UNIVERSITY OF CALIFORNIA
RIVERSIDE

Rapid, Robust Characterization of Subduction Zone Earthquakes

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Geological Sciences

by

Tisha Christine Irwin

June 2017

Thesis Committee:

Dr. Gareth J. Funning, Chairperson
Dr. Nicolas C. Barth
Dr. Christodoulos Kyriakopoulos

The Thesis of Tisha Christine Irwin is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am deeply indebted to Gareth Funning, without whose guidance this project would never have come to fruition.

The same can be said of Christos Kyriakopoulos. I lost count of the hours he sat with me, patiently explaining how to solve some problem that seemed intractable.

Nic Barth was not only a great teacher, but gave lots of great advice. Thank you.

I am forever grateful to Jerlyn Świątłowski and John Conrad. I couldn't have asked for better colleagues with whom to share an office. Skittles and sarcasm just may be the best office supplies.

To everyone else in the Earthquake Processes and Geophysics group: thanks for making this biologist/physician turned geophysicist feel welcome.

This work is dedicated to the two people without whose support I would never have been able to make it through this endeavor.

Mom, you always encouraged me to follow my dreams, even when they were as crazy as giving up a good career to move home and go back to school. Thanks for making sure I never ran out of toilet paper or laundry detergent.

Matt, vi pensis ke mi nur volis lerni ĉi tiun lingvon por amuzo. Vi devas labori por tio. Vi faras min pli feliĉa ol mi pensis ebla. Dankon pro esti kun mi per tiu. Via amo faras tiu tempo pli tolerebla. Mi amas vin.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Rapid Characterization of Earthquakes

Earthquake characterization, that is, the determination of earthquake source parameters, is an important component of earthquake response by the scientific community. In addition to promoting understanding of the physics of earthquake rupture initiation and propagation, the availability of source parameters has practical implications for the affected population. Knowledge of hypocenter location, fault geometry, magnitude, and slip distribution can be used to provide earthquake early warnings (EEW) [1] and tsunami warnings [2], to predict damage, fatalities, and economic losses [3], and to forecast aftershocks [4] and subsequent regional seismicity [5]. Each of these applications has a different time frame during which it is useful: an EEW system must produce a warning within seconds if it is to be beneficial, while tsunami warnings can be helpful minutes to hours after an earthquake. Aftershock forecasts, on the other hand, have a useful lifetime extending many months after a mainshock.

Forecasting aftershocks typically involves calculation of static stress changes on receiver faults following a mainshock [4]. These calculations are sensitive to the input fault geometry and slip distribution model [6]. While the urgency of EEW and tsunami warning systems dictates the use of real-time seismometer and global positioning system (GPS) networks, aftershock forecasting can take advantage of more robust models produced in the days following a major earthquake.

Interferometric synthetic aperture radar (InSAR) is a satellite-based remote sensing method that can be used to measure coseismic surface deformation and thence to infer fault location, geometry, and slip distribution. Since its first successful application to a large earthquake [7], InSAR has matured to become a major tool in the earthquake scientists kit. While the speed of data availability is primarily limited by the satellite orbital period, a recently launched mission promises to make data available within a few days of an event, as will be discussed in detail in Chapter 2.

Given the delay between earthquake and InSAR data availability, a reasonable question to pose is: what advantages does InSAR have that make it a valuable complementary technique in earthquake analysis? First and foremost, InSAR offers a combination of spatial resolution and coverage unmatched by any other geodetic technique available. The images I used in this study cover a width of 250 km and an essentially unlimited length, with a resolution of better than 5 m $\times$ 20 m in range and azimuth. Even dense networks of continuous GPS stations cannot approach this resolution, and few seismically active regions even have them [8]. Thus InSAR is ideal for studying regions with elevated seismic hazard.

Waveform seismology has been the dominant means for studying earthquakes for the last century, and seismometers still play the primary role in EEW and tsunami warning systems. In fact, the United States Geological Survey (USGS) produced a preliminary slip distribution model derived from W-phase inversion less than two hours after the 2011 Tohoku M_w 9.0 earthquake [9]. However, a priori information on slab geometry was required in order to produce a model that agreed with all available waveform data. Systematic discrepancies between slip models derived from seismic data and those derived from InSAR have been well documented [10, 11]; of particular interest are centroid location discrepancies, because InSAR produces a deformation map that is highly accurate in terms of fault location. While the potential causes of these discrepancies are still under investigation, the systematic bias in some regions, e.g., the South American subduction zone, where seismic waveform solutions are consistently located tens of kilometers trenchward of InSAR solutions, suggests that we should not rely solely on seismic waveform analysis for earthquake source characterization.

1.2 Subduction Zones and Their Earthquakes

Subduction zones are products of the convergence of tectonic plates. The apparent simplicity of a dense oceanic plate sinking into the mantle belies the geophysical, geodynamic, mineralogical, and morphological complexity of these regions. Despite decades of intense study by a wide array of earth scientists, many questions remain to be answered. In this section, I will give a brief overview of the current state of knowledge about subduction zones and their earthquakes.

Subduction zones occur at the sites of downwelling of mantle convection cells and they are the main sites of continent growth. The subducting lithospheric plate is composed of oceanic crust and its underlying lithospheric mantle, along with a variable thickness of sediments. The mantle is composed of peridotite, a rock whose most abundant component minerals are polymorphs of olivine. The phase transitions of olivine are responsible for the seismic velocity discontinuities that define the boundaries between the upper mantle and transition zone, and between the transition zone and lower mantle. These mineralogical changes also influence the nature of earthquakes that occur in the Wadati-Benioff zone.

Globally, subduction zone megathrust earthquakes account for upwards of 90 % of the seismic moment released since 1900 (figure 1.1); in fact, just six such events contributed nearly half of the moment release in that time period [12, 13]. However, megathrust earthquakes, those caused by slip along the shallow plate interface, are not the only earthquakes to occur in subduction zones. Shallow earthquakes, at depths less than 70 km, also occur in the downgoing plate at the outer rise and in the deformed overriding plate. Intermediate depth earthquakes, from 70 km to 300 km, and deep earthquakes, from 300 km to 700 km, occur within the subducting slab; these earthquakes define the Wadati-Benioff zone. However, this study is concerned with megathrust earthquakes occurring on the shallow plate interface, typically at depths of less than 25 km, which represent the greatest seismic hazard in subduction zone regions.

A cursory examination of the geometry of a subduction zone reveals the reason for this elevated hazard: the shallow dip of the plate interface creates a potentially enormous area that can slip. In addition, depression of the geothermal gradient pushes the

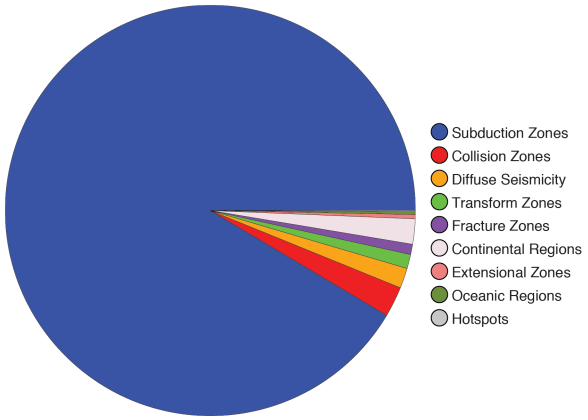


Figure 1.1: Global seismic moment release by tectonic setting. Subduction zones account for more than 90 % of the moment release worldwide. After [13].

brittle-ductile transition deeper than is found in other crustal regions [14]. Seismic moment is determined by the product of average slip, $\bar{s}$, and fault area, A: $M_0 = \mu A \bar{s}$. Thus even modest amounts of slip over large areas can release tremendous seismic moment. Furthermore, vertical displacement of the seafloor is responsible for the generation of tsunami, which are frequently responsible for more casualties and damage than ground shaking from the earthquake itself.

Comprehensive understanding of the earthquake cycle at subduction zones is hampered by the short instrumental record and variable seismic behavior among different subduction zones. The degree of coupling of the plate interface varies widely [15], and some subduction zones are segmented such that earthquakes repeatedly rupture the same segments or combinations of segments, for example, the Nankai Trough in Japan [16]. Aseismic slip occurs at most, if not all, subduction zones, and may occur at any time in the earthquake cycle [17]. Although it is not possible to predict the timing of a megathrust earthquake, their

characteristics and associated hazard make them attractive targets for hazard mitigation strategies. Thus my research is focused on these events.

1.3 InSAR Observation of Earthquakes

Because it is not possible to study an earthquake at the source, i.e., within the slipping fault, we must infer source parameters from remote observations. These observations are typically of two general types: observations of seismic waves radiated from the source, and observations of surface deformation. While a few authors have reported inversion of coseismic gravity changes for source parameters, e.g., [18], the application of this technique is limited to great earthquakes. Observations of coseismic surface deformation date to the late nineteenth century [16] and have become increasingly sophisticated as technologies have matured. The traditional geodetic techniques of triangulation, trilateration, and leveling have largely been supplanted by space geodetic methods, including geodetic-quality GPS networks and InSAR.

GPS networks are less labor-intensive than surveying methods, can continuously monitor deformation, and can cover a greater spatial extent [19]. GPS networks have contributed greatly to our understanding of plate tectonics and the earthquake cycle, particularly in measuring tectonic plate velocities and interseismic deformation and strain accumulation. Unfortunately, in many seismically active regions, GPS networks are sparse to nonexistent because of the financial and logistical challenges of instrumenting these areas. Thus there are relatively few places on Earth where GPS data are sufficiently dense for high-quality source characterization. InSAR is the technique of choice for filling this gap.

I will provide a detailed technical discussion of InSAR in Chapter 2; in this section, I will highlight some of the limitations and capabilities of InSAR when used to measure coseismic deformation.

InSAR uses the phase information of two collocated radar images to determine the line-of-sight distance, or "range change", from the radar to the surface. Because the radar measures distance, it is most sensitive to displacements that are either horizontal but parallel to its line of sight, or vertical; it is largely insensitive to horizontal displacements that are perpendicular to its line of sight. If the two images are acquired on different dates and the surface has moved in the interim, there will be a change in phase in the pixels corresponding to the areas of surface deformation. In fact, this interferometric phase is the sum of the phase contributions from several factors: change in position of the satellite between image acquisitions, topography, atmospheric propagation delay, change in characteristics of the surface represented in the pixel, and displacement [20]. These contributions can be represented mathematically:

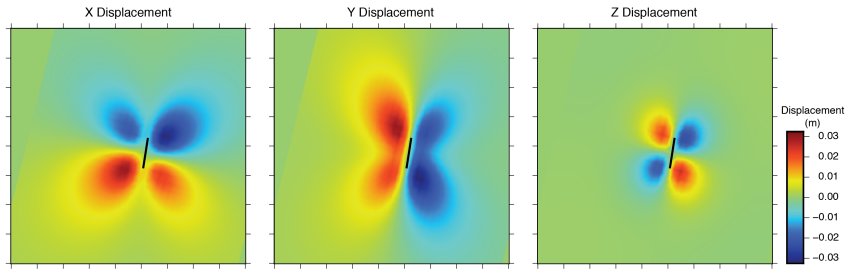
$$\Delta\phi = \Delta\phi_{pos} + \Delta\phi_{topo} + \Delta\phi_{atm} + \Delta\phi_{pixel} + \Delta\phi_{disp} \tag{1.1}$$

While it is the phase change due to displacement that is of interest, the other contributions must be corrected for or acknowledged when interpreting these images; these factors will be discussed further in Chapter 2. Furthermore, there are practical limits on the ranges of surface deformations detectable by InSAR [21]; these limits are largely defined by characteristics of the radar system.

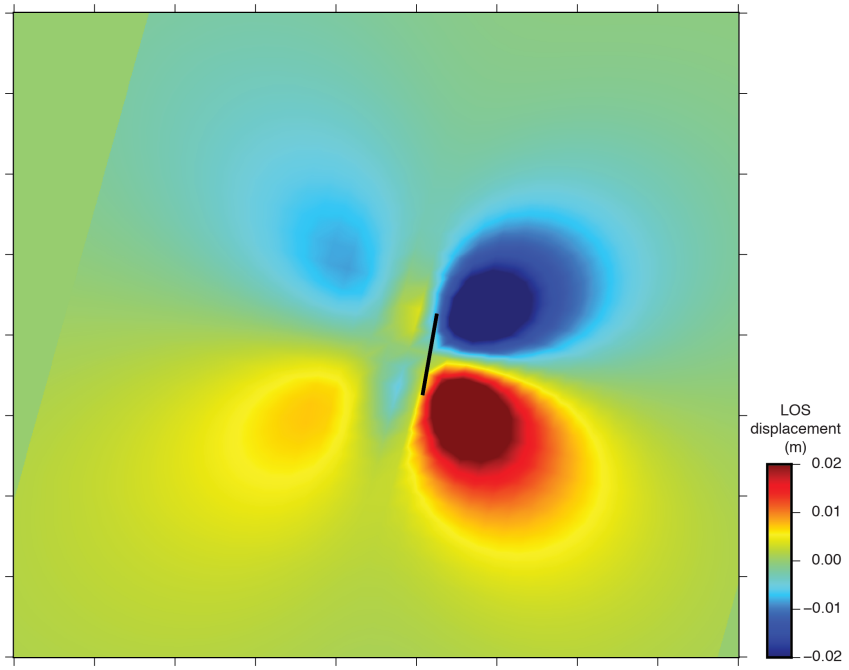
The spatial extent of the deformation must be many times larger than the ground extent of a single pixel, and the deformation gradient, or the range change per pixel, must

be within about 10^{-7} to 10^{-3} strains [21]. Finally, the achievable resolution of the phase difference places a lower limit on the detectable range change. When measuring coseismic deformation, the vector surface displacements must be of sufficient magnitude and in the proper orientation to fall within these limits. Thus the minimum detectable magnitude depends on the centroid depth and focal mechanism.

One can see how these parameters affect the detected displacement by examining synthetic interferograms. For a simple strike-slip earthquake (figure 1.2), the deformation pattern shows four lobes of displacement, each of which is the superposition of vertical and horizontal displacements projected to the radar's line of sight, i.e., the line-of-sight displacement is the dot product of the displacement vector and the line-of-sight vector. The amount of displacement in each lobe "seen" by the radar depends on the fault's orientation relative to the satellite's trajectory. For thrust earthquakes (figure 1.3), one particularly useful observation is the location of the sign change from displacement toward to away from the radar, as this approximates the surface projection of the downdip rupture extent. At this location, vertical displacements change from uplift to subsidence. In subduction zone megathrust events, much of the deformation pattern, including this sign change, is offshore and thus invisible to InSAR. This represents one of the biggest challenges to using InSAR for subduction zone megathrust earthquakes. However, the availability of a priori fault geometry models nevertheless allows us to infer realistic slip models despite this limitation.

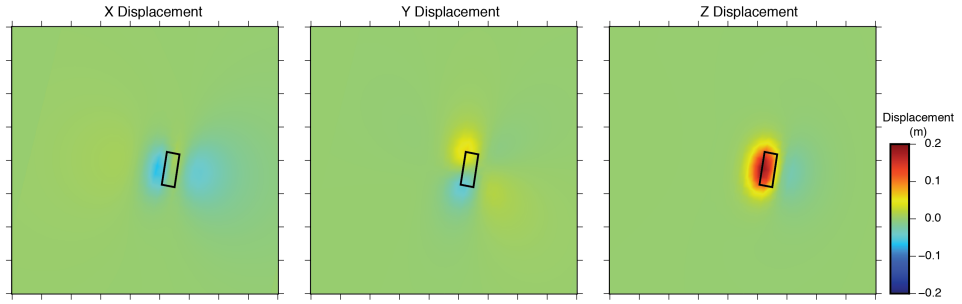


(a) X, Y, and Z components. East, north, up are positive.

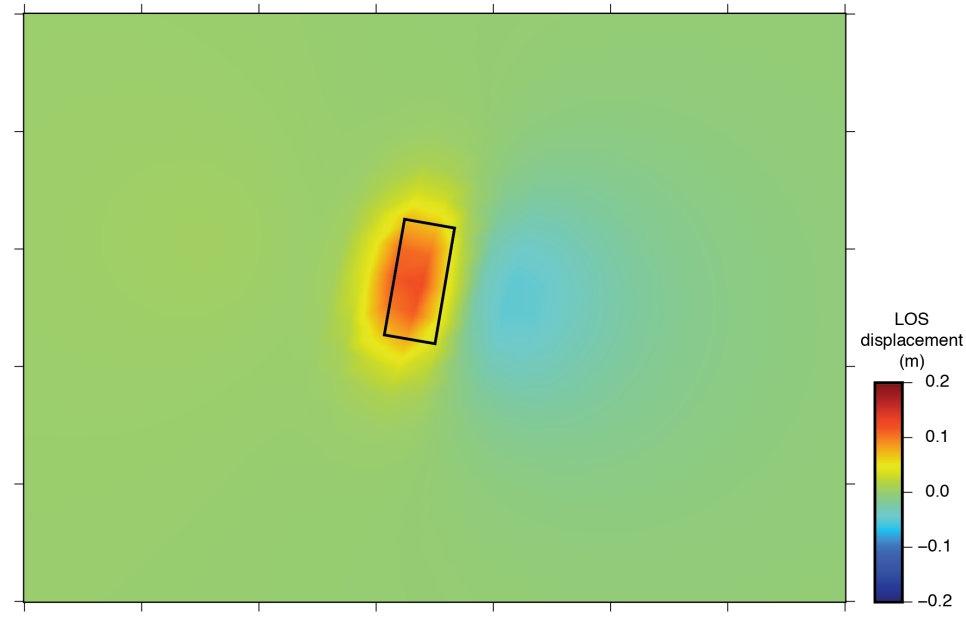


(b) Line-of-sight displacement.

Figure 1.2: Synthetic interferograms for a right-lateral strike-slip earthquake. The X, Y, and Z components of displacement are consistent with those expected from this sense of slip. The pattern is dominated by southward displacement of the surface to the east of the fault; the X and Z components show the rotation and uplift or subsidence expected at the fault tips. Black line is projection of fault.



(a) X, Y, and Z components. East, north, up are positive.



.

(b) Line-of-sight displacement.

Figure 1.3: Synthetic interferograms for a thrust earthquake. The patterns is dominated by vertical displacement: uplift in the area of the fault projection (black box) and subsidence beyond the down-dip extent of rupture.

Chapter 2

InSAR

2.1 Introduction

InSAR is a geodetic technique that uses satellite radar imaging to measure surface deformation. Satellite radar imaging is a form of active remote sensing that uses microwave radiation to illuminate a portion of Earth's surface; the reflected signal is processed to create high-resolution imagery. The satellite platform provides an unobstructed view of the surface, and active illumination provides day-or-night, all-weather imaging capability, as clouds are transparent to radar.

In this chapter I will develop the mathematics involved in creating an image from a radar signal and using these images to measure surface deformation. I will then describe specific features of the data used in this study and how their processing differs from typical InSAR processing. In Appendix A, I provide an introduction to radar systems in general.

2.2 SAR Image Formation

In this section, I will describe the synthetic aperture concept and introduce the reader to the various modes of synthetic aperture radar (SAR) acquisition. I will then define a coordinate system for a SAR acquisition and discuss the signal processing steps required to produce an image from a series of radar returns.

2.2.1 Synthetic Aperture Concept

The antenna footprint, or that area of the surface illuminated by the radar beam, is an area of finite extent. As the satellite travels along its trajectory, the footprint moves along the surface and scatterers on the surface move into, through, and out of the mainbeam. Thus a slightly different area is illuminated with each pulse, and the returns from each scatterer are smeared over several pulses. The angular width of the mainbeam in azimuth determines the number of pulses over which each scatterer is illuminated and therefore how smeared each return is. In real-aperture radar, each pulse is processed to form a separate profile in the range dimension. Hence, the width of the mainbeam determines the azimuth resolution of a real-aperture radar image: the narrower the mainbeam, the better the resolution. As discussed in section A.1.2, the diameter D of the antenna aperture determines the width of the mainbeam; the larger the aperture, the narrower the beam. In terms of azimuth beamwidth, the relevant dimension is antenna length, L. The azimuth beamwidth is also range-dependent, so the azimuth resolution, ΔA, is

$$\Delta A \approx R\theta_3 \approx R\frac{\lambda}{L} \tag{2.1}$$

In a SAR system, the returns collected over some along-track distance D_{SAR}, called the synthetic aperture, are coherently combined, thus the effective SAR beamwidth is

$$\theta_{SAR} = \frac{\lambda}{2D_{SAR}} \tag{2.2}$$

and the azimuth resolution becomes

$$\Delta A \approx \frac{R\lambda}{2D_{SAR}} \tag{2.3}$$

Thus a longer synthetic aperture results in better resolution. Since $D_{SAR} = vT_d$, where T_d is the dwell time, a longer dwell time will produce a longer synthetic aperture and finer azimuth resolution.

SAR Acquisition Modes

There are several modes of acquisition for SAR systems. The simplest is stripmap mode (figure 2.1a), in which a fixed antenna continually transmits pulses as the satellite travels along its trajectory. In stripmap mode, the range resolution is determined by the chirp bandwidth (section A.2.3). The integration angle, θ_{int}, is the angular extent over which pulses can be coherently combined, and is

$$\theta_{int} = \frac{D_{SAR}}{R} \tag{2.4}$$

The integration angle is no larger than the real antenna beamwidth; thus

$$\Delta A_{stripmap} = \frac{\lambda}{2\theta_{int}} \geq \frac{\lambda}{2\theta_3} \geq \frac{L}{2} \tag{2.5}$$

Thus the azimuth resolution in stripmap mode is determined by the physical antenna size. One can also express the azimuth resolution in terms of Doppler bandwidth: recalling the

equation for Doppler bandwidth (equation A.43) and assuming no squint,

$$B_D = \frac{2v\theta_3}{\lambda} = \frac{2v}{L} \tag{2.6}$$

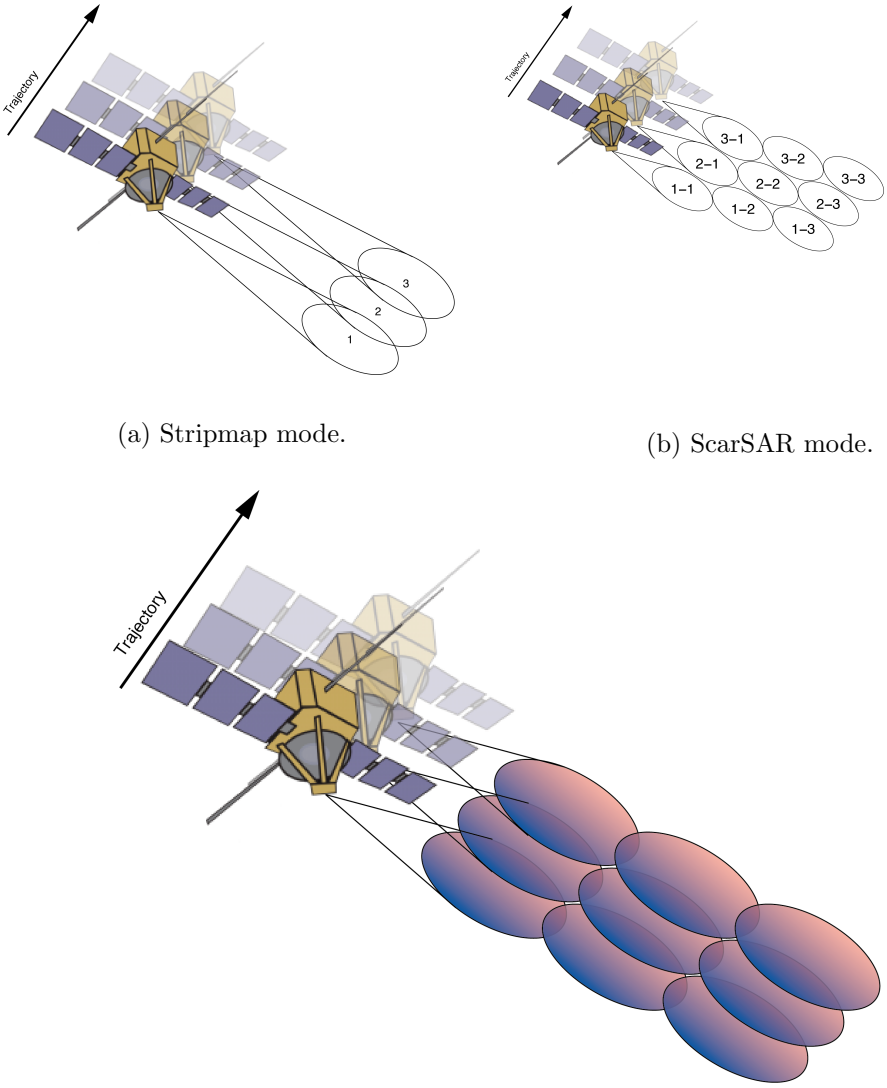
and

$$\Delta A \geq \frac{v}{B_D} \tag{2.7}$$

Thus a wider Doppler bandwidth results in finer azimuth resolution.

In order to increase the swath width, and thereby shorten the revisit interval, ScanSAR mode was developed (figure 2.1b). In this mode, a phased array antenna (section A.1.2) is used to steer the beam in elevation. At each elevation setting, corresponding to a different range extent, a sub-swath is acquired in a fashion similar to stripmap mode. However, while the spatial coverage of the surface is continuous, not every point on the surface is illuminated by the mainbeam. This limits the Doppler bandwidth of each target, reducing the azimuth resolution; it also creates amplitude variations within each sub-swath, an effect called scalloping. These effects require near-perfect scan alignment between image acquisitions in order to allow interferometry.

In an effort to eliminate this scalloping effect while maintaining a wide swath, the newest generation of SAR instruments can operate in Terrain Observation by Progressive Scans (TOPSAR) mode (figure 2.1c). In addition to steering the beam in elevation, these instruments also steer the beam in azimuth, producing a variable squint angle (the angle between the normal to the satellite's velocity vector and the antenna's line-of-sight (LOS) direction to the target) that moves from backward to forward over the course of each sub-swath acquisition. Thus each point on the surface is illuminated by the mainbeam. Consequences of this acquisition mode will be further discussed in section 2.4.1.



(a) Stripmap mode.

(b) ScarSAR mode.

(c) TOPSAR mode.

Figure 2.1: Synthetic aperture radar acquisition modes. In stripmap mode, the fixed antenna continually acquires images along the satellite's trajectory. In ScanSAR mode, the antenna is electronically steered in elevation, acquiring images in non-overlapping sub-swaths. In TOPSAR mode, the antenna is steered in elevation and azimuth; sub-swaths overlap both in range and in azimuth and each sub-swath has variable squint.

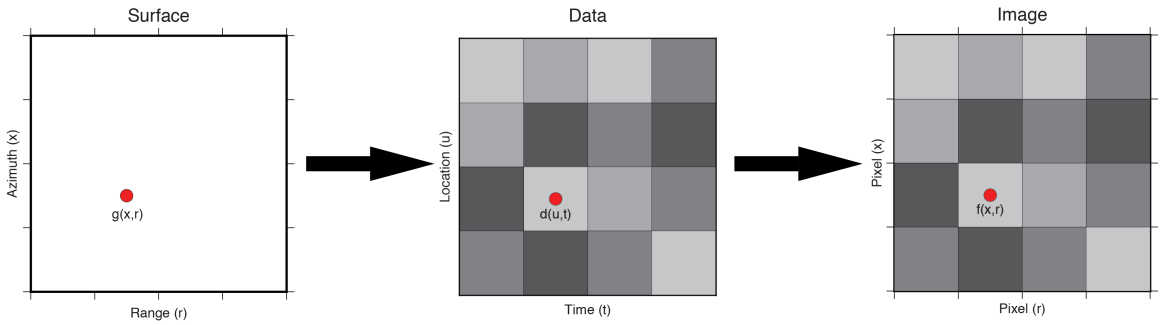


Figure 2.2: Illustration of coordinate systems used to refer to: scatterer location on the surface, $g(x, r)$; data, $d(u, t)$; and pixel in image, $f(x, r)$.

