferograms formed from two SLC burst sequences. After combination and flattening, the bursts form a continuous interferogram as shown on the right, with the left sequence contained in the box at the top as indicated. This interferogram, as well as those of the other subswaths can then be processed as usual.

allows for a fine sampling of atmospheric variability, leading to robust correction algorithms. Given that important areas of the world have significant issues with decorrelation and atmospheric water vapor, interferometry research incorporates methods to reduce the problem to one more similar to continuous GPS. Corner reflectors have long been used as coherent calibration sources for interferometric systems, as they are observable over a wide range of angles and have a well-defined phase center. Usai was the first to note that some man-made structures can behave much like corner reflectors, as though they are continuously reliable coherent scatterers over time and independent of the interferometric baseline (Usai, 1997). By identifying points in a series of radar images that maintain their coherence over time, we can create a network of phase measurements over time and space that are directly analogous to GPS measurements, though sampled once per month

rather than continuously. Ferretti and colleagues (Ferretti *et al.*, 2000, 2001) were the first to systematize and popularize the treatment of these discrete point networks for long-term trend analysis. A number of groups have pursued similar techniques (Wegmuller, 2003; Hooper *et al.*, 2004). **Figure 14** illustrates the way this technique overcomes the atmospheric error issues.

In practice, it is challenging to determine the location of time-coherent scatterers because the phase of each point is comprised of a topographic component, a displacement component, and noise components, all of which are different in each image of a series of images:

$$\phi_{\Delta t_i} = \phi_{\text{topo},\Delta t_i} + \phi_{\text{def},\Delta t_i} + \phi_{\text{atm},\Delta t_i} + \phi_{\text{noise},\Delta t_i} \quad [47]$$

where Δt_i is the time interval for interferogram i, and the interferometric phase, $\phi_{\Delta t_i}$, is broken into its principal components. Though the topography is assumed to be static over time, the phase term
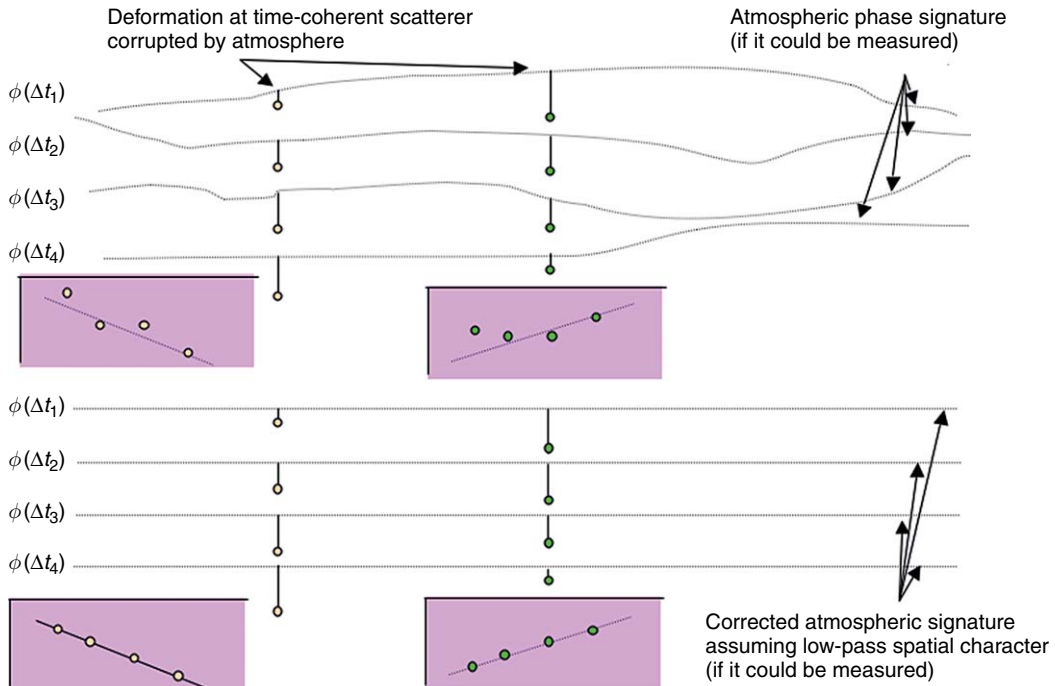
Figure 14 The 'permanent scatterer' technique identifies time-coherent scatterers by estimating the contributions of topography, deformation, and atmospheric delay to the phase under model constraints through correlation maximization. Topography is assumed to be static (with the interferometric phase proportional to baseline), deformation is assumed to follow some functional form (e.g., linear or sinusoidal with time), and atmospheric delay is assumed to vary randomly in time and with long spatial wavelength.

$\phi_{\text{topo},\Delta t_i}$ is variable because each interferogram has a different perpendicular component of the baseline $B_{\perp,i}$. Similarly, the other component terms will change over time. If we simply coalign all available images, then estimate the correlation over time for each point using eqn [37] (with expectation approximated by time averaging), the variability of the phase over time will lead to an estimate of zero correlation. Ferretti *et al.* (2000, 2001) chose to identify time-coherent points by using a brightness threshold, as points with high mean brightness over time have small phase dispersion. This requires that images are radiometrically calibrated, such that variations in the brightness are due to scene variations, not radar system variations (e.g., radar antenna pointing variability).

Once these initial points are identified, it is necessary to find a solution for each phase term for all *i* that maximizes the correlation estimate. The maximization procedure involves a search over reasonable domains of expected phase values, which can be intensive. For example, if a DEM is available, the topographically induced phases are known, except for errors introduced by the DEM

errors, so the search range for topography should be bounded in a region around the known topography by that error. The deformation is parametrized by a functional form such as linear or sinusoidal, and the solution of the parameters is part of the maximization. The atmospheric error is assumed to be spatially slowly varying, and uncorrelated over time.

This technique has been employed with success in urban areas where man-made time-coherent scattering points abound. Results include the measurement of subsidence rates of individual buildings at the level of less than $1 \, \text{mm} \, \text{yr}^{-1}$, and seasonal effects due to groundwater withdrawal and recharge (Ferretti *et al.*, 2000; Colesanti *et al.*, 2003). Applications to problems of geophysical interest are limited because the areas of interest are generally larger and sufficient time-coherent scatterers frequently do not exist in nonurban environments. Burgmann *et al.* (2006) combined GPS-derived horizontal velocities and permanent scatterer InSAR estimates of uplift in the San Francisco Bay Area to track tectonic uplift in areas not subject to seasonal effects, at an accuracy of better than $1 \, \text{mm} \, \text{yr}^{-1}$.

Other methods have been developed to exploit long time series available for over a decade of SAR observations, but without the restriction of using only time-coherent point scatterers, and without specific assumptions of the model of deformation. By limiting data to only those scene combinations where the baseline is well below the critical baseline, one can expand the area of usable image pixels that are from natural surfaces, but are coherent in time as well (Lundgren *et al.*, 2001; Berardino *et al.*, 2002; Schmidt and Burgmann, 2003; Lanari *et al.*, 2004b). In these methods, each collection of images allows the generation of a dense spatial network of points (for well-correlated areas, potentially every image pixel could be included in the network). For each point, interferometric pairs formed from various combinations of images yield phase differences over available time intervals in the time series where correlation is good, with the shortest sampling interval being the satellite repeat period. From these phase difference measurements for each network point, an inversion is carried out to reconcile all differences. The inversion adjusts each interferogram by a constant to bring all differences into agreement, and attempts to integrate the phase differences in a piecewise linear fashion, minimizing the distance between phase estimates from one time step to the next. The inversion can be carried out with constraints that smooth the estimates of phase. It is also possible to incorporate parametrized model functions in the inversion, for example, for seasonal effects, if desired. These techniques derive directly from GPS network inversion methods, with the substantive difference being the much higher density of samples in space, but much lower density of samples in time, for InSAR. These approaches are further discussed later in the chapter.

3.12.3.3   Speckle Tracking and Pixel Tracking

InSAR provides the satellite LOS component of relative displacements within an image. Albeit at lower sensitivities, we can also measure the along-track component of the displacement field using cross-correlation techniques. This is frequently referred to as speckle tracking, pixel tracking, or range and azimuth offsets. For crustal deformation purposes, we frequently only consider the along-track or azimuth offsets (AZOs) which are a purely horizontal component of the displacement field and are by construction perpendicular to the LOS phase-based
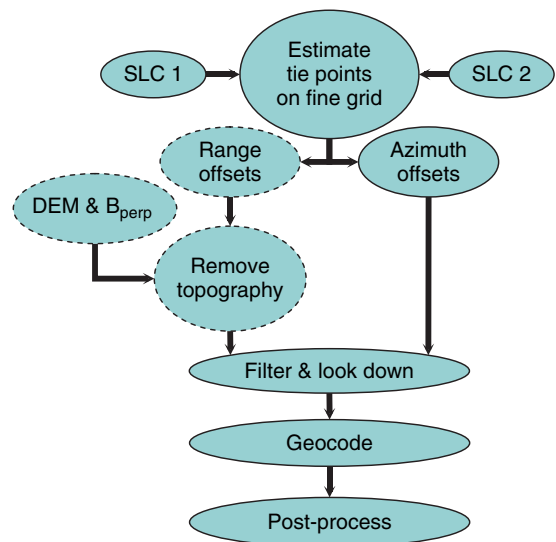
Figure 15   Generic processing flow diagram for generating pixel-tracking measurements. We assume here that the products of the associated differential InSAR processing have already been made.

observations. This approach is strictly speaking not an interferometric approach, but we include it here since it uses the same single look complex (SLC) images used in forming interferograms (**Figure 15**).

At its most basic level, this approach is simply automatic tie-point estimation and frequently uses the same algorithms and software used to calculate tie points when coregistering two SLCs to form an interferogram or to coregister a DEM-based height map to an interferogram to remove topographic effects (e.g., Michel *et al.*, 1999a, 1999b; Michel and Rignot, 1999; Joughin, 2002; Short and Gray, 2004; Gray *et al.*, 2001). Given a master and slave SLC and a rough guess at their relative gross offsets, a chip of $N \times M$ pixels (typically several tens of pixels on a side) is cross-correlated from both images to provide range and azimuth offset with subpixel accuracy. This process can then be repeated over a grid. Typical sensitivities are about somewhere between 1/10 and 1/30 of a pixel dimension. For the ERS satellites, these numbers correspond to about 10–20 cm sensitivity, with the advantage that they do not need to be phase unwrapped. Since the measurement is done using ensembles of pixels, the spatial resolution is substantially degraded relative to standard interferometric observations. This sensitivity of the measurement is limited by the dimensions of a pixel. While the range offsets provide the same component of the displacement field as the

interferometric phase, it is usually much less sensitive and is therefore not usually used. However, there are times, when correlation is low, that the phase-based measurements will not work, but the range offsets can (e.g., Peltzer *et al.*, 1999). In such cases, the range offsets need to be corrected for topographic effects in a way similar to what is done for the phase measurements. The AZO measurement complement the interferometric phase measurements, and therefore always provide useful observations when the expected displacements are large enough to be detected.

Algorithmically, the approach is the same as that used to generate tie points or offset measurements using optical imagery (Michel *et al.*, 1999a, 1999b). With InSAR data, we can achieve extremely accurate estimates of offsets with no inherent contrast in the mean radar backscatter. Since SAR data exhibit high-contrast speckle characteristics from pixel to pixel (Appendix 1) that are the same in interferometrically coherent SAR images, allowing good matching.

Cross-correlation can be performed using the complex images directly or using just the amplitude of the complex image. When complex images are used, we are in essence computing small interferograms and assessing the quality of the fringes therein. Because the phase is used, when the coherence is good, this method can lead to very tight constraints on the pixel offsets. However, when the coherence is poor, the phase contribution leads to extremely poor correlation estimates even when common surface features that are well correlated are present. This is truly a 'speckle-tracking' approach in that if the speckle differs between images, it will not track well. Cross-correlation of amplitude images tracks features, and so would be more accurately called 'feature-tracking' or 'pixel-tracking'. Because speckle appears as high-contrast features from pixel to pixel, amplitude matching also tracks speckle, though it lacks the tightness of the match of complex image correlation when the coherence is good. But it performs very well when images are rich with surface features in areas of poor interferometric correlation. Experimentation with both approaches show that very little matching accuracy is lost by correlating amplitude rather than complex images.

The complex or amplitude image is typically over-sampled by a factor of 2 before the cross-correlation; this is needed to avoid spectral aliasing of the images. The peak of the cross-correlation surface is identified therefore at half-pixel spacing as a first approximation to the offset location. The correlation surface is then interpolated to find the correlation peak with finer granularity. Accuracies have been empirically estimated at about 1/30 of a pixel, on the order of 10–15 cm for typical SAR systems.

We illustrate the use of pixel tracking with an example taken from the 1999 M_w 7.6 Chi–Chi earthquake in Taiwan. This event was a thrust earthquake where the footwall had relatively little topography, while the hanging wall is extremely rugged. Standard InSAR produced clean fringes in the footwall region, but is completely decorrelated in the hanging wall (**Figure 16**). In contrast, the AZO observations can be made for the entire image, with the caveat that strong spatial (median) filtering was required to remove outliers (**Figure 17**).

Instead of using speckle or pixel tracking methods, as described above to infer the along-track component of displacement, Bechor and Zebker (2006) proposed splitting the aperture normally used in forming a single interferogram into separate interferograms using the forward and backward squinting SLCs (relative to the nominal squint angle for the standard SLC). These two interferograms can then in turn be differenced to produce a map of along-track displacements. They showed accuracies on the order of a few centimeters when the interferometric coherence was very good. In areas of lower coherence, the phase difference-derived estimates were comparable to typical AZO estimates.

3.12.4   A Best-Practices Guide to Using InSAR and Related Observations

Our general goal with geodetic imaging is to discover new crustal deformation processes and to estimate the value and uncertainties of controlling parameters associated with these as well as previously recognized mechanisms. The approaches adopted for using these data vary depending on the extent and type of observations available and the geophysical target. Here, we address some of the approaches that have found some utility in different situations. Beyond the exploratory mode of just looking at the images to discover things, we can consider how to best combine the images, deal with long time series, and how to incorporate these observations in parameter estimation schemes in a computationally efficient manner. In all cases, we must consider the different components of the measurements that act as apparent errors in the geodetic signal. From the perspective of modeling tectonic processes, the primary sources of error come from inadequacies in our knowledge of the satellite orbits
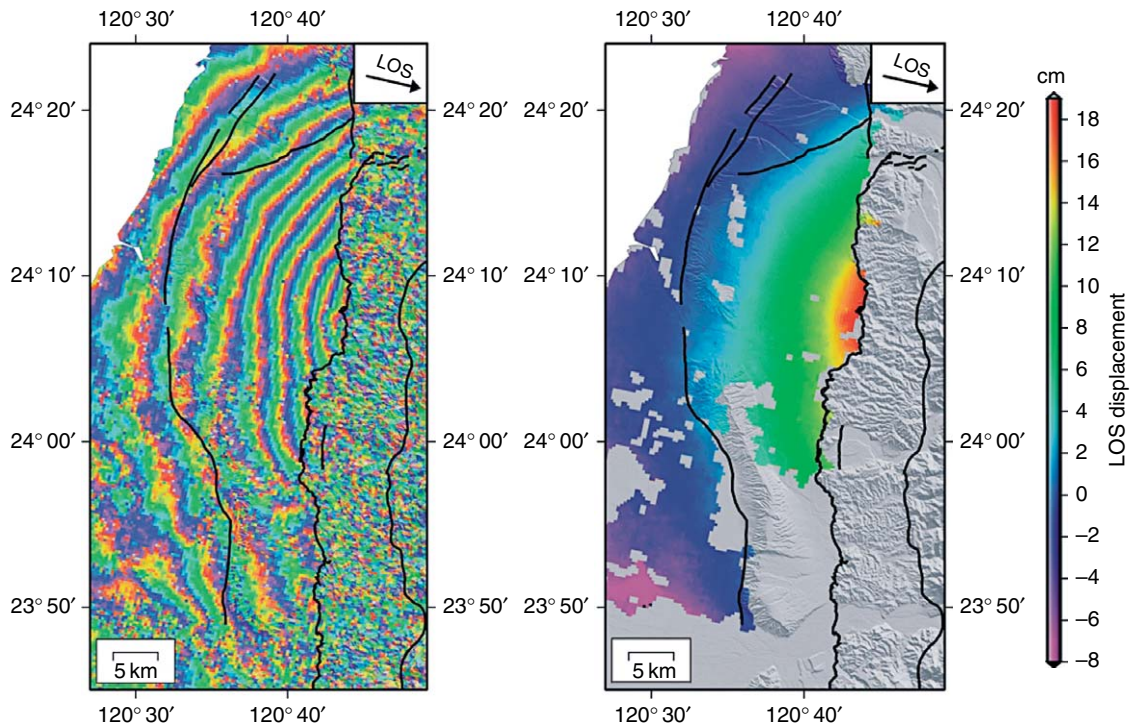
Figure 16 Wrapped at 2.8 cm per fringe (left) and unwrapped (right) interferogram for the 1999 M_W 7.6 Chi-Chi, Taiwan continental thrust earthquake using ERS C-band data from 5/5/99 to 9/23/99. Black lines indicate major fault traces. The Chulengpu Fault is the primary slipping structure for the Chi-Chi earthquake and forms the eastern limit of the unwrapped phase. The phase is successfully unwrapped only on the footwall where the topographic gradients and vegetation cover are dramatically less than in the hanging wall. Adapted from Levy F, Hsu Y, Simons M, LePrince S, and Avouac JP (2005) Distribution of coseismic slip for the 1999 Chi Chi Taiwan earthquake: New data and implications of varying 3D fault geometry. *EOS Transactions of the American Geophysical Union* 86: B1305.

and propagation delays accumulated in the ionosphere and troposphere. Here, we begin with a summary of error sources, since that impacts all uses of the data. At a minimum, knowledge of the error structure (i.e., the covariance structure) is important in order to correctly execute any estimation of geophysical parameters. Furthermore, in certain circumstances, a given component of the error may be reduced using a variety of techniques.

3.12.4.1 Interferometric Baseline Errors

The primary impact of orbital errors is to induce long-wavelength phase ramps associated with incorrect removal of topographic effects. If $B_\perp$ is long, errors in $B_\perp$ can also lead to short-wavelength errors in regions of rough topography. Typically, apparent long-wavelength deformation gradients associated with errors in $B_\perp$ are dealt with by removing a best-fit bilinear or biquadratic polynomial ramp from both the observations and the models before comparing

the two (e.g., Pritchard *et al.*, 2006; Pritchard and Simons, 2006), or by using independent data, usually GPS observations, to either re-estimate the interferometric baseline or to constrain the longest wavelengths of the physical model (e.g., Peltzer *et al.*, 2001; Simons *et al.*, 2002). Alternatively, if the deformation signal is known to be localized within the interferogram, one can flatten the image by assuming zero deformation in the far field.

In order to merge GPS observations with InSAR data, we must take care how these data are combined. The surface deformation field frequently shows a strong cyclical seasonal component (e.g., Dong *et al.*, 2002). This seasonal signal is exemplified by the GPS time series in **Figure 18**, which shows a greater than 1 cm peak-to-peak seasonal displacement. Of course, with InSAR data, one is sensitive to spatial gradients in these seasonal signals on the scale of the InSAR image, which will usually be smaller than implied by these kind of absolute GPS observations. If one uses GPS to refine estimates of $B_\perp$ then one should use the
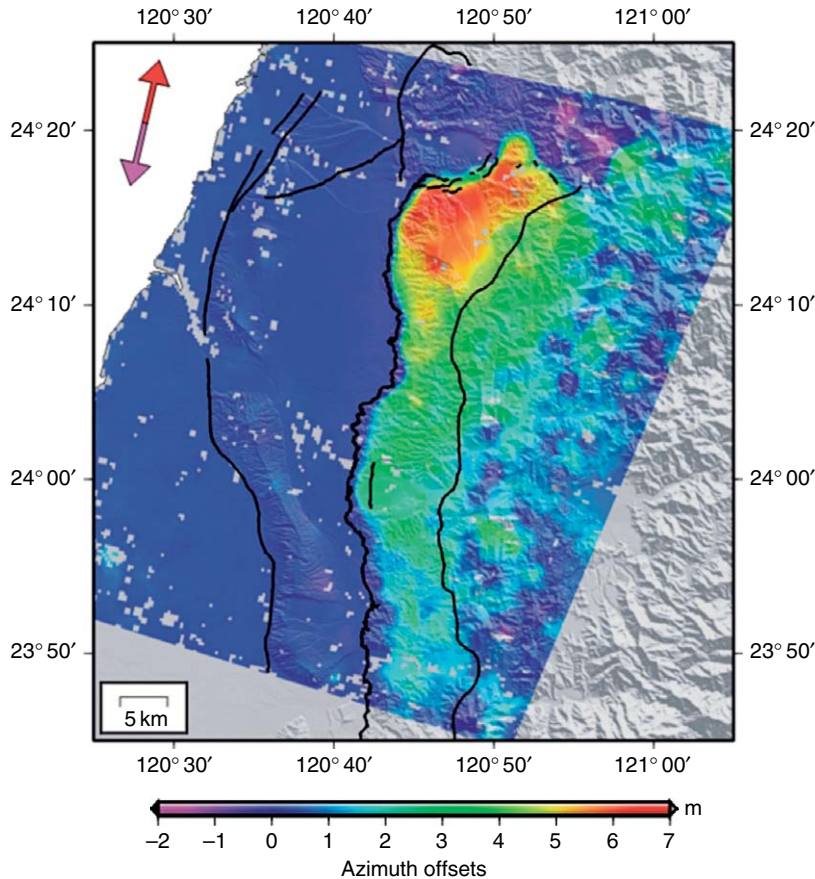
Figure 17 Azimuth offset-based horizontal displacements for the 1999 M_W 7.5 Chi-Chi, Taiwan continental thrust earthquake derived from ERS C-band data from 1/21/99 to 9/23/99. Initial offset measurements are estimated using a correlation window of 23 × 160 pixels equivalent to a resolution of 320 m. A bilinear ramp component of the displacement field is constrained to match that estimated using available GPS data. These offsets are then spatially filtered with a 3-km-wide median filter. Adapted from Levy F, Hsu Y, Simons M, LePrince S, and Avouac JP (2005) Distribution of coseismic slip for the 1999 Chi Chi Taiwan earthquake: New data and implications of varying 3D fault geometry. *EOS Transactions of the American Geophysical Union* 86: B1305.

