CHAPTER 5

Interferometric synthetic-aperture radar (InSAR)

Daniel Dzurisin and Zhong Lu

Geodesists are, for the most part, a patient and hardworking lot. A day spent hiking to a distant peak, hours spent waiting for clouds to clear a line-of-sight between observation points, weeks spent moving methodically along a level line – such is the normal pulse of the geodetic profession. The fruits of such labors are all the more precious because they are so scarce. A good day spent with an electronic distance meter (EDM) or level typically produces fewer than a dozen data points. A year of tiltmeter output sampled at ten-minute intervals constitutes less than half a megabyte of data. All of the leveling data ever collected at Yellowstone Caldera fit comfortably on a single PC diskette! These quantities are trivial by modern data-storage standards, in spite of the considerable efforts expended to produce them.

Armed with a few hard-won data points, the geodesist must hope that they accurately characterize essential features of the deformation field, without introducing artifacts caused by sparse network coverage or bench mark instability. This is a problem at many volcanoes, where geodetic networks tend to be sparse relative to the complexity of the deformation field. Owing to difficult logistics, safety concerns, or limited resources, many networks are neither dense enough in proximal areas to show the details of shallow-seated deformation nor extensive enough distally to capture deformation from deeper sources (Chapter 11).

What a boon it would be to train a magical geodetic camera on a deforming volcano and take a picture of the entire deformation field, rather than trying to piece it together bench mark by bench mark! Imagine a technique capable of producing a detailed snapshot of the deformation field with centimeter accuracy, over lateral dimensions of tens of kilometers, without a requirement for ground access to the field area, even at night or in bad weather. Too good to be true? Not necessarily, at least under favorable conditions that exist at many of the world's volcanoes. In fact, remotely sensed, remarkably detailed images of volcano deformation started appearing in some of the world's most prestigious research journals in the mid-1990s, creating a buzz among normally stodgy geodesists – and for good reason.

Under ideal conditions, radar images acquired by satellites can provide more information about a deforming volcano than even the most intensive ground-based geodetic surveys. For a geodesist, this is remote sensing at its very best! For example, the radar interferogram in Figure 5.1 reveals that a shallow-dipping dike intruded the southwest flank of Fernandina Volcano in the Galápagos Islands and fed a lava flow that reached the sea – a conclusion confirmed by the accompanying SPOT satellite image.[1] Closer examination and modeling reveal that the maximum surface uplift

[1] The art of interpreting interferograms might seem mysterious at first, but the basic principles are not difficult to understand. Think of the pattern of concentric color bands on Fernandina's southwest flank (Figure 5.1) as a contour map of surface deformation, with each band representing 2.83 cm of slant-range change in the direction of the SAR satellite (mostly uplift in this case). Greater the number of bands, greater the amount of deformation. In this example, there are about 27 bands, so the maximum range change is about 75 cm. The pattern is elongate in the northeast–southwest direction, which means that the source has similar geometry and orientation (i.e., a NE–SW-striking dike). There are many more color bands on the southeast side of the dike than on its northwest side, so the dike dips at a shallow angle to the southeast. (*cont.*)

was 0.75 m and that the dike was about 3.8 km long, 2.3 km high, 0.86 km thick, with a 34° dip to the southeast (Jónsson *et al*., 1999). All without ever setting a bench mark or setting up a tripod! Let's take a closer look at this geodetic camera, shall we?

A breakthrough remote-sensing technique called interferometric synthetic-aperture radar (InSAR) burst on the scene with all the drama of a major earthquake.[2] The 8 July 1993 issue of *Nature* highlighted on its front cover a striking color image of the displacement field produced by the *M* 7.3 Landers earthquake, which struck about 150 km east of Los Angeles, California, on 28 June 1992 (Figure 5.2). In the accompanying article, Massonnet *et al*. (1993) introduced InSAR as a geodetic imaging technique and provided geodesists with a revolutionary new tool. Simply stated, InSAR applies interferometric image-processing techniques to two or more synthetic-aperture radar (SAR) images of the same area to measure ground-surface deformation during the time interval spanned by the image acquisitions. Ground-based InSAR had been used earlier to study the topography of the Moon (Shapiro *et al*., 1972), and satellite interferometry had been used to detect swelling of the ground produced by selective watering of fields in California's Imperial Valley (Gabriel *et al*., 1989). Not until the Landers interferogram made its splashy debut, however, did InSAR's potential for mapping ground deformation become widely recognized within the geodetic or volcanological communities.

Beneath the dazzle, InSAR rests on a foundation of classical electromagnetic theory and radar-engineering principles. A thorough treatment is beyond the scope of this chapter, but a brief explanation is in order before moving on to some recent applications of InSAR to volcanoes. More detailed information on SAR and InSAR is available in several books and review articles, including Elachi (1988), Curlander and McDonough (1991), Rodriguez and Martin (1992), Gens and Genderen (1996), Bamler and Hartl (1998), Henderson and Lewis (1998), Madsen and Zebker (1998), Massonnet and Feigl (1998), Bürgmann *et al*. (2000), Rosen *et al*. (2000), Zebker *et al*. (2000), Hensley *et al*. (2001), and Hanssen (2002).

5.1 RADAR PRINCIPLES AND TECHNIQUES

Radio and radar technology in various forms has been around for more than a century. In 1873, Scottish physicist and mathematician James Clerk Maxwell published his seminal work *A Treatise on Electricity and Magnetism*. Included in the work were four elegant partial differential equations describing the nature of electricity and magnetism – the now-famous Maxwell equations.[3] Among the far-reaching implications was the existence of electromagnetic waves traveling at the speed of light – the theoretical basis for the invention of radio and radar. In the latter part of the nineteenth century, German physicist Heinrich Rudolf Hertz expanded Maxwell's theory and performed a series of experiments that led to the development of the wireless telegraph and the radio. In 1904, Christian Hülsmeyer received a German patent for the *Telemobiloskop*, or Remote Object Viewing Device, which could detect ships at sea over ranges of up to 3 km. Sir Robert Watson-Watt successfully demonstrated the detection of an aircraft by a radio device in 1935, paving the way for

(*cont*.) The fact that the dike vented to feed a lava flow is harder to see, but careful inspection reveals a chaotic zone that interrupts the color bands and extends to the sea in the lower left corner of the interferogram. To the trained eye this reveals a resurfacing event that can be interpreted as a new lava flow, which is clearly visible in the SPOT © 1995 optical image on the right.

[2] InSAR, like most technical fields, is well stocked with specialized and potentially confusing terminology. The meanings of many unfamiliar terms should become clear as the chapter unfolds. If the language becomes impenetrable, the reader is encouraged to consult the Glossary (at the end of this book) or one of the references listed at the end of the next paragraph. The term 'interferometric synthetic-aperture radar (InSAR)' is a case in point. 'Interferometric' refers to the interaction that occurs between sinusoidal waves when they encounter one another. You might recall from an introductory physics course that the resulting 'interference' can be either constructive or destructive (i.e., the waves either reinforce or diminish one another). Among the upshots of this phenomenon are multicolored oil slicks and radar interferograms. 'Synthetic-aperture' is an engineering term. Frustrated by practical impediments to building a *really big* radar antenna, enterprising engineers conjured up an imaginary one that works! Read on.

[3] The fundamental behavior of time-varying electric and magnetic fields is encompassed in the following four differential equations (cgs): (1) $\nabla \cdot E = 4\pi\rho$, (2) $\nabla \times E = -(1/c)(\partial B/\partial t)$, (3) $\nabla \cdot B = 0$, and (4) $\nabla \times B = (4\pi/c)J + (1/c)(\partial E/\partial t)$, where $\nabla\cdot$ is the divergence, $\nabla\times$ is the curl, E is the electric field, B is the magnetic field, ρ is the charge density, c is the speed of light, and J is the vector current density. Expressed in integral form, they are known as (1) Gauss's law, (2) Faraday's law, (3) the absence of magnetic monopoles, and (4) Ampere's law. Armed with these four equations and sufficient mathematical acumen to understand the meanings of 'divergence' and 'curl,' you could have invented radio, radar, the incandescent light bulb, GPS, and InSAR. It's not volcano science, but it's impressive nonetheless.

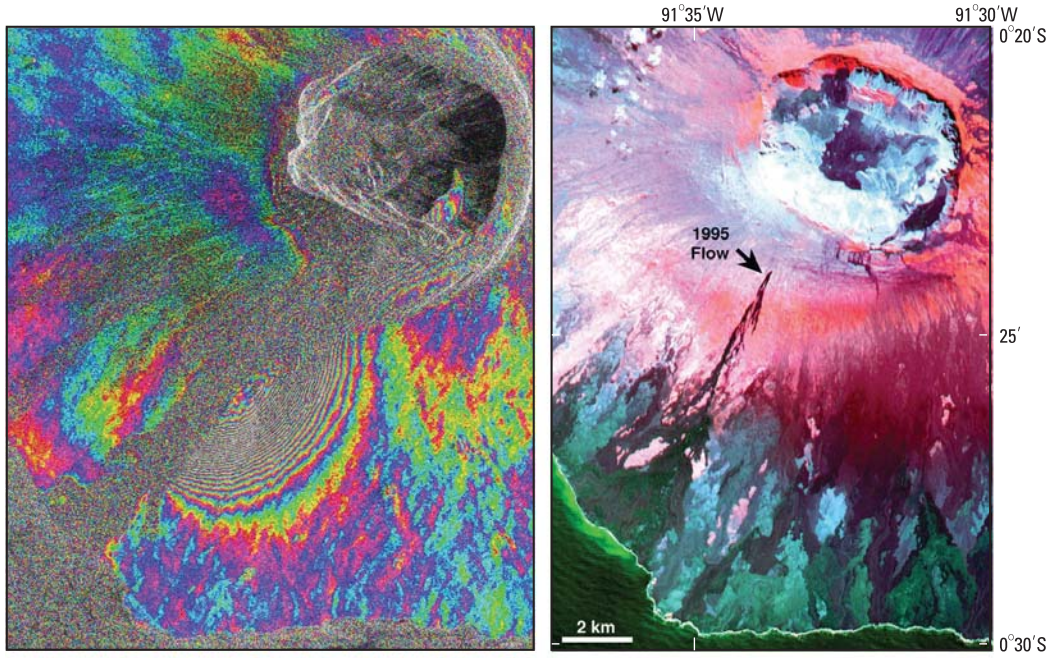

Figure 5.1. Radar interferogram (*left*) and SPOT false-color optical image (*right*) of Fernandina Volcano, Galápagos, showing the effects of an eruption on the volcano's southwest flank during January–April 1995. The interferogram was formed from ERS images acquired on 12 September 1992, and 30 September 1997 (Jónsson *et al*., 1999). Concentric color bands near the center of the image on the left are interferometric fringes that represent a maximum of about 75 cm of slant-range decrease (2.38 cm per fringe), mostly surface uplift, caused by magmatic intrusion. The resulting lava flow, which appears chaotic in the interferogram owing to a lack of phase coherence, is clearly visible in the SPOT image acquired on 6 July 1995. White areas in the upper left of the SPOT image are meteorological clouds. ERS interferogram created by Harold Garbeil and provided by Peter Mouginis-Mark (both at University of Hawaii). SPOT data copyright SPOT Image Corporation 1995.

development in the UK of the first operational radar network, called Chain Home. The Chain Home, which became operational in 1937, played a key air-defense role during the Battle of Britain (1939–1941)[4].

The term *radar* is derived from '***ra***dio ***d***etection ***a***nd ***r***anging', a phrase that hints at a few of radar's many applications, both practical and scientific. Radar systems make use of the microwave portion of the electromagnetic spectrum, at frequencies ranging from 1 billion to 1,000 billion cycles per second (1–1,000 GHz), which correspond to wavelengths from 30 cm to 0.3 mm. Short-wavelength radar signals are more sensitive to small range changes, but they do not penetrate clouds or vegetation as effectively as longer wavelength signals. This tradeoff requires a compromise between high resolution with less penetration at short wavelengths and low resolution with more penetration at long wavelengths. For imaging applications, the frequency range from 1 to 12 GHz, which encompasses X-band ($\lambda \sim 3$ cm), C-band ($\lambda \sim 5$ cm), and L-band ($\lambda \sim 20$ cm), is most useful (Table 5.1).

All radar systems employ a radio transmitter that sends out a beam of microwaves either continuously or in pulses. Two characteristics of radar that are important for volcano monitoring and many other applications are: (1) unlike optical and infrared systems that are inherently passive (i.e., they rely on natural reflected or radiated energy originating at the source), radar is an active sensor that provides its

[4] Here are some other interesting facts about radar excerpted with slight modification from the BBC online encyclopedia, h2g2 (http://www.bbc.co.uk/dna/h2g2/alabaster/A591545, *The History of Radar*). The oldest radar system was developed millions of years ago and is still used worldwide: the ultrasonic sensor of a bat. Bats emit a short 'cry' from their noses, receiving the echo with a set of two antennae (ears). A bat's echolocation system does not use electromagnetic waves, but the working principle is the same as that of a modern radar, with a chirped signal, target-tracking by Doppler estimation, pulse-repetition frequency (PRF) agility, terrain avoidance function, and fine-angle measurement based on the monopulse principle. The oldest radar warning device also was developed millions of years ago. Tiger moths (which frequently appear on bats' menus) are equipped with ears that can detect and jam the ultrasonic signal of a bat, and they have developed tactics to evade a bat's attack. Thus, electronic combat also came into being a long time ago. To my knowledge, *Homo sapiens* can lay sole claim to InSAR, however.

own illumination; and (2) owing to their longer wavelength, radar signals penetrate water clouds, diffuse ash clouds, and sparse to moderate vegetation better than visible light, enabling limited 'see-through' capability for objects that are opaque at optical wavelengths. Because radar is an active microwave system, it is equally effective in darkness and daylight, and during bad weather or good. This is a tremendous advantage for volcano monitoring, which requires round-the-clock operations during periods of unrest. Currently, this is possible only at volcanoes where an airborne radar could be deployed for an extended period of time – a rare luxury in most parts of the world. Nearly global coverage is available from a few operational SAR satellites, but their orbit repeat cycles are too long for intensive monitoring (Table 5.2). In the foreseeable future, though, daily surveillance of the world's volcanoes and other dynamic surface features such as faults and sea ice by a constellation of SAR satellites is an achievable goal.

Most radar systems can be classified as either *tracking* or *imaging* systems. Tracking radars determine the distance to an object based on the time required for a radar pulse moving at the speed of light to make the round trip from the transmitter to the object and back. If the object is moving with respect to the transmitter, its velocity can be determined from the frequency of the reflected signal, which differs from that of the transmitted signal as a result of the Doppler effect. If the receiver is arranged to reject echoes that have the same frequency as the transmitter and amplify only echoes with different frequencies, it shows only moving targets. Air traffic control radars and police speed-detectors, for example, work on this principle.

Imaging radar systems, as the name implies, are designed to record spatial reflectivity information from illuminated targets for display as a synthetic image, not unlike a photograph.[5] When microwaves strike a target, part of the energy is reflected back to its source where it can be received, amplified, and processed. The distance to and nature of the target are determined by the timing and character of the received signal. All targets do not reflect microwaves equally. The strength of the reflection depends on the size, shape, roughness, orientation, and dielectric

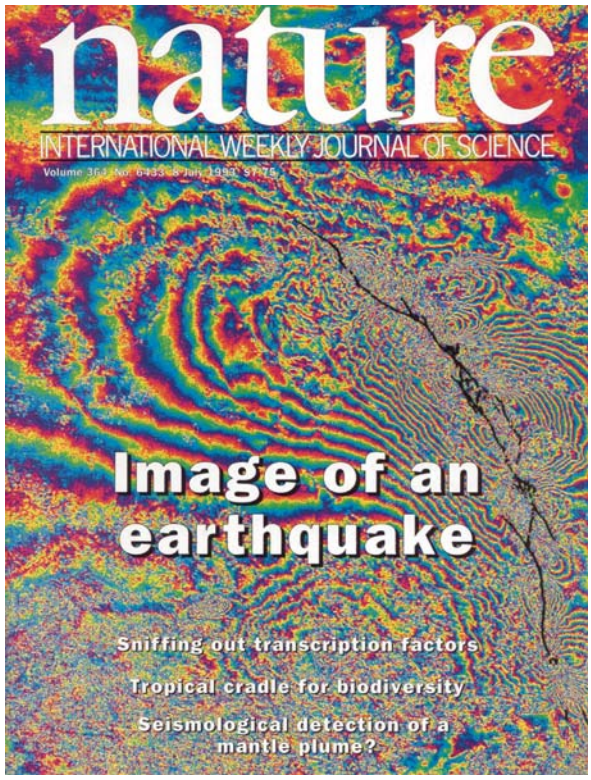

Figure 5.2. This historic satellite radar interferogram, which appeared on the cover of the 8 July 1993 issue of *Nature*, depicts in remarkable detail surface displacements caused by the *M* 7.3 Landers, California, earthquake on 28 June 1992 (Massonnet *et al*., 1993). Each color cycle from violet to red represents 2.8 cm of slant-range change. Black lines indicate the fault geometry as mapped in the field. Reprinted by permission from *Nature*, volume 364, no. 6433, 8 July 1993, copyright 1993 Macmillan Publishers Ltd.

constant (strongly influenced by moisture content) of the target. Metal objects are the best reflectors; wood and plastic produce weaker reflections. Turbulent seawater and lake ice are good reflectors; roads and highways are poor ones. Surfaces that are rough at the scale of the radar wavelength generally are brighter in radar images than smooth ones, because some of the roughness elements are oriented perpendicular to the incoming signal and reflect energy back toward the source. With smooth surfaces, most of the energy is deflected forward, away from the source, which causes them to appear dark. For this reason, a body of water appears dark on a calm day when its surface is smooth, and bright on a windy day when wave action causes a greater proportion of the incoming signal to be reflected back toward the radar.

To understand how a detailed image is assembled from a jumble of radar echoes, let us first consider the principles of real aperture radar. Then we will see

[5] We will see later that radar images also contain valuable information about surface material properties and satellite-to-ground range. The processed range data underlie the colorful fringes in radar interferograms (Section 5.2).

Table 5.1. Frequency and wavelength bands commonly used for imaging radar systems.

Frequency	Wavelength	Designation	Example
1–2 GHz	30–15 cm	L-Band	JERS-1, SIR-C, Seasat
2–4 GHz	15–7.5 cm	S-Band	NEXRAD weather radar (U.S.A.), Digital Audio Radio Satellite (DARS)
4–8 GHz	7.5–3.75 cm	C-Band	SIR-C, ERS-1, ERS-2 RADARSAT-1 Envisat
8–12 GHz	3.75–2.50 cm	X-Band	SIR-C/X-SAR

how a combination of physics, mathematics, and human ingenuity can transform a small radar antenna into a very large one that serves as a lens for our geodetic camera.

5.1.1 Real-aperture imaging radar systems

Unlike the cameras used to make vertical aerial photographs, imaging radar systems point to the side rather than straight down so the arrival path of the radar signal is oblique to the surface being imaged. This is necessary to better distinguish among targets located at different distances from the radar. Side-looking radar systems are classified as either *real-aperture* (called SLAR for side-looking airborne radar or SLR for side-looking radar) or *synthetic-aperture* radar (SAR).

SLR imaging systems use a long, straight antenna mounted on a moving platform, usually an aircraft or satellite, with the antenna's longitudinal axis parallel to the flight path (Figure 5.3). The antenna emits microwave pulses in a beam directed perpendicular to the flight path and obliquely downward toward the surface of the Earth. These pulses strike the surface along a narrow swath and are scattered, usually in many directions, including the direction back to the radar. The antenna captures a very small fraction of the energy it transmits, typically less than 0.000 1%, from the returning pulse echoes. The echoes arrive at the antenna at different times, depending mainly on the distance from the antenna to countless natural reflectors on the ground.[6] By keeping track of the arrival times, an SLR profiles its distance from the ground in the direction perpendicular to the flight path (i.e., the cross-track or range direction). Motion of the aircraft or satellite advances the profile in the direction parallel to the flight path (i.e., the along-track or azimuth direction). By overlapping successive swaths, a large area on the ground can be mapped at a resolution of a few tens of meters (Curlander and McDonough, 1991; Madsen and Zebker, 1998; Hanssen, 2002). The resulting images contain information about surface slopes (i.e., topography) and other factors that influence radar reflectivity, including roughness and moisture content.

The dimensions of the antenna determine the spread of the radar beam and thus the size of its footprint on the ground. The angular beam width is $\beta_r \approx \lambda/D$ in the range direction and $\beta_a \approx \lambda/L$ in the azimuth direction, where D and L are the width and length of the antenna, respectively. The corresponding dimensions of the antenna footprint are:

$$W_r \approx \frac{\lambda R_m}{D \cos\theta_m} \tag{5.1}$$

and

$$W_a \approx \frac{\lambda R_m}{L} \tag{5.2}$$

where W_r is the width of the footprint in the range direction (i.e., the swath width), W_a is the width of the footprint in the azimuth direction, λ is the wavelength of the radar, R_m is the slant range from the antenna to the middle of the footprint, and θ_m is the incidence angle at the middle of the footprint (Figure 5.3).[7] For the European Remote-Sensing satellites (ERS-1 and ERS-2), $D = 1\,\text{m}$, $L = 10\,\text{m}$,

[6] Propagation delays in the ionosphere or troposphere also affect the round-trip travel times of the radar pulses. Inhomogeneities in the ionospheric electron-density or tropospheric water vapor concentration result in spatially variable delays that produce fringes in interferograms, which can be mistaken for ground deformation. Usually, such path-delay effects can be identified and discounted by comparing two or more interferograms of the same scene created from independent pairs of images (Section 5.2.8).

[7] The edges of the antenna footprint are not sharp, so its dimensions are somewhat arbitrary. The energy pattern in radar systems usually is considered to be constant between the half power (3 dB) angles, in which case the theoretical beam widths are $\beta_r = 0.886\lambda/D$ and $\beta_a = 0.886\lambda/L$. For ERS-1 and ERS-2, the beam widths defined in this way are $\beta_r = 2.870°$ and $\beta_a = 0.287°$ (Hanssen, 2002, p. 25).

Table 5.2. Satellite imaging radar systems.