2.2.2 Coordinate Systems

Before describing the image formation process, it is helpful to define coordinate systems for the scene, data, and image. A radar geometry coordinate system will be defined where the azimuth axis is parallel to the satellite's trajectory, while the range axis is parallel to the nominal 90° look direction of the antenna (figure 2.2). A scatterer on the surface will be located at a position x along the azimuth axis and r along the range axis; thus its position is $g(x, r)$. The acquisition process samples the scene, yielding a two-dimensional matrix of voltage values where each pulse is sampled at some time interval directly corresponding to a slant range value; these values are indexed by time, t. The sequence of pulses, each of which was transmitted while the satellite was at a different position along the azimuth axis, is indexed by the satellite position, u; thus the data is $d(u, t)$. Finally, the image formation process converts the data into a two-dimensional matrix of pixels, each with an azimuth-range location; thus the pixel location in the image is $f(x, r)$.

Because Fourier analysis is an integral part of SAR image processing, it is helpful to define how the coordinates in $d(u,t)$ are defined in the Fourier domain. The time dimension, t, is represented in the Fourier domain as temporal frequency, ω, in units of Hz. The spatial dimension, u, is represented as spatial frequency, or wavenumber, k_u (section A.1.1), in units of $\mathrm{rad\,m^{-1}}$. I will show in the next section that this value is equal to

$$k_u = -\frac{4\pi}{\lambda_c}\left(\frac{x}{r}\right) \tag{2.8}$$

where λ_c is the wavelength corresponding to the center frequency of the chirp pulse.

2.2.3 Point Spread Response

The most fundamental function in SAR image formation is the point spread response (PSR) [22]. This defines the slant range, R, to a scatterer given the locations of the satellite and scatterer. By simply applying the Pythagorean theorem:

$$R(u;x,r) = \sqrt{(u-x)^2 + r^2} \tag{2.9}$$

(figure 2.3). This function has the form of a hyperbola; as the radar repeatedly interrogates the scatterer, the range to the scatterer is at a maximum when the scatterer first enters the antenna footprint, decreases to a minimum when $u = x$ and the scatterer is broadside to the antenna, and increases back to the maximum until the scatterer exits the footprint. Furthermore, the echo for any point target will be spread over several range and Doppler bins over the course of the dwell; this is referred to as range cell migration. Thus the goal of image formation algorithms is to focus the return for a single point scatterer into a single pixel.

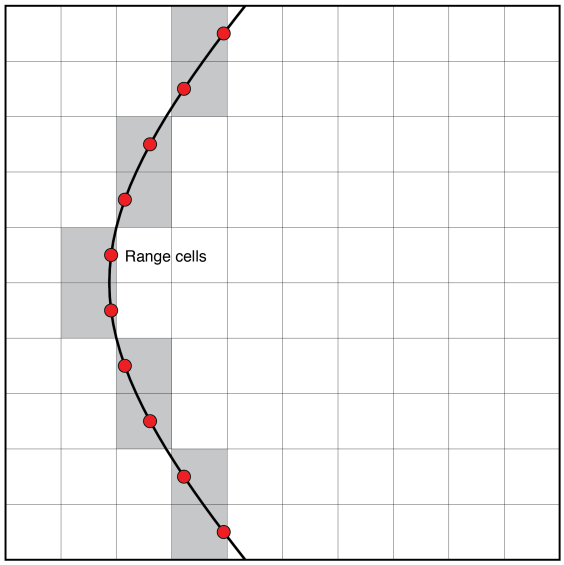


Figure 2.3: Point spread response and range cell migration. Each column represents a range cell. Echoes (dots) move through several range cells during the acquisition.

The shape of the hyperbola is constant for all scatterers at the same value of r, but changes for differing values of r; in other words, the PSR is spatially invariant in x, but spatially variant in r. The PSR can be expressed in terms of time by substituting the value of $R(u;x,r)$ into equation A.1:

$$t(u;x,r) = \frac{2}{c}\sqrt{(u-x)^2 + r^2} \tag{2.10}$$

Furthermore, by considering the reflection from a point scatterer to be an impulse function,

$$h(u,t;x,r) = \delta_D\left(t - \frac{2}{c}\sqrt{(u-x)^2 + r^2}\right) \tag{2.11}$$

To define the relationship between spatial frequency and scatterer location, we can first define how the time delay to its echo changes as a result of satellite motion. By expanding the terms under the radical, approximating using the binomial expansion, and

ignoring the terms that are independent of satellite position, we see that

$$\Delta t \approx \frac{2}{c}\left(\frac{u^2}{2r} - \frac{ux}{r}\right) \tag{2.12}$$

The phase progression associated with this time delay is

$$\Delta\phi = 2\pi f_c \Delta t \approx \frac{4\pi}{\lambda_c}\left(\frac{u^2}{2r} - \frac{ux}{r}\right) \tag{2.13}$$

where f_c is the center frequency of the chirp pulse. The spatial frequency is the rate of change of phase with platform motion, i.e., the derivative of $\Delta\phi$ with respect to u:

$$k_u = \frac{\partial}{\partial u}\Delta\phi(u) = \frac{4\pi}{\lambda_c}\left(\frac{u}{r} - \frac{x}{r}\right) \tag{2.14}$$

Thus for a given scatterer at (x, r), the value of k_u varies linearly with platform location. Furthermore, there is a unique mapping from spatial frequency to Doppler frequency:

$$f_d = -\frac{1}{2\pi}k_u v \tag{2.15}$$

so that each scatterer has a linear variation in Doppler frequency.

2.2.4 Matched Filtering

Matched filtering is a means by which to increase the signal-to-noise ratio (SNR). For a time-domain impulse response $s(t)$, the general matched filter has the form

$$h(t) = s^*(-t) \tag{2.16}$$

It is thus the time-reversed complex conjugate of the signal. Filtering of signal data $d(t)$ can be implemented as a conjugation in the time domain:

$$y(t) = d(t) * h(t) = \int_{-\infty}^{\infty} d(\alpha)h(t-\alpha)d = \int_{-\infty}^{\infty} d(\alpha)s^*(\alpha - t)d\alpha \tag{2.17}$$

where α is a dummy variable of integration. This is mathematically equivalent to correlation with the complex conjugate; it maximizes the recovered signal in the presence of noise by producing the largest integral at the lag where the signal and the filter perfectly align [23]. Alternatively, matched filtering can be implemented as a multiplication in the frequency domain, where the matched filter is $H(\omega) = S^*(\omega)$ so that

$$Y(\omega) = D(\omega)H(\omega) = D(\omega)S^*(\omega) \tag{2.18}$$

where $D(\omega)$ and $S(\omega)$ are the Fourier transform (FT) of $d(t)$ and $s(t)$, respectively.

2.2.5 Range Compression

Range compression is the application of a matched filter over the course of a single pulse; the received waveform, $x_r(t)$, is a time-delayed, phase-shifted, and amplitude-scaled replica of the transmitted pulse, $x(t)$:

$$x_r(t) = be^{j\phi}x(\alpha - t_d)x^*(\alpha - t)d\alpha \tag{2.19}$$

where b is the scaled amplitude and t_d is the time delay. Thus, the output of the matched filter is:

$$y(t) = \int_{-\infty}^{\infty} be^{j\phi}x(\alpha - t_d)x^*(\alpha - t)d\alpha \tag{2.20}$$

In terms of its spectral components:

$$y(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} be^{j\phi}X(\omega)e^{-j\omega t_d}X^*(\omega)e^{j\omega t}d\omega \tag{2.21}$$

where $X(\omega)$ is the FT of the transmitted waveform. In the case of a linear frequency modulated (LFM) waveform, the matched filter output is:

$$y(t) = \left(1 - \frac{|t|}{\tau}\right)\frac{\sin\left(1 - \frac{|t|}{\tau}\right)\pi Bt}{\left(1 - \frac{|t|}{\tau}\right)\pi Bt} \tag{2.22}$$

which approximates the product of a triangle function and a sinc function, defined over an interval $-\tau \leq t \leq \tau$. The first null occurs when the argument of the sine function is equal to π, or at $t \approx \pm\frac{1}{B}$. Extending the explanation of range resolution from section A.2.2, two objects at different ranges can be resolved as long as the mainlobes of their matched filter outputs do not overlap, so $\Delta R = \frac{c}{2B}$.

2.2.6 Azimuth Compression

Following range compression, azimuth compression focuses the image in azimuth. Several algorithms are available, each with its own advantages and disadvantages in terms of fidelity and computational complexity. The most widely used is the Range-Doppler Algorithm (RDA), so called because it operates in the time (which maps to range) and spatial frequency (which maps to Doppler) domain. However, prior to azimuth compression, the range cell migration mentioned in section 2.2.3 must be corrected so that all the energy returned from a single target is focused into a single range bin.

After range cell migration correction, the RDA begins with the PSR:

$$R = \sqrt{(u - x^2) + r^2} \tag{2.23}$$

By defining r to be an offset, Δr, relative to the center of the image, r_0, such that $r = r_0 + \Delta r$, and assuming that the dimensions of the image are much smaller than the slant range to the center, i.e., $r_0 \gg \Delta r$ and $r_0 \gg |u - x|$, one can approximate the PSR to be

$$R \approx \sqrt{(u - x)^2 + r_0^2} + \Delta r \tag{2.24}$$

or

$$h(u, t; x, r_0 + \Delta r) \approx \delta_D\left(\left(t - \frac{2}{c}\Delta r\right) - \frac{2}{c}\sqrt{(u - x)^2 + r_0^2}\right) \tag{2.25}$$

In this form, the PSR is spatially invariant in both dimensions; that is, the PSR at r is simply a shifted version of that at r_0. Therefore a single PSR can be calculated for the entire image by setting x equal to 0:

$$h(u,t;r_0) = \delta_D\left(t - \frac{2}{c}\sqrt{u^2 + r_0^2}\right) \tag{2.26}$$

Because applying a matched filter via convolution in the time domain is more computationally expensive than multiplication in the frequency domain, even after accounting for the forward and inverse Fourier transforms, the RDA can be implemented as follows: perform fast Fourier transformation (FFT) of the data to $D(k_u,\omega)$; choose a suitable value of r_0; calculate the corresponding PSR; subject it to FFT to generate $H(k_u,\omega;r_0)$; multiply the two spectra; and compute the inverse fast Fourier transform (IFFT) of the product. It is therefore necessary to compute a frequency-domain form of the PSR.

This can be accomplished by first time-shifting the PSR, thereby eliminating a zero-padded time interval, such that

$$\begin{aligned} H(k_u,\omega;r_0) &= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}\delta_D\left(t + \frac{2}{c}r_0 - \frac{2}{c}\sqrt{u^2+r_0^2}\right)e^{-j\omega t}dt\right)e^{-jk_u u}du \\ &= A_1 e^{jr_0\left(\frac{\omega}{c/2} - \sqrt{(\frac{\omega}{c/2})^2 - k_u^2}\right)} \end{aligned} \tag{2.27}$$

where the A_1 term is often ignored. Therefore, the PSR has a phase of

$$\phi_{PSR} = r_0\left(\frac{\omega}{c/2} - \sqrt{\left(\frac{\omega}{c/2}\right)^2 - k_u^2}\right) \tag{2.28}$$

The range frequency is due to the carrier frequency and is therefore much larger than the azimuth spatial frequency, which is due to Doppler shift, so $2\omega c \gg k_u$. Furthermore, we can define the frequency ω to be the sum of the center frequency, ω_0 and an offset, $\Delta\omega$, and

assume the bandwidth to be much smaller than the center frequency. Thus we can use the first three terms of the Taylor expansion to approximate the phase of the PSR:

$$\phi_{PSR} \approx \frac{cr_0k_u^2}{4\omega_0} - \frac{cr_0k_u^2}{4\omega_0^2}\Delta\omega + \frac{cr_0k_u^2}{4\omega_0^3}\Delta\omega^2 \tag{2.29}$$

so that

$$H(k_u, \omega = \omega_0 + \Delta\omega; r_0) \approx e^{j\frac{cr_0k_u^2}{4\omega_0}} e^{-j\frac{cr_0k_u^2}{4\omega_0^2}\Delta\omega} e^{j\frac{cr_0k_u^2}{4\omega_0^3}\Delta\omega^2} \tag{2.30}$$

In this form of the matched filter, the first term is a phase modulation that is quadratic in k_u but independent of radiofrequency, the second term is a linear phase modulation over the chirp bandwidth, and the third term causes defocussing of returns which worsens at the top and bottom of the image.

This defocussing can be compensated for by a modification of the range-compression step described in section 2.2.5. Rather than perform pulse compression in the (u, ω) domain, it can be performed in the (k_u, ω) domain, allowing for correction of the defocussing effect; this is called secondary range compression.

2.3 Interferometric SAR

Interferometry exploits the phase difference between two coherent SAR images to measure range change, and thus surface deformation, on the scale of a fraction of a radar wavelength. In this section I will describe the steps involved in interferogram formation as well as the underlying mathematics, and I will discuss some of the limitations of the technique when used for detection and measurement of surface deformation.

2.3.1 Acquisition Geometry

Before discussing interferogram formation, it will be helpful to define some terms related to the geometry of acquisition of multiple SAR images. The two images used for interferogram formation may be acquired at different times, from different positions, or both. When measuring surface deformation, we are interested in range change over a time interval, but it is essentially impossible for the platform to be in the identical location at the two acquisition times, thus both factors come into play. The time difference is the temporal baseline, while the distance between the positions is the spatial baseline, or simply the baseline. The spatial baseline can be decomposed into the perpendicular baseline, the component perpendicular to the line of sight, and the parallel baseline, perpendicular to the satellite's trajectory (figure 2.4). The angle between the line-of-sight vector and the vertical at the radar is the look angle, while that between the line-of-sight vector and the vertical at the surface is the incidence angle. The incidence angle varies across the image due to both the radar beamwidth and Earth's curvature.

2.3.2 Image Registration

Because the spatial baseline is unlikely to be zero, the pixels in the two images are unlikely to correspond perfectly. Because each pixel is the coherent sum of the complex returns from many scatterers within that pixel, a shift of even a fraction of a pixel from one image to the next will result in a summation of the returns from a different group of scatterers. Therefore, the first step in interferogram formation is precise, subpixel-level

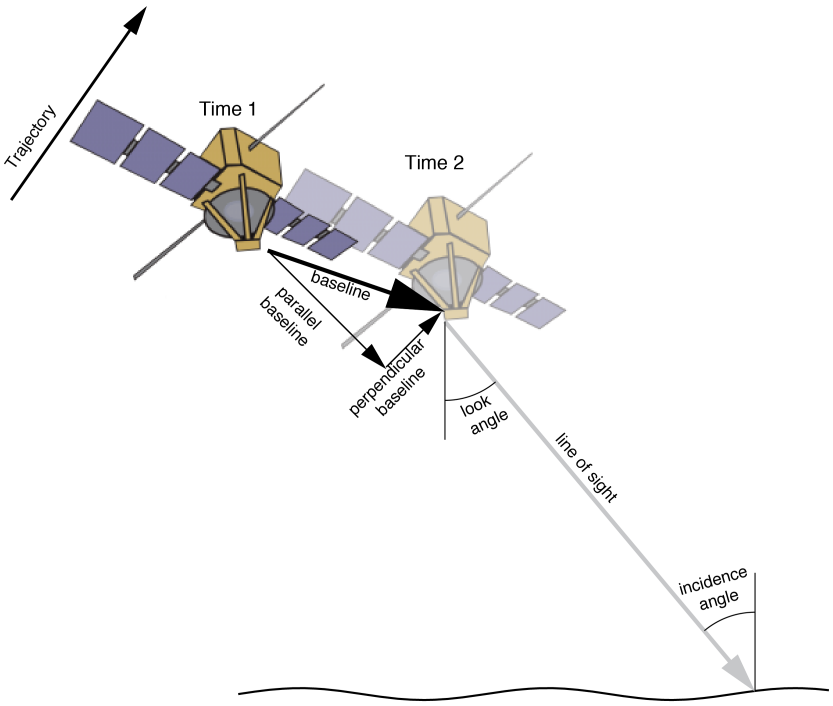


Figure 2.4: SAR acquisition geometry. Baseline is the distance between satellite positions at the two data acquisitions and is decomposed into parallel and perpendicular components. Look angle is the angle between the line-of-sight vector and the vertical at the antenna; incidence angle is between the line-of-sight vector and the vertical at the surface.

registration of the images to one another. Typically, with one image of the pair designated the master and the other the slave, the slave is warped and resampled to match the master.

Many image registration methods are available. Typically the image registration process has a coarse registration step followed by a fine registration step. The two steps may use the same method but with different parameters, or may use different methods. One commonly used method is to divide the two images into subimages and compute the normalized cross-correlation between corresponding subimages. If the peak cross-correlation value is at the center of the window, the two subimages are perfectly registered; if not, an offset is computed. The offsets for all such windows are used to calculate the coefficients of a polynomial which is then used in an affine transformation of the slave. Because subpixel precision is required, the fine coregistration step typically includes an oversampling step to accomplish transformations by fractions of a pixel.

2.3.3 Interferogram Formation

As previously stated, the complex return from a resolution cell is the coherent sum of the returns from all of the individual scatterers within that resolution cell. In reality, the phase of the return includes a contribution from the reflectivity of each scatterer and a contribution from the propagation delay due to path length. In interferometry, we generally assume that the spatial baseline between the two acquisition locations is small enough that the reflectivity is unchanged between the two geometries; if the baseline is too large, the interferogram will suffer from decorrelation, to be discussed in section 2.3.5. Therefore, we wish to isolate the phase contribution from the propagation delay in order to calculate the

path length, then difference the values in the two images to determine the change in path length.

This is accomplished by pixel-wise multiplication of the complex values of one image by the complex conjugate of the second image:

$$I(x,r) = M(x,r) \cdot S^*(x,r) = A_M A_S e^{j(\phi_M - \phi_S)} \tag{2.31}$$

where I is the interferogram, M is the master image, and S is the slave image. Consequently, the phase of the interferogram is $\phi_I = \phi_M - \phi_S$; it can be mapped to range as

$$\phi_I = \frac{4\pi}{\lambda}\left(r_M - r_S\right) = \frac{4\pi}{\lambda}\delta r \tag{2.32}$$

(equation A.27). The value of δr is a relative measure, as it is difficult to know a priori how many full wavelengths lie between the radar and the surface. Thus the value of ϕ_I is known only modulo 2π and is represented in a wrapped interferogram where $-\pi < \phi_I \leq \pi$ (figure 4.2). Incorporating terms related to the acquisition geometry: $\phi_I = -\frac{4\pi}{\lambda}(\hat{l} \cdot \vec{B})$ where $\hat{l}$ is the unit vector in the LOS direction and $\vec{B}$ is the vector along the baseline from the master radar location to the slave radar location.

At this point, the interferometric phase contains a contribution from the phase variation due to the shape of the surface. This phase is typically removed in two steps: flattening subtracts the phase calculated from a reference ellipsoid, e.g. WGS84 (figure 2.5a). The interferometric phase after this step is

$$\phi_{flat} = -\frac{4\pi}{\lambda}\left(\hat{l} \cdot \vec{B} - \hat{l_0} \cdot \vec{B}\right) \tag{2.33}$$

where $\hat{l_0}$ is the LOS unit vector to a point at a reference elevation. Now, the interferometric phase contains a contribution from topography (figure 2.5b). One can use this information to

construct a digital elevation model (DEM). However, when the goal is deformation mapping, this phase must be subtracted as well; this is called differential InSAR (DInSAR). Typically, a DEM from another source is used to predict the phase contribution from topography:

$$\phi_{flat} \approx -\frac{4\pi}{\lambda} B_{\perp} \frac{\delta h}{r_0 \sin\theta_0} \tag{2.34}$$

where δh is the topographical relief, $B_{\perp}$ is the perpendicular baseline, r_0 is the slant range to the reference elevation point, and θ_0 is the look angle to the reference elevation point. It is clear from this equation that the perpendicular baseline is a crucial variable in determining the quality of the deformation map. This is often expressed as the altitude of ambiguity, or the elevation change required to produce one full cycle of phase change, or fringe:

$$h_a = \frac{\partial h}{\partial \phi} = \frac{\lambda r_0 \sin\theta_0}{4\pi B_{\perp}} \tag{2.35}$$

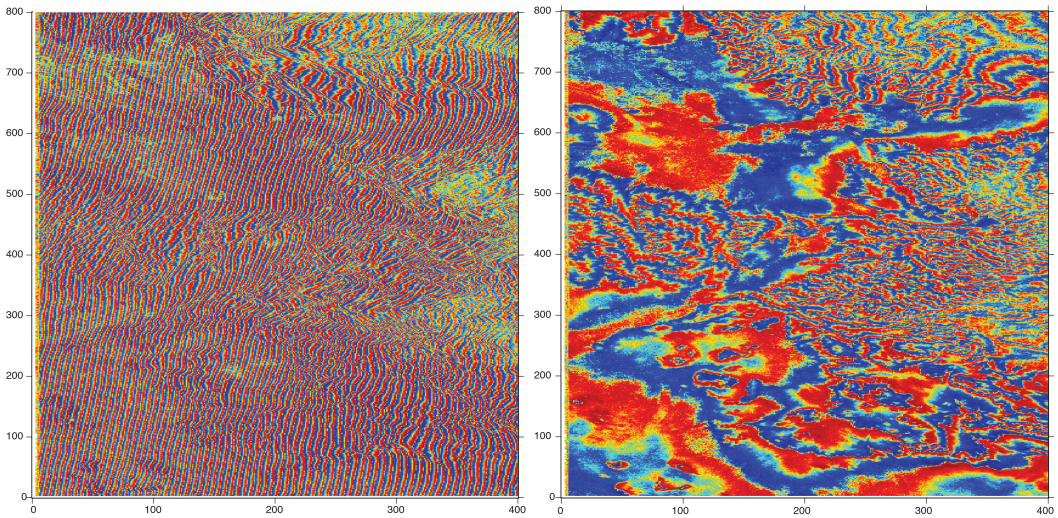
Thus a large perpendicular baseline creates many fringes for a small elevation change. Finally, this leaves the phase contribution due to surface displacement (figure 2.5c):

$$\phi_{disp} = \frac{4\pi}{\lambda} \delta r_{disp} \tag{2.36}$$

where r_{disp} is the surface displacement in the LOS direction.

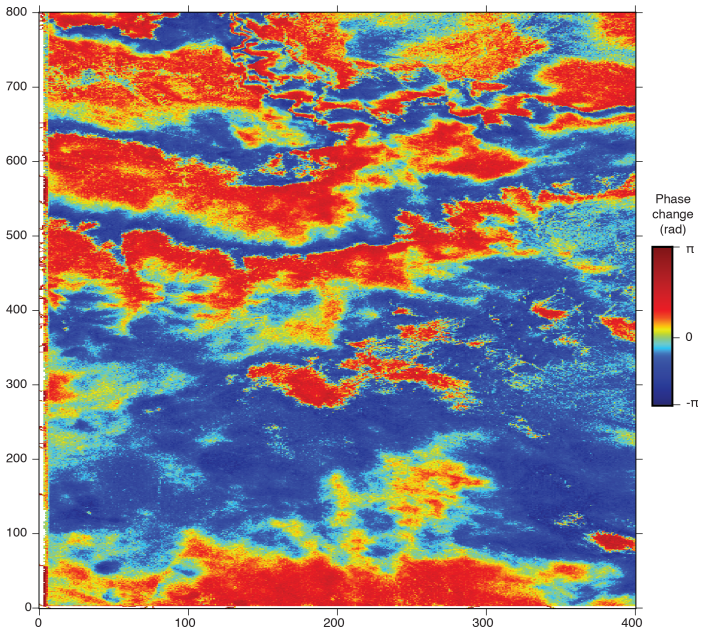
2.3.4 Phase Unwrapping

The wrapped interferogram can be interpreted as a contour map of range change where each fringe represents a range change of $\pi/2$ meters. Geophysical interpretation is facilitated by calculating the absolute values of range change, so the next step in the InSAR process is phase unwrapping, where the modulo 2π phase values are converted to a



(a) Before flattening.

(b) After flattening.



(c) After removing topographic phase.

Figure 2.5: Steps in the formation of the interferogram. After multiplying the master by the complex conjugate of the slave, the phase due to the shape of the surface must be subtracted.

continuous scale of phase values. Phase unwrapping is based on phase gradients between neighboring pixels so it cannot recover absolute range change values in the absence of other information; it merely resolves the ambiguity of the wrapped phase. In theory, an interferogram can be unwrapped by simply integrating the phase gradients between adjacent pixels throughout the interferogram. However, any pixel-to-pixel gradient greater than π radians will cause aliasing and an error that propagates through large regions of the interferogram. Several unwrapping algorithms have been developed to deal with this problem; only the two most commonly used are described here.

The branch cut algorithm is based on the assumption of path-independent integration of phase values in adjacent pixels. In theory, integration around a closed path should be path-independent and always yield a value of zero. However, phase gradients exceeding π radians lead to aliasing and non-zero integrals; the signs of these so-called residues are path-dependent (figure 2.6). Therefore, integration around a single residue would yield different results depending on the path; this is an undesirable situation. The branch-cut algorithm handles this problem by linking residues until a neutral tree is formed, with equal numbers of positive and negative residues; this tree is then a barrier across which integration cannot proceed [24]. One disadvantage of this technique is that noisy areas of the interferogram often contain many residues and barriers can form closed loops, rendering the data in these areas inaccessible to any integration path, resulting in an unwrapped interferogram containing areas with null values which requires manual intervention to form bridges between these areas.

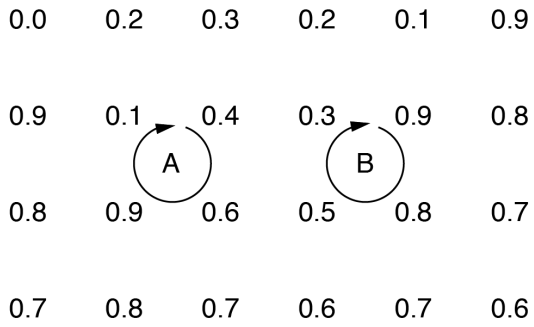


Figure 2.6: An example of residues in the phase unwrapping process. The numbers represent fractions of a cycle. In path A, the gradient between 0.1 and 0.9 will be aliased, resulting in a positive residue. If the integration path is reversed, the residue will be negative. In path B, the gradient between 0.3 and 0.9 will be similarly aliased, resulting in a negative residue; the reverse path would result in a positive residue. Modified from [25].

The second widely used unwrapping algorithm is based on network flow theory. A network flow model contains nodes and arcs (figure 2.7); in phase unwrapping implementations, nodes represent residues as in the branch cut algorithm. It is then a matter of linking these nodes with suitable arcs to form optimal integration paths [25]. Because network flow theory is used in many domains, many algorithms exist for optimizing the solutions to these problems. For example, in the minimum cost flow (MCF) problem, a cost or weight is assigned to each possible arc, then an iterative solver finds the combination of arcs that links all the nodes while minimizing the total cost of the path. One algorithm for solving this problem is the minimum spanning tree (MST) algorithm. A spanning tree is a directed path linking all nodes, where each arc is the shortest possible path linking those two nodes.