GPS data before any seasonal signals have been removed. Conversely, if one assumes the long wavelength in the data are unconstrained and uses the GPS to constrain the model directly, then clearly the seasonal signal should be removed first. This latter approach is safest since the seasonal signal in the GPS is not completely due to surface displacements, and similarly, there are long-wavelength errors in interferograms caused by tropospheric delays and not by errors in $B_\perp$.

3.12.4.2 Propagation Delays

The atmosphere and ionosphere introduce propagation phase and group delays to the SAR signal. In repeat-track systems, the propagation effects can be more severe. The refractive indices of the atmosphere and ionosphere are not homogeneous in space or time. For a spaceborne SAR, the path delays can be very large, depending on the frequency of the radar (e.g., greater than 50 m ionospheric path delay at L-band), and can be quite substantial in the differenced phase that comprises the interferogram (many centimeters differential tropospheric delay, and meter-level ionospheric contributions at low frequencies). These effects in RTIs were first identified by Massonnet *et al.* (1993) and later by others (Goldstein, 1995; Massonnet and Feigl, 1995; Tarayre and Massonnet, 1996; Rosen *et al.*, 1996; Zebker *et al.*, 1997; Hanssen, 2001; Lohman and Simons, 2005; Gray *et al.*, 2000; Mattar and Gray, 2002). Ionospheric delays are dispersive, so
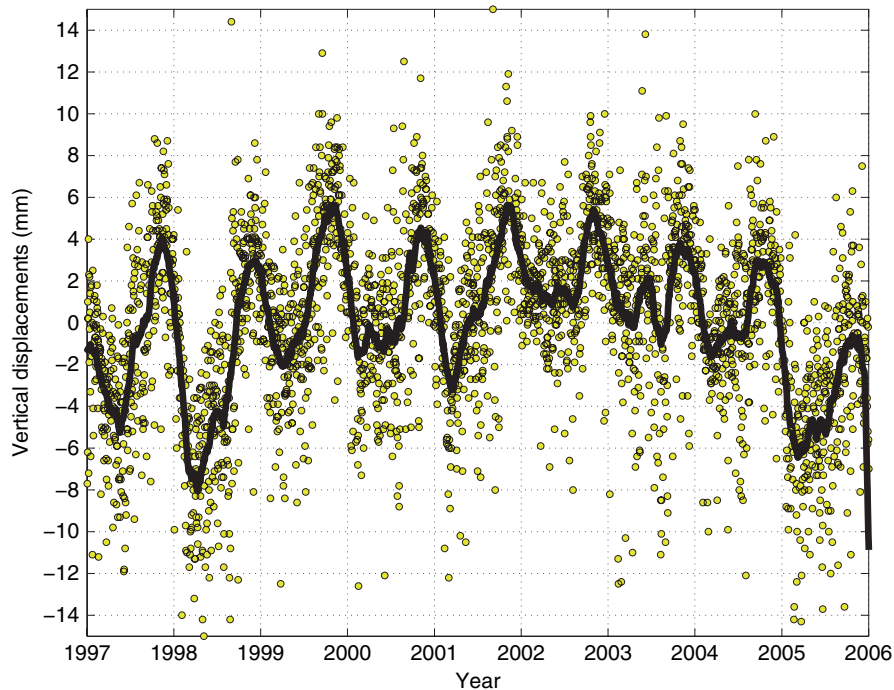
Figure 18   Dots indicate vertical GPS time series from the continuous GPS site TABL located on the central San Andreas Fault, CA. The displacements have been detrended and have the common mode removed. The solid line indicates the running 2-month average. Data from http://sopac.ucsd.edu.

frequency-diverse measurements can potentially help mitigate the effect, as with two-frequency GPS systems. Tropospheric delays are nondispersive, and can mimic topographic or surface displacement effects. Schemes for distinguishing tropospheric effects from other effects have been proposed (Massonnet and Feigl, 1995), and use of averaging interferograms to reduce atmospheric noise is common (Zebker et al., 1997; Fujiwara et al., 1998). Given even a purely horizontally stratified troposphere, one expects topographically correlated delays due to differences in total water content. Given a simple model of the distribution of water and a few points of calibrations from GPS estimates of zenith wet delay (ZWD) or from radiosondes, one can attempt to correct for this effect (Delacourt et al., 1998); unfortunately, excursions from horizontal stratification are the norm, and in fact, in any given interferogram, it is common to see phase variations that correlate with topography but with amplitude (and even sign) that vary from feature to feature.

There are a variety of more involved approaches to deal with the problem of path delays associated with spatial and temporal variations in tropospheric water vapor content. One can estimate statistically

their impact on the estimates of tectonic deformation and account for these in the data covariances when modeling. Alternatively, one can attempt to explicitly model and remove these delays.

In terms of estimating the appropriate statistics, several studies have attempted to define the variance and distance-dependent covariance of these delays (e.g., Massonnet and Feigl, 1995; Goldstein, 1995; Zebker et al., 1997; Williams et al., 1998; Hanssen, 2001; Emardson et al., 2003; Li et al., 2004, 2006b; Lohman and Simons, 2005). Using estimates of ZWD from GPS data in Southern California, Emardson et al. (2003) assume an isotropic (azimuth independent) model and find that the variance, σ, between two locations varies as $\sigma = cL^{\alpha} + kH$, where L and H are the relative horizontal and vertical separation distances between locations. The values of c, α, and k are estimated to be 2.5, 0.5, and 4.8, respectively. The value of α is expected to be more or less region independent, but the value of c may vary between regions. This model must then be scaled to account for the InSAR LOS direction. Limitations of this model include the assumption that the height dependence is independent of absolute height, and that there are no atmospheric

variations at scales smaller than the spacing in the GPS networks, which could influence the estimate of the height dependence (Xu *et al.*, 2006). Lohman and Simons (2005) find reasonable agreement between the estimates from Emardson *et al.* (2003) with empirical estimates from real interferograms (**Figures 19** and **20**). Adopting this statistical approach assumes that one has a large ensemble of interferograms to use in one's analysis. If one has only a few image pairs, then the statistics of small numbers comes into play when interpreting a specific feature in a given interferogram. Furthermore, if one removes the long wavelengths from a given interferogram, then one is also affecting the estimates of σ. These difficulties suggest that at a minimum, one should estimate the full covariance empirically, if necessary, removing an initial model of the phenomena of interest first (Lohman and Simons, 2005).

Since GPS can measure both the ZWD and its spatial gradient, it is possible to use the GPS observations to make a wet delay image to correct individual interferograms (Williams *et al.*, 1998; Xu

et al., 2006; Li *et al.*, 2004, 2006b; Onn and Zebker, 2006). This approach requires estimates of the delay made at or near the time of acquisition of the radar images. The viability of this approach clearly increases with density of GPS sites. Onn and Zebker (2006) demonstrate the use of time series of ZWD combined with a frozen flow assumption and observations of wind speed, to increase the spatial resolution of the images of wet delay beyond just the station spacing. Essentially, by assuming that the pattern of the wet delay is changing slowly, they can use estimates of ZWD before and after the image acquisition times to predict variations between GPS stations. To date, these approaches have not taken advantage of the spatial gradient of the ZWD that can also be estimated when processing the GPS delay. While the spacing of the GPS network will always remain a fundamental limitation, since the wet delay has a red spatial spectrum, any correction at medium and long wavelengths should significantly reduce effects due to tropospheric delays. Unfortunately, there are only a few, albeit important,
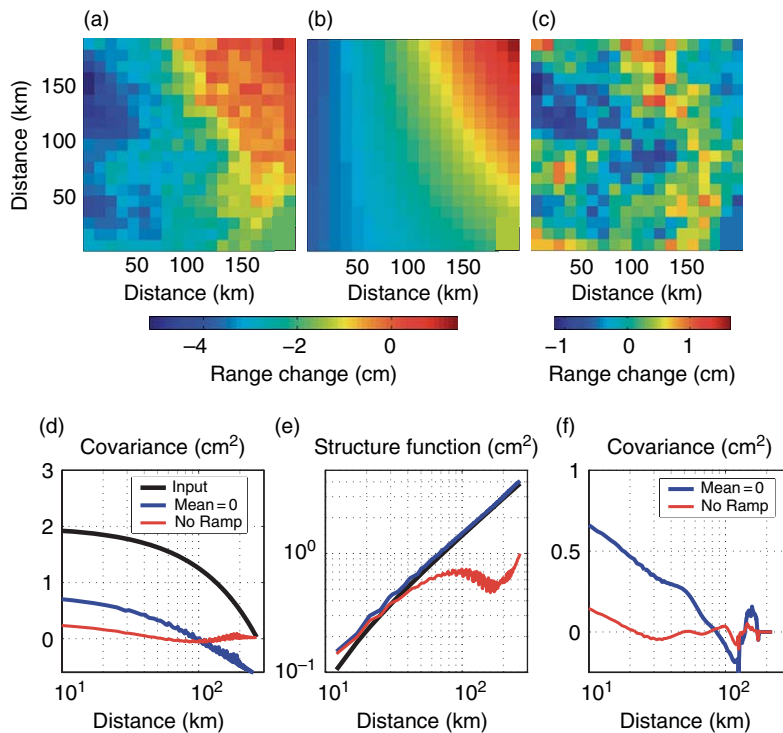
Figure 19 Effect of ramp removal on character of noise. (a) One realization of noise with power-law structure function. (b) Best-fit quadratic ramp to (a). (c) Remaining signal after ramp removal. (d) Input covariance (thick line) and inferred covariance for synthetic noise in (a) after removal of mean (blue) and quadratic ramp (red). (e) Structure function for same scenario. (f) Covariance inferred for a real interferogram, after removal of the mean and with and without the removal of the best-fitting quadratic ramp. Figure from Lohman RB and Simons M (2005) Some thoughts on the use of InSAR data to constrain models of surface deformation. *Geochemistry Geophysics Geosystems* 6 (doi:10.1029/2004GC00084).
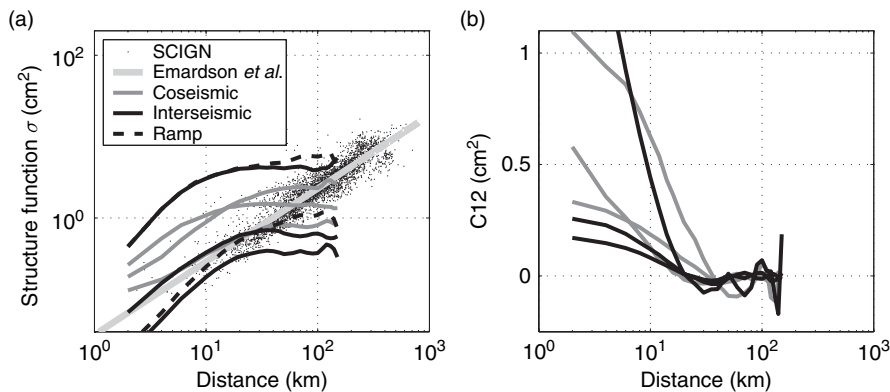
Figure 20   Spatial structure function versus distance for interferograms spanning the Hector Mine earthquake (black), and interferograms spanning time periods before or after the earthquake (gray). Dashed lines indicate the structure function for the interferogram before a quadratic ramp was removed. GPS data (small dots) are from the Southern California Integrated GPS Network (SCIGN) (adapted from Emardson et al. (2003)). The thick gray line indicates the power-law fit from Emardson et al. (2003). (b) Covariance versus distance for the coseismic and interseismic interferograms from (a). (a) Adapted from Emardson TR, Simons M, and Webb FH (2003) Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal of Geophysical Research* 108 (doi:10.1029/2002JB001781).

regions of the world where the GPS network is sufficiently dense to make any of these methods viable.

The lack of dense GPS networks in all regions of geodetic interest motivates the need for satellite-based estimates of the wet delay. Significant progress has been made to this end using observations from near-infrared imaging radiometers such as NASA's MODIS (Li *et al.*, 2005) and the MERIS instrument on ESA's ENVISAT satellite (Li *et al.*, 2006a; Puyssegur *et al.*, 2007). The latter has the advantage of estimating the water delay in the same viewing geometry and at the same time as the acquisition of the radar data. Unfortunately, imaging radiometers are limited to use in the daytime and cloud-free conditions. However, if clouds are not pervasive, then the resulting holes can be interpolated. Thermal infrared measurements could potentially be used for night-time observations. Ideally, use of image-based estimates would be done in combination with GPS estimates of wet delay where and when available.

Thus far, we have only addressed direct measurements of the wet delay. Recently, progress has been made using high-resolution weather models to estimate the spatial variation in wet delay (Foster *et al.*, 2006; Puyssegur *et al.*, 2007). Foster *et al.* (2006) demonstrate the use of the MM5 weather model (Grell *et al.*, 1995) to correct interferograms from the big island of Hawaii. In this case, the model is updated every 12 h using local meteorological data,

and estimates of wet delay are made every hour and interpolated to the time of acquisition for each radar image. Using this model, they can generally reduce the variance by about a half at wavelengths of 30 km, although they find limitations in the model in regions of high topographic gradient. Puyssegur *et al.* (2007) go one step further and combine the MM5 model with MERIS observations. In the near future, where available, imaging radiometer and GPS wet delay observations will be combined with physical models to produce the best estimates of wet delay possible. Indeed, one would expect that any future dedicated InSAR mission would produce and distribute estimates of wet delay for each image acquired. Clearly, once the best estimates of wet delay are removed, it is straightforward to empirically estimate the residual covariance structure on an interferogram-by-interferogram basis (Lohman and Simons, 2005).

3.12.4.3   Stacking Single-Component Images

Independent of whether or not one can correct interferograms for path delay effects, if the primary geophysical target is a single event that occurred quickly (in terms of satellite InSAR, this implies anything lasting less than the orbit repeat time) or alternatively is a gradual process occurring at constant rate, it may be useful to increase the signal-to-noise ratio by stacking (i.e., averaging) multiple

interferograms. Such stacks have the advantage of reducing effects due to tropospheric delays since these are uncorrelated on timescales of more than a day (e.g., Hanssen, 2001; Emardson et al., 2003) and thus are useful to discover small signals (e.g., Peltzer et al., 2001; Lyons and Sandwell, 2003; Gourmelen and Amelung, 2005; Pritchard et al., 2006). Stacking images can also reduce computational burden in parameter estimation schemes by reducing the amount of observations (e.g., Pritchard et al., 2006).

These stacks can be made either in radar coordinates (choosing one interferogram as the master) or in geocoded (latitude and longitude) coordinates. Before stacking, one needs to account for long-wavelength errors or phase ramps associated with errors in estimates of the interferometric baseline for each interferogram. The phase ramp is usually parametrized as a bilinear or biquadratic polynomial function of azimuth and range (or geographic coordinates). Neglecting to deal with the ramp can result in biased stacks. If several GPS displacement vectors are available for the region of interest, they can be used to resolve the ramp parameters either before or after forming the stack.

The simplest and most common brute force stack is made by just adding all the interferograms together. In this approach, regions of decorrelation in the resultant stack will be the union of decorrelated regions in the individual interferograms. Therefore, one may choose not to include an individual interferogram if it has an excessively large amount of decorrelation. An example of such a single LOS image stacks for a short-duration event comes from images of the coseismic displacements from deeper and smaller earthquakes in Chile (Pritchard et al., 2006). In most cases, it is advisable to take a more formal approach to stacking.

When considering making stacks, one should separate the case of a single rapid event from that of linear secular deformation. In the case of a single rapid event, the stack is made in terms of displacements. Scenes are averaged together in a straightforward fashion – ideally, using a weighted average, where the weighting is the inverse covariance matrix of the observations.

The full covariance is constructed of three primary parts: (1) the variance of a given pixel on the diagonal, (2) the intra-image pixel-to-pixel covariance primarily due to ionospheric and tropospheric path delays, and (3) intra-image covariances associated with use of a single common image in two separate interferometric pairs (Emardson et al., 2003). Assuming

ergodicity, one estimates the pixel variance empirically using estimates of the local phase variance over a small patch of pixels. Doing this for each image is important since different interferograms or offset images may have had different amounts of spatial filtering applied to them, as well as different amount of temporal decorrelation. The intra-image pixel-to-pixel covariance should also be computed empirically since the degree of filtering and long-wavelength ramp removal may vary from image to image.

Usually, this covariance is assumed to depend only on the relative distance between two given pixels (Hanssen, 2001; Emardson et al., 2003; Lohman and Simons, 2005). In contrast to the inter-image pixel covariance, the intra-image covariance can be derived analytically (Emardson et al., 2003). When calculating a displacement stack with the full covariance matrix, one would normally pose the problem as a least-squares problem. The use of a purely least-squares approach relying on an L2 norm, is somewhat debatable since phase noise is not necessarily Gaussian. It may in fact be more robust to use an L1 norm, which is equivalent to asking for the median and not the mean displacement. Regardless of the adopted norm, using a formal approach that includes the local estimate of variance permits one to include all available data regardless of the level of decorrelation.

For a constant rate process, the estimation of the rate at any given pixel is best described using a bit more formality. The underlying equation is simply

$$\rho_i = T_i v + \epsilon_i \qquad [48]$$

where ρ_i, T_i, v, and i are the observed range change, time span, rate, and error, respectively, for a pixel in the ith image. We can rewrite this equation as the linear system for the ensemble of images:

$$d = \theta v + E \qquad [49]$$

where d, θ, and E are the vectors of observations, time spans, and errors, respectively. This system has a general weighted least-squares solution

$$v = (\theta^t W \theta)^{-1} \theta^t W d \qquad [50]$$

where as described earlier, the weight matrix, W, is the inverse of the full data covariance matrix. This approach naturally deals with regions that are decorrelated in some but not all of the images. The same issue described above relating to L2 versus L1 norms applies here.

While not recommended, one could assume that the pixel variances are constant within a given displacement image and between different displacement images, as well as ignore all the covariances. These frequently adopted but unnecessary simplifications lead to the explicit least-squares solution:

$$v = \frac{\sum_{i=1}^{N} \Delta T_i \Delta \rho_i}{\sum_{i=1}^{N} \Delta T_i \Delta T_i} \qquad [51]$$

In a study of postseimic deformation from historic earthquakes in the Basin and Range Province of the western US, Gourmelen and Amelung (2005) remove the long-wavelength ramp from individual component interferograms, then add the residual interferograms up and divide the total displacement by the summed total time span of all the interferograms. This is equivalent to converting all the individual interferograms to rates, then averaging the individual rates, weighted by their respective time spans. In other words

$$v = \frac{\sum_{i=1}^{N} \Delta \rho_i}{\sum_{i=1}^{N} \Delta T_i} = \frac{\sum_{i=1}^{N} (\Delta \rho_i / \Delta T_i) \Delta T_i}{\sum_{i=1}^{N} \Delta T_i} \qquad [52]$$

Equation [52] is only the same as the least-squares solution shown in eqn [51] if all the interferograms span equal time periods. In that special case, then

$$v = \frac{1}{N \Delta T} \sum_{i=1}^{N} \Delta \rho_i \qquad [53]$$

In a study of slow interseismic deformation in the Eastern California Shear Zone, Peltzer *et al.* (2001) effectively adopted eqn [53]. Given that they use interferograms of approximately the same duration (about 4 years), they simply averaged the rates (unweighted by time), which is close to the least-squares solution assuming constant and uncorrelated data errors.

