I. The Effect of Volcanic Aerosols on Ultraviolet Radiation in Antarctica

II. A Novel Method for Enhancing Subsurface Radar Imaging using Radar Interferometry

Thesis by

Steven R. Tsitas

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1998

(submitted January 7, 1998)

Abstract

The theory of radiative transfer is used to explain how a stratospheric aerosol layer may, for large solar zenith angles, increase the flux of UV-B light at the ground. As previous explanations are heuristic and incomplete, I first provide a rigorous and complete explanation of how this occurs. I show that an aerosol layer lying above Antarctica during spring will decrease the integrated daily dose of biologically weighted irradiance, weighted by the erythema action spectrum, by only up to 5%. Thus after a volcanic eruption, life in Antarctica during spring will suffer the combined effects of the spring ozone hole and ozone destruction induced by volcanic aerosols, with the latter effect only slightly offset by aerosol scattering.

I extend subsurface radar imaging by considering the additional information that may be derived from radar interferometry. I show that, under the conditions that temporal and spatial decorrelation between observations is small so that the effects of these decorrelations do not swamp the signature expected from a subsurface layer, the depth of burial of the lower surface may be derived. Also, the echoes from the lower and upper surfaces may be separated. The method is tested with images acquired by SIR-C of the area on the Egypt/Sudan border where buried river channels were first observed by SIR-A. Temporal decorrelation between the images, due to some combination of physical changes in the scene, changes in the spacecraft attitude and errors in the processing by NASA of the raw radar echoes into the synthetic aperture radar images, swamps the expected signature for a layer up to 40 meters thick. I propose a test to determine whether or not simultaneous observations are required, and then detail the radar system requirements for successful application of the method for both possible outcomes of the test. I also describe in detail the possible applications of the method. These include measuring the depth of burial of ice in the polar regions of Mars, enhancing the visibility of buried features and, most importantly, the ability to map soil moisture in arid regions of the earth at high spatial resolution.

Table Of Contents

List of Tables

Chapter I

Summary

In Chapter II I use the theory of radiative transfer to explain how a stratospheric aerosol layer may, for large solar zenith angles, increase the flux of UV-B (290-320 nm) light at the ground. I also determine the biological significance of this numerically predicted, but not yet observed, effect. The motivation behind this research is that volcanic eruptions can inject large amounts of aerosol into the stratosphere, increasing the flux of UV-B light at the surface for large solar zenith angles with possibly negative biological consequences.

This predicted enhancement of UV-B light at the ground occurs even though aerosols increase the reflection of sunlight to space. As previous explanations are heuristic and incomplete, I first provide a rigorous and complete explanation of how this occurs. This effect makes Antarctica during spring the most susceptible place on earth to the scattering effect of volcanic aerosols, due to the combined effect of the spring ozone hole and the large solar zenith angles characteristic of this time of year.

I show that an aerosol layer lying above Antarctica during spring will decrease the integrated daily dose of biologically weighted irradiance, weighted by the erythema action spectrum, by only up to 5%. Hence the effects of any significant destruction of ozone induced by volcanic aerosols will not be offset by aerosol scattering. Thus after a volcanic eruption, life in Antarctica during spring will suffer the combined effects of the spring ozone hole and ozone destruction induced by volcanic aerosols, with the latter effect only slightly offset by aerosol scattering.

In Chapter III I extend subsurface radar imaging by considering the additional information that may be derived from radar interferometry. I show that, *under the conditions that temporal and spatial decorrelation between observations is small so that the effects of these decorrelations do not swamp the signature expected from a subsurface layer*, with at least three radar images of the same scene taken from parallel imaging tracks, the depth of burial of the lower surface may be derived. Also, the echoes from the lower and upper surfaces, combined in a single radar image, may be separated.

The reason I carried out this research is that this method has a number of important practical applications that are unique in being based on the way the method gives direct information about the subsurface environment, such as the depth and the strength of the lower and upper echoes, using the geometry of the situation. Normally, complicated models that predict the strength of the radar return at various incidence angles, polarization combinations and wavelengths, and that contain, say, the depth of the subsurface layer as a variable, are used to obtain such information, which is obviously a very indirect way of doing so that is subject to many more assumptions the interferometric method uses.

The applications of the method, unique by not relying on complicated models with many variables, include measuring the depth of burial of ice in the polar regions of Mars, enhancing the visibility of buried features and two independent applications for soil moisture mapping, at the high spatial resolution afforded by synthetic aperture radar, in arid regions of the Earth.

I present a method of solution, validated by extensive simulations, for both the depth of burial and the separation of echoes that is optimal, in the least squares sense that the values found are those from which the data would most likely come. Furthermore, this solution is obtained without requiring knowledge of "tie points"; locations of known altitude usually required to use radar interferometry for practical purposes.

I test the method using three images acquired from parallel orbital tracks by the Shuttle Imaging Radar (SIR-C) in 1994 of the eastern edge of the Selima Sand Sheet on the Egypt/Sudan border (Figure III-21). This area intersects with that imaged by SIR-A in 1981 when buried river channels were first observed [*McCauley et al.*, 1982], and is the best possible data available to confirm the method. However, I show that temporal decorrelation between the images swamps the expected signature for a layer up to 40 meters thick.

This temporal decorrelation, between observations acquired on consecutive days, is due to some combination of physical changes in the scene, changes in the spacecraft attitude and errors in the processing by NASA of the raw radar echoes into the synthetic aperture radar images. Assuming the elimination of the last factor, successful application of this method in the future may or may not require simultaneous observations to avoid the effects of physical changes in the scene. In section III.17 I propose a test to determine whether or not simultaneous observations are required, and then detail the radar system requirements for successful application of the method for both possible outcomes of the test. I also describe in detail the possible applications of the method, most importantly the ability to map soil moisture in arid regions of the earth at high spatial resolution.

Chapter II

The Effect of Volcanic Aerosols on Ultraviolet Radiation in Antarctica

II.1 Introduction

Recent numerical solutions of the equation of radiative transfer [*Michelangeli et al.*, 1989; *Michelangeli et al.*, 1992; *Tsay and Stamnes*, 1992] indicate that, at large solar zenith angles, scattering by a stratospheric aerosol layer can increase the flux of ultraviolet light reaching the surface. An explanation of this effect, which has not yet been observed, has been offered by *Davies* [1993]. Basically, at large solar zenith angles, scattering from a stratospheric aerosol layer provides a shortcut for ultraviolet light, which is strongly absorbed by ozone, to reach the ground. This scattered light suffers much less attenuation than the direct solar beam, which travels a long slant path through the atmosphere. Hence there is an increase in flux at the ground, even though some of the scattered light escapes to space, increasing that flux also. *Davies* uses a model of the atmosphere which treats the aerosol as a single scattering layer, but neglects Rayleigh scattering by the atmosphere in any approximation, to demonstrate that this mechanism can lead to an enhancement of surface flux. However, at ultraviolet wavelengths the atmosphere is strongly Rayleigh scattering (the Rayleigh scattering optical depth in the zenith direction is 1 at 305 nm), and so it has not been shown if the mechanism proposed by *Davies* is the one responsible for surface flux enhancement in a realistic, Rayleigh scattering atmosphere.

II.2 Explanation of the Mechanism of Enhancement

To investigate this effect, I use a 30-stream radiative transfer model of the atmosphere that is based on the numerical solution of the equation of radiative transfer as formulated by *Michelangeli et al.* [1992], which is suitable for treating an aerosol layer. Our numerical model includes Rayleigh scattering by air, absorption by ozone, aerosol scattering and absorption and assumes a Lambert surface. The attenuation of the direct solar beam is treated using spherical geometry, while the multiple scattering is calculated for a plane-parallel atmosphere [*Froidevaux et al.*, 1985]. Rayleigh scattering and ozone absorption cross sections, and the solar flux incident upon the top of the atmosphere, are taken from the *World Meteorological Organization* [1986] report. Volcanic aerosols are assumed to scatter with the Henyey-Greenstein phase function [*Henyey and Greenstein*, 1941], with an asymmetry parameter of 0.75 and single scattering albedo of 0.99 [*Vogelmann et al.*, 1992].

I validated our model by calculating the diffuse reflection and transmission functions for light incident upon a plane-parallel slab of scatterers, and comparing the results with the tables published by *van de Hulst* [1980], for 120 widely varied sets of values of the relevant parameters. The 120 calculated diffuse transmission functions differ from *van de Hulst's* tabulated values by an average of 0.13% and a maximum of 1.5%. Corresponding values for the 120 diffuse reflection functions are 0.33% and 5.6%, respectively, with all the differences greater than 0.60% occurring when the incident beam is normal to the slab, and the diffuse reflection function is calculated in the direction back along the incident beam. Since I only use our model to calculate the flux transmitted to the surface at large solar zenith angles, the results presented in this paper have a fractional error of less than 1-2%.

I have used our numerical model to calculate the factor, E, by which the surface flux is enhanced by a volcanic aerosol layer, assumed to spread uniformly from 25-26 km altitude [*DeLuisi et al.*, 1983] with an optical depth of 0.4 [*Valero and Pilewskie*, 1992]. This factor, calculated for a solar zenith angle of 80°, is shown in the second column of Table II-1. A surface albedo of 0.05 was used in the calculations, and I assumed that the atmosphere is the *U.S. Standard Atmosphere* [1976], with the amount of ozone scaled to 200 DU. The first column of Table II-1 lists the approximate center wavelengths of the wavelength intervals from the *World Meteorological Organization* [1986] report. τ_{O3} is the optical depth of the atmosphere, in the zenith direction, due to ozone absorption. The fourth and fifth columns show the percentage of downward traveling light that is due to direct (unscattered) sunlight at 26 km altitude and at the surface, respectively, in the clear atmosphere. The last column gives the enhancement factor, E, for surface irradiance calculated using *Davies'* [1993] model (the fraction of τ_{O3} beneath the aerosol layer is 0.672, independent of wavelength).

I do not make the comparison with *Davies'* model to show its inaccuracy, for *Davies'* model is obviously intended to be qualitative. But its purpose is to demonstrate the correctness of his intuitive explanation for the mechanism responsible for the enhancement in a realistic atmosphere. The point I wish to make by the comparison is that *Davies'* model fails to reproduce even qualitatively the decrease in enhancement predicted at the shortest ultraviolet wavelengths. Also, since *Davies'* model does not include Rayleigh scattering by the atmosphere, it assumes that in the clear atmosphere the radiation at the surface is just the direct radiation, attenuated by ozone absorption. The penultimate column of Table II-1 shows that this is not a valid approximation throughout the wavelength range of the enhancement. For these two reasons I claim that *Davies'* explanation for the enhancement has not been shown to apply to a realistic, Rayleigh scattering atmosphere.

To rigorously demonstrate the mechanism responsible for enhancement in a realistic, Rayleigh scattering atmosphere, I plot the direct, diffuse and total irradiance as a function of altitude both with and without an aerosol layer. Two diffuse irradiances are plotted. One (dashed curve) is the cosine weighted integration of the diffuse intensity over the upward hemisphere, representing the flux of diffuse light traveling up through the atmosphere. The other (dotted-dashed curve) is the cosine weighted integration of the diffuse intensity over the downward hemisphere, representing the flux of diffuse light traveling down through the atmosphere. The net downward irradiance is the direct irradiance plus the diffuse irradiance integrated over the downward hemisphere minus the diffuse irradiance integrated over the upward hemisphere. I consider the wavelength range 303.0-307.7 nm, for which the surface flux enhancement factor is 2.22, as given in Table II-1. The upward diffuse, downward diffuse, direct and net downward irradiance as a function of altitude in the aerosol free atmosphere are shown in Figure II-1. The introduction of an aerosol layer of optical depth 0.4, extending from 25-26 km altitude, causes the upward diffuse, downward diffuse, direct and net downward irradiance to become the functions of altitude shown in Figure II-2. These irradiances are normalized to the solar irradiance incident perpendicularly upon the top of the atmosphere.

By comparing the Figures we see that the aerosol layer increases the downward diffuse irradiance at 25 km altitude by a factor of 5. This is due to the conversion of direct sunlight into diffuse (scattered) light by the aerosol. Since the total irradiance at the surface is overwhelmingly due to downward diffuse light, the decrease in direct sunlight at the surface due to the aerosol layer is more than compensated for by the increase in downward diffuse light. Hence, as long as the radiation at the altitude of the top of the aerosol layer is primarily direct, enhancement of surface flux occurs by the mechanism described by *Davies* [1993]. However, *Davies'* model fails to predict the decrease in enhancement seen at the shortest ultraviolet wavelengths, because such wavelengths are both strongly Rayleigh scattered and ozone absorbed by the atmosphere. Therefore, little of the radiation is direct sunlight by the altitude of the top of the aerosol layer (most is scattered light), as can be seen from the fourth column of Table II-1. Hence scattering by the aerosol layer provides no shortcut to the ground, and, because of back scatter to space, the flux at the surface actually decreases.

I extended *Davies'* model to include Rayleigh scattering by the atmosphere in the single scattering approximation, assuming the same type of atmosphere and aerosol as was considered in calculating the results presented in Table II-1. In this case no enhancement of surface flux occurs in any of the wavelength intervals considered in Table II-1. This is because a single scattering calculation always underestimates the diffuse light reaching the surface; since the enhancement is due to an increase in diffuse light, no enhancement will be predicted if the model significantly underestimates the transmission of diffuse light to the surface.

II.3 Biological Significance

Large solar zenith angles and the ozone hole make Antarctica during spring the most susceptible place on earth to the scattering effect of volcanic aerosols. Accordingly, I have used our numerical model to calculate the effect of a volcanic aerosol layer on indigenous life in Antarctica during spring. For a representative day in Antarctica during the spring ozone hole (October 7), and for a representative region near the edge of the pack ice where most indigenous life is located (70°S), I calculate the flux, from 290-420 nm, transmitted to the surface both with and without a stratospheric aerosol layer. The baseline atmosphere is the *U.S. Standard Atmosphere* [1976], with the amount of ozone scaled to 200 DU to represent the effect of the ozone hole (about 35% below normal). I assume that the lower boundary of the atmosphere is the ocean, with an albedo of 0.05 [*Doda and Green,* 1980; *Blumthaler and Ambach,* 1988], assumed constant from 290-420 nm. The flux is calculated for 10 solar zenith angles between sunrise and noon (solar zenith angle = 64.5°).

The spectral irradiance (flux per unit wavelength interval) at each solar zenith angle is convolved with an action spectrum. The integrated daily dose of radiation [*Lubin et al.,* 1992] is approximated by the sum of the 10 convolutions weighted by the length of time each solar zenith angle is assumed to exist for. The convolution is calculated from 290-420 nm in 0.1 nm steps. It is necessary to calculate the change in the integrated daily dose of radiation, rather than just the change in the mid-day dose, because about half the daily dose of radiation is acquired at times other than mid-day, and the enhancement is much larger at these solar zenith angles than at mid-day.

The action spectrum is the relative effectiveness of different wavelengths in causing a biological response. I use the action spectrum for DNA [*Setlow*, 1974] for reference. This action spectrum is probably not representative of the effect on any biological organism. However, many previous studies of the transmission of ultraviolet radiation through the atmosphere use this action spectrum, and so I have included it for comparison purposes only. The action spectra I have chosen to be representative of the biological response of actual organisms are the phytoplankton action spectrum [*Mitchell*, 1990; *Lubin et al.*, 1992], and the erythema action spectrum [*Diffey*, 1987]. The choice of a phytoplankton action spectrum is obvious, since the oceans off Antarctica contain abundant phytoplankton. I chose the erythema action spectrum as a generic representation of the possible biological response of the eyes of birds and the eyes and skin of marine mammals which may come close to the ocean surface, or out of the water entirely onto the ice. These three action spectra differ mainly in the strength of the UV-A (320-420 nm) "tail" of the spectra relative to the weighting of the action spectra in the UV-B (290-320 nm). The DNA action spectrum has a negligible UV-A tail, the erythema UV-A tail is more prominent relative to its UV-B weighting, and the phytoplankton action spectrum has the strongest UV-A tail relative to UV-B weighting. Since the flux of UV-B is in general increased by the aerosol layer, while the flux of UV-A is decreased by the aerosol layer, the change in the integrated daily dose of radiation depends very strongly on which action spectrum is used to weight the spectral irradiance. The action spectrum with the weakest UV-A tail relative to UV-B weighting (DNA) will produce the most pronounced increases in the integrated daily dose of biologically weighted irradiance. By contrast, the action spectrum with the strongest relative UV-A tail (phytoplankton) is expected to show a decrease in the integrated daily dose of biologically weighted irradiance, because the decrease in UV-A, relatively strongly weighted by the action spectrum, outweighs the increase in UV-B.

The change in the integrated daily dose of biologically weighted irradiance, caused by a stratospheric aerosol layer being introduced into the baseline atmosphere, is given in Table II-2. The columns labeled "PP", "ERY" and "DNA" refer to weighting by the phytoplankton, erythema and DNA action spectra, respectively. The range of aerosol layer optical depths (t) bounds values observed after volcanic eruptions [*Hofmann*, 1987; *Valero and Pilewskie*, 1992]. The aerosol is assumed to be uniformly distributed within 2 altitude ranges representative of observations made after eruptions, 12-24 km [*McCormick et al.*, 1984] and 25-26 km [*DeLuisi et al.*, 1983], the latter altitude range observed at Mauna Loa about 1 week after the El Chichon eruption. This layer spread vertically with time [*DeLuisi et al.*, 1983], so for a similar layer to exist over Antarctica, the eruption would have to occur in or near to Antarctica. Observations of the aerosol cloud injected into the stratosphere by the Mt. Pinatubo eruption [*Osborn et al.*, 1995] indicate that this cloud spread from a narrow initial height range of 22 - 24 km, observed 2 months after the eruption, to a much broader and lower height range of 16-24 km, observed 12 months after the eruption. The observations indicate that the aerosol cloud descended and spread out at a roughly constant rate during this 10 month period. Hence I consider the narrow aerosol range considered in this paper to only be valid for one to two months following an eruption, requiring a volcanic eruption in Antarctica. Since volcanic aerosols may induce ozone destruction [*Vogelmann et al.*, 1992], I repeated the calculations for the 12-24 km layer assuming an 8% ozone column depletion (to 184 DU), due to uniform depletion within the layer. This is representative of ozone column depletions observed at some locations after the El Chichon eruption [*Bojkov*, 1987]. There is not enough ozone between 25 and 26 km altitude for an aerosol induced ozone depletion there to be significant.

For reference, the increase in the integrated daily dose of biologically weighted irradiance, weighted by the DNA action spectrum (BWI-DNA), due to an aerosol layer extending from 12-24 km, with an 8% ozone column depletion (to 184 DU), is given in Table II-2 and may be compared directly with the results of *Vogelmann et al.* [1992]. Their results were calculated for mid-latitudes, where scattering by an aerosol layer decreases the flux of ultraviolet light at the surface, offsetting the effects of aerosol induced ozone destruction [*Vogelmann et al.*, 1992]. *Vogelmann et al.* predict increases of approximately 8-15% for $t = 0.4\text{-}0.1$, which if applied to Antarctica would underestimate the increase in BWI-DNA there by up to a factor of 2.

Volcanic aerosols must induce significant ozone destruction in order for there to be an increase in BWI-DNA at mid-latitudes [*Vogelmann et al.*, 1992]. However, the Mt Pinatubo eruption caused zonal-scale ozone depletions of only 2-4% within the aerosol layer, so that scattering by the aerosols effectively offset the ozone destruction [*Vogelmann* *et al.*, 1992]. By contrast, the last column in Table II-2 shows that a narrow aerosol layer lying above Antarctica during spring can cause large (up to 17%) increases in BWI-DNA, even if the aerosols do not induce any ozone destruction.

Fortunately, however, the effect of a volcanic aerosol layer on the integrated daily dose of biologically weighted irradiance is much less severe when the phytoplankton and erythema action spectra are used. These action spectra are expected to more realistically mimic the biological response of life in Antarctica to ultraviolet radiation. Phytoplankton are actually better off with an aerosol layer overhead, because while the flux of UV-B radiation is increased, the flux of UV-A radiation is decreased, and the relatively strong UV-A weighting of the phytoplankton action spectrum leads to an overall decrease in biological response, even with an 8% reduction of ozone due to aerosol induced ozone destruction. However, the third column of Table II-2 shows that when the irradiance is weighted with the erythema action spectrum, the integrated daily dose of biologically weighted radiation (BWI-ERY) decreases by only up to 5% with the introduction of an aerosol layer. This is not enough to offset the effects of an 8% ozone depletion due to ozone destruction within the aerosol layer (sixth column of Table II-2). Larger aerosol induced ozone depletions will be offset by only up to the same 5%, so it is possible for volcanic aerosols to cause significant increases in BWI-ERY in Antarctica during spring. These increases would be in addition to the effects of the higher than normal levels of ultraviolet radiation in Antarctica caused by the yearly spring ozone hole.