Mission	Agency	Period of operation[1]	Orbit repeat cycle	Frequency	Wavelength	Incidence angle at swath center	Resolution[2]
ERS-1 and ERS-2	European Space Agency (ESA)	July 1991 to March 2000 (ERS-1) April 1995 to present (ERS-2)	35 days[3]	C-band 5.3 GHz	5.66 cm	23 degrees	30 m
JERS-1	National Aeronautics and Space Development Agency of Japan (NASDA) and Ministry of International Trade and Industry (MITI)	February 1992 to October 1998	44 days	L-band 1.275 GHz	23.5 cm	39 degrees	18 m
RADARSAT-1	Canadian Space Agency (CSA)	November 1995 to present	24 days	C-band 5.3 GHz	5.66 cm	10 to 59 degrees	8 m to 100 m
SIR-C/X-SAR	National Aeronautics and Space Administration (NASA), German Space Agency (DARA), and Italian Space Agency (ASI)	9–20 April 1994 (STS-59) 30 September to 11 October 1994 (STS-68)	N/A[4]	L-band 1.249 GHz C-band 5.298 GHz X-band 9.6 GHz	24.0 cm 5.66 cm 3.1 cm	17 to 63 degrees (L- and C-band) 54 degrees (X-band)	10 m to 200 m (30 m typical)

SRTM	National Aeronautics and Space Administration (NASA)	11-22 February 2000 (Space Shuttle mission STS-99)	N/A	C-band 5.3 GHz X-band 9.6 GHz	5.8 cm 3.1 cm	17 to 63 degrees (L-band and C-band) 54 degrees (X-band)	30 m
Envisat	European Space Agency (ESA)	March 2002 to present	35 days	C-band 5.331 GHz	5.63 cm	14 to 45 degrees	30 m
ALOS	Japan Aerospace Exploration Agency (JAXA), formerly National Aeronautics and Space Development Agency of Japan (NASDA)	Launched 24 January 2006	46 days	L-band 1.275 GHz	23.5 cm	8 to 60 degrees	10 m to 100 m

[1] Information was current in June 2005.
[2] The resolution of radar images differs in the range and azimuth directions and depends on the type of data processing (single-look or multi-look). A representative value is shown here for comparison purposes.
[3] To accomplish various mission objectives, the ERS-1 repeat cycle was shortened to 3 days (43 orbits) from 25 July 1991 to 1 April 1992 and from 23 December 1993 to 9 April 1994. From 10 April 1994 to 20 March 1995, the cycle was extended to 168 days (2,411 orbits) to support a radar altimetry mission. At other times for ERS-1 and throughout the mission of ERS-2, the repeat cycle was 35 days (501 orbits). From 17 August 1995 to 2 June 1996, ERS-1 and ERS-2 flew a tandem mission in which ERS-2 followed 30 minutes behind ERS-1 in the same orbital plane. As a result, the same swath on the ground was acquired by ERS-2 one day later than by ERS-1. The short repeat cycle resulted in good interferometric coherence in most areas, which facilitates the production of DEMs. ERS-1 was retired on 10 March 2000. ERS-2 was operational, with some limitations, in June 2005.
[4] Although repeat-pass interferometry was not a primary mission objective, more than one million square kilometers of repeat-pass data were obtained during STS-68.

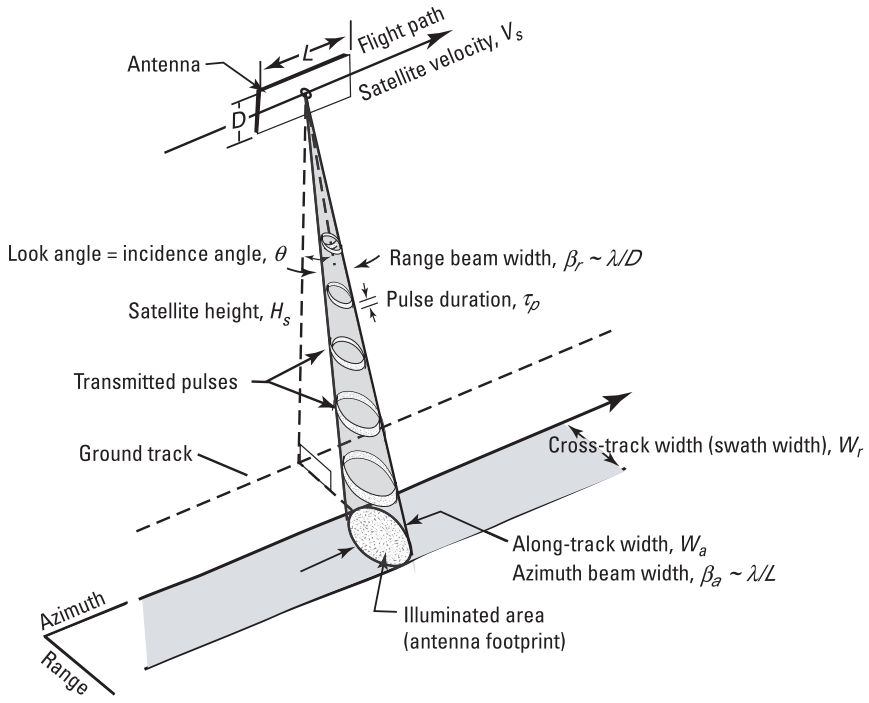

Figure 5.3. Scanning configuration for a right-looking radar, after Olmsted (1993). The dimensions of the antenna (*D* and *L*) combined with the radar wavelength (λ) determine the beam width in the range and azimuth directions: $\beta_r \sim \lambda/D$ and $\beta_a \sim \lambda/L$, respectively. The beam widths, in turn, determine the size of the antenna footprint on the ground.

$R_m \sim 850\,\text{km}$, $\lambda = 5.66\,\text{cm}$ (C band), and $\theta \sim 23^\circ$. The theoretical footprint dimensions are, therefore, $W_r \sim 50\,\text{km}$ and $W_a \sim 5\,\text{km}$. In practice, the beam width in the range direction is broadened to achieve more even energy distribution across the full swath. This results in practical beam widths of $\beta_r = 5.4^\circ$ and $\beta_a = 0.228^\circ$ for ERS-1 and ERS-2, which correspond to $W_r = 102.5\,\text{km}$ and $W_a = 4.8\,\text{km}$ (Hanssen, 2002, p. 26).

5.1.2 Ground resolution of real-aperture imaging radars

The spatial resolution of an imaging radar system depends on different factors in the range and azimuth directions. Intuitively, the range resolution must depend both on the pulse duration and the incidence angle of the radar beam. In the first case, a shorter pulse duration provides a finer probe of the distance from antenna to ground, so range resolution should improve at shorter pulse durations. The situation is akin to measuring the width of this page using a ruler marked off either in millimeters or centimeters – the more-closely-spaced markings give better precision, for the same reason that short-duration radar pulses allow better range resolution than long-duration ones. The angle of incidence of the incoming radar signal must also play a role, because a vertically directed radar would receive return echoes from all objects in the scene at nearly the same time, making it very difficult to distinguish among them. At the other extreme, distances across the ground swath for an SLR with a high angle of incidence (i.e., a grazing beam) would differ by a large amount so the range resolution would be good, but the return signal strength would be poor because most of the signal would be scattered forward away from the radar. The ERS satellites accommodate this tradeoff with a fixed incidence angle of 23° at the middle of the swath. The radar aboard the Canadian RADARSAT-1 satellite can be steered electronically to produce incidence angles from 10° to 59° at the middle of the swath.

Two objects can be distinguished in the range direction only if the leading edge of the pulse echo from the more distant object arrives at the antenna later than the trailing edge of the echo from the nearer object. In other words, the two slant ranges must differ by at least half of the pulse length for the objects to be resolvable (Figure 5.4):

$$\Delta R_s \geq \frac{c\tau_p}{2} \tag{5.3}$$

where ΔR_s is the slant-range difference, c is the speed of light, and τ_p is the pulse duration. Otherwise, the

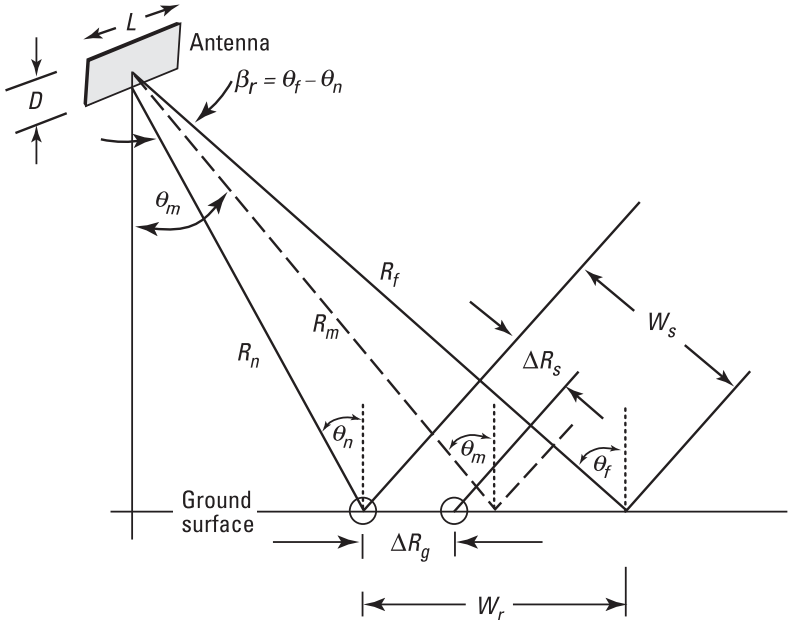

Figure 5.4. The swath width W_r for a real-aperture radar is determined by the wavelength λ, the antenna width D, the slant range at the middle of the antenna footprint R_m, and the incidence angle at the middle of the footprint θ_m. The ground-range resolution ΔR_g is related to the pulse duration τ_p and the incidence angle. R_n and R_f are the slant ranges at near range and far range, respectively, and β_r is the angular beam width in the range direction. See text for details. After Olmsted (1993).

echoes would arrive at the antenna at the end of their round trip separated by less than one pulse length and be indistinguishable. This constraint leads to the following expression for the theoretical ground-range resolution as a function of pulse duration and incidence angle:

$$\Delta R_g = \frac{c\tau_p}{2\sin\theta} \tag{5.4}$$

where ΔR_g is the ground-range resolution and θ is the incidence angle at a particular ground range.

The best theoretical ground-range resolution would be produced by a combination of extremely short pulse duration and large incidence angle. For example, an X-band radar operating at a frequency of 10 GHz with a pulse duration of 10^{-8} s at an incidence angle of 60° would have a theoretical ground-range resolution of 1.7 m. Unfortunately, there is a catch. The quality of the radar image depends on the ratio of the return signal strength to the ambient system noise. The greater the signal-to-noise ratio, the more fidelity the radar image will have. To achieve meter-scale resolution over typical terrain, the pulse duration would have to be too short to generate enough energy to achieve an adequate signal-to-noise ratio. High-incidence angles pose the same problem. For the case of grazing incidence, forward scattering dominates and the return from back scattering is negligible. In other words, engineering constraints and the principles of wave scattering preclude the realization of a short-pulse-duration, high-incidence-angle SLR.

This difficulty can be overcome, however, by using a technique known as *range compression* together with a type of signal processing called *matched filtering*. A thorough description of these techniques can be found in Curlander and McDonough (1991). For our purposes, suffice it to say that the approach is based on transmitting and receiving not a single frequency but a spread of frequencies in the microwave range. The goal is to correlate specific outgoing and incoming pulses – a daunting task when one considers that a typical X-band radar operates at a frequency of 10 billion cycles per second with a pulse duration of ten one-billionths of a second! To solve this problem, radar engineers use some nifty mathematics and a special kind of pulse called a *chirp*. The frequency of a typical radar chirp increases linearly with time, say from 9.99 GHz to 10.01 GHz. In the auditory frequency range, such a pulse would sound like a chirp – hence the term. When the returned echo of a linearly frequency modulated pulse (a chirp) is correlated with the known transmitted signal, the autocorrelation function is nearly zero except for a narrow spike that corresponds to the round trip travel time of the pulse. The echo is thus 'compressed' into a form that can be readily identified with an electronic match filter tuned to the appropriate range of travel times.

With these techniques it is possible to achieve a ground-range resolution given by:

$$\Delta R_g = \frac{c}{2B\sin\theta} \tag{5.5}$$

where B is the frequency bandwidth of the transmitted radar pulse. The range resolution can be improved by increasing the frequency bandwidth of the pulse or the incidence angle, as long as the signal-to-noise ratio remains high enough. Comparing the ground resolution without range compression (5.4) with that after range compression (5.5), we see that the improvement in ground-range resolution is equal to $\tau_p \times B$, which is commonly called the *pulse compression ratio*. For ERS-1 and ERS-2, $\tau_p = 37.1\,\mu s$, and $B = 15.5\,MHz$, so $\Delta R_g = 31\,m$ at near range ($\theta \sim 18^\circ$), 25 m at mid-range, and 22 m at far range ($\theta \sim 26^\circ$). The pulse compression ratio is about 575, so the use of a frequency modulated chirp waveform and the matched filtering technique equivalently 'compresses' the ERS-1, ERS-2 radar pulse from 37.1 μs into 64.5 ns. Nominally, the ground range resolution ΔR_g is slightly larger than the sampling spacing $c/(2f_s \sin\theta)$, where f_s is the sampling frequency in the range direction. For ERS-1 and ERS-2, $f_s = 18.96\,MHz$, so the sampling spacing is about 26 m at near range, 20 m at mid-range, and 18 m at far range.

We can see that the range resolution is controlled by the type of frequency modulated waveform and the way in which the returned signal is compressed. Both SLR and SAR systems resolve targets in the range direction in the same way. What distinguishes SLR from SAR is resolution in the azimuth direction (Curlander and McDonough, 1991).

In the azimuth direction, two objects at the same range can only be distinguished if they are not both illuminated by the beam at the same time. Therefore, the best azimuth resolution attainable with a real-aperture radar is, from (5.2) (Curlander and McDonough, 1991):

$$\Delta A_g = W_a = \frac{\lambda R_m}{L} \tag{5.6}$$

where ΔA_g is the ground-azimuth resolution, akin to the ground-range resolution ΔR_g in (5.5). A 3-m-long antenna operating at a wavelength of 5.6 cm (C-band) on an aircraft flying 1,400 m above terrain with a beam incidence angle $\theta_m = 30^\circ$ ($R_m \sim 1,600\,m$) has an azimuth resolution of about 30 m. The same C-band SLR operated from orbit at a slant range of 850 km from terrain would produce an azimuth resolution of about 16 km, which is much too coarse for most applications. To reduce this number to 30 m would require an antenna about 1.6 km long, which is not likely to be deployed in orbit anytime soon. The solution to this dilemma lies in conjuring up a huge synthetic antenna by taking advantage of the along-track motion of a real antenna and using the Doppler effect to identify signals reflected from specific parts of the antenna footprint.

5.1.3 Synthetic-aperture radar

Carl A. Wiley (1965) was the first to realize that the frequency spread in the echo signal produced by the Doppler effect could be used to synthesize a much larger antenna aperture than is achievable with real-aperture radars (hence the term *synthetic-aperture radar*, or SAR) (Figure 5.5). This breakthrough led to dramatic improvement in the resolution of SLRs, paving the way for satellite-borne SARs with meter-scale resolution over swath widths of 50–150 km.[8]

Relative motion between an observer and a source of waves causes a frequency shift known as the Doppler effect. When a source emitting waves with frequency f approaches or recedes from the observer at velocity v, the observed frequency f' is:

$$f' = \sqrt{\frac{1 + (v/c)}{1 - (v/c)}} \times f \tag{5.7}$$

where c is the speed of light and v is positive for an approaching source and negative for a receding source. Thus, an approaching source exhibits a higher frequency and a receding source a lower frequency than the frequency emitted by the source. This effect is familiar to anyone who has listened to a passing train whistle or emergency siren, which decreases in pitch as its velocity relative to the observer changes from positive (approaching) to negative (receding).

How can we take advantage of this classical principle of wave propagation to transform a small real antenna into a huge synthetic one? Notice from Figure 5.5 that targets in the radar illumination footprint which are located at the same range but different azimuths subtend slightly different angles with respect to the look direction of the antenna. As

[8] The resolution of SAR images differs in the range and azimuth directions and depends on how the raw data are processed (i.e., single-look or multi-look processing). For ERS-1 and ERS-2, the range and azimuth resolutions of single-look complex images are about 25 m and 5 m, respectively. The resolution of ERS multi-look images with 1 × 5 pixel averaging is about 30 m. Averaging several multi-look images has the desired effect of suppressing speckle, albeit at the expense of degraded spatial resolution. For more information about multi-look processing, see 'multi-look' and 'speckle' in the **Glossary**, or one of the books listed near the beginning of this chapter.

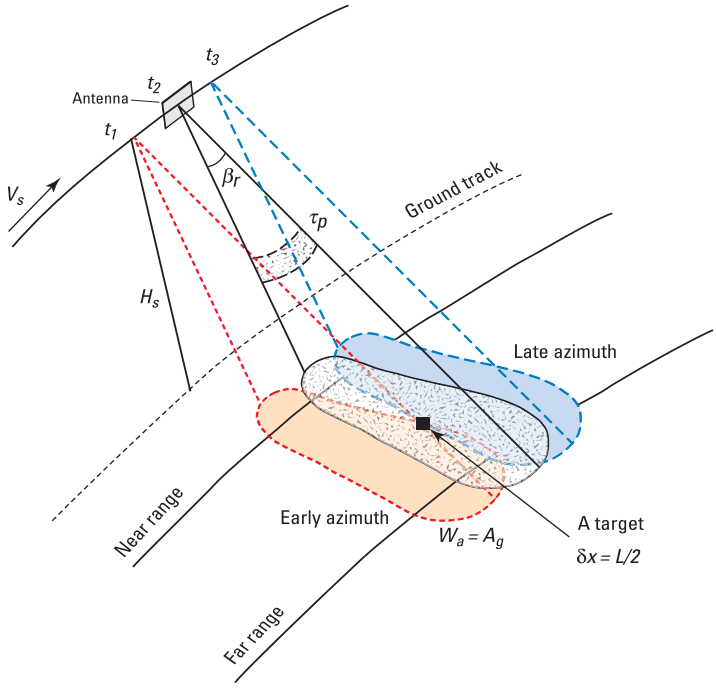

Figure 5.5. At any time t_2, a real-perture radar of length L illuminates a footprint on the ground of dimension $W_a = \lambda R_m/L$ in the azimuth direction. A synthetic-aperture radar takes advantage of the fact that the same target is illuminated continuously and remains in the footprint from time t_1 to t_3 as the radar travels along its flight path at velocity V_s. All of these echoes are used to synthesize a much longer antenna with an azimuth resolution improved from $A_g = W_a$ to $\delta x \sim L/2$ (Curlander and McDonough, 1991). For ERS-1 and ERS-2 ($L = 10$ m), the use of synthetic aperture improves the azimuth resolution by about three orders of magnitude, from $W_a \sim 4.8$ km to $\delta x \sim 5$ m.

a result, they have slightly different velocities at any given moment relative to the antenna. At the leading edge of the footprint the radar is approaching the target, while at the trailing edge it is receding. Therefore, the signal echoed from each target will have its frequency shifted a different amount from the transmitted frequency. Recall that the limitation on azimuth resolution for real-aperture radars derives from the requirement that, to be resolvable, two objects at the same range can not be within the beam's footprint at the same time. The Doppler effect provides a means of distinguishing two such objects, because the antenna will see a return from each that is frequency shifted by an amount that can be attributed to a specific azimuth within the antenna footprint. The procedure is known as *azimuth compression*. Through the wonders of range and azimuth compression, each of myriad echoes from resolution cells throughout the footprint can be assigned unique coordinates in both range and azimuth to produce a 'focused' SAR image (Figure 5.6).

The situation is complicated, however, by the fact that the round-trip travel time for pulses from the ERS-1 and ERS-2 radars ($R_m \sim 850$ km) is nearly 10 times the pulse repetition interval.[9] The SAR antenna switches between transmit and receive modes and back again after each pulse, but at any given instant the received signal is not from the last pulse transmitted, which has yet to reach the ground, but rather from an earlier pulse which has echoed from numerous resolution cells in the illuminated footprint. In this respect, the ERS-1 and ERS-2 radars are akin to a juggler who keeps 10 balls in the air by tossing #1, #2, #3 . . . #10, then catching #1, tossing #11, catching #2, tossing #12, etc. The feat is all the more remarkable when one considers that the ERS-1 and ERS-2 radars transmit and receive an average of 1,680 pulses *every second*! Still, it gets worse. Because the antenna footprint (100 km × 5 km) is much larger than the size of a resolution cell (20 m × 4 m), each sample of the return signal includes information from many different cells on the ground, and each cell contributes to many different samples. Rather than a melodious

[9] The pulse repetition interval (PRI) is the reciprocal of the pulse-repetition frequency (PRF), which for ERS-1 and ERS-2 is about 1,680 Hz. The time required for a pulse to make the SAR-to-surface round trip is $2R_n/c$, where R_n is the slant-range distance to near range and c is the speed of light. The number of pulses that could be transmitted before the echo of the first pulse is received is: $1 + 2R_n/c \cdot \text{PRF} = 10.3$.

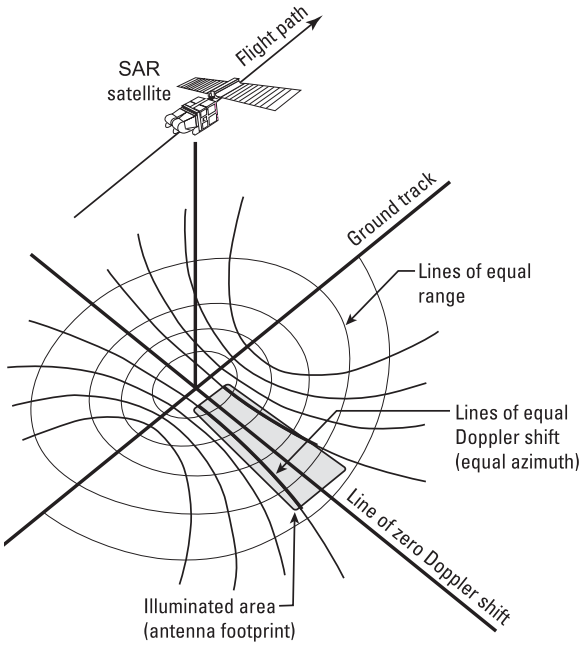

Figure 5.6. SAR images are focused by assigning range and azimuth coordinates to the echoes received from each resolution cell on the ground, which are delineated by the intersections of lines of equal range and lines of equal Doppler shift. The ERS-1 and ERS-2 satellites emit one radar pulse for each 4 m traveled along the flight path. The resulting single-look images have range and azimuth resolutions of 20–30 m and ~5 m respectively (Wright, 2000). Single-look images are resampled using 1 × 5 pixel averaging to produce standard ERS multi-look (see the Glossary at the back of this book) images, which have square pixels and ~30 m ground resolution in both range and azimuth.

sequence of well-defined chirps, the return signal is actually a cacophony of blurred echoes! How can a SAR data processor possibly make sense of such babble?