In general, given today's computational resources, there is no reason not to use the full weighted least-squares estimation (e.g., eqn [50]). In particular, this approach allows a more rigorous estimate of both rates and their associated errors. When a final stack is complete, if desirable, it is then possible to create a mask based on the final estimated variances.

3.12.4.4 InSAR Time Series

We can adopt more complicated models than the linear one just described. For instance, given the potential for seasonal signals (e.g., **Figure 18**; Amelung *et al.*, 1999; Hoffmann *et al.*, 2001; Schmidt and Burgmann, 2003), it may be desirable to augment eqn [48] by explicitly including a seasonal variation in the estimation process. Indeed, given enough temporal sampling, there exists the possibility for a wide range of time-series approaches, effectively mimicking all the approaches adopted in GPS analysis.

There are a couple of variants to current GPS analysis approaches. The first approach aims to make the cleanest time series possible for later mechanical modeling. This approach can include estimates of site-specific signals including seasonal signals (frequently assumed to be sinusoidal), random walk and flicker noise at each site, and spatially correlated processes such as daily variations in reference frame estimates (e.g., Dong *et al.*, 1998, 2003, 2006). Generally, the temporal and spatial analyses are separated in these approaches. A more complex underlying temporal evolution can also be adopted including a superposition of behaviors including linear rates, coseismic steps, and postseismic exponential or logarithmic decays. At a minimum, these more complicated model terms reduce biases in estimating seasonal and reference frame contributions and can always be added back in for subsequent mechanical modeling. Of course, the inferred model terms, such as coseismic steps, are themselves useful as input into subsequent mechanical models.

An alternative approach uses data from a network of GPS sites to simultaneously solve for time-varying parameters of a specified mechanical model and the aforementioned nontectonic signals. This approach is exemplified by the extended network inversion filter (ENIF) (Segall and Matthews, 1997; McGuire and Segall, 2003; Fukuda *et al.*, 2004).

The parallels between GPS and InSAR time-series analysis are numerous. Obviously, the physical processes of interest are the same, and they are sensitive to similar nontectonic processes such as seasonal effects. Other similarities exist such as the ambiguity in absolute displacements with InSAR data, which is equivalent to reference frame error in GPS. Despite these parallels, InSAR time-series analysis has a few unique challenges. In terms of nontectonic signals, we may want to remove the effects of the troposphere (empirically or by physical model) and we have errors due to inaccurate removal of topographic effects stemming from inaccurate orbits. With typical suites of interferograms from a given orbital track, we may also face incompletely connected chains of dates. For instance, given images acquired at times *A*, *B*, *C*, and *D*, we may be able to

form interferograms I_{AB}, I_{BC}, and I_{CD}, or perhaps we only have I_{AC} and I_{BD}. In the later case, we would like to be able to construct a continuous time record.

To deal with these issues, we have the same variants in modeling approaches for InSAR time series as just described with GPS data. One approach attempts to decompose a suite of interferograms into its component time intervals (e.g., *AB*, *BC*, and *CD*), and to solve for the incremental displacements. Ambiguities in combining unconnected pairs (or collections of pairs) requires some form of regularization of the time series. One approach, based on use of singular value decomposition (SVD), finds a time series that fits the available interferograms while minimizing the implied velocities in any given underlying time interval (Berardino *et al.*, 2002). This approach is equivalent to minimizing the temporal gradient of the deformation field. Alternatively, one could use explicit Laplacian damping to minimize the roughness of the temporal evolution. The approach described by Berardino *et al.* (2002) begins with unwrapped interferograms, and assumes that they have been tied to some stable reference point to define a common phase bias. This approach is a pixel-by-pixel algorithm at its core (important for computational purposes), but they apply subsequent spatial and temporal filtering to provide cleaner time series. An SVD or principal component approach may also be useful to isolate atmospheric signals in large InSAR time series from spatially and temporally coherent deformation fields (Ballatore, 2006). Examples of models using large time series of InSAR data include deformation at volcanoes (Lundgren *et al.*, 2001, 2004), in the Santa Clara Basin (California) (Schmidt and Burgmann, 2003), in the Los Angeles Basin (California) (Lanari *et al.*, 2004a), and faulting in the Asal Rift (Djibouti) (Doubre and Peltzer, 2007).

As with GPS data, an alternative approach to time-series modeling involves explicit use of a physical model. For example, Pritchard and Simons (2006) studied time-dependent postseismic slip and seismic slip on the subduction zone megathrust in northern Chile assuming a simple model of time-dependent slip on a fault plane, and imposing Laplacian smoothing in time in order to tie together unconnected groups of interferograms. Alternatively, they could have used a Kalman filter approach, analogous to the ENIF approach used with GPS data; however, the number of interferograms available did not support the increased complexity of the modeling approach. Of course, the mechanical modeling approach benefits directly from the availability of any additional temporally continuous data such as can be provided by GPS observations.

3.12.4.5   Vector Displacements

Another form of stacking can be useful when data from multiple viewing geometries are available. In particular, we may want to (1) compare geodetic imaging data directly with other single-component geodetic data, such as leveling or electronic distance measurements (EDM), without going through a physical model, (2) get a better grasp on what are purely vertical versus horizontal displacements, (3) as in the single-component stacks described above, combine as many images of the geophysical event to reduce noise and look for unexpected (undiscovered) processes, and (4) reduce the total amount of data used in a parameter estimation task. In these cases, it may be useful to combine displacement images from three or more viewing geometries to construct the best estimate of the east, north, and vertical displacement field. The input data need not be homogeneous in type. For instance, one could combine right-looking images from ascending and descending orbits and pixel offset estimates from one or more tracks to resolve the full 3-D displacement field (Fialko *et al.*, 2001a, 2005). Similarly, one could use observations from different overlapping beams of adjacent orbital tracks (Wright *et al.*, 2004). The whole system would then be weighted by the covariance W described in the previous section. If sufficient data are available, then one can explicitly include estimation of ramp parameters, which as before, are best constrained if at least a few independent geodetic observations are available. Estimates of the full 3-D coseismic displacement field exist for several earthquakes including the 1999 M_w 7.1 Hector Mine earthquake (Fialko *et al.*, 2001a) and the 2003 M_w 6.5 Bam earthquake in Iran (Fialko *et al.*, 2005; Funning *et al.*, 2005) (**Figure 21**). With respect to future mission design, Wright *et al.* (2004) points out the reduced sensitivity of 3-D deformation maps to north–south motions for missions with near-polar orbits.

While we may not need to explicitly make 3-D decompositions of the displacement fields, having multiple components of the displacement fields are clearly important to constrain physical models such as earthquake and volcano source models (Fialko *et al.*, 2001b; Lohman *et al.*, 2002). For example, for all earthquake models, when only one LOS component is observed, there will be a tradeoff between the
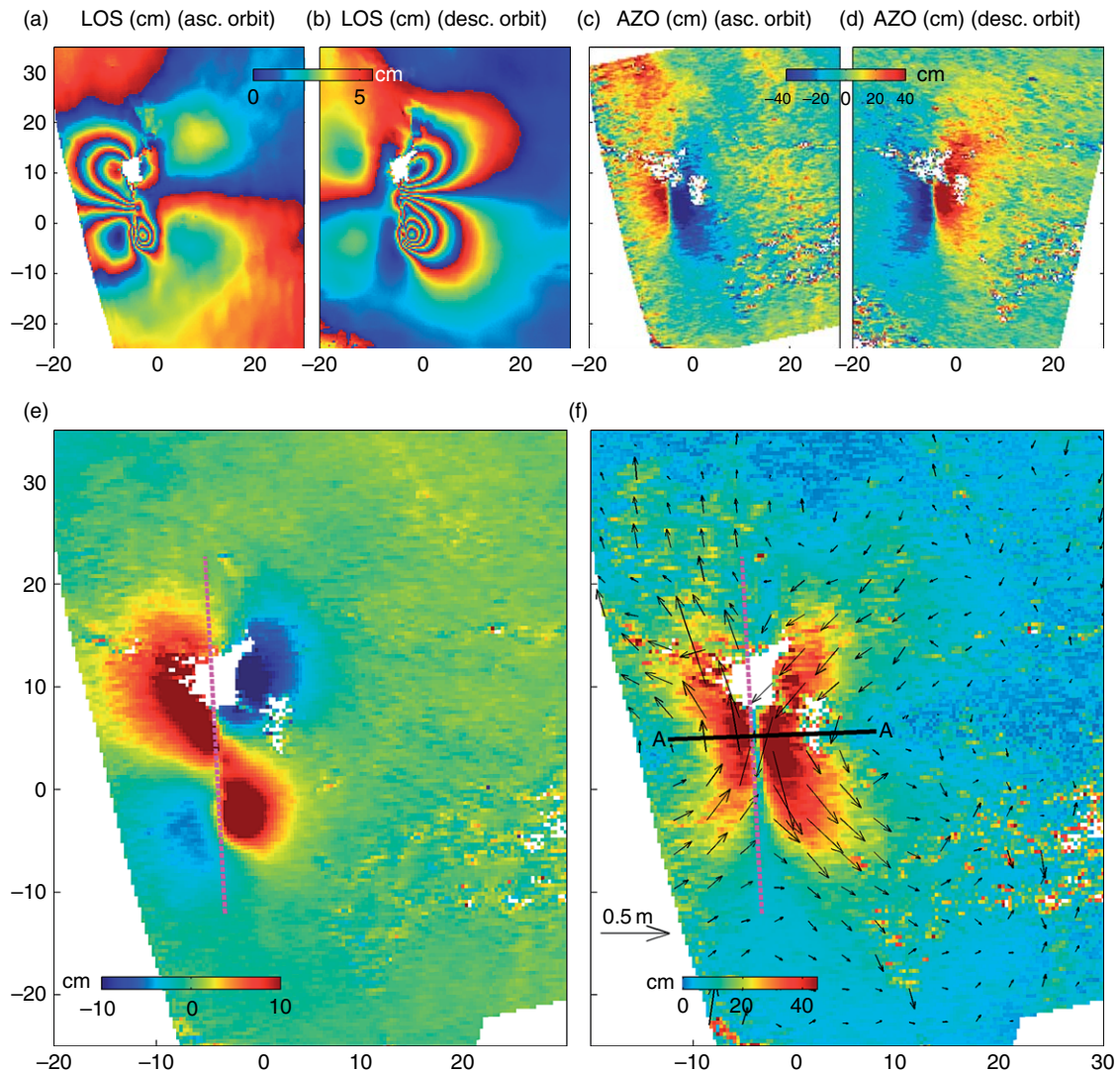
Figure 21 Phase and offset measurements for the 2003 M_W 6.5 Bam, Iran shallow strike-slip earthquake. Images are from the C-band ENVISAT ASAR instrument. The coordinate axes are in kilometers, with the origin at 58.4° E, 29 N. Colors denote displacements in centimeters. (a) Interferogram for 11/16/2003–1/25/2004, ascending orbit. (b) Interferogram for 12/3/2003–2/11/2004, descending orbit. (c) Azimuthal offsets, ascending orbit. (d) Azimuthal offsets, descending orbit. Derived vertical (e) and horizontal (f) components of the surface displacement field. Arrows show the subsampled horizontal displacements. Dashed line shows the surface projection of the fault plane inferred from the inverse modeling of the data. Figure from Fialko Y, Sandwell D, Simons M, and Rosen P (2005) Three-dimensional deformation caused by the Bam, Iran, earthquake and the origin of shallow slip deficit. *Nature* 435: 295–299.

amplitude and rake of the fault slip. This tradeoff will be further exasperated for small earthquakes if one does not know the location well (Lohman *et al.*, 2002). Generally, we note that it may be dangerous to over-simplify the problem by assuming *a priori* that the deformation field is either purely vertical or purely horizontal in nature. For instance, in large strike-slip systems, there is a temptation to assume that the near-fault displacement field is purely fault parallel.

However, even the largest strike-slip systems frequently have vertical or fault normal deformation associated with them that is either tectonic in origin (e.g., Klinger *et al.*, 2005) or due to hydrologic effects associated with faults acting as flow barriers (Bell *et al.*, 2002). The presence of any vertical displacements is particularly problematic, given the high sensitivity to vertical deformation in most InSAR data. Indeed, any ignored vertical displacement will

corrupt estimates of the horizontal displacement, where the amplitude of this corruption is amplified by at least $1/\tan\theta$, where θ is the angle of incidence. This effect increases as the fault strike approaches the azimuth of the orbital track.

### 3.12.4.6   The Luxury of Sampling – Rationale and Approach

Geodetic imaging can produce a very large number observations – of order 10^6 pixels. In many cases, one wants to use these observations to constrain parameters from a mechanical model, such as distributed fault slip in an elastic space. In practice, the computational expense of the inverse problem can be computationally bound by the number of Green functions one need calculate. In the case of linear problems where the source geometry is known, the Green functions may not be known analytically (e.g., as with 3-D variations in material properties) or one may still want to impose nonlinear constraints, both of these requirements can be expensive to implement. Even worse, for nonlinear problems, such as when we do not know the geometry of the source, then we must recalculate the Green functions at each iteration.

Having spatially continuous observations provides us the opportunity to use a selected subset of these observations. This possibility leads to the question of the optimal subset of observations to pick. It is most precise to view this problem as spatially variable averaging and sampling. Proposed approaches fall in two classes: in the first, the image sampling is done based on properties of the data themselves (Jonsson *et al.*, 2002; Simons *et al.*, 2002), and in the second, the sampling is controlled by the character of the model (Lohman and Simons, 2005). Jonsson *et al.* (2002) and Simons *et al.* (2002) proposed similar approaches that rely on successive subdivision of the deformation image. In both approaches, the image is cut into quadrants and a low-order best-fit surface is removed from the phase field in each quadrant. For a given quadrant, if the residual is greater than a prescribed threshold, the quadrant is further divided into four new quadrants, with the process allowed to continue until a minimum quadrant size is reached. For both approaches, one should then use the mean (or median) of the pixels in the quadrant and assign the value to the center of the quadrant. Jonsson *et al.* (2002) and Simons *et al.* (2002) differ in that the former considers the residual after removing the mean from each quadrant, why as the latter removes a bilinear function. Both approaches work well, but for the same number of points, the Simons *et al.* (2002) approach does

a better job at constraining a given fault slip model. Underlying the Simons *et al.* (2002) approach is the recognition that any physical model can at a minimum produce a linearly varying displacement field, and thus the ability to constrain detailed behavior of a given model lies in the curvature of the deformation field. In essence, we are attempting to choose a data set that includes as many points as possible while maintaining a nearly diagonal data resolution matrix. Both of these approaches are limited by the inherent noise in the observations, and can give spurious regions of high sampling (e.g., far away from a target fault) if there are unwrapping errors or regions of decorrelation.

An alternative approach to image sampling involves using a best-guess initial model parametrization (Lohman and Simons, 2005). A given linear model can be written as $Gm = d$, where G is the design matrix, m is the vector of model parameters, and d is the vector of observations. This has a generalized solution $m_{est} = G^{-g}d$, where G^{-g} is the generalized inverse and the model and data resolution matrices, R and N, can be written as $R = G^{-g}G$ and $N = GG^{-g}$ (e.g., Menke, 1989). Lohman and Simons (2005) propose to find a distribution of samples that has the most points while keeping N nearly diagonal. The algorithm is similar to that of Simons *et al.* (2002) and Jonsson *et al.* (2002) in the use of successive division into quadrants, with the difference being that at each stage, an estimate of N is made, and refinement stops when N becomes sufficiently nondiagonal. As with the other approaches, for any given quadrant, this approach uses the weighted mean (or median) value, where the weight takes into account the pixel variances and covariances. This approach is ideal if one has a reasonable first guess at the model geometry. It is of course also sensitive to the assumed model parametrization – in this case the size of each fault patch. Regardless, all these schemes permit one to constrain models with about 1% of the original data without significant loss of information. It is worth emphasizing that even where the final sample spacing is large, the data variance will be relatively low, since more points go into this estimate than in regions of finer spacing, thus information is not lost.

These spatially variable averaging/sampling approaches are particularly important when modeling shallow sources with the potential of causing complex deformation patterns. A common example of this class of problems comes from shallow earthquakes. **Figure 22** demonstrates the difference in sampling patterns that result from the different approaches described here. This figure also
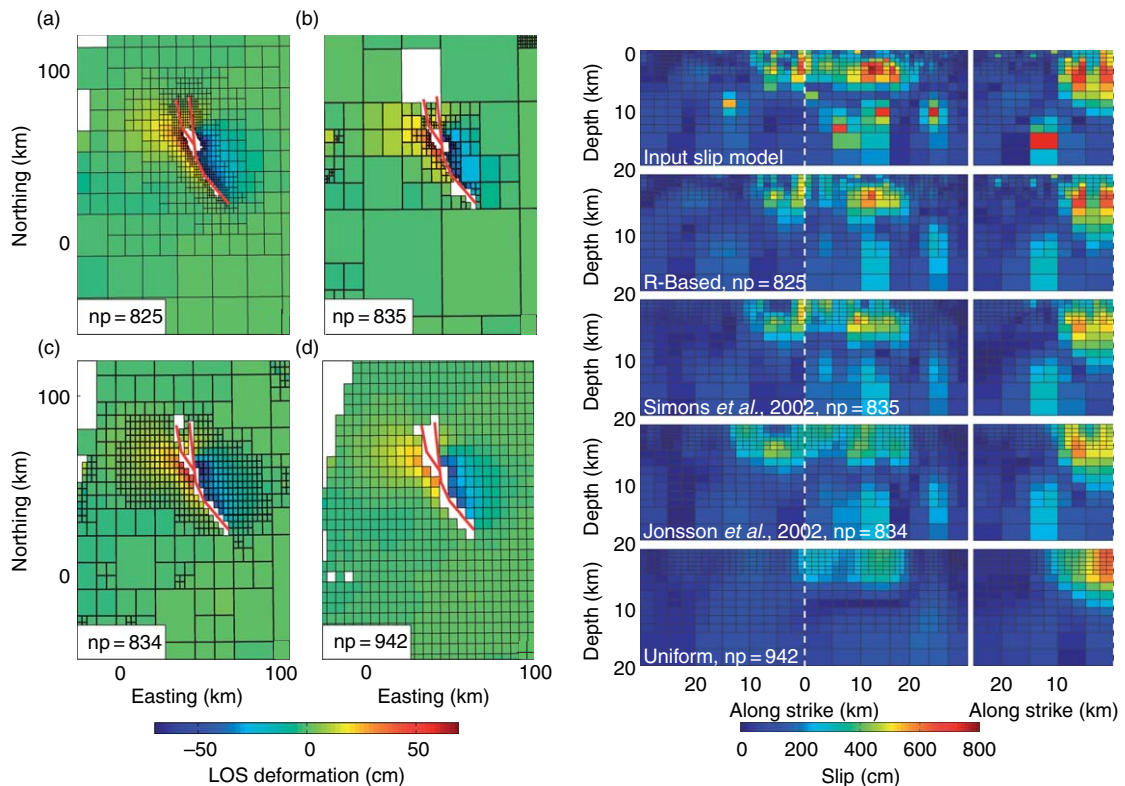
Figure 22 Left: Examples of different interferogram sampling approaches. Results are shown using (a) the data resolution-based algorithm from Lohman and Simons (2005), (b) a variable sampling proportional to data curvature approach from Simons *et al.* (2002), (c) a variable sampling proportional to data gradient Jonsson *et al.* (2002), and (d) uniform sampling. Right: Comparison of inferred slip distribution when using the different sampling approaches shown at left. In each case, the total number of points retained is similar. The fault geometry is shown at left by the red lines. The inferred slip is similar for all inversions, but the resolution of shallow slip features increases from bottom to top. Lohman RB and Simons M (2005) Some thoughts on the use of InSAR data to constrain models of surface deformation. *Geochemistry Geophysics Geosystems* 6 (doi: 10.1029/2004GC00084).

demonstrates the difference in the ensuing model resolution. The choice of algorithm clearly impacts the ability to distinguish variations in the shallowest parts of the model (Lohman and Simons, 2005).

3.12.4.7   Decorrelation as Signal

Interferometric temporal decorrelation is usually viewed at least as a nuisance and sometimes as a complete barrier to making useful displacement measurements. However, in some cases, the spatial distribution of decorrelation, as well as its temporal evolution, can serve as useful indicators of geological processes. There are many causes of temporal decorrelation that are not necessarily of interest, among these are the impact of weather-related processes (rain, snow, etc.). Of more interest is decorrelation caused by rearrangement of the scatterers within a pixel associated with intense shaking,

damage, etc. Examples of the use of such decorrelation include mapping of active lava flows (Zebker *et al.*, 1996), as well as near-fault decorrelation near shallow earthquake ruptures in the 1995 Kobe, Japan earthquake (Rosen *et al.*, 2000), the 1999 Hector Mine, California earthquake (**Figure 23**; Simons *et al.*, 2002), and the 2003 Bam, Iran earthquake (Fielding *et al.*, 2005). This latter form of decorrelation has been found to clearly outline fault traces that slip in earthquakes and can serve as an obvious way to constrain surface fault geometry, and so should be used to guide postearthquake field studies. In general, InSAR decorrelation images provide a synoptic view of the spatial extent (along-strike and cross-strike) of faulting that is not easily achieved from the ground.