Table II-1. Enhancement of Flux by a Stratospheric Aerosol Layer

Wavelength	E	τ_{O3} (nm)	% Direct (26 km)	% Direct (surface)	E-*Davies'* model
292	0.919	5.6	14	9.4×10^{-9}	1.5×10^{6}
296	1.22	3.1	65	1.6×10^{-4}	8.8×10^{2}
301	2.22	1.7	86	4.1×10^{-2}	1.6×10^{1}
305	2.22	0.89	92	0.47	2.0
310	1.56	0.47	93	1.2	0.76
315	1.18	0.23	94	1.9	0.48
320	1.01	0.11	94	2.7	0.39
325	0.939	0.050	95	3.5	0.36

Table II-2. Percentage Change in the Integrated Daily Dose of Biologically Weighted Irradiance (Positive = Increase), Due to an Aerosol Layer Lying Above Antarctica

	12-24 km (200 DU)			12-24 km (184 DU)			25-26 km (200 DU)		
τ	PP	ERY	DNA	PP	ERY	DNA	PP	ERY	DNA
0.0	0	0	0	1.8	11	19	0	0	0
0.1	-3.0	-1.5	0.0	-1.2	9.4	20	-2.4	0.8	4.8
0.2	-5.8	-2.7	0.3	-4.0	9.0	21	-4.7	1.7	9.3
0.3	-8.4	-4.9	-0.8	-6.5	7.9	22	-6.7	2.6	13
0.4	-11	-5.1	0.4	-8.9	4.5	18	-8.6	3.4	17

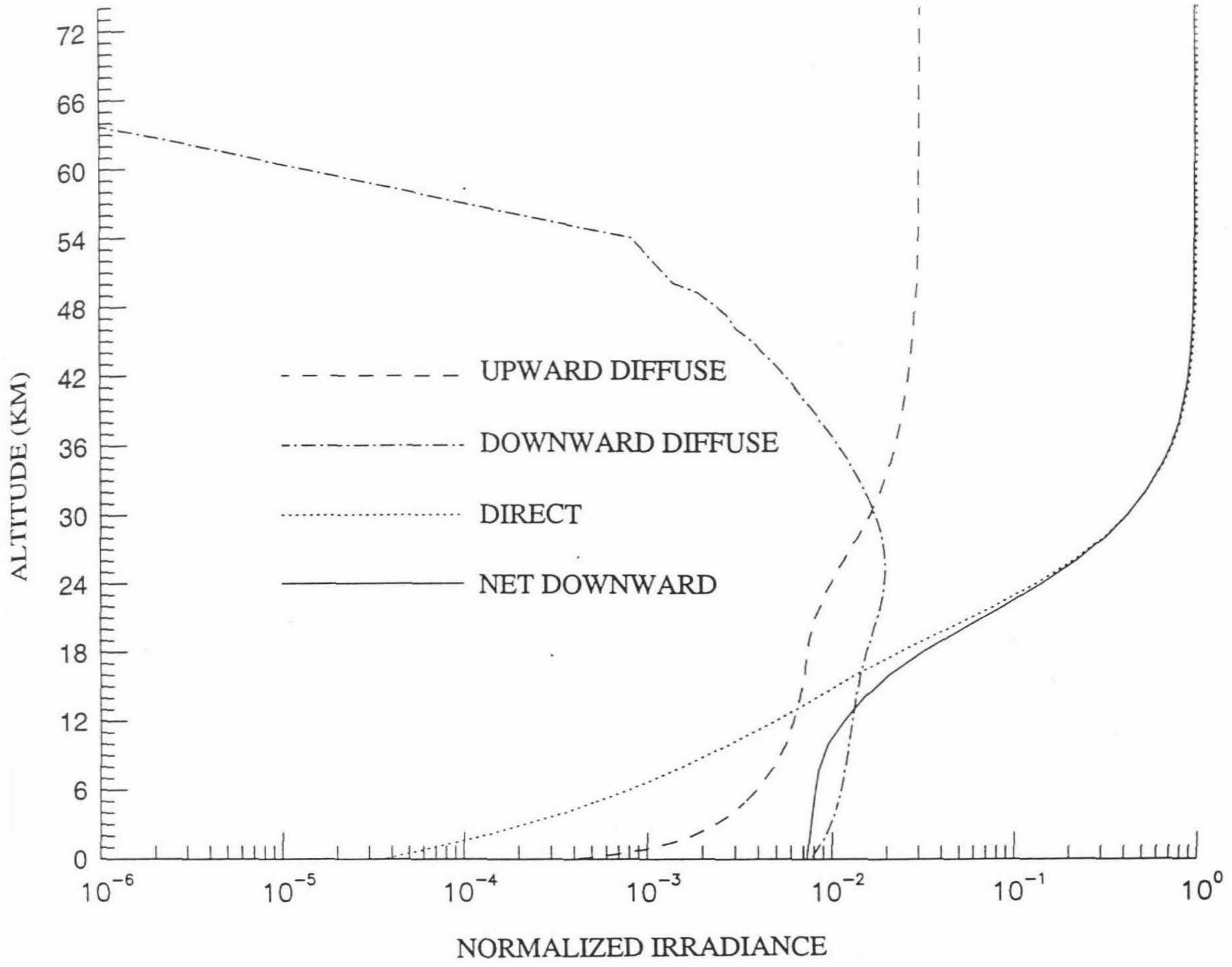


Figure II-1. The upward diffuse, downward diffuse, direct and net downward irradiance, normalized to the solar irradiance incident perpendicularly upon the top of the atmosphere, in the wavelength range 303.03-307.7 nm as a function of altitude in an aerosol-free *U.S. Standard Atmosphere* [1976], weighted to 200 DU. Net downward = direct + downdiffuse - updiffuse. The solar zenith angle is 80°.

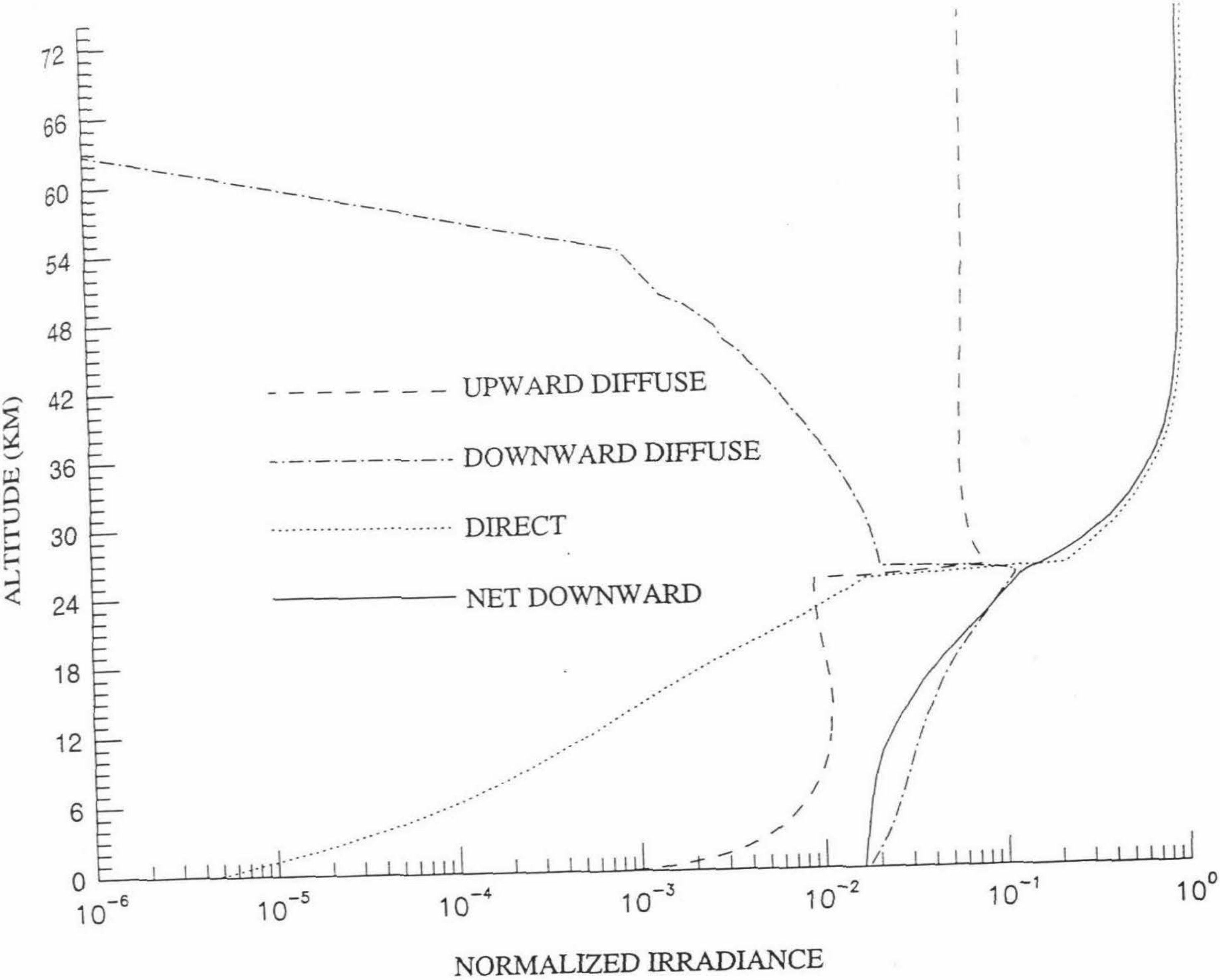


Figure II-2. With the introduction of an aerosol layer of optical depth 0.4, extending from 25-26 km altitude, the upward diffuse, downward diffuse, direct and net downward irradiance become the functions of altitude shown. The aerosol layer increases the flux at the surface by more than a factor of 2 by converting direct sunlight into diffuse (scattered) light.

Appendix

A1. A New Method for Validating a Multiple Scattering Radiative Transfer Model

The main method I used to validate our model was to calculate diffuse reflection and transmission functions for light incident upon a plane-parallel slab of scatterers, and compare the results with the tables published by *van de Hulst* [1980]. This method was used by *Michelangeli et al.* [1992] to validate their multiple scattering radiative transfer model. It should be noted that *Michelangeli et al.* [1992] state that the transmission function they calculate includes the direct (unscattered) beam when the emergence angle equals the incidence angle of the light incident upon the slab. In fact, the transmission function given in the tables by *van de Hulst* [1980] is clearly a diffuse transmission function that does not include the direct beam. Therefore, it is both a diffuse reflection and transmission function that must ultimately be calculated by ones model for comparison with the tables published by *van de Hulst* .

I have thought of a new way to validate a multiple scattering radiative transfer model that has not been proposed before. This may be used in addition to the methods used by previous authors to provide an additional test of the validity of ones model. The idea is to calculate the amount of diffuse light analytically in the single scattering approximation, and compare this with the results from the multiple scattering model being validated, using a small enough optical depth that the single scattering approximation is valid. I first derive the single scattering analytical solution and offer some relevant insights into radiative transfer theory, since neither the derivation nor the insights have been found anywhere in the literature. I then give an example of a comparison between the analytical solution and our multiple scattering radiative transfer model.

A1.1 Derivation of the Single Scattering Analytical Solution

We start with the equation of radiative transfer in a plane-parallel atmosphere, which is equation 1.63 in *Liou*,

$$\mu\frac{dI(\tau;\mu,\phi)}{d\tau} = I(\tau;\mu,\phi) - J(\tau;\mu,\phi) \tag{A1.1}$$

The diffuse intensity I is the flux of diffuse light per unit steradian at an optical depth τ, which is measured from the top of the atmosphere down. μ and ϕ define the direction of the diffuse beam, where μ is the cosine of the zenith angle and ϕ is the azimuth angle. The source function J represents the addition to the diffuse beam in the direction (μ , ϕ) due to light being scattered into that direction by the infinitesimal layer $d\tau$. We multiply both sides of this equation by $(1/\mu)e^{-\tau/\mu}$ and integrate in optical depth from 0 to τ:

$$I(\tau;\mu,\phi)e^{-\tau/\mu} - I(0;\mu,\phi) = \int_0^\tau -d\tau' e^{-\tau'/\mu}\,\frac{J(\tau';\mu,\phi)}{\mu} \tag{A1.2}$$

For downward traveling diffuse radiation μ is negative and $I(0;\mu,\phi) = 0$ since there is no diffuse radiation incident upon the top of the atmosphere. We see by inspection that

$$J(\tau';\mu,\phi) = \pi F_o e^{-\tau'/\mu_o}\,\tilde{\omega}\left(\frac{P(\mu,\phi;-\mu_o,\phi_o)}{4\pi}\right) \tag{A1.3}$$

where πF_0 is the solar flux incident upon the top of the atmosphere in the direction of the direct solar beam, and μ_o is the cosine of the angle between the zenith and a vector pointing towards the sun (which lies in the opposite direction of the direct solar beam). ϕ_o

is the azimuth angle of the direct solar beam, and so the direction of the direct solar beam, which points down, is $(-\mu_o,\phi_o)$. Assuming that the atmosphere is uniform, the single scattering albedo of every infinitesimal layer within the atmosphere is $\tilde{\omega} = \frac{\tau_s}{\tau}$, where τ_s is the scattering optical depth of the atmosphere in the zenith direction and τ is the total (scattering + absorption) optical depth of the atmosphere in the zenith direction. $\tilde{\omega}$ is the fraction of the extinction of a beam by an infinitesimal layer that is due to scattering. We also define $P(\mu',\phi';\mu,\phi)/_{4\pi}$ to be the probability of light scattering into the unit solid angle about the direction (μ',ϕ') from the direction (μ,ϕ). Evaluating the integral we get

$$I(\tau;\mu,\phi) = \pi F_o\, \tilde{\omega} \left(\frac{P(\mu,\phi;-\mu_o,\phi_o)}{4\pi}\right) \frac{\mu_o}{\mu + \mu_o}(e^{-\tau/\mu_o} - e^{\tau/\mu}) \tag{A1.4}$$

This is the downward diffuse intensity at an optical depth τ. The downward diffuse irradiance at this optical depth is then

$$F_D = \int_{\Omega'} -\mu I(\tau;\mu,\phi)\, d\Omega \tag{A1.5}$$

where Ω' is the downward facing hemisphere. Substituting the expression for I in the integrand and writing the integral over solid angle as an integral over μ and ϕ we get

$$F_D = \int_{\mu=0}^{1} \int_{\phi=0}^{2\pi} \pi F_o\, \tilde{\omega} \left(\frac{P(-\mu,\phi;-\mu_o,\phi_o)}{4\pi}\right) \frac{\mu_o}{\mu_o - \mu}\, (e^{-\tau/\mu_o} - e^{-\tau/\mu}) d\mu\mu d\phi \tag{A1.6}$$

Now

$$\begin{aligned} P(-\mu,\phi;-\mu_o,\phi_o) &= P(\mu,\phi;\mu_o,\phi_o) \\ &= \sum_{m=0}^{N} \sum_{l=m}^{N} \left(2 - \delta_{0,m}\right) \alpha'_l \frac{(l-m)!}{(l+m)!} P_l^m(\mu) P_l^m(\mu_o) \cos m(\phi_o - \phi) \end{aligned} \tag{A1.7}$$

where

$$\alpha'_l : P(\cos\alpha) = \sum_{l=0}^{N} \alpha'_l\, P_l(\cos\alpha) \tag{A1.8}$$

where α is the scattering angle, defined as the angle between the direction of travel of the incident and scattered radiation. The α'_l are the expansion coefficients for the expansion of the phase function in terms of Legendre polynomials.We have

$$\alpha'_l = \begin{Bmatrix} \delta_{l,0} + \frac{1}{2}\delta_{l,2} & \textit{Rayleigh scattering} \\ (2l+1)g^l & \textit{aerosol scattering} \end{Bmatrix} \tag{A1.9}$$

where we have used the Kronecker delta notation to write α'_l for Rayleigh scattering in compact form (for Rayleigh scattering $\alpha'_0 = 1, \alpha'_1 = 0, \alpha'_2 = 0.5; \alpha'_n = 0 \;\; for\; n \geq 3$). For aerosol scattering I have assumed the Henyey-Greenstein phase function, where g is the asymmetry parameter. Substituting this expansion of the phase function in terms of associated Legendre polynomials into the expression for F_D we get (arbitrarily putting ϕ_0 = 0)

$$F_D = \pi F_o \mu_o \frac{\tilde{\omega}}{2} \sum_{l=0}^{N} \alpha'_l P_l(\mu_o) \int_{\mu=0}^{1} \frac{\mu}{\mu_o - \mu} (e^{-\tau/\mu_o} - e^{-\tau/\mu}) P_l(\mu) d\mu \tag{A1.10}$$

The quantity $\pi F_o \mu_o$ is the flux of light incident *perpendicularly* upon the top of the atmosphere, and so $F_D / \pi F_o \mu_o$ is the downward diffuse irradiance at the bottom of the atmosphere, normalized to the flux of light incident perpendicularly upon the top of the atmosphere, calculated with the assumption of single scattering:

$$\frac{F_D}{\pi F_o \mu_o} = \frac{\tilde{\omega}}{2} \sum_{l=0}^{N} \alpha'_l P_l(\mu_o) \int_0^1 \frac{\mu}{\mu_o - \mu} (e^{-\tau/\mu_o} - e^{-\tau/\mu}) P_l(\mu) d\mu \tag{A1.11}$$

A1.2 The Meaning of Single Scattering

The amount of diffuse intensity extinguished by an infinitesimal layer of *total* optical depth $d\tau$ is the diffuse intensity incident upon the top of the infinitesimal layer times $d\tau/\mu$ (μ is now defined to be the negative of the cosine of the zenith angle, and so is a positive number for the downward traveling diffuse radiation we are considering). The source of that diffuse intensity incident upon the top of the infinitesimal layer is the single scattering of sunlight in higher infinitesimal layers. The point is that this diffuse intensity incident upon the top of the infinitesimal layer suffers attenuation due to the *total* - absorption *plus* scattering - optical depth of the infinitesimal layer. The equation of radiative transfer as formulated for single scattering still includes this extinction of single scattered light due to further scattering. That is, the further scattering of single scattered sunlight is taken into account in the single scattering approximation! The difference is that, unlike a multiple scattering model, this further scattering is treated as a simple absorption, and so always overestimates the actual extinction. Hence a single scattering calculation always underestimates the diffuse light reaching the surface.

A1.3 A Comparison of the Single Scattering Analytical Solution with our Multiple Scattering Numerical Model

The optical depths of the US Standard Atmosphere above 26 km in the wavelength range 290-294 nm are 0.0255 and 1.59 for Rayleigh scattering and ozone absorption, respectively. This gives a single scattering albedo of 0.016 for the assumed uniform atmosphere above 26 km. This is small enough that a single scattering calculation of the diffuse light above 26 km altitude should be reasonably accurate, allowing our multiple scattering numerical model to be validated by comparing the models predictions for the flux of diffuse light with that calculated with the single scattering analytical solution.

In order to make the comparison between our numerical model and the single scattering solution, I repeated the calculations using our numerical model with the assumptions that the atmosphere is uniform above 26km (but has a realistic varying composition below 26 km), that the radius of the earth is very large so that the attenuation of the direct beam occurs as if in a plane-parallel atmosphere, and that the surface albedo is zero. All of these assumptions serve to allow a fair comparison between our numerical model and the single scattering calculation, because the single scattering calculation assumes a uniform, plane-parallel atmosphere with no reflection from below. All the fluxes reported here from our numerical model and from the single scattering calculation have been normalized to the direct flux incident perpendicularly upon the top of the atmosphere.

In the following Table A1.1 I include a comparison between the results of our multiple scattering numerical model (the one used in our paper) and the single scattering analytical calculation at 9 intervening optical depths between the top of the atmosphere and the 1.62 total optical depth of 26 km altitude. In this Table only one direct flux is listed, since the direct flux is the same as calculated by our model or the single scattering calculation. The first two columns list the total optical depth and the approximate altitude at which that optical depth occurs. The third column gives the direct downward irradiance at that altitude. The fourth and fifth columns list the downward diffuse irradiance as calculated by our multiple scattering numerical model and by the analytical single scattering calculation respectively. All of the direct and diffuse irradiances are normalized to the direct irradiance incident perpendicularly upon the top of the atmosphere. The penultimate column shows the percentage amount by which the single scattering diffuse irradiance is below the diffuse irradiance as calculated by our multiple scattering numerical model. The last column shows the percentage of downward traveling light at that altitude which is due to direct light.

The percentage of direct light at 26 km altitude from the last line of the Table (13.3%) is slightly less than the 14% reported in Table II-1 of the main text of this chapter because the direct light suffers more attenuation in the flat atmosphere I have assumed here than in the realistic spherical atmosphere that was assumed in the calculations carried out in the main text.

The maximum percentage difference between the diffuse fluxes calculated using our numerical model and the single scattering calculation is 1.8%, with the diffuse flux calculated with the single scattering solution always less than that calculated with the multiple scattering numerical model, as expected. Notice that the percentage difference decreases as we go higher in the atmosphere. This is because as we go higher in the atmosphere the optical depth due to scattering of the layer we are calculating the diffuse flux through decreases, improving the assumption of single scattering (the chance of a photon suffering a second scattering decreases as the number of scatterers decreases).

This comparison provides an additional confirmation of the validity of our multiple scattering numerical model, and demonstrates a new method by which a multiple scattering numerical model may be validated.

Table A1.1

τ	alt (km)	direct	diffuse (m.s.)	diffuse (s.s.)	% diff	% direct
0.2	68	0.3160	3.739×10^{-3}	3.711×10^{-3}	-0.749%	98.8%
0.4	62	9.984×10^{-2}	3.550×10^{-3}	3.522×10^{-3}	-0.789%	96.6%
0.6	56	3.155×10^{-2}	2.753×10^{-3}	2.726×10^{-3}	-0.981%	92.0%
0.8	50	9.969×10^{-3}	2.033×10^{-3}	2.009×10^{-3}	-1.18%	83.1%
1.0	44	3.150×10^{-3}	1.489×10^{-3}	1.469×10^{-3}	-1.34%	67.9%
1.2	38	9.954×10^{-4}	1.097×10^{-3}	1.080×10^{-3}	-1.55%	47.6%
1.4	32	3.145×10^{-4}	8.147×10^{-4}	8.010×10^{-4}	-1.68%	27.9%
1.6	26.5	9.938×10^{-5}	6.105×10^{-4}	5.994×10^{-4}	-1.82%	14.0%
1.62	26	9.011×10^{-5}	5.959×10^{-4}	5.850×10^{-4}	-1.83%	13.3%

A2. A Proposed Method for Observing the Enhancement of Ultraviolet Light at the Surface by a Stratospheric Aerosol Layer.

Here I propose a method for quantitatively confirming that a stratospheric aerosol layer increases the flux of ultraviolet light at the surface for large solar zenith angles. *Zeng et al.* [1994] compared measurements of the ratio of diffuse light to direct light at the surface made prior to and after the Mt. Pinatubo eruption with calculations of these ratios using the optical properties of the atmosphere at both times, including the stratospheric aerosols present after the eruption. Calculation and measurement mostly agree; however, I do not consider this to be evidence for observation of an increase in flux of ultraviolet light at the surface at large solar zenith angles due to the presence of a stratospheric aerosol layer.

Inspection of Figure 9 of *Zeng et al.* [1994] shows that the measured ratio of diffuse to direct light increases with decreasing wavelength in a similar fashion for both the pre and post Mt. Pinatubo eruption measurements; the only difference is a constant shift between the two curves, presumably due to the different ozone column amounts present for both measurements, since this constant shift extends out to 450 nm, the maximum wavelength shown in the Figure, at which there could be no enhancement in the ratio of diffuse to direct light due to the effect in question, especially considering that the solar zenith angle for both observations is only around 23°. Therefore, the constant shift between the two curves of diffuse/direct light for pre and post eruption measurements cannot be due to the effect, and the increase in this ratio with decreasing wavelength also cannot be due to the enhancement effect because this increase occurs for the pre-eruption measurement also, where there is no stratospheric aerosol. Therefore, this comparison of measurement with calculation provides no evidence for the enhancement effect.

Forster et al. [1995] found an unexplained discrepancy between modeled and measured values of the flux of UV-B light at the surface, measured at Reading, United Kingdom with clear skies. They found that including stratospheric volcanic aerosols in their model improved agreement between their model and the measurements for large solar zenith angles by increasing the flux of UV-B light at the surface predicted by the model by as much as 6%. They explain there could be an instrument based reason for the discrepancy also, and the fact that including volcanic aerosols in their model improved agreement with their observations is not confirmation of the effect, since the existence of volcanic aerosols must be an independent fact.

In light of this history, I may state that a stratospheric aerosol layer increasing the flux of ultraviolet light at the surface, for large solar zenith angles, is predicted but has not yet been observed, and so I propose here a possible method for observation. After the next major volcanic eruption that injects significant quantities of volcanic aerosols into the stratosphere (El Chichon and Mt. Pinatubo being the most recent such eruptions in 1982 and 1991, respectively), I propose that observations be made at sunrise and sunset of the flux of downward traveling diffuse light at the surface.