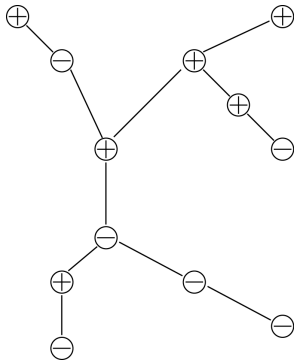


Figure 2.7: An example of the network model of phase unwrapping. Each node represents a residue. The arcs connecting the nodes represent the optimal path that links all nodes resulting in a total value of zero. After [25].

Because spanning trees are not permitted to close in on themselves, these algorithms have the advantage of providing a solution for the entire interferogram.

2.3.5 Correlation

Recalling that the return signal for each pixel is the coherent sum of reflections from many individual scatterers, it is easy to imagine many possible ways that sum may change between acquisition of two images. Mathematically, the change is quantified over some set of pixels as

$$\gamma = \frac{|M \cdot S^*|}{\sqrt{|M|^2 |S|^2}} \tag{2.37}$$

where $0 \leq |\gamma| \leq 1$ is called coherence; a coherence of 1 indicates perfect correlation between the two images. Clearly, low coherence between two pixels renders any geophysical interpretation of interferometric phase unreliable. The causes of decorrelation can be represented

as:

$$\gamma = \gamma_{noise} \cdot \gamma_{spatial} \cdot \gamma_{temporal} \tag{2.38}$$

The first cause of decorrelation is noise, particularly thermal noise in the receiver. Typically, the SNR of SAR systems is high enough that decorrelation due to thermal noise is minimal.

The second cause, spatial or geometric decorrelation, is due to the viewing geometry of the two acquisitions. Because each scatterer within the ground resolution cell is at a slightly different range and orientation when viewed from two different locations, the change in viewing geometry causes each scatterer to contribute a slightly different phase to the coherent sum in the pixel. If the baseline is too large, the pixel will decorrelate. One can calculate a critical baseline beyond which the pixel will decorrelate:

$$B_{c_\perp} = \frac{\lambda R}{2 \cdot \Delta y \cdot \tan \delta} \tag{2.39}$$

where y is the range extent of the ground resolution cell and δ is the nominal grazing angle (figure A.9). If we then define spatial coherence as

$$|\gamma_{spatial}| = 1 - \frac{B_\perp}{B_{c_\perp}} \tag{2.40}$$

it is clear that we require a perpendicular baseline much smaller than the critical baseline in order to avoid decorrelation.

Finally, temporal decorrelation is a function of temporal baseline, the time interval between image acquisitions. It is due to a physical change in the scatterers between the acquisitions. For example, vegetated areas may change due to growth or loss of foliage, or agricultural fields may be plowed or irrigated. In figure 2.8, the area in the magenta box was burned by a wildfire in the interval between image acquisitions; the burned area can

be clearly seen as an irregular area of low coherence. Of particular interest when using InSAR to study coseismic surface deformation is decorrelation due to damage caused by the earthquake; collapse of buildings, triggered landslides, liquefaction, and surface rupture will produce decorrelation.

2.3.6 Georeferencing

The final step in the InSAR processing chain is georeferencing. Until this point, the interferogram has been in radar coordinates, where the horizontal axis corresponds to range and the vertical axis corresponds to azimuth; the top left pixel is the first detected pixel. SAR satellites are in near-polar, sun-synchronous, low Earth orbits; they can image a surface location on both ascending passes, where the satellite crosses the equator from south to north, and on descending passes, crossing the equator from north to south [26]. Thus interferograms displayed in radar coordinates will appear to be "flipped" vertically for ascending passes and horizontally for descending passes.

It is not only more convenient to display an interferogram in a standard cartographic coordinate system, but it allows for mosaicking multiple interferograms to form one image. This is accomplished by first transforming the radar coordinates to Cartesian coordinates, which requires knowledge of the satellite's orbital parameters, then projecting to a standard map projection, e.g., a Mercator projection. Finally the image is resampled to a uniform grid pattern (figure 2.9).

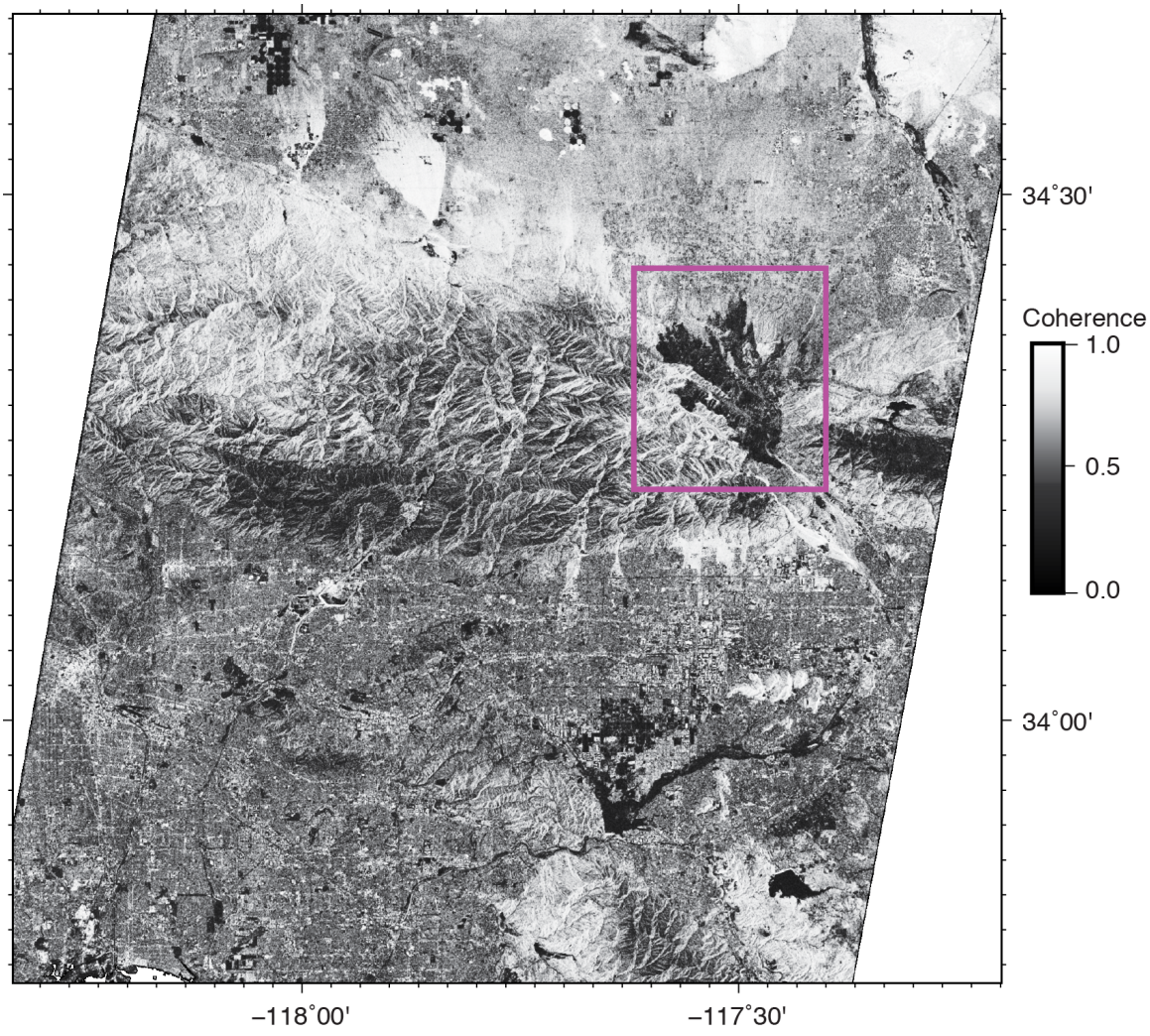
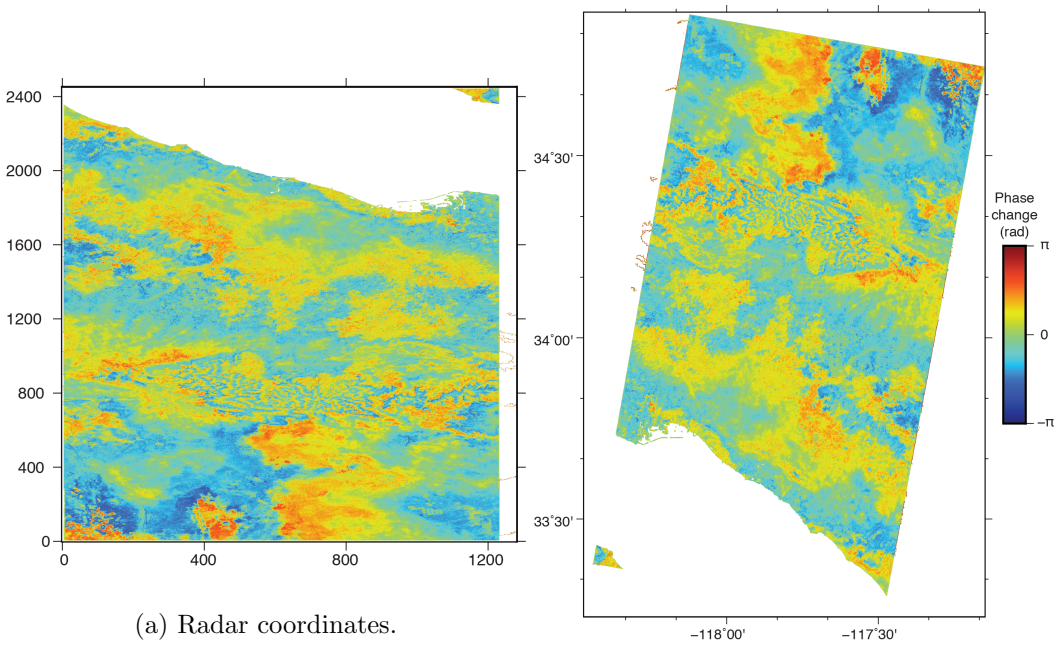


Figure 2.8: Coherence image from an interferogram in the Los Angeles, California region spanning 24 days. The magenta box outlines the irregular area burned by a wildfire, evident by its low coherence. In comparison, buildings in the urbanized area show high coherence.



(a) Radar coordinates.

(b) Georeferenced and resampled.

Figure 2.9: The final step in interferogram formation is georeferencing, where the interferogram is converted from radar coordinates (range, azimuth) to geographic coordinates (longitude, latitude). In this descending interferogram, the image appears to be flipped horizontally.

2.4 Sentinel-1

The data used for this thesis were acquired by the Sentinel-1A satellite belonging to the European Space Agency (ESA). In this section I will give a brief history of the Sentinel-1 mission and describe the unique characteristics of the mission. I will then describe the processing I employed for this study, with particular emphasis on the steps that differ from typical SAR processing as described in earlier sections.

2.4.1 Mission

Sentinel-1 is part of the ESA Global Monitoring for Environment and Security initiative [27]. There are a total of six Sentinel missions, either presently operational or in development, each with its own observation capabilities and objectives: land, sea, atmosphere, and air quality observation using radar, visible, infrared, and ultraviolet wavelengths [28]. The Sentinel-1 mission is a two-satellite constellation (Sentinel-1A and Sentinel-1B) using C-band (table A.1) radar to make land and ocean observations. Both satellites are in near-polar, sun-synchronous, low Earth orbit with their orbits offset by 180°. Each satellite has a 12-day exact repeat interval, so the constellation has a maximum return time, including both ascending and descending passes, of three days. This short return interval allows greater flexibility in interferometric pair formation while minimizing the risk of temporal decorrelation. Sentinel-1A was launched on April 3, 2014; Sentinel-1B was launched on April 25, 2016.

TOPSAR Mode

The Sentinel-1 instruments can operate in one of four imaging modes; the default mode for imaging over land is Interferometric Wide Swath (IW) mode, also known as TOPSAR. As described in section 2.2.1, this is a variation on ScanSAR that employs electronic beam scanning in azimuth to maintain a wide swath while eliminating the scalloping effect [29]. The antenna cycles through three elevation settings, corresponding to three sub-swaths in the final image. The block of radar returns acquired in each cycle through a sub-swath is called a burst, and the antenna is steered in azimuth over the course of each burst, with the squint angle varying by no more than $\pm 0.7^\circ$ (figure 2.10). There is a small zone of overlap at the beginning and end of each burst, so that the overlap area is imaged from both a forward-squinted and backward-squinted geometry.

The azimuth beam steering allows every target to be illuminated by the mainbeam, but the effective footprint is smaller than the equivalent stripmap mode footprint. Thus the azimuth resolution of TOPSAR mode is

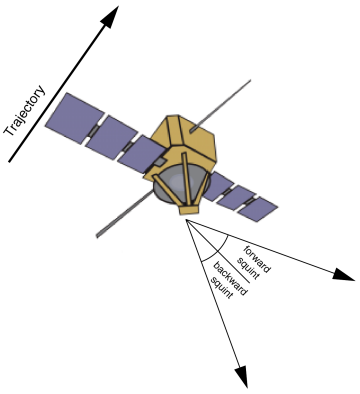
$$\Delta A_{TOPSAR} \approx \frac{\alpha D}{2} \tag{2.41}$$

where

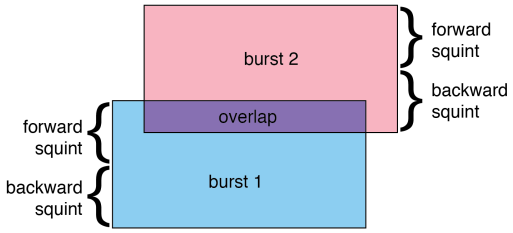
$$\alpha = 1 + \frac{R_0 |k_\psi|}{v} \tag{2.42}$$

and k_ψ is the antenna azimuth steering rate. The azimuth steering also introduces a Doppler centroid rate,

$$k_a = -\frac{2v}{\lambda} k_\psi \tag{2.43}$$



(a) TOPSAR mode squint.



(b) TOPSAR burst geometry.

Figure 2.10: TOPSAR mode acquisition geometry. During acquisition of each burst, electronic antenna steering in azimuth produces a variable squint angle, so that half of the burst is acquired with a backward squint while half is acquired with a forward squint. The area of overlap between two bursts is thus acquired with both forward and backward squints.

This must be taken into account during the image formation process. The antenna steering introduces another complication in image formation: because the Doppler frequency is sampled at the pulse repetition frequency (PRF) (section A.2.2), a fixed antenna creates an unambiguous Doppler bandwidth of no more than $\pm PRF/2$. However, antenna steering, and the consequent variable squint angle, creates a Doppler bandwidth much larger than the PRF. Consequently, the data are folded, or aliased, in the frequency domain, an effect that must be corrected by the image formation processor.

The variable Doppler centroid also affects the interferogram formation process. Coregistration of the two images in the interferometric pair depends on knowledge of the satellite's orbital parameters. There is always some small error in these parameters but in the stripmap case, these errors would introduce a constant misregistration, thus a constant

phase offset, for the entire interferogram, and would be inconsequential. In the TOPSAR case, however, the same constant misregistration would produce a linear phase ramp across the burst. Thus adjacent bursts would have phase discontinuities at their boundaries [30]. Therefore, an additional step is required in the coregistration process where the spectral separation between the forward- and backward-squinted portions of the overlap area is calculated. Any phase difference beyond that predicted by the spectral separation represents a misregistration error that must be corrected; this technique is called enhanced spectral diversity (ESD).

A further consequence of TOPSAR mode is one that can be exploited in the geophysical interpretation of TOPSAR interferograms. Any target that is viewed from two different squint angles will have two different Doppler spectra. In terms of mission design, this effect must be accounted for so that each area of the surface is imaged from the same geometry. On the other hand, any Doppler shift, and therefore any phase change, in the properly coregistered burst overlap areas beyond that expected from the variable squint angles, must be due to change in location along the azimuth axis [31]. This allows the ESD technique to be used to recover along-track deformation signals which stripmap interferograms are largely insensitive to.

2.4.2 Data

ESA has committed to free and rapid access to Sentinel-1 data in various formats. For this study, I downloaded the level 1 single look complex (SLC) images [32]. Each file contains a single slice divided into three sub-swaths, each of which contains nine bursts. Each burst is processed by ESA as a separate SLC image, and the bursts are combined into

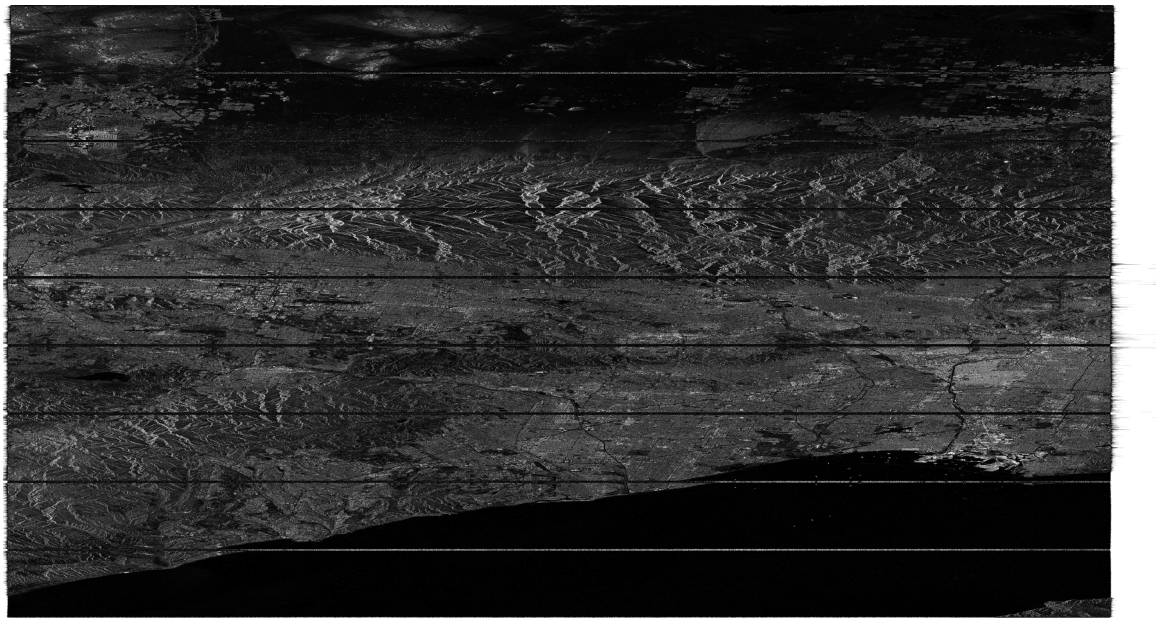


Figure 2.11: Nine single look complex images (bursts) separated by blank lines. There is a small zone of overlap between adjacent bursts.

a slice with blank lines between each burst (figure 2.11). I designated the image acquired prior to the mainshock as the master image, and the image acquired after the mainshock as the slave.

2.4.3 Processing

Image Processing

I processed the SAR images using the Sentinel-1 Toolbox software package [33]. I coregistered each of the three sub-swath images of the slave to the master using both geometric and ESD-based coregistration. I formed an interferogram for each pair of sub-swath images, then joined the bursts by removing the intervening blank lines. I merged

the three interferograms into one image, subtracted the topographic phase, and applied a Goldstein phase filter. I then reduced the file size by spatial averaging. Finally, I used the minimum spanning tree algorithm of Snaphu [34] to unwrap the interferogram, then georeferenced the unwrapped interferogram.

After processing adjacent slices from the same satellite track and with the same acquisition dates, I mosaicked them into a single interferogram. I extracted the overlapping area between two adjacent slices from both images and resampled one to match the pixel spacing of the other. I computed the phase difference between the two overlapping areas by pixel-wise subtraction and division of the difference by 2π to find the modal integer multiple of 2π. I then added this 2π multiple to the entire extent of one of the images and mosaicked the two interferograms to form a single image. I repeated this as many times as required to encompass the entire displacement field in one interferogram. Finally, because it is reasonable to assume zero deformation in the far field, I added an arbitrary value to the entire interferogram to give a zero value in the far field.

Line-of-Sight Vectors

I calculated LOS vectors using the formula of [35]. The LOS vector for each observation point is:

$$LOS = [\sin\theta\cos\alpha\prime, \sin\theta\sin\alpha\prime, \cos\theta] \tag{2.44}$$

where θ is the incidence angle and $\alpha\prime$ is the squint angle. The incidence angle is provided as a raster for each sub-swath, while the squint angle is calculated from the satellite heading, azimuth steering rate, and PRF as provided in the metadata accompanying each file; because these values are different for each sub-swath, I calculated the LOS vectors for each

sub-swath individually. While each pixel of the interferogram has a different LOS vector, data reduction (section 2.4.4) of the interferogram means that the LOS vector must be approximated for a wide area of the interferogram. Therefore, I used the value of the LOS vector at the center of each area to represent the entire area.

2.4.4 Data Reduction

A single interferogram may contain more than 10^7 pixels, making inversion of the data impractical. Therefore, interferograms are downsampled to reduce the number of data points brought into the inversion. This reduction must be balanced with the need to maintain enough resolution to permit meaningful interpretation of the inversion results. Several downsampling algorithms have been proposed; two will be discussed here.

Regardless of the downsampling algorithm used, each point resulting from the downsampling process requires a LOS vector. I found that vector by importing the geographic coordinates of the centers of each area into Sentinel-1 Toolbox. I sampled each raster of LOS vectors (x, y, and z components for each sub-swath) at these points and exported the values. In some cases, the center coordinates were in areas of overlap between two or more sub-swaths, such that two or more LOS vectors applied to each point. In these cases, I averaged the values of the LOS vectors.

Quadtree Decomposition

The most commonly used downsampling algorithm is quadtree decomposition. This is similar to that used in many image compression algorithms. In this algorithm, the entire image is divided into four quadrants. If the variance of the values within a quadrant

exceeds a predetermined threshold, it is subdivided into quadrants. The process is repeated until either a subdivided area has a variance below the threshold, or the area reaches a predetermined minimum size. The result is a downsampled image that contains few large areas where the original data were spatially invariant, and many small areas where the data were highly variable. This preserves the variability associated with high deformation gradients. The disadvantage of this method is that is tends to preserve fine sampling in noisy areas in the far field.

For this study, I implemented quadtree decomposition by exporting the interferogram in NetCDF format with a geographic coordinate system. I projected it to Transverse Mercator with a projected coordinate system to maintain consistency with other components of the model. I imported the interferogram into Matlab [36] where I used a custom script to perform the quadtree decomposition. I then calculated the coordinates of the corners and centers of each region in both projected and geographic coordinate systems.

Resolution-based Resampling

Quadtree decomposition uses the actual interferogram data to produce a downsampling scheme. In order to achieve the goal of rapid characterization, however, it would be preferable to have a downsampling scheme determined prior to an earthquake so that inversion can proceed more quickly. A plausible alternative is the resolution-based resampling algorithm of [37]. In this algorithm, a precomputed matrix of Green's functions, $\mathbf{G}$, is used to compute a data resolution matrix, $\mathbf{N} = \mathbf{G}\mathbf{G}^{-g}$ where $\mathbf{G}^{-g}$ is the generalized inverse of $\mathbf{G}$: $\mathbf{G}^{-g} = (\mathbf{G}^{T}\mathbf{G})^{-1}\mathbf{G}^{T}$ [38]. The first set of Green's functions is based on a coarse sampling of the original interferogram. Whenever any area has a value in the diagonal of

N that exceeds a predetermined threshold, that area is subdivided into four new areas and the process is repeated.

In this study, I implemented resolution-based resampling by determining the maximum extent of an interferogram likely to encompass the entire deformation field. I initially divided this area into an 8×8 grid and calculated the coordinates for the center of each sample by comparing to an actual interferogram covering the same area. I used these coordinates to determine LOS vectors as in section 2.4.3 and in a forward model to create the matrix of Green's functions (Chapter 3). I augmented this matrix with a smoothing matrix, where the smoothing parameter was the same as that used in the preferred inversion of the quadtree decomposition (Chapter 3). I computed the data resolution matrix as described above and compared its values to a predetermined threshold. I repeatedly subdivided any areas exceeding the threshold into quadrants and repeated the process until either every area had a value on the diagonal of **N** below the threshold or the areas reached a predetermined minimum size.

Chapter 3

Fault Modeling

3.1 Introduction

In order to accomplish the goal of rapid characterization of an earthquake, an accurate model of the ruptured fault must be available prior to the event. Fortunately, models are available for most of the world's subduction zones as the Slab 1.0 set of models. In this section, I will describe the Slab 1.0 geometries and the steps required to ingest them into my models. I will also describe the discretization method I used and the details of the modeling process.

3.2 Slab 1.0

3.2.1 Background

The Slab 1.0 models provide fault geometries for most of the world's subduction zones. They were prepared by the USGS [39] and can be downloaded from their website

[40]. The models assimilate data from a variety of sources. Seismogenic zone geometries are constrained using seismicity catalogs filtered to include well-located earthquakes with thrust focal mechanisms. The geometries of the shallow portions of the slab interfaces, which are much less seismically active, were determined using active-source seismic studies, bathymetry, and trench sediment thickness data. For each subduction zone, two-dimensional profiles were constructed every 10 km along-strike. These profiles were then smoothed and averaged to create three-dimensional surfaces. From these surfaces, a grid of points with longitude, latitude, and depth coordinates was interpolated to a regular 0.5° spacing.

This comprehensive approach is more robust than previous attempts to model subduction zone geometries because the incorporation of active-source seismic, bathymetry, and sediment thickness data allow better modeling of the shallow interface, where earthquakes represent a greater seismic hazard. In contrast, the model's reliance on seismic catalogs results in relatively poor resolution of the deep interface. In fact, a complete lack of deep seismicity in some subduction zone regions, e.g., southern Chile, means that the model is undefined in these regions. However, deep-focus earthquakes are not a significant source of seismic hazard, so this limitation is inconsequential for the purposes of my study.

Examination of the Slab 1.0 models indicates that the seismogenic zone of most subduction zone interfaces is wider than had previously been inferred, because previous models relied on the dip of the focal plane as determined by the global centroid moment tensor (GCMT) method, which tends to overestimate dip in subduction zones [39]. A shallower dipping interface would necessarily lead to a wider seismogenic zone, and shallow slab

dip is a necessary feature for the occurrence of giant subduction zone thrust earthquakes [41]. The robustness of the Slab 1.0 dip values is confirmed by comparing the maximum moment magnitude observed in various subduction zones to their calculated seismogenic fault widths. As discussed in Chapter 1, wider seismogenic zones should produce higher magnitude earthquakes, and this correlation is higher with Slab 1.0 models than with previous subduction zone models. This finding has important implications for seismic hazard estimation. Since subduction zone seismic cycles are longer than in other tectonic settings, the record of subduction megathrust earthquakes is incomplete, and hazard has likely been underestimated in many regions, as was the case in Japan in 2011 [42]. Better models of the seismogenic zones of subduction interfaces would allow better calculation of seismic hazard in these regions.

3.2.2 Data

I downloaded the Slab 1.0 depth grids from the USGS website, where they are provided in NetCDF format. Each point consists of longitude, latitude, and depth coordinates. Longitude and latitude are in the WGS84 geographic coordinate system; depth is in kilometers relative to sea level, which is assigned a depth of 0 km.

3.2.3 Processing

I will ultimately incorporate the fault models into a three-dimensional model of the surface, so the coordinate systems of each component must be consistent. Therefore, I converted the fault model to a transverse Mercator projection using a suitable central

meridian so that all model components would be in the same map projection, and used a projected coordinate system in meters using the Generic Mapping Tools (GMT) [43].