A more important potential societal benefit of decorrelation measurements is their use immediately after an earthquake, landslide, or eruption to map out
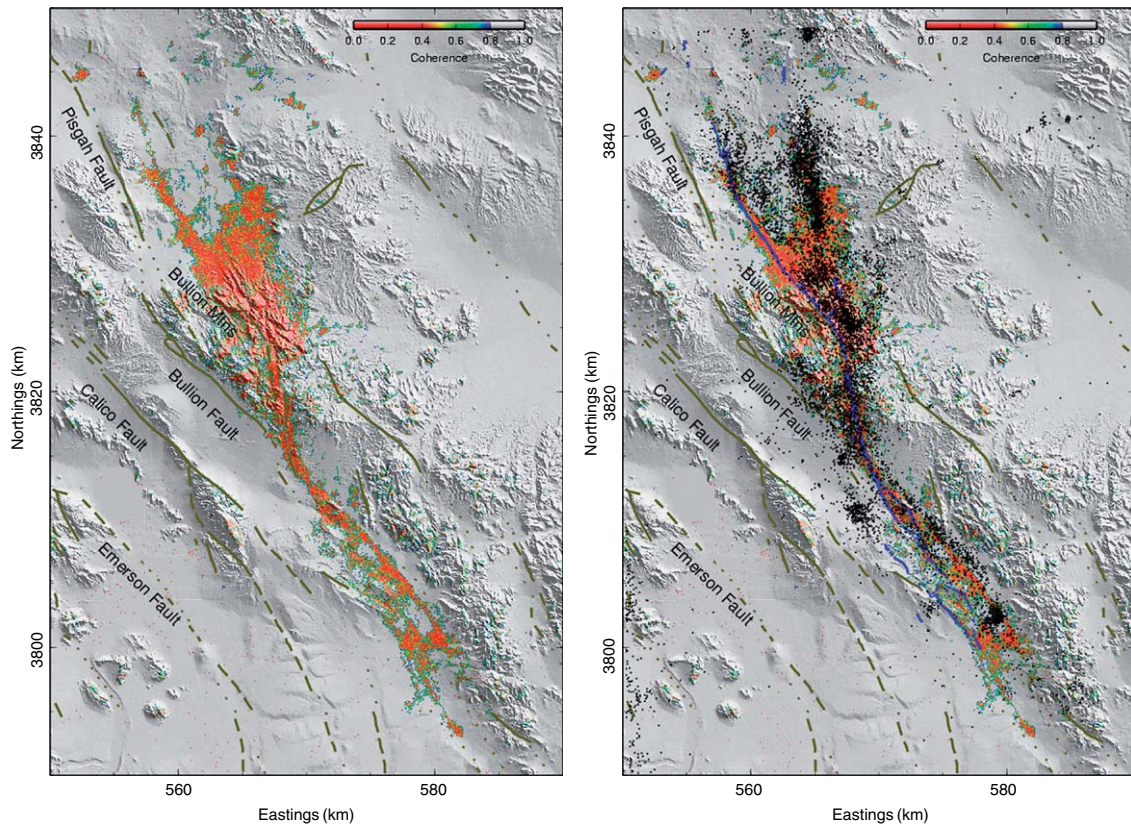
Figure 23 Interferometric decorrelation from the 1999 M_w 7.1 Hector Mine, California earthquake. Only areas with correlation less than 0.8 are shown. Images are from the C-band ERS satellite and span 35 days. Same as left but with aftershocks and the fault trace as mapped in the field. Figure modified from Simons M, Fialko Y, and Rivera L (2002) Coseismic deformation from the 1999 M_w 7.1 Hector mine, California, earthquake as inferred from InSAR and GPS observations. *Bulletin of the Seismological Society of America* 92(4): 1390–1402.

regions of damage (e.g., Fielding *et al.*, 2005). Frequently, existing communication systems are incapacitated by the event, and local inhabitants are not able to report the level of devastation. Rapidly made decorrelation maps produced during or soon after the event could be used to guide emergency response.

The usefulness of a decorrelation map is limited by 'uninteresting' sources of decorrelation not associated by the event. In particular, all forms of temporal decorrelation, especially if not monotonic, can mask the correlation image. Obvious examples include rain, snow fall, and vegetation changes occuring during the time span of the image pair. The most important ways to improve the usefulness of decorrelation measurements include rapid revisit times to limit the possibility of extraneous processes and the use of long-wavelength InSAR (e.g., L-band as opposed to C-band).

3.12.5   The Link between Science and Mission Design

InSAR and other space geodetic techniques are primarily designed to measure the displacement of Earth's surface over time. The particular characteristics of the measurements are tied to the specific implementation of the InSAR instrument and the mission characteristics. Clearly, there is a close link between the science that can be done with an InSAR mission and the design of that mission. It is important for the research community to understand this link in order to accomplish their goals. A simple example of this would be avoiding the use of data from a mission that provides an image over an area once per month, when the phenomenon of interest has changes that occur on timescales of days. In this section, we describe the flow from scientifically driven measurement needs to the basic parameters of an InSAR

mission. With such an understanding, it is then possible to characterize the performance of existing missions and productively discuss the design trades for future systems.

It is worth stating that InSAR measurements are a poor proxy for what scientists really would like to know about geophysical systems. The desired starting point to address larger geophysical questions – for example, what are the mechanics of earthquakes and how do fault systems interact? – would be measurements of the state of the crust, its pressure, temperature, and distribution of material properties throughout medium, to use as input to geophysical models. These *in situ* measurements at depth are impossible to obtain, so scientists model them through observations of the motions of Earth's surface. So it is important initially to understand the sensitivity relationship between model parameters and surface displacements. For example, if any reasonable change in a model parameter changes the modeled surface displacement by 1 mm, there is limited value in measuring deformation to only 1 cm accuracy.

So the question arises: what can we reasonably expect to measure with an InSAR system, and conversely, what is required of that system in order to advance science? A repeat-pass InSAR system measures the range displacement of any image element through a differencing of the phase from one epoch to another. Using a time series of observations, a repeat-pass InSAR system measures a spatial distribution of range displacements in discrete time intervals. To use InSAR displacement measurements in geophysical models, the measurements must have adequate displacement accuracy, both absolute and relative, spatial resolution, spatial coverage, and temporal sampling. These requirements differ for each specific scientific investigation.

For a given system, we have seen that the accuracy of the range displacement measurement is determined by noise induced by decorrelation of the radar echoes, by random phase delays introduced by propagation effects through the time-variable atmosphere, and by systematic knowledge uncertainties in the radar path delays and orbit. In addition, the incidence angle and azimuth angle of the observation can affect the accuracy of the measurement greatly; measuring a horizontal displacement with a system that looks steeply down toward nadir is not desirable.

As described earlier, decorrelation is comprised of principally three components: thermal decorrelation is related to the noise level relative to the signal level

of the radar system; geometric decorrelation is related to the arrangement of scatterers on the surface and how they change with differences in time or viewing geometry; and other decorrelation terms that derive from noise (e.g., quantization, ambiguities, sidelobes) that is dependent on the signal level itself. There is a strong functional dependence of these decorrelation terms on system parameters such as power, antenna size, wavelength, etc., and great interplay among them.

The range resolution of a SAR system is inversely proportional to waveform bandwidth. The required range resolution is usually set by the scale size of the ground feature that is being mapped. For InSAR systems, however, there is a relationship between system bandwidth and desirable interferometric baselines. Finer resolution implies less decorrelation due to nonzero baselines. Thus, even if the final map resolution desired is only 100 m, one might require a system to have 10 m resolution in range so that the constraints on the repeat-orbit accuracy are manageable.

The along-track resolution of a conventional strip map SAR system is equal to half the along-track length of the antenna, independent of range and wavelength. This is illustrated schematically in **Figure 24**, in which the antenna size is grossly exaggerated in size relative to the antenna beam width and the range, so that the salient characteristics can be viewed on a single page. The resolution along-track is determined by the spread in frequency content a given surface element experiences as the SAR beam illuminates it. Because of its beam width, the radar signal experiences Doppler shifts across its beam, such that the received echo contains a spectrum of information. The beam width is given by $\phi_D = \lambda/L$, where λ is the wavelength of the radar tone and L is the length of the antenna along-track. The frequency bandwidth associated with the Doppler shifts within this beam width is given by $\Delta f_{D,t} = 2v\phi_D/\lambda = 2v/L$. In spatial coordinates, the spatial frequency bandwidth is $\Delta f_{D,x} = 2/L$, which implies a resolution of $L/2$. Thus, while wavelength and antenna length determine the beam width of the antenna, the spatial frequency content is not dependent on the wavelength in SAR.

For ScanSAR systems the resolution is determined by the length of the burst of pulses in a given scan, which is generally related to the number of ScanSAR beams. For spotlight SAR systems, the resolution is determined by the length of time over which the observation is made.
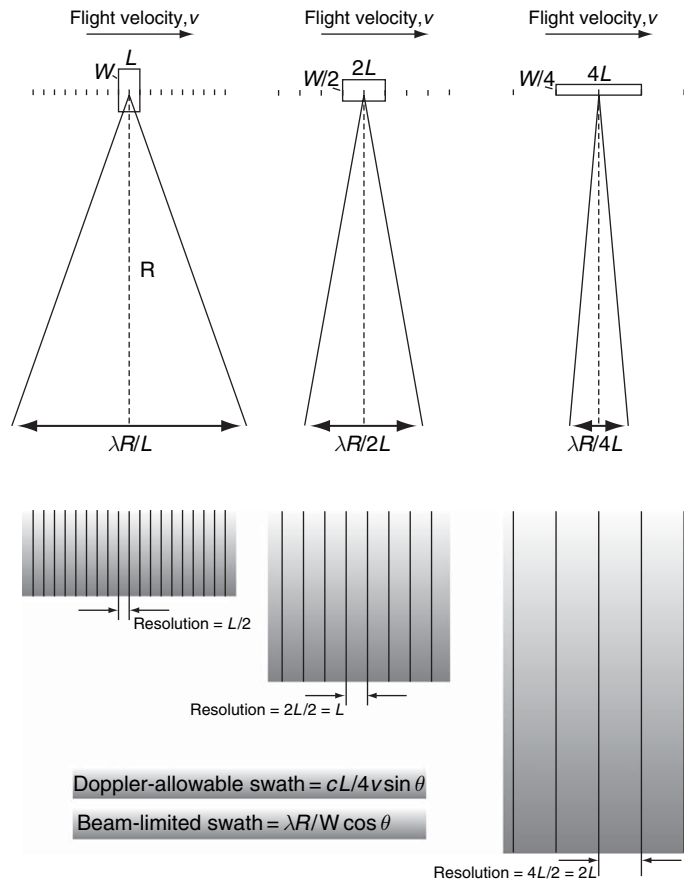
Figure 24 Illustrating the trade-off between swath extent (coverage) and resolution for strip mode SAR systems. In order to image a wide swath in strip mode, the antenna must be long in the along-track dimension. However, the resolution along track is half the antenna dimension, so the resolution is degraded as the swath width increases.

In addition to the thermal noise present in a radar system, there are a number of noise sources that play a significant role in the design of a radar and the accuracy one can expect to achieve for InSAR displacements. SAR systems point to an angle off nadir to avoid echoes from both sides of the nadir track: such echoes would be ambiguously combined in the SAR receiver and could not be distinguished from each other. Even with off-nadir pointing, the transmitted and received energy cannot be completely localized in time as the signal spreads throughout the illumination area, resulting in a wide range of time over which a given echo can return (as well as some energy from the opposite side of nadir in some cases). Because the radar transmits pulses of energy over 1000 times per second for typical space-borne systems in order to properly sample the Doppler spectrum, it can often occur that energy from time intervals outside the area of interest defined by this sampling rate, for example, from a previous or later

pulse, arrives at the receiving antenna at the same time as the desired echo energy. These corrupting echoes, generally occurring at much lower amplitude, are called range ambiguities. The magnitude of these are controlled by the pulse rate – generally lower pulse rate allows more time to collect all echoes from a pulse – and by shaping the illumination area by manipulating the shape of the antenna pattern.

On the other hand one cannot lower the pulse rate below the point where the along-track pulses become undersampled. As we have seen, to create a narrow Doppler spectrum, we desire a long antenna in the along-track direction. Physical constraints on the size of the flight system, as well as a common desire for reasonably high resolution, limit this size, and therefore limit the minimum pulse rate. The illumination pattern in the along-track direction also has extent beyond the nominal beam width (antenna sidelobes), so the pulse sampling rate will naturally cause

aliasing of some energy from beyond the along-track antenna main beam extent.

To first order, then, if a wide swath is desired, then a low pulse rate must be chosen to allow enough time between pulses for the received echo to be unambiguously acquired. This then requires that the Doppler bandwidth, and hence the antenna beam width, be narrow, which then mandates a long antenna in the along-track direction. Furthermore, the antenna must be limited in size in the elevation direction to create a wide-enough beam to receive energy from the wide swath of interest.

For a particular spacecraft altitude, the swath size is determined by these ambiguity constraints for most practical spaceborne systems. As the radar antenna of fixed beam width is pointed at greater distances off nadir, the swath illuminated on the ground becomes broader from projection effects, but the usable swath extent is usually narrower because of ambiguities. This then means that the antenna must be larger in the elevation dimension to limit the swath to maintain performance.

These effects then influence mission design for an InSAR system. Scientists studying deformation want to be able to observe any point on the Earth at some regular interval. They also would like rapid repeat coverage to be able to track rapid changes of the Earth. Suppose the mission requirement is to repeat an orbit every 8 days. For exact repeat, this requires an integer number of orbits in this time. There is an 8-day repeat polar orbit at about 760 km altitude that contains 115 orbits in 8 days. This then implies that the separation of the orbit tracks at the equator will be $2\pi 6378/115 \approx 340$ km. Thus, the SAR must be able to cover 340 km of swath, either all at once with a very wide swath from a very long and skinny antenna (giving very low resolution), or using multiple smaller beams with smaller swaths covering different angles off nadir at different times. ScanSAR, where the radar sends a collection of pulses illuminating one subswath, then electronically steers the antenna to the next subswath off nadir and sends the next collection of pulses, and so on with multiple beams, allows full coverage of the wide swath in one pass again, at the expense of resolution, and somewhat degraded ambiguity performance.

These design space possibilities – frequency, resolution, antenna size, orbit altitude and control, system power, viewing angle, repeat period, observation modality – are the playground of scientists, working with system designers, to optimize a mission to capture meaningful geophysical signals in the presence of the noise sources that are present in radar measurements. Space faring nations are increasingly relying on SAR and InSAR for scientific discovery and monitoring, with the trend moving away from large multimode systems to simpler instruments that do a few things well. One of these missions will no doubt be dedicated InSAR mission, a true geodetic instrument in the model of GPS, but with global reach and dramatically improved denser coverage of the Earth.

Acknowledgments

Appendix 1

Radar Imaging

The power of radar interferometry for geodesy stems from its high-resolution images that are generated from a controlled coherent light source. The coherence of the measurements ensures that the phase associated with each complex image element contains both the round-trip geometric path length from the radar sensor to the surface and back, and the ensemble phase associated with the coherent summation of scattering within the image element. The imaging process is quite involved and is described in detail in many other places, for example, Elachi (1988), Raney (1998), and Franceschetti and Lanari (1999). However some basic description of the process will frame the discussion of interferometry. This appendix is a compilation of the basics from a variety of sources, and is designed to cover all the essentials for understanding radar images, and their phase, in one convenient narrative.

The most common radar waveform in civilian spaceborne radars is 'chirp' coding:

$$f(t) = A\cos\left(\omega_0 t + \frac{B_r}{2\tau_p}t^2\right)\mathrm{rect}\left(\frac{t}{\tau_p}\right) \qquad [54]$$

This waveform is transmitted typically over 1000 times per second and its echo is received, digitized, and transmitted to a recorder. The presence of a point target on the ground surrounded by a surface with no inherent reflectivity causes the radar system to respond with the following received signal:

$$r(t, R) = p(R)g(R)\cos\left(\omega_0(t - 2R/c) + \frac{1}{2}\frac{B_r}{\tau_p}(t - 2R/c)^2\right)$$
$$\times \mathrm{rect}\left(\frac{t - 2R/c}{\tau_p}\right) \qquad [55]$$

where for now, $p(R)$ and $g(R)$ are generic functions to represent $1/R^2$ loss in power and backscatter at R, respectively. They will be defined with greater specificity shortly. Note if we think of the radar as a system, the point target can be thought of as a delta function input, implying that $r(t, R)$ would be the system impulse response.

If numerous point targets are arranged in range, then the total received signal at any given time will be the sum of point-target returns over these ranges:

$$r(t) = \int_R p(R)g(R)\cos\left(\omega_0(t - 2R/c) + \frac{1}{2}\frac{B_r}{\tau_p}(t - 2R/c)^2\right)$$
$$\times \mathrm{rect}\left(\frac{t - 2R/c}{\tau_p}\right)\mathrm{d}R \qquad [56]$$

This equation states that the received signal is the convolution of the ground backscatter, modulated by the geometric decay in power with range, with the transmitted chirp. Viewing the transmitted chirp as the radar system impulse response, this equation is a typical expression for the output of a linear system driven by $p(R)g(R)$.

The response of the radar to the presence of a line of scatterers in range, represented in its complex-valued form $r_z(t)$, is

$$r_z(t) = K\int_{t'} p(t')g(t')\mathrm{e}^{\mathrm{j}\left(\omega_0(t - t') + \frac{1}{2}\left(B_r/\tau_p\right)(t - t')^2\right)}$$
$$\times \mathrm{rect}\left(\frac{t - t'}{\tau_p}\right)\mathrm{d}t' \qquad [57]$$

where K is an arbitrary constant and $t' = 2R/c$. The convolution of the surface with the chirp smears out the information from any given scatterer over the extent τ_p. The goal is to recover the original $g(t')$ from $r_z(t)$.

Matched filtering

Suppose

$$p(t')g(t') = g_0\mathrm{e}^{\mathrm{j}(\omega_0 t' - 2kR_0)}\delta(t' - t_0) \qquad [58]$$

where $R_0 = ct_0/2$, implying there is a single-point scatterer on the ground at range R_0. In the radar receiver, normally one of the first hardware functions is to essentially remove the rapidly varying phase term $\exp(\mathrm{j}\omega_0 t')$ by a process called 'heterodyning'. The received signal is multiplied by a reference signal $\exp(-\mathrm{j}\omega_0' t')$, with $\omega_0 \approx \omega_0'$ so the narrowband received signal (the bandwidth is $B \ll \omega_0$) is near zero frequency, in the so-called 'video band'. (In practice there can be several stages of filtering and heterodyning to achieve the video signal, and often it is arranged so the spectrum of the real signal has its positive and negative frequencies symmetrically arranged around zero, each centered at $B/2$ and $-B/2$, respectively.) For our purposes, we let $\omega_0 = \omega_0'$, so that

$$p(t')g(t') = g_0\mathrm{e}^{-\mathrm{j}2kR_0}\delta(t' - t_0) \qquad [59]$$

with the understanding that the 'carrier' has been removed. Then

$$r_{z\delta}(t) = g_0\mathrm{e}^{-\mathrm{j}2kR_0}\mathrm{e}^{\mathrm{j}(1/2)\left(B_r/\tau_p\right)(t - t_0)^2}\mathrm{rect}\left(\frac{t - t_0}{\tau_p}\right) \qquad [60]$$

$$= g_0\mathrm{e}^{-\mathrm{j}2kR_0}h(t - t_0) \qquad [61]$$

Since $r_{z\delta}$ is the response of the radar to an impulsive 'signal' source (i.e., the ground), we recognize

$$h(t) = \mathrm{e}^{\mathrm{j}(1/2)\left(B_r/\tau_p\right)t^2}\mathrm{rect}\left(\frac{t}{\tau_p}\right) \qquad [62]$$
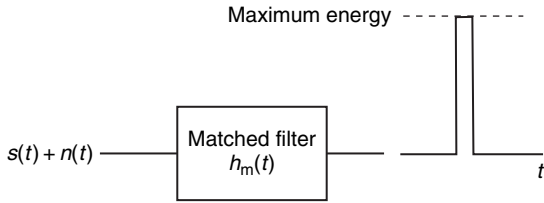
as the impulse response of the radar system. For this system, h is just the transmitted signal. The Fourier Transform of h is known as the system transfer function $H(\omega)$. One of the properties of Fourier transforms is that the convolution of functions in the time domain is equivalent to the product of functions in the frequency domain

$$r_z(t) = \int g(t')h(t - t')\mathrm{d}t' \leftrightarrow R_z(\omega) = G(\omega)H(\omega) \quad [63]$$

Since we are interested in recovering $g()$ from $r_z()$, it is clear that if we can multiply $R_z(\omega)$ by the reciprocal of $H(\omega)$, the inverse Fourier transform will retrieve $g()$. It turns out that $H(\omega)$ is of the form

$$H(\omega) \approx \mathrm{e}^{\mathrm{j}\kappa\omega^2} \qquad [64]$$

Figure 25 Illustration of the matched filter concept, where a signal in noise is optimally detected by passing it through a filter matched to its characteristics. In our case $s(t)$ is what we called the impulse response of the radar system $h(\ )$.

which implies $R_z(\omega)H^*(\omega) = G(\omega)$, that is, we can filter r_z with the conjugate spectrum of the impulse response to recover g.