By making observations at sunrise and sunset, large solar zenith angles can be realized at any latitude. Therefore, this technique allows observations of the effect to be made from anywhere in the world. In addition to measurements of the flux of downward traveling diffuse light at the surface, the solar flux per unit wavelength interval (spectral irradiance) incident upon the top of the atmosphere must be known, since the flux at the surface is proportional to the flux at the top of the atmosphere. Therefore, the flux incident upon the top of the atmosphere must be known otherwise if an increase in flux is observed beyond that expected for a standard solar irradiance at the top of the atmosphere, it will be unknown whether this increase is due to the effect in question or simply because the incident solar flux at that wavelength happened to be larger than the standard amount.

That the flux at the surface is proportional to the flux incident upon the top of the atmosphere may not be immediately obvious, but this follows from considering the flux (either direct, upward or downward diffuse) to be a certain value f at the surface for a given value of the flux of sunlight incident perpendicularly upon the top of the atmosphere, F. Now if the flux of sunlight incident perpendicularly upon the top of the atmosphere is now some multiple of F, $\mathrm{n}F$, then it can be seen that for each part F of the flux, the flux at the surface is f, so that for a total flux $\mathrm{n}F$, the flux at the surface is $\mathrm{n}f$. This holds as long as the n "components" of size F of the total incident flux $\mathrm{n}F$ do not affect each other's transmission through the atmosphere, for example by heating the atmosphere which affects absorption cross sections and therefore the transmission of light. If non linear effects such as these are ignored, then it follows that the flux of light at the surface is proportional to the flux of light incident upon the top of the atmosphere.

In addition, the amount of ozone as a function of altitude above the observing site must be known, as well as the albedo of the surface around the observing site. With the incident solar spectral irradiance, ozone density as a function of altitude and surface albedo known, the downward directed diffuse flux at the surface in the UV-B (290-320 nm) may be easily calculated numerically for a clear atmosphere. The diffuse fluxes calculated this way represent the non-volcanic case; that is, the diffuse fluxes that would have been observed on this day if there had been no volcanic eruption. For simplicity, so there are no complications affecting observation of the effect, I assume a clear atmosphere, meaning no clouds and no significant tropospheric aerosols. This requires that the observations of diffuse flux at the surface after the volcanic eruption be made in a cloudless sky with little tropospheric aerosols. The observations of diffuse fluxes can then be compared with the values calculated for a clear atmosphere; the effect will show up as an enhancement in the diffuse flux at certain UV-B wavelengths, with a characteristic diminishment at both shorter and longer wavelengths.

To illustrate this method, I calculate the downward diffuse light at the surface for a clear *U.S. Standard Atmosphere* [1976] with 344.446 DU and a UV-B surface albedo of 0.05, which happens to be characteristic of either (snow free) land or sea [*Doda and Green*, 1980; *Blumthaler and Ambach*, 1988]. I calculate this light for both a clear atmosphere, as I would if I were making observations of the downward diffuse light after a volcanic eruption, but instead of observations of the downward diffuse light after an eruption I use calculations of the flux of downward diffuse light assuming a volcanic aerosol layer of optical depth 0.4 [*Valero and Pilewskie*, 1992] spread uniformly from 25-26 km [*DeLuisi et al.*, 1983]. The volcanic aerosols are assumed to scatter with the Henyey-Greenstein phase function [*Henyey and Greenstein*, 1941], with an asymmetry parameter of 0.75 and single scattering albedo of 0.99 [*Vogelmann et al.*, 1992].

In the Table A2.1 I show the ratio of "observed" flux after the eruption f_o, to the flux calculated for a clear atmosphere, f_c, using the properties of the atmosphere and surface that exist when the post eruption "observations" are made. Notice that it is not necessary to specify the solar irradiance incident upon the top of the atmosphere (F); since both f_o and f_c are proportional to F, the ratio of f_o to f_c is independent of F. I show f_o/f_c for 8 wavelength intervals from the *World Meteorological Organization* [1986] report whose approximate center wavelengths are shown and extend from 292 to 325 nm, which is basically the UV-B wavelength range. The variation of fo/fc over these wavelength intervals is shown for three solar zenith angles; 85^o, 80^o and 75^o.

As can be seen quantitatively from Table A2.1, the wavelength interval at which the enhancement peaks decreases as the solar zenith angle decreases. When the sun is higher in the sky, the condition that the slant path of the direct beam below the aerosol layer is optically very thick in absorption is met at shorter wavelengths, since the slant path is less inclined to the vertical and so travels through less path length of atmosphere. Conversely, when the sun is low in the sky, the slant path is very long, and so the absorption optical depth along the slant path becomes quite large at relatively longer wavelengths, meaning that the conditions necessary for enhancement occur at longer wavelengths.

In Table A2.2 I have given the downward directed diffuse flux at the surface, in Photons cm^{-2}s^{-1}, for the atmosphere with the volcanic aerosol layer. For the solar flux incident upon the top of the atmosphere in the 8 wavelength intervals, I have used the *World Meteorological Organization Report* [1986]. The purpose of this Table is to show the wide range in magnitudes of the transmitted diffuse flux at the surface, which would have to be measured, as a function of wavelength and solar zenith angle. The observer's ability to measure diffuse fluxes on the magnitudes indicated will dictate the wavelengths and solar zenith angles at which the observer should attempt to observe the enhancement.

The observations proposed here, made after the next major volcanic eruption that injects significant quantities of aerosols into the stratosphere, would detect the enhancement through the increase of downward directed diffuse light compared with that predicted numerically for a clear atmosphere. The characteristic behavior of the enhancement as a function of wavelength and solar zenith angle would be the telltale signs of the effect qualitatively; of course, if quantitative confirmation were desired, the observations could be compared quantitatively with the ratios predicted numerically using the observed optical proprieties of the volcanic aerosol layer.

Table A2.1. The Enhancement of Downward Directed Diffuse Flux at the Surface by a Volcanic Aerosol Layer at Various Solar Zenith Angles

	Solar Zenith Angle		
Wavelength (nm)	85°	80°	75°
292	0.880	0.881	0.897
296	0.905	0.925	1.22
301	0.937	1.36	2.56
305	1.14	2.40	2.47
310	1.71	2.17	1.62
315	1.71	1.48	1.23
320	1.28	1.15	1.09
325	1.01	1.01	1.04

Table A2.2. Downward directed diffuse flux at the surface (solar flux incident perpendicularly upon the top of the atmosphere) in Photons $cm^{-2}s^{-1}$ in the atmosphere with a volcanic aerosol layer.

	Solar Zenith Angle		
Wavelength (nm)	85°	80°	75°
289.9 - 294.1	$5.6x10^5$ ($3.0x10^{13}$)	$8.8x10^5$ ($6.0x10^{13}$)	$1.3x10^6$ ($9.0x10^{13}$)
294.1 - 298.5	$9.0x10^7$ ($3.0x10^{13}$)	$1.7x10^8$ ($5.9x10^{13}$)	$3.3x10^8$ ($8.8x10^{13}$)
298.5 - 303.0	$2.6x10^9$ ($2.8x10^{13}$)	$7.5x10^9$ ($5.6x10^{13}$)	$2.6x10^{10}$ ($8.3x10^{13}$)
303.0 - 307.7	$4.7x10^{10}$ ($3.7x10^{13}$)	$2.1x10^{11}$ ($7.3x10^{13}$)	$6.4x10^{11}$ ($1.1x10^{13}$)
307.7 - 312.5	$4.3x10^{11}$ ($4.3x10^{13}$)	$1.9x10^{12}$ ($8.6x10^{13}$)	$4.5x10^{12}$ ($1.3x10^{14}$)
312.5 - 317.5	$2.2x10^{12}$ ($4.7x10^{13}$)	$7.6x10^{12}$ ($9.4x10^{13}$)	$1.5x10^{13}$ ($1.4x10^{14}$)
317.5 - 322.5	$6.0x10^{12}$ ($5.2x10^{13}$)	$1.6x10^{13}$ ($1.0x10^{14}$)	$3.2x10^{13}$ ($1.5x10^{14}$)
322.5 - 327.5	$1.2x10^{13}$ ($6.1x10^{13}$)	$3.0x10^{13}$ ($1.2x10^{14}$)	$5.2x10^{13}$ ($1.8x10^{14}$)

References

Blumthaler, M., and W. Ambach, Solar UVB-albedo of various surfaces,*Photochem. Photobiol.,* 48, 85-88, 1988.

Bojkov, R. D., The 1983 and 1985 anomalies in ozone distribution in perspective, *Mon. Weath. Rev.,* 115, 2187-2201, 1987.

Davies, R., Increased transmission of ultraviolet radiation to the surface due to stratospheric scattering, *J. Geophys. Res.,* 98, 7251-7253, 1993.

DeLuisi, J. J., E. G. Dutton, K. L. Coulson, T. E. DeFoor, and B. G. Mendonca, On some radiative features of the El Chichon volcanic stratospheric dust cloud and a cloud of unknown origin observed at Mauna Loa, *J. Geophys. Res.,* 88, 6769-6772, 1983.

Diffey, B. L., A comparison of dosimeters used for solar ultraviolet radiometry, *Photochem. Photobiol.,* 46, 55-60, 1987.

Doda, D. D., and A. E. S. Green, Surface reflectance measurements in the UV from an airborne platform. Part 1, *Applied Optics,* 19, 2140-2145, 1980.

Forster, P. M. D., K. P. Shine, and A. R. Webb, Modeling ultraviolet-radiation at the earths surface .2. Model and instrument comparison, *J. Applied Meteorology*, 34, 2426-2439, 1995.

Froidevaux, L., M. Allen, and Y. L.Yung, A critical analysis of CLO and O3 in the mid-latitude stratosphere,*J. Geophys. Res.,* 90, 12,999- 13,029, 1985.

Henyey, L. C., and J. L. Greenstein, Diffuse radiation in the galaxy, *Astrophys. J.,* 93, 70-83, 1941.

Hofmann, D. J., Perturbations to the global atmosphere associated with the El Chichon volcanic eruption of 1982, *Rev. Geophys.*, 25, 743-759, 1987.

Lubin, D., B. G. Mitchell, J. E. Frederick, A. D. Alberts, C. R. Booth, T. Lucas, and D. Neuschuler, A contribution toward understanding the biospherical significance of Antarctic ozone depletion, *J. Geophys. Res.*, 97, 7817-7828, 1992.

McCormick, M. P., T. J. Swissler, W. H. Fuller, W. H. Hunt, and M. T. Osborn, Airborne and ground-based lidar measurements of the El Chichon stratospheric aerosol from 90^{o} N to 56^{o} S, *Geofis. int.*, 23, 187-221, 1984.

Michelangeli, D. V., M. Allen, and Y. L. Yung, El Chichon volcanic aerosols: impact of radiative, thermal, and chemical perturbations, *J. Geophys. Res.*, 94, 18,429-18,443, 1989.

Michelangeli, D. V., M. Allen, Y. L. Yung, R.-L. Shia, D. Crisp, and J. Eluszkiewicz, Enhancement of atmospheric radiation by an aerosol layer, *J. Geophys. Res.*, 97, 865-874, 1992.

Mitchell, B. G., Action spectra of ultraviolet photoinhibition of Antarctic phytoplankton and a model of spectral diffuse attenuation coefficients, in *Response of Marine Phytoplankton to Natural Variations in UV-B Flux,* edited by G. Mitchell, I. Sobolev, and O. Holm-Hansen, Appendix H, Chemical Manufacturers Association, Washington, D.C., 1990.

Osborn, M. T., R. J. DeCoursey, C. R. Trepte, D. M. Winker, and D. C. Woods, Evolution of the Pinatubo volcanic cloud over Hampton, Virginia, *Geophys. Res. Lett.*, 22, 1101-1104, 1995.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN The Art of Scientific Computing, Second Edition*, Cambridge University Press, Cambridge CB2 IRP, pp. 38-39, p. 247, 1992.

Setlow, R. B., The wavelengths in sunlight effective in producing skin cancer: A theoretical analysis, *Proc. Natn. Acad. Sci.. USA,* 71, 3363- 3366, 1974.

Tsay, S. C., and K. Stamnes, Ultraviolet-radiation in the Arctic: the impact of potential ozone depletions and cloud effects, *J. Geophys. Res.*, 97, 7829-7840, 1992.

U.S. Standard Atmosphere, *NOAA-S/T 76-1562*, U.S. Government Printing Office, Washington D.C., 1976.

Valero, F. P. J., and P. Pilewskie, Latitudinal survey of spectral optical depths of the Pinatubo volcanic cloud - derived particle sizes, columnar mass loadings, and effects on planetary albedo, *Geophys. Res. Lett.*, 19, 163-166, 1992.

van de Hulst, H. C., *Multiple light scattering, volume 2*, Academic, San Diego, pp. 412-443, 1980.

Vogelmann, A. M., T. P. Ackerman, and R. P. Turco, Enhancements in biologically effective ultraviolet radiation following volcanic eruptions, *Nature,* 359, 47-49, 1992.

World Meteorological Organization, Atmospheric Ozone 1985: Assessment of our understanding of the processes controlling its present distribution and change, *Rep.*

16, chap. 7, pp. 349-392, Global Ozone Research and Monitoring Project, Geneva, 1986.

Zeng, J., R. McKenzie, K. Stamnes, M. Wineland, and J. Rosen, Measured UV spectra compared with discrete ordinate method simulations, *J. Geophys. Res.*, 99, 23019-23030, 1994.

Chapter III

A Novel Method for Enhancing Subsurface Radar Imaging using Radar Interferometry

III.1 Introduction

Radar imaging of buried surfaces that cannot be seen in visible light is a phenomenon that can occur due to the much longer wavelengths (of order 10cm) of imaging radar compared to visible light (around 0.5 micron). The radar wave Rayleigh scatters easily around particles less than about 1/10th of its wavelength the same way visible light Rayleigh scatters from nitrogen molecules in the atmosphere; Rayleigh scattering by nitrogen molecules, much smaller than the wavelength of visible light, makes the sky blue but it doesn't obscure the Sun.

There are two major factors that can prevent a strong signal from being observed from a buried surface: liquid water [*Blom et al.*, 1984] and the presence of particles, defined as objects with a different dielectric constant from the surroundings, of size greater than about one tenth of the radar wavelength [*Roth and Elachi*, 1975]. Liquid water is a good absorber of electromagnetic radiation of radar wavelengths, and so penetration of the radiation into the water is limited even if the water forms an infinite medium. If instead the water forms drops comparable in size to the radar wavelength then it will both absorb and strongly scatter the radiation, and penetration is limited due to both effects. Indeed, any "drop" of substance with a dielectric constant different from the surrounding medium of size greater than about one fifth of the radar wavelength will strongly scatter the radiation, and therefore limit penetration, even if absorption within the "drop" is negligible.

The reason is the same reason that clouds only become visible when water vapor condenses into droplets; the water vapor was always there, but only when droplets form does the cloud suddenly become visibly, meaning it scatters much more light. Drops of liquid water and particles of a size greater than about one fifth of the radar wavelength essentially act like clouds, greatly scattering light of radar wavelengths much more than water vapor or smaller particles. The reason [*Feynman et al.*, 1963] is that when atoms are clumped together in a agglomerate less than about a wavelength, the electrons in the atoms move in phase in response to the electric field of the electromagnetic radiation, because for distances of, say, one tenth of a wavelength the phase of the wave is approximately constant. Being driven in phase, the electromagnetic fields radiated from the atoms combine in phase, so the amplitudes add, and for N atoms the scattered amplitude is N fold increased. The scattered intensity is then N^2 fold increased, while if the N atoms were spread out in space the scattered intensity from each would add, since they would not be in phase, and so there is only a N fold increase in intensity.

Compared to small pebbles of a similar size, so the enhancement in scattering due to the agglomeration effect described above is the same, water droplets interact particularly strongly with radar waves because water molecules are very dipolar, and so they make excellent linear antennas along whose axis the electrons may be accelerated by the electromagnetic radar wave. The accelerated electrons radiate electromagnetic radiation, and their motion is additionally damped by interaction with the electrons from neighboring molecules. These two effects are manifested as the scattering and absorption of the incident radar wave, respectively.

Of course, if the water molecules form an infinite medium the scattered radiation from each water molecule combines in such a way to cancel out in all directions *except* for the incident direction, so there is no net effect due to the scattering. If the water molecules do not form a continuous, infinite medium but a half space, then at the boundary we know there will be a reflected and transmitted wave. If the molecules just form a discrete drop then there is no cancellation of reradiated radiation in all directions except for the incident, as in the case of an infinite medium, or except for the transmitted and reflected directions, in the case of a half space. Instead, there is scattered radiation in all directions, with the scattering effect strongly enhanced for agglomeration of water molecules of a size greater than about one fifth of the radar wavelength; drops of liquid water. If the drops get very big then they start to approximate infinite half spaces, with cancellation of the scattered radiation in all directions other than the reflected and transmitted directions. Hence there is a range of drop sizes for most effective scattering, calculated by *Roth and Elachi* [1975] to be generally 0.1λ - 100λ.

Solid water, while an agglomeration only slightly less dense than liquid, actually interacts very weakly with radar waves. In ice, electrons in individual water molecules are tightly bound in covalent bonds to the electrons of neighboring water molecules; these strong bonds, like stiff springs, essentially prevent the oscillation of a molecule's electrons along the dipole axis by the radar electromagnetic wave. Hence there is little scattering or absorption of the radar wave.

Therefore the conditions for subsurface imaging to occur are that the overlying material must not contain within its volume too many particles of a size greater than about one fifth of the radar wavelength, or even more importantly much liquid water. Otherwise, the echo from the lower surface will be strongly attenuated due to scattering or scattering and absorption, respectively, of the incident wave. In this case the echo from the lower surface will not be distinguishable from the thermal noise (reflected and emitted radiation from the surface due to the temperature of the sky and ground, respectively, as well as the receiver thermal temperature). Even if the echo were distinguishable from the thermal noise, it may not be distinguishable from the echo due to the dielectric discontinuity at the upper surface. Thus, I add the final condition for subsurface imaging to occur, that the upper surface not be too "radar bright".

III.2 Quantifying the Echo Due to the Dielectric Discontinuity at the Upper Surface

While the requirement that the surface not be too "radar bright" has been identified before [*Blom et al.*, 1984], the magnitude of the surface echo compared to the lower echo has not been previously discussed. Here I calculate the relative magnitudes for a typical situation for the first time, using the approach developed by *Elachi et al.* [1984] who calculate the effect a covering layer can have on the echo from the lower surface. *Elachi et al.* [1984] showed that, due to reduction of the angle of incidence and wavelength, a covering layer can actually enhance the lower echo, compared to a bare surface, for depths up to an appreciable fraction of a skin depth. However, *Elachi et al.* [1984] only considered the effect a covering layer with a *completely smooth* upper surface has. The effect of a covering layer with an upper surface that was as rough as the lower surface - assumed by *Elachi et al.* [1984] to be only slightly rough after all - was not discussed.

A natural surface can be simply modeled as the combination of large facets with small scale roughness superimposed [*Elachi*, 1987]. The radar echo from each of the facets is the same as the antenna pattern of small rectangular antenna, a three dimensional lobe, and since most natural surfaces have mean surface slopes less than about 30° [*Elachi*, 1987], in order to be within the lobes near normal incidence angles are required. At higher incidence angles the echo lies outside the boundaries of the facet lobe and the echo from the small scale roughness dominates. The small scale roughness can be considered as a Fourier series of sinusoidal components of different amplitudes, and there will be a dominant resonance scattering, the strongest Bragg resonance, from the component whose wavelength is such that echoes from neighboring crests are one wavelength out of phase, combining constructively.

Using this model for the radar scatter from a natural surface at a non-near normal incidence angle, the direct polarization (same send and receive polarizations) power back scatter cross section, σ, from a slightly rough surface is given by [*G. R. Valenzuela*, 1967; *G. R. Valenzuela*, 1968]

$$\sigma_u = f_{ij}\left(\varepsilon_1, \lambda, \theta\right) W_u\left(2k \sin \theta\right) \tag{III.1}$$

for the upper surface and

$$\sigma_l = T_i\left(\theta\right) T_j\left(\theta\right) e^{-2L/d \cos \theta'} f_{ij}\left(\frac{\varepsilon_2}{\varepsilon_1}, \frac{\lambda}{\sqrt{\varepsilon_1}}, \theta'\right) W_l\left(2k \sqrt{\varepsilon_1} \sin \theta'\right) \tag{III.2}$$

for the lower surface, where the geometry of the two surfaces is illustrated in Figure III-1.

We require that the surface be only slightly rough because the derivation of the above expressions is predicated on the assumption that the phase of the wave is constant over the scale of the small scale roughness, requiring this scale to be small compared to the radar wavelength.

ε_1 and ε_2 are the real part of the dielectric constant of the medium overlying the lower surface and the material comprising the lower surface, respectively. The wavelength of the radar wave in the atmosphere above the medium overlying the lower surface is λ (the wave number $k = 2\pi/\lambda$), and this wave impinges upon the medium with the angle of incidence θ. The wavelength and angle of incidence are reduced to $\frac{\lambda}{\sqrt{\varepsilon_1}}$ and θ' respectively within the medium, in which the skin depth is d, where θ' is given approximately (because the upper surface is not completely smooth) by Snell's law, $\frac{\sin\theta}{\sin\theta'} = \sqrt{\varepsilon_1}$.

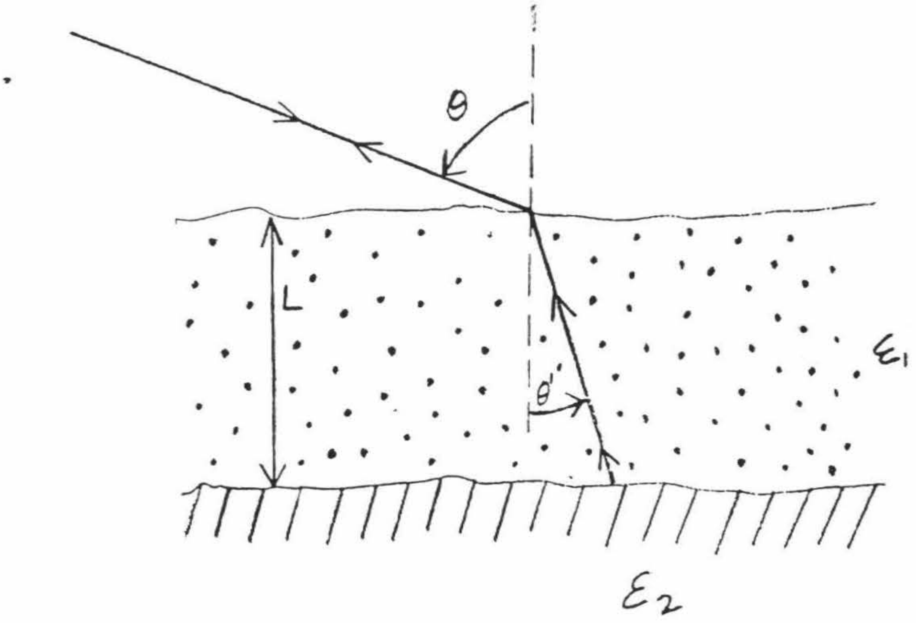


Figure III-1: The geometry of subsurface imaging. ε_1 and ε_2 are the real part of the dielectric constant of the medium overlying the lower surface and the material comprising the lower surface, respectively. The wavelength of the radar wave in the atmosphere above the medium overlying the lower surface is λ (the wave number $k = 2\pi/\lambda$), and this wave impinges upon the medium with the angle of incidence θ.