The spatial extents of the Slab 1.0 models greatly exceed those that would be expected to rupture in a single earthquake. For example, the model for South America extends 50° in latitude. Furthermore, the computational load required to process such extensive models is beyond what could be considered reasonable for a desktop computer. Moreover, the resolution of the Slab 1.0 models is finer than is required by the fault discretization method I am using in this thesis. Therefore, I cut the models to a spatial extent which encompassed the rupture length inferred by the USGS Finite Fault Model [44] for a relevant earthquake in the model region. I then downsampled them and imported the remaining points into Cubit [45] for meshing.

3.3 Triangular Dislocation Elements

3.3.1 Background

Elastic Dislocation Theory

Elastic dislocation theory provides the foundation for modeling many engineering and geophysical problems. In brief, elastic dislocation theory describes the displacements, stresses, and strains produced by a dislocation on a finite surface within an elastic medium [46]. In this sense, a dislocation is a displacement discontinuity across the dislocation surface; in geophysical terms, a dislocation represents slip on a fault. The goal, then, is to develop a mathematical representation of the surface manifestations of this dislocation. Surface displacement may be composed of any combination of rigid body translation, strain,

and rotation. Several assumptions and approximations make the development of these mathematical descriptions relatively straightforward. For example, it is commonly assumed that the medium in which the dislocation is embedded is linearly elastic, homogeneous, and isotropic; this produces symmetry in the strain and stress tensors.

The most common method of deriving the full-space solutions is by imposing conditions of equilibrium and conservation of mass and linear momentum [47]. These solutions are extended to the half-space using the method of images, in which a fictitious dislocation is placed above the surface in order to satisfy the boundary condition that the surface be traction-free; in other words, no stresses or strains are imposed on the atmosphere by the solid surface. In both cases, the solutions require computation of Green's functions, which relate the displacements at an observation point to a force at a source point. The Green's functions for a half-space were derived decades ago. It is then a straightforward, albeit tedious, task to incorporate them into the proper equations. While several authors have tackled this problem, the first publication to provide an error-free enumeration of these equations for the point source and the finite rectangular source was by Okada [48].

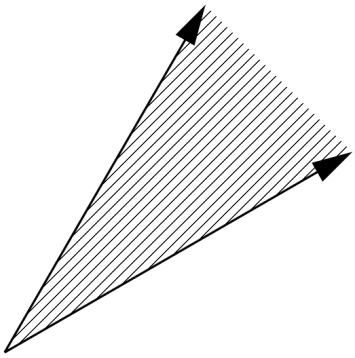
The solution for surface displacements due to an arbitrary finite source could potentially be determined by integrating the solution for a point source over the extent of the fault. However, these computations would be exceedingly difficult in the general case. A common approach is thus to discretize the fault surface into elements of limited spatial extent and simple geometry. Each element, or patch, is subjected to uniform slip, simplifying the computation of surface displacement from each patch. The total displacement at any observation point is then the sum of the contributions from every patch. Because the

solutions for a rectangular source are easily included in computer programs and scripts, this approach is the most widely used in earthquake fault modeling studies [49].

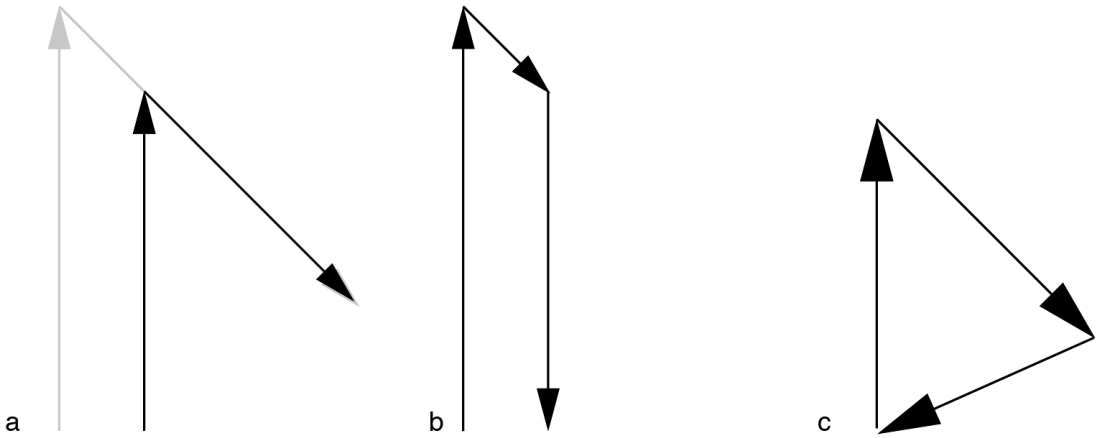
Angular and Triangular Dislocations

While discretizing planar faults using rectangular elements is easily accomplished, it is impossible to accomplish gap-free tessellation of a curved surface using rectangles. Therefore, using rectangular dislocation elements to discretize a fault with substantial variation in strike or dip results in unphysical gaps and stress singularities at these gaps [50]. In contrast, gap-free tessellation of curved surfaces is possible using triangular elements [51]. The problem, then, is to determine the solutions for surface displacements due to slip on a triangular fault patch.

The approach begins with the angular dislocation in a full space [52]. An angular dislocation is composed of two semi-infinite dislocation lines meeting at a point and bounding a semi-infinite surface with a constant dislocation (figure 3.1a). As before, the full-space solutions were extended to the half space [53] by adding a fictitious mirror dislocation above the surface to cancel the surface tractions generated by the actual dislocation. A dislocation segment can then be constructed by superposing two coplanar angular dislocations, where the legs of each angular dislocation meet at equal angles, and the two slip vectors have opposite signs [49]; thus the coincident semi-infinite legs cancel one another and the remaining dislocation corresponds to a semi-infinite plane along one side of the triangular element (figure 3.1b). Three such dislocation segments can be superposed in such a way that



(a) An angular dislocation. The shaded area represents the semi-infinite surface of dislocation.



(b) Two superposed angular dislocations of opposite sign (a) form one side of a triangular dislocation element and the semi-infinite plane below it. Three such superposed dislocations form a triangular dislocation (c).

Figure 3.1: Angular and triangular dislocations.

the adjacent semi-infinite dislocations at the triangle vertices cancel one another, leaving only the dislocation due to the triangular surface.

The solutions described above all require transformation of Cartesian coordinates to a coordinate system relative to the dislocations in which the components of the slip vector are defined. In addition, the orientation and superposition of the six angular dislocations must be carried out precisely so that all the relevant components are properly cancelled. I therefore chose to use the Matlab code accompanying Meade [50], whose algorithms are based on Comninou and Dundurs [53], rather than develop my own. It should be noted, however, that these algorithms are not without problems. In fact, there are singularities in the displacement field along the dislocation lines, as would be expected, as well as directly above the vertices of the triangle [54]. While Nikkhoo and Walter outline a method for eliminating these singularities, I chose to use the code written by Meade and confirm by visual inspection of the forward model (section 3.4.1) that none of my observation points coincided with the locations of these singularities.

3.3.2 Meshing

I meshed the fault model in Cubit as follows. I represented each point as a vertex with x-, y-, and z-coordinates corresponding to the longitude, latitude, and depth, respectively, of each point in the depth grid. I drew spline curves along vertices with constant x- or y-coordinates producing two sets of perpendicular curves, which I used to create a surface. I meshed this surface with triangles of constant size, then exported the x-, y-, and z-coordinates of each triangle vertex, as well as the identification of the vertices comprising each triangle, to a text file. In addition, I created a surface that encompassed the projection

of the slab interface to an elevation of 0 m but extended a sufficient distance in the x and y directions to fully enclose all likely surface displacements. I meshed this surface with rectangles of constant size and exported the x-, y-, and z-coordinates of the nodes.

3.4 Modeling Surface Displacements

3.4.1 Forward Modeling

Prior to computing a forward model in Matlab, I validated my implementation as follows. Using the examples provided in [48], I decomposed the rectangular example fault surfaces into two triangles and used their coordinates in a call to Meade's algorithms for surface displacements [50]. As discussed in section 3.3.1, there are singularities in the displacement fields directly overlying the vertices of the triangular dislocation elements; the observation coordinates provided in Okada lie over these vertices. I therefore chose alternative observation points and found that the displacements agreed to within 10^{-6} of those calculated using Okada's equations for the same observation points. I also devised several fictitious faults of varying geometries and slip vectors to confirm that the displacements were as expected. The synthetic interferograms produced by two of these faults are shown in figures 1.2 and 1.3.

To compute the forward models, I used the coordinates of the triangle vertices exported from Cubit (section 3.3.2) to calculate surface displacements resulting from a uniform slip of 1 m and constant rake on all triangular dislocation elements. I plotted the displacements of a flat observation surface as separate east-west, north-south, and vertical

displacements to confirm that the sense and magnitude of the displacements matched what would be expected for the input fault geometry and rake.

I incorporated surface topography into the model by sampling a DEM [55] at the observation points determined by subsampling of an interferogram as described in section 2.4.4 and using these coordinates to calculate surface displacements. I then used the LOS vectors for each of these observation points, determined as described in section 2.4.3, to convert the east-west, north-south, and vertical displacements to LOS displacements. These values were assembled into the design matrix to be used in the inversion process described in the next section.

3.4.2 Inversion

The inversion process I used is derived from the general least squares solution of the linear inverse problem. If we represent the matrix of Green's functions as $\mathbf{G}$, the vector of estimated model parameters as $\mathbf{m}^{\text{est}}$, and the vector of observations as $\mathbf{d}$, then $\mathbf{G}\mathbf{m}^{\text{est}} = \mathbf{d}$, where $\mathbf{m}^{\text{est}}$ is the vector of model parameter estimates, and the least squares solution is $\mathbf{m}^{\text{est}} = [\mathbf{G}^{\text{T}}\mathbf{G}]^{-1}\mathbf{G}^{\text{T}}\mathbf{d}$ [56]. Because I assumed a fixed rake of 90°, I used a nonnegative least squares solver so as to avoid retrograde slip. In addition, I performed smoothing of the solution by computing a matrix, $\mathbf{S}$, containing the scale-dependent umbrella approximation of the Laplacian (∇^2) for each triangular element [51]. The weight, κ, given to the smoothing matrix was determined by trial and error to simultaneously minimize the L_2-norm of the misfit and the model roughness. The smoothing matrix was included in the inversion such that $\kappa\mathbf{S}\mathbf{m}^{\text{est}} = \mathbf{0}$.

Because the length of the fault model far exceeds that of the interferogram, I wished to disallow slip on the elements at the ends of the fault. A large amount of slip on fault elements at great distances can produce small displacements; however, this would be unrealistic. I therefore included a matrix with a value of 1000 at element (i, i), where i is the triangular element I wished to restrict from slipping. I included this matrix, $\mathbf{F}$, in the inversion such that $\mathbf{F}\mathbf{m}^{\text{est}} = \mathbf{0}$.

I also included in the inversion a solution for offset of the zero values and for orbital tilts. These tilts are the result of the inclusion of inaccurate satellite orbital parameters in the image registration process. The errors were reduced by solving for and subtracting a plane from the modeled displacements [57].

Finally, the data points were weighted by including an inverse variance-covariance matrix (VCM) in the inversion. The VCM was calculated as in [57]; each element of the VCM represents the covariance of each data point with every other data point. Thus its inverse, $\mathbf{E}^{-1}$, serves to give higher weight to data points which show less spatial correlation with others. The final form of the inversion is:

$$\begin{bmatrix} \mathbf{G}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{G} + \kappa^2\mathbf{S}^{\mathrm{T}}\mathbf{S} \\ \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{m}^{\text{est}} & a & b & c \end{bmatrix}^T = \begin{bmatrix} \mathbf{G}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{d} \\ \mathbf{0} \end{bmatrix} \tag{3.1}$$

where $\mathbf{G}$ has been augmented to [$\mathbf{G}$ 1 x y], $\mathbf{S}$ has been augmented to [$\mathbf{S}$ 0 0 0], and $\mathbf{F}$ has been augmented to [$\mathbf{F}$ 0 0 0]. The first row is a weighted damped least squares solution [56] where the inverse VCM provides the weight and the smoothing matrix provides the damping. The values of (x, y) are the projected coordinates of each observation. The parameters a, b, and c at the end of $\mathbf{m}^{\text{est}}$ are the solutions to the offset and orbital tilt. In

order to choose the best value of κ, I plotted the roughness, $||\mathbf{S}\mathbf{m}^{\mathrm{est}}||$, vs. misfit, $||\mathbf{e}^{\mathrm{T}}\mathbf{E}^{-1}\mathbf{e}||_2^2$, where $\mathbf{e} = \mathbf{G}\mathbf{m}^{\mathrm{est}} - \mathbf{d}$, from several iterations and chose a value that minimized both.

In addition to inverting a solution from a single interferogram, I jointly inverted ascending and descending interferograms. This inversion has the form:

$$\begin{bmatrix} \mathbf{G}_{\mathrm{a}}^{\mathrm{T}}\mathbf{E}_{\mathrm{a}}^{-1}\mathbf{G}_{\mathrm{a}} + \mathbf{G}_{\mathrm{d}}^{\mathrm{T}}\mathbf{E}_{\mathrm{d}}^{-1}\mathbf{G}_{\mathrm{d}} + \kappa^2\mathbf{S}^{\mathrm{T}}\mathbf{S} \\ \\ \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{m}^{\mathrm{est}} & a_{\mathrm{a}} & b_{\mathrm{a}} & c_{\mathrm{a}} & a_{\mathrm{d}} & b_{\mathrm{d}} & c_{\mathrm{d}} \end{bmatrix}^{\mathrm{T}}$$

$$= \begin{bmatrix} \mathbf{G}_{\mathrm{a}}^{\mathrm{T}}\mathbf{E}_{\mathrm{a}}^{-1}\mathbf{d}_{\mathrm{a}} + \mathbf{G}_{\mathrm{d}}^{\mathrm{T}}\mathbf{E}_{\mathrm{d}}^{-1}\mathbf{d}_{\mathrm{d}} \\ \\ \mathbf{0} \end{bmatrix} \tag{3.2}$$

where the subscripts a and d indicate the values for the ascending and descending interferograms, respectively, and $\mathbf{G}$, $\mathbf{S}$, and $\mathbf{F}$ were augmented as for the single inversion case.

Chapter 4

Illapel Earthquake

In this chapter I will describe the Illapel earthquake of September 16, 2015 and the seismotectonic setting in which it occurred. I will summarize the published models of the earthquake and then provide information about the data I used in this study as well as my results.

4.1 Introduction

The South American subduction zone spans more than 3500 km and is divided into two segments by the Chile Rise. North of the Chile Rise, the Nazca plate subducts beneath the South American plate at a rate of $6.8\,\mathrm{cm\,yr^{-1}}$; to the south, the Antarctic plate subducts at a rate of $1.8\,\mathrm{cm\,yr^{-1}}$ [58]. The Chilean part of the subduction zone has produced several of the largest earthquakes measured instrumentally and has a long historical record of great earthquakes. Prior to this event, the most recent great earthquakes occurred in 2010 and 2014. The M_w8.8 Maule earthquake of 2010 ruptured a 500 km long portion of

the subduction thrust, generating a tsunami with up to 29 m of run-up [59]. The M_w8.1 Iquique earthquake of 2014 ruptured approximately 150 km [60]. These two events, which occurred to the south and north, respectively, of the Illapel rupture, ruptured previously recognized seismic gaps [61, 62]. However, a substantial moment deficit remained along the subduction thrust after these events [62].

The M_w8.3 Illapel earthquake occurred on September 16, 2015 (figure 4.1). It generated a tsunami with up to 11 m of run-up near the source [63]. While as many as one million people were evacuated from coastal areas in anticipation of a tsunami [64], there were fewer than 20 fatalities. The 2015 rupture was approximately coincident with that of a 1943 M_w8.2 earthquake [63]. The hypocenter was located at $-31.5952°$, $-71.6728°$ and 22.4 km depth [65]. Barnhart, et. al. [66] infer a significant amount of afterslip occurring over an area that partially overlaps that of the coseismic slip.

4.2 InSAR

4.2.1 Sentinel-1A Data and Processing

I constructed interferometric pairs from Sentinel-1A SLC images acquired on August 26 and September 19, 2015, in the ascending direction, and on August 24 and September 17, 2015 in the descending direction (table 4.1). In each case, I designated the image predating the mainshock the master and that postdating the mainshock the slave. The postseismic interval includes 383 aftershocks with magnitudes ranging from 3.7 to 7.0 in

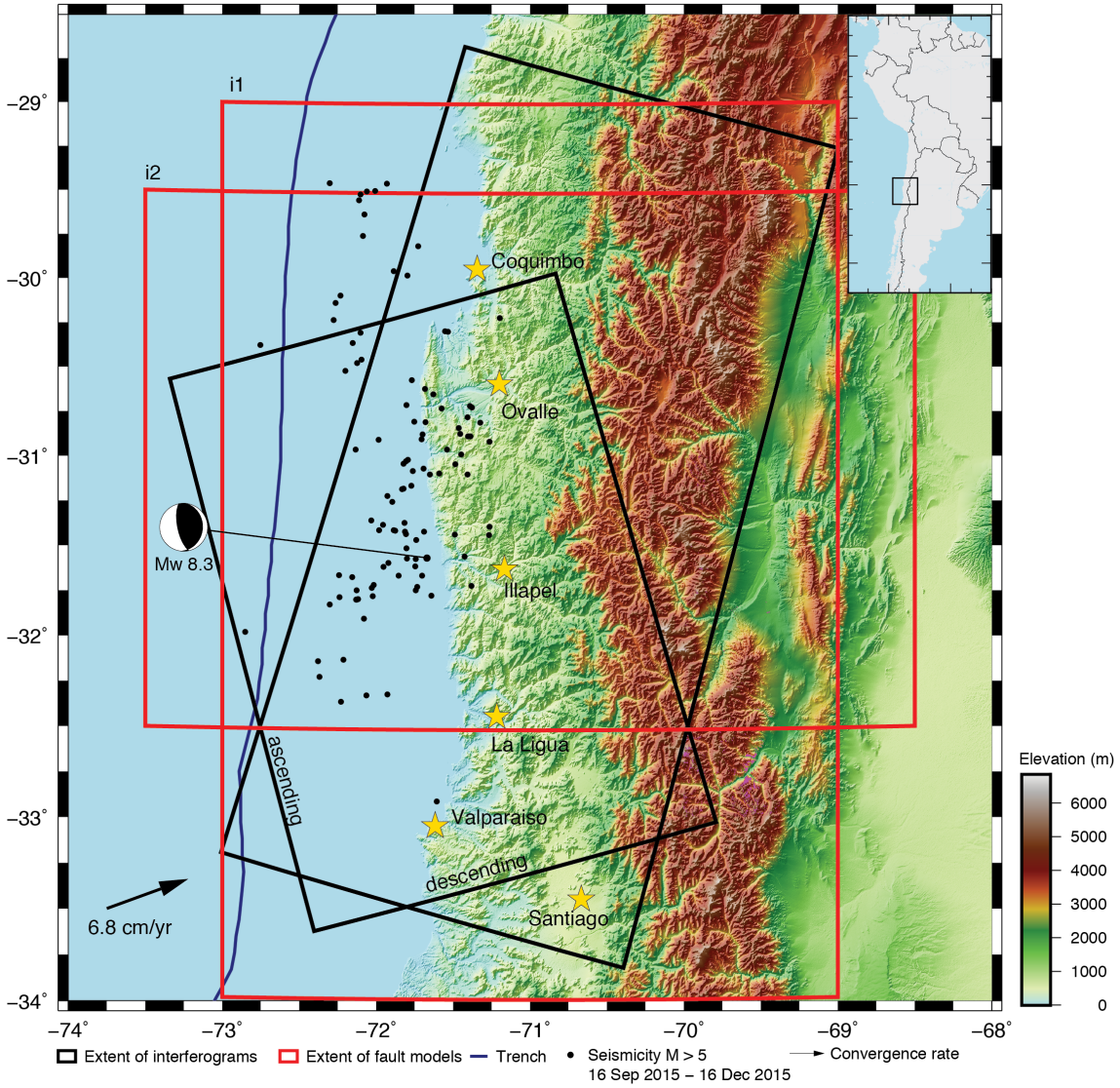


Figure 4.1: Shaded relief map of Illapel region of Chile showing the epicenter of the M_w8.3 earthquake, significant aftershocks in the three months after the mainshock, extents of the ascending and descending interferograms used in this study, and extents of the fault models used in this study.

Table 4.1: Sentinel-1A data used in this study.

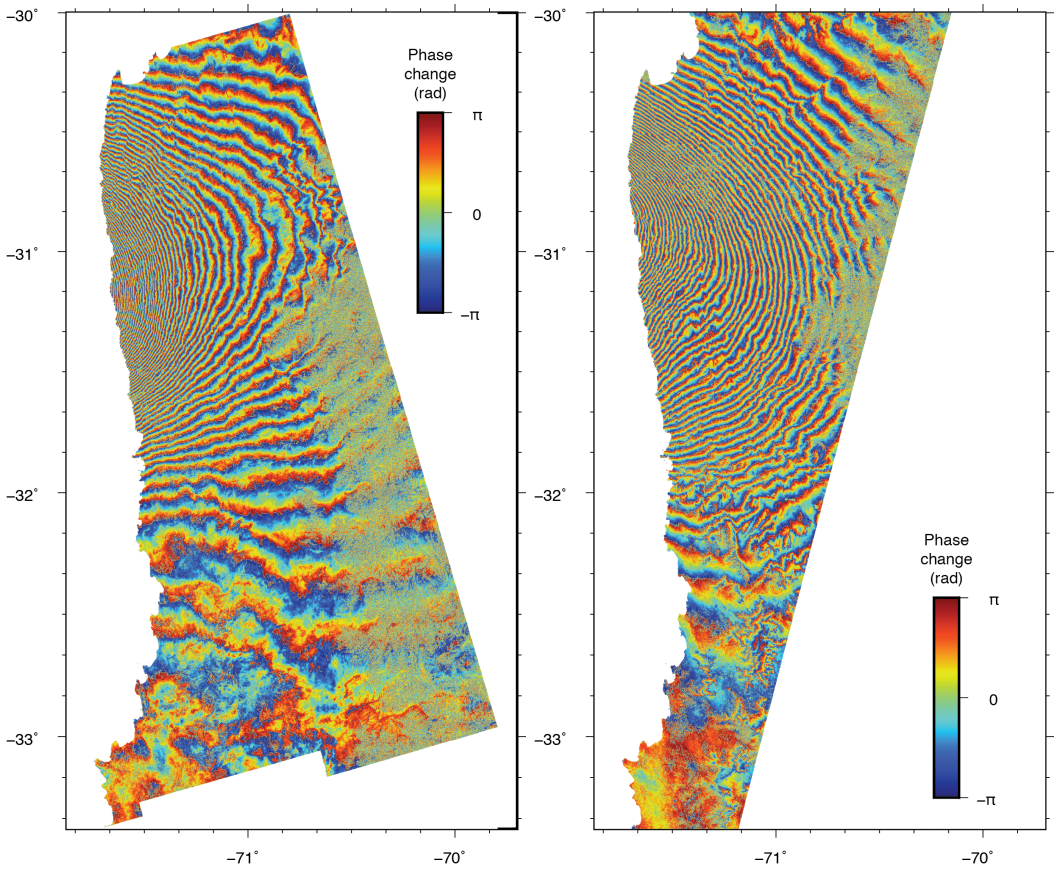
Acquisition Date				Pixels	
Master	Slave	Baseline (m)	Slices	Full	Reduced
		Ascending			
2015-08-26	2015-09-19	72.68	9 to 10	11 450 656	1091
		Descending			
2015-08-24	2015-09-17	114.89	12 to 14	17 891 412	1319

the ascending interferogram, and 29 aftershocks with magnitudes ranging from 4.5 to 7.0 in the descending interferogram [67].

I processed the SLC images and interferograms in Sentinel-1 Toolbox as described in section 2.4.3. The wrapped and unwrapped interferograms are shown in figures 4.2 and 4.3, respectively. To construct the ascending interferogram, I assembled two slices; for the descending interferogram, I assembled three. I performed quadtree decomposition of the unwrapped ascending interferogram with a variance threshold of 4π and a minimum size of 32 by 32 pixels; this reduced the size of the interferogram from 11 450 656 pixels to 1091 pixels (figure 4.4a). The descending interferogram was data reduced using the same parameters, reducing the image size from 17 891 412 pixels to 1319 pixels (figure 4.4b).

4.2.2 Results

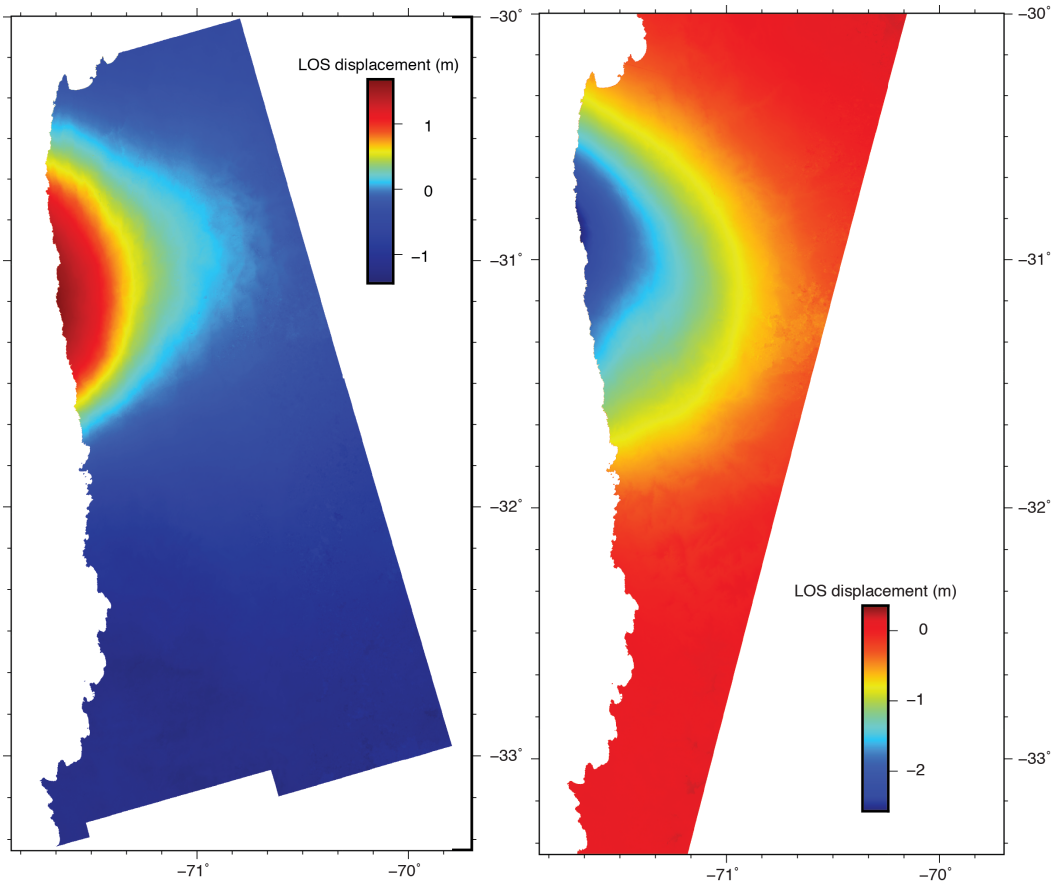
The wrapped interferograms from both the ascending and descending passes show a clear fringe pattern centered near the city of Illapel. As is expected for thrust earthquakes in a subduction zone, the deformation pattern is incomplete, in that there is no evidence of a change in the sense of displacement (see figure 1.3b). Converting the unwrapped



(a) Ascending.