This specific result is a special case from matched filter theory, where an optimal filter is designed to best detect a transmitted waveform in the presence of noise (**Figure 25**). The matched filter of a signal $s(t)$ has an impulse response

$$h_{\mathrm{m}}(t) = s^*(-t) \qquad [65]$$

In our case, $s(\ )$ is what we defined as radar impulse response $h(\ )$, so $h_{\mathrm{m}}(t) = h^*(-t)$. Thus the recovered, or 'compressed', signal r_{zc} for our delta function is given by

$$r_{zc}(t) = \int r_z(t')h_{\mathrm{m}}(t-t')\mathrm{d}t' \qquad [66]$$

$$r_{zc\delta}(t) = g_0 e^{-\mathrm{j}2kR_0} \int h(t'-t_0)h_{\mathrm{m}}(t-t')\mathrm{d}t' \qquad [67]$$

$$= g_0 e^{-\mathrm{j}2kR_0} \int h(t'-t_0)h^*(t'-t)\mathrm{d}t' \qquad [68]$$

So we see that convolution with the matched filter is equivalent to correlation with the conjugate of the function itself. This makes some sense intuitively, in that as we slide the filter along our received signal and integrate, they will match best and give the highest integral when they are aligned with lag $t = t_0$ and when all phase terms are cancelled, making the integrand always positive. Since the Fourier transform of $h^*(-t)$ is just $H^*(\omega)$, we see that the result above considering the specific form of h in eqn [64] is actually a more general result. Namely, we can in general filter r_z with the conjugate spectrum of H to retrieve g, or equivalently (**Figure 25**),

$$r_{zc}(t) = \int r_z(t')e^{-\mathrm{j}(1/2)(B_r/\tau_{\mathrm{p}})(t'-t)^2}\mathrm{rect}\left(\frac{t'-t}{\tau_{\mathrm{p}}}\right)\mathrm{d}t' \quad [69]$$

Expanding eqn [68], we have

$$r_{zc\delta}(t) = g_0 e^{-\mathrm{j}2kR_0} \int e^{\mathrm{j}(1/2)\,(B_r/\tau_{\mathrm{p}})\,(t'-t_0)^2}\mathrm{rect}\left(\frac{t'-t_0}{\tau_{\mathrm{p}}}\right)$$
$$\times e^{-\mathrm{j}\,(1/2)(B_r/\tau_{\mathrm{p}})(t'-t)^2}\mathrm{rect}\left(\frac{t'-t}{\tau_{\mathrm{p}}}\right)\mathrm{d}t' \qquad [70]$$

If we ignore the rect functions, and assume infinite integration limits, the integration becomes fairly straightforward because the quadratic phase terms involving the integration variable cancel:

$$r_{zc\delta}(t) = g_0 e^{-\mathrm{j}2kR_0}e^{\mathrm{j}(1/2)(B_r/\tau_{\mathrm{p}})(t_0^2-t^2)}$$
$$\times \int e^{-\mathrm{j}(B_r/\tau_{\mathrm{p}})(t't_0)}e^{\mathrm{j}(B_r/\tau_{\mathrm{p}})(t't)}\mathrm{d}t' \qquad [71]$$

$$= g_0 e^{-\mathrm{j}2kR_0}e^{\mathrm{j}(1/2)(B_r/\tau_{\mathrm{p}})(t_0^2-t^2)} \int e^{\mathrm{j}(B_r/\tau_{\mathrm{p}})t'(t-t_0)}\mathrm{d}t' \qquad [72]$$

$$= g_0 e^{-\mathrm{j}2kR_0}e^{-\mathrm{j}(1/2)(B_r/\tau_{\mathrm{p}})(t^2-t_0^2)}\delta(t-t_0) \qquad [73]$$

So we see that we exactly recover the delta function with an infinite bandwidth signal and impulse response. If we were to leave the impulse response bandwidth infinite but apply the windowing of the received signal, then the integration limits above are limited to the pulse extent τ_{p}, and we obtain the usual sinc result:

$$r_{zc\delta}(t) = g_0 e^{-\mathrm{j}2kR_0}e^{-\mathrm{j}(1/2)(B_r/\tau_{\mathrm{p}})(t^2-t_0^2)}\frac{\sin \pi B(t-t_0)}{\pi B(t-t_0)} \quad [74]$$

where $B = B_r/2\pi$ (**Figure 26**). The first null is at the expected location of $1/B$, defining the resolution of the compressed signal (**Figure 26**).
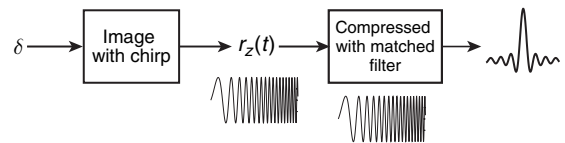
When both rect functions are considered in the integral, the quadratic phase term cancels and the sinc function is modified. The integration is carried out in three regimes as shown in **Figure 27**.
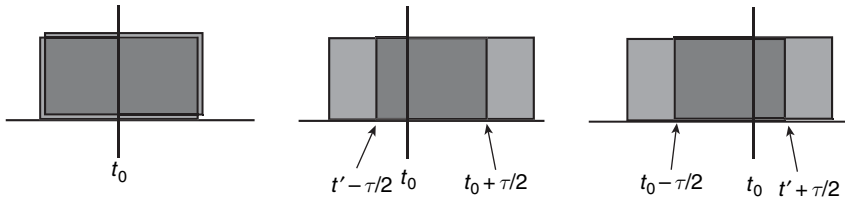
Using these limits,

$$r_{zc\delta}(t) = g_0 e^{-\mathrm{j}2kR_0}\mathrm{rect}\left(\frac{t-t_0}{\tau_{\mathrm{p}}}\right)\frac{\sin \pi B(t-t_0)(\tau_{\mathrm{p}}-|t-t_0|)}{\pi B(t-t_0)}$$
$$[75]$$

When $t - t_0 \ll \tau_{\mathrm{p}}$ this reduces to the usual expression without the quadratic phase:

$$r_{zc\delta}(t) \approx g_0 e^{-\mathrm{j}2kR_0}\frac{\sin \pi B(t-t_0)}{\pi B(t-t_0)} \qquad [76]$$

Figure 26 Illustration of the matched filter for the delta-function scene and linear FM chirp. The compressed signal is close to a sinc function.

Figure 27 The three regimes of integration as the windowed received signal and matched filter have zero, postive and negative lag. Modified from Franceschetti G and Lanari R (1999) *Synthetic Aperture Radar Processing*. Boca Raton, FL: CRC Press.

Radar signal return from a 3-D surface

We examine what comprises $p()$ and $g()$ in eqn [57] in an imaging radar system looking at the Earth, where we do not necessarily have a neat line of scatterers arranged in range. Instead we have a spatial distribution of scatterers within the beam on a 3-D topographically modulated surface (refer back to **Figure 2** for the SAR imaging geometry). The receive echo will integrate over the azimuth direction and the range pulse at any instant. For a real aperture radar, that is one for which the resolution in azimuth is dictated by the beam width of the antenna, the point-target concept extends to an integrated backscatter value over azimuth $g'(R)$ and the same arguments and equations for the simple point targets described above are obtained.

Figure 28 shows the imaging scenario looking down on the Earth from above the sensor. This more clearly shows that the received echo energy at any given instant within the pulse will effectively be the integrated return from every scatterer within the instantaneous pulse extent on the ground in range and azimuth, weighted by the transmitted chirp waveform. Because the chirp signal is extended in range, the response from any given scatterer is also extended in range. Variations in range relative to a nominal flat surface within the beam induced by topography or surface relief will lead to geometric distortion in the range trace: the measurement is fundamentally an integration (convolution) over phase fronts (i.e., constant range).

Being more specific about the function $g()$, following Raney (1998), we define the geometry of the phase front incident on a elemental surface with reference to **Figure 29**. The surface element is like the facet model, bounded in range by a distance ΔR, and presents a cross-sectional area normal to the incident wave of $da\,dz$, where $da = R\,d\phi$ and $dz = \Delta R/\tan\theta$, where θ is the look angle.

With this definition, we can define the relationship between the normalized back scattering
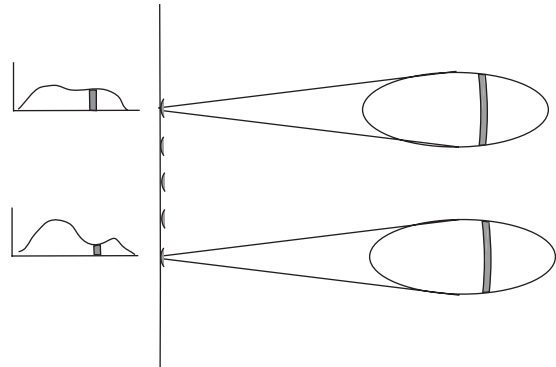
Figure 28 Radar imaging as viewed looking down at the Earth from above the sensor. The sweep of the pulse across the radar beam leads to a return echo trace for each pulse as depicted on the left. Every IPP, a pulse is sent and a trace recorded. The collection of pulses, that is, the along-track history of range profiles, is therefore a 2-D representation of the surface, albeit poorly resolved. There is considerable redundant information for any given surface element in the pulses because the beam samples each point many times as it travels along track (e.g., pulses every 5 m observed over a beam extent of 5 km). Through matched filtering techniques, a fine resolution image of the surface can be generated.
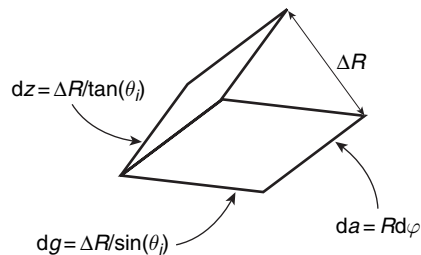
Figure 29 Depiction of a cell in range defined by two phase fronts separated by range distance ΔR. This could be the transmit pulse length in range units, for example. Modified from Raney RK (1998) Radar fundamentals: Technical perspective. In: Henderson FM and Lewis AG (eds.) *Manual of Remote Sensing Volume 2, Principles and Applications of Imaging Radar*, 3rd edn. Hoboken, NJ: J. Wiley.

cross-subsection σ_0 and the reflectivity normal to the look direction:

$$|\Gamma(a,\ R)|^2\,\mathrm{d}a\,\mathrm{d}z = \frac{\sigma_0(a,\ R)}{\sin\theta}\,\mathrm{d}a\Delta R \qquad [77]$$

This equation defines Γ. In general, the reflectivity is a complex quantity, taking into account phases from random scatterers and differential path delays within the elemental area.

The received field from this elemental area is then

$$r(t) = \frac{P_T^{1/2}G\lambda}{(4\pi)^{3/2}R^2}\Gamma(R,\ \phi)R\,\mathrm{d}\phi\,\mathrm{d}z\,p(\theta(R),\ \phi)$$

$$\times\,\mathrm{e}^{-\mathrm{j}2kR}\mathrm{e}^{\mathrm{j}(1/2)\left(B_r/\tau_p\right)(t-2R/c)^2}\mathrm{rect}\!\left(\frac{t-2R/c}{\tau_p}\right) \qquad [78]$$

At every instant $t-2R/c$ all the signals that arrive at the receiver add up. We have constructed the elemental area on the phase front such that an integration over these elements is at fixed $2R/c$. Thus, we can define the integration over the phase front to be a function of time that is a generalization of the range-only function $g()$ studied earlier:

$$g_\Gamma(R) = \mathrm{e}^{-\mathrm{j}2kR}\int\!\!\int p(\theta(R),\ \phi)\Gamma(R,\ \phi)R\,\mathrm{d}\phi\,\mathrm{d}z \qquad [79]$$

If the scatterers were all arranged on a single phase front at range R, then the received signal would be

$$r(t) = \frac{P_T^{1/2}G\lambda}{(4\pi)^{3/2}R^2}g_\Gamma(R)\mathrm{e}^{\mathrm{j}(1/2)\left(B_r/\tau_p\right)(t-2R/c)^2}\mathrm{rect}\!\left(\frac{t-2R/c}{\tau_p}\right)$$

$$\qquad [80]$$

For scatterers arranged over the entirety of the surface, we have an additional integration over all phase fronts:

$$r(t) = \frac{P_T^{1/2}G\lambda}{(4\pi)^{3/2}R^2}\int g_\Gamma(R)\mathrm{e}^{\mathrm{j}(1/2)\left(B_r/\tau_p\right)(t-2R/c)^2}$$

$$\times\,\mathrm{rect}\!\left(\frac{t-2R/c}{\tau_p}\right)\mathrm{d}R \qquad [81]$$

This is a convolution of the transmitted pulse with the integrated phase front. The range compression process (matched filtering) described above returns to us a record of the total return from phase fronts resolved in time to $1/B$ or in range to $c/2B$. The next step is to exploit the redundancy in the range signal as a function of pulse number to build resolution in the along-track dimension.

Azimuth signal and aperture synthesis

Real aperture radars are very useful for a variety of applications, generally those which require detection of a bright object in a broad region as is common in military applications, or where a low-resolution averaged measure of surface roughness or dielectric is sufficient. Oceans and ice sheets are generally featureless, so high resolution is not often needed. Large-area estimates of backscatter can be interpreted as wind speed, for example, on global scales, with the sensor dealing with only modest amounts of data.

Often, however, especially for land applications, fine-resolution images are important because the natural spatial variability of the surface is rapid. So it would be very nice if we could overcome the azimuth beam limit on resolution without building an enormously long antenna. It turns out that it is possible using the techniques of aperture synthesis.

Returning momentarily to a point target on the ground, consider a train of pulses transmitted as a spacecraft travels along a straight line path (**Figure 30**) above a flat Earth. The pulse is transmitted and subsequently received at a slightly later time, but for the purpose of understanding the concepts, we can imagine a 'stop-and-shoot' model where the radar to target to radar distance is $2R_i$ for a given pulse i. The transmitted and received pulses are shown schematically in **Figure 30**, illustrating the changing range from pulse to pulse.

In this geometry,

$$R_i = \sqrt{R_0^2 + (x_i - x_0)^2}, \quad x_i = vt_i, \quad t_i = i\,\mathrm{PRI} + t_0$$

$$\qquad [82]$$

where R_0 is the shortest range from the flight track to the target, also called 'broadside', and x_0 is the azimuth at that location. Thus the phase factor we have been carrying along for a return from a broadside pulse, $\exp(-j2kR_0)$, must now be updated from pulse to pulse. After demodulation and range compression, the received signal from a point target for pulse i is

$$r_{zc\delta,x_i}(t) = g_0 p(x_i - x_0)\mathrm{e}^{-\mathrm{j}2kR_i(x_i)}\mathrm{sinc}\{\pi B(t - t_i)\} \qquad [83]$$

where $t_i = 2R_i/c$. This is the range-compressed form eqn [76] with an adjusted range term, and an explicit function $p()$ denoting the antenna pattern of the radar antenna. Writing this entirely in terms of range and azimuth coordinates gives

$$r_{zc\delta,x_i}(R_f) = g_0 p(x_i - x_0)\mathrm{e}^{-\mathrm{j}2kR_i(x_i)}\mathrm{sinc}\left\{\frac{\pi}{\Delta R}(R_f - R_i)\right\}$$

$$\qquad [84]$$
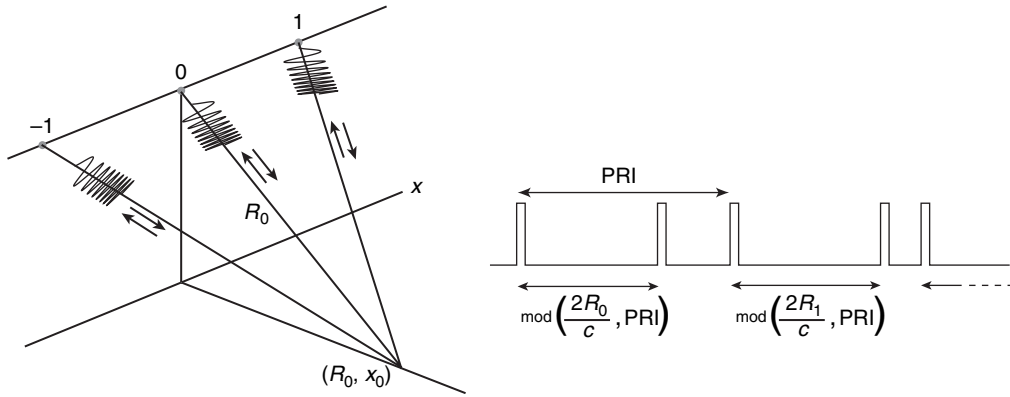
Figure 30 Illustration of range history in azimuth of the pulsed radar. Ignoring the finite speed of light, the radar can be thought of as transmitting and receiving from each point on orbit. Because of the finite extent of the azimuth beam, a single point target on the ground will be seen in a succession of received pulse echoes, with decreasing-to-increasing range location as the spacecraft flies by.

where R_f is the range variable, also called 'fast time' because it represents the time variation of the signal across a single pulse, and x_i is the azimuth location of a given pulse, also called 'slow time' since it characterizes the slower variation of the range trace from pulse to pulse.

If the transmitting antenna is length L in azimuth, then the target will be illuminated for an azimuth distance (3 dB) of

$$X_{\mathrm{ill}} = \frac{\lambda}{L} R_0 \qquad [85]$$

The echo field from the point target will be sampled $N = X_{\mathrm{ill}}/\Delta x_{\mathrm{p}}$ times while it is in the beam, where $\Delta x_{\mathrm{p}} = v\,\mathrm{PRI}$ is the pulse spacing on the flight track (**Figure 31**). The field is sampled effectively by a linear array of identical antennas spaced by Δx_{p}. It is shown in the homework (and any number of books) that the antenna pattern of a linear array of this sort is the product of the antenna pattern of the physical antenna and the array factor. The physical antenna is small, leading to a broad sinc-like beam in azimuth. The array factor is also sinc-like, but is much narrower

$$AF = \frac{\sin\big((N/2)k\Delta x_{\mathrm{p}} \sin\theta\big)}{\sin\big((1/2)k\Delta x_{\mathrm{p}} \sin\theta\big)} \qquad [86]$$

In the limit as $\Delta x_{\mathrm{p}} \to 0$, and for small θ

$$AF \to N\frac{\sin((k/2)X_{\mathrm{ill}}\theta)}{(k/2)X_{\mathrm{ill}}\theta} \qquad [87]$$
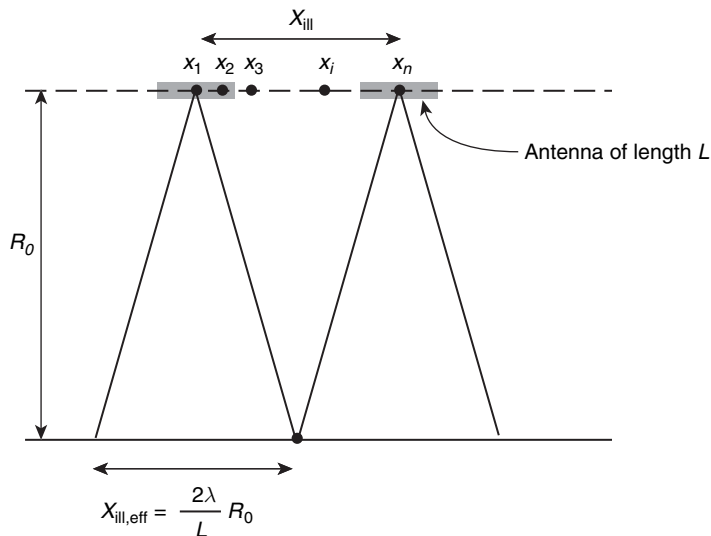
Figure 31 Point target illuminated by an antenna of length L. The 3 dB extent of the beam is $\lambda R_0/L$, but because two-way propagation introduces a doubling of the range and consequently the phase, the effective extent of illumination is $2\lambda R_0/L$.

which has the expected 3 dB beam width of $\theta_{3\,dB} = \lambda / X_{ill}$. Thus the sampled linear array in space serves as a very long antenna of length X_{ill}. Without careful deliberation, we might conclude that the resolution in azimuth we could achieve is given by its effective 3 dB beam extent $\Delta X = \theta_{3\,dB} R_0 = L$; however, we must take into account the two-way propagation of the radar signal. The above arguments are correct for an antenna synthesized to illuminate an area at range R_0. However, since $2R_0$ is the actual distance propagated, we can think of the beam extent for illumination as actually at twice the distance, that is, the phase variation across the aperture is as if the antenna were at twice the distance. Thus,

$$X_{ill,\,eff} = \frac{\lambda}{L} 2R_0 \qquad [88]$$

and the resolution allowed by an antenna of this length is

$$\Delta X = L/2 \qquad [89]$$

This remarkable result states that the resolution of a synthetic aperture system is independent of range and velocity, and is just half the physical antenna length. As the range is increased, the synthetic aperture increases in length and the angular extent of synthetic beam narrows in proportion to maintain fixed resolution. If the physical antenna decreases in size for a fixed range, the illuminated area increases, increasing the synthetic aperture and narrowing the synthetic beam width. Since the range is fixed, the resolution becomes finer in proportion to the reduction in L.