The following analogy is not perfect, but useful nonetheless. Imagine yourself standing at center stage in a very large auditorium, with 100 people seated across the front row. The room is quiet until suddenly all 100 people start talking simultaneously. From your central vantage point, you hear only a din of indistinct voices. Recalling your high school physics, you stroll to the side of the stage and then pace deliberately across the front while listening intently to the chatter. This time, you have the benefit of hearing a time history of each voice, including subtle frequency changes produced by your motion relative to the speakers. Distinct messages begin to emerge from the racket. Everyone continues to talk throughout your transit of the stage, but your motion allows you to distinguish 100 different voices as you pass by. In our SAR analogy, you have just sorted the echoes from near-range pixels by azimuth.

Now imagine that every seat in the auditorium is occupied: 100 seats by 100 rows. Having all 10,000 people start to talk simultaneously would not be analogus, however, because a SAR is an active system (i.e., it initiates the action by transmitting a pulse to the ground and then listening for the echoes). So let us say that you instruct the audience to close their eyes and respond in kind each time they hear you clap your hands, which you do repeatedly as you pace across the stage. The speed of sound is finite, so people in the last row hear your first clap a short time after those in the first row, and the sound of their responses takes a correspondingly longer time to reach you. You hear a rolling series of handclaps from first row to last as the 'echoes' of your own clap are sorted by range. You can imagine the racket that would be created by the audience's responses to your repeated handclaps as you strolled across the stage, but in theory you could sort the sounds by seat and row (i.e., by azimuth and range) and attribute them to 10,000 separate sources (resolution cells). A SAR processor goes through an analogous procedure to create a focused SAR image, albeit with more mathematical rigor and considerably less panache!

Let us get back to the math. As the radar antenna moves along its flight path its velocity relative to points on the ground changes, inducing a corresponding Doppler shift in the frequency of the echoes (Curlander and McDonough, 1991, p. 17):

$$f_d = \frac{2 \times (v_s \sin \varphi)}{\lambda} \approx \frac{2 v_s x}{\lambda R} \tag{5.8}$$

where f_d is the Doppler shift, v_s is the antenna's velocity relative to a point on the ground, φ is the angle from a line perpendicular to the ground track and is defined on the plane constituted by the flight direction and the point on the ground, x is the along-track distance measured from the point of closest approach, R is the slant range, and the factor of 2 arises from the two-way travel time (Figure 5.7). Thus, the Doppler frequency of the echoes can be used to determine the along-track coordinate (azimuth) of a particular echo sample:

$$x = \frac{\lambda R f_d}{2 v_s} \tag{5.9}$$

As the target passes through the antenna beam, it exhibits a constantly changing Doppler shift that is directly related to the position of the radar and the azimuth position of the target. For each azimuth

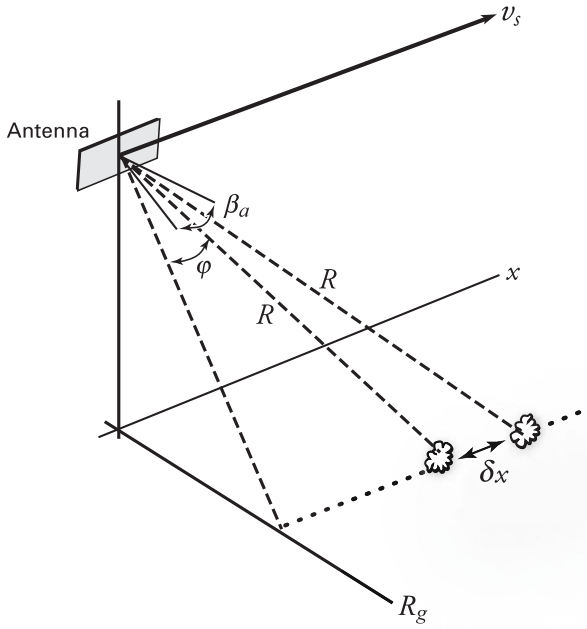

Figure 5.7. Azimuth resolution of a synthetic-aperture radar, after Curlander and McDonough (1991). v_s is the velocity of the satellite or aircraft, β_a is the angular beam width in azimuth, φ is measured from a line perpendicular to the ground track, R is the slant range, R_g is the ground range, x is the along-track distance measured from the point of closest approach, and δx is the azimuth resolution of the SAR.

position there is a unique Doppler history. So the echoes from multiple targets at the same range that are illuminated simultaneously by the beam can be discriminated on the basis of their Doppler shifts. The focused azimuth resolution δx of a SAR depends on the precision of the Doppler shift measurement:

$$\delta x = \left(\frac{\lambda R}{2v_s}\right) \times \delta f_d \tag{5.10}$$

where δf_d, the measurement resolution of the Doppler shift, is nominally the inverse of the time span Δt, for which the target remains illuminated (Figure 5.5):

$$\Delta t = t_3 - t_1 = \frac{A_g}{v_s} = \frac{R\lambda}{Lv_s} \tag{5.11}$$

Therefore, the azimuth resolution of a focused SAR image is approximately (Curlander and McDonough, 1991, p. 20):

$$\delta x = \left(\frac{\lambda R}{2v_s}\right) \times \left(\frac{Lv_s}{R\lambda}\right) = \frac{L}{2} \tag{5.12}$$

This simple result seems surprising because it shows that the azimuth resolution of a SAR improves as the antenna size *decreases*, which is opposite to the rule for a real-aperture radar. The smaller, the better would seem to be the rule in this case. This seems like good news, but again there is a catch. The same antenna must act both as a transmitter and as a receiver, so it must alternate between transmitting and listening modes. The effective gain of the antenna is proportional to the square of its real aperture, so reducing the antenna size to improve the azimuth resolution produces a tradeoff with the maximum obtainable signal-to-noise ratio (Henderson and Lewis, 1998). A very small antenna might produce excellent azimuth resolution, but with an unacceptably poor signal-to-noise ratio. The determining factor is the pulse repetition interval PRI = 1/PRF, where PRF is the pulse repetition frequency, because the difference in time between echoes from the near range $2R_n/c$ and far range $2R_f/c$ must be less than the time between pulses (Figure 5.4):

$$W_s = R_f - R_n \leq \frac{c}{2 \times \text{PRF}} \tag{5.13}$$

Otherwise, far-range echoes from one pulse would be confused with near-range echoes from the next. The maximum swath width is therefore:

$$W_g \approx \frac{c}{2 \times \text{PRF} \sin\theta} \tag{5.14}$$

Hence, large swath widths require small PRFs. For ERS-1 and ERS-2, PRF = 1,640–1,720 Hz and the maximum theoretical swath width is about 230 km.

To focus a SAR image in azimuth, a SAR processor must be able to relate an observed incremental phase change to a specific Doppler frequency. Curlander and McDonough (1991) show that this requires the bandwidth of the Doppler signal $B_d \approx (2\beta_a v_s/\lambda)$ to be less than the PRF. Recalling that $\beta_a \approx (\lambda/L)$ and $\delta x = (L/2)$, we can write:

$$\frac{2v_s}{L} = \frac{v_s}{\delta x} < \text{PRF} \tag{5.15}$$

So in the time between successive pulses 1/PRF the antenna must travel a distance along its flight path of no more than half its own length $L/2$. Combining this result with that from the previous paragraph, we see that although the azimuth resolution δx is increased for smaller antennae, smaller antennae require larger pulse repetition frequencies, which reduce the swath width (Wright, 2000, p. 26). These relationships place a lower bound on the size of a practical SAR antenna. Most SAR systems designed for Earth orbit use an antenna 1–4 m wide and 10–15 m long, with a look angle in the range 10 to 60 degrees to illuminate a footprint

of 50–150 km in range by 5–15 km in azimuth, producing a single-look ground resolution of 4–10 m in azimuth and 10–20 m in range (Table 5.2).

5.1.4 Characteristics of SAR images

A focused SAR image contains information about the SAR-to-surface range, the radar beam's two-way transit through the atmosphere, and its interaction with a myriad of small reflectors on the ground. Each illuminated pebble, tree, and fence post leaves a unique imprint on the reflected signal. In addition, because microwaves penetrate the surface for a distance that depends on their wavelength and on the dielectric constant of the surface material, the reflected signal also contains information about the shallow subsurface – primarily about the electrical conductivity, which in turn is a strong function of water content, but also about any buried reflectors such as a gravel layer, lost coins and keys, and so forth. When the signal scattered from a given resolution cell on the ground is received back at the SAR, it contains a microwave fingerprint of the cell encoded as amplitude (i.e., intensity or brightness) and phase information, which is mathematically distinctive.[10] This information can be recorded as a single complex number $Ae^{i\phi}$, with A representing the amplitude of the echo and ϕ representing its phase.[11]

For illustrative purposes, raw SAR data can be displayed as an image by assigning values to adjacent pixels equal to the real part and imaginary part of each successive echo sample, starting in the upper left corner of the image (i.e., pixel $1 = A_1 \cos\phi_1$, pixel $2 = A_1 \sin\phi_1$, pixel $3 = A_2 \cos\phi_2$, pixel $4 = A_2 \sin\phi_2$, etc.).[12] Each line in such an image corresponds to a single pulse echo. The average PRF for ERS-1 and ERS-2 is 1,680 Hz and the return signal is sampled at $f_s = 18.96$ MHz. The maximum number of samples per line in a raw ERS image is therefore: 18.96 MHz/1,680 Hz = 11,285 samples. In practice, the ERS-1 and ERS-2 radars spend only about half of the time between successive pulses in sampling the return signal, so the raw images contain 5,616 samples per line. Each raw image includes about 27,000 lines, which corresponds to 16 s of continuous data collection. At 16 bits/sample, each image comprises about 300 Mbytes. The samples in a raw image do not correspond to resolution cells on the ground until the data are focused by a SAR processor. Not surprisingly, an image constructed in this way appears random. When the amplitude and phase data are displayed separately, the result is only marginally better (Figure 5.8). This is because each echo contains information from many resolution cells on the ground, which are spread out in both range and azimuth. As a result, the information from individual cells is 'smeared' across the images. The unfocused phase image is entirely random, because each pixel includes contributions from many elementary reflectors in each of many cells, which combine to produce a random phase signature.

A SAR processor sharpens the unfocused image in range using the matched filter technique described earlier. The process is called range compression because information from each echo that had been spread from near range to far range is compressed into specific range bins. The resulting amplitude image resembles the target area, but still is unfocused in azimuth (Figure 5.9). The final step is to take advantage of the differing Doppler shifts among echo samples from the same range to compress the data in azimuth as well. A fully focused SAR amplitude image is similar in many ways to an aerial photograph. The dimensions on the ground of a standard, focused ERS-1 or ERS-2 image are about 110 km in range and about 110 km in azimuth.

There are some important differences between radar-amplitude images and vertical aerial photographs. Radar images are subject to foreshortening, layover, and shadowing. If the ground surface is level, the echo from a far-range reflector arrives at the SAR later than the echo from a near-range reflector. The time difference Δt is mapped correctly into a ground-range difference and the distance between the two reflectors is properly rendered in the radar image. However, if the surface slopes toward the SAR, Δt is reduced and the distance

[10] It is important to note that the term 'phase' is used here to denote a dimensionless number from 0 to 2π (phase is modulo 2π for a sinusoidal wave), which differs from the usage in Chapter 4 for GPS.

[11] A complex number has both a real part and an imaginary part. In this case, $Ae^{i\phi} = A\cos\phi + iA\sin\phi$, where A is the amplitude of the signal scattered back to the SAR antenna from a given resolution cell on the ground, ϕ is the corresponding phase, $A\cos\phi$ is the real part of $Ae^{i\phi}$, and $A\sin\phi$ is the imaginary part. Thus, both the amplitude and phase of the return signal are recorded in the complex number $Ae^{i\phi}$.

[12] For ERS-1 and ERS-2, the raw data are stored as 16 bit integers with bits 1–8 (byte 1) corresponding to the real part of the complex-valued echo and bits 9–16 (byte 2) corresponding to the imaginary part. Bytes 1 and 2 are sometimes called the I-channel and Q-channel (in-phase, quadrature), respectively. I suspect that SAR engineers call this sort of ingenuity 'fun.' I'm just glad there are real volcanoes on which to train their ingenious, synthetic, partly imaginary, geodetic camera!

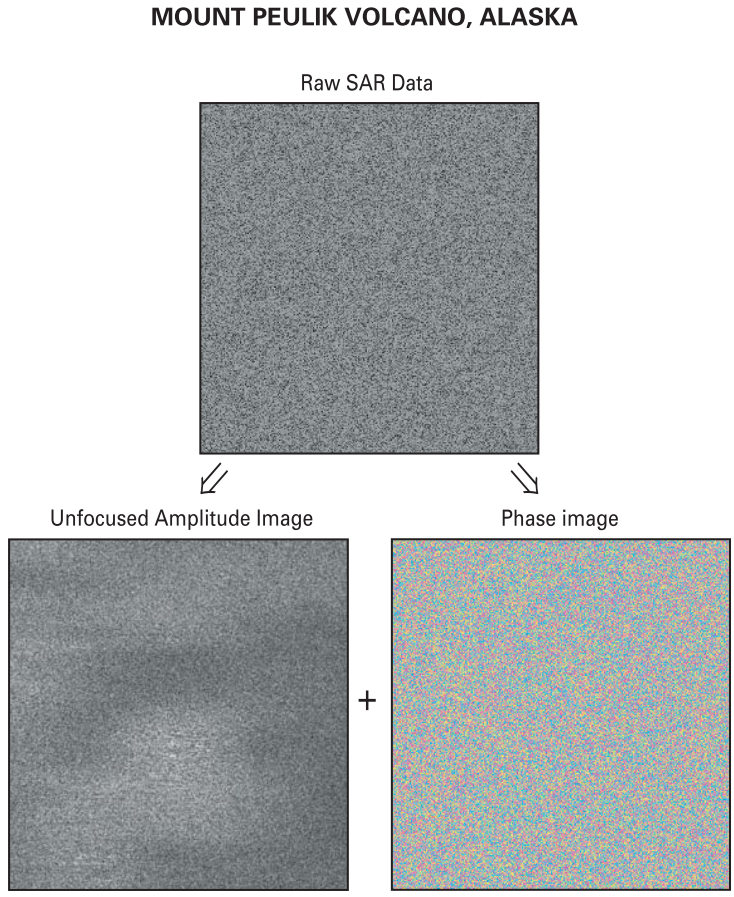

Figure 5.8. Raw SAR data consist of amplitude and phase information from time-sequential samples of the echo signal produced by reflection of transmitted pulses from the ground surface. This information can be represented by a complex number $Ae^{i\phi}$, with A representing the amplitude of the echo and ϕ representing its phase. See text for details. Until the data are focused by a SAR processor, the samples do not correspond to resolution cells on the ground, because each cell contributes to many samples and each sample includes information from many cells. Here, raw SAR data are displayed as an image by assigning the real and imaginary parts of the complex-valued signal from each sample ($A\cos\phi$ and $A\sin\phi$, respectively) to two adjacent pixels. The resulting image (*top*) appears random. When the amplitude and phase data are separated, the result is unfocused amplitude and phase images (*bottom left and right*, respectively). The amplitude image does not resemble the target area because each echo contains information from many resolution cells on the ground, which are spread out in range and azimuth. The random nature of the phase image reflects the fact that each resolution cell includes a unique montage of radar reflectors (e.g., pebbles, shrubs, fence posts), each of which contributes a different phase shift to the echo signal. These data for Mount Peulik Volcano, Alaska, were acquired by ERS-1 on 4 October 1995. North is approximately at the top.

between the reflectors is artificially compressed in the ground-range direction. This type of spatial distortion is known as foreshortening, and the feature appears to lean toward the radar in the image. If the surface slope is equal to the incidence angle, the incoming wavefronts are parallel to the surface so they arrive at both reflectors simultaneously. In this case, Δt is zero and the two reflectors appear on top of each other in the image. Steeper slopes cause layover, which occurs when the echo from the top of a slope arrives at the antenna before the echo from the bottom. The top gets mapped on the wrong side of the bottom. This is a particular problem for systems with a small incidence angle, such as ERS-1 and ERS-2, which produce layover wherever the surface slope exceeds about 23° (18° at near range or 26° at far range). Shadowing occurs when intervening topography blocks the line-of-sight from the radar antenna to the ground. This is more common for systems with large incidence angles, including RADARSAT-1 in some of its imaging modes. The effects of foreshortening can be removed if the surface topography is known, but the information from areas of layover or shadowing cannot be recovered.

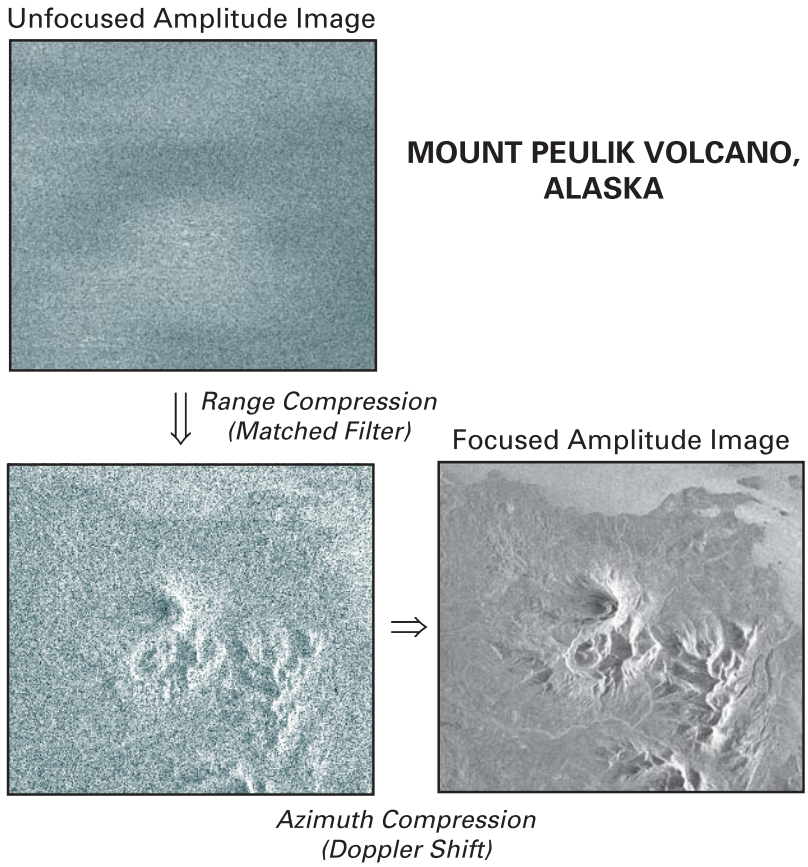

Figure 5.9. A SAR processor focuses the image in range using a matched filter, which sorts the echo signal into range bins in a process called range compression. The result is an image that resembles the target area but is not yet focused in azimuth (*middle*). Motion of the SAR with respect to the target area causes the echo signals to be Doppler shifted by an amount that depends on azimuth, which allows the SAR processor to focus the image in azimuth as well. The resulting focused image (*right*) resembles an aerial photograph. Foreshortening in the direction of the incoming radar signal, from the right in this case, is apparent on steep east-facing slopes. North is approximately at the top.

5.2 PRINCIPLES OF SAR INTERFEROMETRY

We have seen that it is possible to map the surface of the Earth with a SAR from an aircraft or satellite at meter-scale resolution, but to produce a digital elevation model (DEM) or to image centimeter-scale surface displacements requires some additional cleverness and mathematical manipulation. First, we will need a pair of overlapping radar images taken from slightly different vantage points, analogous to a stereo pair of aerial photographs used to produce a topographic map (Chapter 6). Overlapping radar images can be obtained in two ways. First, an aircraft or satellite can be equipped with two spatially separated antennas. One of the antennas serves as both transmitter and receiver, while the other serves as a second receiver. The images produced by the antennas are similar but subtly different owing to the different viewing geometries.

The National Aeronautics and Space Administration (NASA) used this single-pass, dual antenna approach for its Shuttle Radar Topography Mission (SRTM), which mapped approximately 80% of Earth's land surface between 60°N and 56°S during an 11-day flight in February 2000 (Farr and Kobrick, 2000; Williams, 2002). For SRTM, the SIR-C/X-SAR radar system was augmented by secondary C-band and X-band receive (slave) antennas mounted at the tip of a 60-m boom that could be extended from Space Shuttle Endeavor's cargo bay to form a single-pass interferometer (Bamler *et al*., 1996; Jordan *et al*., 1996; Farr and Kobrick, 2000). In this configuration, Endeavor/SRTM was the largest rigid structure ever flown in space – a record that

will not be eclipsed until the International Space Station grows to its planned 108-m wingspan.[13]

Another way to obtain overlapping radar images is to observe the same land area at least twice with the same radar by returning to nearly the same vantage point at different times. The European Space Agency (ESA) used this repeat-pass approach with a C-band radar system aboard its ERS-1 and ERS-2 satellites. From August 1995 to May 1996, these satellites flew a unique tandem mission in which ERS-2 followed approximately 30 minutes behind ERS-1 in the same orbital plane. Both orbits had a 35-day repeat cycle, which meant that once every 35 days, the same land area was observed first by ERS-1 and then, one day later, by ERS-2. Any surface changes that might have occurred during the one-day observation interval generally were small, so coherence was maintained in most areas and the tandem images are well suited to making DEMs (Section 5.2.5). The vertical accuracy of DEMs derived from ERS tandem pairs can be $\leq$10 m across an entire 100 km $\times$ 100 km image and, under favorable conditions, $\leq$2 m for local areas.

5.2.1 Co-registration of overlapping radar images

The next step in producing an interferogram is to co-register the two radar images so that we can difference (subtract) the phase information from corresponding pixels. Usually, this is possible if the image-acquisition points are separated by less than about 1 km for ERS-1 or ERS-2. The separation of image-acquisition points, or the distance between the SAR trajectories at the times of the two image acquisitions, is called the baseline. For longer baselines, differences caused by topography and viewing geometry can be too severe for coherence to persist. In such cases, the images are said to be spatially decorrelated. If the spatial coherence is good, we can use both the amplitude and phase information to co-register the images. The two sets of phase data are random but in a similar way, so they are helpful if they are spatially coherent. Otherwise, the amplitude data alone are generally sufficient. Think of the amplitude and phase data as two layers of a single image. These layers are perfectly co-registered for each image, but corresponding pixels in the two images are offset from one another owing to the different viewing geometries and topography. Conceptually, we would like to slide the two images over one another and distort them to achieve the best possible alignment of corresponding pixels.