(b) Descending.

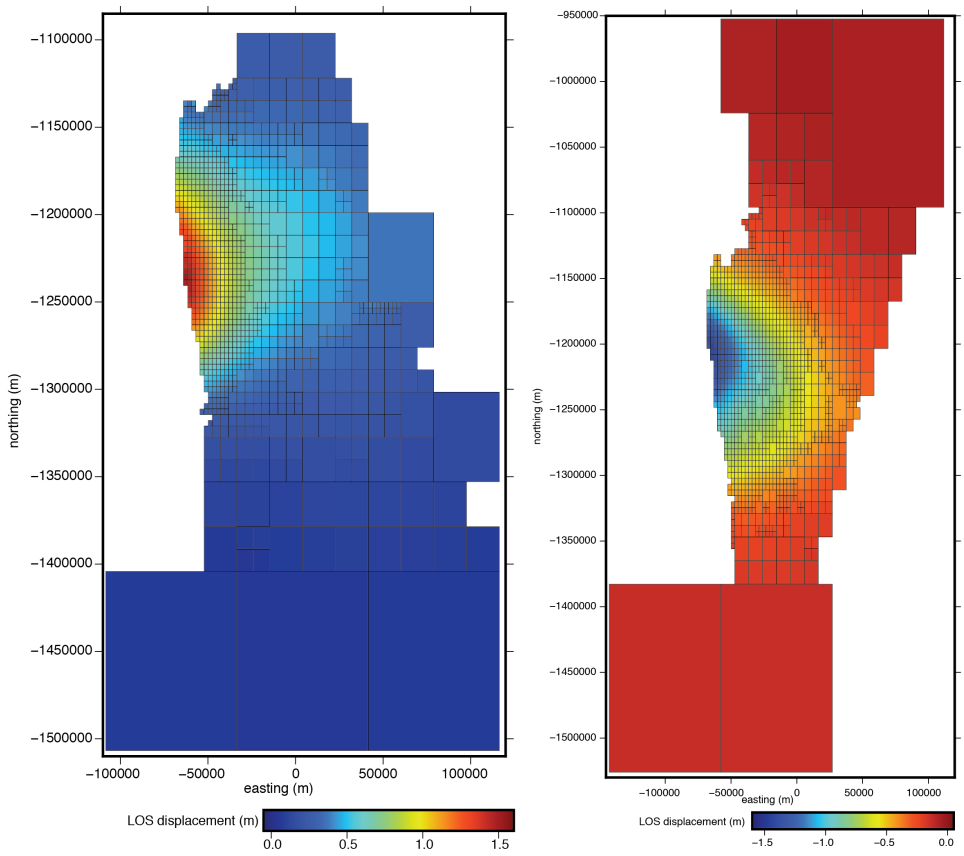
Figure 4.2: Wrapped interferograms. Fringe pattern is centered near the city of Illapel. Decorrelation in the far-field corresponds to mountainous regions.



(a) Ascending.

(b) Descending.

Figure 4.3: Unwrapped interferograms. The ascending and descending interferograms have opposite signs of displacement, suggesting that displacement is dominated by the horizontal component.



(a) Ascending.

(b) Descending.

Figure 4.4: Quadtree decomposed interferograms. The number of data points in each interferogram has been reduced to less than 0.01 % of the number in the original. In addition, each has been adjusted to produce a value of 0 in the far field.

Table 4.2: Fault models used in this study.

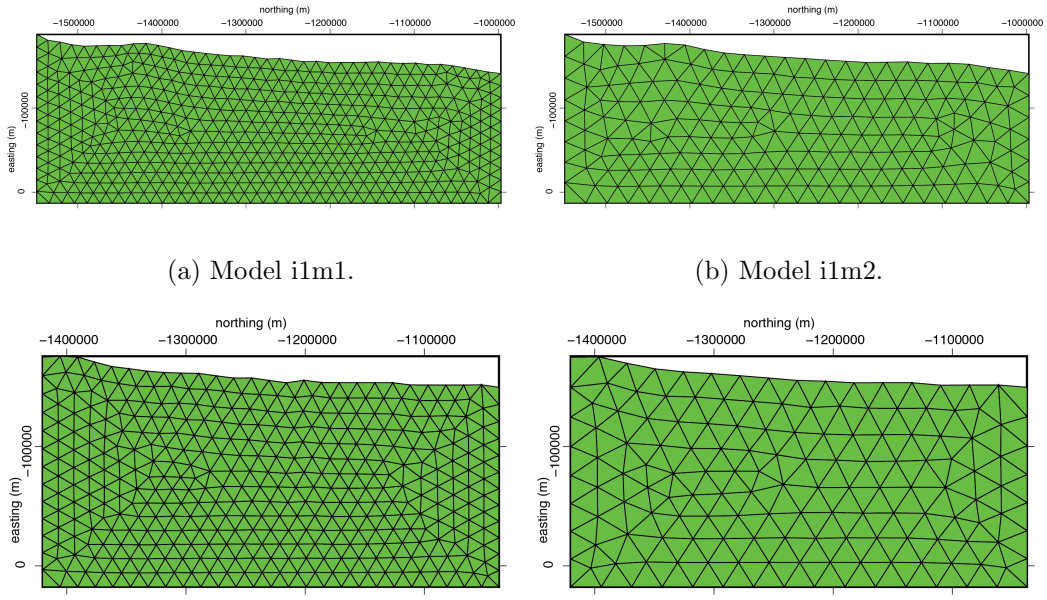
Model	Fault Limits Latitude	Depth (km)	Area (km^2)	Mesh Size (km)	Elements
i1m1	−34° to −29°	66	101 215	15	1036
i1m2				25	369
i2m1	−32.5° to −21.5°	64	71 191	15	731
i2m2				25	262

interferogram to LOS displacement indicates a maximum displacement of approximately 1.5 m. The sign of the displacement in the ascending interferogram is opposite that in the descending, indicating that deformation is predominantly horizontal toward the west.

4.3 Fault Modeling

4.3.1 Slab 1.0 Data and Processing

I downloaded Slab 1.0 models for the South American subduction zone as a depth grid in NetCDF format and as a set of depth contours in ASCII format. I processed the slab models as described in section 3.2.3. The South American subduction zone model extends over 50° in latitude, so I reduced the model size by cutting the extent of the model to an area which fully encloses the slip model inferred from the USGS Finite Fault Model for the Illapel mainshock [65]. In order to determine the best size for the fault, I cut the fault to two different sizes, each of which I meshed with triangles of two different sizes, resulting in four total models (table 4.2, figure 4.5).



(a) Model i1m1.

(b) Model i1m2.

(c) Model i2m1.

(d) Model i2m2.

Figure 4.5: Meshed fault models. See table 4.2 for geographic limits and mesh sizes.

4.3.2 Results

I performed forward and inverse modeling as described in section 3.4. In order to illuminate the effects of each component in the inversion, I inverted each of the two interferograms, both individually and jointly, with each of the four fault models. Table 4.3 lists the components included in each of the inversion types. Results are provided in tables 4.4 through 4.9.

Table 4.3: Inversion types for results given in tables 4.4 through 4.9.

Type	Description
Unweighted	without inverse VCM
Weighted	with inverse VCM
A	smoothing
B	smoothing, restricted ends and bottom
C	smoothing, solved for offset and tilt
D	smoothing, restricted ends and bottom, solved for offset and tilt

Table 4.4: Unweighted inversion results for fault model i1.

Model, IFG	Inversion Type	Moment (N m), Magnitude (M_w)	Max. Slip (m), Depth (km)	95 % Moment Area (km^2)
i1m1, asc.	A	4.24×10^{21}, 8.42	6.55, 9.53	59 352
	B	3.59×10^{21}, 8.37	7.68, 9.53	41 501
	C	2.31×10^{21}, 8.24	4.84, 15.89	41 687
	D	2.41×10^{21}, 8.26	5.97, 11.84	28 527
des.	A	2.98×10^{21}, 8.32	7.54, 9.47	59 345
	B	2.76×10^{21}, 8.29	7.57, 9.47	46 463
	C	2.43×10^{21}, 8.26	7.14, 9.47	38 870
	D	2.32×10^{21}, 8.24	7.28, 9.47	27 023
joint	A	3.28×10^{21}, 8.34	9.32, 8.12	42 597
	B	3.31×10^{21}, 8.35	9.99, 8.12	33 740
	C	4.17×10^{21}, 8.41	8.43, 9.47	59 955
	D	3.66×10^{21}, 8.38	8.55, 8.12	48 689
i1m2, asc.	A	4.35×10^{21}, 8.43	6.87, 11.24	59 171
	B	3.64×10^{21}, 8.37	7.77, 12.33	42 773
	C	2.31×10^{21}, 8.24	4.88, 17.16	41 667
	D	2.43×10^{21}, 8.26	6.10, 12.33	27 352
des.	A	3.03×10^{21}, 8.32	7.54, 9.27	59 979
	B	2.77×10^{21}, 8.29	7.55, 9.27	47 284
	C	2.55×10^{21}, 8.27	7.29, 9.27	45 975
	D	2.35×10^{21}, 8.25	7.30, 9.27	30 622
joint	A	3.43×10^{21}, 8.36	10.11, 8.43	41 258
	B	3.53×10^{21}, 8.37	11.46, 8.43	32 657
	C	4.64×10^{21}, 8.44	8.86, 8.43	59 940
	D	3.97×10^{21}, 8.40	9.02, 8.43	49 508

Table 4.5: Weighted inversion results for fault model i1.

Model, IFG	Inversion Type	Moment (N m), Magnitude (M_w)	Max. Slip (m), Depth (km)	95 % Moment Area (km^2)
i1m1, asc.	A	3.94×10^{21}, 8.40	14.72, 41.65	38 045
	B	3.00×10^{21}, 8.32	17.14, 15.89	29 488
	C	3.63×10^{21}, 8.37	26.28, 41.65	30 180
	D	2.46×10^{21}, 8.26	15.55, 16.91	28 318
des.	A	1.74×10^{21}, 8.16	11.53, 19.14	27 949
	B	1.85×10^{21}, 8.18	11.56, 19.14	28 043
	C	1.54×10^{21}, 8.12	10.97, 19.14	24 070
	D	1.40×10^{21}, 8.10	11.10, 19.14	20 633
joint	A	4.63×10^{21}, 8.44	47.76, 8.12	19 868
	B	5.18×10^{21}, 8.48	66.32, 8.12	16 863
	C	4.35×10^{21}, 8.43	15.36, 8.12	42 373
	D	4.37×10^{21}, 8.43	34.22, 8.12	29 349
i1m2, asc.	A	3.07×10^{21}, 8.32	20.04, 17.09	30 834
	B	3.07×10^{21}, 8.32	21.47, 17.09	26 690
	C	2.76×10^{21}, 8.29	14.60, 17.16	25 922
	D	2.43×10^{21}, 8.26	17.81, 17.09	24 295
des.	A	1.88×10^{21}, 8.18	9.32, 17.09	31 089
	B	2.98×10^{21}, 8.32	9.13, 17.09	42 864
	C	1.61×10^{21}, 8.14	9.01, 17.09	25 978
	D	1.51×10^{21}, 8.12	9.13, 17.09	23 774
joint	A	4.81×10^{21}, 8.45	57.85, 8.43	15 999
	B	5.13×10^{21}, 8.47	56.76, 8.43	16 013
	C	4.50×10^{21}, 8.44	27.23, 8.43	32 613
	D	4.52×10^{21}, 8.44	38.90, 8.43	24 842

Table 4.6: Misfits for model i1 inversions.

		i1m1				i1m2			
Misfit	Inversion	A	B	C	D	A	B	C	D
					Unweighted				
RMS_a,	single	1.68	1.73	1.29	1.39	1.48	1.56	1.30	1.35
(cm)	joint	3.09	3.64	1.80	1.77	2.85	3.37	1.58	1.61
RMS_d,	single	1.52	1.48	1.59	1.43	1.21	1.22	1.19	1.21
(cm)	joint	2.66	3.59	1.75	2.09	2.30	3.14	1.42	1.82
$WRSS_a$	single	2009	2064	1913	1954	2006	2062	1916	1961
	joint	2404	2712	2023	2006	2330	2592	2033	2034
$WRSS_d$	single	1539	1557	1548	1529	1500	1518	1496	1495
	joint	1988	2282	1732	1853	2021	2323	1669	1802
					Weighted				
RMS_a,	single	1.09	1.02	1.18	1.05	1.16	1.08	1.20	1.06
(cm)	joint	3.17	3.11	2.41	2.17	3.08	3.01	2.14	2.05
RMS_d,	single	0.776	0.868	0.801	0.882	0.879	0.991	0.909	0.989
(cm)	joint	1.72	1.92	1.62	1.68	1.63	1.77	1.56	1.65
$WRSS_a$	single	1494	1486	1538	1524	1508	1505	1550	1543
	joint	2000	2052	1768	1783	2017	2070	1792	1807
$WRSS_d$	single	1026	1040	1042	1052	1114	1143	1132	1141
	joint	1229	1250	1196	1202	1235	1276	1173	1192

Overall, I prefer the i2m1 fault model, as the moment and maximum slip show the most consistent agreement with published models of this earthquake. The slip patterns are shown in figures 4.6 through 4.11.

Examining these results reveals some patterns in the effects of the various components included in the inversion. For example, preventing the ends and bottom of the fault from slipping generally decreases the moment in the unweighted inversions, but increases it in the weighted inversions. It generally increases the maximum slip and decreases the area encompassing 95 % of the moment, indicating that it concentrates the slip into a smaller area, as would be expected. It also increases the misfit, except in the weighted inversions of the ascending interferogram.

Table 4.7: Unweighted inversion results for fault model i2.

Model, IFG	Inversion Type	Moment (N m), Magnitude (M_w)	Max. Slip (m), Depth (km)	95 % Moment Area (km^2)
i2m1, asc.	A	3.59×10^{21}, 8.37	6.61, 11.45	39 883
	B	3.36×10^{21}, 8.35	7.74, 9.70	30 154
	C	2.04×10^{21}, 8.21	4.58, 20.11	28 195
	D	2.20×10^{21}, 8.23	5.41, 14.50	24 398
des.	A	2.66×10^{21}, 8.28	7.49, 9.72	37 401
	B	2.55×10^{21}, 8.27	7.50, 9.72	29 005
	C	2.31×10^{21}, 8.24	7.12, 10.20	27 126
	D	2.21×10^{21}, 8.23	7.13, 10.20	22 575
joint	A	3.16×10^{21}, 8.33	9.01, 8.23	37 373
	B	3.12×10^{21}, 8.33	9.62, 8.23	28 733
	C	2.97×10^{21}, 8.31	7.91, 9.72	40 420
	D	2.64×10^{21}, 8.28	8.00, 8.23	27 153
i2m2, asc.	A	3.62×10^{21}, 8.37	6.75, 11.42	39 781
	B	3.36×10^{21}, 8.35	7.68, 11.42	29 862
	C	2.00×10^{21}, 8.20	4.48, 17.19	28 158
	D	2.23×10^{21}, 8.23	5.63, 11.42	23 596
des.	A	2.67×10^{21}, 8.28	7.48, 9.31	37 374
	B	2.54×10^{21}, 8.27	7.53, 9.31	29 885
	C	2.30×10^{21}, 8.24	7.23, 11.34	26 788
	D	2.21×10^{21}, 8.23	7.25, 11.34	22 103
joint	A	3.28×10^{21}, 8.34	9.57, 9.40	36 808
	B	3.28×10^{21}, 8.34	10.82, 9.40	28 096
	C	3.04×10^{21}, 8.32	8.41, 9.40	40 047
	D	2.73×10^{21}, 8.29	8.62, 9.40	27 369

Table 4.8: Weighted inversion results for fault model i2.

Model, IFG	Inversion Type	Moment (N m), Magnitude (M_w)	Max. Slip (m), Depth (km)	95 % Moment Area (km^2)
i2m1, asc.	A	3.37×10^{21}, 8.35	14.43, 17.55	31 138
	B	3.18×10^{21}, 8.34	16.93, 17.55	25 604
	C	1.85×10^{21}, 8.18	10.98, 17.55	22 895
	D	1.85×10^{21}, 8.18	15.25, 17.55	21 001
des.	A	2.18×10^{21}, 8.23	10.48, 20.13	28 985
	B	2.64×10^{21}, 8.28	21.00, 7.82	23 969
	C	1.62×10^{21}, 8.14	10.44, 20.13	26 050
	D	1.53×10^{21}, 8.12	10.49, 20.13	21 682
joint	A	4.31×10^{21}, 8.42	35.61, 8.20	20 436
	B	4.94×10^{21}, 8.46	61.62, 8.20	15 896
	C	2.56×10^{21}, 8.27	9.20, 30.80	30 997
	D	2.42×10^{21}, 8.26	12.35, 10.20	23 733
i2m2, asc.	A	3.38×10^{21}, 8.35	12.68, 17.19	33 768
	B	3.22×10^{21}, 8.34	17.20, 15.59	27 185
	C	1.89×10^{21}, 8.18	7.35, 17.19	25 960
	D	1.99×10^{21}, 8.20	14.34, 17.19	19 900
des.	A	2.27×10^{21}, 8.24	8.39, 17.57	28 435
	B	2.57×10^{21}, 8.27	15.55, 8.50	25 221
	C	1.66×10^{21}, 8.15	8.05, 17.57	25 617
	D	1.59×10^{21}, 8.13	8.20, 17.57	21 865
joint	A	4.45×10^{21}, 8.43	52.49, 8.57	17 356
	B	4.80×10^{21}, 8.45	50.17, 8.57	16 616
	C	2.56×10^{21}, 8.27	7.79, 21.93	32 427
	D	2.73×10^{21}, 8.29	21.89, 8.57	22 496

Table 4.9: Misfits for model i2 inversions.

		i2m1				i2m2			
Misfit	Inversion	A	B	C	D	A	B	C	D
					Unweighted				
RMS_a,	single	1.68	1.76	1.34	1.48	1.56	1.72	1.28	1.39
(cm)	joint	3.05	3.61	1.95	1.93	2.81	3.32	1.69	1.69
RMS_d,	single	1.57	1.63	1.48	1.54	1.37	1.39	1.22	1.24
(cm)	joint	2.53	3.54	1.92	2.30	2.19	3.05	1.61	2.01
$WRSS_a$	single	2032	2123	1911	1950	2043	2157	1900	1960
	joint	2365	2756	1969	1920	2300	2630	1971	1933
$WRSS_d$	single	1636	1694	1549	1543	1606	1671	1526	1524
	joint	1992	2299	1945	2096	2043	2363	1946	2101
					Weighted				
RMS_a,	single	1.08	1.02	1.14	1.07	1.30	1.17	1.28	1.12
(cm)	joint	3.21	3.14	3.13	2.99	3.19	3.03	3.18	2.75
RMS_d,	single	0.794	0.875	0.787	0.875	0.919	1.03	0.920	1.00
(cm)	joint	1.60	1.79	2.02	2.11	1.79	1.87	2.05	2.16
$WRSS_a$	single	1494	1494	1535	1532	1566	1566	1599	1580
	joint	1933	2038	1780	1799	2001	2105	1831	1852
$WRSS_d$	single	1072	1083	1059	1072	1172	1208	1162	1172
	joint	1245	1286	1336	1473	1284	1363	1395	1535

Removing the offset and orbital ramp generally decreases the moment and maximum slip; it decreases the area of 95 % moment release except in the joint inversions, where it increases this area. It decreases the misfit except in some of the weighted inversions. Adding the restriction of the ends and bottom further decreases the moment and area, while increasing the maximum slip. Again, this is to be expected if fewer elements are allowed to slip. This has a varying effect on the misfit, improving it in some cases, worsening it in others.

There are also interesting patterns when comparing the ascending, descending, and joint inversions. For example, inversion of the ascending interferogram tends to produce higher moment and deeper maximum slip than with the descending interferogram. Given

the longer postseismic interval of the ascending interferogram, the presence of afterslip would be expected to produce these findings.

One pattern for which I have no explanation is the systematic up-dip migration of the slip pattern in the joint inversions as compared to inversion of the ascending or descending interferogram alone. This can be seen in both the weighted and unweighted inversions. An additional feature is present in the weighted single inversions where the fault ends and bottom were prevented from slipping: a cluster of patches with very slip at the up-dip edge (figures 4.7b and 4.9b).

One likely explanation for the variability in the slip patterns can be seen by examining the model resolution matrix in figure 4.12. This matrix is

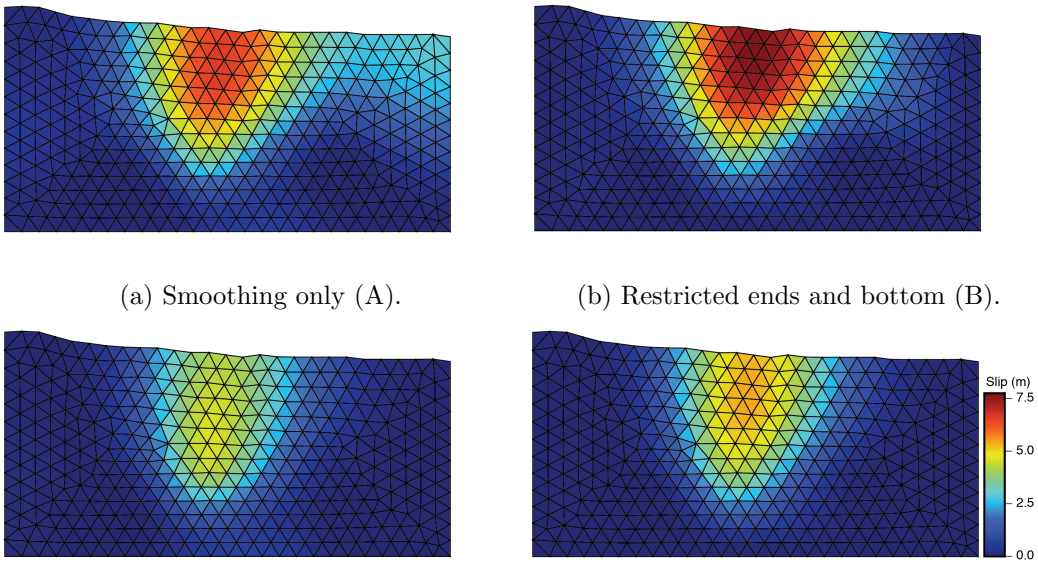
$$\mathbf{R} = \left[\mathbf{A}_\mathrm{s}^\mathrm{T}\mathbf{A}_\mathrm{s}\right]^{-1}\mathbf{A}^\mathrm{T}\mathbf{A} \tag{4.1}$$

where

$$\mathbf{A}_\mathrm{s} = \begin{bmatrix} \mathbf{A} \\ \\ \kappa\mathbf{S} \end{bmatrix} \tag{4.2}$$

The up-dip portion of the fault model has very low resolution. This is not unexpected as this area lies offshore and is quite distant from the observation points on land.

Another consideration is that the matrix of Green's functions is rank deficient. This would make the inversions numerically unstable, resulting in slip distributions which seem very unrealistic. To explore this possibility, I performed another set of inversions with higher weight given to the smoothing matrix; these results are in tables 4.10 and 4.11 and



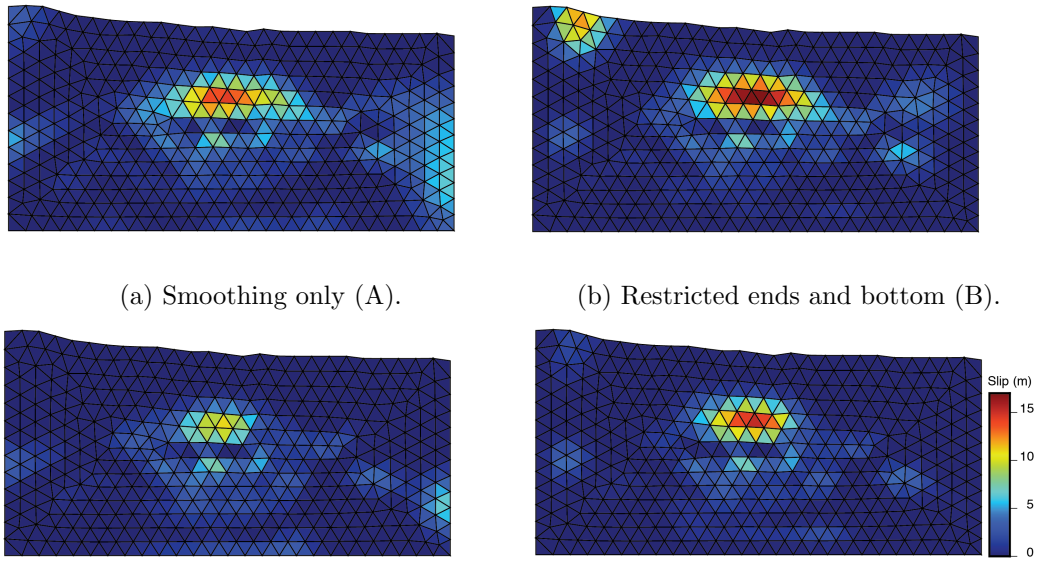
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.6: Slip models from ascending, unweighted inversions with lower smoothing value.



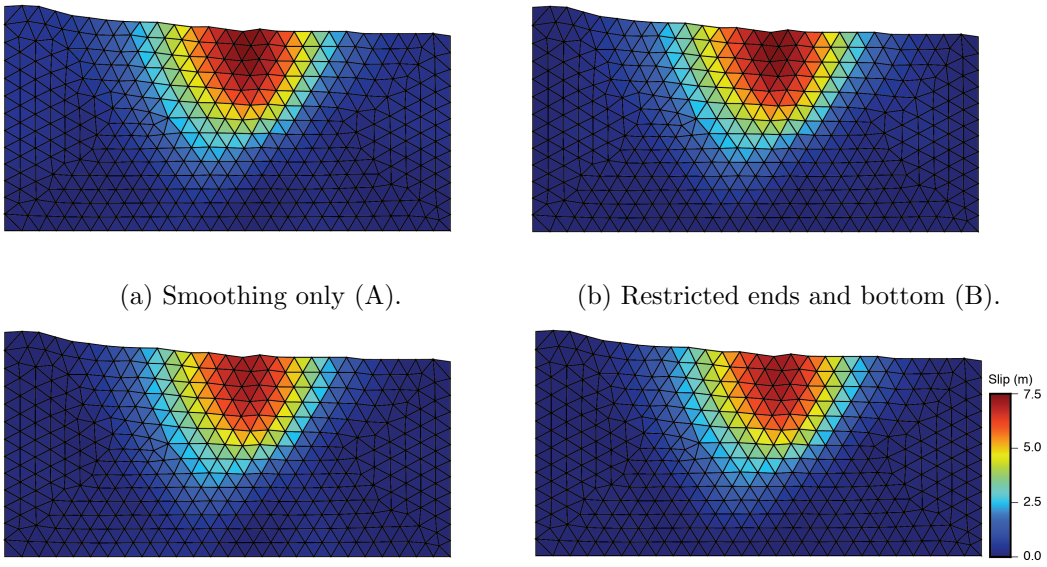
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.7: Slip models from ascending, weighted inversions with lower smoothing value.