Consider R_i as a continuous variable of the azimuth coordinate

$$R(x) = \sqrt{R_0^2 + (x - x_0)^2} \qquad [90]$$
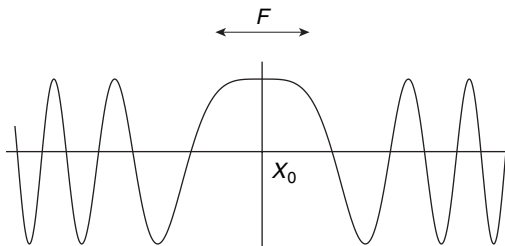
$$\approx R_0 \left(1 + \frac{(x - x_0)^2}{2R_0^2} \right) \qquad [91]$$

$$\approx R_0 + \frac{(x - x_0)^2}{2R_0} \qquad [92]$$

The constant phase term $\exp(-j2kR_0)$ is very difficult to measure absolutely (billions of cycles), so for now we will ignore it and focus on the spatially varying component of the phase

$$e^{-j2k(x - x_0)^2 / 2R_0} = e^{-j\pi(x - x_0)^2 / F^2} \qquad [93]$$

where $F^2 = \lambda R_0 / 2$. F is known as the Fresnel zone and it is the distance along the synthetic array where the wave front is within $\pi/4$ radians of phase (see **Figure 32**). Typical Fresnel zone size is on the order of 100–200 m. Since the array is coherent over this length, points simply summed together in azimuth will add coherently and form a meaningful, albeit low-resolution representation of the surface in azimuth:

$$\pi \frac{(x - x_0)^2}{F^2} < \frac{\pi}{4} \rightarrow x - x_0 < \frac{F}{2} \qquad [94]$$

The range-compressed pulse history can now be written as

$$r_{zc\delta}(x_i, R_f) = g_0 p(x_i - x_0) e^{-j2kR_0} e^{-j\pi(x - x_0)^2 / F^2}$$
$$\times \text{sinc} \left\{ \frac{\pi}{\Delta R} \left[R_f - \left(R_0 + \frac{(x_i + x_0)^2}{2F^2} \right) \right] \right\} \qquad [95]$$

where we explicitly call out the functional dependence on x_i in the argument to $r_{zc\delta}$.

The form of the azimuth signal is rather similar to the range signal before compression. Using the same approach to recover the point target in azimuth, we can define a matched filter in azimuth that is a similar conjugate chirp signal:
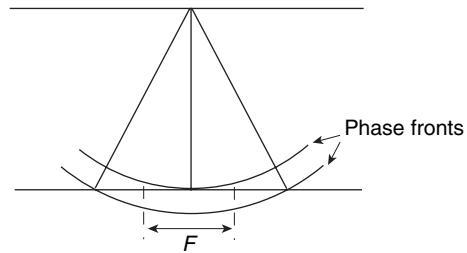
Figure 32 Illustration of the Fresnel zone, where the phase from an incident wave front varies by less than $\pi/4$ radians across the zone (right). Symmetric chirp at left is the real part of the azimuth chirped phasor. It is nearly constant over the Fresnel zone. The synthetic array is coherent over this length and points simply summed in azimuth will add coherently, and will be resolved at this length.

$$b_{am}(x) = b_a^*(-x) = p(-x)e^{j\pi x^2/F^2} \qquad [96]$$

The doubly compressed point-target signal is then

$$r_{zcc\delta}(x, R) = g_0 e^{-j2kR_0} \text{sinc}\left\{\frac{\pi}{\Delta R}[R - R_0]\right\} \int p(x' - x_0)$$
$$\times e^{-j\pi(x'-x_0)^2/F^2} p(x' - x)e^{j\pi(x'-x)^2/F^2}\, dx' \qquad [97]$$

where we ignored the x-dependent term in the range sinc function because processors generally take care of this term before performing the matched filter in azimuth. If we allow $p()$ to be a rect-function of extend X_{ill} as we did for range, then we will obtain exactly the same expression as we did for range, and the doubly compressed point-target signal becomes

$$r_{zcc\delta}(x, R) = g_0' e^{-j2kR_0} \text{sinc}\left\{\frac{\pi}{\Delta R}[R - R_0]\right\}$$
$$\times \text{sinc}\left\{\frac{\pi(x - x_0)X_{ill}}{F^2}\right\} \qquad [98]$$

Since in this case we explicitly took into account the two-way propagation in the phase, X_{ill} is the actual illuminated extent, not the double-length effective extent used in the heuristic argument earlier. The first null of the azimuth response is where

$$x - x_0 = \frac{F^2}{X_{ill}} \qquad [99]$$
$$= \frac{\lambda R_0/2}{\lambda R_0/L} = \frac{L}{2} \qquad [100]$$

We take this as the 3 dB resolution in azimuth, and we see it is identical to the heuristically derived resolution.

Expressed in terms of the range resolution $\Delta R = c/2B$ and azimuth resolution $\Delta X = L/2$, the overall impulse response of the radar system and matched filtering operations is

$$r_{zcc\delta}(x, R) = g_0' e^{-j2kR_0} \text{sinc}\left\{\frac{\pi}{\Delta R}[R - R_0]\right\}$$
$$\times \text{sinc}\left\{\frac{\pi}{\Delta X}[x - x_0]\right\} \qquad [101]$$

For a surface described by a general reflectivity function in range and azimuth $\Gamma(x, R)$ related to the surface backscatter cross-subsection through eqn [77], the matched filter response will be the linear convolution with the impulse response in eqn [101]:

$$\Gamma_{zcc}(x, R) = \int e^{-j2kR'}\Gamma(x', R')\text{sinc}\left\{\frac{\pi}{\Delta R}[R - R']\right\}$$
$$\times \text{sinc}\left\{\frac{\pi}{\Delta X}[x - x']\right\}dx'\, dR' \qquad [102]$$

Range-Doppler images

SAR images generated in this style of filtering are often called 'range-Doppler' images. The quadratic variation of the phase of a point target with time along track leads to a linear frequency variation:

$$\phi(t) = -\frac{\pi}{F^2} v^2(t - t_0)^2 \qquad [103]$$
$$\omega(t) = -2\pi\frac{v^2}{F^2}(t - t_0) \qquad [104]$$
$$f_D(t) = -\frac{v^2}{F^2}(t - t_0) \qquad [105]$$

Note the Doppler frequency is positive as the sensor approaches the target $(t < t_0)$. The Doppler bandwidth is the totality of frequency content for any target. This is limited by the beam width in azimuth, $T_{ill} = X_{ill}/v$, giving

$$f_{D,ill} = \frac{2}{\lambda R}v^2\frac{X_{ill}}{v} = \frac{2v}{L} \qquad [106]$$

This is in time units (Hz). In spatial units, $f_{Dx,ill} = 2/L$, which is just the reciprocal of the spatial resolution.

Another way of describing this Doppler bandwidth is through the Doppler equation $f_D = -2\vec{v} \cdot \hat{l}/\lambda$, where $\vec{v}$ is the vector velocity of the sensor, and $\mathbf{l}$ is the direction from the sensor to a point on the ground. At any instant of time, there is a one-to-one correspondence between the azimuth position of a target and the Doppler frequency (**Figure 33**). The antenna beam illuminates an area on the ground, with each target at a given angle off boresight having a unique Doppler frequency. The Doppler frequency can be written as $f_D = -2v\cos\theta_{az}/\lambda$, where θ_{az} is the angle formed by the velocity vector $\vec{v}$ and the look vector $\hat{l}$. Defining $\theta_{b.s.}$ as the angle measured relative to broadside (that is, perpendicular to the velocity vector), and assuming that the antenna's boresight is oriented at broadside, $f_D = -2v\sin\theta_{b.s.}/\lambda$. The total Doppler bandwidth at that instant is given by the Doppler frequencies at the edge of the beam. The total beam extent is then
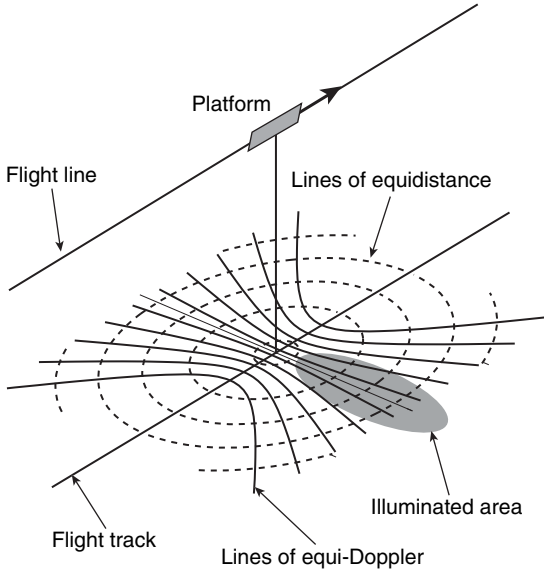
$$2\theta_{b.s.} = X_{ill}/R \qquad [107]$$
$$= \lambda R/(LR) \qquad [108]$$
$$= \lambda/L \qquad [109]$$

as expected. Thus, the Doppler bandwidth $f_{D,ill}$, is also $2v(2\sin\theta_{b.s.})/\lambda = 2v/L$.

At any given instant, the received signal is comprised of the full complement of Doppler frequencies

Figure 33 Range-Doppler coordinates of the surface as seen by a sensor at an instant of time. Targets in a band of Doppler frequencies contribute to the signal return. It is up to the azimuth matched filtering operation to sort out the targets into unique Doppler bins. Modified from Elachi C (1988) *Spaceborne Radar Remote Sensing: Applications and Techniques*. New York: IEEE Press.

because targets are distributed throughout the beam, each with their own Doppler frequency. In forming the synthetic aperture at some reference time, we are rearranging each of the targets in the range-Doppler image to lie at the Doppler frequency appropriate to that geometry at that reference time (**Figure 33**).

To be more explicit, we pick a collection of pulses that will be used to form our synthetic aperture, noting that this establishes a position in space at a time instant as a reference for defining Doppler frequencies. We can also model the ground as a collection of point targets arranged in azimuth, such that

$$r_{zc\delta,\,\text{tot}} = \sum_{k=1}^{N} A_k e^{-j\pi(x-x_k)^2/F^2} \operatorname{sinc}\frac{\pi}{\Delta R}\left(R - R_0 - (x-x_k)^2/2R_0\right)$$
[110]

For the sake of this discussion, we allow the sinc function to become a delta function in range and ignore its variability with x, which is small and correctable. Thus,

$$r_{zc\delta} \approx \sum_{k=1}^{N} A_k e^{-j\pi(x-x_k)^2/F^2}$$
[111]

The azimuth compression process can be written then as

$$
\begin{aligned}
r_{zcc\delta}(x) &= \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} r_{zc\delta}(x') e^{j\pi(x'-x)^2/F^2}\,dx' \\
&= \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} \sum_{k=1}^{N} A_k e^{-j\pi(x'-x_k)^2/F^2}\, e^{j\pi(x'-x)^2/F^2}\,dx' \\
&= e^{j\pi x^2/F^2} \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} \sum_{k=1}^{N} A_k e^{-j\pi x_k^2/F^2} \\
&\quad \times e^{j2\pi x' x_k/F^2}\, e^{-j2\pi x' x/F^2}\,dx' \\
&= \sum_{k=1}^{N} e^{j\pi x^2/F^2} \int_{x-X_{\text{ill}}/2}^{x+X_{\text{ill}}/2} A_k' e^{-j2\pi x'(f_{Dx}-f_{k,\,Dx})}\,dx' \quad [112]
\end{aligned}
$$

where $f_{Dx} = x/F^2$ is the Doppler spatial frequency. From this arrangement of terms, we can recognize that the compressed azimuth signal is just the sum of Fourier transforms of the individual complex exponentials of complex amplitude A_k'. Because the argument of the exponential is the Doppler frequency, we see the transform domain is the Doppler domain, and each exponential (being representative of a scatterer at a particular azimuth location x_k relative to the reference location) transforms to an individual Doppler frequency f_{kDx}. Thus for a general continuous surface, the summation becomes an integration, and we see the compression process consists of focusing the array by applying a quadratic phase correction, followed by a transform that sorts the scatterers into correct Doppler bins for that time reference.

We represent and execute the imaging as a convolution in spatial variables because the processed data sample spacing is independent of range on input and output. In the focussing-Fourier-transform approach, the Doppler sample spacing is controlled by the implementation of Fourier transform, which is usually accomplished by fast Fourier transform (FFT). In this case, if the input data are sampled at a spacing $L/2$, the Doppler spacing will be $2/L/N$, where N is the number of points in the FFT.

Other Doppler considerations

If the antenna pattern was isotropic and we have a continuous time system, the phase and Doppler of a single point target would be as depicted in **Figure 34**. Imposing an antenna pattern where the antenna pattern is pointing 'broadside,' that is orthogonal to the velocity vector, the phase and Doppler history would be truncated according to the extent of the beam (see **Figure 34**).
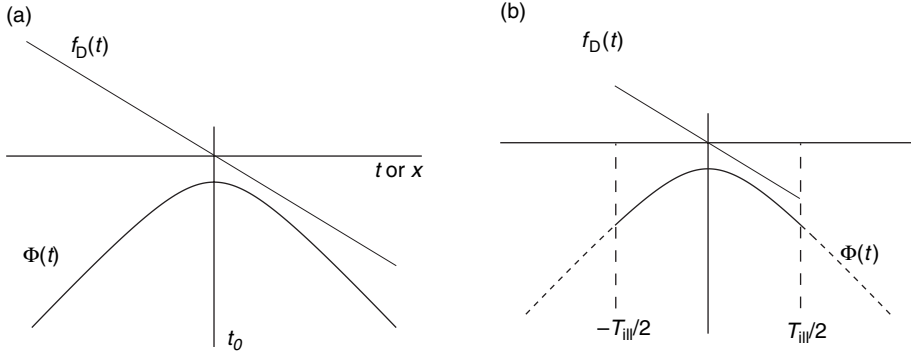
Figure 34 Phase and Doppler history over time for a point target. (a) The trace for a system with an isotropic antenna: the trace keeps going. (b) The finite illumination time of a finite-dimension antenna truncates the phase and Doppler history.
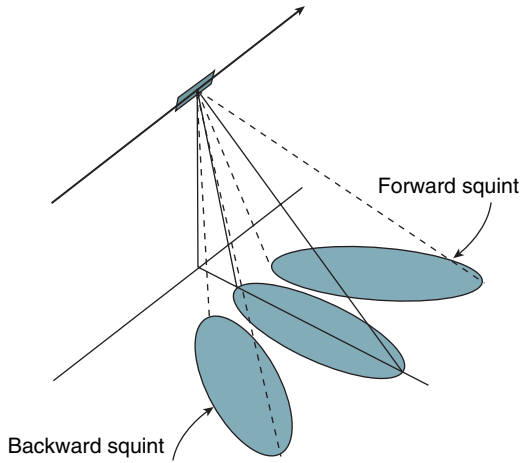
Figure 35 Illustration of forward and backward squint.

It is most commonly the case that the finite extent beam is 'squinted' forward or backward with respect to the velocity vector of the sensor. For an aircraft this is most commonly due to cross-winds causing the fuselage to crab. For spacecraft, there is a natural latitude-dependent angular offset between the velocity vector of the inertial orbit and the rotating Earth below, causing a natural squint (**Figure 35**).

The effect of this squint on the phase and Doppler history of a target is shown in **Figure 36**. Note the shape of the phase or Doppler does not change, just the range of phases or frequencies considered. Thus, there is a unique relationship between the pointing and the location of the scatterers. In processing data, the extent of the antenna beam (Doppler bandwidth) and the 'centroid' of the Doppler, or the degree of squint as represented by the center of the Doppler bandwidth, are specified in one way or another.

Figure 37 shows that the spectra are shifted to be centered on the Doppler spread associated with a particular squint. It is important to know the squint to ensure that the processing is centered on the proper part of the azimuth spectrum. The centroid specified in processing defines the direction from which the ground is imaged after processing.

The squint can be derived from information provided by the spacecraft and radar manufacturer. Often spacecraft have sensitive position and attitude sensors that record the information needed to calculate the direction of the antenna boresight. When this information is not available, the Doppler centroid itself can be estimated from the data. This estimated centroid can be used directly in the processing. The centroid generally varies across range, so often the centroid is specified as a function of range.

To understand the arrangement of spectral energy for a collection of scatterers, consider a collection of point-target scatterers all at the same range, but at different azimuths:

$$
\begin{aligned}
r_{zc\delta,\text{tot}} = &\, A_0 e^{-j\pi(x-x_0)^2/F^2} \\
&\times \operatorname{sinc} \frac{\pi}{\Delta R}\left(R - R_0 - (x-x_0)^2/2R_0\right) \\
&+ A_1 e^{-j\pi(x-x_1)^2/F^2} \\
&\times \operatorname{sinc} \frac{\pi}{\Delta R}\left(R - R_0 - (x-x_1)^2/2R_0\right) + \cdots \\
&+ A_n e^{-j\pi(x-x_n)^2/F^2} \operatorname{sinc} \frac{\pi}{\Delta R}\left(R - R_0 - (x-x_n)^2/2R_0\right)
\end{aligned}
$$
$$[113]$$

If we ignore the antenna pattern modulation of target energy and denote the azimuth spectrum of an individual scatter response as

$$
F(\omega) = \text{F.T.}\left\{ e^{-j\pi x^2/F^2} \operatorname{sinc} \frac{\pi}{\Delta R}\left(R - R_0 - x^2/2R_0\right)\right\}
$$
$$[114]$$
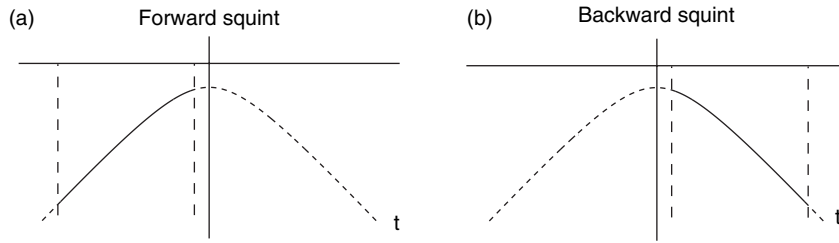
Figure 36   Phase history over time for a point target. (a) The trace for a forward squinted system. (b) The trace for a backward squinted system. Only phase is shown, but with reference to the previous figures, the Doppler history is similarly constrained. Forward (backward) squinted systems see more positive (negative) Doppler frequencies.
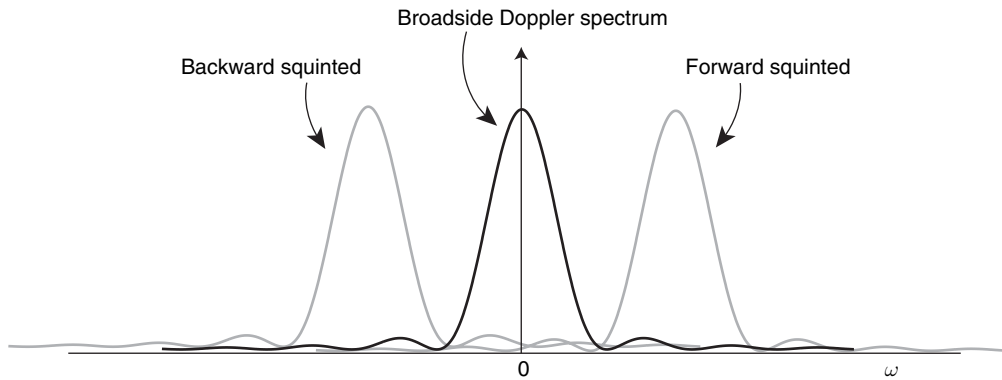
Figure 37   Hypothetical Doppler spectra of broadside, forward and backward squinted signals. For a collection of scatterers over the beam pointed with a particular squint, all scatterers experience the same Doppler frequency spread, so the spectra would be as shown, controlled in extent and magnitude by the antenna pattern in azimuth.

where F.T. denotes the Fourier Transform, the spectrum of $r_{zc\delta, \text{tot}}$ would be

$$R_{zc\delta, \text{tot}}(\omega) = F(\omega)\left(A_0 e^{-j\omega x_0} + A_1 e^{-j\omega x_1} + \cdots + A_n e^{-j\omega x_n}\right)$$

$$[115]$$

Thus, except for a phase ramp in azimuth frequency space, all scatterers have the same spectrum. Clearly, the process of applying the matched filter neutralizes this common component, and the linear phase terms in frequency distinguish the targets in position upon inverse transformation. This again shows the spatial-Doppler duality. The x dependence of the energy's localization in range $x^2/2R_0$ is called the 'range migration' and is seen to be part of the function $F(\omega)$ in the frequency domain. Hence any effects this may introduce can be dealt with for all targets simultaneously by manipulating the spectrum.