In practice this procedure is accomplished by a computer, which uses cross correlation to determine the offsets between corresponding pixels. In one co-registration algorithm, a Fast Fourier Transform (FFT) is first run on a down-sampled version of each single-look complex (SLC) image.[14] The FFT of the first image is multiplied by the conjugate of the FFT of the second image to compute the correlation matrix in the frequency domain, and an inverse FFT is run on the product to return to the spatial domain (Figure 5.10). The location of the peak in the correlation matrix approximates the overall offset between the images. This result is refined by selecting corresponding, full-sampled portions from near the center of each image, using the approximate offset determined earlier, and repeating the procedure. Finally, the images are divided into sections called *chips*[15] and the correlation procedure is repeated on corresponding chips to determine the offsets between each of them as a function of position in the image. By fitting a polynomial to the results (second-order is sufficient in most cases), co-registration at the sub-pixel level can be achieved throughout the entire image. The second image is then re-sampled into the viewing geometry of the first so the pixels in both images correspond precisely.

[13] SRTM's primary objective was to produce a worldwide DEM with 1 arc-second ($\sim$30 m) spatial resolution, 16 m absolute height accuracy, and 10 m relative height accuracy (Farr and Kobrick, 2000). Plans to distribute global SRTM data were sharply curtailed after 11 September 2001 for reasons of US national security. SRTM data with 30 m spatial resolution are currently available for the US, and with 90 m resolution for all of North America and South America. Data for other continents will follow. Portions of the 30 m dataset outside of the US are available for scientific applications by special request. SRTM DEM data can be accessed at *http://seamless.usgs.gov/*.

[14] The FFT is an algorithm for computing the Fourier transform of a set of discrete data values, thereby converting data from the time domain to the frequency domain. To reduce the computing time required for the FFT when co-registering SAR images, the images are down-sampled in both range and azimuth directions. For example, the resolution cell size for ERS-1 and ERS-2 single-look images is 20–30 m in range and $\sim$5 m in azimuth. Using every second column and every tenth row in the SLC images greatly reduces the time required for the FFT.

[15] This usage of the term chip differs from that in the Global Positioning System (GPS) realm, where it refers to part of a binary sequence called a pseudorandom-noise code that is used to modulate GPS carrier signals (Section 4.3.1).

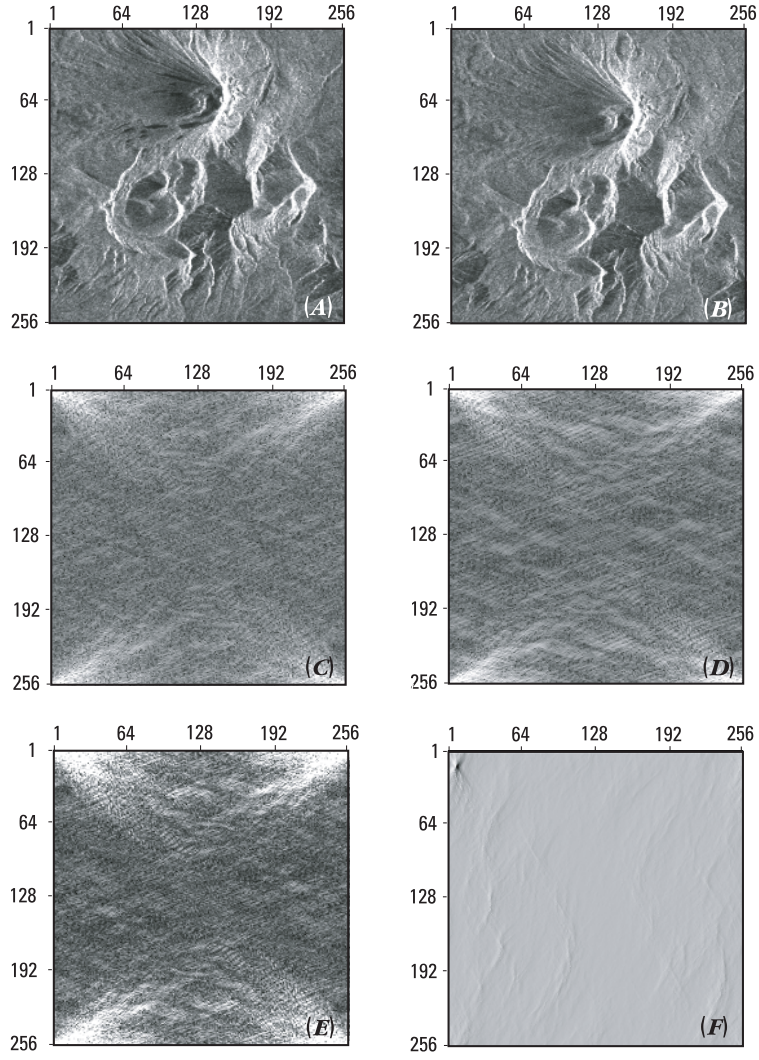

Figure 5.10. A Fast Fourier Transform (FFT) algorithm for co-registering SAR images. (A) Image 1 in the spatial domain; (B) Image 2 in the spatial domain; (C) FFT of Image 1 in the frequency domain (unit of 1/256); (D) FFT of Image 2 in the frequency domain (unit of 1/256); (E) product of the FFT of Image 1 with the conjugate of the FFT of Image 2 (in the frequency domain with a unit of 1/256); (F) inverse FFT of the product in (E). The coordinates of the peak near the upper left corner in (F) indicate the horizontal and vertical offsets in the spatial domain beween Image 1 and Image 2; in this case, 8 and 15 pixels, respectively. Readers interested in digital signal processing might want to consult an introductory textbook on the subject, such as Lyons (2004).

If the preceding paragraph was a painful reminder of how much math you have forgotten since college, do not worry. The algorithms have all been worked out and the details can be fairly transparent to users who prefer it that way. Alternatively, interested readers might want to consult an introductory textbook on digital signal processing, such as Lyons (2004), to learn how to commute between time and frequency domains by way of such conveyances as the FFT, inverse FFT, and complex conjugate of FFT. Here, let us forge ahead to the fun part – processing the pictures from our geodetic camera.

5.2.2 Creating the interferogram

When two images have been successfully co-registered, we can produce an interferogram simply by differencing the phase values of corresponding pixels (Figure 5.11). How can the difference between two random images be anything other than random? The random character of the phase images arises from the fact that the phase value associated with each resolution cell on the ground is determined in large part by the properties of numerous elementary reflectors within the cell, which combine to produce a unique phase signature.

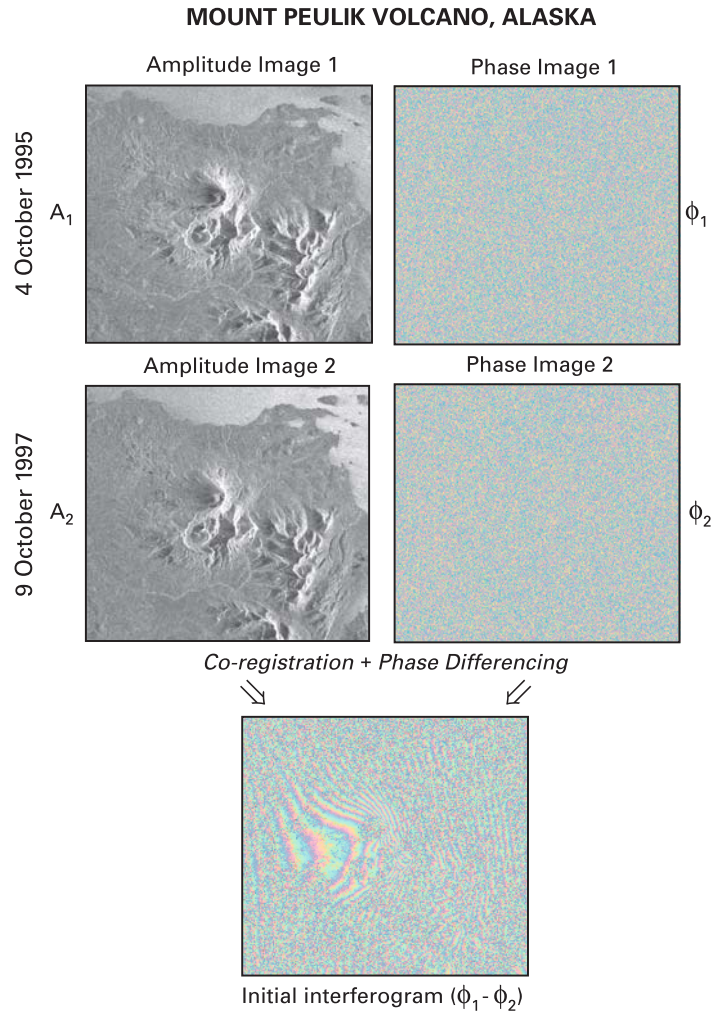

Figure 5.11. Information from two SAR images of the same target area (*top and middle*) acquired at different times and from slightly different vantage points (perpendicular baseline $b = 35$ m, altitude of ambiguity $h_a = 279$ m) can be combined to produce an interferogram by differencing the phase values after the images have been co-registered (*bottom*) . The resulting interferogram contains fringes produced by the differing viewing geometries (orbital fringes), topography (topographic fringes), any path delays present in the images, and surface displacements (deformation fringes). These images of Mount Peulik Volcano, Alaska, were acquired by ERS-1 on 4 October 1995 (*top*) and ERS-2 on 9 October 1997 (*middle*). North is approximately at the top. On the ground, the dimensions of the area shown are about 28 km from left to right by 25 km from top to bottom.

If these properties do not change appreciably between the acquisition times of the first and second images, their random contributions vanish in the difference image (i.e., the interferogram) and we are left with nonrandom, useful information.

An interferogram produced in this way includes contributions from the different viewing geometries of the two parent images, topography, path delays owing to different atmospheric conditions, noise, and any range changes caused by surface deformation during the interval spanned by the image acquisitions. It is this last bit of information that we would like to tease out of the data. First, though, we have to remove or account for the other sources of phase differences in the interferogram.

5.2.3 Removing the effects of viewing geometry and topography

The difference in viewing geometry between two images acquired from slightly different vantage points produces a regular pattern of phase differences between the images. If the target area were perfectly flat, these differences would manifest themselves in the interferogram as a series of nearly parallel bands called orbital fringes (Figure 5.12).

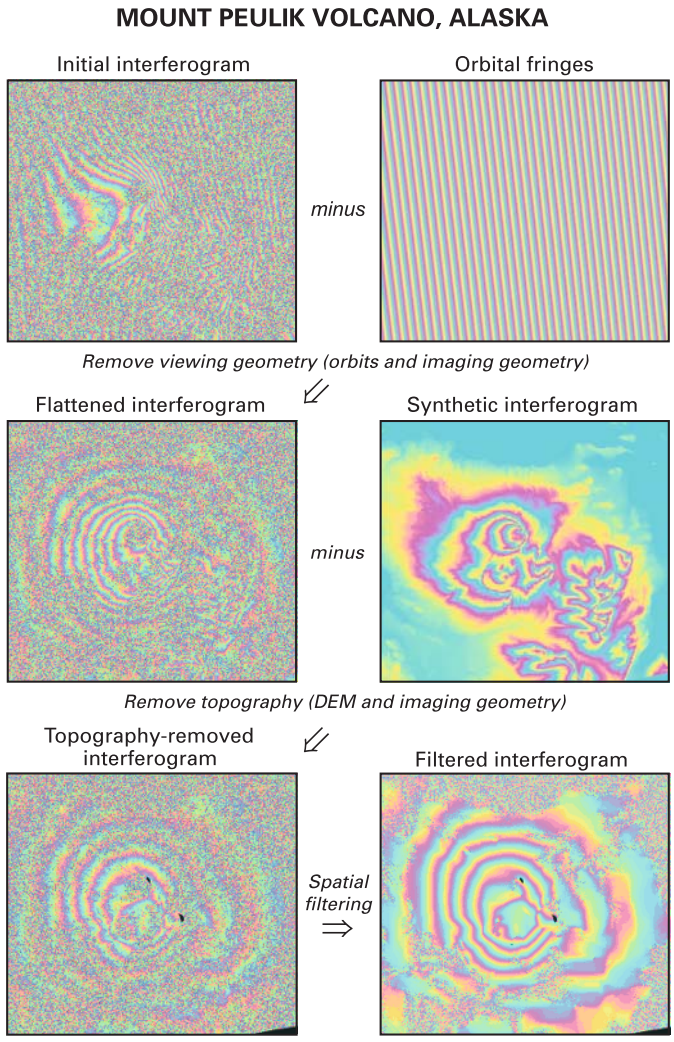

Figure 5.12. Orbital fringes can be computed from the known trajectory and imaging geometry of the SAR and removed from the initial interferogram by subtraction (*top*) to produce a flattened interferogram. Similarly, topographic fringes can be removed by subtracting a synthetic interferogram (*middle*) based on the known topography (Section 5.2.3) to produce a topography removed interferogram, which includes fringes produced by surface displacements (deformation fringes) and any path delays present in the images, plus noise. Spatial filtering suppresses the noise and accentuates the fringes (*lower right*).

These can be computed and removed based on the known trajectory and imaging geometry of the SAR, along with the small additional effect of Earth's curvature. The result is called a flattened interferogram.

Surface topography also contributes to the phase images, which is helpful if our goal is to produce a DEM but superfluous if our real interest is ground deformation. In the latter case, the effect of topography can be removed from a flattened interferogram by first constructing a synthetic interferogram based on known topography (i.e., on a DEM). A useful concept in this regard is the altitude of ambiguity h_a which is the amount of surface height difference that would produce exactly one topographic fringe:

$$h_a = \frac{H\lambda \tan\theta}{2b} \tag{5.16}$$

where H is the SAR altitude (about 800 km for most ERS-1 and ERS-2 orbits), λ is the wavelength of the SAR, θ is the incidence angle, and b is the perpendicular component of the baseline with respect to the incidence angle θ. For example, $h_a = 10\,\text{m}$ means that, for a given image pair, each 10 m of surface height difference would produce one topographic fringe in the interferogram, which would resemble a topographic map with a contour interval of 10 m.

Given h_a for a particular image pair, it is relatively straightforward to produce a synthetic interferogram that simulates topographic fringes.[16] Subtracting the synthetic interferogram from the observed flattened interferogram, we are left with a topography removed interferogram that contains only the effects of noise, path delays, and (most importantly) surface deformation (Figure 5.12).

Any errors in the DEM used to produce the synthetic interferogram will remain in the topography removed interferogram, so for deformation studies it is important to choose image pairs with a large altitude of ambiguity (i.e., short baseline). An interferogram formed from such pairs will have low sensitivity to DEM errors but full sensitivity to surface displacements. For example, a 30 m error in the DEM would result in 3 residual fringes in a topography removed interferogram if $h_a = 10$ m, but only one-third of a fringe if $h_a = 90$ m. Conversely, if the goal is to produce a DEM from the interferogram, a small h_a value (long baseline) is desirable for better sensitivity to topography. For ERS-1 and ERS-2, a useful rule of thumb is $h_a(\text{m}) \approx 9,400/b$ (m).

A topography removed interferogram can be spatially filtered to reduce the effect of short-wavelength noise, and displayed over an amplitude image to facilitate visual correlation of deformation fringes with surface features (Figure 5.13(A) and (B)). A final step to enhance the utility of the interferogram is to rectify (warp) the image to a geographical coordinate system, so it can be displayed over a shaded relief image derived from the DEM or with other georeferenced map features (Figure 5.13(C) and (D)). The flattened, topography removed, filtered interferogram in Figure 5.13, for example, shows 6 concentric fringes that represent about 17 cm of range decrease (mostly uplift) centered on the southwest flank of Mount Peulik Volcano, Alaska. The time spans of several interferograms include an uplift episode that occurred between October 1996 and September 1998 – a period that was aseismic at the volcano, but included an intense earthquake swarm near Becharof Lake, 30 km to the northwest, starting in May 1998 (Lu *et al.*, 2002a).

The ability to georeference SAR images and interferograms (i.e., to assign geographical coordinates in a known reference frame to each pixel in an image) opens the door to the wonderful world of geographic information systems (GISs). All geodetic and ground-deformation data are inherently geospatial, because they pertain to specific locations on the surface of the Earth; for example, the ellipsoidal height at point *A*, height difference from point *B* to point *C*, or change in baseline vector from point *D* to point *E*. All forms of geodetic and deformation data, including SAR images and interferograms, can and should coexist within a GIS with data pertaining to topography, geology, geophysics, geochemistry, hazard zones, infrastructure, land-management jurisdictions – all of the information pertinent to monitoring volcanoes, studying volcanic processes, assessing volcano hazards, and communicating effectively with colleagues, public officials, and the general public.

Housed in a robust and accessible geospatial database, SAR and InSAR data can be put to many uses beyond volcano monitoring – limited mainly by one's imagination – from map-making to hazards analyses and source inversions. As repeated SAR coverage of the globe increases and InSAR monitoring of the world's volcanoes becomes routine, the need to store and manipulate large geospatial datasets will only become more acute. The tool of choice for the foreseeable future will be GIS, and that suits volcano geodesy quite well.

5.2.4 Two-pass, three-pass, and four-pass interferometry

The method described above is called two-pass interferometry because it requires two overlapping radar images plus a DEM (Massonnet and Feigl, 2000). The two-pass method has the advantage of requiring only two spatially coherent radar images, but it also requires an accurate DEM from some other source (e.g., photogrammetry, digitized contours from a topographic map, SRTM). Typically, a DEM with a cell size of 30–90 m is adequate for this purpose. This is a drawback in many parts of the world where high-resolution DEMs are not yet available. Alternatively, two or more SAR images can be used to generate a DEM (Section 5.2.5),

[16] The altitude of ambiguity for a given viewing geometry h_a is analogus to the contour interval for a topographic map (i.e., it is the surface height difference encompassed by one interferometric fringe or by successive topographic contours, respectively). A synthetic interferogram can be derived from known topography (e.g., a DEM) by first contouring the height data using a contour interval equal to h_a, then assigning modulo-2π phase values to the cells between successive contours. If a suitable DEM is not available, one can be produced from a flattened interferogram of the same area, using an independent image pair (Section 5.2.5).

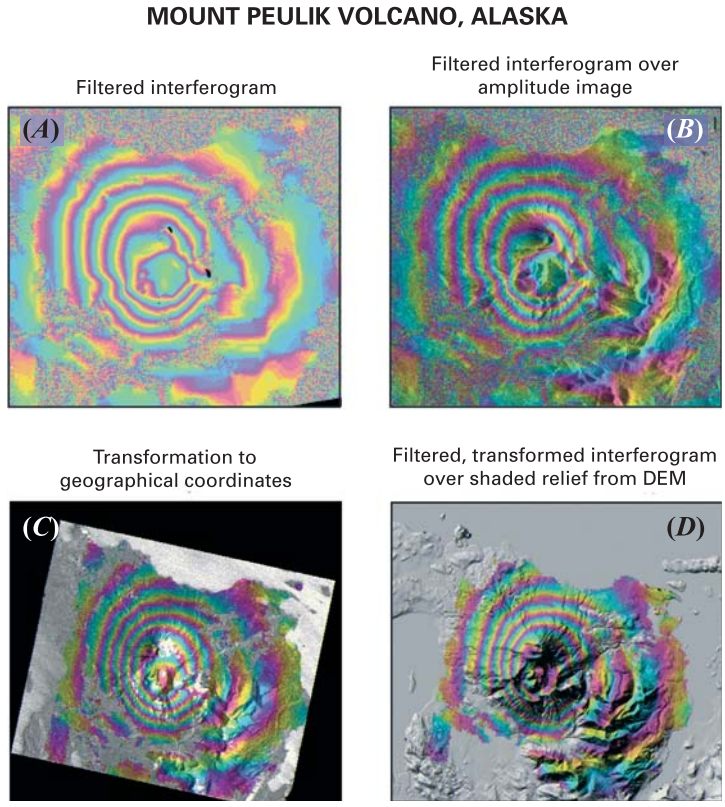

Figure 5.13. Displaying an interferogram over an amplitude image facilitates visual correlation of fringes with surface features (***B***). Images ***A*** and ***B*** are displayed in the SAR viewing geometry. Because the orbit and imaging characteristics of the SAR are known, the interferogram and amplitude image can be georeferenced (i.e., rectified to geographical coordinates) for consistency with standard map products (***C***) (Williams, 2002), and the rectified interferogram can be displayed over a shaded relief image produced from a digital elevation model (***D***). Each fringe corresponds to 2.8 cm of range change (mostly uplift in this case). The concentric pattern indicates ~17 cm of uplift centered on the southwest flank of Mount Peulik Volcano, Alaska, which must have occurred sometime during the interval spanned by the interferogram (October 1995–October 1997). Other interferograms for overlapping time periods bracket the inflation episode, which was aseismic at the volcano, between October 1996 and September 1998 (Lu *et al*., 2002a). The dimensions of the area shown are about 28 km from left to right by 25 km from top to bottom.

which then can be used to remove topographic effects from other interferograms of the same area. This technique requires only radar data, but the requirement for multiple, spatially coherent images of the same scene is a disadvantage for areas where coverage is relatively sparse (e.g., near the equator) and where the number of useful images is limited further by seasonal rainfall, snowfall, vegetation, or other factors that destroy coherence.

In three-pass interferometry, three SAR images are used to form two topographic interferograms; one is used to predict and remove the topographic fringe pattern from the other (Gabriel *et al*., 1989; Zebker *et al*., 1994). Usually, one image is selected as the reference image (call it ***A***) and two interferograms are formed between the reference image and the other two images (***B*** and ***C***). In other words, one interferogram is formed from ***A*** and ***B***, and another is formed from ***A*** and ***C***. Because the two interferograms are referenced to the same geometry, extra registration and re-sampling procedures are not necessary.

Four-pass interferometry is similar to three-pass, with the exception that different reference geometries are used to generate the two interferograms (i.e., the two interferograms do not share a common image). Consequently, one of the two interferograms needs to be re-sampled to the geometry of the other, and the topographic contribution in one of the interferograms needs to be removed using the other. The three-pass or four-pass method works well when no suitable DEM is available, but each requires extra image-processing steps relative to the two-pass method. Given the considerable number-crunching power of modern

PCs and workstations, this is more of a bother than a real impediment. The two-pass method is more straightforward, but requires a DEM from some other source. Nonetheless, the two-pass method is likely to become the technique of choice in most cases when the entire worldwide DEM from SRTM becomes widely available.