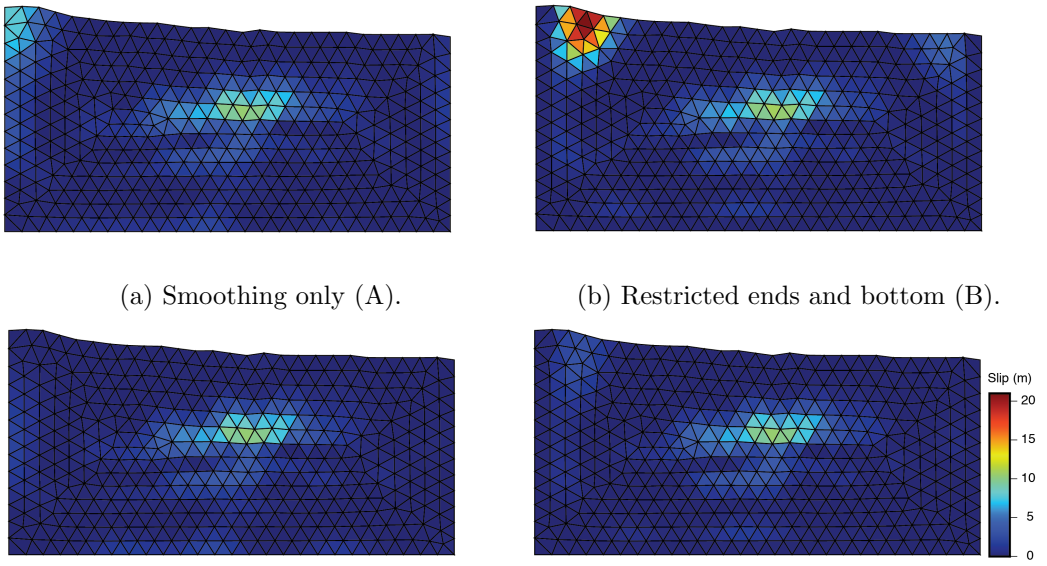
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.8: Slip models from descending, unweighted inversions with lower smoothing value.



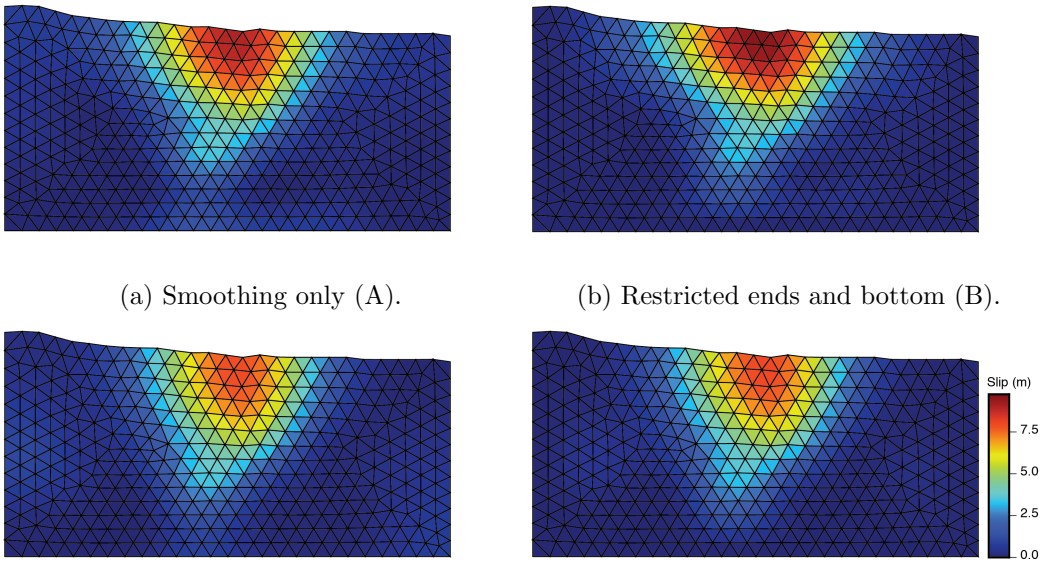
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.9: Slip models from descending, weighted inversions with lower smoothing value.



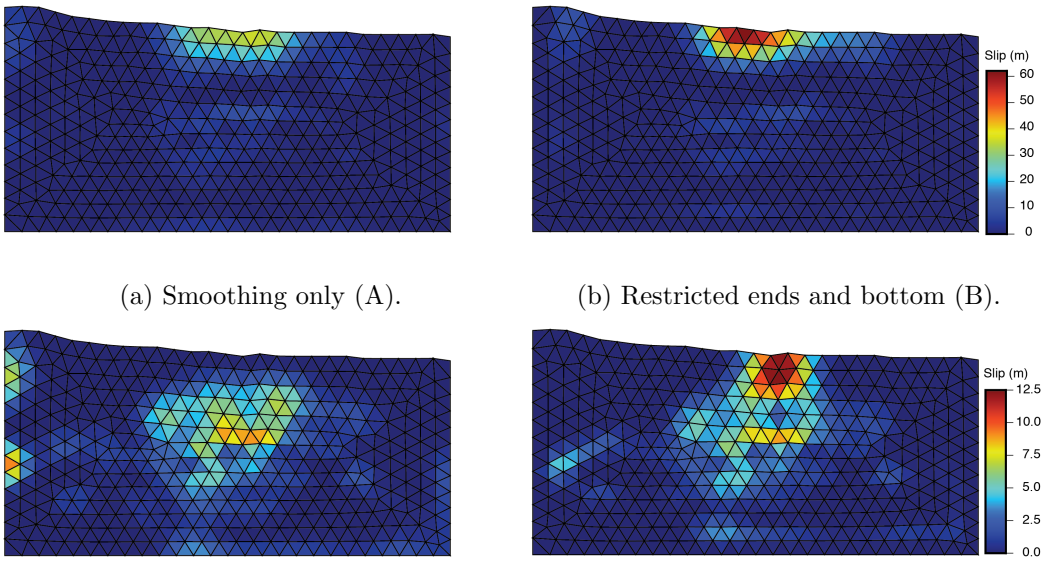
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.10: Slip models from joint, unweighted inversions with lower smoothing value.



(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.11: Slip models from joint, weighted inversions with lower smoothing value.

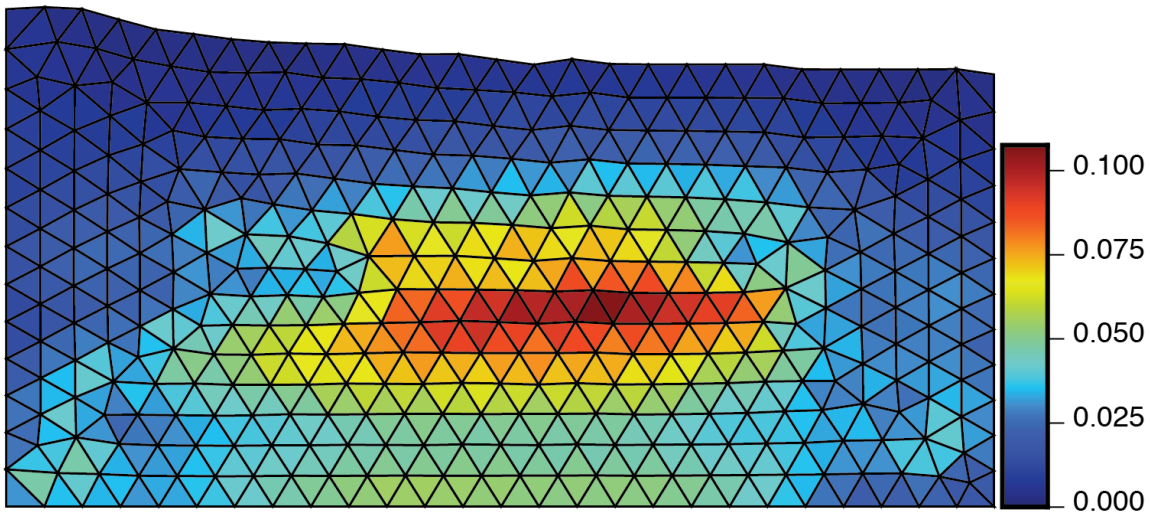


Figure 4.12: Model resolution with lower smoothing value.

figures 4.13 through 4.18. While this did not eliminate all of the problems encountered with the lower smoothing value inversions, the slip patterns did appear more physically plausible.

One feature that persisted throughout most of the inversions was a downdip extension of slip at the southern end of the main slip patch; this is most clearly seen in figure 4.10. This feature likely represents a true asperity as it is present in the geodetic inversion of [66] as well as the W-phase inversion of [65].

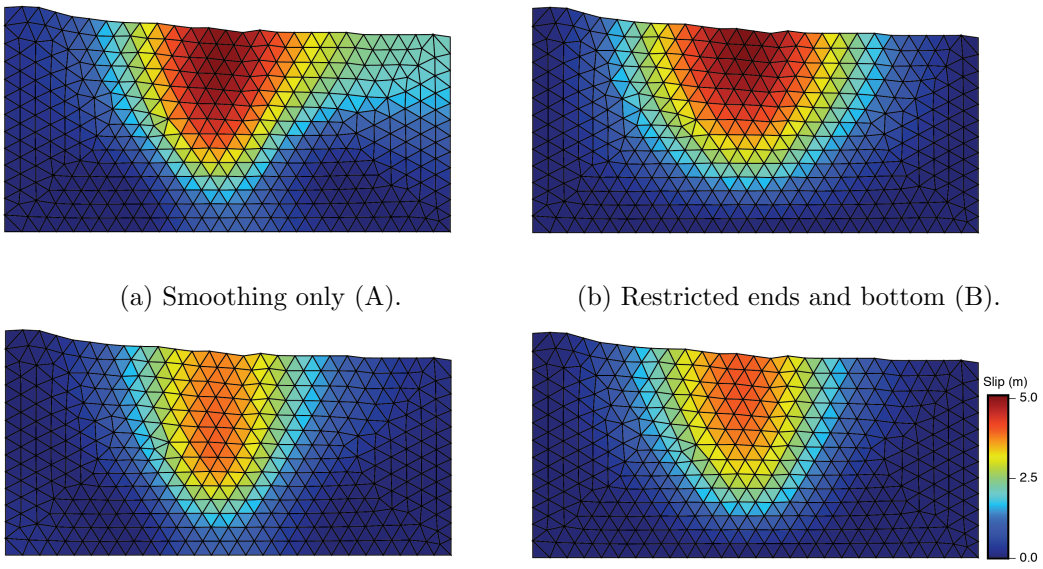
The recomputed model resolution with the higher smoothing value (figure 4.19) has lower values overall but shows a somewhat different pattern in that the resolution is more evenly distributed, especially in the down-dip portion of the fault.

Table 4.10: Inversion results for fault model i2m1 with higher smoothing value.

IFG	Inversion Type	Moment (N m), Magnitude (M_w)	Max. Slip (m), Depth (km)	95 % Moment Area (km^2)
unweighted,	A	3.20×10^{21}, 8.34	5.07, 10.26	41 823
asc.	B	2.95×10^{21}, 8.31	5.09, 10.26	37 203
	C	2.19×10^{21}, 8.23	3.98, 21.45	35 177
	D	2.09×10^{21}, 8.21	4.06, 11.45	31 165
des.	A	2.81×10^{21}, 8.30	6.27, 8.70	39 401
	B	2.56×10^{21}, 8.27	6.38, 8.70	30 385
	C	2.70×10^{21}, 8.29	6.43, 9.72	40 867
	D	2.37×10^{21}, 8.25	6.40, 9.72	28 501
joint	A	2.86×10^{21}, 8.30	7.05, 8.70	31 122
	B	2.86×10^{21}, 8.30	6.95, 8.23	29 806
	C	2.75×10^{21}, 8.29	6.89, 9.72	36 525
	D	2.59×10^{21}, 8.28	6.67, 10.28	26 878
weighted,	A	3.30×10^{21}, 8.35	5.04, 13.62	46 673
asc.	B	3.52×10^{21}, 8.36	9.60, 8.67	29 668
	C	2.16×10^{21}, 8.22	3.73, 29.37	46 718
	D	2.09×10^{21}, 8.21	5.90, 10.26	24 217
des.	A	2.77×10^{21}, 8.29	6.73, 9.72	41 805
	B	2.54×10^{21}, 8.27	6.75, 9.72	32 549
	C	2.69×10^{21}, 8.29	6.18, 12.15	47 794
	D	2.36×10^{21}, 8.25	6.42, 9.72	31 868
joint	A	3.14×10^{21}, 8.33	7.49, 8.70	43 222
	B	3.61×10^{21}, 8.37	12.52, 8.20	28 698
	C	2.84×10^{21}, 8.30	6.41, 12.15	44 336
	D	2.61×10^{21}, 8.28	6.99, 8.70	33 218

Table 4.11: Misfits for model i2m1 inversions with higher weighting value.

		Unweighted				Weighted			
Misfit	Inversion	A	B	C	D	A	B	C	D
RMS_a,	single	3.41	9.28	2.82	6.12	4.41	3.82	3.51	3.40
(cm)	joint	4.83	6.59	2.94	3.97	4.83	5.61	3.49	4.44
RMS_d,	single	6.02	5.62	3.14	4.14	2.95	3.18	3.05	3.26
(cm)	joint	5.10	7.41	3.16	4.83	4.12	4.90	3.33	3.75
$WRSS_a$	single	4182	2996	2075	2202	2047	2051	1990	1973
	joint	2705	3146	2086	2034	2276	2485	2012	1998
$WRSS_d$	single	2229	2108	1927	1886	1610	1716	1728	1673
	joint	2225	2722	2162	2649	1940	2582	1828	2151



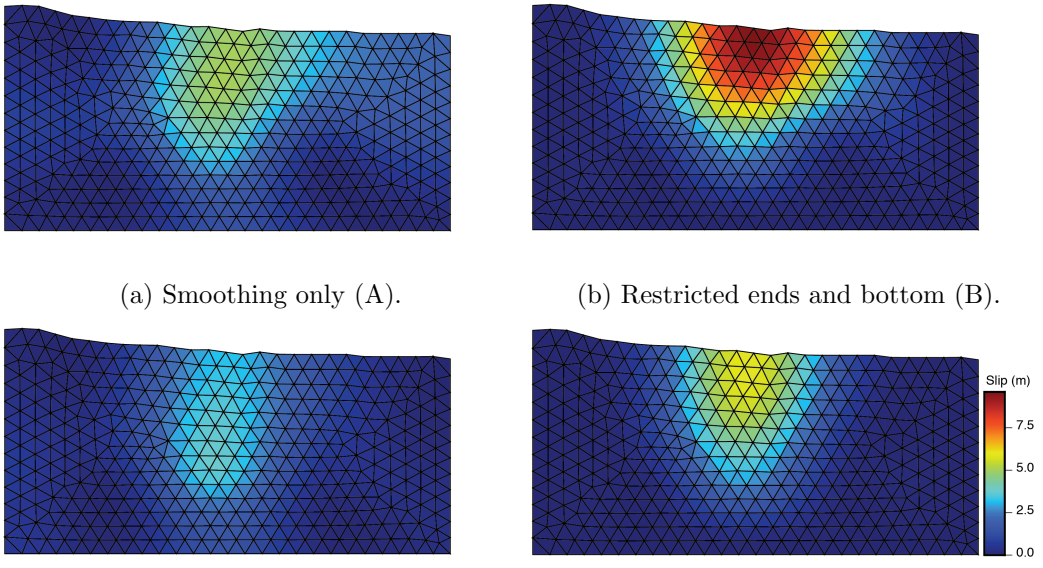
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.13: Slip models from ascending, unweighted inversions with higher smoothing value.



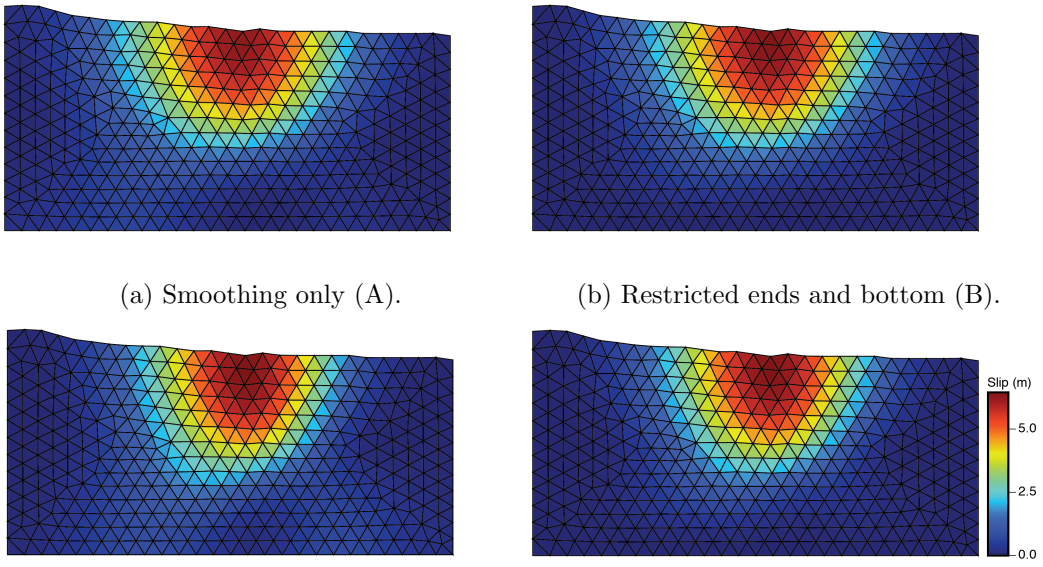
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.14: Slip models from ascending, weighted inversions with higher smoothing value.



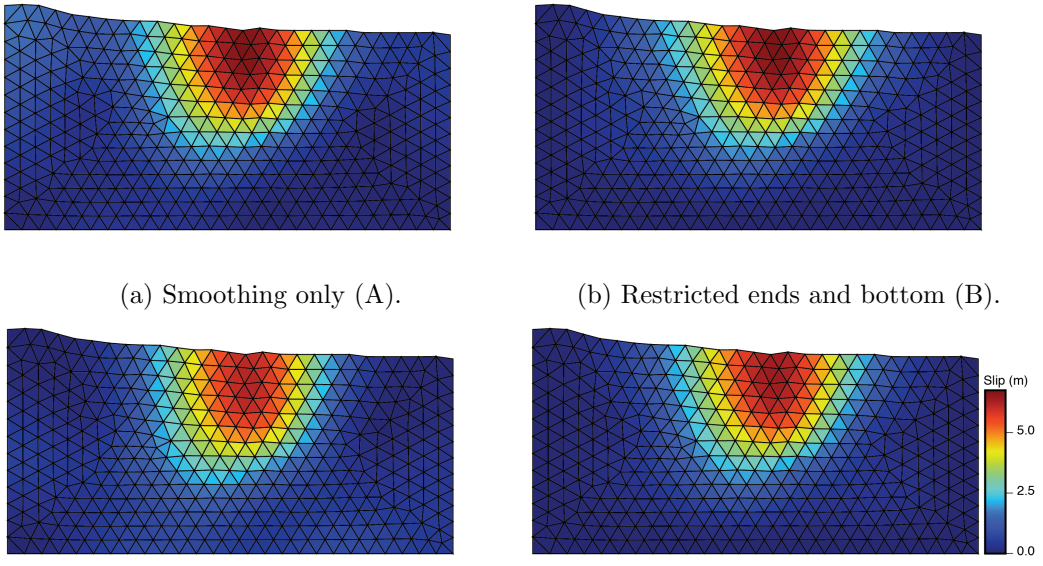
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.15: Slip models from descending, unweighted inversions with higher smoothing value.



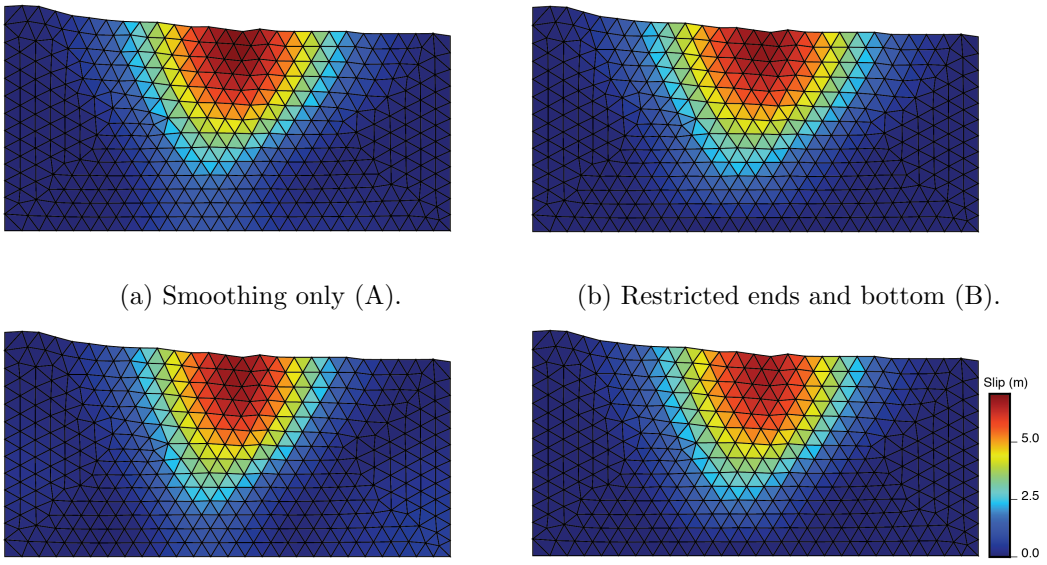
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.16: Slip models from descending, weighted inversions with higher smoothing value.



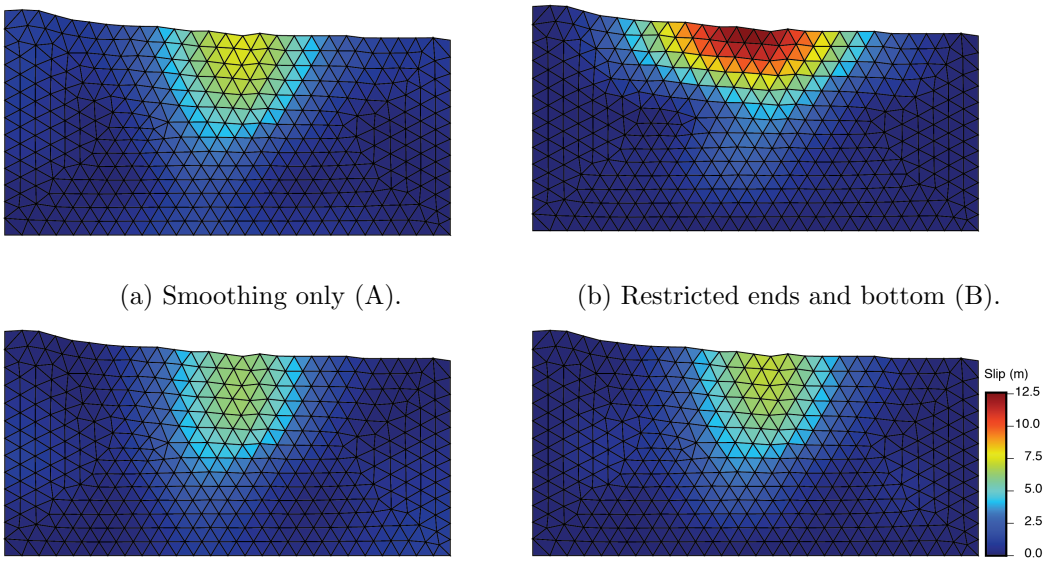
(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.17: Slip models from joint, unweighted inversions with higher smoothing value.



(a) Smoothing only (A).

(b) Restricted ends and bottom (B).

(c) Solved for ramp (C).

(d) Restricted and solved for ramp (D).

Figure 4.18: Slip models from joint, weighted inversions with higher smoothing value.

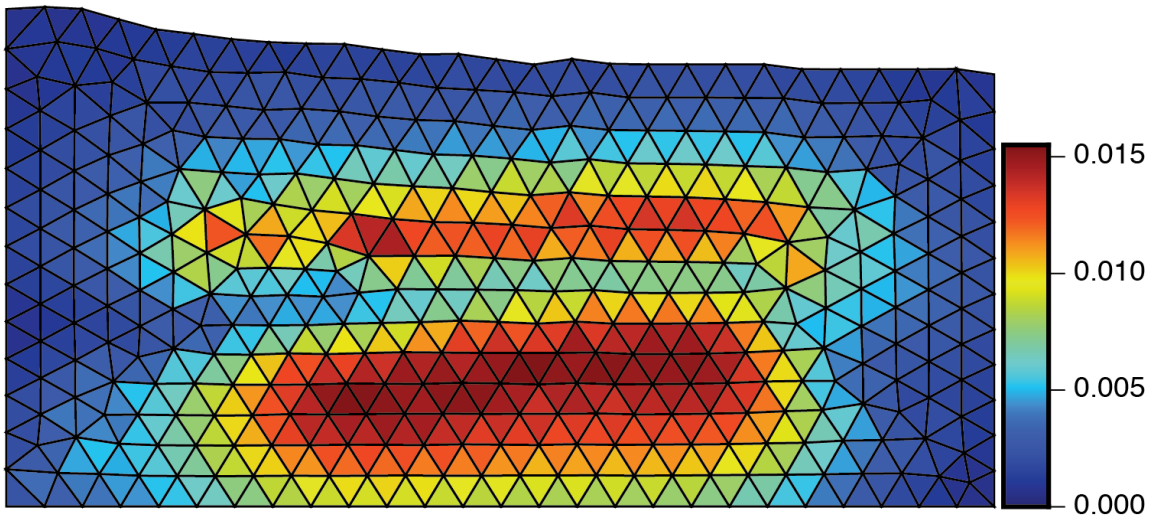


Figure 4.19: Model resolution with higher smoothing value.

Chapter 5

Conclusions

Characterization of earthquake source parameters is a crucial component of the scientific response to an earthquake. Reducing the time required for this process will improve the utility of these studies for the public. In this thesis, I have focused on developing a set of tools by which earthquake source parameters can be rapidly inferred, thereby allowing quicker turnaround on downstream products such as aftershock forecasts.

The Illapel earthquake was the first great earthquake to occur after the ESA's Sentinel-1A satellite became fully functional. Thus it provided an opportune test case for developing these tools using the latest generation of SAR imagery. At the time of the event, there were few software tools widely available for processing these images. This presented a major challenge in the early stages of this project as I struggled with poorly-documented software which frequently crashed during processing. This situation has since improved a great deal; not only has functionality been added to the Sentinel-1 Toolbox software, but

it has become much more stable. Furthermore, other software packages are now available that can process these images, e.g. ISCE.

The fact that the Illapel earthquake was a subduction zone megathrust event posed another set of challenges, but also presented an opportunity to demonstrate the utility of rapid characterization in this most dangerous tectonic setting. Part of the deformation pattern produced by a subduction thrust earthquake is offshore and therefore invisible to radar. This vertical deformation of the seafloor is of particular interest when evaluating tsunami potential; it also helps constrain fault geometry and up-dip slip distribution. This challenge can be partly overcome by using a pre-determined fault geometry model, i.e. Slab 1.0, and assuming a pure thrust mechanism.

The final step in achieving the goal of rapid characterization would be pre-determination of an interferogram downsampling scheme and computation of the matrix of Green's functions. I have shown that a triangular mesh of the non-planar fault surface can be used fairly easily with a published algorithm for calculating surface displacements. A downsampling method that relies on the data resolution, such as that of [37], is an attractive option as it can be implemented before an earthquake; in fact, it was used in one published model of the Illapel earthquake [66].

One persistent stumbling block in this project was the numerical instability of these inversions. Increasing the model smoothing did help to relieve these issues to a certain extent, however, the solutions remain somewhat unsatisfactory. While I was able to arrive at solutions with geodetic moments and peak slip values consistent with those published by

others, the difficulty in reaching a stable nonnegative least squares solution is a factor that should be kept in mind when inverting future subduction zone earthquakes.