W_l and W_u are the roughness spectra - the spectral energy density - of the lower and upper surfaces, respectively. These roughness spectra are a function of wavenumber k and give the energy per unit wave number of the sinusoidal components of the surface roughness with wave number infinitesimally close to k. Though having no physical meaning when describing the waves that a rough surface can be decomposed into, the energy associated with a wave is the square of its amplitude, so the spectral energy density could be renamed the spectral "amplitude squared" density in this case. The spectral density gives information about the amplitudes of the different Fourier components that the small scale roughness of each surface can be decomposed into. W_l and W_u are evaluated at the wave number that gives the lowest order, and strongest, Bragg resonance, which is $2k \sin\theta$ for the upper surface and $2k\sqrt{\varepsilon_1}\sin\theta'$ for the lower surface. From Snell's law we see that these wave numbers are the same, so the relative values of W_l and W_u are the relative values of the two surfaces spectral energy densities evaluated at the same wave number, $2k \sin\theta$. If the surfaces are equally rough at this wave number - if the sinusoidal component of wave number $2k \sin\theta$ of each surface has the same amplitude - the spectral densities will be the same.

$T_i(\theta)$ is the power transmission function for the transmission of polarization i through the upper surface. For a completely smooth upper surface $T_i(\theta)$ is given by

$$T_V(\theta) = \frac{\sin 2\theta \sin 2\theta'}{\sin^2(\theta+\theta')\cos^2(\theta-\theta')} \qquad \text{(III.3)}$$

for vertical polarization and

$$T_H(\theta) = \frac{\sin 2\theta \sin 2\theta'}{\sin^2(\theta+\theta')} \qquad \text{(III.4)}$$

for horizontally polarized radiation [*Born and Wolf*, 1975]. While we have assumed that the upper surface is slightly rough, for such a surface there are still well defined reflected and transmitted beams, albeit slightly lobe shaped [*Elachi*, 1987], and so the above expressions for the transmission function will still be valid.

This follows intuitively by considering that the error associated with this approximation results from the fact that some radiation is scattered into directions other than the smooth surface transmitted and reflected directions, and so there is less energy carried away in the direction of the smooth surface reflected and transmitted beams than in the case of a smooth surface. However, for a slightly rough surface we know that the specular reflection is much brighter than the scattered return (seen from an airplane, the specular reflection of the sun off slightly rough water is much brighter than the surrounding water). Furthermore, the dielectric constant of the materials through which subsurface imaging is achieved is typically around 3, and for such values the transmission coefficient is near 1 and therefore much more energy is transmitted than reflected. Since little energy is specularly reflected compared to transmitted through the surface, and the amount of energy diffusely scattered at the surface is much less than the amount specularly reflected, it follows that the amount of energy diffusely reflected from the surface is very much less than the energy transmitted though the surface. The result is that very little power is scattered from a slightly rough surface compared with the power transmitted through the surface, and so the error associated with using the smooth surface expression for the transmission function is small.

$f_i(\varepsilon,\lambda,\theta)$ is a function dependent upon the send polarization i and receive polarization j, and is given by

$$f_{HH}(\varepsilon,\lambda,\theta) = 4\pi k^4 \cos^4\theta \left[\frac{\varepsilon - 1}{\left(\cos\theta + \sqrt{\varepsilon - \sin^2\theta}\right)^2} \right]^2 \tag{III.5}$$

for direct horizontal polarization ($i=H$, $j=H$) and

$$f_{VV}(\varepsilon,\lambda,\theta) = 4\pi k^4 \cos^4\theta \left[\frac{(\varepsilon - 1)\left\{\varepsilon + (\varepsilon - 1)\sin^2\theta\right\}}{\left(\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}\right)^2} \right]^2 \tag{III.6}$$

for direct vertical polarization ($i=V$, $j=V$).

The cross polarization (different send and receive polarizations) power back scatter cross section from a slightly rough surface is given by [*G. R. Valenzuela*, 1967; *G. R. Valenzuela*, 1968]

$$\sigma_u = f_{HV}(\varepsilon_1,\lambda,\theta) I_u \tag{III.7}$$

for the upper surface and

$$\sigma_l = T_H(\theta)\, T_V(\theta)\, e^{-2L/d\cos\theta'} f_{HV}\left(\frac{\varepsilon_2}{\varepsilon_1}, \frac{\lambda}{\sqrt{\varepsilon_1}}, \theta'\right) I_l \tag{III.8}$$

for the lower surface, where I_l and I_u are the surface roughness auto correlation functions of the lower and upper surfaces, respectively. $f_{HV}(\varepsilon,\lambda,\theta)$ is a function defined by

$$f_{HV}(\varepsilon,\lambda,\theta) = 2\pi k^8 \cos^4\theta \frac{(\varepsilon - 1)^4 \left(\varepsilon - \sin^2\theta\right)}{\left(\cos\theta + \sqrt{\varepsilon - \sin^2\theta}\right)^2 \left(\varepsilon\cos\theta + \sqrt{\varepsilon - \sin^2\theta}\right)^2} \tag{III.9}$$

If the lower and upper surfaces are equally rough they will have the same spectral energy densities W and therefore the same surface roughness auto correlation functions I. In this case the surface roughness terms cancel out and the ratio R_{ij} of the power back scatter cross section of the upper surface to the power back scatter cross section of the lower surface for a polarization combination *ij* given by

$$R_{ij}\left(\varepsilon_1,\varepsilon_2,\lambda,\theta\right) = \frac{\sigma_u}{\sigma_l} = \frac{f_{ij}\left(\varepsilon_1,\lambda,\theta\right)}{T_i\left(\theta\right)\, T_j\left(\theta\right)\, e^{-2L/d\cos\theta'} f_{ij}\left(\frac{\varepsilon_2}{\varepsilon_1},\frac{\lambda}{\sqrt{\varepsilon_1}},\theta'\right)} \tag{III.10}$$

where it is understood that θ' is a function of θ via Snell's law, $\frac{\sin\theta}{\sin\theta'} = \sqrt{\varepsilon_1}$.

In Figures III-2, III-3 and III-4 I plot R_{HH}, R_{VV} and R_{HV} ($=R_{VH}$), respectively, as a function of ε_1 and ε_2 over a ± 1 range for each dielectric constant about the values ε_1 = 2.5 and ε_2 = 8.0. The central value of ε_1 is characteristic of the dielectric constant of many powdered rocks, while the central value of ε_2 is characteristic of solid mafic rocks [*Campbell and Ulrichs*, 1969]. The sand in the Selima Sand Sheet on the Egypt/Sudan border through which subsurface imaging of dry river channels was achieved by SIR-A has a measured dielectric constant of about 3.5 [*McCauley et al.*, 1982]. I take $\theta = 50^{\circ}$, typical for the Shuttle Imaging Radar, and the depth of the over layer equal to the skin depth, $L=d$.

From these Figures we see that with direct polarization the echo from the upper surface is comparable in size to the echo from the lower surface. This upper echo, uncorrelated with the lower echo and of a similar magnitude, would effectively make the signal to noise equal to one regardless of the thermal noise and prevent subsurface imaging. Therefore, subsurface imaging at depths near the skin depth at direct polarization will be impossible when the upper surface is as rough as the lower surface. If the upper surface is less rough the upper echo is smaller, and imaging will be possible but obviously degraded

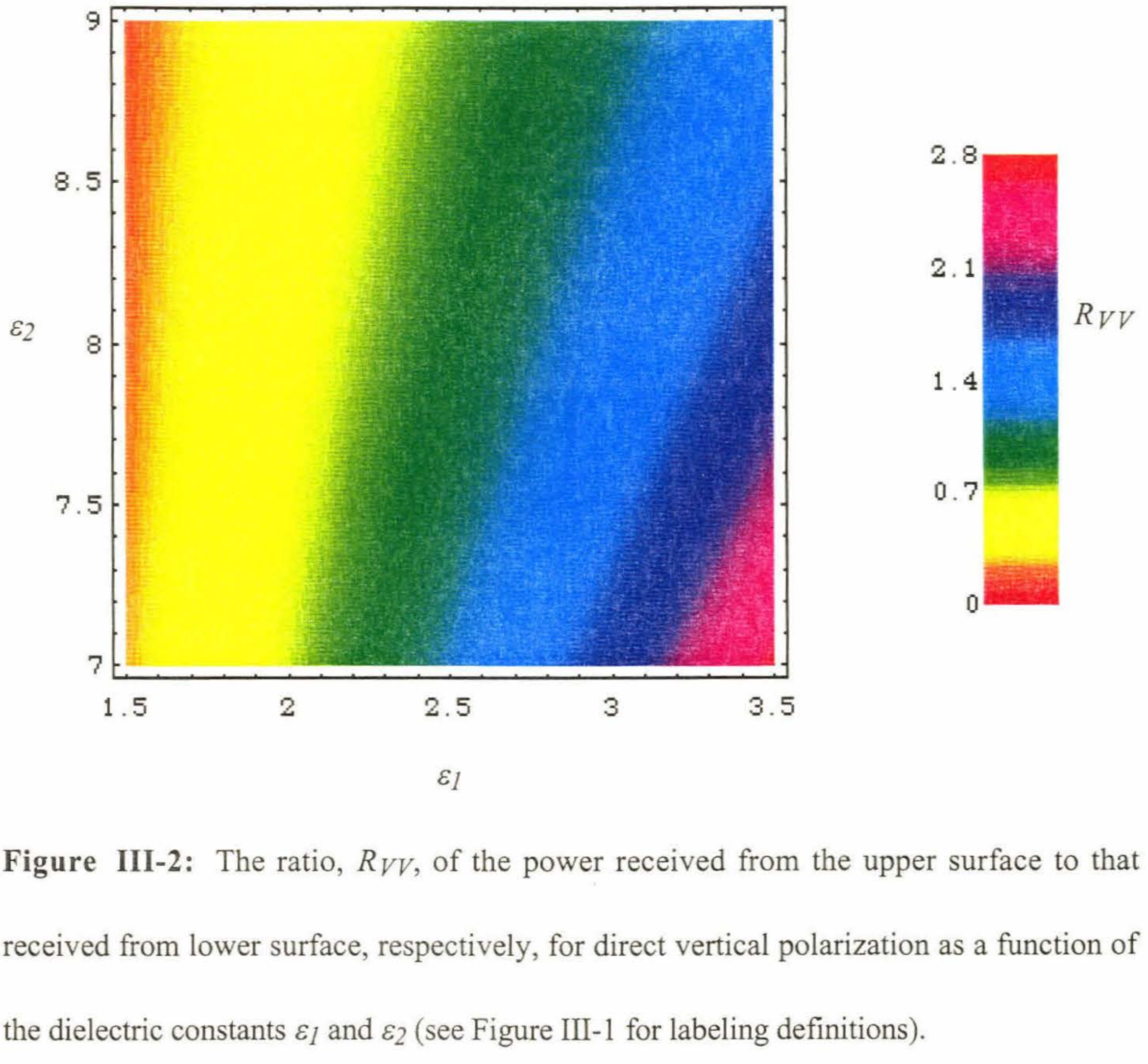


Figure III-2: The ratio, R_{VV}, of the power received from the upper surface to that received from lower surface, respectively, for direct vertical polarization as a function of the dielectric constants ε_1 and ε_2 (see Figure III-1 for labeling definitions).

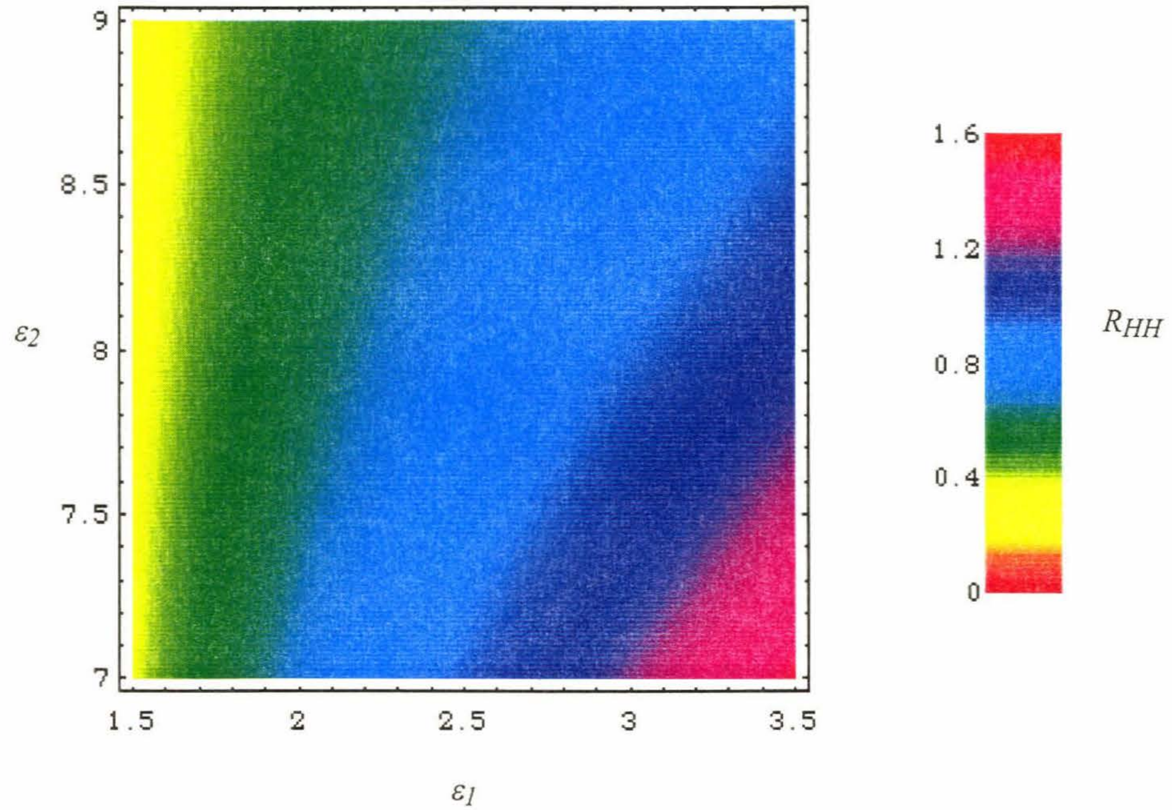


Figure III-3: The ratio, R_{HH}, of the power received from the upper surface to that received from lower surface, respectively, for direct horizontal polarization as a function of the dielectric constants ε_1 and ε_2 (see Figure III-1 for labeling definitions).

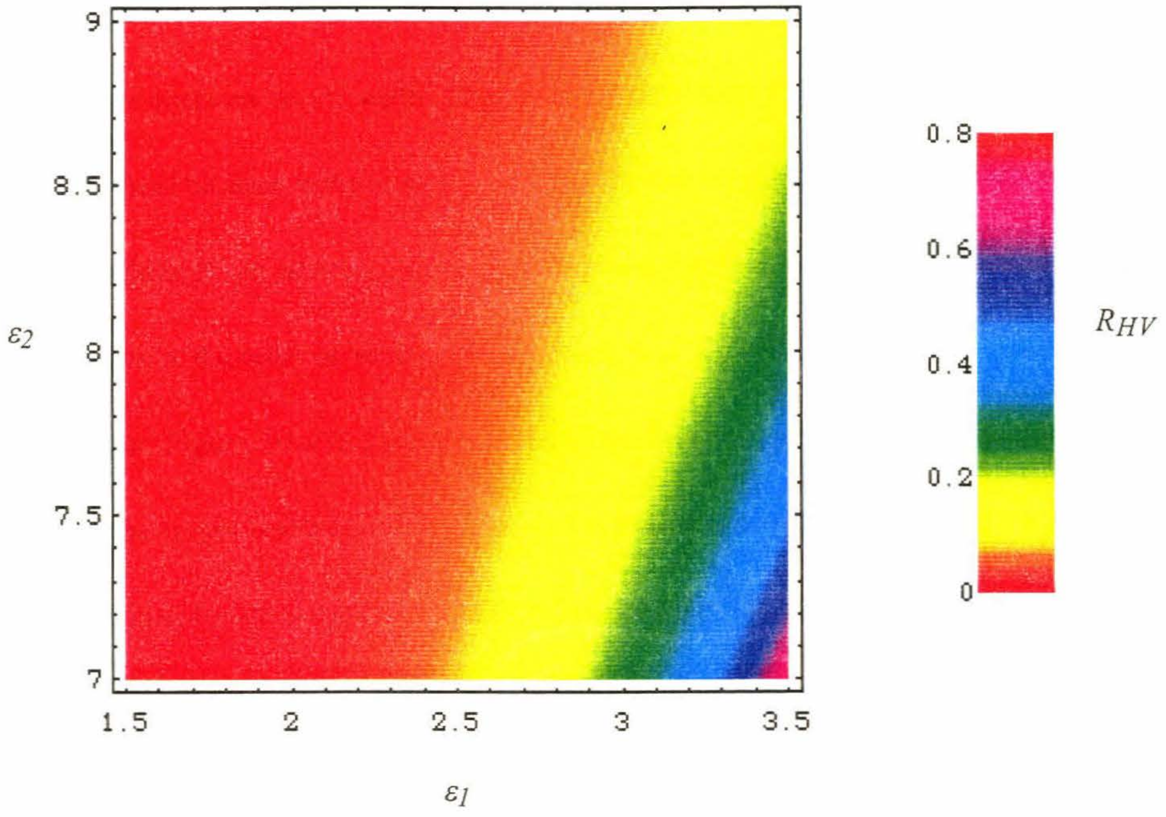


Figure III-4: The ratio, R_{HV}, of the power received from the upper surface to that received from lower surface, respectively, for cross polarization as a function of the dielectric constants ε_1 and ε_2 (see Figure III-1 for labeling definitions).

according to how strong the upper echo is, that is, how rough the upper surface is.

The fact that subsurface imaging on a regional scale is only possible in the very driest regions of the earth, where skin depths are the deepest, shows that in general for the wavelengths typical of imaging radar the skin depth does not reach down deep enough to reach buried surfaces on the earth. Therefore, by definition when subsurface imaging is achieved it will be done at the skin depth, except for the very driest places on earth such as the Selima Sand Sheet where the depth penetrated may only be a fraction of the skin depth, which can be of the order of meters there [*McCauley et al.*, 1982].

From Figure III-4 we see that subsurface imaging with cross polarization will minimize the power scattered by the upper surface relative to the lower surface we wish to image. In fact, for the nominal values $\varepsilon_1 = 2.5$ and $\varepsilon_2 = 8.0$, R_{HV} is only 0.04, whereas $R_{HH} = 0.77$ and $R_{VV} = 0.93$. Hence we have discovered how to minimize the effects of surface roughness on subsurface imaging, by using cross polarization. If direct polarization is used and the depth is near the skin depth, which is likely as I discussed above, then only if the upper surface is very smooth relative to the lower surface will the lower surface be imaged. In nature it will happen much more often that the upper surface has a roughness somewhere between somewhat rough and comparably rough relative to the lower surface than being very smooth relative to the lower surface, simply because there are more possible states in the former scenario. Therefore, if we limit ourselves to regions where the upper surface is very smooth relative to the lower surface in order to avoid the effects of surface roughness, we have greatly reduced the area over which subsurface imaging could be achieved if the surface roughness is taken into account, for example by imaging in cross polarization.

In fact, subsurface imaging was achieved through the Selima Sand Sheet by SIR-A with HH polarization [*McCauley et al.*, 1984; *Elachi et al.*, 1984]. This must mean the upper surface is very smooth, and indeed a very smooth upper surface was reported by field expeditions sent there to confirm the subsurface imaging [*McCauley et al.*, 1984]. Therefore, in this region the analysis by *Elachi et al.* [1984], predicated on the assumption of a very smooth upper surface, is valid. This analysis indicates that a covering layer can be beneficial to subsurface imaging for covering depths up to an appreciable fraction of the skin depth. However, there should also be areas within this region that have a rougher upper surface, and since I have SIR-C data with *VV* polarization, which gives the strongest surface echo (Figure III-2), it is likely that in these areas subsurface imaging is prevented due to a strong upper surface echo.

Though we have discovered that by imaging with cross polarization we can minimize the effects of surface roughness on subsurface imaging, in reality one would not expect an appreciable echo at cross polarization, because at least a double scattering is required to create a cross polarized echo, and this would not be very common for the surfaces characteristic of where subsurface imaging occurs. Therefore, we are forced to observe with direct polarization and deal with the strong surface echo that results.

Finally, now that I have discovered that there is such a strong surface echo in direct *VV* polarization, I exploit this fact in section III.19 to form the basis of a method for measuring soil moisture by measuring the depth of burial of significant moisture, which should effectively form a lower surface, using the ubiquitous upper surface echo.

III.3 Subsurface Imaging in the Presence of an Upper Surface Echo

Now that we have discovered that there will often be an upper surface echo as strong as the lower surface echo, and that it is desirable to separate the echoes and determine the depth of burial, I will describe a method for accomplishing these tasks. Figure III-5 illustrates the geometry of subsurface imaging in the presence of an upper surface echo, where three images are made from paths that are possibly displaced but definitely parallel. From such images, if the displacement of the imaging paths perpendicular to the line of sight is not too great, the change in phase of the radar echo, for both the magnitude and phase of a radar echo are recorded, may be related to the geometry of the situation. This is the technique of radar interferometry.

In Figure III-5 I show ray paths of the radar wave to the lower and upper surfaces from which echoes are recorded. The point here is that an imaging radar is "imaging" because the echoes received can be separated as coming from different parts of a two dimensional imaging plane. This separation is done by Synthetic Aperture Radar using range (actually time of flight) and Doppler shift, with the result that echoes which return with range and Doppler shift within a certain interval are are assumed to come from the same imaging point. Figure III-5 illustrates that the geometry of the situation is such that there is an echo from the upper surface that arrives back at the radar at the same time as an echo from the lower surface; these two echoes are indistinguishable to the radar and their sum is assigned to be the echo from this range and Doppler cell (pixel).