5.2.5 DEMs derived from InSAR

Because the observed SAR-to-target range depends mainly on the known vehicle track and the unknown topography along the ground track, data from a SAR can be used to create a DEM of the terrain it illuminates. The procedure is identical to the one described above for producing a topography removed interferogram, except in this case we would like to eliminate any contribution to the interferogram from surface deformation so we can extract just the topographic information. For this reason, it is advantageous to choose an image pair that spans the shortest possible time interval. For ERS-1 and ERS-2, tandem pairs with one-day separation are ideal for this purpose. If a tandem pair for the area of interest is not available, a small multiple of the 35-day repeat cycle for each satellite is usually adequate, except in areas of very rapid surface change or deformation. Relatively long baselines correspond to the smallest altitudes of ambiguity and therefore to the best sensitivity to topography. Baselines between 1,000 m and 300 m, which correspond to $h_a \sim 10$ m to $h_a \sim 30$ m, respectively, are suitable for DEM production. For baselines longer than ~1,000 m, the difference in viewing geometry between two images usually is too severe for the images to be spatially coherent.

To produce a DEM from a flattened interferogram, we need to keep track of the integer number of phase cycles between points using a procedure called phase unwrapping. The phase of the interferogram, which is directly related to the topography through the altitude of ambiguity, is only measured modulo 2π (i.e., values range from 0 to 2π and then repeat). To calculate the elevation of each point, it is necessary to establish an elevation at some point in the interferogram and to add the correct integer number of phase cycles to each phase measurement. The 'known' elevation should be taken from the best source available (e.g., a spot elevation on a topographic map or a surveying result for the summit of a prominent peak in the area). The problem of solving this 2π ambiguity is called phase unwrapping. Several approaches to this problem have been proposed (e.g., Goldstein *et al.*, 1988; Zebker and Lu, 1998; Costantini, 1998). Phase unwrapping is relatively easy in areas of high coherence. It is a matter of counting fringes from a 'known' starting point to other points in the image, then multiplying by the altitude of ambiguity (for a flattened interferogram to calculate elevation difference) or by the radar half-wavelength (for a topography removed interferogram to calculate apparent range-change). The procedure is akin to counting contour lines on a topographic map from a labeled contour to some other point, then multiplying by the contour interval to determine the elevation difference. However, where decorrelation or layover is a problem, it can be very difficult to unwrap an interferogram correctly. In such cases, it is sometimes necessary to fill holes in the unwrapped interferogram with lower resolution data from an existing DEM.

We have seen that a good DEM is an essential component of our geodetic camera, but there are other important applications, too. For example, the information contained in a DEM can be represented as a shaded-relief image, which simulates an oblique aerial photograph. The shaded-relief image in Figure 5.14 was produced from an InSAR-derived DEM; it shows in considerable detail a large debris avalanche deposit from Socompa Volcano in north Chile. The 40-km-long deposit was emplaced about 7,000 years ago when the northwestern flank of the volcano collapsed (Wadge *et al.*, 1995). Unlike a photograph, a shaded-relief image is 'illuminated' by simulated lighting that can be manipulated to accentuate features of particular interest (e.g., low-angle lighting from the north or south in an equatorial area to accentuate east–west trending ridges or valleys). Thus, shaded-relief images can reveal surface morphology that might otherwise be obscured in heavily vegetated areas. Combined with an empirical model of how lahars or lava flows interact with topography, a high-resolution DEM can be used to produce inundation-hazard maps quickly, efficiently, and with relatively little work in the field area (Iverson *et al.*, 1998). In areas of rapid geomorphic change, sequential DEMs can provide quantitative information about the sites and rates of such processes as erosion and deposition by rockfalls (Mills, 1992), lava-dome growth (Mills and Keating, 1992), eruption volume estimation (Lu *et al.*, 2003b), or glacier evolution (Schilling *et al.*, 2002, 2004).

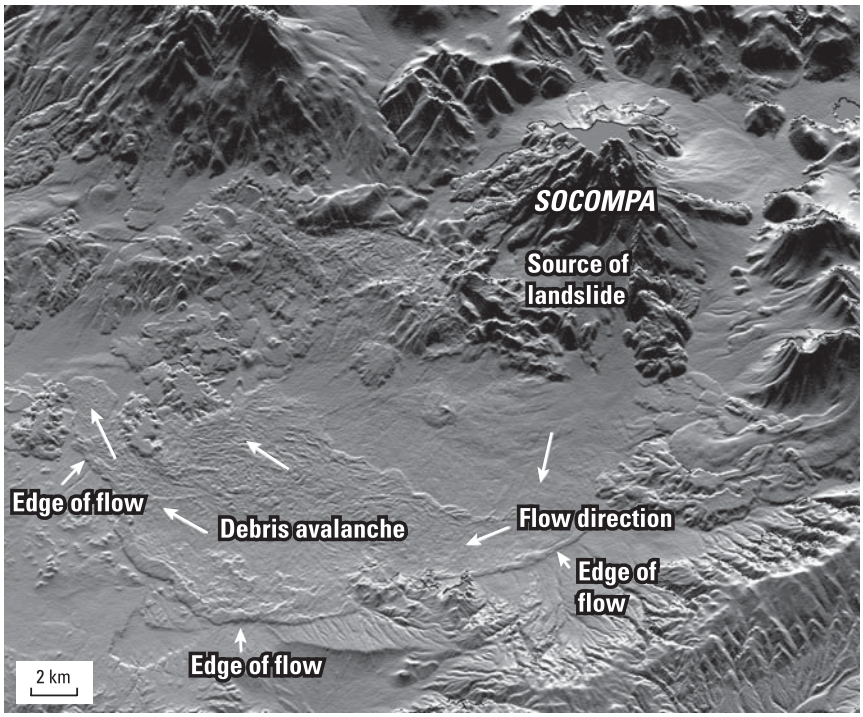

Figure 5.14. Shaded-relief image of Socompa Volcano and a prehistoric debris avalanche deposit in north Chile. The image was created from a DEM, which was produced by radar interferometry using a tandem pair of images from the ERS-1 and ERS-2 satellites. Image produced by Harold Garbeil and provided by Peter Mouginis-Mark, both at the University of Hawaii.

5.2.6 Lidar, InSAR, and photogrammetry – a potent remote-sensing triad

Additional sources of DEMs for InSAR and other applications include SRTM, digitization of conventional topographic maps, photogrammetry (Chapter 6), and lidar surveys. *Lidar*, an acronym for ***li***ght ***d***etection ***a***nd ***r***anging, is similar to radar except that a lidar system emits laser-light pulses instead of microwave chirps. The most widespread scientific application of lidar is in the field of atmospheric remote sensing (see, e.g., Weitkamp, 2005), but an aircraft equipped with lidar, GPS, and an inertial navigation system can also be used to map surface elevations with decimeter-scale accuracy, even in heavily vegetated terrain. How is this possible, given that: (1) most lidars operate at near-infrared wavelengths that are 10^4 to 10^5 times *shorter* than the microwave wavelengths used for radars (Figure 6.1), and (2) electromagnetic signal penetration generally *decreases* with wavelength?

The key to lidar's success in this regard is persistence. Lidar systems emit thousands of pulses per second, so each resolution cell on the surface is illuminated repeatedly. The size of the resolution cell is much larger than the slant-range resolution of the lidar (typically a few meters versus ~0.1 m), which means that the return signal potentially includes contributions from multiple reflectors in each cell that can be resolved in slant-range. The lidar system determines round-trip travel times for pulses arriving from each cell. The earliest returns correspond to the shortest slant range between the lidar and any reflector in a particular cell. In densely forested areas, for example, the first returns to arrive are from treetops; later ones are from branches, leaves, or other reflectors at successively lower levels in the canopy. Eventually, if all goes well, at least a few pulses make it to the ground surface and back to the lidar. These tardiest returns from each cell are used to produce a DEM, because they effectively 'strip away' any vegetation and reveal the topography of the underlying ground surface.

It should come as no surprise that the foregoing description is grossly oversimplified – if it were that easy, lidar surveys wouldn't be so expensive.[17] In the real world, several complications arise. Steep, uneven terrain is an obvious problem. If the

[17] The cost of lidar data varies greatly depending on factors such as the size and location of the project, horizontal postings (point density), vertical accuracy requirements, and type of data product requested. The standard product is (*x*, *y*, *z*) point data; derivative products including DEMs, digital terrain models (DTMs), and digital contours add to the cost. In 2005, the average cost for point data was $400 to $800 per square kilometer for 2- to 3-meter postings. Production of a DEM with 1-m postings in steep terrain could increase the cost by an order of magnitude or more.

surface height within a resolution cell varies by more than the slant-range resolution of the lidar, the return times of signals scattered from different parts of the cell will vary accordingly. The upshot is added uncertainty in the cell-height estimate. Patches of extremely dense vegetation present a similar problem. If none of the returns from a given resolution cell penetrates the canopy completely (i.e., the lidar cannot see the ground for the trees), the height estimate for that cell is biased upward.

Another fact that should not be surprising is that lidar engineers and digital-signal-processing experts have come up with clever ways to suppress the undesirable characteristics of raw lidar data, and to extract remarkably accurate approximations to the topography of the bare Earth. Sub-decimeter accuracy with 1-m postings is within the capability of modern lidar systems, and lidar coverage is expanding rapidly. Among lidar's recent successes in solid Earth science is delineation of fault scarps beneath heavy vegetation in the active Seattle fault zone, which strikes through downtown Seattle in the densely populated Puget Lowland of western Washington State (Harding and Berghoff, 2000; Haugerud *et al.*, 2003).

Lidar, InSAR, and photogrammetry are complementary remote-sensing approaches to DEM generation that, when integrated, can yield a robust and widely useful geospatial data product. InSAR provides geometric fidelity, lidar supplies highly accurate 'point-level' data, and photogrammetry provides realistic terrain visualization (Althausen *et al.*, 2004).

5.2.7 Range-change resolution of InSAR

To better appreciate the potential of InSAR, let us quickly review some essential points from the preceding sections. Recall that the phase value for each pixel in a radar image is determined by many factors, including a phase shift caused by the radar beam's interaction with surface materials. The radar wavelength is small compared with the size of a resolution cell on the ground, which typically contains hundreds of individual reflectors with differing complex reflection coefficients, each of which causes a different phase rotation or delay. The complex returns from all of these reflectors are summed to give the phase and amplitude contribution of the cell as a whole. The instantaneous SAR footprint is large compared with the cell size, so many cells contribute to each pulse echo. The echoes from many cells arrive simultaneously at the SAR, which samples the echo signal between successive chirps and passes the time-and-frequency information for each sample to a SAR processor for deciphering. The processor unscrambles the raw signal and assembles a focused SLC image with pixels that correspond to specific resolution cells on the ground. Because the phase value for each pixel depends on a unique montage of reflectors in the corresponding resolution cell, the phase image is random.

Here is where things seemingly get 'magical'. When two such random images of the same scene are co-registered and differenced, the randomness disappears (assuming phase coherence has been maintained) and distinctive patterns emerge with information about viewing geometry, surface topography, and ground deformation. When the effects of viewing geometry and topography are removed, the remaining pattern of phase differences in some cases reveals subtle ground deformation and thus the locations and shapes of the causative sources buried kilometers beneath the surface. It is not really magic, but the analogy to pulling a rabbit out of a hat seems apt. Let us analyze this trick more carefully before examining some snapshots of deforming volcanoes.

We have seen that our geodetic camera is capable of imaging topography with $\sim$10 m resolution, but how sharp is it when the subject matter is much more subtle (e.g., millimeter-scale surface displacements)? Consider what would happen if part of the target area moved toward or away from the SAR by exactly half of the radar wavelength, measured along the SAR-to-target (i.e., slant-range) direction, during the time between two image acquisitions.[18] The round-trip distance would have changed by one full wavelength, which means that signals reflected from the same resolution cell in the two images (after corrections for viewing geometry and topography) would be exactly in phase back at the SAR. The phase difference between corresponding pixels in the two images, and, therefore, the corresponding data value in the interferogram formed from the two images, would be zero (or, equivalently, 2π). The same would be true if the surface moved by $n\lambda/2$, where n is any integer. For cells that moved by an intermediate amount, $(n\lambda/2) \leq \Delta R \leq ((n+1)\lambda/2)$,

[18] The directionality of the SAR-to-target vector is commonly denoted by a unit vector (i.e., a vector of unit length pointing in the direction of the radar beam). The range vector is the product of this unit vector and the scalar range (i.e., the line-of-sight distance from SAR to target (Chapter 8)).

the data values in the interferogram would range from $0 < \Delta\phi < 2\pi$.

Now imagine that we use the color red to depict those pixels in the interferogram where $\Delta\phi = 0$ and violet for those areas where $\Delta\phi$ is slightly greater than zero. Let us also assign a continuous spectrum of colors from violet to red to pixels with intermediate phase differences (i.e., $0 \leq \Delta\phi \leq 2\pi$). Because the phase of a sinusoidal wave is modulo 2π (i.e., the waveform repeats every 2π or 360°), we would not know a priori whether a given violet pixel had moved $\Delta R = 0, \lambda/2, 3\lambda/2$, etc. If we assumed that most of the scene did not move at all and counted the number of violet-to-red color cycles (i.e., fringes) that formed a closed pattern, we could deduce the total amount of movement in much the same way that we interpret the contours on a topographic map.

Imagine that a volcano subsided in a concentric pattern centered at the summit by a maximum amount corresponding to several radar wavelengths (Figure 5.15). The resulting interferogram would show a concentric, repeating pattern of colors from red-to-green-to-violet in the direction of the summit. Each complete color cycle would represent differential movement of the ground surface by $\lambda/2$, and we could count the number of cycles (i.e., fringes) to determine the total amount of movement. If, instead, the surface had moved *toward* the SAR (i.e., uplift), the color pattern would be the reverse, from violet-to-green-to-red toward the summit (Figure 5.16).

For a C-band SAR operating at a wavelength of 5.6 cm, each fringe in the interferogram corresponds to a range change of $\lambda/2 = 2.83$ cm. By interpolation, we could in theory measure surface displacements corresponding to a small fraction of a single fringe, perhaps as little as 1 mm. In practice, various sources of image noise, mostly phase delays caused by inhomogeneity in the troposphere and ionosphere, limit the range-change accuracy of C-band interferometry to 1–10 mm under favorable conditions. Combined with spatial coverage from orbiting SARs that is typically on the order of 100 km × 100 km for each scene, this means that our dream of a geodetic camera is not so far-fetched as it might have seemed at first thought.

A crucial assumption in the foregoing discussion is that nothing in the target area changed during the time between successive radar images other than the surface height. This is not necessarily true over timescales of days to years, which are of interest for volcano monitoring. Over that period of time, vegetation might grow or die, rainfall might change the soil moisture, snow might accumulate or melt, or the surface might be affected by erosion or deposition. If such non-deformation changes are large with respect to the radar wavelength, our attempt to co-register the images and form a useful interferogram might fail. In this case, the images are said to lack coherence or be temporally decorrelated, and our geodetic camera would be hopelessly out of focus.

5.2.8 Coping with decorrelation and atmospheric-delay anomalies

The main limitations of repeat-pass SAR interferometry (i.e., temporal decorrelation and artifacts caused by atmospheric-delay anomalies), are subjects of intensive study. By understanding these phenomena better, researchers hope to mitigate their effects and thus make InSAR more accurate and widely applicable. Not all of the factors that contribute to temporal decorrelation are well understood, but it's clear that most are inescapable (i.e., soils erode and deposit, trees move in the wind, ice forms and melts, etc.). The best means available to extend the period over which successive radar images of the same area maintain coherence are to: (1) work in lava or other rocky terrains, or in urban areas, where temporal decorrelation generally is less problematic (not always possible if your target is a specific volcano), or (2) use a longer wavelength SAR. The penetration depth of a radar pulse increases with wavelength, so a longer wavelength pulse interacts with scatterers that are buried more deeply and is therefore less susceptible to changes occurring at the surface. In vegetated areas, a long-wavelength pulse is more likely to reach the ground, which increases the likelihood of obtaining coherent images on successive passes. A short-wavelength pulse is more likely to interact with branches, leaves, grasses, etc., that are much more fickle as scatterers. As a result, a target area that is virtually inaccessible to C-band interferometry might yield good results at L-band. Alternatively, an area might stay coherent for a matter of days at X-band, months at C-band, and years at L-band. Whereas most InSAR studies of volcanoes to date have relied on C-band images from ERS-1 and ERS-2, the fact that so many of the world's volcanoes are vegetated means that L-band is a better choice for future volcano–InSAR missions. This is true in spite of the fact that L-band's longer wavelength produces correspondingly less range-change resolution, a

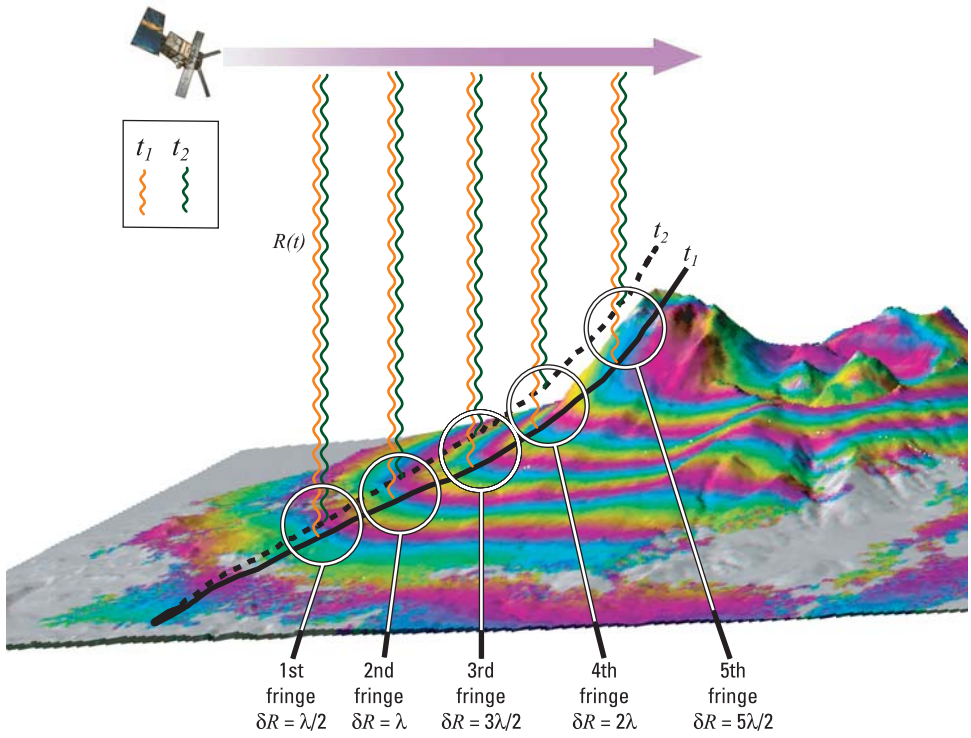

Figure 5.15. An inflating volcano (or any other landform) produces a pattern of concentric fringes in a radar interferogram from which the effects of viewing geometry and topography have been removed (Massonnet, 1997) . If the volcano's height profile changes from the solid line to the dashed line between the acquisition times, t_1 and t_2, of two radar images, the range $R(t)$ from the SAR to the surface will decrease by half the radar wavelength ($\delta R = \lambda/2$) in some areas, by $\delta R = \lambda$ in other areas, $\delta R = 3\lambda/2$ in others, etc. When the images are differenced, each half-wavelength change produces one complete interference fringe, which is portrayed as a spectrum of colors from red to violet (or violet to red, the choice is arbitrary) in the interferogram. The result is a concentric pattern of fringes around the volcano, akin to contours of range change with a contour interval of $\lambda/2$. Point-source subsidence would produce a similar pattern of fringes but with the opposite color sequence (i.e., from violet to red in the direction of increasing subsidence) (Figure 5.16). The satellite passes at times t_1 and t_2 need not be sequential; they could be separated by up to a few years if the surface retains interferometric coherence.

disadvantage that is outweighed by the prospect of better coherence in vegetated areas.

Even when temporal decorrelation is not a problem, atmospheric-delay anomalies can cause artifacts in radar interferograms that complicate their interpretation. Delay anomalies originate in the troposphere or ionosphere and are caused by inhomogeneities in water content, temperature, pressure, or electron density. They cannot be avoided entirely, but often their effects can be recognized and dealt with. For example, Massonnet and Feigl (1998) noted that interferograms made from night-time scenes show fewer atmospheric artifacts than those from daytime scenes, presumably because the atmosphere generally is more stable at night. If they are available, night-time images are therefore preferable.

How can path-delay anomalies be recognized and distinguished from surface deformation if both phenomena produce fringes in coherent interferograms? Fortunately, this is relatively straightforward if more than one interferogram is available for the same target area and approximately the same time period. By comparing multiple interferograms of the same scene, several types of potential artifacts can be recognized and discounted. For example, imagine that we form 3 coherent interferograms, AB, BC, and AC, from radar images A, B, and C. For simplicity, let us assume that the data-acquisition geometry is such that each of the interferograms has the same low sensitivity to any errors in the DEM (i.e., small b, large h_a). Assume also that image B was acquired during a period of heavy thunderstorm activity in the target area. The concentration of raindrops in a storm cell can interact with the radar beam to alter the phase of the echo signal and thus produce a path-delay anomaly, which appears as a closed pattern of fringes in interferograms AB and BC. The anomalies will have similar shapes and sizes but opposite signs,

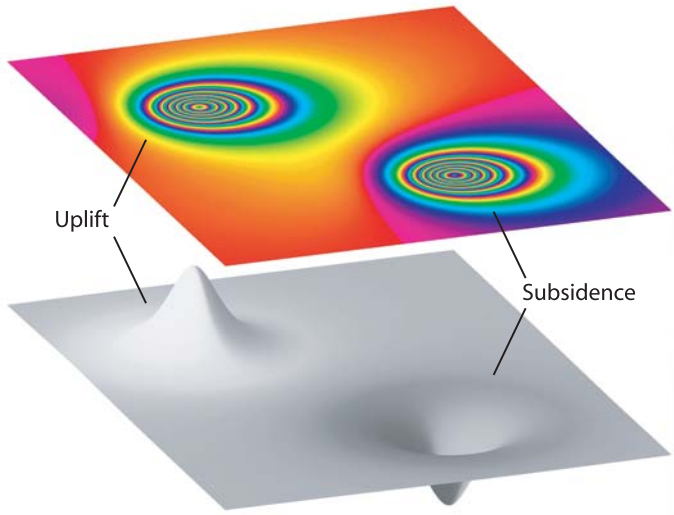

Figure 5.16. In a topography-removed interferogram, each fringe can be represented by a color band that spans the spectrum from red to violet, or *vice versa*. Each fringe corresponds to a change in the SAR-to-surface range by $\lambda/2$, where λ is the wavelength of the radar. For C-band radars like those aboard ERS-1, ERS-2, Radarsat, and Envisat, each fringe corresponds to 2.8 cm of range change. In this case, decreasing range (i.e., mostly uplift) is denoted by the color progression red--yellow--green--blue--violet; the opposite progression denotes increasing range (subsidence). Fringes can be thought of as contours of range change, akin to the contours on a topographic map. The fringe patterns corresponding to an axisymmetric source (e.g., a sphere or vertical pipe, see Chapter 8) are not circular because the radar is side-looking and therefore sensitive to both vertical and horizontal surface displacements. The fringes are more closely spaced where the vertical and horizontal components of motion are additive in the range direction (i.e., where the displacements are upward and toward the radar or downward and away). Accordingly, the patterns shown here are appropriate for a right looking SAR. Graphic provided by Charles W. Wicks, Jr., USGS.

which suggests they originated in the common image *B*. Interferogram *AC* will be free of the anomaly, because the storm effects only appear in image *B*. How can we be sure that the anomalies in *AB* and *BC* were caused by the atmosphere and not by ground deformation?