References

[1] Böse, M, Allen, R, Brown, H, Gua, G, Fischer, M, Hauksson, E, Heaten, T, Hellweg, M, Liukis, M, Neuhauser, D, Maechling, P, Solanki, K, Vinci, M, Henson, I, Khainovski, O, Kuyuk, Carpio, M, Meier, M.-A, & Jordan, T. (2014) in *Early Warning for Geological Disasters: Scientific Methods and Current Practice*, eds. Wenzel, F & Zschau, J. (Springer), pp. 49–69.

[2] Ohta, Y, Kobayashi, T, Tsushima, H, Miura, S, Hino, R, Takasu, T, Fujimoto, H, Iinuma, T, Tachibana, K, Demachi, T, Sato, T, Ohzono, M, & Umino, N. (2012) Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: Application to the 2011 Tohoku-Oki earthquake (Mw 9.0). *Journal of geophysical research* **117**, B02311.

[3] Jaiswal, K. S & Wald, D. J. (2013) in *Handbook of Seismic Risk Analysis and Management of Civil Infrastructure Systems*. (Woodhead Publishing), pp. 839–869e.

[4] Parsons, T, Segou, M, Sevilgen, V, Milner, K, Field, E, Toda, S, & Stein, R. S. (2014) Stress-based aftershock forecasts made within 24 h postmain shock: Expected north San Francisco Bay area seismicity changes after the 2014 M=6.0 West Napa earthquake. *Geophysical research letters* **41**, 2014GL062379.

[5] Toda, S & Stein, R. S. (2013) The 2011 M=9.0 Tohoku oki earthquake more than doubled the probability of large shocks beneath Tokyo. *Geophysical research letters* **40**, 2562–2566.

[6] Steacy, S, Marsan, D, Nalbant, S. S, & McCloskey, J. (2004) Sensitivity of static stress calculations to the earthquake slip distribution. *Journal of geophysical research* **109**, B04303.

[7] Massonnet, D, Rossi, M, Carmona, C, Adragna, F, Peltzer, G, Feigl, K, & Rabaute, T. (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* **364**, 138–142.

[8] Bock, Y & Webb, F. H. (2012) MEaSUREs Solid Earth Science ESDR System (http://geoapp03.ucsd.edu/gridsphere/gridsphere). Accessed: 2016-11-4.

[9] Hayes, G. P. (2011) Rapid source characterization of the 2011 M_w 9.0 off the Pacific coast of Tohoku Earthquake. *Earth, Planets and Space* **63**, 529–534.

[10] Weston, J, Ferreira, A. M. G, & Funning, G. J. (2011) Global compilation of interferometric synthetic aperture radar earthquake source models: 1. Comparisons with seismic catalogs. *Journal of geophysical research* **116**, B08408.

[11] Weston, J, Ferreira, A. M. G, & Funning, G. J. (2012) Systematic comparisons of earthquake source models determined using InSAR and seismic data. *Tectonophysics* **532–535**, 61–81.

[12] Lay, T. (2015) The surge of great earthquakes from 2004 to 2014. *Earth and planetary science letters* **409**, 133–146.

[13] Pacheco, J. F & Sykes, L. R. (1992) Seismic moment catalog of large shallow earthquakes, 1900 to 1989. *Bulletin of the Seismological Society of America* **82**, 1306–1349.

[14] Stein, S & Stein, C. A. (1996) in *Subduction Top to Bottom*, Geophysical Monographs, eds. Bebout, G. E, Scholl, D. W, Kirby, S. H, & Platt, J. P. (American Geophysical Union) Vol. 96, pp. 1–17.

[15] Pacheco, J. F, Sykes, L. R, & Scholz, C. H. (1993) Nature of seismic coupling along simple plate boundaries of the subduction type. *Journal of geophysical research* **98**, 14133–14159.

[16] Yeats, R. S, Sieh, K, & Allen, C. R. (1997) *The Geology of Earthquakes.* (Oxford University Press, New York).

[17] Schwartz, S. Y & Rokosky, J. M. (2007) Slow slip events and seismic tremor at circum-Pacific subduction zones. *Reviews of geophysics* **45**, RG3004.

[18] Wang, L, Shum, C. K, Simons, F. J, Tassara, A, Erkan, K, Jekeli, C, Braun, A, Kuo, C, Lee, H, & Yuan, D.-N. (2012) Coseismic slip of the 2010 Mw 8.8 Great Maule, Chile, earthquake quantified by the inversion of GRACE observations. *Earth and planetary science letters* **335–336**, 167–179.

[19] Blewitt, G. (2007) in *Geodesy*, Treatise on Geophysics, ed. Herring, T. (Elsevier) Vol. 3, pp. 351–390.

[20] Salvi, S, Stramondo, S, Funning, G. J, Ferretti, A, Sarti, F, & Mouratidis, A. (2012) The Sentinel-1 mission for the improvement of the scientific understanding and the operational monitoring of the seismic cycle. *Remote sensing of environment* **120**, 164–174.

[21] Massonnet, D & Feigl, K. L. (1998) Radar interferometry and its application to changes in the Earth's surface. *Reviews of Geophysics* **36**, 441–500.

[22] Showman, G. A. (2012) in *Principles of Modern Radar: Advanced Techniques*, eds. Melvin, W. L & Scheer, J. A. (SciTech Publishing) Vol. 2, pp. 259–335.

[23] Simons, M & Rosen, P. A. (2015) in *Treatise on Geophysics (Second Edition)*, ed. Schubert, G. (Elsevier, Oxford) Vol. 3, pp. 339–385.

[24] Goldstein, R. M, Zebker, H. A, & Werner, C. L. (1988) Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Science* **23**, 713–720.

[25] Chen, C. W & Zebker, H. A. (2000) Network approaches to two-dimensional phase unwrapping: intractability and two new algorithms. *Journal of the Optical Society of America. A, Optics, image science, and vision* **17**, 401–414.

[26] Olsen, R. C, ed. (2007) in *Remote Sensing from Air and Space*. (Society of Photo-Optical Instrumentation Engineers, Bellingham, WA), pp. 103–115.

[27] Torres, R, Snoeij, P, Geudtner, D, Bibby, D, Davidson, M, Attema, E, Potin, P, Rommen, B, Floury, N, Brown, M, Traver, I. N, Deghaye, P, Duesmann, B, Rosich, B, Miranda, N, Bruno, C, L'Abbate, M, Croci, R, Pietropaolo, A, Huchler, M, & Rostan, F. (2012) GMES Sentinel-1 mission. *Remote sensing of environment* **120**, 9–24.

[28] European Space Agency. (2016) Missions - Sentinel Online (http://sentinel.esa.int/web/sentinel/missions). Accessed: 2016-10-27.

[29] De Zan, F & Monti Guarnieri, A. (2006) TOPSAR: Terrain Observation by Progressive Scans. *IEEE transactions on geoscience and remote sensing: a publication of the IEEE Geoscience and Remote Sensing Society* **44**, 2352–2360.

[30] Prats-Iraola, P, Scheiber, R, Marotti, L, Wollstadt, S, & Reigber, A. (2012) TOPS Interferometry With TerraSAR-X. *IEEE transactions on geoscience and remote sensing: a publication of the IEEE Geoscience and Remote Sensing Society* **50**, 3179–3188.

[31] Grandin, R, Klein, E, Métois, M, & Vigny, C. (2016) Three-dimensional displacement field of the 2015 M_w8.3 Illapel earthquake (Chile) from across- and along-track Sentinel-1 TOPS interferometry. *Geophysical research letters* **43**, 2016GL067954.

[32] European Space Agency. (2016) Sentinel-1A Data (http://vertex.daac.asf.alaska.edu/). Accessed: 2016-7-15.

[33] European Space Agency. (2016) Sentinel-1 Toolbox.

[34] Chen, C. W & Zebker, H. A. (2001) Two-dimensional phase unwrapping with use of statistical models for cost functions in nonlinear optimization. *Journal of the Optical Society of America. A, Optics, image science, and vision* **18**, 338–351.

[35] González, P. J, Bagnardi, M, Hooper, A. J, Larsen, Y, Marinkovic, P, Samsonov, S. V, & Wright, T. J. (2015) The 2014–2015 eruption of Fogo volcano: Geodetic modeling of Sentinel-1 TOPS interferometry. *Geophysical research letters* **42**, 2015GL066003.

[36] The MathWorks Inc. (2015) MATLAB.

[37] Lohman, R. B & Simons, M. (2005) Some thoughts on the use of InSAR data to constrain models of surface deformation: Noise structure and data downsampling. *Geochemistry, Geophysics, Geosystems* **6**, Q01007.

[38] Menke, W. (2012) in *Geophysical Data Analysis: Discrete Inverse Theory.* (Academic Press, Boston), pp. 69–88.

[39] Hayes, G. P, Wald, D. J, & Johnson, R. L. (2012) Slab1.0: A three-dimensional model of global subduction zone geometries. *Journal of geophysical research* **117**, B01302.

[40] U. S. Geological Survey. (2012) Slab Models for Subduction Zones (http://earthquake.usgs.gov/data/slab/). Accessed: 2016-7-7.

[41] Schellart, W. P & Rawlinson, N. (2013) Global correlations between maximum magnitudes of subduction zone interface thrust earthquakes and physical parameters of subduction zones. *Physics of the Earth and Planetary Interiors* **225**, 41–67.

[42] Stein, S, Geller, R. J, & Liu, M. (2012) Why earthquake hazard maps often fail and what to do about it. *Tectonophysics* **562–563**, 1–25.

[43] Wessel, P, Smith, W. H. F, Scharroo, R, Luis, J, & Wobbe, F. (2015) Generic Mapping Tools.

[44] Ji, C, Helmberger, D. V, Wald, D. J, & Ma, K.-F. (2003) Slip history and dynamic implications of the 1999 Chi-Chi, Taiwan, earthquake. *Journal of geophysical research* **108**, 2412.

[45] Sandia Corporation. (2011) Cubit.

[46] Steketee, J. A. (1958) On Volterra's dislocations in a semi-infinite elastic medium. *Canadian Journal of Physics* **36**, 192–205.

[47] Segall, P. (2010) *Earthquake and Volcano Deformation.* (Princeton University Press).

[48] Okada, Y. (1985) Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America* **75**, 1135–1154.

[49] Jeyakumaran, M, Rudnicki, J. W, & Keer, L. M. (1992) Modeling slip zones with triangular dislocation elements. *Bulletin of the Seismological Society of America* **82**, 2153–2169.

[50] Meade, B. J. (2007) Algorithms for the calculation of exact displacements, strains, and stresses for triangular dislocation elements in a uniform elastic half space. *Computers & geosciences* **33**, 1064–1075.

[51] Maerten, F, Resor, P, Pollard, D, & Maerten, L. (2005) Inverting for Slip on Three-Dimensional Fault Surfaces Using Angular Dislocations. *Bulletin of the Seismological Society of America* **95**, 1654–1665.

[52] Yoffe, E. H. (1960) The Angular Dislocation. *Philosophical Magazine* **5**, 161–175.

[53] Comninou, M & Dundurs, J. (1975) The angular dislocation in a half space. *Journal Of Elasticity* **5**, 203–216.

[54] Nikkhoo, M & Walter, T. R. (2015) Triangular dislocation: an analytical, artefact-free solution. *Geophysical Journal International* **201**, 1119–1141.

[55] U.S. Geological Survey. (2016) Shuttle Radar Topography Mission (http://earthexplorer.usgs.gov/). Accessed: 2016-7-15.

[56] Menke, W. (2012) in *Geophysical Data Analysis: Discrete Inverse Theory (Third Edition)*. (Academic Press, Boston), pp. 39–68.

[57] Funning, G. J, Parsons, B, Wright, T. J, Jackson, J. A, & Fielding, E. J. (2005) Surface displacements and source parameters of the 2003 Bam (Iran) earthquake from Envisat advanced synthetic aperture radar imagery. *Journal of geophysical research* **110**, B09406.

[58] Ranero, C. R, von Huene, R, Weinrebe, W, & Reichert, C. (2006) in *The Andes*, Frontiers in Earth Sciences, eds. Oncken, O, Chong, G, Franz, G, Giese, P, Götze, H.-J, Ramos, V. A, Strecker, M. R, & Wigger, P. (Springer Berlin Heidelberg), pp. 91–121.

[59] Yue, H, Lay, T, Rivera, L, An, C, Vigny, C, Tong, X, & Báez Soto, J. C. (2014) Localized fault slip to the trench in the 2010 Maule, Chile Mw = 8.8 earthquake from joint inversion of high-rate GPS, teleseismic body waves, InSAR, campaign GPS, and tsunami observations. *Journal of Geophysical Research, [Solid Earth]* **119**, 2014JB011340.

[60] Lay, T, Yue, H, Brodsky, E. E, & An, C. (2014) The 1 April 2014 Iquique, Chile, Mw 8.1 earthquake rupture sequence. *Geophysical research letters* **41**, 2014GL060238.

[61] Hayes, G. P, Bergman, E, Johnson, K. L, Benz, H. M, Brown, L, & Meltzer, A. S. (2013) Seismotectonic framework of the 2010 February 27 Mw 8.8 Maule, Chile earthquake sequence. *Geophysical Journal International* **195**, 1034–1051.

[62] Hayes, G. P, Herman, M. W, Barnhart, W. D, Furlong, K. P, Riquelme, S, Benz, H. M, Bergman, E, Barrientos, S, Earle, P. S, & Samsonov, S. (2014) Continuing megathrust earthquake potential in Chile after the 2014 Iquique earthquake. *Nature* **512**, 295–298.

[63] Melgar, D, Fan, W, Riquelme, S, Geng, J, Liang, C, Fuentes, M, Vargas, G, Allen, R. M, Shearer, P. M, & Fielding, E. J. (2016) Slip segmentation and slow rupture to the trench during the 2015, Mw8.3 Illapel, Chile earthquake. *Geophysical research letters* **43**, 2015GL067369.

[64] BBC News. (2015) Chile quake triggers mass evacuation and tsunami alert - BBC News (http://www.bbc.com/news/world-latin-america-34275783). Accessed: 2017-3-20.

[65] Hayes, G. P. (2015) USGS Finite Fault Model - Illapel, Chile Earthquake September 16, 2015 (http://earthquake.usgs.gov/earthquakes/eventpage/us20003k7a). Accessed: 2016-7-12.

[66] Barnhart, W. D, Murray, J. R, Briggs, R. W, Gomez, F, Miles, C. P. J, Svarc, J, Riquelme, S, & Stressler, B. J. (2016) Coseismic slip and early afterslip of the 2015 Illapel, Chile, earthquake: Implications for frictional heterogeneity and coastal uplift. *Journal of Geophysical Research, [Solid Earth]* **121**, 2016JB013124.

[67] U.S. Geological Survey. (2016) USGS Earthquake Catalog (http://earthquake.usgs.gov/earthquakes/search/). Accessed: 2016-12-8.

[68] Richards, M. A, Scheer, J. A, & Holm, W. A, eds. (2010) *Principles of Modern Radar: Basic Principles*. (SciTech Publishing) Vol. 1.

Appendix A

Radar Systems

In this appendix, I introduce the reader to the basics of radar systems, emphasizing those aspects relevant to interferometric synthetic aperture radar (InSAR). A much more comprehensive treatment can be found in [68].

A.1 Radar Basics

In the most basic sense, radar systems use electromagnetic (EM) waves to detect the presence of distant objects. The fundamental components of a radar system are: the transmitter, which generates the EM waves; the antenna, which broadcasts the waves into the atmosphere and captures their echoes reflected from the distant objects; the receiver, which amplifies the received waves and converts them to digital form; and the signal processor, which analyzes the digitized signals. When the same antenna is used for transmission and detection (a monostatic configuration), the transmitter and receiver must be electronically isolated to prevent damage of the extremely sensitive receiver circuits by the

high-power products of the transmitter. This is generally accomplished by means of a switch that disables the receiver during transmission.

In addition to simple target detection, radar systems can measure several aspects of target location, the simplest being range, or distance, to the target. Since EM waves propagate at the speed of light, c, measuring the two-way travel time, ΔT, of a radar pulse allows calculation of range, R:

$$R = \frac{c\Delta T}{2} \tag{A.1}$$

The physics of EM waves impose limits on the capabilities of radar systems, so a basic description of them is required to understand radar imaging.

A.1.1 Electromagnetic Waves

Electromagnetic waves are periodic oscillations of electric and magnetic fields that are orthogonal to one another and that propagate in a direction orthogonal to both. As with other wave phenomena, they are defined by their amplitude, E, wavelength, λ, wavenumber, k, frequency, f, period, T, and angular frequency, ω (figure A.1). These quantities are not independent of one another, but are related as follows:

$$\begin{aligned} k &= \frac{2\pi}{\lambda} \\ f &= \frac{1}{T} \\ \omega &= 2\pi f \\ \lambda f &= c \end{aligned} \tag{A.2}$$

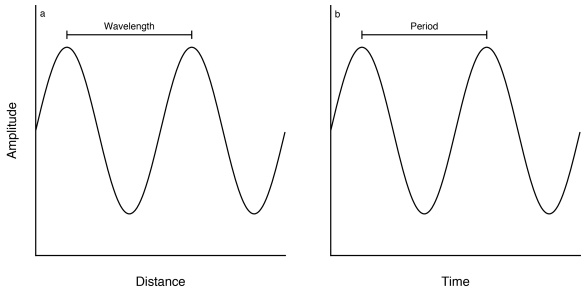


Figure A.1: The wavelength (a) and period (b) of a wave.

Table A.1: Radar frequency bands commonly used in SAR applications.

Band	Frequency Range
L	1 GHz to 2 GHz
S	2 GHz to 4 GHz
C	4 GHz to 8 GHz
X	8 GHz to 12 GHz

Radar systems typically operate in the 300 MHz to 35 GHz frequency range; for convenience, this range is divided into bands. Those commonly used in synthetic aperture radar (SAR) are listed in table A.1.

The wave's amplitude at location z and time t is expressed mathematically as

$$E = E_0 \cos(kz - \omega t + \phi) \tag{A.3}$$

where E_0 is the peak amplitude and ϕ is the initial phase, an arbitrary value. The entire argument of the cosine function is called the phase, and two waves are said to be in-phase when their phases are equal. Two waves which were in-phase can become out-of-phase (figure A.2) if they propagate over differing distances. When two or more waves of the same frequency are present at the same time and location, the resultant wave is the superposition of the individual waves. If the waves are in-phase, they will constructively interfere and the

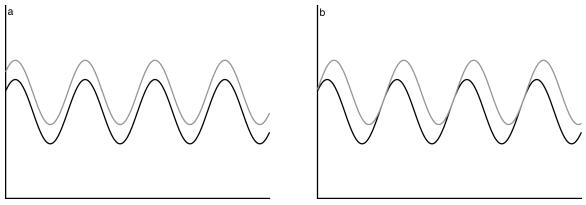


Figure A.2: In-phase (a) and out-of-phase (b) waves. The out-of-phase waves do not have coincident peaks and troughs.

resulting amplitude will equal the sum of the individual amplitudes; if out-of-phase, they will destructively interfere and the resulting amplitude will be less than their sum.

Another important quantity is intensity: the power per unit area of the wavefront; it is equivalent to power density, with units of $\mathrm{W\,m^{-2}}$. The wavefront defines the set of all points in space where the wave amplitude is equal at a given point in time. For an isotropic wave, where power is emitted equally in all directions, a situation not present in radar systems, the wavefront is a sphere, and adjacent spherical wavefronts are separated by the wavelength. The intensity, Q, is

$$Q = \frac{P}{4\pi R^2} \tag{A.4}$$

where P is the power and R is the radius of the sphere; thus intensity decreases with the square of distance. At a sufficiently large distance, the spherical wavefront can be approximated as a plane wave.

A.1.2 EM Wave Interactions with Matter

The EM waves in a radar system interact with matter in the antenna, in the atmosphere, and in the target. At the antenna, this interaction is in the form of diffraction,

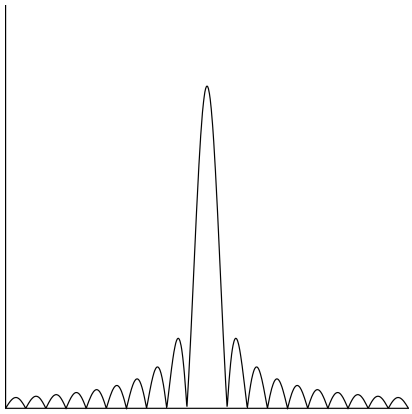


Figure A.3: A sinc function.

a type of interference. Huygens' principle states that a wavefront is composed of infinitely many point sources of wavelets which constructively interfere to produce further wavefronts. If an isotropic planar wavefront were to encounter a large plate with an opening, or aperture, of much greater diameter than the wavelength, many of these wavelet sources would pass through the aperture and there would be significant destructive interference beyond the bounds of the aperture, producing little to no diffraction and resulting in a shaped beam exiting the aperture with a distinct, high-power mainlobe and low-power sidelobes. Conversely, if the aperture diameter is much smaller than the wavelength, few wavelet sources pass through the aperture producing a great deal of diffraction, resulting in an isotropic wave exiting the aperture. In the case of a radar antenna, the aperture is not an opening in a plate, but an array of emitters. The shape of the emitted wave is a sinc, ($\frac{\sin x}{x}$) function (figure A.3), and the width of the mainlobe, which contains most of the wave's energy, is a function of the ratio of wavelength to aperture diameter: large apertures (relative to wavelength) produce narrow mainlobes, while small apertures produce wide mainlobes.

If an antenna consists of an array of individual emitters, each of which radiates EM waves in-phase with one another, then at some distance along a line normal to the antenna surface, each of those waves will have traveled the same distance and will remain in-phase. The amplitudes will constructively interfere, producing a plane wave whose amplitude is the sum of the individual amplitudes. At any angular distance away from this line, the path lengths from the individual emitters will no longer be equal, the waves will be out-of-phase, and their amplitudes will destructively interfere. By manipulating the phases at the emitters, a plane wave can be produced at any arbitrary angle off the normal; this is the basis for electronic scanning of phased array antennas. Phase shifter components associated with each emitter accomplish this manipulation.

Interactions between EM waves and the atmosphere include attenuation, refraction, and dispersion. Atmospheric attenuation, caused by absorption and scattering, is negligible at the frequencies typically used in Earth-observing radar systems. However, refraction of EM waves on passing through media with different dielectric properties is a significant source of interference in these systems. Refraction is quantified by the index of refraction, n, equal to $n = \frac{c}{v}$ where v is the phase velocity of the medium. The index of refraction of the troposphere, that layer of the atmosphere closest to Earths surface, generally decreases with increasing altitude. Thus, Snell's law predicts that EM waves should bend toward the surface in a radially stratified atmosphere. However, substantial lateral heterogeneity can exist due to variations in air temperature, atmospheric pressure, and partial pressure of water vapor. These variations can themselves produce variable refractions, resulting in errors in range calculation. Conditions which are especially problematic in this

respect include temperature inversion (increasing temperature with altitude) and moisture lapse (decreasing humidity with altitude). In addition, atmospheric turbulence can produce areas of varying index of refraction covering a small spatial extent. Measuring these variations in the index of refraction is difficult, therefore they are seldom accounted for in radar image processing.

Interactions between EM waves and targets produce scattering phenomena. The incident EM wave induces an electric field in the target, which then produces its own EM waves, as described by Maxwell's equations. The extent to which incident EM energy is re-radiated or absorbed depends on the dielectric properties of the target material. The manner in which the EM wave is scattered depends on the roughness, or variation in height relative to the EM wavelength, of the surface. If the surface is smooth, the angle of reflection is equal to the angle of incidence, resulting in specular scattering (figure A.4). Conversely, if the surface is rough, the energy is reflected at a variety of angles, resulting in diffuse scattering. The radar antenna can "capture" only that energy which is reflected towards it. Thus the radar "visibility" of a target is a function of its composition and shape as well as its size and orientation with respect to the antenna.

A.2 Radar Waveforms

Radar waveforms can be of several forms, and the waveform chosen is the primary factor determining the performance of the radar system. In this section, I will describe the characteristics of radar waveforms and their effects on radar system performance. I will then justify the use of pulse compression waveforms in radar imaging systems.

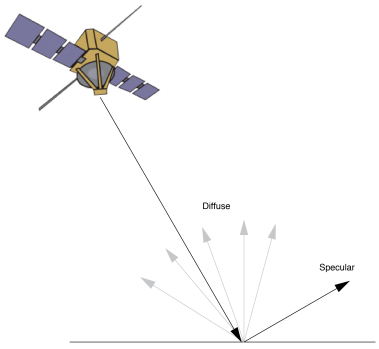


Figure A.4: Types of radar scattering. Specular scattering occurs when radar waves reflect from smooth surfaces; diffuse scattering occurs with rough surfaces.

A.2.1 Radar Waveform Characteristics

Radar waveforms can be classified as continuous wave or pulsed wave. In continuous wave operations, the transmitter is continually transmitting; this is very rarely used with monostatic configurations as there would be no means by which to isolate the receiver circuitry from the damaging effects of the high-power transmitted wave. Pulsed wave waveforms consist of EM waves of short duration, or pulse width, τ, during which the receiver is disabled, followed by longer intervals during which the receiver is enabled and the system can detect returned signals. The entire cycle of transmit time and "listening" time is called the pulse repetition interval (PRI) (figure A.5); the number of complete transmit/receive cycles in one second is the pulse repetition frequency (PRF), where

$$PRF = \frac{1}{PRI} \tag{A.5}$$

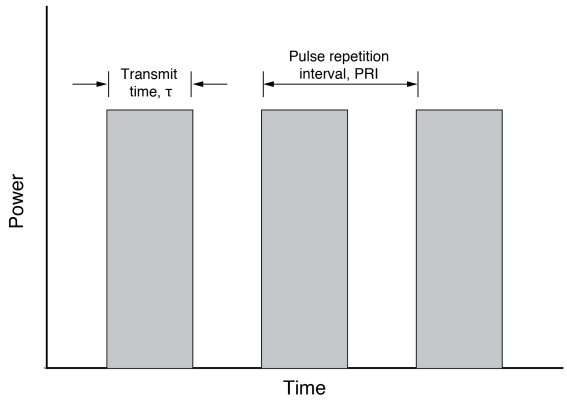


Figure A.5: Transmit time, τ, is the time period during which the antenna is transmitting. Pulse repetition interval, PRI, is the time period between the beginning of one pulse and the beginning of the next. Pulse repetition frequency, PRF, is $\frac{1}{PRI}$.

The fraction of the entire cycle during which the system is transmitting is the duty cycle, d_t, where

$$d_t = \frac{\tau}{PRI} = \tau \cdot PRF \tag{A.6}$$

The average power, P_{avg}, of the transmitted wave is

$$P_{avg} = P_t \cdot d_t = P_t \cdot \tau \cdot PRF \tag{A.7}$$

where P_t is the peak transmitted power.

A.2.2 Radar System Capabilities

The waveform characteristics described in the previous section control many aspects of radar system design and performance. The sampling interval of the analog-to-digital conversion (ADC) of the received signal must not exceed the pulse width or returned signals may be lost. Because time is directly related to range to target, each sample represents a

specific range increment, or range bin, that can be measured. Furthermore, if the two-way travel time to a target exceeds the PRI, the signal returned from a first transmitted pulse will be received after a second pulse is transmitted, creating a range ambiguity. Therefore, the PRI must be long enough, and therefore the PRF low enough, to allow return from all targets of interest; the unambiguous range, R_{ua} is

$$R_{ua} = \frac{c}{2 \cdot PRF} \tag{A.8}$$

Virtually all modern radar systems are coherent, meaning that both the amplitude and phase of the received signal are measured; noncoherent systems measure only amplitude. The additional information provided by the phase is what allows use of radar for measurement of surface deformation. Because the relative phase of the received signal compared to the transmitted signal is determined by the path length, a change in phase between subsequent measurements, for example before and after an earthquake, indicates surface motion in the intervening time (figure A.6), assuming there has been no change in scatterer characteristics.