I may therefore write the amplitude of the same pixel observed from positions ξ, ψ and ζ as

$$x = l + u + noise_x \tag{III.11}$$

$$y = le^{iA} + ue^{i(A+\Delta)} + noise_y \tag{III.12}$$

$$z = le^{iC} + ue^{i(C+\Delta')} + noise_z \tag{III.13}$$

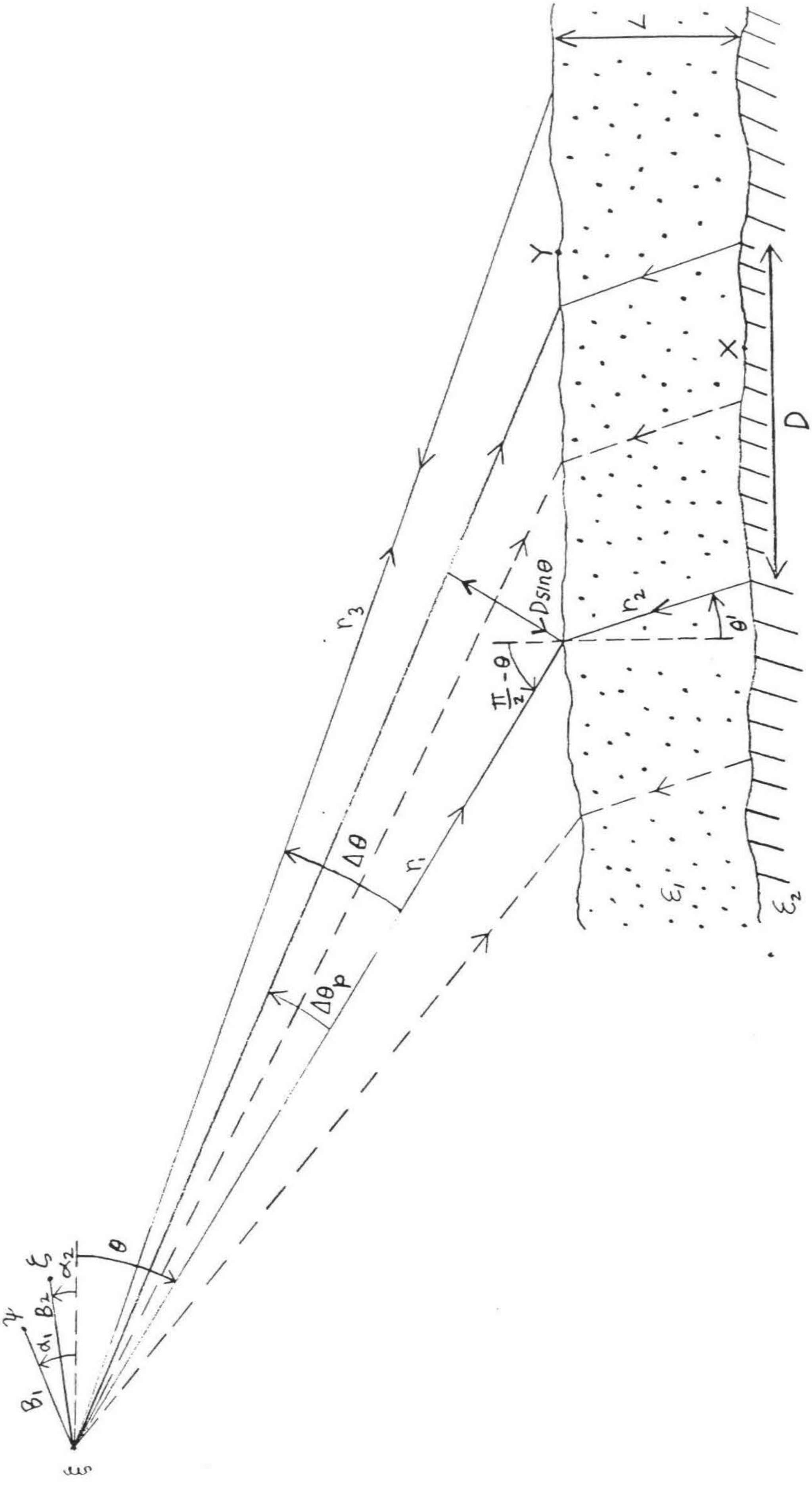
where the complex echoes from the lower and upper surfaces observed from ξ are l and u respectively. There are noises associated with each observation and A and A+Δ are phase *differences* between the lower and upper echoes, respectively, measured at ψ and those measured at ξ. Similarly, C and C+Δ' are phase *differences* between the lower and upper echoes, respectively, measured at z and those measured at ξ.

Figure III-5: The geometry of subsurface imaging in the presence of an upper surface echo. The solid line labeled r_1 to the upper surface, and then labeled r_2 as it continues, refracted, to the lower surface, is the echo from the middle of a certain pixel. The dashed rays to either side of this ray represent the echoes from the extremities of this pixel; that is, all echoes from the lower surface bounded by the two dashed rays are recorded within the time resolution of the radar and are assigned to come from the same pixel.

Displaced a distance D across track is the solid line representing the echo from the lower surface of the middle of the next pixel across track. It subtends an angle $\Delta\theta_p$ relative to the echo from the lower surface of the middle of the previous pixel.

The solid line labeled r_3 is the echo from the upper surface that is received at the same time as, but subtends an angle $\Delta\theta$ to, the echo from the lower surface with the rays labeled r_1 and r_2. It is therefore assigned to the same pixel as the echo with rays labeled r_1 and r_2.

If there was no refraction, the echoes from the lower and upper surface that are received at the same time would come from positions X and Y, respectively. The effect of refraction for values of ε_1 and theta typical of the SIR-C data used in this research is to increase the value of $\Delta\theta$ by about 30% compared with its refraction free value. This is illustrated to scale in Figure III-5.


L
Y
X
D
r_3
r_2
θ'
$D\sin\theta$
$\frac{\pi}{2}-\theta$
$\Delta\theta$
r_1
$\Delta\theta_p$
ε_1
ε_2
θ
B_1
ψ
α_1
B_2
ξ
α_2

I emphasize that these are phase *differences* between lower or upper echoes observed from different positions, and that the effects on the phase of the echo due to scattering from the lower or upper surface or due to propagation through the layer are the same no matter which of the three positions the radar wave is transmitted and received from.

The phase differences between the lower and upper echoes observed from the three different positions can be seen from the geometry illustrated in Figure III-5 to be defined by the equations

$$\frac{4\pi B_1}{\lambda}\cos\left(\theta+\alpha_1\right)+\varphi_y=2\pi N+A \tag{III.14}$$

$$\frac{4\pi B_1}{\lambda}\cos\left(\theta-\Delta\theta+\alpha_1\right)+\varphi_y=2\pi N+A+\Delta \tag{III.15}$$

$$\frac{4\pi B_2}{\lambda}\cos\left(\theta+\alpha_2\right)+\varphi_z=2\pi N'+C \tag{III.16}$$

$$\frac{4\pi B_2}{\lambda}\cos\left(\theta-\Delta\theta+\alpha_2\right)+\varphi_z=2\pi N'+C+\Delta' \tag{III.17}$$

The first term on the left side of each equation is the phase difference that results from the change in path length when the ray from the lower surface is observed from ψ, the ray from the upper surface is observed from ψ, the ray from the lower surface is observed from ζ and the ray from the upper surface is observed from ζ, respectively. The phases φ_y and φ_z are constants that represent possible constant phase shifts between ψ and ξ and ζ and ξ, respectively. Constant phase shifts such as these would occur because the phase of the outgoing wave as it is sent from the radar from position ξ is not necessarily the same as when it is sent from ψ and ζ.

The right hand side of each equation is the sum of an integer multiple of 2π plus the left hand side of the equation modulo 2π. This recognizes that the measurable phase difference is the actual phase difference modulo 2π, since the phase differences always appear in the equations as the exponent of the complex exponential which is periodic with period 2π. We may derive an expression for Δ by expanding the cosine term in (III.15) to first order in $\Delta\theta$, because this angle is very small, because the distance between the radar and the imaged area, r_3, is large compared to the depth of the layer:

$$\frac{4\pi B_1}{\lambda}\left(\cos\left(\theta+\alpha_1\right)\cos\Delta\theta+\sin\left(\theta+\alpha_1\right)\sin\Delta\theta\right)+\varphi_y=2\pi N+A+\Delta$$

$$\therefore\frac{4\pi B_1}{\lambda}\left(\cos\left(\theta+\alpha_1\right)1+\sin\left(\theta+\alpha_1\right)\Delta\theta\right)+\varphi_y=2\pi N+A+\Delta$$

$$\therefore\frac{4\pi B_1}{\lambda}\cos\left(\theta+\alpha_1\right)+\varphi_y+\frac{4\pi B_1}{\lambda}\sin\left(\theta+\alpha_1\right)\Delta\theta=2\pi N+A+\Delta$$

$$\therefore \text{ substituting (III.14) we get } 2\pi N+A+\frac{4\pi B_1}{\lambda}\sin\left(\theta+\alpha_1\right)\Delta\theta=2\pi N+A+\Delta$$

$$\therefore\Delta=\frac{4\pi B_1}{\lambda}\sin\left(\theta+\alpha_1\right)\Delta\theta \tag{III.18}$$

We may derive a similar expression for Δ':

$$\Delta'=\frac{4\pi B_2}{\lambda}\sin\left(\theta+\alpha_2\right)\Delta\theta \tag{III.19}$$

Putting the times of flight for the two rays shown in Figure III-5 equal we get

$$r_1+r_2\sqrt{\varepsilon_1}=r_3 \tag{III.20}$$

where I have used the fact that the radar wave travels a factor $\sqrt{\varepsilon_1}$ slower in the layer covering the lower surface than in the atmosphere above. Expressing the altitude of the radar above the upper surface in two different ways we get

$$r_1\sin\theta=r_3\sin\left(\theta-\Delta\theta\right) \tag{III.21}$$

The depth of the layer can be written as

$$r_2\cos\theta'=L \tag{III.22}$$

and Snell's law gives us

$$\sin\left(\frac{\pi}{2}-\theta\right)=\sin\theta'\sqrt{\varepsilon_1}$$

$$\therefore\ \cos\theta=\sin\theta'\sqrt{\varepsilon_1} \tag{III.23}$$

Combining equations (III.20) - (III.23) we can eliminate r_2, r_3 and θ' to get

$$\Delta\theta = \frac{\sin\theta}{\cos\theta\left(1 + \frac{r_1\sqrt{\varepsilon_1 - \cos^2\theta}}{L\,\varepsilon_1}\right)} \qquad \text{(III.24)}$$

Since $r_1 >> L$ we have to a very good approximation

$$\Delta\theta = \frac{L}{r_1\cos\theta}\left(\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) \qquad \text{(III.25)}$$

I have written $\Delta\theta$ this way because the term in parentheses $\rightarrow$ 1 as $\varepsilon_1 \rightarrow 1$, and the resulting expression can easily be seen to be correct for the case of no refraction. For $\varepsilon_1 = 2.5$, typical of the dielectric constant of many powdered rocks [*Campbell and Ulrichs*, 1969], and $\theta=40^\circ$ (incidence angle = 50°, which is the typical incidence angle for the Shuttle Imaging Radar) this term in parentheses is 1.16. This rises to 1.32 for $\varepsilon_1 = 3.5$, which is the measured dielectric constant of the sand through which subsurface imaging of dry river channels was achieved by SIR-A [*McCauley et al.*, 1982].

In Figure III-6 I plot the term in parentheses, now defined as $F(\varepsilon_1,\theta)$, for values of ε_1 varying from 1 to 4 and values of the angle θ varying from 10° to 60° (the incidence angle $\equiv 90^\circ - \theta$, so this corresponds to a range of incidence angles from 30° to 80°). Since I will show that the estimates for the separated lower and upper echoes and the depth of burial improve as $\Delta\theta$ increases, we see that the effect of refraction can be quite beneficial (except at large incidence angles).

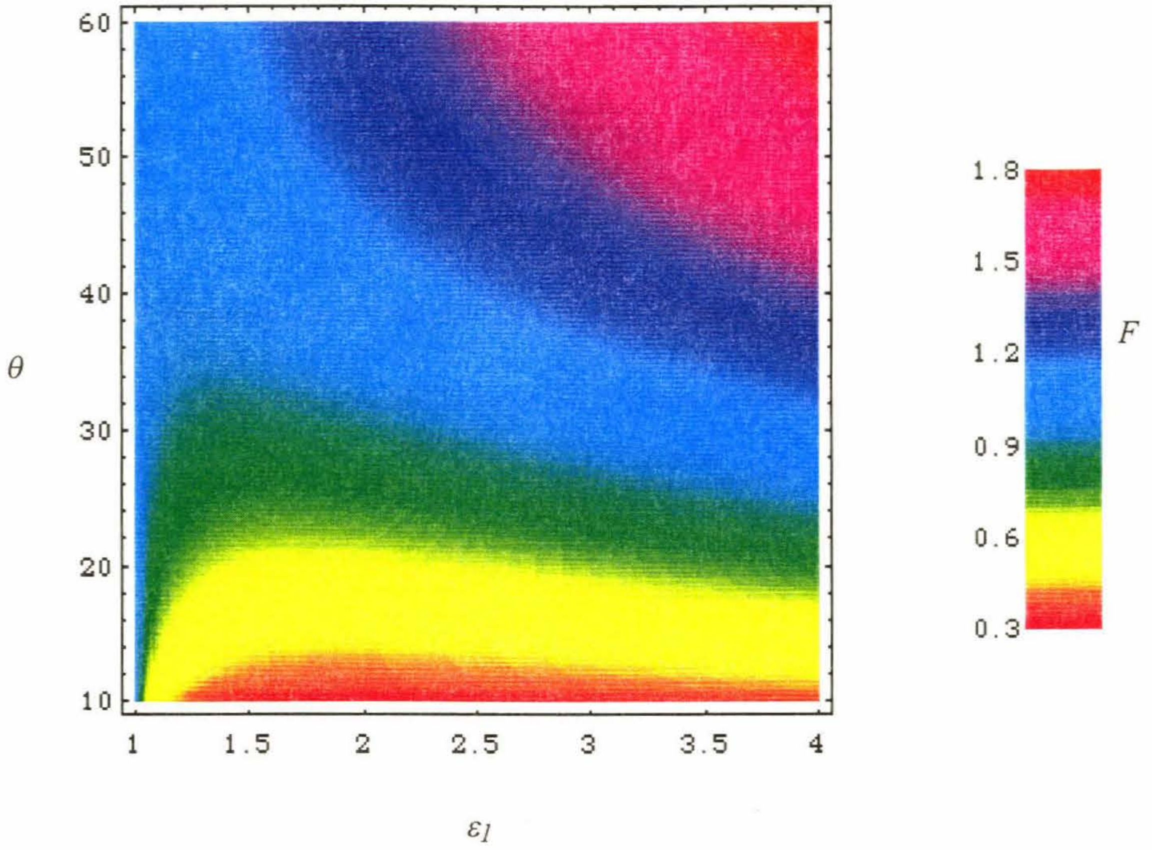


Figure III-6: The factor *F* by which $\Delta\theta$ is changed from its refraction free value.

$$F(\varepsilon_1,\theta)=\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1-\cos^2\theta}}$$

The incidence angle $\equiv 90^{\circ} - \theta$, so a range of incidence angles from 30° to 80° is shown in the above Figure (the SIR-C incidence angle is 50°). I will later show that the estimates for the separated lower and upper echoes and the depth of burial improve as $\Delta\theta$ increases, so we see that the effect of refraction can be quite beneficial (except at large incidence angles).

Substituting equation (III.25) into (III.18) we get an expression for Δ in terms of the geometry of the situation and the dielectric constant and depth of the layer:

$$\Delta = \frac{4\pi B_1}{\lambda}\sin\left(\theta + \alpha_1\right)\frac{L}{r_1\cos\theta}\left(\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) \tag{III.26}$$

A similar expression is obtained for Δ':

$$\Delta' = \frac{4\pi B_2}{\lambda}\sin\left(\theta + \alpha_2\right)\frac{L}{r_1\cos\theta}\left(\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) \tag{III.27}$$

Since $B_i \sin\left(\theta + \alpha_i\right)$ is the component of the i^{th} baseline (i=1 for the ξ - ψ baseline and i=2 for the ξ - ζ baseline) perpendicular to the line of sight to the pixel, we will write $B_i \sin\left(\theta + \alpha_i\right) = B_{\perp_i}$, allowing (III.26) and (III.27) to be rewritten more informatively as

$$\Delta = \frac{4\pi B_{\perp_1}}{\lambda}\frac{L}{r_1\cos\theta}\left(\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) \tag{III.28}$$

$$\Delta' = \frac{4\pi B_{\perp_2}}{\lambda}\frac{L}{r_1\cos\theta}\left(\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) \tag{III.29}$$

with the result that $\Delta = \frac{B_{\perp_1}}{B_{\perp_2}}\Delta'$.

It is apparent from equations (III.14) - (III.18) that Δ and Δ' result from a slightly different look angle to the upper surface compared to the lower surface. Because the phase difference between, for example, the lower echoes measured at position ψ and ξ depends on the exact look angle to the lower surface via the path length distance difference, a different look angle to the upper surface results in a different path length difference between echoes received at ψ and ξ and therefore a different phase difference. We see that the lower and upper echoes, indistinguishable in a single radar image, now have something about themselves that is different - the phase difference - when we look at multiple images.

Clearly then, the larger that Δ or Δ' is the more distinct is the difference between the lower and upper echoes from images ξ and ψ or images ξ and ζ, respectively. Therefore, the larger that one of these phases is the better we should be able to separate the lower and upper echoes and determine the depth of burial. From (III.28) and (III.29) we see that these phases are proportional to the depth of burial and the perpendicular component of the baseline, and inversely proportional to the wavelength. These then are the three parameters which will determine, for a given level of noise, how well the lower and upper echoes can be separated and the depth determined.

III.4 The Origin of Noise in the Multiple Observations of a Pixel

Contaminating each observation of the pixel is an associated noise. This noise causes deviation of the radar echo from the model terms on the right hand side of equations (III.11) - (III.13) and can have four distinct physical origins:

(i) reflected solar radiation, thermal emission from the ground and thermal emission within the receiver in the spectral region of the blackbody spectra of the sun, ground and receiver, respectively, near the observation wavelength. Together this is known as thermal noise.

(ii) multiple observations of the pixel will cause the distance between the observation point and each sub resolution facet that makes up the pixel to change compared with some reference observation of the pixel. If the multiple observations are made from locations displaced perpendicularly to the line of sight to the pixel from the reference observation then the distance to each sub resolution facet changes by a different amount, and the voltage vector raised in the antenna by the echo from that facet is rotated in phase (due to the different time of flight) by a different amount compared to the voltage vector from other facets. Hence the total voltage vector, being the vector sum of the voltage vectors from each facet, changes in phase and magnitude in general compared to the voltage vector observed from the reference observation. This change in the echo compared with the reference observation is known as spatial or baseline decorrelation.

(iii) If the multiple observations of a pixel are not made simultaneously, then there is the possibility of physical changes in the pixel between observations, which can be thought of as changes in the sub resolution facets which make up the pixel. Hence the vector sum of the echoes from each facet is different compared with some reference observation, and this change is called temporal decorrelation

(iv) If the multiple observations are not made simultaneously, there is the possibility of different path lengths through the atmosphere, and hence a spurious rotation in the phase of the echo, due to different amounts of water vapor in the troposphere [*Goldstein*, 1995].

If the observations are made near the same local time then the solar illumination and ground temperature will be approximately the same for each observation, and so the thermal noise will be equal in size, on average, for each observation. In the next section I describe fundamentally why a radar echo is described as a draw from a zero mean Gaussian distribution. For the same reasons thermal noise is also described in the same way, for it is also the sum of a myriad of tiny echoes (for example the reflected solar radiation from each facet within a pixel), each drawn from some well behaved distribution for the size of the echoes. I will assume that the observations are made near the same local time (as is the case for the SIR-C data with which I tested the method) and so the noise terms are independent and described as draws from a zero mean Gaussian distribution of a certain constant width.

Temporal and spatial decorrelation only have meaning relative to some reference observation; if there is only one observation, there is no such thing as temporal or spatial decorrelation.Therefore, to discuss these noises we must define a reference observation, and I define the second, or y, observation as such. This reference observation will be contaminated by thermal noise, but the other two observations, x and z, will be contaminated by noise representing change in the echoes due to spatial and temporal effects as well as thermal noise. I can characterize the noise associated with the temporal and spatial effects by considering that, with the effects of one or both of temporal and spatial decorrelation included, the resulting echo is still just the sum of a multitude of echoes from the sub resolution facets. The sub resolution facets may be altered (temporal decorrelation) or the echo from each may be differentially rotated in phase (spatial decorrelation), leading to a different total echo.

One can see intuitively that this echo, when plotted as a vector in complex voltage space, will have some random orientation relative to the reference observation echo. If the effects of temporal and spatial decorrelation are slight, the two vectors will lie close together, while if there is great temporal and/or spatial decorrelation the two vectors will be randomly positioned in complex space. Either way, the difference between the two vectors, which is defined as the spatial and temporal decorrelation noise, will be randomly oriented in complex voltage space. In fact, one can see that this vector has real and imaginary parts drawn from a zero mean Gaussian distribution of a certain width.This can be seen mathematically by noting that the sum (or difference) of two independent variables each drawn from a zero mean Gaussian distribution is a number that is described by a zero mean Gaussian distribution (whose standard deviation is the square root of the sum of the squares of the standard deviation of each of the two variables). This may also be seen more intuitively by considering that the difference of the two vectors can be thought of as the sum of the difference of the echoes from each sub resolution facet. These little difference vectors are randomly oriented in complex space, since the effect of spatial and temporal decorrelation will be random, and will have lengths drawn from some distribution with a zero minimum and some well defined maximum, depending on the amount of decorrelation. The sum of a multitude of such difference vectors is just a vector whose real and imaginary parts are random draws from a zero mean Gaussian distribution, as I explain in the next section.

There is also noise associated with the possibility of different amounts of water vapor in the troposphere for each observation, which results in a different path length for each observation. In testing the method with SIR-C data, however, I show that the predominant source of noise, which prevents the method from detecting the presence of a layer, is temporal decorrelation. My proposed solution to this problem of excessive temporal decorrelation between observations is to make the observations simultaneously, and this will also eliminate the possibility of different amounts of water vapor in the troposphere for each observation. Therefore, since temporal decorrelation is the major source of noise, and eliminating this noise also eliminates the effect of different amounts of water vapor, there is no need to separately consider the effects of different amounts of water vapor in the troposphere.

For the x and z observations of the pixel, there will then be a total noise term that is the sum of three zero mean Gaussian noise vectors, representing thermal, spatial and temporal decorrelation. The total vector is the sum of these, so we see that in the most general case equations (III.11) - (III.13) hold, but the width of the noise terms are different. The noise contaminating the reference observation y is just the thermal noise, while the noise terms contaminating x and z are the sum of the thermal noise contaminating each observation plus the spatial and temporal decorrelation between the x and z observation, respectively, and the y observation.

I will show later in testing the method with SIR-C data of the Egypt/Sudan border where subsurface imaging occurs that temporal decorrelation between observations produces an effect over 20 times the signature expected from a layer.This effect is so large that it is pointless to try to model the temporal decorrelation in the hopes of separating it from the effect of a layer; we would never know whether the residual signature was due to a layer or a fluctuation in the much larger effect of temporal decorrelation. By demonstrating that even data considered to be highly correlated suffers too much temporal decorrelation between observations to detect a realistic thickness layer I argue that temporal decorrelation prevents detection of a subsurface layer, and that observations must be made simultaneously to eliminate this effect.