If we are fortunate enough to have another image of the study area, *D*, acquired at approximately the same time as *B* but not during the period of storm activity, we can form interferograms *AD*, *BD*, and *DC*. The storm-induced anomaly will appear only in *BD*, tying its origin more closely to image *B*. If the anomaly appears exclusively in interferograms formed from image *B*, we have good reason to discount it as a tropospheric propagation-delay effect. This strategy can fail if delay anomalies are persistent or recurring, as might be the case near volcanoes with large topographic relief. Tall volcanoes can influence local weather and set up quasi-steady patterns of water vapor concentration both horizontally and vertically (e.g., windward clouds and leeward clear skies, or summit cap clouds and sunny flanks). Such quasi-steady tropospheric patterns might be especially troublesome at large volcanoes such as Mount Etna, Italy (Beauducel *et al.*, 2000) and Mauna Loa, Hawai′i.

Emardson *et al.* (2003) pointed out that propagation delays arising in the ionosphere and troposphere might have different small-scale variability and decorrelation times, and that the effect of the ionosphere could be mitigated in the future by deploying a dual-frequency or split-band radar system. For the neutral atmosphere (troposphere), they suggested stacking (i.e., averaging) successive interferograms as a reasonable approach to increasing the signal-to-noise ratio. Stacking takes advantage of the fact that any geodetic signal is likely to persist through several interferograms, whereas tropospheric noise is more nearly random over timescales longer than a day (with some exceptions as noted above). Stacking interferograms tends to enhance the geodetic signal and average out the noise.

The standard deviation of the neutral-atmosphere-induced noise σ is approximated by (Emardson *et al.*, 2003):

$$\sigma = cL^{\alpha} + kH \tag{5.17}$$

where L and H are the differential length and height in kilometers and, for southern California, typical values of c, α, and k (determined empirically) are 2.5, 0.5, and 4.8, respectively. Using Mount Shasta as an

example, $L \sim 20$ km and $H \sim 3$ km, so $\sigma \sim 26$ mm or $\sim$1 fringe at C-band. Emardson *et al.* (2003) contend that the value of α is largely site-independent, but the value of c depends on the variability of the atmospheric water vapor content at the site of interest. Experience with Aleutian volcanoes suggests that σ might be as large as 2–3 fringes in some cases (Lu *et al.*, 2000c). The greater the atmospheric noise level, the greater the need for multiple interferograms. Emardson *et al.* (2003) estimated that resolving a deformation rate of $1\,\text{mm}\,\text{yr}^{-1}$ over a distance of 10 km with radar images acquired every 7 days would take 2.2 years, and that resolving the same rate over 100 km would take 4.8 years. Thus, a small constellation of radar satellites dedicated to interferometry could track deformation from magmatic and tectonic sources worldwide with a precision rivaling that of other geodetic techniques.

Now consider the case of anomalies that appear in interferograms *AB*, *BC*, and *AC* at the same location but with differing amplitudes. No single image could be responsible for all three anomalies, but they all could be topographic artifacts rather than deformation signals. This is especially true if the altitudes of ambiguity are small with respect to the accuracy of the DEM. For example, if *AB* has an altitude of ambiguity twice as large as *BC* and the anomaly in *BC* has twice as many fringes as the one in *AB*, there is good reason to be suspicious. In this case, we could try to find a fourth image to form other interferograms with greater altitudes of ambiguity.

Failing this, we could employ another trick of the InSAR trade. By adding (or subtracting) co-registered interferograms with altitudes of ambiguity of opposite (or same) sign, we can synthesize an interferogram from two ghost images that were never acquired! This is a useful technique to increase the effective altitude of ambiguity and thus to produce an interferogram with less sensitivity to topography. The same technique can be used to decrease the effective h_a and thus to increase the topographic sensitivity. This can be helpful if the desired product is a DEM rather than a topography removed interferogram (Massonnet and Feigl, 1998).

In summary, three factors that can limit InSAR's usefulness for geodetic studies are temporal decorrelation, path-delay anomalies, and topographic artifacts caused by DEM errors. Decorrelation is less of a problem in lava or other rocky terrains, and in urban areas, owing to relatively sparse vegetation and the presence of permanent scatterers such as blocky lava, boulders, pavement, or buildings. Elsewhere, including at sparsely vegetated volcanoes where deformation rates are high, useful interferograms sometimes can be produced from images separated by a relatively short time interval (e.g., the 35-day repeat cycle for ERS-1 and ERS-2). Another approach is to select images acquired during the same season of successive years, when decorrelation caused by snow or rainfall is likely to be at a minimum (Lu and Freymueller, 1998). At several Aleutian volcanoes, C-band coherence has been observed to persist for 3–5 summers in some areas (Lu *et al.*, 2002b,c). Future L-band SAR missions could greatly extend InSAR's applicability by reducing the problem posed by temporal decorrelation at shorter wavelengths. Artifacts caused by path-delay anomalies or DEM errors can be identified in a relatively straightforward way if several interferograms are available for the same area. InSAR is fast becoming an operational tool suitable for use at many of the world's volcanoes, and its utility will surely grow as next-generation InSAR missions are realized.

5.2.9 Volcano InSAR studies: a growing list of success stories

Fortunately, experience has shown that many real-world surfaces stay correlated for periods of months to a few years, long enough for most volcano-monitoring purposes, even in areas of sparse-to-moderate vegetation. For example, coherent interferograms over periods of 1 to 4 years have been produced for Mount Etna, Italy (Massonnet *et al.*, 1995; Lanari and others*et al.*, 1998); Long Valley Caldera, California (Thatcher and Massonnet, 1997); Yellowstone Caldera, Wyoming (Wicks *et al.*, 1998); Okmok (Lu *et al.*, 1998), Akutan (Lu *et al.*, 2000a, 2005), Westdahl (Lu *et al.*, 2000b, 2004), Makushin (Lu *et al.*, 2002c), and Kiska (Lu *et al.*, 2002d) Volcanoes, Alaska; Piton de la Fournasie, La Réunion, Indian Ocean (Sigmundsson *et al.*, 1999); several volcanoes in the Galápagos Islands (Jonsson *et al.*, 1999; Amelung *et al.*, 2000) and in Chile (Pritchard and Simons, 2002); and for a growing list of other volcanoes worldwide (e.g., Zebker *et al.*, 2000). These volcanoes span a wide range of climatic conditions and vegetative cover, suggesting that many more of the world's volcanoes are amenable to study using InSAR, especially at L-band wavelengths.

However, there have been unsuccessful attempts as well, especially at volcanoes with dense vegeta-

tion, rugged topography, or permanent snow-and-ice cover (Zebker *et al.*, 2000). The volcanoes of the Cascade Range in the western USA present difficult challenges in this regard, as do many volcanoes in tropical or subtropical regions. Even at the volcanoes where InSAR has been successful, it generally has been necessary to use only images acquired in the same season to avoid temporal decorrelation caused by changes in the land surface or snow cover. Together with the 35-day orbital repeat cycle of the ERS-1 and ERS-2 satellites, this essentially limits the useful repeat cycle for InSAR monitoring at most volcanoes to periods of either 1–3 months during the summer or to one or more years (e.g., successive summers). Nonetheless, there have been some striking successes that bode well for the future of InSAR studies at volcanoes. A few of these are discussed below. Several others are described in excellent reviews of the subject by Massonnet and Feigl (1998) and Zebker *et al.* (2000).

5.3 EXAMPLES OF INTERFEROMETRIC SAR APPLIED TO VOLCANOES

5.3.1 Mount Etna

Spurred by the remarkable success of InSAR in capturing coseismic displacements from the 1992 Landers earthquake (Massonnet *et al.*, 1993), InSAR specialists soon turned their attention to another class of promising targets – volcanoes. The first positive results were for Mount Etna during the course of an effusive eruption from 14 December 1991 to 31 March 1993. Massonnet *et al.* (1995) produced a series of interferograms from ERS-1 images that revealed progressive subsidence of the edifice from 17 May 1992 to 24 October 1993. A series of best-fit models based on sequential interferograms showed that the volume of subsidence increased linearly with time until the end of the eruption, when the subsidence also stopped. The overall best fit was obtained from a deflation source located 2 ± 0.5 km east of the summit at 16 km depth.

The InSAR results for Mount Etna are supported by an abundance of observational data. During the latter part of the 1991–1993 eruption, ERS-1 acquired 13 images of Mount Etna during ascending orbits (i.e., traveling north) and 16 images during descending orbits. Only images from the same family of orbits can be combined to form interferograms. The ascending images give 78 potential interferometric combinations, and the descending images give 120 independent combinations. Of these, Massonnet *et al.* (1995) selected 32 ascending and 60 descending combinations with favorable data-acquisition geometries to minimize any residual topographic errors resulting from the DEM. About 30 of the 92 favorable combinations were coherent enough to be analyzed and only 12 were used for modeling. An important lesson here is that no single interferogram is likely to be definitive. It should be standard procedure to analyze as many potential image pairs as possible to overcome problems caused by temporal decorrelation, path-delay anomalies, or DEM errors.

Following the initial InSAR study of Mount Etna, Lanari *et al.* (1998) produced additional interferograms of the volcano for several intervals between September 1992 and October 1996. The 1992–1993 interferograms revealed a subsidence source similar to that inferred from the earlier study but at shallower depth (9 km versus 16 km). Following the end of an effusive eruption in March 1993, a period of uplift ensued until the start of another eruption in July 1995. The uplift source was deeper (>11 km), which is consistent with the idea that deflation at shallow levels is followed by inflation at deeper levels as the volcano recharges with magma from below. Sequential interferograms showed steady inflation following the end of the 1991–1993 eruption, which accelerated in the months preceding the next outbreak in July 1995. This was the first InSAR study to detect a reversal from deflation to inflation and the first to track pre-eruptive inflation.

The foregoing interpretations were questioned by Beauducel *et al.* (2000), who argued that tropospheric effects could account for most of the reported ground deformation during 1992–1998. They estimated tropospheric delays in 238 interferograms of -2.7 to $+3.0$ (± 1.2) fringes, and suggested that any volume changes in the Mount Etna magma reservoir were within the observational uncertainty. The volcano's imposing size (3,300 m) undoubtedly is a factor, because the edifice extends through a large range in atmospheric conditions and influences local weather. It is not uncommon for delay anomalies to produce 2–3 fringes in interferograms of Aleutian volcanoes (Lu *et al.*, 2000c), which are subject to notoriously poor weather. Better understanding of path delays would improve InSAR's reliability for volcano monitoring, especially at tall volcanoes capable of influencing local weather patterns.

5.3.2 Long Valley Caldera, California

Several striking successes followed on the heels of the initial InSAR studies of Mount Etna. At the Long Valley Caldera in eastern California, InSAR was used to map uplift of the resurgent dome and surrounding caldera floor in more detail than had been possible with repeated geodetic surveys (Chapter 7). A multifaceted episode of tectonic and magmatic unrest at Long Valley began in October 1978 with the occurrence of a M 5.7 earthquake about 15 km southeast of the caldera rim. This was followed on 25–27 May 1980, by a swarm-like sequence of four M 6 and thousands of smaller earthquakes near the south caldera rim, and by the onset of inflation of the resurgent dome (Hill *et al.*, 1985a,b). Dozens of small earthquake swarms occurred during the ensuing two decades, including especially strong swarms beneath the south moat in January 1983 (Savage and Cockerham, 1984) and December 1997–January 1998 that may have been accompanied by intrusions. In 1989, a swarm of small earthquakes beneath Mammoth Mountain that included dozens of long-period events probably marked another intrusion (Hill *et al.*, 1990). The swarm was preceded by more than 2 months of increased extensional strain within the caldera (Langbein *et al.*, 1993) and was followed several months later by the start of tree-kill caused by increased emission of carbon dioxide (Farrar *et al.*, 1995).

Repeated leveling and two-color EDM surveys at Long Valley revealed persistent uplift and extension across the caldera since 1980 at rates that varied with time (Chapter 7). Maximum uplift of the resurgent dome from 1975 to 1997 was more than 70 cm, and from mid-1983 to mid-1998 a two-color EDM line across the resurgent dome extended by nearly 40 cm at rates that sometimes exceeded 10 cm yr^{-1}. Most models of the 1980–1998 deformation have in common a spherical or ellipsoidal source 5–8 km beneath the resurgent dome and a fault or dike beneath the south moat. Some studies also include a second, deeper ellipsoidal source beneath the south moat or a dike beneath Mammoth Mountain (Langbein *et al.*, 1995, 1993, 1987a,b; Langbein, 1989; Savage, 1988; Savage and Cockerham, 1984; Rundle and Whitcomb, 1986).

Thatcher and Massonnet (1997) were first to study the Long Valley region with InSAR. They constructed several interferograms from ERS-1 and ERS-2 images, and obtained coherent results with image pairs spanning 2-year and 4-year intervals. The interferograms revealed a pattern of surface deformation centered in the caldera and generally consistent with that deduced from repeated leveling and two-color EDM surveys. Synthetic interferograms based on the ellipsoidal-inclusion models of Langbein *et al.* (1995), which were developed from the leveling and EDM results, showed good agreement with the observed interferograms. This demonstrated that InSAR could produce essentially the same result as field-intensive geodetic surveys, with much better spatial resolution (albeit with less time resolution in this case).

Simons *et al.* (1999) combined sequential InSAR, continuous GPS, and two-color EDM measurements at Long Valley to constrain time-dependent parameters for various deformation sources needed to explain the observed deformation pattern. The InSAR dataset included ERS-1 and ERS-2 images acquired at sub-yearly intervals from 1992 to 1999. Topographic effects were removed from the interferograms using a DEM with 5 m cell size derived from the NASA/JPL airborne SAR instrument (TopSAR). Inversion[19] of the InSAR data suggested a spherical or ellipsoidal source 9 ± 1 km beneath the resurgent dome, deeper than the best-fit model of Langbein *et al.* (1995) based on inversion of two-color EDM and leveling data (5.5 to 7 km depth). The discrepancy might reflect the much greater spatial aperture of the interferograms, which capture far-field deformation from a deeper source more effectively than the less extensive EDM and leveling networks.

Fialko *et al.* (2001b) came to a similar conclusion for the 1997–1998 inflation episode at Long Valley. Their modeling of ERS-1 and ERS-2 interferograms, which collectively span the time interval from 6 June 1996 to 12 July 1998, showed that the observed uplift pattern ($\sim$11 cm maximum) could be explained by an inflating sill at $\sim$12 km depth or an elongate prolate spheroid (i.e., a 'pluton-like body') at $\sim$8 km depth beneath the resurgent dome. Joint inversion of the InSAR and two-color EDM results suggested that the inferred magma body is a steeply dipping prolate spheroid with a depth of 7–9 km beneath the dome and an aspect ratio greater than 2 : 1. Although the exact nature of

[19] Inversion is a mathematical procedure for determining best-fit model parameters from a set of observations, starting with a generalized forward model and known or assumed boundary conditions. For related information, see *inversion (numerical modeling)* in the Glossary (at the back of this book), and Chapter 8.

the deformation source(s) might be obscured by such complications as crustal heterogeneity and deviations from linear elasticity, the InSAR results at Long Valley have helped to constrain the range of viable models and thus have contributed to a better understanding of unrest at one of the most intensively monitored calderas on Earth.

5.3.3 Yellowstone Caldera, Wyoming

Encouraged by the success of InSAR at Long Valley, two colleagues and I applied the same approach to the Yellowstone Caldera, where leveling surveys had shown both uplift and subsidence for several decades. There have been at least seven M 6 earthquakes and one M 7 earthquake in the Yellowstone region during historical time, and swarms of smaller earthquakes are commonplace. The vigor of Yellowstone's hydrothermal system, epitomized by the regular spouting of Old Faithful geyser, is well known throughout the world. Less obvious to the casual visitor, but equally fascinating, are the ups and downs of the caldera floor as revealed by repeated leveling and GPS surveys (Chapter 7). Comparison of leveling surveys in 1923 and 1975–1977 first detected uplift within the caldera that continued through 1984 at an average rate of 1–2 cm yr^{-1}. When the uplift paused during 1984–1985, parts of the caldera floor had risen more than 90 cm since 1923 (Dzurisin and Yamashita, 1987). Annual leveling surveys thereafter showed subsidence at rates of 1–3 cm yr^{-1} through 1995, raising the question of when uplift would resume (Dzurisin *et al.*, 1990, 1994). InSAR provided the answer, and some surprises too.

As expected from the leveling results, interferograms formed from ERS-1 and ERS-2 images show a pattern of subsidence within the caldera during 1992–1993 and 1993–1995 (Wicks *et al.*, 1998) (Figures 5.17 and 5.18). From August 1992 to June 1993, more than 3 cm of subsidence accumulated in the northeast part of the caldera near the Sour Creek resurgent dome (Figure 5.18(A)). Subsidence continued from June 1993 to August 1995 (more than 4 cm), but surprisingly the center of deformation shifted to the Mallard Lake dome in the southwest part of the caldera (Figure 5.18(B)). Interferograms for the entire 1992–1995 interval show about 6 cm of subsidence centered between the two domes, consistent with leveling surveys in the northeast part of the caldera in 1992, 1993, and 1995 (Dzurisin *et al.*, 1999). This was the first time that a shift in the deformation center had been recognized at Yellowstone. But the surprises continued.

An interferogram for the period from August 1995 to September 1996 shows that the Sour Creek dome rose about 2 cm while the Mallard Lake dome was still subsiding slightly (Figure 5.18(C)). The renewal of uplift following a decade of subsidence was confirmed by an interferogram for the period from July 1995 to June 1997, which shows over 3 cm of uplift extending throughout the central part of the caldera (Figure 5.18(D)), and by leveling in 1998. Whereas leveling surveys spanning more than seven decades had failed to detect any movement of the deformation center at Yellowstone or a change from subsidence to uplift, InSAR accomplished both in very short order. Even before the implications of these discoveries had fully settled in, we were treated to yet another surprise.

Figure 5.19 shows Yellowstone interferograms for the periods from September 1996 to September 2000 and from August 2000 to September 2001. An elliptical uplift centered along the north caldera rim about 10 km south of Norris Geyser Basin is beautifully portrayed in the 1996–2000 image, which spans part of the July 1995 to June 1997 period of renewed uplift in the caldera (Figure 5.18(D)). The Norris uplift extends about 30 km east–northeast from Madison Junction to Canyon Junction, and 40 km north–northwest from the center of the caldera. It is tucked neatly between the two resurgent domes, which had been identified as separate deformation centers just a few years earlier. The 2000–2001 interferogram shows that the Norris uplift continued at an average rate of ~3 cm yr^{-1} (i.e., ~1 fringe yr^{-1}) while the central part of the caldera floor, including both resurgent domes, subsided at a comparable rate (Wicks *et al.*, 2002c). More recent InSAR results indicate that the 2000–2001 pattern persisted through summer 2002 but had stopped by summer 2003.[20] In less than a decade, three distinct deformation sources have been active at Yellowstone, and two of the three have produced both

[20] The year-to-year range changes are small enough to have been caused by tropospheric-path delays, as discussed by Beauducel *et al.* (2000) for Mount Etna, but three lines of evidence point instead to ground deformation as the cause: (1) the fringes around Norris Geyser Basin appear in several independent interferograms spanning similar time periods, (2) the Norris area is relatively flat so there is no reason to suspect a strong topographic influence on tropospheric water vapor concentration, and (3) the uplift inferred from InSAR was corroborated by comparison of leveling results from 1987 and 2004 (CVO, unpublished data).

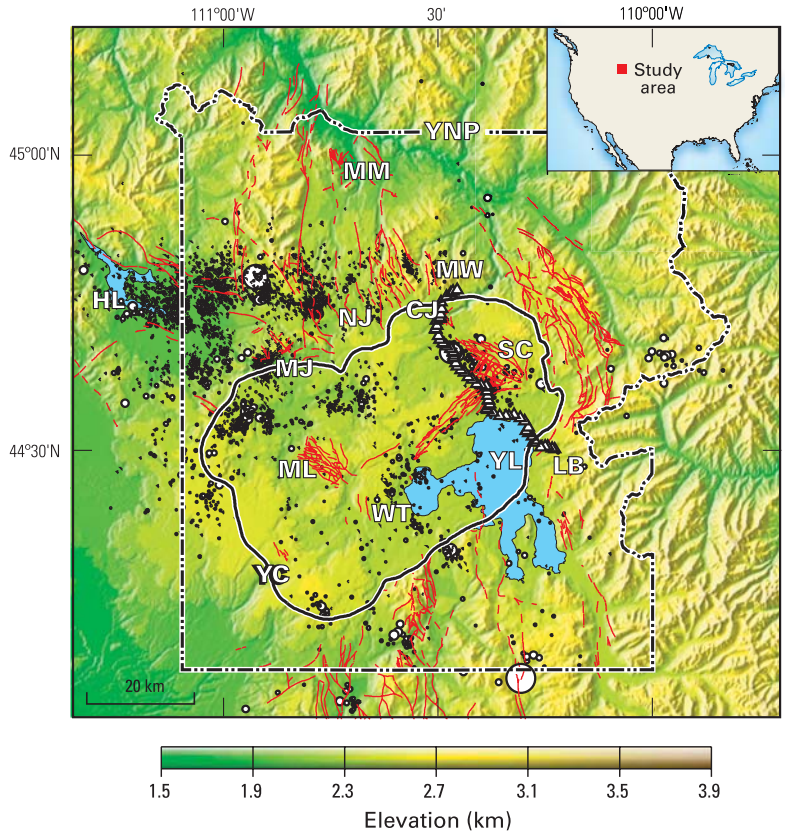

Figure 5.17. Topography of the Yellowstone region derived from USGS DEMs, after Wicks *et al*. (1998). YNP (dash-dot line), Yellowstone National Park boundary; YC (solid black line), Yellowstone Caldera boundary; MM, Mammoth; HL, Hebgen Lake; MW, Mount Washburn; CJ, Canyon Junction; MJ, Madison Junction; NJ, Norris Junction; SC, Sour Creek resurgent dome; ML, Mallard Lake resurgent dome; WT, West Thumb; LB, Lake Butte; YL, Yellowstone Lake. String of white-filled black triangles running northwest from LB to MW represents bench marks along a level line that was measured annually from 1983 to 1998, except 1994, 1996, and 1997 (Dzurisin *et al*., 1994). Red lines show faults mapped (solid) and inferred (dashed) by Christiansen (1984). Earthquake epicenters from the University of Utah Seismographic Stations' Yellowstone National Park Earthquake Catalogs for 1983--1995 are shown as white-filled black circles and black dots, with symbol size proportional to earthquake magnitude. The three largest earthquakes shown are *M* 5.1 (28 August 1995) on the southern boundary of YNP, *M* 4.9 (26 March 1994) ~20 km east of HL, and *M* 4.8 (24 September 1994) on the west edge of SC along the level line; the smallest earthquakes shown are *M* ~0.0.

uplift and subsidence. Thanks to InSAR, we now know that the dynamics of unrest are far more complicated than we anticipated!