We ignored the 'range migration' term $(x - x_0)^2/2R_0$ in the argument to the sinc function in range, but frequently this is not advisable. The processing of data

on an orthogonal grid is enforced by the need for high-speed computations, and the most efficient processing methods are those that are separable in range and azimuth and for which FFT-based convolutions can be performed. For broadside imaging, the term can be ignored when $X_{\text{ill}}^2/2R_0$ is a fraction of the range resolution ΔR. In other words, when the migration over the full extent of the azimuth matched filter (equivalent to the beam extent on the ground) is smaller than a range-resolution element, the azimuth matched filter will catch all the energy for a given scatterer in a single range bin.

The sampled azimuth spectrum and range migration

Since the range migration correction terms depend on the Doppler frequency, one must know the correct Doppler frequency to properly compress the imagery. This is also true for computing the correct azimuth matched filter; however, the sampled nature of the azimuth signal adds an interesting wrinkle to the processing that is often confusing. Since the

(a)

Antenna spectrum

Spectrum replica

$-$PRF/2 PRF/2

f

(b)

$-$PRF/2   PRF/2   3 PRF/2

f
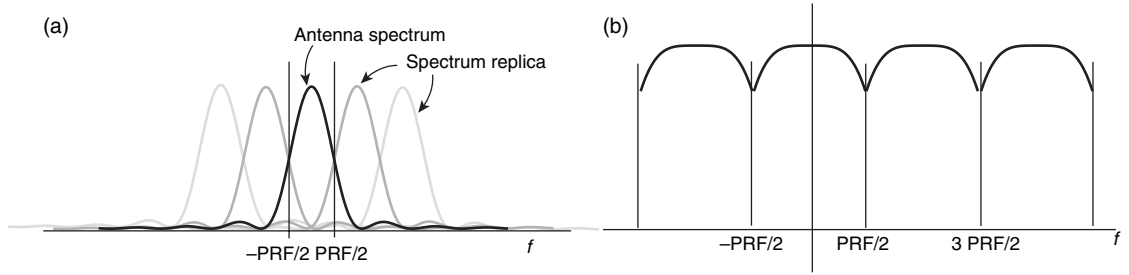
Figure 38   The azimuth signal is sampled at the PRF. The resulting discrete time signal in azimuth has a circular spectrum with all energy at frequency higher than the PRF aliased into the circular bandwidth of the PRF. (a) Illustrates the replicas of the continuous spectrum that are aliased, and (b) the resulting circular spectrum.

azimuth signals are inherently derived from pulses acquired along track, the azimuth signal is a discrete time signal sampled at the PRF. The actual discrete time spectrum is then a circular spectrum, with all energy at frequencies higher than the PRF aliased into the range of one PRF (**Figure 38**).

Thus without knowledge of the geometry of the spacecraft and the observation, it is often difficult to know the actual Doppler centroid, since one usually estimates it from the spectrum only in the range of the PRF. One can process the data with an incorrect Doppler centroid, off by some number of PRFs (also called 'ambiguities') for instance, and the imagery will look almost correct. Since the signal is sampled, the matched filter spectrum for the actual centroid will be the same as for the incorrect centroid as long as the centroid is wrong by an integer number of ambiguities. However, the range migration correction is in terms of absolute Doppler, so if the wrong ambiguity is used, the images will be poorly focused in range (because energy is spread in range for each scatterer) and in azimuth (because the compression uses incomplete information for each scatterer).

If the attitude of the platform is known, then the Doppler can be computed from

$$f_\mathrm{D} = -\frac{2\vec{v} \cdot \hat{l}}{\lambda} = -\frac{2v}{\lambda} \sin\theta \sin\phi_\mathrm{sq} \qquad [116]$$

where $\vec{v}$ is the velocity vector, $\hat{l}$ is the look vector from antenna to target, θ is the look angle, and ϕ_sq is the squint angle. If the attitude parameters are not known, then one has to estimate the ambiguity. This can be done by trial and error – examine the focus for a variety of integer ambiguity choices – or by one of several automated techniques. A popular method of ambiguity estimation is to split the azimuth spectrum into two pieces and process each with a particular Doppler ambiguity assumption. If the correction is

the right one, both side bands will be adjusted to the correct range and the two images produced will be properly registered in range. If the correction is not right, there will be a residual shift between the two images in range. From the magnitude of this shift it is often possible to estimate the correct ambiguity. Unfortunately the imagery are lower resolution in azimuth and defocused in range, so estimating the shift is sometimes difficult.

Speckle

Radar images are naturally noisy because of the coherent interaction of the electro-magnetic wave with scatterers within a resolution cell, known as speckle. Speckle is an important part of the radar literature, as it is an important limitation of the imagery that many people have worked hard to minimize. It is also an important concept for interferometry, which exploits the spatially fixed properties of speckle to extract phase differences between images. The normalized backscatter cross-section σ_0 is an average quantity for a naturally varying random surface. For homogenous surfaces made up of a continuum of scatterers, the backscatter cross-section is a measure of the roughness of the surface and its natural reflectivity, as shown in a previous section. Viewing the SAR imaging process as a lens (**Figure 39**), after imaging, the random (complex) surface reflectivity is convolved with the sinc-function impulse response of the imaging system, accounting for the random differential path length introduced by surface roughness.

In a given resolution element, we consider the sum of the collection of small independent scatterers, where each scatterer has its own amplitude and phase
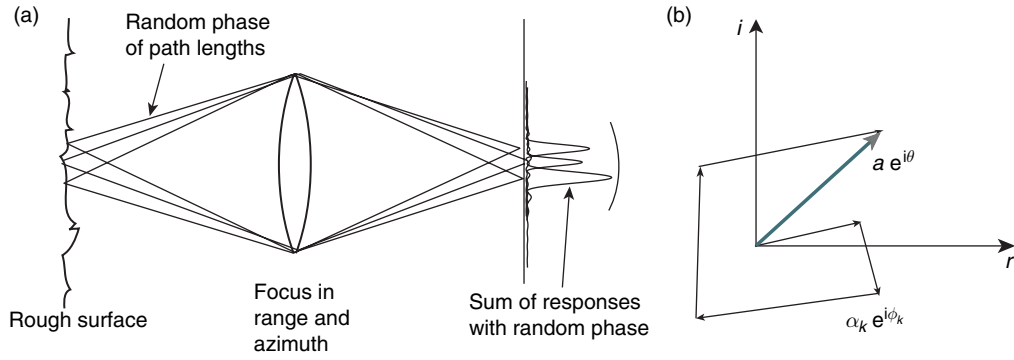
Figure 39 Illustration of the concept of radar imaging of a random surface (a) and the coherent sum of independent scatterers as a phasor (b).
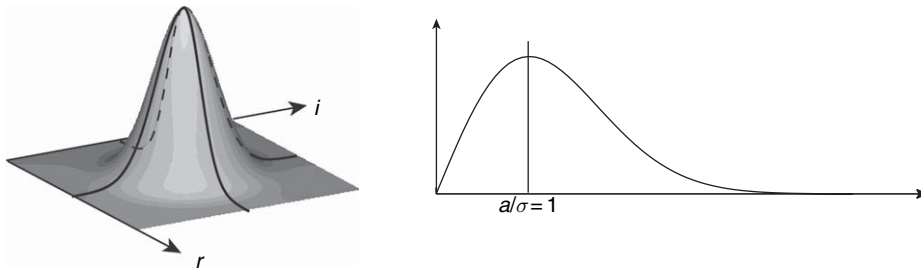
Figure 40 The real and imaginary parts of a SAR image are circularly Gaussian (*left*). The amplitude of the SAR pixel is Rayleigh distributed, with peak at $a = \sigma$.

$$ae^{j\theta} = \sum_{k=1}^{N} \alpha_k e^{j\phi_k} \qquad [117]$$

If α_k and ϕ_k are statistically independent of each other and α_k are identically distributed with mean $\bar{\alpha}$ and second moment $\overline{\alpha^2}$, and ϕ_k are uniformly distributed in the interval $[-\pi, \pi]$, then one can show that for large N

$$r = a \cos \theta \qquad [118]$$

$$i = a \sin \theta \qquad [119]$$

are Gaussian-distributed zero-mean random variables with joint probability density function

$$p_{RI}(r, i) = \frac{1}{2\pi\sigma^2} e^{-(r^2+i^2)/2\sigma^2} \qquad [120]$$

where $\sigma^2 = \overline{\alpha^2}/2$ (**Figure 40**). The variance of the real and imaginary parts of the SAR image are related to the mean power of the intrinsic scatterers but are zero mean random variables themselves. As such for homogeneous areas, averaging many complex pixels together in a region will eventually reduce the signal to zero on average and not improve the quality of the image. This is a common mistake among beginners. It

is the amplitude of the image, a, or its intensity (power), $I = a^2$, that interests us.

For the amplitude $a = \sqrt{r^2 + i^2}$, we can use the rules of transformation of random variables (multiply the distribution by the Jacobian of the transformation and substitute in the new variables) to get the distribution

$$p_A(a) = \frac{a}{\sigma^2} e^{-a^2/2\sigma^2}, \quad a \geq 0 \qquad [121]$$

which is known as the Rayleigh distribution (**Figure 40**). The phase is uniformly distributed over the interval $(-\pi, \pi)$

$$p_\Theta(\theta) = \frac{1}{2\pi}, \quad -\pi < \theta < \pi \qquad [122]$$

The mean and variance of the amplitude are

$$\bar{a} = \sqrt{\frac{\pi}{2}} \sigma \qquad [123]$$

$$\sigma_a^2 = \left(2 - \frac{\pi}{2}\right)\sigma^2 \qquad [124]$$

Clearly the mean amplitude of the image scales with the variability in the image, which is related to the mean backscattered power $\overline{\alpha^2}$ of the scatterers

comprising the resolution element. The brighter the image, the noisier the image, and vice-versa. Thus, at some signal level such that the signal level returned is sufficiently larger than the thermal noise in an image, having a more powerful radar does not necessarily improve image quality.

The intensity, or power, of the image $I = a^2$ has an exponential distribution

$$p_I(I) = \frac{1}{\bar{I}} e^{-I/\bar{I}}, \quad I \geq 0 \qquad [125]$$

with mean and variance

$$\bar{I} = 2\sigma^2, \quad \sigma_I^2 = 4\sigma^4 \qquad [126]$$

The mean power is equal to its standard deviation, as expected from the mean amplitude scaling with σ. In the display and analysis of an image with only 'one look,' meaning processed at full resolution using the full synthetic aperture, the SNR of the intensity image is just I since $\bar{I}^2/\sigma_j^2 = 1$; the brighter the image, the noisier. In the design of a radar system, one must have sufficient power to receive a return from the surface above the thermal noise (satisfying the radar equation, where $\overline{\alpha^2}$ is related to the back-scattering cross-section), but not so high that we are wasting power. For interferometry, we will soon see however that more power in general is better!

The radiometric resolution is defined as the ability to discriminate surfaces of different brightnesses. It has been defined to be

$$Q = 10 \log \left(\frac{\bar{I} + \sigma_I}{\bar{I} - \sigma_I} \right) \qquad [127]$$

which is something like the distance apart in amplitude of discriminable amplitudes. For a single-look image, that distance is infinite. In order to improve this situation, it is often necessary to average many pixels together. The act of averaging N independent (Gaussian, but also approximately true for non-Gaussian) random variables with identical distributions reduces the variance by a factor of N. Thus for

$$I' = \frac{1}{N} \sum_{k=1}^{N} I_k \qquad [128]$$

we have $\bar{I}' = \bar{I}$, $\sigma_{I'}^2 = \sigma_I^2/N$, and the radiometric resolution becomes

$$Q = 10 \log \left(\frac{\bar{I}' + \sigma_{I'}}{\bar{I}' - \sigma_{I'}} \right) = 10 \log \left(\frac{\sqrt{N}+1}{\sqrt{N}-1} \right) \qquad [129]$$

The resolution distance approaches 0 for arbitrarily large N. This averaging technique is often called 'taking looks' because it can be shown to be equivalent to segmenting the azimuth spectrum into separate frequency bands, forming a power image from each band, then stacking the images in an average. Since azimuth band corresponds to a particular look direction relative to a target on the ground, this technique is called taking looks.

Of course averaging together a large number of pixels will lower the intrinsic resolution of the image, so designing the system to have a larger bandwidth is desirable if radiometric resolution is important. Note also that there is a limit to how far one can go to improve the statistics. The improvement above assumes that all the pixels come from the same distribution. If the scene is variable in intrinsic brightness, not all the benefits will be realized.

If there is a bright target within a resolution cell, then there is a mean value associated with the Gaussian random variables (**Figure 41**). This complex random variable has a Rician distribution given by

$$p_A(a) = \frac{a}{\sigma^2} e^{-(a^2+s^2)/2\sigma^2} I_0 \left(\frac{as}{\sigma^2} \right), \quad a \geq 0 \qquad [130]$$
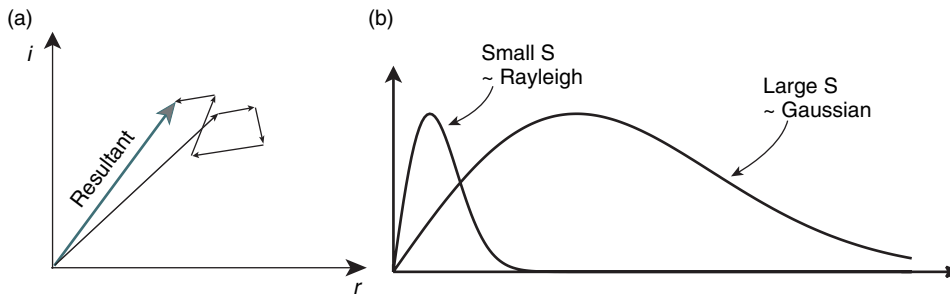
Figure 41 (a) A phasor diagram with a bright specular target (within a scattering cell. (b) The Rician distribution of the amplitude.

where I_0 is the modified Bessel function of the first kind and s is the specular target strength. This distribution approaches a Gaussian distribution for large s relative to σ. The distribution of the phase, which is uniform for $s=0$, becomes more peaked as s grows, eventually approximating a Gaussian also:

$$p_\Theta(\theta) \to \frac{s/\sigma}{\sqrt{2\pi}} e^{-s^2\theta^2/2\sigma^2} \qquad [131]$$

Often specular targets are used for calibration of the radar. If the specular target is not sufficiently bright, then the amplitude and phase will be corrupted by the background random scatterers.

Another way to model the surface is in terms of a general reflectivity function that is statistical in nature. Since

$$\Gamma_{zcc}(x', r') = \iint dr\, dx \Gamma(x, r)$$
$$\times \operatorname{sinc}\left(\frac{\pi}{\Delta X}(x'-x)\right)\operatorname{sinc}\left(\frac{\pi}{\Delta R}(r'-r)\right) [132]$$

if Γ is a statistical function, the power that is meaningful is the expected power scattered back from the resolution cell, which is related by scale factors and geometric normalizations to $E\{\Gamma\Gamma^*\}$. Writing this out explicitly, we have

$$E\{\Gamma_{zcc}\Gamma_{zcc}^*\}(x', r') = \int dx_1 \int dx_2 \int dr_1 \int dr_2$$
$$\times E\{\Gamma(x_1, r_1)\Gamma^*(x_2, r_2)\} \qquad [133]$$

$$= \operatorname{sinc}\left(\frac{\pi}{\Delta X}(x'-x_1)\right)\operatorname{sinc}\left(\frac{\pi}{\Delta X}(x'-x_2)\right) \qquad [134]$$

$$= \operatorname{sinc}\left(\frac{\pi}{\Delta R}(r'-r_1)\right)\operatorname{sinc}\left(\frac{\pi}{\Delta R}(r'-r_2)\right) \qquad [135]$$

If the scatterers that comprise the reflectivity are uncorrelated, then the expected value of the product of the surface with itself when not coaligned would be zero:

$$E\{\Gamma(x_1, r_1)\Gamma^*(x_2, r_2)\} = E\{\Gamma(x_1, r_1)\}E\{\Gamma^*(x_2, r_2)\} = 0$$
$$[136]$$

However, when the surface is self-aligned

$$E\{\Gamma(x_1, r_1)\Gamma^*(x_2, r_2)\} = |\Gamma(x_1, r_1)|^2 \qquad [137]$$

More succinctly,

$$E\{\Gamma(x_1, r_1)\Gamma^*(x_2, r_2)\} = E\{|\Gamma(x_1, r_1)|^2\}$$
$$\times \delta(x_1-x_2)\delta(r_1-r_2) \qquad [138]$$

Using this expression in eqn [135] simplifies the integrals

$$E\{\Gamma_{zcc}\Gamma_{zcc}^*\}(x', r') = \int dx \int dr\, E\{|\Gamma(x, r)|^2\} \quad [139]$$

$$\operatorname{sinc}^2\left(\frac{\pi}{\Delta X}(x'-x)\right)\operatorname{sinc}^2\left(\frac{\pi}{\Delta R}(r'-r)\right) \qquad [140]$$

$$= E\{|\Gamma|^2\}\Delta X\Delta R \qquad [141]$$

For homogeneous surfaces, we can expect the mean power to be essentially constant over the resolution cell, so we can take the expectation outside the integration. Also the variables x' and r' can profitably be set to zero since there is no dependence on space. Then the integration over the sinc functions produces a constant, which is proportional to the area under them. This turns out to be ΔX and ΔR. So we see that $E\{\Gamma_{zcc}\Gamma_{zcc}^*\}$ is related to the normalized backscattering cross-section σ_0 through the function Γ.

Doppler and the Interferometric Baseline

The precise definition of interferometric baseline and phase, and consequently the topographic mapping process, depends on how the SAR data comprising the interferometer are processed. Consequently, a brief overview of the salient aspects of SAR processing is in order.

Processed data from SAR systems are sampled images. Each sample, or pixel, represents some aspect of the physical process of radar backscatter. A resolution element of the imagery is defined by the spectral content of the SAR system. Fine resolution in the range direction is achieved typically by transmitting pulses of either short time duration with high peak power, or of a longer time duration with a wide, coded signal bandwidth at lower peak transmit power. Resolution in range is inversely proportional to this bandwidth. In both cases, the received echo for each pulse is sampled at the required radar signal bandwidth.

For ultra-narrow pulsing schemes, the pulse width is chosen at the desired range resolution, and no further data manipulation is required. For coded pulses, the received echoes are typically processed with a matched filter technique to achieve the desired range resolution. Most spaceborne platforms use chirp-encoding to attain the desired bandwidth and consequent range resolution, where the frequency is linearly changed across the pulse.

Resolution in the azimuth, or along-track, direction, parallel to the direction of motion, is achieved by synthesizing a large antenna from the echoes received from the sequence of pulses illuminating a target. The pulses in the synthetic aperture contain an unfocused record of the target's amplitude and phase history. To focus the image in azimuth, a digital 'lens' that mimics the imaging process is constructed, and is applied by matched filtering. Azimuth resolution is limited by the size of the synthetic aperture, which is governed by the amount of time a target remains in the radar beam. The azimuth beam width of an antenna is given by $\theta_{BW} \equiv k\lambda/L$, where λ is the wavelength, L is the antenna length, and k is a constant that depends on the antenna ($k = 1$ is assumed in this chapter). The size of the antenna footprint on the ground in the azimuth direction is approximately given by

$$l_{az} = \rho\theta_{BW} = \rho\frac{\lambda}{L} \qquad [142]$$

where ρ is the range to a point in the footprint.