With temporal decorrelation eliminated, we are left with thermal and spatial decorrelation. I show in section III.19 that spatial decorrelation may be minimized by making the pixel width narrow compared with the layer depth, and propose a way to do this, leaving thermal decorrelation, which I will show can be dealt with by making the layer thick in phase by using a long baseline. *Having identified a regime where the signature from a layer is not swamped by temporal or spatial decorrelation effects*, I will therefore concentrate on solving the equations (III.11) - (III.13) in this regime, that is, assuming only thermal noise is present and creating a method for dealing with this thermal noise.

III.5 Extracting the Lower and Upper Echoes and Depth of Burial from Multiple Radar Images

Equations (III.11) to (III.13), which define multiple radar images of a scene where subsurface imaging occurs in the presence of an upper echo, contain 8 unknowns - l, u, A, C, Δ and Δ'. Both l and u are complex and therefore contain two unknowns each. These equations contain 6 known quantities; x,y and z, also complex. Clearly, while equations

(III.11) to (III.13) may look superficially simple, they are in fact literally complex, nonlinear and under specified, and their solution in the presence of noise for the separated echoes l and u and the phases A, C, Δ and Δ' is non trivial.

Because for a single pixel there are more unknowns than known quantities, we are obviously going to have to assume that some variables do not vary from pixel to pixel. That way if we average over a group of pixels within which certain variables are assumed constant, it might be possible to gain more known quantities than unknowns and at least then a solution may be possible.

We can immediately reject l and u from consideration as constant over a group of pixels. The reason is that a radar echo from a surface, which is what l and u are from the lower and upper surfaces, respectively, can be thought of as the vector sum of a myriad of tiny vectors that represent the complex echoes (magnitude and phase) from small facets that the surface can be assumed to be composed of. The orientation of each of the tiny voltage vectors (the echo is received by the radar antenna in which a complex voltage is raised) is random, reflecting the random orientation and distribution of the facets. The length of these tiny vectors will be described by some distribution reflecting the orientation of the facets and the dielectric constant of the surface, and one can see intuitively that the distribution will extend from zero length tapering off above some maximum value, and will be well behaved in between. Therefore, the total voltage vector can be described as an independent random walk in both the real and imaginary directions, where the length of each step of the random walk in each direction is described by some well behaved distribution function and there are very many steps taken.

In this case the position of the random walker is described by the Gaussian distribution function [*Feynman et al.*, 1963], and one can see that the exact position and orientation of the small facets will change from pixel to pixel even for another pixel that is the same in an average sense. Therefore, the steps taken by the random walk in each direction are different and the final position will be totally different, so that even for a surface that is constant in any physical sense of the word (that is, not truly constant on a microscopic scale which would be bizarre for a natural surface) the echo from different parts of the surface will be wildly different; in fact, the real and imaginary parts will be random draws from a zero mean Gaussian distribution. This variation, even from a "constant" surface, is the phenomenon called "speckle", and reflects the fact that the echo from a single pixel literally contains no physically useful information; only the average of the echoes from a number of pixels has any physical meaning. This gives the standard deviation of the Gaussian distribution from which the real and imaginary components of the echo are drawn.

As a result, the values of l and u will most definitely *not* be constant among neighboring pixels. We are left with considering the phases A, C, Δ and Δ'. Fortunately, it is plausible that these phases are constant over a group of neighboring pixels. If both the lower and upper surfaces are flat over this group of pixels, then Δ and Δ' are constant because the depth of the layer is constant. The only change in A and C is that due to the geometry of a flat plane, which can be removed because with radar data we know the dimensions of a pixel. Hence if we assume no topographic variation of either the upper or lower surfaces over a patch of pixels, then the values of A, C, Δ and Δ' are constant within the patch. If the patch contains N pixels, we see that the first pixel contributes 8 unknowns, l, u, A, C, Δ and Δ', and 6 known quantities (the measured data); x,y and z. However, every additional pixel in the patch now only contributes an additional 4 unknowns, the unique values of l and u for that pixel (leading to a total of $8+(\mathrm{N}-1)\times 4=4(\mathrm{N}+1)$ unknowns), while still offering 6 new known quantities; the data x,y and z for that pixel (leading to a total of 6N known quantities). Hence we see that with at least two pixels we will have as many known quantities as unknown (12), and every additional pixel adds 2 more known quantities over unknowns, providing a redundancy that will allow solution in the presence of noise.

III.6 Using "Tie points" to Reduce the Number of Unknowns

"Tie points" are so called because they are points of known altitude within the image; they tie the altitude at that point. If we have such points then that is the same as knowing the look angle θ to that point, since the altitude of the radar above the point (that is, the tie points are points of known altitude relative to the altitude of the radar) may be converted into look angle via $\sin\theta=\frac{altitude}{r_1}$. Now equations (III.14) - (III.17) can be rewritten as

$$\frac{4\pi B_1}{\lambda}\cos\left(\theta+\alpha_1\right)+\left(\varphi_y-2\pi N\right)=A \tag{III.30}$$

$$\frac{4\pi B_1}{\lambda}\cos\left(\theta-\Delta\theta+\alpha_1\right)+\left(\varphi_y-2\pi N\right)=A+\Delta \tag{III.31}$$

$$\frac{4\pi B_2}{\lambda}\cos\left(\theta+\alpha_2\right)+\left(\varphi_z-2\pi N'\right)=C \qquad \text{(III.32)}$$

$$\frac{4\pi B_2}{\lambda}\cos\left(\theta-\Delta\theta+\alpha_2\right)+\left(\varphi_z-2\pi N'\right)=C+\Delta' \qquad \text{(III.33)}$$

where I have grouped the constant phase shifts with the integer multiples of 2π to form one unknown. The constant phase shifts are of course constant over the image, and while the integer multiples of 2π do vary across the image, their variation can be accounted for by "counting fringes": counting how many fringes - cycles of 2π in the phase difference between x and y or x and z - have been cycled through relative to the tie points. Therefore, the second term on the right hand side of each equation is effectively a constant across the image, and there are two such unknown constants, $\left(\varphi_y-2\pi N\right)$ and $\left(\varphi_z-2\pi N'\right)$.

There are 4 other unknowns that are constant across the image; the orbital parameters B_1, α_1, B_2 and α_2. Hence there are 6 unknowns constant across the image, and if one has 3 tie points at places where there is only one surface echo, say a rocky outcrop where subsurface imaging cannot occur, then there are 6 equations ((III.30) and (III.32) for each tie point) for the 6 unknowns, with the phases A and C directly measured from the data, since there is only a single surface involved. The angle θ at each tie point is known so these equations may be solved for the constants, and then everywhere else in the image, including places where there is subsurface imaging, I can now, for a given value of A, work out θ from (III.30). Putting this value of θ into (III.32) I can obtain the value of C. That is, I have C given A. From (III.28) and (III.29) we have

$$\Delta=\frac{B_1\sin\left(\theta+\alpha_1\right)}{B_2\sin\left(\theta+\alpha_2\right)}\Delta' \qquad \text{(III.34)}$$

so we see that even just knowing the orbital parameters allows us to express Δ' in terms of Δ, since only approximate knowledge of θ is needed, which we always have.

Tie points require knowledge of the altitude at three locations, which is often problematic, and require additional bookkeeping, but reduces the number of phase unknowns from 4 to 2. In this case three images of the same pixel have 6 unknowns, l, u, A and Δ, which equals the number of known quantities, the data x, y and z. However, in order to solve for the unknowns in the presence of noise some redundancy is needed, more known quantities than unknowns, so we will still have to average over a patch of pixels, assuming constant topography. Once we have at least 2 pixels for the case of no tie points (4 unknown phases) or just 1 pixel if we have tie points (2 unknown phases) , then each additional pixel provides 2 more known quantities than unknowns. That is, if we have no tie points, we can gain the same accuracy in the solution as if we had tie points by including just one more pixel in our patch than when we have tie points, for this one extra pixel provides the additional 2 more known quantities than unknowns that tie points offer.

Since there is no functional improvement in accuracy (since I will be averaging over > 20 pixels and so the additional accuracy afforded by including one more pixel in the average is not significant) and tie points require topographic data which can be problematic to obtain for the earth and unavailable on other planets, I do not use tie points.

III.7 Separating the Echoes and Determining the Depth of Burial

I will assume then that A, C, Δ and Δ' are constant over a patch of N neighboring pixels, which we have seen is the same as assuming that the upper and lower surfaces are flat within this patch, and solve for the 4(N+1) unknown quantities, A, C, Δ, Δ', l_i and u_i within this patch. l_i and u_i are the lower and upper echoes, respectively, from the i^{th} pixel within the patch. The N pixels afford 6N known quantities, the data x_i, y_i and z_i, where i = 1,...,N, which are the voltages raised in the radar antenna due to echoes from the i^{th} pixel received at positions x, y and z of Figure III-5.

Applying equations (III.11) - (III.13) to each pixel within the patch we get 6N equations (considering both the real and imaginary components) that express the data x_i, y_i and z_i in terms of the unknown quantities, A, C, Δ, Δ', l_i and u_i:

$$x_i = l_i + u_i + noise_{x_i} \tag{III.35}$$

$$y_i = l_i e^{iA} + u_i e^{i(A+\Delta)} + noise_{y_i} \tag{III.36}$$

$$z_i = l_i e^{iC} + u_i e^{i(C+\Delta')} + noise_{z_i} \tag{III.37}$$

where i = 1,...,N. We want to solve these 6N equations for A, C, Δ, Δ', l_i and u_i. By doing so we will have separated the lower and upper echoes (having solved for l_i and u_i, respectively) for each pixel within the patch, having started with data for each pixel that is a combination of these echoes. Since the upper echo is uncorrelated with the lower echo, one can see that its addition to the lower echo is like adding noise, and if the magnitudes are comparable the lower surface will not be clearly imaged. Therefore, separation of the upper and lower echoes is like removing noise from the lower echo, improving the image of the lower surface.

By solving for Δ for the patch we see from equation (III.28) that we could get the depth of burial, L, if we knew the value of the remainder of the right hand side of (III.28); that is, the quantity

$$\frac{4\pi B_{\perp_1}}{\lambda}\frac{1}{r_1\cos\theta}\left(\frac{\varepsilon_1\sin\theta}{\sqrt{\varepsilon_1-\cos^2\theta}}\right) \tag{III.38}$$

To evaluate the term in parenthesis we can use a generic value for ε_1, say $\varepsilon_1 = 2.5$ [*Campbell and Ulrichs*, 1969], since we see from Figure III-6 that this term is fairly constant against changes in ε_1 at constant θ, for most angles. θ and r_1 are known accurately enough from the raw data for the purpose of evaluating (III.38), and the wavelength of the radar wave, λ, is of course known, but the baseline term, $B_{\perp_1}$, is unknown.

One might think that the location of the imaging positions x, y and z would be known as a matter of course, but in the case of the Space Shuttle, data from which I will use later, the error in its known orbital position is of the order of the distance between the displaced orbits, on the order of 100 meters [*M. Kobrick*, personal communication, 1996]. Since low earth orbiting satellites can accurately receive the Global Positioning Satellite (GPS) signal [*Kursinski et al.*, 1996], use of a GPS receiver on the Space Shuttle would allow its location to be known to within meters (for a GPS receiver of military specification) , and $B_{\perp_1}$ could be reconstructed accurately enough from the raw data. In our case I could instead determine $B_{\perp_1}$ through the use of tie points, discussed previously. However, I want to avoid using tie points because such accurate topographic knowledge does not exist for many regions of the earth, let alone other planets where we may wish to also apply the method.

Instead, I will estimate $B_{\perp_1}$ using the fact that the width of the pixels along the ground in the direction of increasing range is known. Figure III-5 shows two dashed line rays on either side of the ray to the lower surface with incidence angle $\frac{\pi}{2} - \theta$. These two rays represent echoes from the lower surface that arrive back at the radar antenna at times $\pm \Delta t$ from the time that the echo along the ray with incidence angle $\frac{\pi}{2} - \theta$ is received. That is, all echoes received during the time interval $2\Delta t$ are assumed to come from the same pixel, and so all echoes received from within a range interval $c\,\Delta t$ are assumed to come from the same pixel, where c is the speed of the radar wave as it travels between the radar and the upper surface. This length $c\,\Delta t$ is the range resolution of the radar, and its projection along the ground, $D = c\,\Delta t \cos\theta$, is the width of a pixel along the ground in the range direction. We can see that refraction in the layer has no effect on the width of the pixel since the travel times of all the rays bounded by the dashed lines are identical within the layer. Since c is assumed to be the speed of light in a vacuum, Δt is a parameter of the radar that is known and θ is known accurately enough from the geometry of the situation such that the error in D is small, D is known and I can use this knowledge of D by considering the change in phase of pixels as we move in the direction of increasing range.

The decrease in look angle as we move one pixel across in the direction of increasing range is shown in Figure III-5 as $\Delta\theta_p$. From the geometry of the situation illustrated in Figure III-5 we see that

$$\Delta\theta_p = \frac{D \sin\theta}{r_1} \tag{III.39}$$

If A, defined by equation (III.14), is the phase difference between the lower echo measured at y and at x, respectively, for the pixel with look angle θ, then for the next pixel across in the direction of increasing range the phase difference will be $A + \Delta_p$, where

$$\Delta_p = \frac{4\pi B_{\perp_1}}{\lambda} \Delta\theta_p \qquad \text{(III.40)}$$

where I have used equation (III.18) with $\Delta\theta_p$ substituted for $\Delta\theta$ and $B_i \sin\left(\theta + \alpha_i\right) = B_{\perp_i}$.

The phase of the quantity $x_i{}^*y_i$ is the phase difference between the i^{th} pixel measured from positions x and y, respectively. This phase difference, shown as a color representing values from 0 to 2π, may be plotted for the pixels and the resulting fringe patterns that can result are the familiar interferometric fringes, whose shape contains within it topographic information [*Zebker and Goldstein*, 1986]. If the surface has moved between observations (for example an ocean surface or glacier, with the time between observations measured in fractions of a second and days, respectively) then the shape of the fringes manifests a combination of topographic and motion information, with the latter dominant for a flat surface [*Goldstein et al.*, 1993].

If the upper surface is so smooth that we only have a lower echo then the phase of $x_i{}^*y_i$ is A, and moving across the image in the direction of increasing range will cause A to vary by the amount per pixel given by equation (III.39). Eventually A will change by 2π which corresponds to a single interferometric fringe. By counting the number of fringes N_f cycled through due to moving N_p pixels across track, the phase change per pixel, $2\pi N_f/N_p$, can be accurately determined (assuming the region is basically flat so there are no kinks in the fringes due to topography). Equating this phase change per pixel with equation (III.40), and substituting in equation (III.39) for $\Delta\theta_p$, we get an expression for $B_{\perp_1}$:

$$B_{\perp_1} = \frac{N_f}{N_p} \frac{\lambda r_1}{2D \sin \theta} \tag{III.41}$$

If the upper surface is so smooth that we only have a lower echo, this expression allows us to determine $B_{\perp_1}$ directly from the data, without the use of tie points or accurate knowledge of the spacecraft's position.

If the upper surface is of comparable roughness to the lower surface, so we have comparable upper and lower echoes, then equation (III.41) can still be used to determine $B_{\perp_1}$, where N_f is still the number of fringes cycled through with a displacement of N_p pixels in the direction in increasing range. This follows from considering that the phase difference between the x_i and y_i of equations (III.35) and (III.36), respectively, will be A if $u_i = 0$, and B if $l_i = 0$. One can intuitively see that that the phase will be somewhere in between A and B when l_i and u_i are both non-zero. Due to the phenomenon of speckle previously discussed, it will happen that some pixels have near zero values for one of l_i or u_i, even if the standard deviations of the Gaussian distributions from which the real and imaginary parts of l_i and u_i, respectively, are drawn are not near zero. Therefore, I expect the phase difference between x_i and y_i to retain the underlying fringe pattern of the lower (or upper) surface, but with the addition of "noise" representing the phase oscillating between A and B as speckle makes one or the other of l_i or u_i close to zero for that pixel.

I confirm this intuition by simulating images acquired at positions ξ and ψ, using equations (III.35) and (III.36) to determine the value of x_i and y_i for each pixel. I assume a phase change of 0.1 radians per pixel in the direction of increasing range for the phase of both the upper and lower surfaces. This is the phase change of a single surface due solely to the geometry of the situation, which is a manifestation of the decrease in look angle due to a displacement in the direction of increasing range. For a single surface this phase change will result in interferometric fringes will a spacing of about 63 pixels. In simulating x_i and y_i using equations (III.35) and (III.36) I use values of l_i and u_i that are complex numbers whose real and imaginary parts are all random draws from a zero mean Gaussian distribution with standard deviation equal to 1. Therefore, the upper and lower echoes are taken to be equal in strength, on average. The value of A at the pixels with minimum range is taken as 2 radians, and the value of Δ will be taken as either 0 (one surface) or 1 radian.

I simulate noise_x and noise_y from equations (III.35) and (III.36) as complex numbers whose real and imaginary parts are all random draws from a zero mean Gaussian distribution with standard deviation equal to n. n is taken as either 0 (no noise) or 0.3 (13.5 dB signal to noise ratio).

Figures III-7 to III-10 shows images of the phase of x^*y for this simulated data, where the directions of increasing range and increasing distance along the imaging path are shown. The number of pixels shown is 500 in the direction of increasing range times 500 in the direction of increasing distance along the imaging path. In Figure III-7 $n=0$ and $\Delta=0$, representing the baseline case of the fringes from a single surface with no noise. Figure III-8 shows the case $n=0$ and $\Delta=1$, simulating fringes from two surfaces with no noise.

Figures III-9 and III-10 are repeats of Figures III-7 and III-8, respectively, but with n=0.3. We see that in all cases, including the realistic noisy images, the addition of an upper echo does add to the noise in the fringes, but the basic underlying spatial fringe frequency is not altered, confirming our intuition. Therefore, I may use equation (III.41) to estimate $B_{\perp_1}$ even when there are both lower and upper surface echoes.

With $B_{\perp_1}$ determined directly from the data via equation (III.41), we can solve for the depth of burial L given Δ using equation (III.28). Therefore, solving equations (III.35) - (III.37), for i = 1,...,N (a total of 6N equations predicated on the assumption that both the lower and upper surfaces of the patch of N pixels is flat), gives us the *separated* lower and upper echoes for each pixel i within the patch, and also gives us the depth of the layer, L, for the patch, via equation (III.28). Hence, *assuming that the correlation between any two observations is approximately equal, which requires that the images suffer from only thermal decorrelation and no appreciable spatial or temporal decorrelation*, we will have accomplished our goal of separating the lower and upper echoes and determining the depth of burial, in absolute terms, all without any a priori knowledge of the scene we are imaging or of the radar's position at the three imaging locations ξ, ψ and ζ.

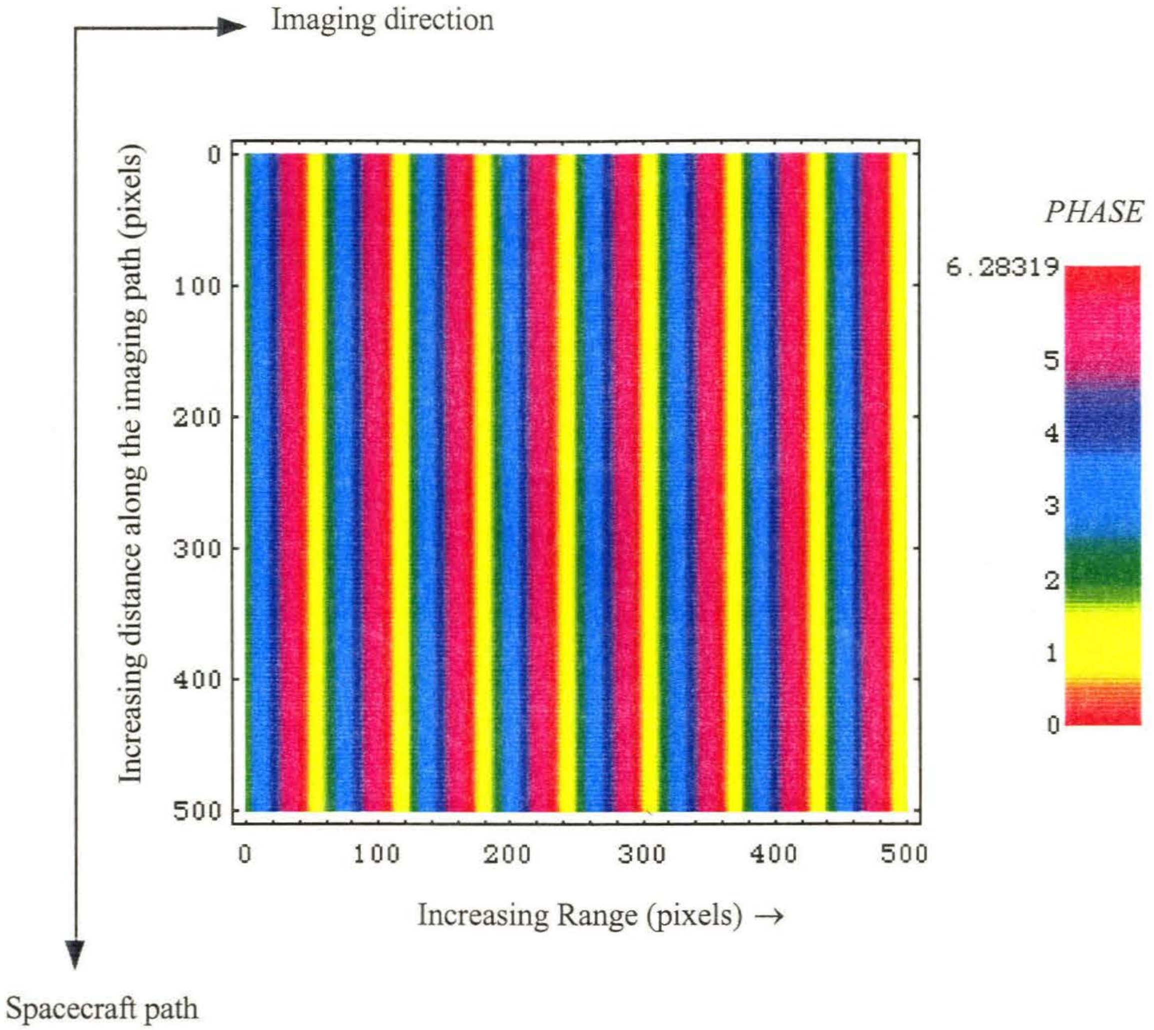


Figure III-7: The phase of x*y for the simulated noise free data (n=0) from a single surface (Δ=0). A phase change of 0.1 radians per pixel in the direction of increasing range is assumed. This phase change results in interferometric fringes with a spacing of about 63 pixels.