No comprehensive model has been proposed to explain all of the InSAR results at Yellowstone, although the essential ingredients have been identified from earlier work. There is abundant geophysical evidence for the existence of a rhyolitic magma body in the crust beneath Yellowstone that is sustained by basaltic intrusions from the mantle (Smith and Braile, 1994; Smith and Siegel, 2000; Christiansen, 2001). The existence of a vigorous hydrothermal system is evident from the countless thermal features that grace Yellowstone National Park. It, too, requires basaltic heat input to have persisted since Yellowstone's most recent eruption ~70,000 years ago (Fournier and Pitt, 1985; Dzurisin *et al*., 1990). The hydrothermal system consists of a deep zone in which pore-fluid pressure is near lithostatic and a shallow zone in which pore pressure is hydrostatic (Fournier and Pitt, 1985; Fournier, 1991, 1999).[21] The two zones are separated by an impermeable, self-sealed layer created by mineral deposition and quasi-plastic flow at a depth of about 5 km. Pore-fluid pressure in the shallow zone is always less than lithostatic because a

[21] Lithostatic pressure refers to the vertical pressure at a point in the Earth's crust caused by the weight of the overlying column of rock or soil. Pore fluids in the lithostatically pressured zone are isolated from the surface by impermeable rock or a self-sealed zone. Hydrostatic pressure is due to the weight of groundwater at higher levels in the system. In highly fractured or porous rocks where groundwater flows freely, the total pressure approximates the hydrostatic pressure in this upper zone.

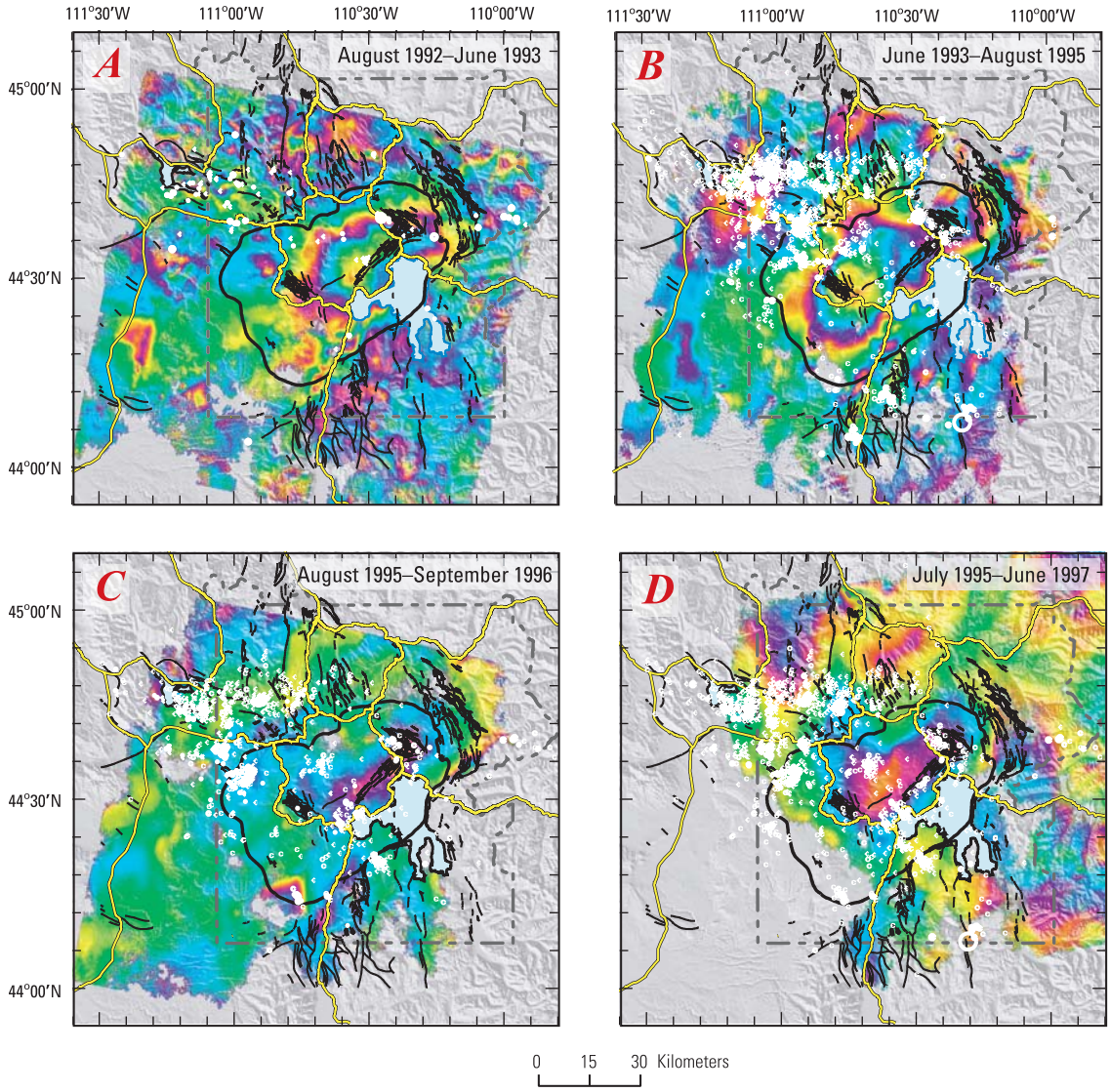

Figure 5.18. Successive topography-removed interferograms of the Yellowstone region from 1992 to 1997, shown over shaded relief (Wicks *et al*., 1998; Dzurisin *et al*., 1999; Wicks *et al*., 2002c). Range of colors from violet to red corresponds to 2.8 cm (one C-band fringe) of range change (mostly uplift or subsidence). Dashed line, Yellowstone National Park boundary; heavy black line, Yellowstone Caldera; fine black lines, faults; yellow lines, roads; white circles, earthquake epicenters (sized by magnitude) for each time interval. (A) From August 1992 to June 1993, subsidence was centered in the northeast part of the Yellowstone Caldera near the Sour Creek resurgent dome (Figure 5.17). (B) The subsidence center shifted to the southwest part of the caldera near the Mallard Lake resurgent dome from June 1993 to August 1995. (C) From August 1995 to September 1996, a small amount of subsidence persisted near the Mallard Lake dome, but the Sour Creek dome began to rise slightly. (D) Uplift spread throughout the central part of the caldera from July 1995 to June 1997. Note that the color progression inside the caldera is reversed between D (uplift) and either A or B (subsidence).

complex system of microfractures and other fluid passageways connects it to the surface. For this reason, pressure changes in the hydrostatically pressured zone cannot be the cause of widespread uplift or subsidence. Beneath the self-sealed layer the deep hydrothermal system is near lithostatic pressure, so any change in pore-fluid pressure has the potential to raise or lower the ground surface.

Based on the interferograms shown in Figure 5.18, Wicks *et al*. (1998) proposed that uplift and subsidence inside the caldera are caused by pressure changes in two interacting fluid reservoirs, one under each of the resurgent domes at 8 ± 4 km depth. Inflation and deflation are regulated by flow through conduits that connect the reservoirs to each other and to a deeper magmatic pressure

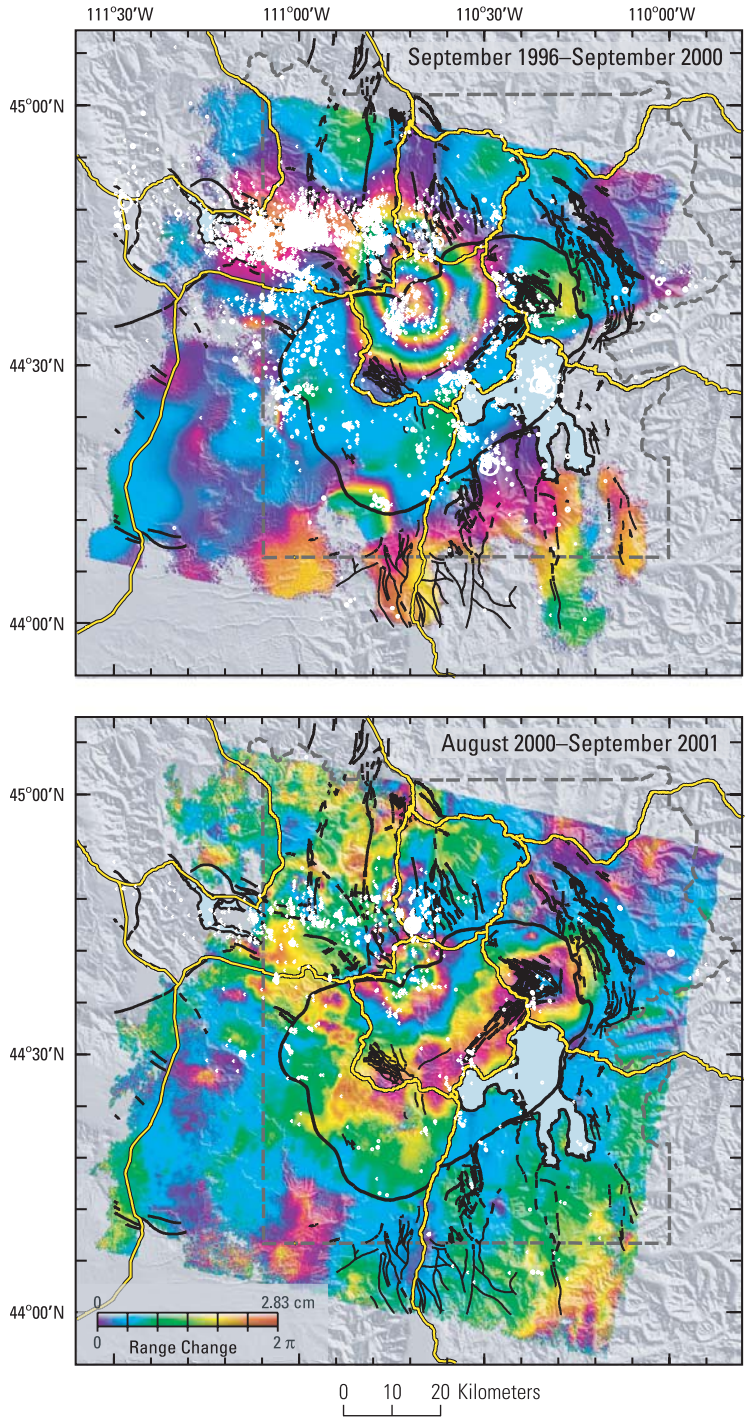

Figure 5.19. Topography-removed interferograms of the Yellowstone region for the period from September 1996 to September 2000 (*top*) and from August 2000 to September 2001 (*bottom*), showing a large uplifted area centered along the north caldera rim ~10° south of Norris Geyser Basin (NJ, Figure 5.17) and, for the later period, subsidence of the central part of the Yellowstone Caldera. White circles and dots represent earthquakes that occurred during the periods spanned by the interferograms, from the University of Utah Seismographic Stations' Yellowstone National Park Earthquake Catalogs. The caldera rim and Yellowstone National Park boundary are shown with heavy solid and dashed black lines, respectively. Mapped and inferred faults are shown with thinner black lines. Yellow lines represent roads. Uplift in the Norris area began in 1997 or 1998 and continued at an average rate of ~3 cm yr^{-1} (~1 fringe yr^{-1}) at least through 2001. Starting in 2000, the uplift was accompanied by subsidence of the central part of the caldera, including the Mallard Lake and Sour Creek resurgent domes (Wicks *et al*., 2002c).

source. According to their model, uplift occurs when part of the deep hydrothermal system becomes overpressured by magmatic volatiles. Subsidence ensues when the self-sealed layer eventually ruptures, releasing volatiles and depressurizing the deep hydrothermal system. Self-sealing and re-pressurization allow the process to repeat indefinitely.

Waite and Smith (2002) proposed a similar mechanism to explain the 1985–1986 earthquake swarm near Madison Junction, which was associated with the onset of subsidence inside the caldera. The activity was noteworthy for several reasons: (1) it was the most energetic earthquake swarm ever recorded at Yellowstone;[22] (2) seismic activity migrated at an average rate of 113 m day^{-1} to the northwest, away from the caldera rim, during the first month of the swarm, and thereafter it migrated gradually from depths of 2–5 km to more than 10 km; (3) the dominant focal mechanisms for the early part of the swarm were oblique-normal strike-slip instead of more typical normal-faulting mechanisms; (4) the maximum principal stress rotated from vertical to horizontal and subparallel to the axis of the swarm; and (5) the swarm was accompanied by increased geyser activity, changes in the clarity and temperature of hot springs, formation of new fumaroles and mud caldrons, and two small steam explosions (Dzurisin *et al.*, 1994). Waite and Smith (2002) attributed these characteristics of the swarm to lateral intrusion of magmatic or hydrothermal fluid from beneath the caldera floor into the Hebgen Lake fault zone.

The 1997–2002 uplift near Norris (12.5 cm maximum) seems to bleed westward toward the Hebgen Lake fault zone along a corridor marked by repeated earthquake swarms (Figure 5.19), which is consistent with the idea that the fault zone is connected in some way to a fluid reservoir beneath the caldera. Whether the fluid is hydrothermal or magmatic remains a mystery. The base of the hydrothermal system is thought to lie at a depth of ~5 km (Fournier, 1991, 1999), whereas the deformation source depths based on interferometry are 4–12 km beneath the resurgent domes (Wicks *et al.*, 1998) and 11–16 km beneath Norris (Wicks *et al.*, 2002c). At least in the latter case, the involvement of Yellowstone's magmatic system seems inescapable.

[22] Excluding the aftershock sequences of the 1959 M_s 7.5 Hebgen Lake earthquake and the 1975 *M* 6.1 Yellowstone Park earthquake.

5.3.4 Akutan Volcano, Alaska

Earthquake swarms are a common occurrence near volcanoes, but often whether the primary cause of the seismicity is magmatic or tectonic cannot be determined. InSAR can help to resolve this ambiguity, as illustrated by the March 1996 earthquake swarm beneath Akutan Island, Alaska. Akutan Volcano, which is situated on the western half of the island, is one of the most active volcanoes in the Aleutian arc with at least 27 separate eruptive episodes since 1790 (Miller *et al.*, 1998). An eruption in 1948 produced measurable amounts of ash in Akutan village, 13 km east of the volcano's summit. In 1978, airline pilots observed large incandescent bombs, some of them 'as big as a car', ejected more than 100 m above the summit and lava flowed down the volcano's north flank. The most recent eruptive activity was a series of small (VEI ~1)[23] steam and ash emissions during March through May 1992 and possibly into 1993.

Beginning early on 11 March 1996, Akutan Island was struck by an intense earthquake swarm that lasted about 11 hours. The swarm consisted of more than 80 earthquakes greater than *M* 3.5; the largest was *M* 5.1. A second, more vigorous swarm on 14 March lasted about 19 hours and included more than 120 earthquakes greater than *M* 3.5. Thousands of smaller earthquakes also occurred during these periods of strong activity. During peak activity, residents of Akutan Island reported feeling continuous low-level ground shaking punctuated by individual shocks about once per minute! The shaking caused considerable anxiety for the people of Akutan, many of whom left the island voluntarily in fear of an eruption. However, no eruption ensued and the level of earthquake activity gradually returned to normal during the next several months.

In July 1996, scientists from the Alaska Volcano Observatory (AVO) discovered fresh ground cracks that extended discontinuously across Akutan Island from near Lava Point on the northwest, past the summit of Akutan Volcano, nearly to the southeast end of the island (Figure 5.20). The cracks broke snowfields and their sides

[23] Volcanic Explosivity Index (VEI): a measure of the size of volcanic eruptions akin to the Richter magnitude scale for earthquakes. The VEI is a 0-to-8 index of increasing explosivity, each interval representing an increase of about a factor of ten. It combines total volume of explosive products, eruptive cloud height, descriptive terms, and other measures (Newhall and Self, 1982).

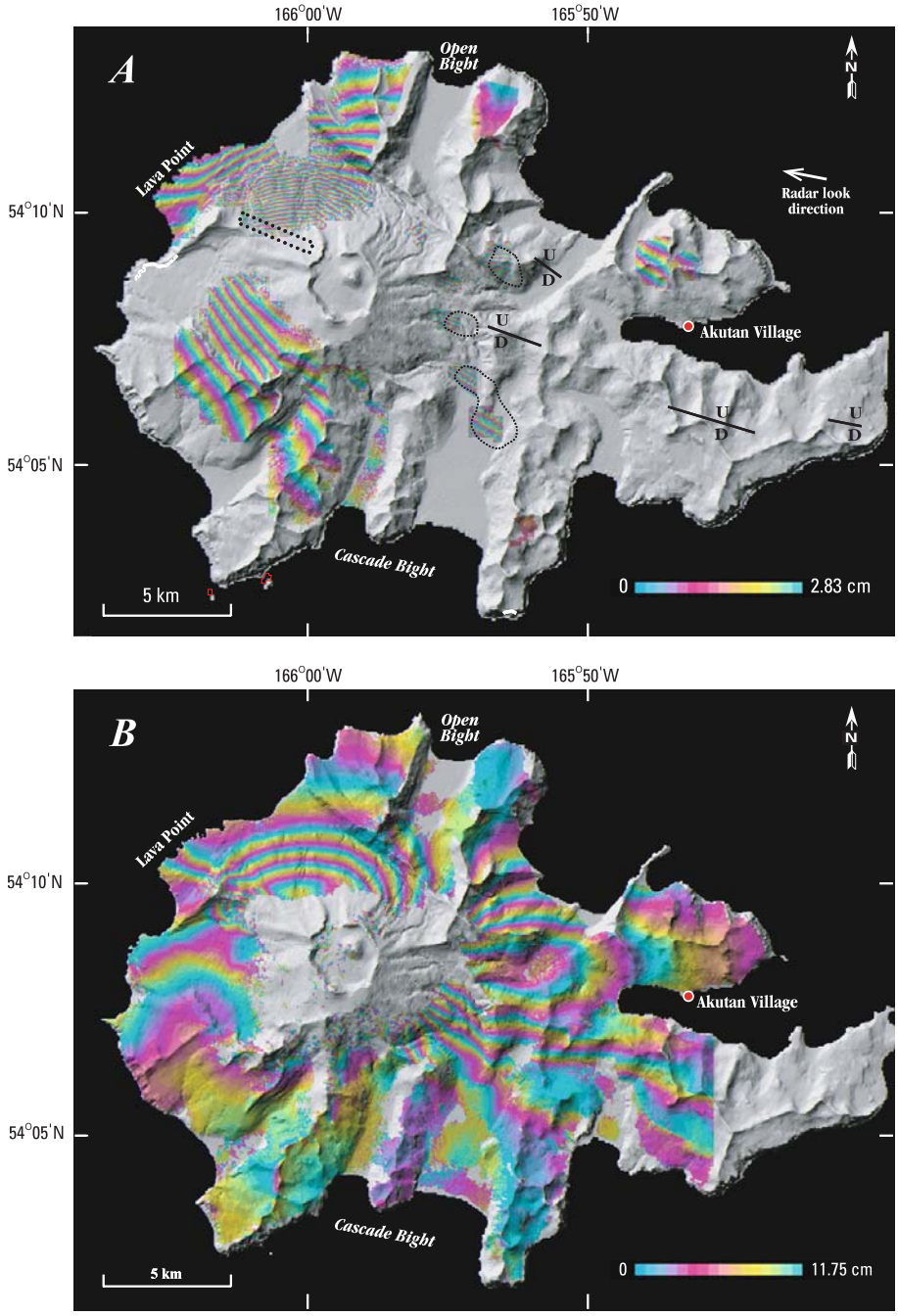

Figure 5.20. (A) Topography-removed interferogram of Akutan Island for the period from 20 August 1993 to 7 October 1996, from C-band images acquired by the ERS-1 satellite (Lu *et al*., 2000a). The breached summit caldera and active cinder cone of Akutan Volcano are visible slightly left of center in the shaded relief image. Dotted rectangle in upper left surrounds a 250-m wide zone of ground cracks that formed on the volcano's northwest flank during an intense earthquake swarm in March 1996. Four lines on the eastern part of the island represent normal faults that reactivated during the swarm, with the upthrown (U) and downthrown (D) sides, respectively. There are more than 21 fringes on the northwest flank of the volcano and more than 8 fringes on the southwest flank, representing ~60 cm and ~23 cm of uplift relative to the shoreline, respectively. Coherence on the eastern two-thirds of the island is patchy, but fringes near Akutan Village and within three areas enclosed by dotted lines indicate downwarping along an axis roughly coincident with the ground cracks and reactivated faults. (B) Coherence is much better in this interferogram produced from L-band images acquired by the JERS-1 satellite on 28 October 1994 and 22 June 1997 (Lu *et al*., 2005). In this case, lower resolution in range of L-band compared to C-band (11.76 cm per fringe versus 2.83 cm per fringe, respectively) is more than offset by improved coherence at L-band.

showed little evidence of erosion, suggesting they had formed recently – almost surely during the March 1996 swarm. The most extensive cracks occurred in a zone roughly 250 m wide and 3 km long between Lava Point and the summit. Local graben structures with net vertical displacements of 30–80 cm suggested that the cracks formed in response to tumescence of the volcano's northwest flank. The cracks on the east side of the island were only a few centimeters wide and apparently formed by activation of Holocene normal faults (Richter *et al.*, 1998). Was the March 1996 activity a tectonic earthquake swarm, a failed eruption, or both?