During the time a target is in the beam, the range and angular direction to the target are changing from pulse to pulse, as shown in **Figure 30**. To generate SAR image, a unique range or angle must be selected from the family of ranges and angles to use as a reference for focusing the image. Once selected, the target's azimuth and range position in the processed image is uniquely established. Specifying an angle for processing is equivalent to choosing a reference Doppler frequency. The bold dashed line from pulse N-2 to the target in **Figure 30** indicates the desired angle or Doppler frequency at which the target will be imaged. This selection implicitly specifies the time of imaging, and therefore the location of the radar antenna. This is an important and often ignored consideration in defining the interferometric baseline. The baseline is the vector connecting the locations of the radar antennas forming the interferometer; since these locations depend on the choice of processing parameters, so does the baseline. For two-aperture cross-track interferometers, this is a subtle point; however, for repeat-track geometries where the antenna pointing can be different from track to track, careful attention to the baseline model is essential for accurate mapping performance.

References

Alsdorf DE, Smith LC, and Melack JM (2001) Amazon floodplain water level changes measured with interferometric SIR-C radar. *Geoscience and Remote Sensing, IEEE Transactions on* 39: 423–431 (ISSN 0196–2892).

Amelung F, Galloway DL, Bell JW, and Zebker H (1999) Sensing the ups and downs of Las Vegas: InSAR reveals structural control of land subsidence and aquifer-system deformation. *Geology* 27: 483–486.

Ballatore P (2006) Synthetic aperture radar interferometry: Separation of atmospheric artifacts from effects due to the topography and the terrain displacements. *Earth Planets Space* 58: 927–935.

Bamler R and Eineder M (1996) ScanSAR processing using standard high-precision SAR algorithms. *IEEE Transactions on Geoscience and Remote Sensing* 34: 212–218.

Bamler R and Hartl P (1998) Synthetic aperture radar interferometry. *Inverse Problems* 14: 1–54.

Bawden GW, Thatcher W, Stein RS, and Hudnut K (2001) Tectonic contraction across Los Angeles after removal of groundwater pumping effects. *Nature* 412: 812–815.

Bechor NBD and Zebker HA (2006) Measuring two-dimensional movements using a single InSAR pair. *Geophysical Research Letters* 33 (doi:10.1029/2006GL026883).

Bell JW, Amelung F, Ramelli AR, and Blewitt G (2002) Land subsidence in Las Vegas, Nevada, 1935–2000: New geodetic data show evolution, revised spatial patterns, and reduced rates. *Environ* 8: 155–174.

Berardino P, Fornaro G, Lanari R, and Sansosti E (2002) A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing* 40: 2375–2383.

Born M and Wolf E (1989) *Principles of Optics, 6th edn.* Oxford: Pergamon Press.

Burgmann R, Hilley G, Ferretti A, and Novali F (2006) Resolving vertical tectonics in the San Francisco Bay area from permanent scatterer InSAR and GPS analysis. *Geology* 34(3): 221–224.

Burgmann R, Rosen PA, and Fielding EJ (2000) Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Annual Review of Earth and Planetary Science* 28: 169–209.

Chen CW and Zebker HA (2001) Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization. *Journal of the Optical Society of America A* 18: 338–351.

Colesanti C, Ferretti A, Novali F, Prati CC, and Rocca F (2003) SAR monitoring of progressive and seasonal ground deformation using the permanent scatterers technique. *IEEE Transactions on Geoscience and Remote Sensing* 41: 1685–1701.

Delacourt C, Briole P, and Achache J (1998) Tropospheric corrections of SAR interferograms with strong topography: Application to Etna. *Geophysical Research Letters* 25: 2849–2852.

Dong D, Fang P, Bock Y, Cheng MK, and Miyazaki S (2002) Anatomy of apparent seasonal variations from GPS-derived site position time series. *Journal of Geophysical Research* 107: 2075.

Dong D, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, *et al.* (2006) Spatiotemporal filtering using principal component analysis and Karhunen–Loeve expansion approaches for regional GPS network analysis. *Journal of Geophysical Research* 111: B03405 (doi:10.1029/2005JB003806).

Dong D, Herring TA, and King RW (1998) Estimating regional deformation from a combination of space and terrestrial geodetic data. *Journal of Geodesy* 72: 200–214 (doi:10.1007/s001900050161).

Dong D, Yunck T, and Heflin M (2003) Origin of the international terrestrial reference frame. *Journal of Geophysical Research* 108: 2200.

Doubre C and Peltzer G (2007) Fluid-controlled faulting process in the Asal rift, Djibouti, from 8 yr of radar interferometry observations. *Geology* 35: 69–72.

Elachi C (1988) *Spaceborne Radar Remote Sensing: Applications and Techniques*. New York: IEEE Press.

Emardson TR, Simons M, and Webb FH (2003) Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal of Geophysical Research* 108 (doi:10.1029/2002JB001781).

Ferretti A, Prati C, and Rocca F (2000) Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 38: 2202–2212.

Ferretti A, Prati C, and Rocca F (2001) Permanent scatterers in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 39: 8–20.

Fialko Y, Simons M, and Agnew D (2001a) The complete (3-D) surface displacement field in the epicentral area of the 1999 M_w 7.1 Hector mine earthquake, California, from space geodetic observations. *Geophysical Research Letters* 28: 3063–3066.

Fialko Y, Simons M, and Khazan Y (2001b) Finite source modeling of magmatic unrest in Socorro, New Mexico, and Long Valley, California. *Geophysical Journal International* 146(1): 191–200.

Fialko Y, Sandwell D, Simons M, and Rosen P (2005) Three-dimensional deformation caused by the Bam, Iran, earthquake and the origin of shallow slip deficit. *Nature* 435: 295–299.

Fielding EJ, Talebian M, Rosen PA, *et al.* (2005) Surface ruptures and building damage of the 2003 Bam, Iran, earthquake mapped by satellite synthetic aperture radar interferometric correlation. *Journal of Geophysical Research* 110 (doi:10.1029/2004JB003299).

Foster J, Brooks B, Cherubini T, Shacat C, Businger S, and Werner CL (2006) Mitigating atmospheric noise for InSAR using a high resolution weather model. *Geophysical Research Letters* 33 (doi:10.1029/2006GL026781).

Franceschetti G and Lanari R (1999) *Synthetic Aperture Radar Processing*. Boca Raton, FL: CRC Press.

Fujiwara S, Rosen PA, Tobita M, and Murakami M (1998) Crustal deformation measurements using repeat-pass JERS-1 synthetic aperture radar interferometry near the Izu Peninsula, Japan. *Journal of Geophysical Research* 103: 2411–2426.

Fukuda J, Higuchi T, Miyazaki S, and Kato T (2004) A new approach to time-dependent inversion of geodetic data using a Monte Carlo mixture Kalman filter. *Geophysical Journal International* 159(1): 17–39.

Funning GJ, Parsons B, Wright TJ, Jackson JA, and Fielding EJ (2005) Surface displacements and source parameters of the 2003 Bam (Iran) earthquake from Envisat advanced synthetic aperture radar imagery. *Journal of Geophysical Research* 110: B09406.

Gabriel AK, Goldstein RM, and Zebker HA (1989) Mapping small elevation changes over large areas: Differential radar interferometry. *Journal of Geophysical Research* 94: 9183–9191.

Gatelli F, Guarnieri AM, Parizzi F, Pasquali P, Prati C, and Rocca F (1994) The wave-number shift in SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 32: 855–865.

Goldstein RM (1995) Atmospheric limitations to repeat-track radar interferometry. *Geophysical Research Letters* 22: 2517–2520.

Goldstein RM, Engelhardt H, Kamb B, and Frolich RM (1993) Satellite radar interferometry for monitoring ice sheet motion: Application to an Antarctic ice stream. *Science* 262: 1525–1530.

Goldstein RM and Zebker HA (1987) Interferometric radar measurement of ocean surface currents. *Nature* 328: 707–709.

Goldstein RM, Zebker HA, and Werner CL (1988) Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Science* 23: 713–720.

Goodman JW (1985) *Statistical Optics*. New York: Wiley-Interscience.

Gourmelen N and Amelung F (2005) Post-seismic mantle relaxation in the Central Nevada seismic belt. *Science* 310: 1473–1476.

Gray AL, Mattar KE, and Sofko G (2000) Influence of ionospheric electron density fluctuations on satellite radar interferometry. *Geophysical Research Letters* 27: 1451–1454.

Gray AL, Short N, Mattar KE, and Jezek KC (2001) Velocities and flux of the Filchner ice shelf and its tributaries determined from speckle tracking interferometry. *Canadian Journal of Remote Sensing* 27: 193–206.

Grell GA, Dudhia J, and Stauffer PJ (1995) A description of the fifth generation Penn State/NCAR Mesoscale Model (MM/5). NCAR Tech. Note 398, National Center for Atmospheric Research Boulder, CO.

Guarnieri AM and Prati C (1996) ScanSAR focusing and interferometry. *IEEE Transactions on Geoscience and Remote Sensing* 3: 1029–1038.

Hanssen RF (2001) *Radar Interferometry: Data Interpretation and Error Analysis*. Dordrecht, Netherlands: Kluwer.

Hoffmann J, Zebker HA, Galloway DL, and Amelung F (2001) Seasonal subsidence and rebound in Las Vegas Valley, Nevada, observed by synthetic aperture radar interferometry. *Water Resources Research* 37: 1551–1566.

Hooper A, Zebker H, Segall P, and Kampes B (2004) A new method for measuring deformation on volcanoes and other natural terrains using insar persistent scatterers. *Geophysical Research Letters* 31: L23611.

Jonsson S, Zebker H, Segall P, and Amelung F (2002) Fault slip distribution of the 1999 M_w 7.1 Hector Mine, California, earthquake, estimated from satellite radar and GPS measurements. *Bulletin of the Seismological Society of America* 92: 1377–1389.

Joughin I (2002) Ice-sheet velocity mapping: A combined interferometric and speckle-tracking approach. *Annals of Glaciology* 34: 195–201.

Joughin I, Winebrenner DP, and Percival DB (1994) Probability density functions for multi-look polarimetric signatures. *IEEE Transactions on Geoscience and Remote Sensing* 32: 562–574.

Klinger Y, Xu XW, Tapponnier P, Van der Woerd J, Lasserre C, and King G (2005) High-resolution satellite imagery mapping of the surface rupture and slip distribution of the M_W 7.8, 14 November 2001 Kokoxili earthquake, Kunlun Fault, northern Tibet, China. *Bulletin of the Seismological Society of America* 95: 1970–1987.

Lanari R, Hensley S, and Rosen PA (1998) Chirp-Z transform based SPECAN approach for phase-preserving ScanSAR image generation. *IEEE Proceedings - Radar, Sonar and Navigation* 145: 254–261.

Lanari R, Lundgren P, Manzo M, and Casu F (2004a) Satellite radar interferometry time series analysis of surface deformation for Los Angeles, California. *Geophysical Research Letters* 31: L23613 (doi:10.1029/2004GL021294).

Lanari R, Mora O, Manunta M, Mallorqui JJ, Berardino P, and Sansosti E (2004b) A small baseline approach for investigating deformations on full resolution differential SAR interferograms. *IEEE Transactions on Geoscience and Remote Sensing* 42: 1377–1386.

Lee J-S, Hoppel KW, Mango SA, and Miller AR (1992) Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing* 32: 1017–1028.

Levy F, Hsu Y, Simons M, LePrince S, and Avouac JP (2005) Distribution of coseismic slip for the 1999 Chi Chi Taiwan

earthquake: New data and implications of varying 3D fault geometry. *EOS Transactions of the American Geophysical Union* 86: B1305.

Li F and Goldstein RM (1990) Studies of multibaseline spaceborne interferometric synthetic aperture radars. *IEEE Transactions on Geoscience and Remote Sensing* 28: 88–97.

Li ZW, Ding XL, and Liu G (2004) Modeling atmospheric effects on InSAR with meteorological and continuous GPS observations: Algorithms and some test results. *Journal of Atmospheric and Solar-Terrestrial Physics* 66: 907–917.

Li ZH, Fielding EJ, Cross P, and Muller JP (2006b) Interferometric synthetic aperture radar atmospheric correction: GPS topography-dependent turbulence model. *Journal of Geophysical Research* 111: B02404.

Li Z, Muller JP, Cross P, Albert P, Fischer J, and Bennartz R (2006a) Assessment of the potential of MERIS near-infrared water vapour porducts to correct ASAR interferometric measurements. *International Journal of Remote Sensing* 27: 349–365.

Li ZH, Muller JP, Cross P, and Fielding EJ (2005) Interferometric synthetic aperture radar (InSAR) atmospheric correction: GPS, moderate resolution imaging spectroradiometer (MODIS), and InSAR integration. *Journal of Geophysical Research* 110: B03410.

Lohman RB and Simons M (2005) Some thoughts on the use of InSAR data to constrain models of surface deformation. *Geochemistry Geophysics Geosystems* 6 (doi:10.1029/ 2004GC00084).

Lohman RB, Simons M, and Savage B (2002) Location and mechanism of the Little Skull Mountain earthquake as constrained by satellite radar interferometry and seismic waveform modeling. *Journal of Geophysical Research* 107: 2118.

Lundgren P, Casu F, Manzo M, *et al.* (2004) Gravity and magma induced spreading of Mount Etna volcano revealed by satellite radar interferometry. *Geophysical Research Letters* 31: L04602 (doi:10.1029/2003GL018736).

Lundgren P, Usai S, Sansosti E, *et al.* (2001) Modeling surface deformation observed with synthetic aperture radar interferometry at Campi Flegrei caldera. *Journal of Geophysical Research* 106: 19355–19366 (doi:10.1029/2001JB000194).

Lyons S and Sandwell D (2003) Fault creep along the southern San Andreas from interferometric synthetic aperture radar, permanent scatterers, and stacking. *Journal of Geophysical Research* 108: 2047.

Massonnet D and Feigl KL (1995) Discrimination of geophysical phenomena in satellite radar interferograms. *Geophysical Research Letters* 22: 1537–1540.

Massonnet D, Rossi M, Carmona C, *et al.* (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364: 138–142.

Mattar KE and Gray AL (2002) Reducing ionospheric electron density errors in satellite radar interferometry applications. *Canadian Journal of Remote Sensing* 28: 593–600.

McGuire JJ and Segall P (2003) Imaging of aseismic fault slip transients recorded by dense geodetic networks. *Geophysical Journal International* 155: 778–788 (doi:10.1111/j.1365–246X.2003.02022.x).

Menke W (1989) *International Geophysics Series, vol.45: Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, rev. edn.

Michel R, Avouac JP, and Taboury J (1999a) Measuring near field coseismic displacements from SAR images: Application to the Landers earthquake. *Geophysical Research Letters* 26: 3017–3020.

Michel R, Avouac JP, and Taboury J (1999b) Measuring ground displacements from SAR amplitude images: Application to the Landers earthquake. *Geophysical Research Letters* 26: 875–978.

Michel R and Rignot E (1999) Flow of Glaciar Moreno, Argentina, from repeat-pass Shuttle Imaging Radar images: Comparison of the phase correlation method with radar interferometry. *Journal of Glaciology* 45: 93–100.

Moore R, Classen J, and Lin YH (1981) Scanning spaceborne synthetic aperture radar with integrated radiometer. *IEEE Transactions on Aerospace and Electronic Systems* 17: 410–420.

Onn F and Zebker HA (2006) Correction for interferometric synthetic aperture radar atmospheric phase artifacts using time series of zenith wet delay observations from a GPS network. *Journal of Geophysical Research* 111: (doi:10.1029/2005JB004012).

Peltzer G, Crampe F, Hensley S, and Rosen P (2001) Transient strain accumulation and fault interaction in the Eastern California Shear Zone. *Geology* 29: 975–978.

Peltzer G, Crampe F, and King G (1999) Evidence of nonlinear elasticity of the crust from the Mw 7.6 Manyi (Tibet) earthquake. *Science* 286: 272–276.

Pritchard ME and Simons M (2006) An aseismic slip pulse in northern Chile and along-strike variations in seismogenic behavior. *Journal of Geophysical Research* 111: (doi:10.1029/2006JB004258).

Pritchard ME, Ji C, and Simons M (2006) Distribution of slip from 11 $M_w > 6$ earthquakes in the northern Chile subduction zone. *Journal of Geophysical Research* 111: (doi:10.1029/ 2005JB004013).

Puyssegur B, Michel R, and Avouac JP (2007) Tropospheric phase delay in InSAR estimated from meteorological model and multispectral imagery. *Journal of Geophysical Research* (in press).

Raney RK (1971) Synthetic aperture imaging radar and moving targets. *IEEE Transactions on Aerospace and Electronic Systems* AE S-7: 499–505.

Raney RK (1998) Radar fundamentals: Technical perspective. In: Henderson FM and Lewis AG (eds.) *Manual of Remote Sensing Volume 2, Principles and Applications of Imaging Radar,* 3rd edn., Hoboken, NJ: J. Wiley.

Rodríguez E and Martin JM (1992) Theory and design of interferometric synthetic-aperture radars. *Proceedings of the IEEE* 139: 147–159.

Rosen PA, Hensley S, Joughin IR, *et al.* (2000) Synthetic Aperture Radar Interferometry. *Proceedings of the IEEE* 88: 333–382.

Rosen PA, Hensley S, Peltzer G, and Simons M (2004) Updated repeat orbit interferometry package released. *Eos Transactions of the American Geophysical Union* 85: 47 (doi:10.1029/2004EO050004).

Rosen PA, Hensley S, Zebker HA, Webb FH, and Fielding EJ (1996) Surface deformation and coherence measurements of kilauea volcano, Hawaii, from SIR-C radar interferometry. *Journal of Geophysical Research* 268: 1333–1336.

Schmidt DA and Burgmann R (2003) Time-dependent land uplift and subsidence in the Santa Clara Valley, california, from a large interferometric synthetic aperture radar data set. *Journal of Geophysical Research* 108: 2416–2428.

Segall P and Matthews M (1997) Time dependent inversion of geodetic data. *Journal of Geophysical Research* 102: 22391–22409.

Short NH and Gray AL (2004) Potential for RADARSAT-2 interferometry: glacier monitoring using speckle tracking. *Canadian Journal of Remote Sensing* 30: 504–509.

Simons M, Fialko Y, and Rivera L (2002) Coseismic deformation from the 1999 M_w 7.1 Hector mine, California, earthquake as inferred from InSAR and GPS observations. *Bulletin of the Seismological Society of America* 92(4): 1390–1402.

Sorenson H (1980) *Parameter Estimation*. New York: Marcel Dekker.

Tarayre H and Massonnet D (1996) Atmospheric propagation heterogeneities revealed by ERS-1 interferometry. *Geophysical Research Letters* 23(9): 989–992.

Tomiyasu K (1981) Conceptual performance of a satellite borne wide swath radar. *IEEE Transactions on Geoscience and Remote Sensing* 19: 108–116.

Touzi R and Lopes A (1996) Statistics of the stokes parameters and the complex coherence parameters in one-look and multilook speckle fields. *IEEE Transactions on Geoscience and Remote Sensing* 34: 519–531.

Usai S (1997) The use of man-made features for long time-scale InSAR. In: *Proceedings IEEE International Geoscience and Remote Sensing Symposium* pages, 1542–1544 Singapore, 1997. cdrom.

Wdowinski S, Amelung F, Miralles-Wilhelm F, Dixon T, and Carande R (2004) InSAR-based hydrology of the Everglades, South Florida. In: *Geoscience and Remote Sensing Symposium, 2004. IGARSS '04. In Proceedings, 2004 IEEE International*, vol. 3, pp. 1870–1873.

Wegmuller U (2003) Potential of interferometry point target analysis using small data stacks. *Proceedings of Fringe '03 workshop*, pp. 1–3. Frascati, Italy (cdrom).

Williams S, Bock Y, and Fang P (1998) Integrated satellite interferometry: Tropospheric noise, GPS estimates and implications for interferometric synthetic aperture radar observations. *Journal of Geophysical Research* 103: 27051–27067.

Wright TJ, Parsons BE, and Lu Z (2004) Toward mapping surface deformation in three dimensions using InSAR. *Geophysical Research Letters* 31(1): L01607.

Xu CJ, Wang H, Ge LL, Yonezawa C, and Cheng P (2006) InSAR tropospheric delay mitigation by GPS observations: A case study in Tokyo area. *Journal of Atmospheric and Solar-Terrestrial Physics* 68(6): 629–638.

Zebker HA, Rosen PA, Hensley S, and Mouganis-Mark P (1996) Analysis of active lava flows on Kilauea Volcano, Hawaii, using SIR-C radar correlation measurements. *Geology* 24: 495–498.

Zebker HA, Rosen PA, and Hensley S (1997) Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps. *Journal of Geophysical Research* 102: 7547–7563.

Zebker HA and Goldstein RM (1986) Topographic mapping from interferometric SAR observations. *Journal of Geophysical Research* 91: 4993–4999.

Zebker HA, Rosen PA, Goldstein RM, Gabriel A, and Werner CL (1994) On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. *Journal of Geophysical Research* 99: 19617–19634.

Zebker HA and Villasenor J (1992) Decorrelation in interferometric radar echoes. *IEEE Transactions on Geoscience and Remote Sensing* 30: 950–959.