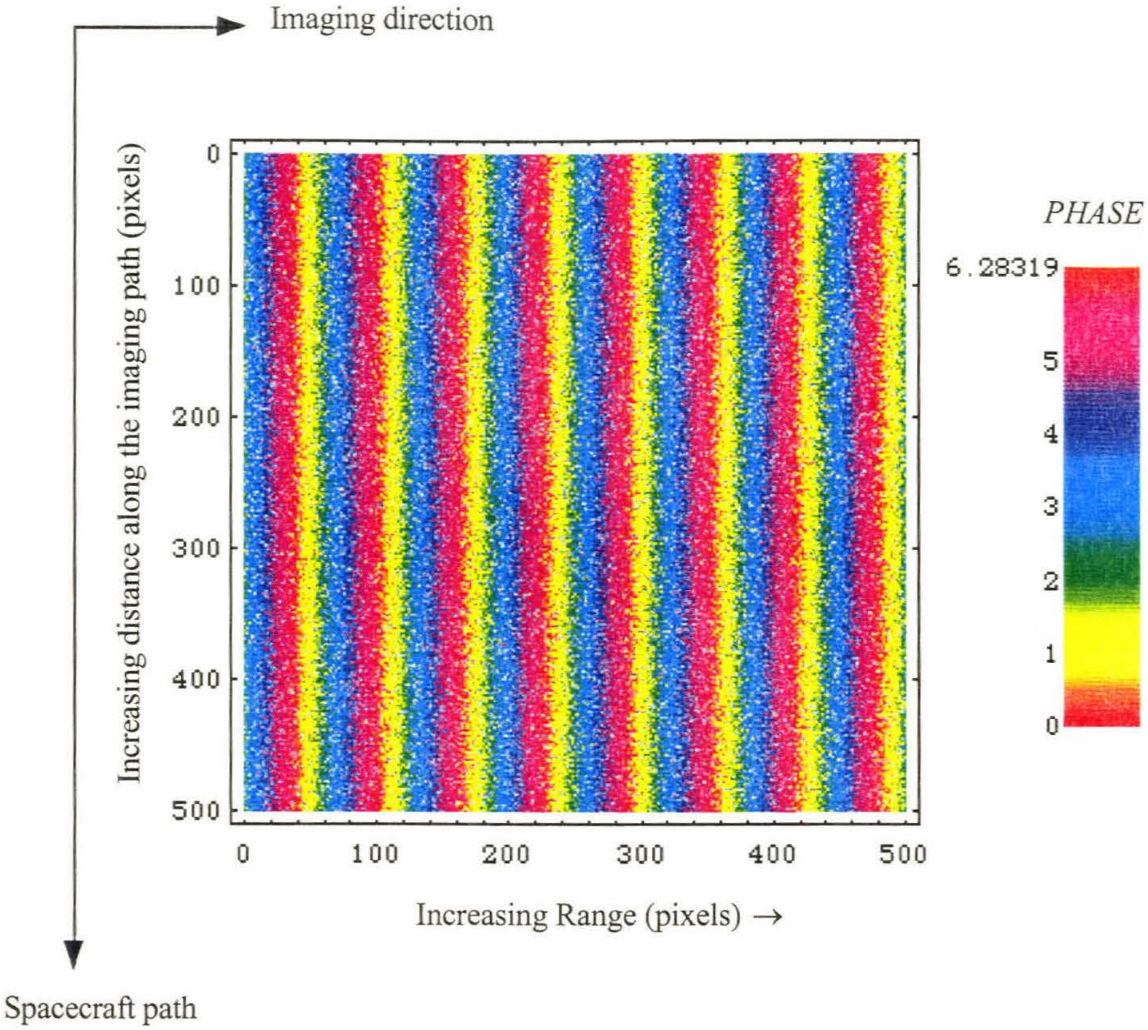


Figure III-8: The phase of x*y for the simulated noise free data (n=0) from two surfaces, with a phase spacing of $\Delta=1$ between the lower and upper surfaces. The strength of the echoes from the lower and upper surfaces are taken to be equal in strength, on average. A phase change of 0.1 radians per pixel in the direction of increasing range for the phase of both the lower and upper surfaces is assumed. This phase change would result in interferometric fringes with a spacing of about 63 pixels for each surface (previous Figure), and it can be seen from the above Figure that the phase of the combined echo from both surfaces has the same basic underlying spatial fringe frequency.

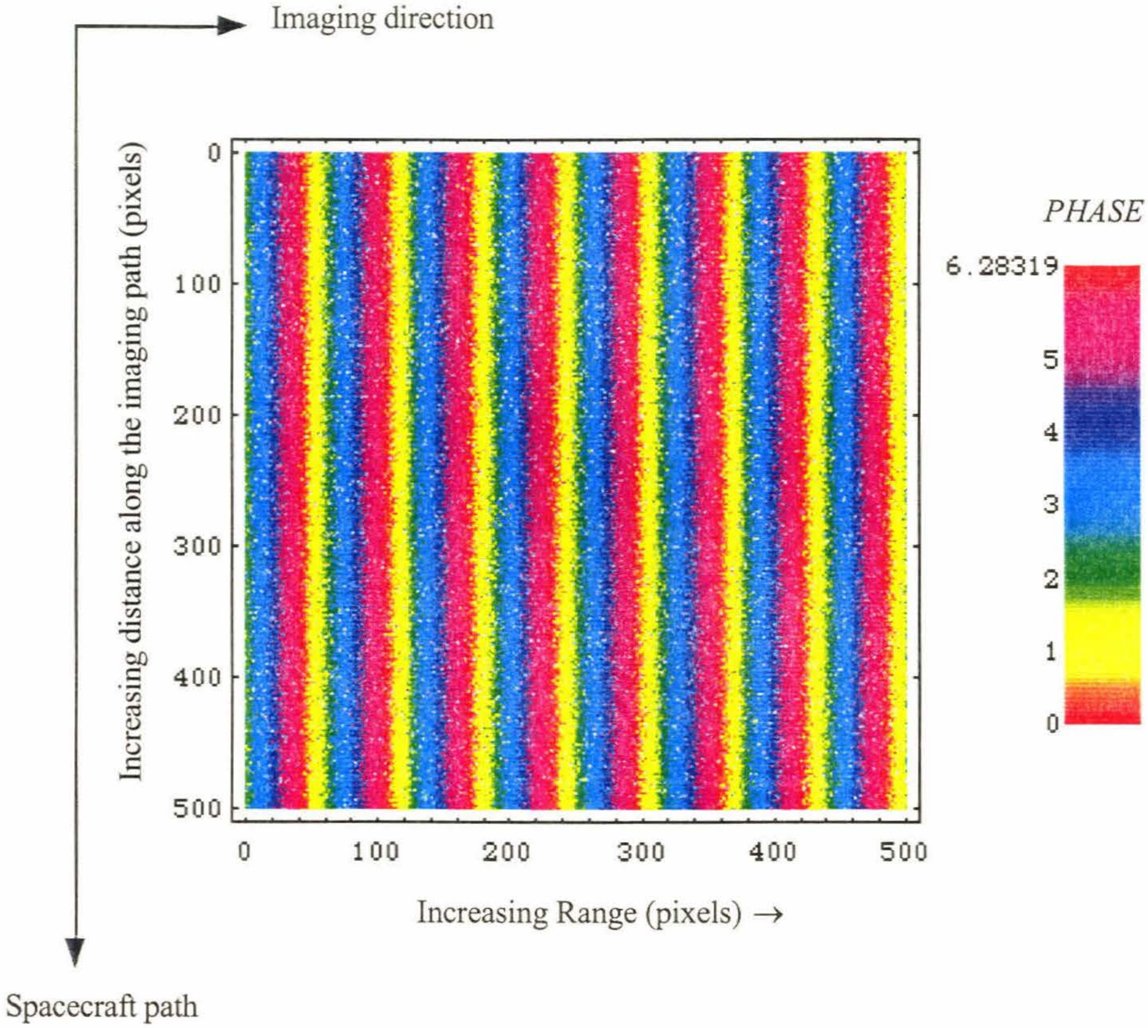


Figure III-9: The phase of x*y for the simulated noisy data (n=0.3) from a single surface (Δ=0). The power signal to noise ratio is approximately 10 dB, a typical value for imaging radar. A phase change of 0.1 radians per pixel in the direction of increasing range is assumed. This phase change would result in interferometric fringes with a spacing of about 63 pixels for a single noise free surface, and it can be seen that the addition of noise does not change the basic underlying spatial fringe frequency.

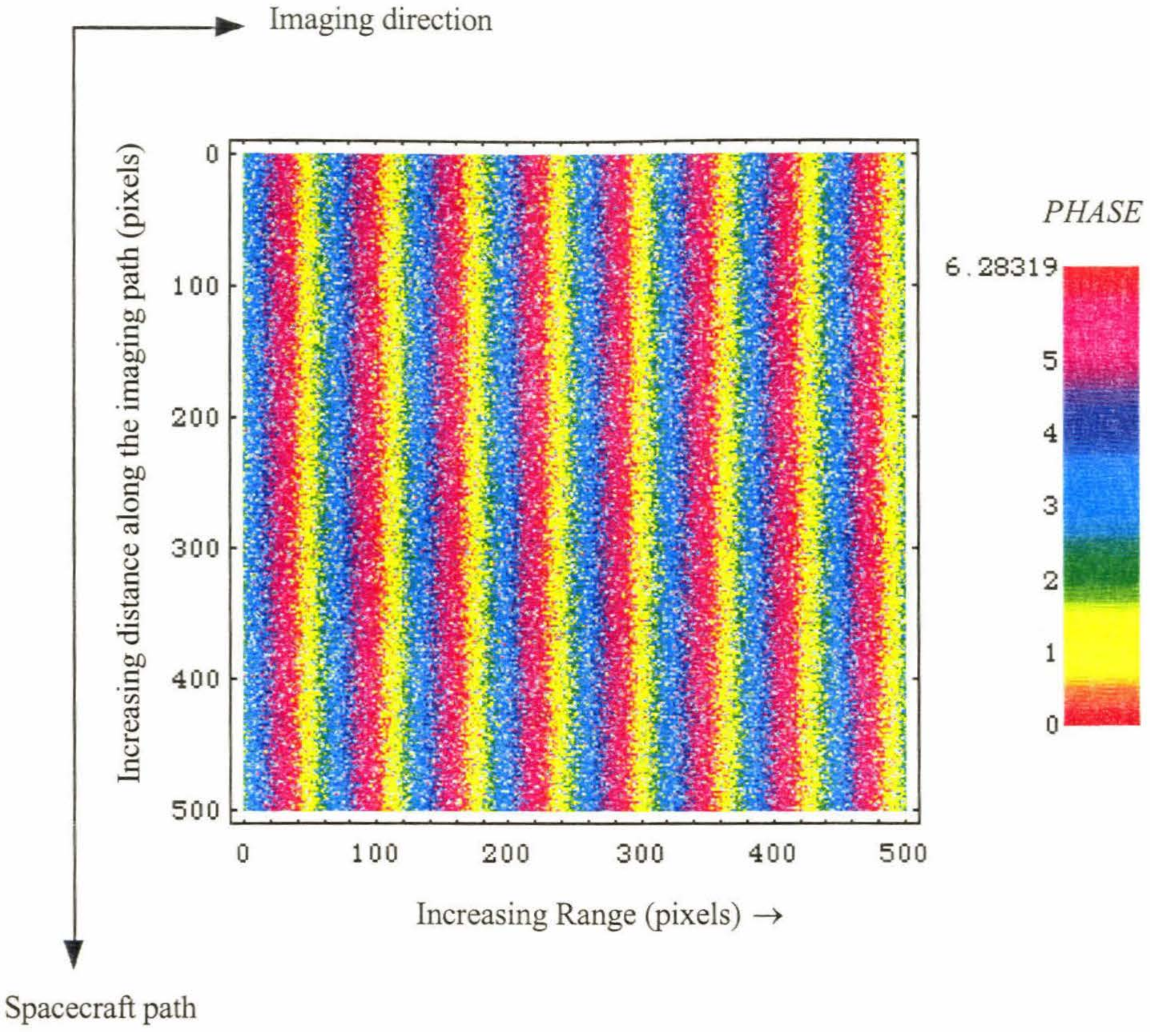


Figure III-10: The phase of x*y for the simulated noisy free data (n=0.3) from two surfaces, with a phase spacing of $\Delta=1$ between the lower and upper surfaces. The strength of the echoes from the lower and upper surfaces are taken to be equal in strength, on average. The power signal to noise ratio is approximately 10 dB, a typical value for imaging radar. A phase change of 0.1 radians per pixel in the direction of increasing range for the phase of both the lower and upper surfaces is assumed. This phase change would result in interferometric fringes with a spacing of about 63 pixels for each surface (Figure III-7), and it can be seen from the above Figure that the phase of the combined echo from both surfaces, with the addition of noise, has the same basic underlying spatial fringe frequency.

III.8 Solution

The noise terms of equations (III.35) - (III.37), $noise_{x_i}$, $noise_{y_i}$ and $noise_{z_i}$, respectively, for i = 1,...,N, which we are assuming represent thermal noise only, have two significant properties:

(i) they are independent of each other and all other terms in the 6N equations

(ii) their real and imaginary parts (for they are complex) are well characterized as random draws from a zero mean *Gaussian* distribution of certain standard deviation

These two properties are evocative of the language used to describe the requirements for a maximum likelihood estimate of the parameters of a function $y(x)$ given data points y'_i: *"Suppose that each data point y_i has a measurement error that is independently random and distributed as a normal (Gaussian) distribution around the "true" model $y(x)$"* [*Press et al.*, 1992].

Pursuing this analogy, we consider, following *Press et al.* [1992], that we *are* fitting a function $y(x) = y(x;a_1...a_M)$ that has M adjustable parameters a_j, $j = 1,...,M$, to N' data points (x_i,y'_i). Even though we are actually fitting discrete equations to data, not a continuous function to a set of data points, in the latter case what we are really doing is fitting the continuous function $y(x)$ *evaluated at the discrete points* x_i to the data y'_i - which is exactly the same as fitting a set of discrete equations to data. Therefore, we may indeed view the solution of the 6N equations (III.35) - (III.37), for i = 1,...,N, for the 4(N+1) unknowns A, C, Δ, Δ', l_i and u_i, as the fitting of a function $y(x;a_1...a_M)$ with M=4(N+1) adjustable parameters a_j to N'=6N data points y'_i. The M=4(N+1) adjustable parameters a_j

are the unknowns A, C, Δ, Δ', l_i and u_i, for i = 1,...,N (l_i and u_i contain 2 unknowns each since they are complex). The N'=6N data points y'_i are the data x_i, y_i and z_i, for i = 1,...,N (x_i, y_i and z_i are all complex and so contain 2 unknowns each).

We can consider the noise associated with each of the 6N equations (III.35) - (III.37), for i=1,...,N, to be the measurement error, and, as I have discussed, the error for each data point (for each equation) y'_i is independently random and is drawn from a zero mean (distributed around the true model (equation) $y(x)$) Gaussian distribution. Furthermore, we will assume that the standard deviations σ of these Gaussian distributions is the same, which should be the case with radar data taken at similar times of the day (as is the case for the data will I use later to test the method), since the thermal noise should be the same, on average, *and we are assuming that thermal noise is the major contributor to the noise contaminating each observation (relatively little spatial and temporal decorrelation).*

For a given set of values of the 4(N+1) unknowns A, C, Δ, Δ', l_i and u_i, for i = 1,...,N, the probability that the data observed would actually have resulted *given this set of values* is the probability that all the measurement errors were such that the difference between the model $y(x_i)$ and the data y'_i for each data point i=1,...,N'=6N is accounted for by the measurement error of each data point. Since the probability of a measurement error lying between y and $y+\Delta y$ is given by the Gaussian probability density times Δy :

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2\right]\Delta y \qquad \text{(III.42)}$$

the probability of all the measurement errors lying within Δy of $y'_i - y(x_i)$, the amount necessary for each measurement error to account for the difference between the model $y(x_i)$ and the data y'_i for each data point i, is given by

$$P \propto \prod_{i=1}^{N'} \left\{ \exp\left[-\frac{1}{2} \left(\frac{y'_i - y(x_i)}{\sigma} \right)^2 \right] \Delta y \right\} \tag{III.43}$$

Since the *measurement errors* are independent, the total probability of all the particular measurement errors for each data point occurring is the product of the probabilities of the measurement error for each data point occurring.

Different sets of values of the 4(N+1) unknowns A, C, Δ, Δ', l_i and u_i, for i = 1,...,N, will have different probabilities, given by equation (III.43), that the data observed would actually have resulted *given these sets of values*. The set of values of A, C, Δ, Δ', l_i and u_i for which this probability is a maximum is the set of values of A, C, Δ, Δ', l_i and u_i *from which the observed data would have most likely come*. Intuitively, but with no mathematical basis whatsoever, as is the case for all maximum likelihood estimates [*Press et al.*, 1992], we identify this set of values from which the data would most likely have come *as the most likely set of values given the data.* The set of values of A, C, Δ, Δ', l_i and u_i for which the probability given by (111.43) is a maximum are the maximum likelihood estimates of these values, and, being the most likely set of values given the data, we can do no better.

Maximizing (III.43) is the same as maximizing its logarithm, or minimizing the negative of its logarithm, which is

$$\left[\sum_{i=1}^{N'} \frac{\left[y'_i - y(x_i)\right]^2}{2\sigma^2}\right] - N \log\left(\Delta y\right) \tag{III.44}$$

The number of data points *N'* and the infinitesimal width Δy are constants, so minimizing (III.44) is equivalent to minimizing the first term only. Since σ is assumed constant, *P* is a maximum when

$$\sum_{i=1}^{N'} \left[y'_i - y(x_i)\right]^2 \tag{III.45}$$

is a minimum. I will seek the set of values of A, C, Δ, Δ', l_i and u_i, for i = 1,...,N, that minimize (III.45).

Substituting the 6N (real and imaginary) values of x_i, y_i and z_i, for i=1,...,N from equations (III.35) - (III.37) into (III.45) as the data values y'_i, and substituting the corresponding right hand side of equations (III.35) - (III.37) into (III.45) for the value of the function $y(x_i)$, we can write (III.45) as

$$\begin{aligned}&\sum_{i=1}^{N} \left\{ \left|x_i - \left(l_i + u_i\right)\right|^2 + \left|y_i - \left(l_i e^{iA} + u_i e^{i(A+\Delta)}\right)\right|^2 + \left|z_i - \left(l_i e^{iC} + u_i e^{i(C+\Delta')}\right)\right|^2 \right\} \\ &\equiv \sum_{i=1}^{N} \chi_i^2 \\ &\equiv \chi^2\end{aligned} \tag{III.46}$$

Now the summation is over the pixels within the patch being averaged over, with each pixel contributing the term χ_i^2 defined in (III.46) towards the total value χ^2 that we wish to minimize (as a function of A, C, Δ, Δ', l_i and u_i, for i=1,...,N).

Given a set of values of A, C, Δ and Δ', the values of l_i and u_i that minimize χ_i^2, called l_{ml_i} and u_{ml_i}, respectively, may be easily determined since χ_i^2 is linear in l_i and u_i when the phases A, C, Δ and Δ' are fixed. Equating the partial derivative of χ_i^2 with respect to each of the 4 real and imaginary components of l_i and u_i to zero, we get 4 equations for these components that can be written as a 2×2 matrix equation of complex numbers:

$$\begin{bmatrix} u_{ml_i} \\ l_{ml_i} \end{bmatrix} = \begin{bmatrix} 3 & 1 + e^{-i\Delta} + e^{-i\Delta'} \\ 1 + e^{i\Delta} + e^{i\Delta'} & 3 \end{bmatrix}^{-1} \begin{bmatrix} x_i + e^{-i(A+\Delta)} y_i + e^{-i(C+\Delta')} z_i \\ x_i + e^{-iA} y_i + e^{-iC} z_i \end{bmatrix} \qquad \text{(III.47)}$$

For a given a set of values of A, C, Δ and Δ', each χ_i^2 is individually minimum, and therefore their sum, χ^2, is minimum, when the local values of l_i and u_i for that pixel are the values l_{ml_i} and u_{ml_i}, respectively, given by (III.47). I may therefore write χ^2 as a function of A, C, Δ and Δ' only, with the understanding that I have substituted l_{ml_i} and u_{ml_i} for l_i and u_i, respectively, so that the value of χ^2 so formed, called $\chi_{\min}{}^2$, is the minimum value of χ^2 for that set of values of A, C, Δ and Δ':

$$\chi_{\min}{}^2(A, C, \Delta, \Delta') =$$
$$\sum_{i=1}^{N} \left\{ \left| x_i - \left(l_{ml_i} + u_{ml_i} \right) \right|^2 + \left| y_i - \left(l_{ml_i} e^{iA} + u_{ml_i} e^{i(A+\Delta)} \right) \right|^2 + \left| z_i - \left(l_{ml_i} e^{iC} + u_{ml_i} e^{i(C+\Delta')} \right) \right| \right.$$

(III.48)

Let $(A_{\min}, C_{\min}, \Delta_{\min}, \Delta'_{\min})$ be the coordinates at which *this* function is a minimum, which I determine using the Fletcher-Reeves-Polak-Ribiere multidimensional conjugate gradient method described by *Press et al.* [1992]. These coordinates, and the corresponding values of l_i and u_i given by

$$\begin{bmatrix} u_i \\ l_i \end{bmatrix} = \begin{bmatrix} 3 & 1 + e^{-i\,\Delta_{\min}} + e^{-i\,\Delta'_{\min}} \\ 1 + e^{i\,\Delta_{\min}} + e^{i\,\Delta'_{\min}} & 3 \end{bmatrix}^{-1} \times \begin{bmatrix} x_i + e^{-i\,(A_{\min} + \Delta_{\min})}\, y_i + e^{-i\,(C_{\min} + \Delta'_{\min})}\, z_i \\ x_i + e^{-i\,A_{\min}}\, y_i + e^{-i\,C_{\min}}\, z_i \end{bmatrix} \tag{III.49}$$

are the set of values of A, C, Δ, Δ', l_i and u_i, for i=1,...,N for which χ^2 is a global minimum, and are therefore the most likely set of these values given the data, which is our solution.

III.9 Confirmation of the Solution by Numerical Simulation

I can test the validity of the solution derived in the previous section by numerically simulating the radar echoes received from a patch of N pixels at three positions x, y and z, according to equations (III.35) - (III.37). I specify the values of A, C, Δ and Δ' that characterize the patch, and the standard deviation σ_n of the Gaussian distribution from which the real and imaginary parts of the noise in equations (III.35) - (III.37) is randomly drawn. I also specify the standard deviations σ_l and σ_u of the Gaussian distributions from which the real and imaginary parts of l_i and u_i, respectively, for i=1,...,N, are randomly drawn.

In this case I am assuming that the lower and upper echoes are constant, on average, across the patch, although the method of solution I described in the previous section is not predicated on this assumption. Our solution *is* predicated on the assumption that A, C, Δ and Δ' are constant within a patch of N pixels that will be averaged over.

However, the lower and upper echoes from each pixel within the patch, l_i and u_i, respectively, for i=1,...,N, could in principle be drawn from different standard deviation distributions for each pixel. However, for simplicity I will assume in the simulations that the echoes are constant, on average, across the patch.