To address this question, Lu *et al.* (2000a) produced several interferograms of Akutan from ERS-1 and ERS-2 images using the two-pass technique and a United States Geological Survey (USGS) DEM. Figure 5.20 shows one of these for the period from 20 August 1993 to 7 October 1996, which brackets the March 1996 earthquake swarm. There are more than 21 fringes on the northwest flank of Akutan Volcano, representing ~60 cm of uplift of the upper part of the volcano relative to the northern shoreline. Similarly, there are more than 8 fringes representing ~23 cm of uplift on the volcano's southwest flank. This result was verified by forming an independent interferogram with images acquired on 4 June 1995 and 9 July 1997. The second interferogram showed essentially the same fringe pattern as the first, so any contribution from path-delay anomalies was negligible. Instead, the fringes map out ground deformation that must have occurred during the period of overlap in the interferograms (i.e., between June 1995 and October 1996), almost surely during the March 1996 earthquake swarm. Coherence on the eastern part of the island is limited to a few small patches, but the fringes in those areas are sufficient to reveal a consistent pattern of deformation.

Several features of the Akutan interferograms are noteworthy. First, the sense of motion was different on the western and eastern parts of the island. Whereas the western part of the island including Akutan Volcano moved upward, the eastern part (i.e., east of a line from Open Bight on the north coast to Cascade Bight on the south coast) moved downward. Second, the fringe pattern is asymmetric on the western flank of the volcano. The fringes are denser (i.e., deformation is more concentrated) north of the ground cracks than they are to the south. The uplift axis aligns with the zone of fresh ground cracks observed on the volcano's northwest flank. Third, fringes in several patches of coherence on the eastern part of the island indicate subsidence along an axis roughly parallel to young normal faults in the area, some of which were reactivated during the swarm. The fringe pattern in this area suggests broad downwarping along a northwest–southeast axis rather than slippage along normal faults, most of which step down to the south.

A best-fit model of the Akutan interferograms includes at least two deformation sources: (1) an east–west-trending, north-dipping dike that intruded to within 500 m of the surface beneath the northwest flank of the volcano to account for concentrated, asymmetric deformation there; and (2) a Mogi-type inflation source ~13 km beneath the northwest flank to account for broad uplift of the volcanic edifice (Lu *et al.*, 2000a). The addition of one or more contracting sources to account for downwarping does not significantly improve the overall fit, although some type of subsidence mechanism clearly is required. Subsidence might have occurred when magma moved out of deep storage within a fault system beneath the island to intrude the northwest flank of the volcano, or it might reflect some other change in the fault system before, during, or after the intrusion (e.g., stress relaxation or pore fluid depressurization).

Lu *et al.* (2005) came to a similar conclusion based on comparison and modeling of 43 interferograms from C-band ERS images and 1 interferogram from L-band JERS images, which collectively span 1992–2002. Notably, the L-band interferogram has greater coherence than the others, especially in areas with loose surface material or thick vegetation. This bodes well for the future of L-band radar interferometry at volcanoes, starting with the successful launch on 24 January 2006 of the L-band Advanced Land Observing Satellite (ALOS) (Section 5.3.7).

For our purposes, the details of the Akutan model are less important than the first-order interpretation that a magma body beneath the volcano inflated and fed a dike that propagated to within 500 m of the surface during the March 1996 earthquake swarm. Tectonic influences probably played an important role as well, but the magmatic character of the event is clear from the InSAR results. Consider how difficult logistically it would have been to obtain a similar result from EDM, GPS, or leveling results. The number of bench marks required to capture details of the deformation pattern shown in Figure 5.20 would have been prohibitive, and access to the island is difficult even during brief Aleutian summers. Placed in a broader context,

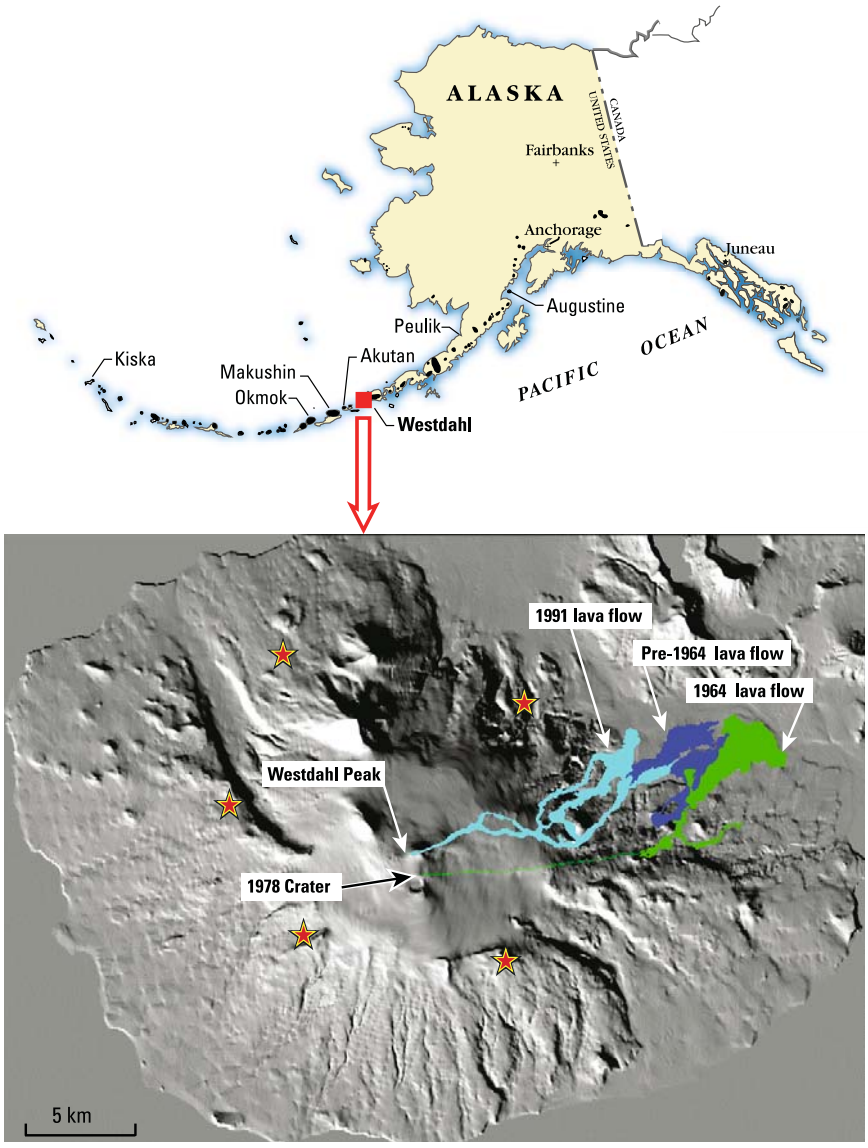

Figure 5.21. (*top*) Map of Alaska including the 2500-km-long Aleutian volcanic arc. Small red square and arrow indicate the area shown below. (*bottom*) Shaded-relief image of Westdahl Volcano. The 1978 crater, Westdahl Peak (1991--1992 vent), and three recent lava flows are labeled. Red stars represent seismic stations near the volcano. Figure slightly modified from Lu *et al*. (2003a).

the Akutan InSAR results suggest that intrusions beneath arc volcanoes are more common than previously thought, because the associated earthquake swarms easily could be mistaken for tectonic activity. Furthermore, as the following two cases illustrate, some volcanoes inflate without any accompanying unusual seismicity, so their activity would likely escape detection entirely were it not for InSAR.

5.3.5 Westdahl Volcano, Alaska

Unlike Akutan, Westdahl Volcano has been seismically quiet since its last eruption in 1991–1992. Nonetheless, InSAR studies by Lu *et al*. (2000b, 2003a) revealed an unmistakable pattern of progressive inflation at Westdahl from 1992 to 1998. This result is noteworthy because it confirms that magma sometimes accumulates and moves beneath volcanoes without causing earthquakes. This raises the possibility of longer term warnings of future eruptions, based on geodetic detection of aseismic inflation, than would be possible otherwise.

Westdahl Volcano is located on Unimak Island in the eastern Aleutian arc, about 85 km southwest of the tip of the Alaska Peninsula and 100 km northeast of Akutan Island (Figure 5.21). Known eruptions of Westdahl occurred in 1964, 1978, and 1991–1992. The 1991–1992 eruption, from a fissure through ice, produced a lava flow that extended about 7 km down the northeast flank, debris flows that reached the sea 18 km from the vent, and ash plumes to 7 km altitude

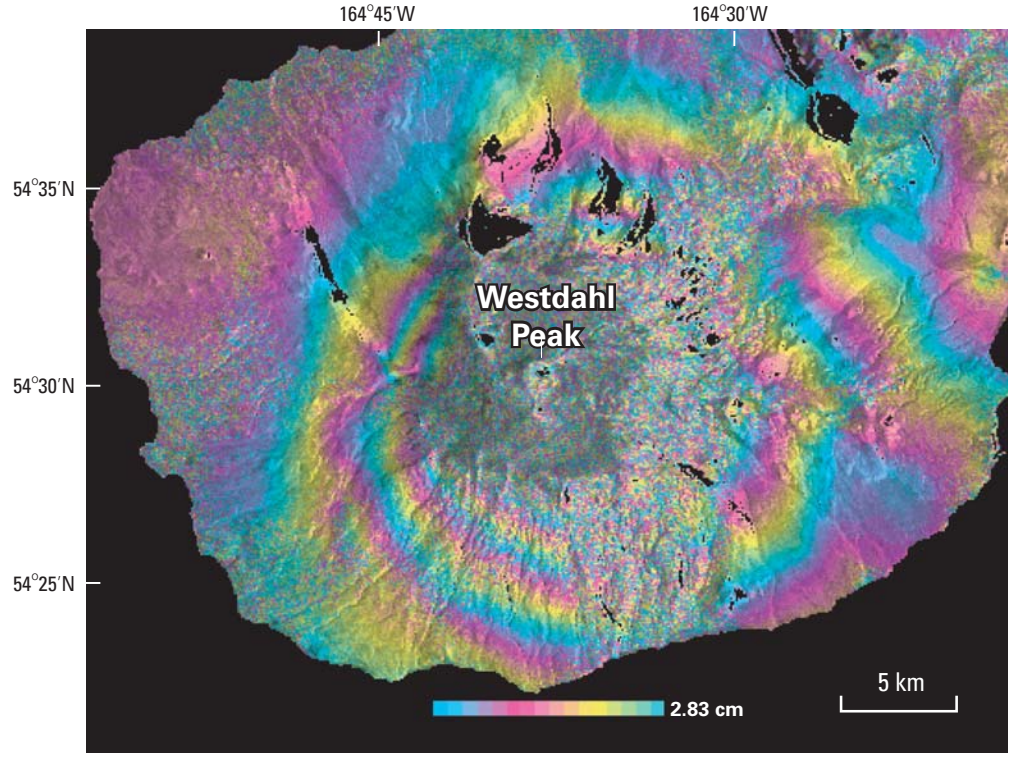

Figure 5.22. Topography-removed interferogram showing inflation at Westdahl Volcano, Alaska, from 21 September 1993 to 9 October 1998 (Lu *et al*., 2000b). Range of colors in the color bar corresponds to 2.83 cm of range change between the ground and ERS satellite. Earthquake activity was at a very low, background level throughout this period.

(Miller *et al*., 1998, pp. 45–46). The AVO installed a five-station seismic network at Westdahl Volcano in July 1998. Local seismic activity remained at a very low background level of about 5 earthquakes per year through 2001.

Motivated primarily by eruptions at nearby Shishaldin Volcano in 1995–1996 and early 1999, Lu *et al*. (2000b, 2003a) produced interferograms for Unimak Island that span the interval from 1992 to 1998 and made a surprising discovery. Unheralded by unusual activity of any kind, Westdahl Volcano began to re-inflate almost immediately after the end of its most recent eruption in January 1992 (Figure 5.22). Several independent interferograms reveal a consistent pattern of uplift centered beneath the summit area. The best-fit model is a Mogi-type source located about 8 km beneath the summit area that inflated by 0.05 km^3 from 1993 to 1998.

The Westdahl result is important for two reasons. First, it was the first time that InSAR revealed inflation of an otherwise quiescent volcano for several years *after* an eruption, and presumably *before* a future eruption. This raises the possibility that InSAR can be used to prospect for volcanic systems where magma is accumulating before other signs of unrest are detected or recognized. If so, scientists could take advantage of the advance information to intensify their monitoring efforts at inflating volcanoes and provide longer term warnings of future eruptions. Second, the inflation sources beneath Westdahl and Akutan are relatively deep (8–13 km), which means that InSAR can be used to identify deforming magma bodies before they rupture toward the surface to feed shallow intrusions or eruptions.

Local seismic activity at Westdahl began to increase gradually in 2002. The pace quickened in 2003, and a small earthquake swarm occurred beneath the volcano on 7 January 2004. Whether the seismicity is directly precursory to the next eruption remains to be seen.

5.3.6 Three Sisters volcanic center, Oregon

Another striking example of aseismic crustal uplift was provided by an InSAR study of the Three Sisters volcanic center in central Oregon, U.S.A.

Three Sisters is a long-lived center of basaltic to rhyolitic volcanism that has produced five large cones of Quaternary age: North Sister, Middle Sister, South Sister, Broken Top, and Mount Bachelor (Figure 2.33). South Sister, the youngest of the five cones, has erupted lavas ranging in composition from basaltic andesite through rhyolite. Most, if not all, of the volcano is probably of late Pleistocene age. The youngest eruptions occurred at a series of vents on the south and northeast flanks of the volcano that erupted rhyolite tephra, lava flows, and domes between about 2,200 and 2,000 years ago (Taylor *et al*., 1987; Scott and Gardner, 1990; Scott *et al*., 2001).

In the spring of 2001, my good friend and USGS colleague, Chuck Wicks, decided to study the Three Sisters area with InSAR, mainly because its eruptive history and setting make it a promising target. Geologically young eruptive products suggest that an active magmatic system may still exist beneath the area, and its eastern part is sparsely vegetated and therefore promising for InSAR. Coincidentally, the area is also one of Chuck's favorite backpacking haunts, so he was familiar with the terrain and curious whether it might be deforming. There had been no reports of unusual activity in the area and, in fact, central Oregon is known for its low level of seismic activity relative to other parts of Cascadia.

Imagine our surprise, then, when some of the first interferograms revealed a striking bullseye pattern of uplift centered about 5 km west of the summit of South Sister. An area about 20 km in diameter had been uplifted as much as 11 cm sometime between August 1996 and October 2000. An elastic point-source model indicates a volume increase by $0.023 \pm 0.003\,\mathrm{km}^3$ located 6.5 ± 0.4 km beneath the surface, presumably a magmatic intrusion (Wicks *et al*., 2001). The deformation field extends beyond the margins of EDM and tilt-leveling networks that were established in 1985 and re-measured in 1986 in anticipation of future activity at South Sister (Yamashita and Doukas, 1987; Iwatsubo *et al*., 1988). Re-observation of these networks during summer 2001 (the former by GPS) corroborated the InSAR results and showed that there had been little or no deformation prior to the time spanned by the interferograms. Subsequent analysis of additional interferograms revealed that the deformation episode started in 1997 or 1998 and that cumulative uplift had reached 14 cm by September 2001 (Figure 5.23).

At the time of this writing (May 2005), the uplift is continuing at a steady rate of about $3\,\mathrm{cm\,yr}^{-1}$ and the area is producing only a few small earthquakes ($M < 3$) per year. The USGS Volcano Hazards Program, through its David A. Johnston Cascades Volcano Observatory (CVO), has enhanced its monitoring of the area in cooperation with the US Forest Service, which manages the Three Sisters Wilderness Area. Both agencies have taken steps to keep the public informed about the activity and its implications for volcano hazards.

5.3.7 The future of volcano InSAR

In the coming years, there are sure to be successful applications of InSAR at many other volcanoes. Just as there was a period of uncertainty over the long-term availability of GPS to civilian users, some doubts are now being expressed over future satellite missions suitable for InSAR. In May 2005, JERS-1 and ERS-1 had been decommissioned after exceeding their design lifetimes. ERS-2 and RADARSAT-1 were still functioning after exceeding their design lifetimes, but problems aboard ERS-2 were limiting its usefulness for InSAR applications. Envisat, the European Space Agency's follow-on to ERS-1 and ERS-2, had completed its third full year in orbit and was functioning normally. The Japan Aerospace Exploration Agency (JAXA, formerly the National Space Development Agency of Japan, NASDA) sucessfully launched its follow-on to JERS-1, called the Advanced Land Observing Satellite (ALOS), on 24 January 2006. Also in 2006, the Canadian Space Agency plans to launch RADARSAT-2. Unfortunately, there seems to be no US InSAR mission on the horizon in the near future, although several proposals have been advanced within NASA.

In the book's final chapter, I try to envisage the near-term future of volcano geodesy, including a much-expanded role for InSAR. With a small constellation of SAR satellites dedicated to interferometry at multiple wavelengths (X-band, C-band, L-band), and with on-board GPS and careful orbit control to reduce spatial decorrelation, it should be possible in the foreseeable future to: (1) monitor most of the world's volcanoes with L-band interferometry (for greater penetration of vegetation) at repeat cycles of a few days or less; (2) resolve the problems currently posed by atmospheric-delay anomalies and lack of coherence; (3) routinely detect magma accumulation or hydrothermal pressurization before the onset of shallow seismicity or other indications of unrest; and (4) identify key volcanoes for intensified monitoring with con-

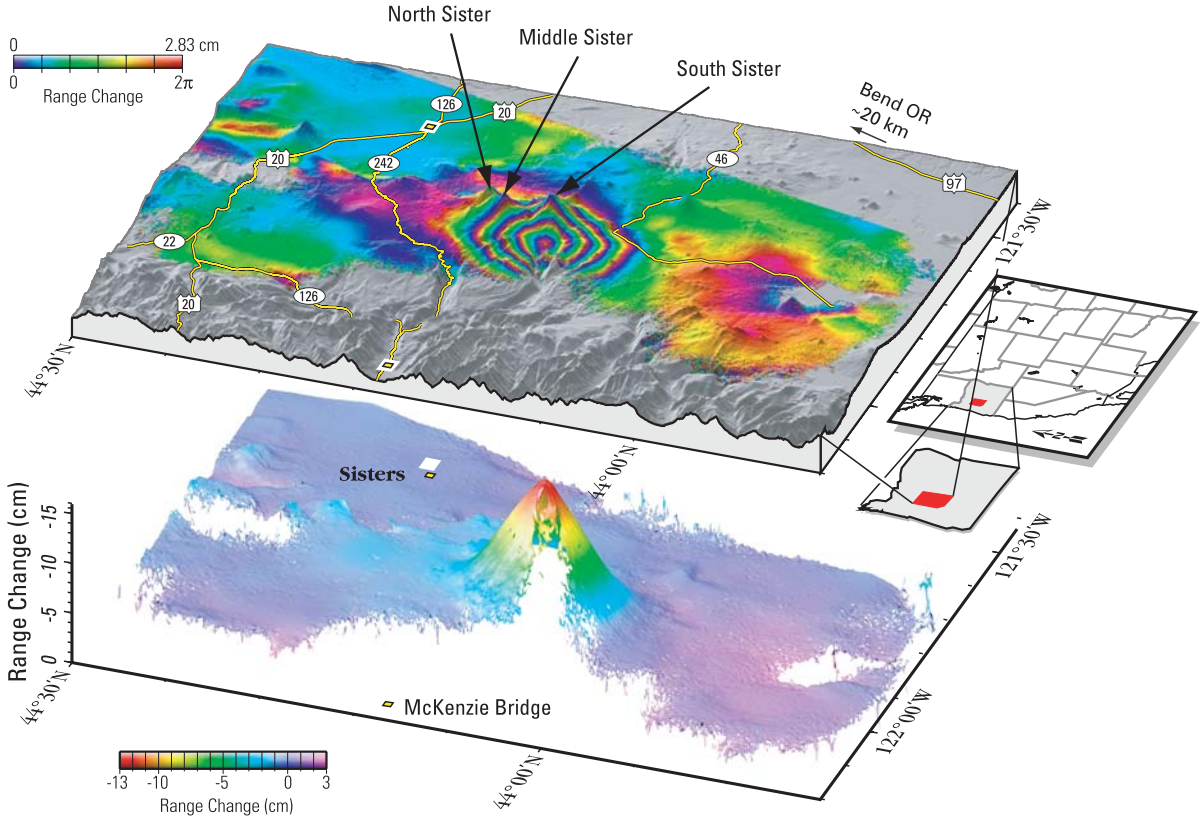

Figure 5.23. Wrapped (*top*) and unwrapped (*bottom*) interferograms of the Three Sisters area (red rectangle in inset), central Oregon Cascade Range, for the period September 1995 to September 2001 (Wicks *et al*., 2002a,b). Approximately five fringes centered 5 km west of South Sister Volcano correspond to ~14 cm of surface movement toward the satellite (mostly uplift) during this period. Analysis of additional time sequential interferograms shows that the uplift started in 1997 or 1998 and proceeded at a steady rate of 3--4 cm yr^{-1} at least through summer 2002. Data from repeated leveling and GPS campaigns, and also from two continuous GPS stations in the deforming area, show that uplift was continuing in May 2005. Modeling indicates that the uplift is caused by inflation of a source 5--7 km deep, presumably a magma body, at a rate of ~0.006 km^3 yr^{-1}.

tinuous GPS, strainmeters, tiltmeters, dense seismic networks, and volcanic gas sensors. Short-term uncertainty notwithstanding, InSAR has established itself as a powerful geodetic tool for studying ground deformation at volcanoes and elsewhere. Given its proven capabilities and vast potential, this geodetic camera is likely to remain in service for a long time to come.

In the meantime, there will be a continuing need in volcanology for other geodetic tools, including campaign and continuous GPS, leveling, photogrammetry, tiltmeters, strainmeters, and gravimeters. InSAR provides an exciting means to identify and study deforming volcanoes worldwide, but there are more riddles left to solve than any single technique or sensor can handle. The best strategy is to attack the unknown on multiple fronts, with as much weaponry as can be brought to bear. Volcanologists often find themselves outgunned, but seldom daunted, in their struggle to understand how volcanoes work.