Doppler Effect

In addition to phase change, relative motion between the radar and the target produces a Doppler shift in the frequency of the wave. The Doppler effect is well known and commonly observed with acoustic waves from moving vehicles, for example, an ambulance siren; the sound from an approaching vehicle is shifted to higher frequencies, while that from a receding vehicle is shifted to lower frequencies. In a radar system, where the radial velocity, v_r, between the radar and the target (i.e., the velocity along the line of sight) is

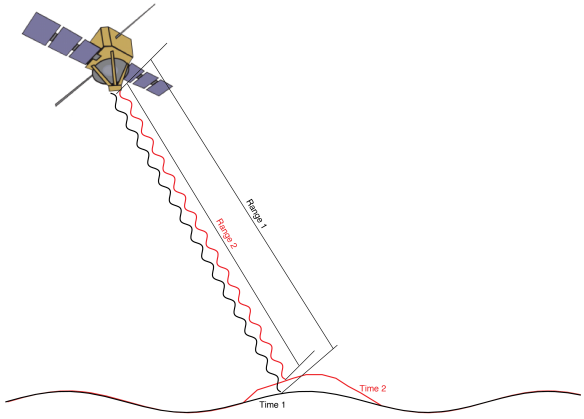


Figure A.6: Detection of surface deformation by range change. Uplift of the ground surface in the interval between acquisitions reduces the range between antenna and surface. This is detected by the phase difference between the two waves.

much less than the speed of light, the Doppler frequency shift, f_d, is given by

$$f_d = \frac{2v_r}{\lambda} \tag{A.9}$$

Analogous to range ambiguity, there may be ambiguity in Doppler frequency measurement. Because the Doppler shift is sampled at the PRF, a PRF that is too low will lead to Doppler ambiguity. The maximum Doppler shift that can be measured unambiguously, as predicted by the Nyquist sampling theorem, is $f_{d_{max}} = \pm\frac{PRF}{2}$. Consequently, range and Doppler measurements impose conflicting demands on radar system design: range ambiguity is avoided by using a low PRF, while Doppler ambiguity is avoided by using a high PRF. In practice, most radar imaging systems use a relatively low PRF.

Doppler frequency is measured by performing spectral analysis of the received signal in every range bin. As is described in detail in section 2.2.1, a SAR system uses Doppler analysis to improve azimuth resolution. Resolution describes the system's ability

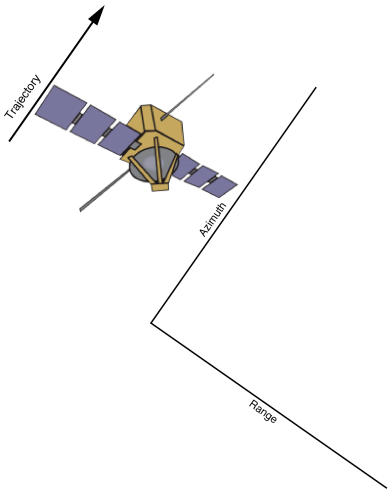


Figure A.7: Radar geometry: range is perpendicular to the satellite's trajectory; azimuth is parallel.

to distinguish closely spaced objects as separate; in radar imaging, resolution is defined along both the range dimension, perpendicular to the satellite's flight path, and the azimuth dimension, parallel to the satellite's flight path (figure A.7).

Resolution

In the range dimension, two targets can be resolved if the two returned signals from a single transmitted pulse can be distinguished as two separate pulses; if the two pulses overlap in time, they cannot be distinguished, so the targets cannot be resolved. These two possibilities are separated by the case where the two returned pulses abut one another, or when the separation between the targets, ΔR, is equal to

$$\Delta R = \frac{c\tau}{2} \tag{A.10}$$

Thus the range resolution for an unmodulated pulse is ΔR. Clearly, better resolution can be achieved by using a shorter duration pulse; however, as described in equation A.7, the average transmitted power is proportional to pulse width. A pulse that is too short will contain too little power to be detected in the presence of noise after two-way propagation; in other words, the signal-to-noise ratio (SNR) will be too low. The solution to this dilemma is to use a pulse compression waveform, described in the next section; it produces a short pulse with high power. Doppler resolution depends on the ability to distinguish peaks in the Doppler spectrum. When the Doppler spectrum has the form of a sinc function, discrimination of two peaks depends on the widths of the mainlobes, which are inversely proportional to the dwell time, or the total time to transmit all the pulses used for the spectral analysis.

A.2.3 Pulse Compression Waveforms

Pulse compression waveforms are used in radar imaging to resolve the conflicting demands of range resolution and SNR. They consist of a short-duration pulse with a wide bandwidth. The short duration maintains fine range resolution, while the wide bandwidth maintains sufficient energy to produce sufficient SNR. While there are several methods for producing a pulse compression waveform, the most commonly used in radar imaging is linear frequency modulation (LFM), also known as the chirp pulse, in which the frequency sweeps through a bandwidth B over the course of the pulse, τ (figure A.8); the ramp rate or sweep rate is $\frac{B}{\tau}$. After generation of the chirp waveform, it is upconverted to the carrier

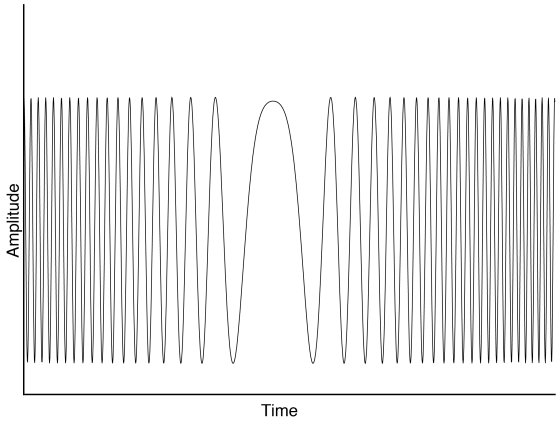


Figure A.8: A linear frequency modulated waveform (LFM) sweeps through a range of frequencies over the course of a pulse.

frequency, f_0, so that the final transmitted waveform is

$$x(t) = A\cos\left(2\pi f_0 t + \pi\frac{B}{\tau}t^2\right), \frac{-\tau}{2} \leq t \leq \frac{\tau}{2} \tag{A.11}$$

The returned signal can be convolved with a matched filter (section 2.2.4) to maximize the SNR and produce a pulse compressed signal. I show in section 2.2.5 that when a pulse compression waveform is used, the range resolution becomes $\Delta R = \frac{c}{2B}$.

A.3 Radar Range Equation

The most significant source of instrument noise in a radar imaging system is thermal white noise in the receiver circuits. Thus the amplitude of the received signal is the main factor determining SNR; it can be modeled with the radar range equation (RRE). In this section, I will develop the radar range equation in a series of steps in order to motivate understanding of the factors that affect SNR.

As stated in equation A.4, the intensity, or power density, Q_i, of an isotropic wave incident on a target at a distance R from an antenna is

$$Q_i = \frac{P_t}{4\pi R^2} \tag{A.12}$$

where P_t is the peak transmitted power. In reality, radar antennas are not isotropic, and the directivity of a transmission antenna, combined with power losses within the antenna subsystem, are represented by a quantity called gain, G_t. This modifies the transmitted power such that

$$Q_i = \frac{P_t G_t}{4\pi R^2} \tag{A.13}$$

After interacting with a target, some of the incident radar energy is re-radiated, or reflected, toward the antenna. As discussed in section A.1.2, the proportion of the incident energy that is reflected toward the antenna depends on the target's size, shape, composition, and orientation with respect to the antenna; these factors are summarized in the radar cross-section (RCS), denoted σ, of the target, with units of m^2. Thus the power reflected by the target, P_{refl}, is

$$P_{refl} = Q_i \sigma = \frac{P_t G_t \sigma}{4\pi R^2} \tag{A.14}$$

This reflected power propagates toward the antenna, diminishing with the square of distance, such that the received intensity, Q_r is

$$Q_r = \frac{P_{refl}}{4\pi R^2} = \frac{P_t G_t \sigma}{(4\pi)^2 R^4} \tag{A.15}$$

To calculate received power, P_r, from received power density, Q_r, it is necessary to multiply by the effective area, A_e, of the receiving antenna, a term which incorporates antenna

efficiency and physical area, and is represented as gain, G_r, such that

$$A_e = \frac{G_r \lambda^2}{4\pi} \tag{A.16}$$

where λ is radar wavelength. Thus the final (idealized) form of the RRE is

$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R^4} \tag{A.17}$$

Realistically, signal loss, e.g., due to component resistive losses or signal processing losses, reduces the actual received power, such that

$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R^4 L_s} \tag{A.18}$$

where L_s represents the total of these losses.

From this form of the RRE, the SNR can be calculated for a coherent radar system:

$$SNR = \frac{P_t G_t G_r \lambda^2 \sigma n_p}{(4\pi)^3 R^4 P_n L_s} \tag{A.19}$$

where n_p represents the number of pulses included in the coherent integration, and P_n represents noise power and equals

$$P_n = kT_0FB \tag{A.20}$$

where k is Boltzmann's constant, T_0 is standard temperature, F is the noise figure (a property inherent to the radar system), and B is the receiver bandwidth. When using pulse compression waveforms, it is customary to use a form of the RRE that includes average transmitted power rather than peak transmitted power. The dwell time, or the time to transmit the pulses included in the coherent integration, is equal to

$$T_d = n_p \cdot PRI = \frac{n_p}{PRF} \tag{A.21}$$

The pulse compressed SNR is equal to the uncompressed SNR multiplied by the pulse width, τ, and the bandwidth, B, so that:

$$SNR_{pc} = SNR_u \cdot \tau \cdot B = \frac{P_{avg} T_d G_t G_r \lambda^2 \sigma}{(4\pi)^3 R^4 k T_0 F L_s} \tag{A.22}$$

where P_{avg} is as calculated in equation A.7.

In radar imaging applications, as opposed to detection systems, the "target" of interest is typically not a single, discrete target, such as a vehicle, but all the individual scatterers within the area illuminated by the radar beam. In other radar applications, these scatterers would be considered unwanted "clutter". Each scatterer in the illuminated area has its own RCS and is at a slightly different range from the radar. Hence, the received signal for each range bin is the vector sum of the complex signals for each scatterer within that range bin. Thus the RRE must be modified such that σ is now

$$\sigma = \sigma^0 \cdot A \tag{A.23}$$

where σ^0 is the surface reflectivity, or RCS per unit area, and A is the illuminated area, equal to

$$A = R \Delta R \theta_3 \sec \delta \tag{A.24}$$

where θ_3 is the half-power beamwidth and δ is the grazing angle (figure A.9). Therefore,

$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma^0 \Delta R \theta_3 \sec \delta}{(4\pi)^3 R^3 L_s} \tag{A.25}$$

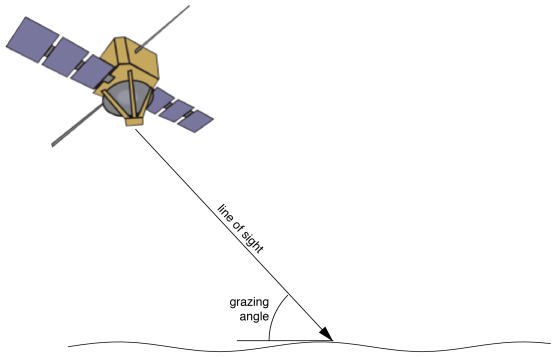


Figure A.9: Grazing angle is between the horizontal and the line-of-sight vector.

A.4 Radar Signal Processing

In this section, I will describe how a complex signal is detected in a radar system, and how that signal is used to measure the two most important quantities in radar imaging: range and Doppler frequency.

A.4.1 Coherent Signal Detection

If a sinusoidal pulse of amplitude A is transmitted at a carrier frequency f_0, with pulse width τ centered at time $t = 0$, and with initial phase ϕ, it can be represented as

$$x(t) = A\cos(2\pi f_0 t + \phi) = Re\Big\{Ae^{j(2\pi f_0 t+\phi)}\Big\} = Re\Big\{Ae^{j\phi}e^{j2\pi f_0 t}\Big\}, \frac{-\tau}{2} \leq t \leq \frac{\tau}{2} \quad \text{(A.26)}$$

where j is the imaginary number, $\sqrt{-1}$. In other words, it has a complex amplitude $Ae^{j\phi}$. If this pulse is then reflected from a target at range R_0, the received pulse will be

$$\begin{aligned} y(t) &= x\left(t - \frac{2R_0}{c}\right) \\ &= A\prime \cos\left(2\pi f_0\left(t - \frac{2R_0}{c}\right) + \phi\right) \\ &= A\prime \cos\left(2\pi f_0 t + \phi - \frac{4\pi R_0}{\lambda}\right) \\ &= Re\left\{A\prime e^{j(\phi - \frac{4\pi R_0}{\lambda})} e^{j2\pi f_0 t}\right\}, \frac{-\tau}{2} + \frac{2R_0}{c} \leq t \leq \frac{\tau}{2} + \frac{2R_0}{c} \end{aligned} \tag{A.27}$$

where $A\prime$ is the amplitude predicted by the RRE. In addition to the time delay of $\frac{2R_0}{c}$, there is a phase shift of $\frac{-4\pi R_0}{\lambda}$. Thus a change in range of $\frac{\lambda}{2}$ is sufficient to cause a phase shift of 2π radians.

Coherent detection is accomplished by splitting this received signal into two paths (figure A.10). The first is mixed with a reference signal, $2\cos(2\pi f_0 t)$ with the result being the sum of $A\prime\cos(2\pi(2f_0)t + \phi\prime)$ and $A\prime\cos\phi\prime$ where $\phi\prime$ is the shifted phase of the received signal. Passage through a low pass filter leaves only the second term, a constant pulse of amplitude $A'\cos\phi\prime$ representing the real part of the complex amplitude $A\prime e^{j\phi\prime}$. This is the output of the in-phase or I channel of the coherent detector.

At the same time, the second path carries the received signal to a second mixer with a reference signal $2\sin(2\pi f_0 t)$, which is 90° out of phase from the first reference signal. Similar to the first channel, the result is the sum of $A\prime\sin(2\pi(2f_0)t + \phi\prime)$ and $A\prime\sin\phi\prime$, with the former being removed by passage through a low pass filter. The latter represents the imaginary part of the complex amplitude and is the output of the quadrature or Q channel.

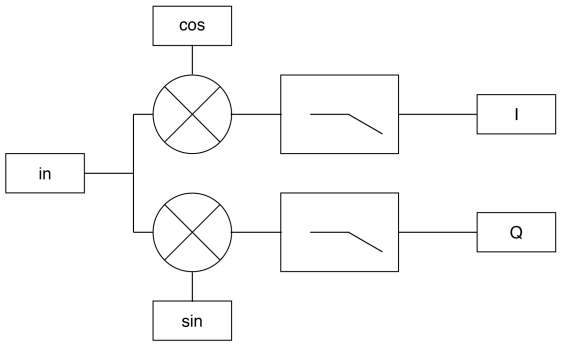


Figure A.10: Coherent signal detection is accomplished by splitting the input signal into two paths. One is mixed with a cosine reference signal to produce the I channel signal; the other is mixed with a sine reference signal to produce the Q channel signal.

Together, the two channels resolve the sign ambiguity which would result from calculating the phase from one channel alone.

In the case of a chirp waveform, the return signal is first mixed to baseband, so that the bandwidth is centered at zero frequency, and coherent detection results in a complex signal $x(t) = Ae^{j\pi\frac{B}{\tau}t^2}$ with a time-varying phase $\phi(t) = \pi\frac{B}{\tau}t^2$.

A.4.2 Measuring Range

The output of the detector is sampled at discrete time intervals over the period between pulses, producing a vector of complex voltage values. Each sample represents a range interval, or range bin. Measuring the time delay at which a return signal is received allows estimation of range to the corresponding scatterer, as described in section A.1.

A.4.3 Doppler Analysis by Fourier Transform

As stated in section A.2.2, Doppler processing is performed by spectral analysis, in which a continuous time-domain signal, $x(t)$, is converted to a frequency-domain representation, either $X(\omega)$ or $X(f)$, using the Fourier transform (FT):

$$X(\omega) = \mathfrak{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}dt \tag{A.28}$$

$$X(f) = \mathfrak{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \tag{A.29}$$

The inverse Fourier transform (IFT) back to the time domain is:

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)e^{j\omega t}d\omega = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}df \tag{A.30}$$

The simplest waveform would be an infinite-length complex sinusoid with frequency f_0:

$$x(t) = Ae^{j2\pi f_0 t} \tag{A.31}$$

Its FT is:

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} Ae^{j2\pi f_0 t}e^{-j2\pi ft}dt \\ &= A\int_{-\infty}^{\infty} e^{-j2\pi(f-f_0)t}dt \\ &= A\cdot\delta_D(f-f_0) \end{aligned} \tag{A.32}$$

where $\delta_D(f)$ is the Dirac delta function; the resulting spectrum consists of a single line (figure A.11) at the frequency f_0 and with amplitude A. If the sinusoid is real-valued, the spectrum will consist of two lines, at $\pm f_0$ with an amplitude of $\frac{A}{2}$.

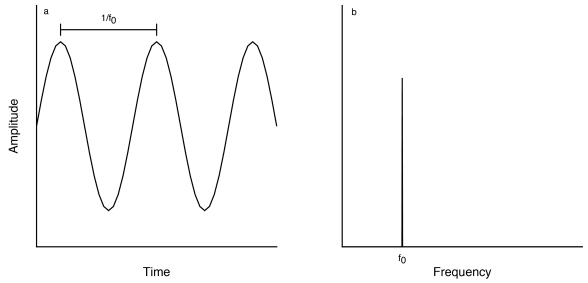


Figure A.11: A sinusoidal wave with a single frequency (a) and its spectrum (b).

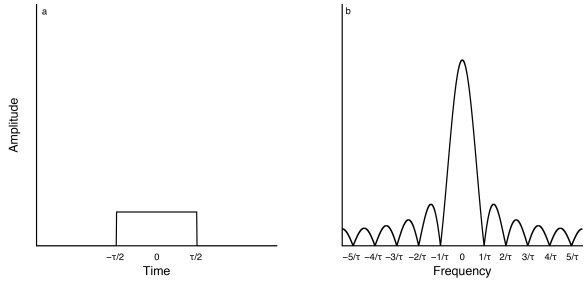


Figure A.12: A pulsed wave (a) and its spectrum, corresponding to a sinc function (b).

For a single pulse of width τ, represented by a boxcar function,

$$p_\tau(t) = \begin{cases} A, \frac{-\tau}{2} \leq t \leq \frac{\tau}{2} \\ 0, otherwise \end{cases} \tag{A.33}$$

and the FT is

$$P_\tau(f) = A\tau \cdot \text{sinc}(\pi f \tau) \tag{A.34}$$

The null values of the sinc function occur at integer multiples of $\frac{1}{\tau}$ (figure A.12). The width of the mainlobe can be defined in several ways, but in any case is inversely proportional to pulse length.

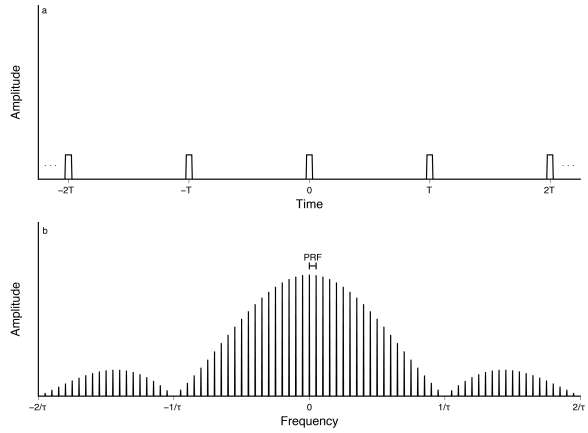


Figure A.13: A waveform consisting of an infinite train of pulses (a) and its spectrum, consisting of a series of lines bounded by a sinc function (b).

An infinite train of such pulses, with each pulse separated by a PRI of T, can be represented as a convolution of the single pulse and the Dirac delta function:

$$p_I(t) = p_\tau(t) * \sum_{n=-\infty}^{\infty} \delta_D(t - nT) \tag{A.35}$$

Its FT is

$$P_I(f) = (A\tau \cdot \mathrm{sinc}(\pi f \tau))(\frac{1}{T} \sum_{k=-\infty}^{\infty} \delta_D(f - k \cdot PRF)) \tag{A.36}$$

$$= \frac{A\tau}{T} \sum_{k=-\infty}^{\infty} \mathrm{sinc}(\pi \tau k \cdot PRF) \cdot \delta_D(f - k \cdot PRF) \tag{A.37}$$

This spectrum is a series of spectral lines of zero bandwidth, separated by PRF, with amplitudes bounded by a sinc envelope whose nulls occur at integer multiples of $\frac{1}{\tau}$ (figure A.13).

Truncating the infinite pulse train to a finite sequence of pulses of duration T_d, called the dwell time or coherent processing interval, can be represented by the multiplica-

tion of the infinite pulse train and a boxcar function:

$$p_F(t) = p_I(t) \cdot p_{T_d}(t)$$

$$p_{T_d}(t) = \begin{cases} 1, \frac{-T_d}{2} \leq t \leq \frac{T_d}{2} \\ 0, otherwise \end{cases} \tag{A.38}$$

Its FT is:

$$P_F(f) = \frac{AT_d\tau}{T} \sum_{k=-\infty}^{\infty} \text{sinc}(\pi\tau k \cdot PRF)\,\text{sinc}(\pi(f - k \cdot PRF)T_d) \tag{A.39}$$

The spectral lines are now sinc functions, with the amplitudes bounded by a larger sinc function.

Finally, the amplitude of the finite pulse train can be modulated by multiplying by a cosine function of frequency f_0:

$$x(t) = p_F(t) \cdot \cos(2\pi f_0 t) \tag{A.40}$$

Its FT is:

$$X(f) = \frac{AT_d\tau}{T}\Bigg\{ \sum_{k=-\infty}^{\infty} \text{sinc}(\pi\tau k \cdot PRF)\,\text{sinc}(\pi(f - f_0 - k \cdot PRF)T_d) + \sum_{k=-\infty}^{\infty} \text{sinc}(\pi\tau k \cdot PRF)\,\text{sinc}(\pi(f + f_0 - k \cdot PRF)T_d) \Bigg\} \tag{A.41}$$

The rather complicated spectrum consists of two sets of sinc functions, centered at $\pm f_0$. The spectral lines are also sinc functions, separated by PRF, and with a bandwidth proportional to $\frac{1}{T_d}$.

This progression illustrates the reciprocal spreading relationship of Fourier transformation: the wider a function in the time domain, the narrower it is in the frequency domain, and vice versa. This relationship is exceedingly important when considering the

issue of Doppler resolution. The spectrum of the return signal from a single target with a Doppler frequency of f_d will have its two sets of sinc functions centered at $-f_0 - f_d$ and $f_0 + f_d$; the spectrum of the return signal from multiple scatterers in an illuminated area, each with a unique Doppler frequency, will be the superposition of multiple sets of these sinc functions. Discriminating each set is possible only if the sinc functions that make up the spectral lines are narrow enough so as not to overlap. Because their bandwidth is proportional to $\frac{1}{T_d}$, a longer dwell time, or equivalently, a longer pulse train, provides better Doppler resolution.

The previous discussion assumed that the function being transformed, $x(t)$, was a continuous function. In fact, the received signal is sampled by an analog-to-digital converter (ADC). Fourier analysis of a signal sampled in the time domain requires the discrete-time Fourier transform (DTFT). Just as the sinc function is the characteristic result of Fourier transformation of a single pulse, the aliased sinc or asinc function is the characteristic result of the DTFT. Its shape is similar to that of the sinc function (figure A.3), but the width of the mainlobe is inversely proportional to the number of samples. This is another example of the reciprocal spreading relationship of the Fourier transform: more samples results in a narrower mainlobe. The result of the DTFT is then sampled again in the frequency domain, resulting in the discrete Fourier transform (DFT). This is implemented using one of the fast Fourier transform (FFT) algorithms; the time-domain signal can be recovered by inverse FFT (IFFT).

The moving satellite imparts a Doppler shift on stationary ground targets, but because of the finite extent of the antenna footprint, that shift is not identical for every

illuminated target. The bandwidth of Doppler frequencies for the illuminated area can be calculated by extending the Doppler frequency equation (equation A.9) to

$$f_d = \frac{2v}{\lambda} \sin \psi \tag{A.42}$$

where ψ is the squint angle, or the angle between the normal to the satellite's velocity vector and the antenna's line-of-sight (LOS) direction to the target (figure A.14). Thus if ψ is 0°, as it would be in a side-looking airborne radar (SLAR) configuration, targets along the LOS direction have no Doppler shift. Conversely, if the squint angle is not 0°, or if the target is not along the LOS direction, the target will have a non-zero Doppler shift. The Doppler bandwidth, or the difference in Doppler frequencies between two targets at the same range, where one is at the leading edge of the antenna beam footprint and the other is at the trailing edge, is

$$\begin{aligned} B_D &= \frac{2v}{\lambda}\big(\sin(\psi + \frac{\theta_3}{2}) - \sin(\psi - \frac{\theta_3}{2})\big) \\ &= \frac{4v}{\lambda} \sin(\frac{\theta_3}{2}) \cos \psi \\ &\approx \frac{2v\theta_3}{\lambda} \cos \psi \end{aligned} \tag{A.43}$$

The center frequency of the Doppler bandwidth is the Doppler centroid, and this value is crucial to the image formation process.

Because Doppler shift is proportional to the sine of ψ, or equivalently, to the cosine of $90^\circ - \psi$, every point on the surface of a cone with a half-angle equal to $90^\circ - \psi$ and centered on the satellite's velocity vector has the same Doppler frequency. For a satellite in level flight, the intersection of this cone with the surface of Earth is a hyperbola called the isodoppler contour. Similarly, every point on the surface of a sphere of radius R centered on

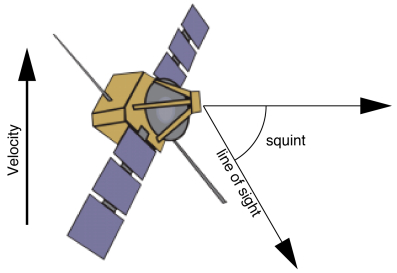


Figure A.14: Squint is the angle between the normal to the satellite's velocity vector and the line-of-sight vector.

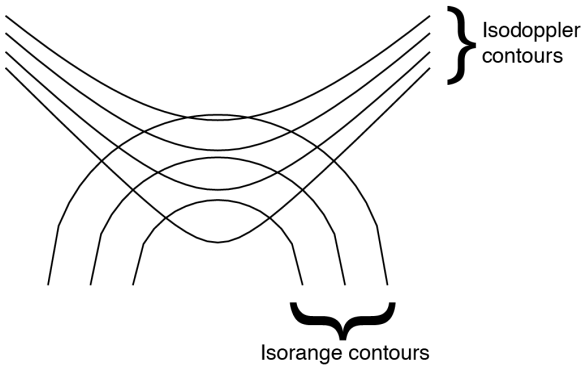


Figure A.15: Isodoppler and isorange contours. Every point on the surface is located at a unique intersection of the two.

the satellite has the same range. The intersection of this sphere with the surface of Earth is a circle called the isorange contour. Therefore, every scatterer can be located by its unique combination of Doppler frequency and range, and lies at the intersection of an isodoppler contour and an isorange contour (figure A.15).