I applied the method of solution to the simulated data and found the most likely values of A, C, Δ, Δ', l_i and u_i, for i=1,...,N, given the simulated data. I averaged the l_i and u_i, for i=1,...,N, to get the standard deviations σ_l and σ_u, respectively, via the equation

$$\sigma_l = \sqrt{\frac{\langle |l_i|^2 \rangle}{2}} = \sqrt{\frac{\sum_{i=1}^{N} l_i l_i^*}{2N}} \tag{III.50}$$

with a similar equation for σ_u. There is a factor of 2 in the denominator because l_i is complex, so there is an estimate for the standard deviation from averaging both the real and imaginary parts.

In Figures III-11 - III-20 I plot the solutions for the most likely values of A, C, Δ, Δ', σ_l and σ_u given the simulated data. These solutions are labeled $\hat{A}$, $\hat{C}$, $\hat{\Delta}$, $\hat{\Delta}'$, $\hat{\sigma}_l$ and $\hat{\sigma}_u$, respectively. For each scenario in the Figure there is a different set of values of the parameters used to create the simulated data; A, C, Δ, Δ', σ_l, σ_u, σ_n and N. The power signal to noise ratio is given by (in dB)

$$SNR = 10 \log \left(\frac{\sigma_l^2 + \sigma_u^2}{\sigma_n^2} \right) \tag{III.51}$$

For each scenario I show 3 plots; $\hat{A}$ vs $\hat{\Delta}$, $\hat{C}$ vs $\hat{\Delta}'$ and $\hat{\sigma}_l$ vs $\hat{\sigma}_u$. Each point within a plot represents the solution for *one* simulated patch of data, containing *N* pixels. Multiple points represent solutions obtained from multiple simulated patches. Each patch is unique because the particular values of the lower and upper echoes and noises for each pixel within the patch are random, constrained only through being drawn from Gaussian distributions with the specified standard deviations. The scatter in the resulting plot of solutions for multiple simulated patches, with the parameters defining the scenario held constant, therefore illustrates the validity and accuracy of the solution; a solution from one simulated patch that happens to be close to the true values of the parameters could occur by chance, but if the solutions for multiple patches are all scattered around the true values of the parameters, rather than being scattered randomly throughout the solution space, then the method is working, and the accuracy of the solution is defined by the width of the scatter.

For the cases of $\Delta' \cong 0.25$ some fine band like structure is visible in the scatter plots. This is an artifact of the method I use to find the global minimum that becomes manifest for thin layers, when the function being minimized is not well behaved. I alleviate this problem by taking many (18) initial guess for the starting values of Δ and Δ' and selecting the solution which has the lowest minimum. One can intuitively see from the plots that if we were to try more starting values the banding would be less pronounced, so the banding is merely a consequence of the finite spacing of the grid of starting values of Δ and Δ' that we begin the minimum finding algorithm from. This is conclusively demonstrated by Figure III-21 in which I repeat the simulation shown in Figure III-20 but using 756 (instead of 18) different possible starting values - the slight banding apparent in Figure III-20 disappears in Figure III-21. For thicker layers no banding is apparent, because the function being minimized is more well behaved, with a more pronounced minimum, as the layer gets thicker.

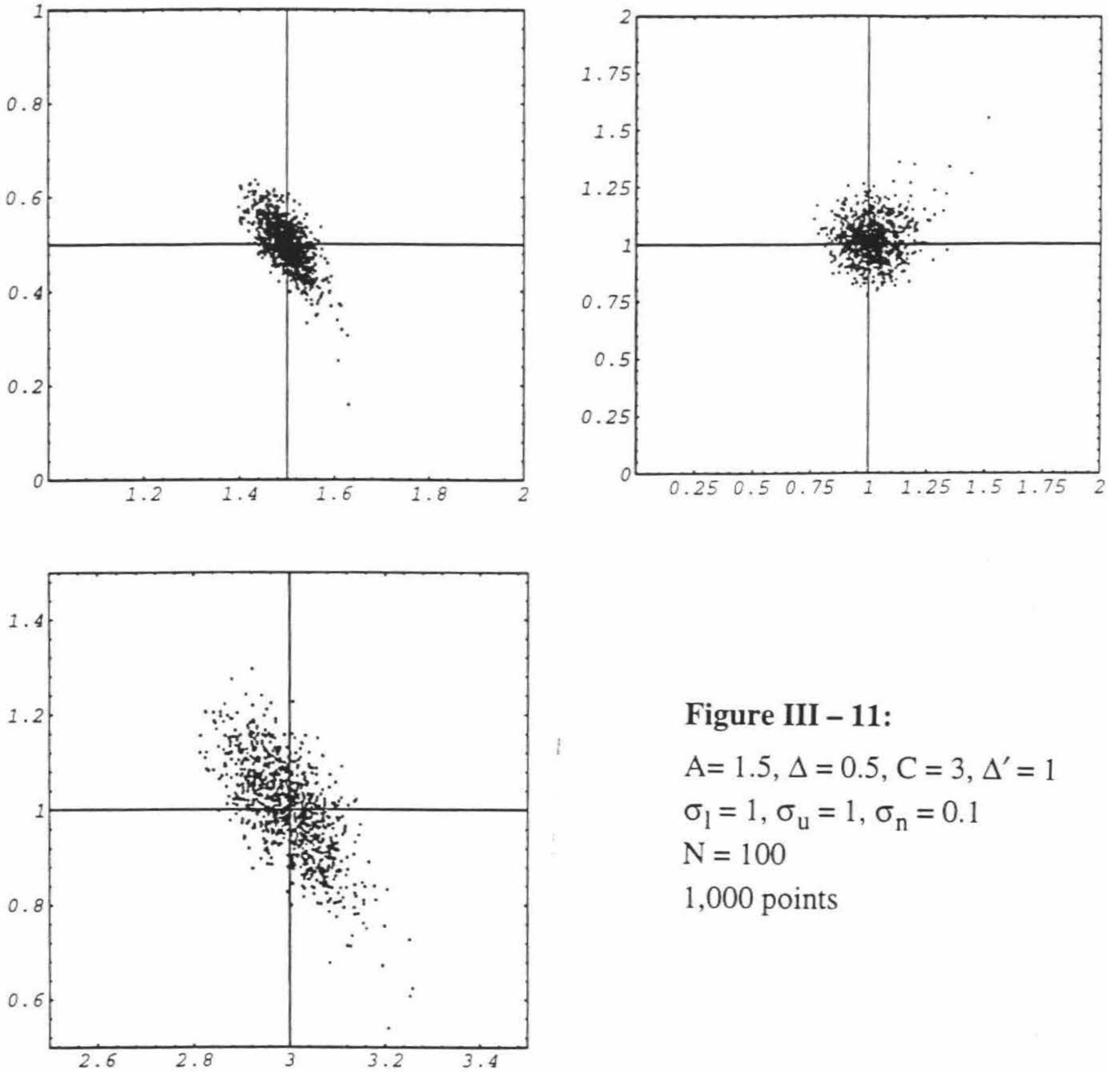


Figure III – 11:

$A = 1.5$, $\Delta = 0.5$, $C = 3$, $\Delta' = 1$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

$N = 100$

1,000 points

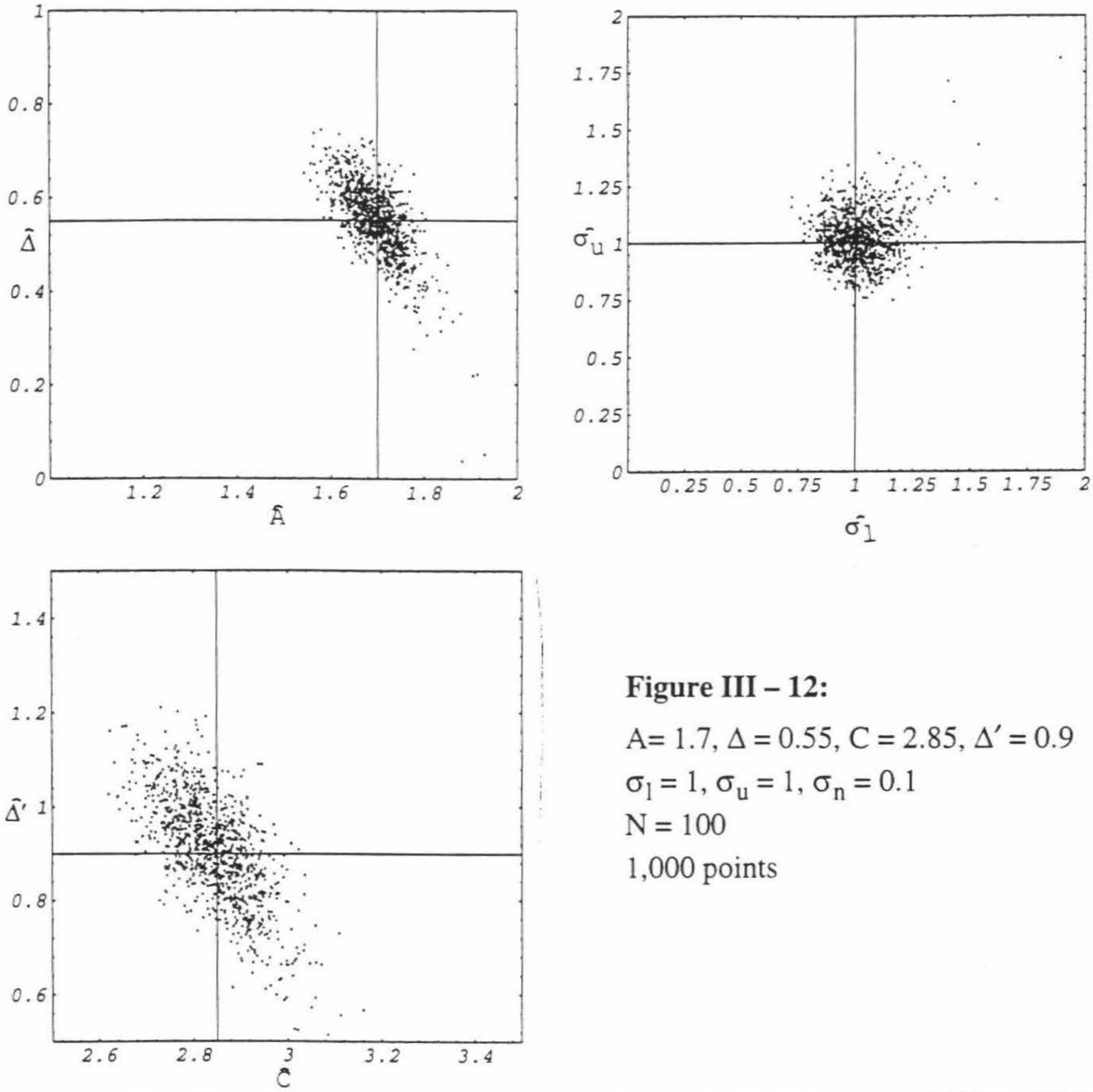


Figure III – 12:

A= 1.7, $\Delta = 0.55$, C = 2.85, $\Delta' = 0.9$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

N = 100

1,000 points

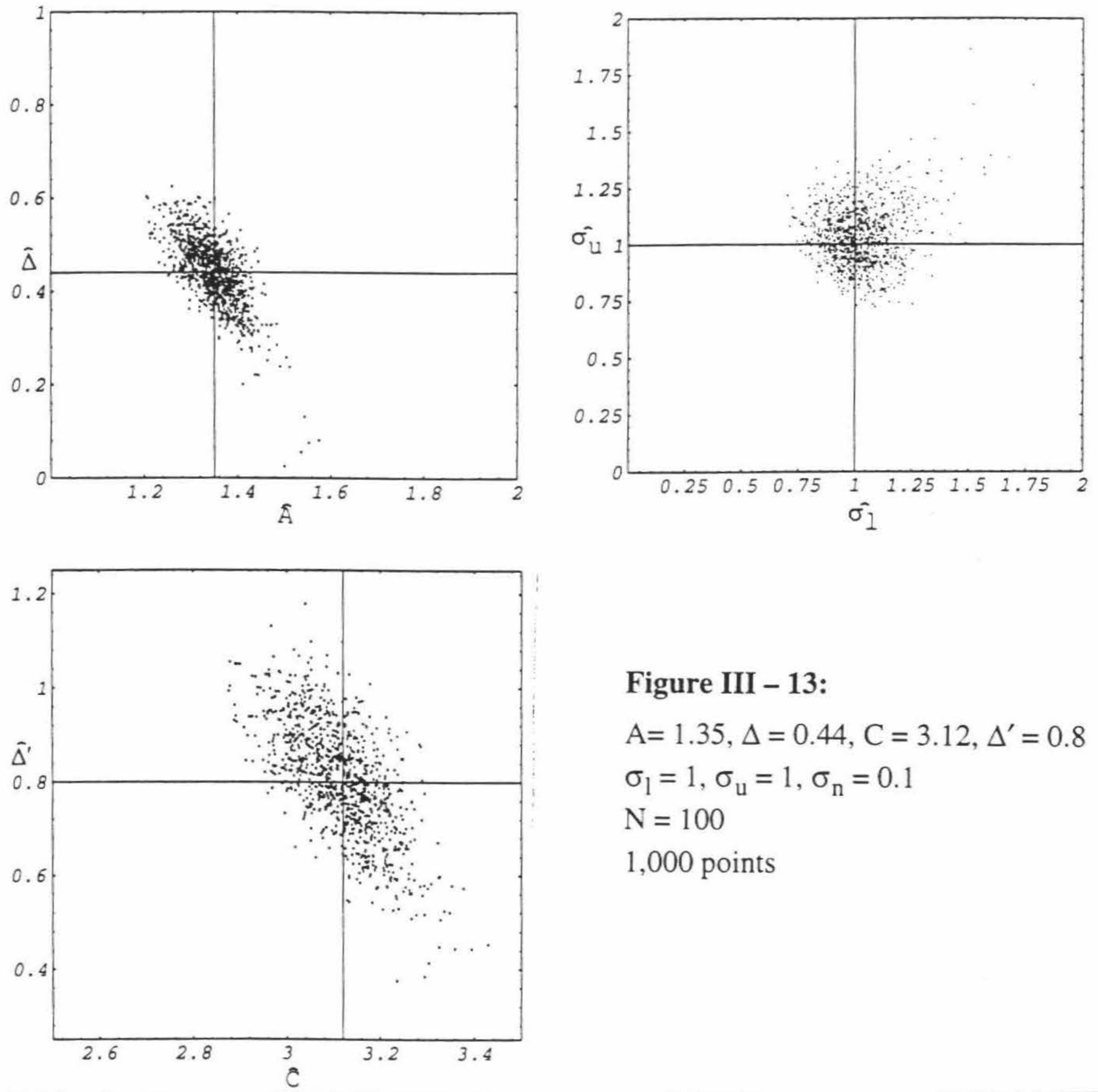


Figure III – 13:

$A = 1.35$, $\Delta = 0.44$, $C = 3.12$, $\Delta' = 0.8$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

$N = 100$

1,000 points

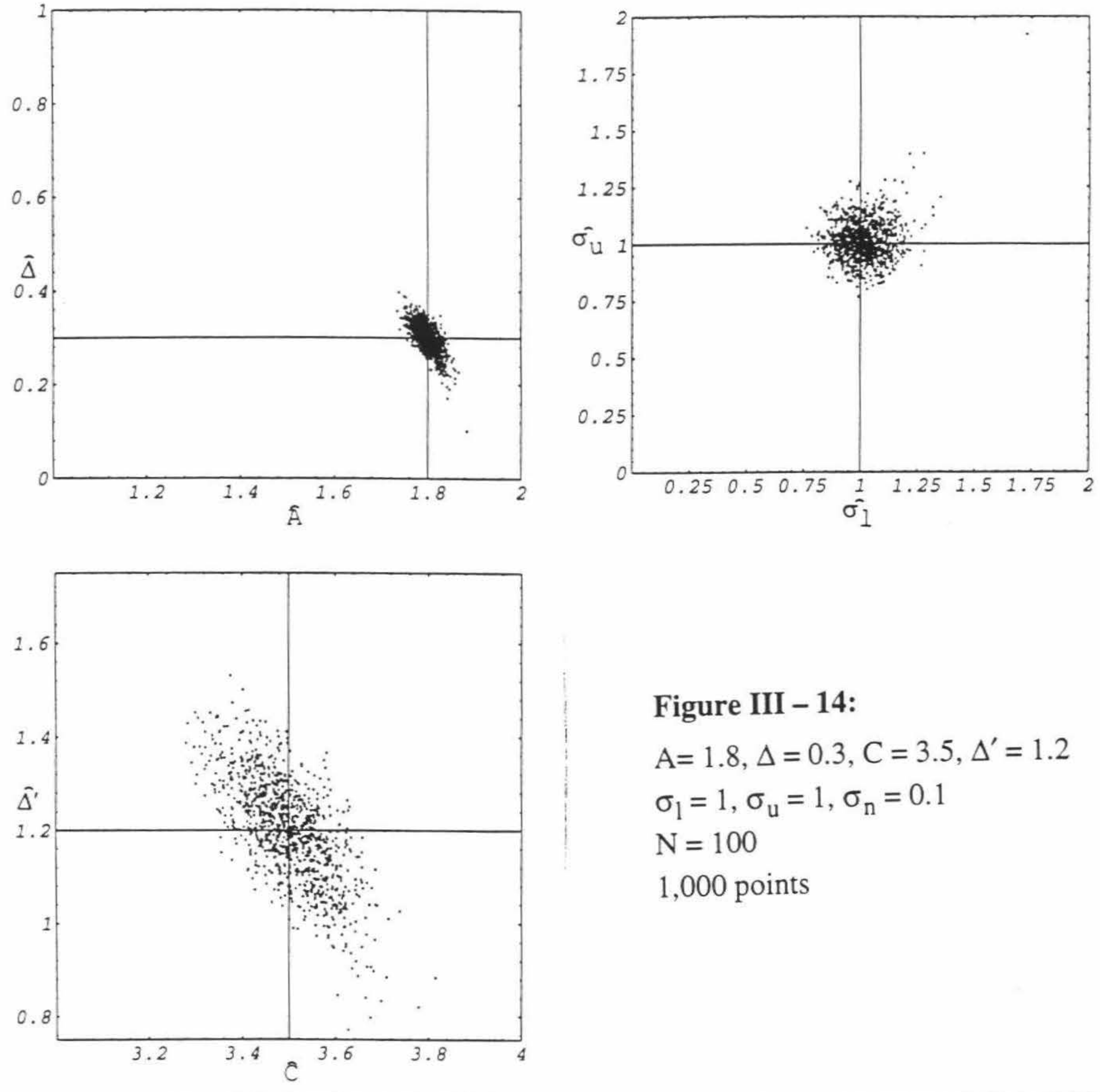


Figure III – 14:

A= 1.8, $\Delta = 0.3$, C = 3.5, $\Delta' = 1.2$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

N = 100

1,000 points

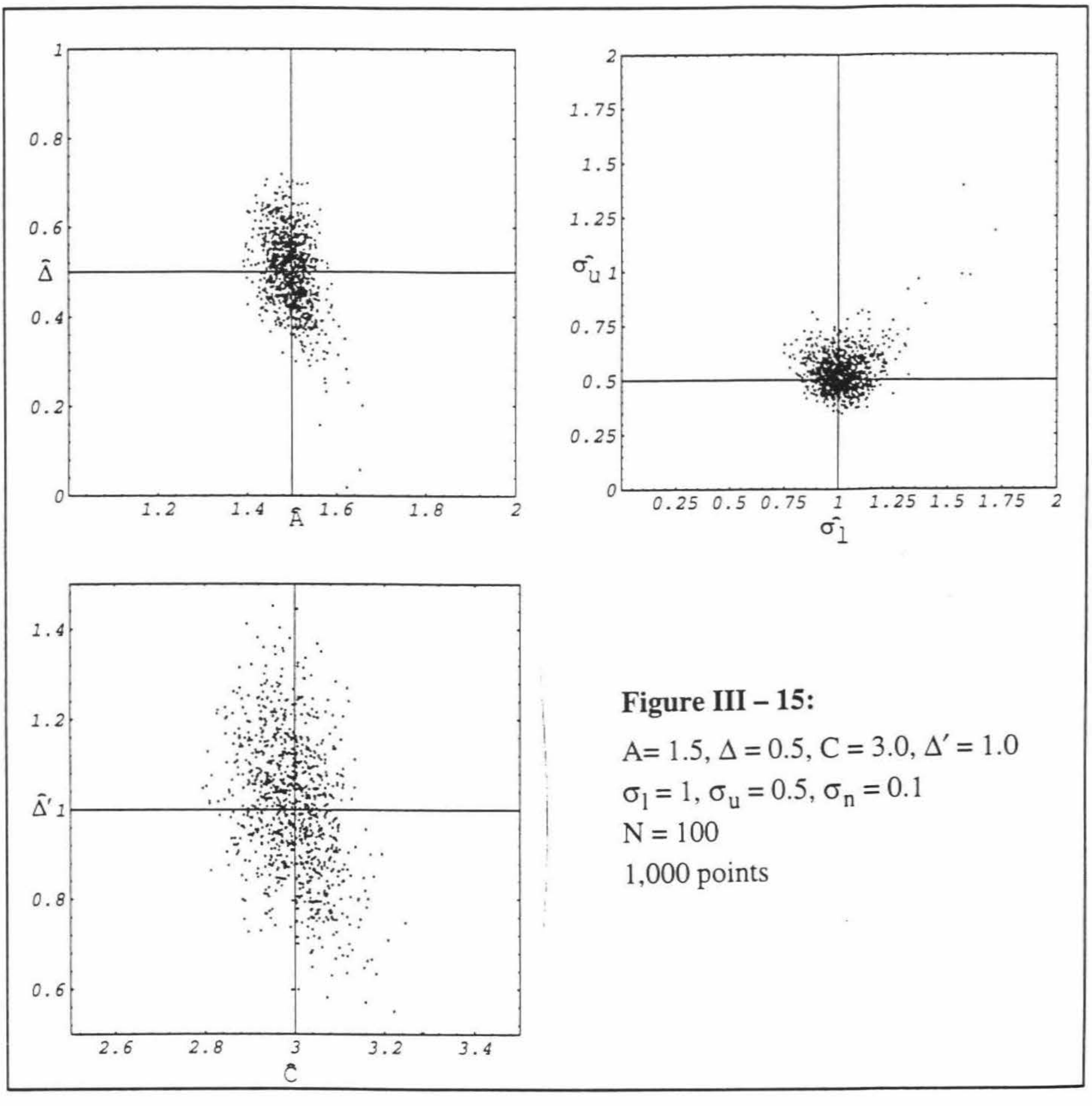


Figure III – 15:

A= 1.5, $\Delta = 0.5$, C = 3.0, $\Delta' = 1.0$

$\sigma_l = 1$, $\sigma_u = 0.5$, $\sigma_n = 0.1$

N = 100

1,000 points

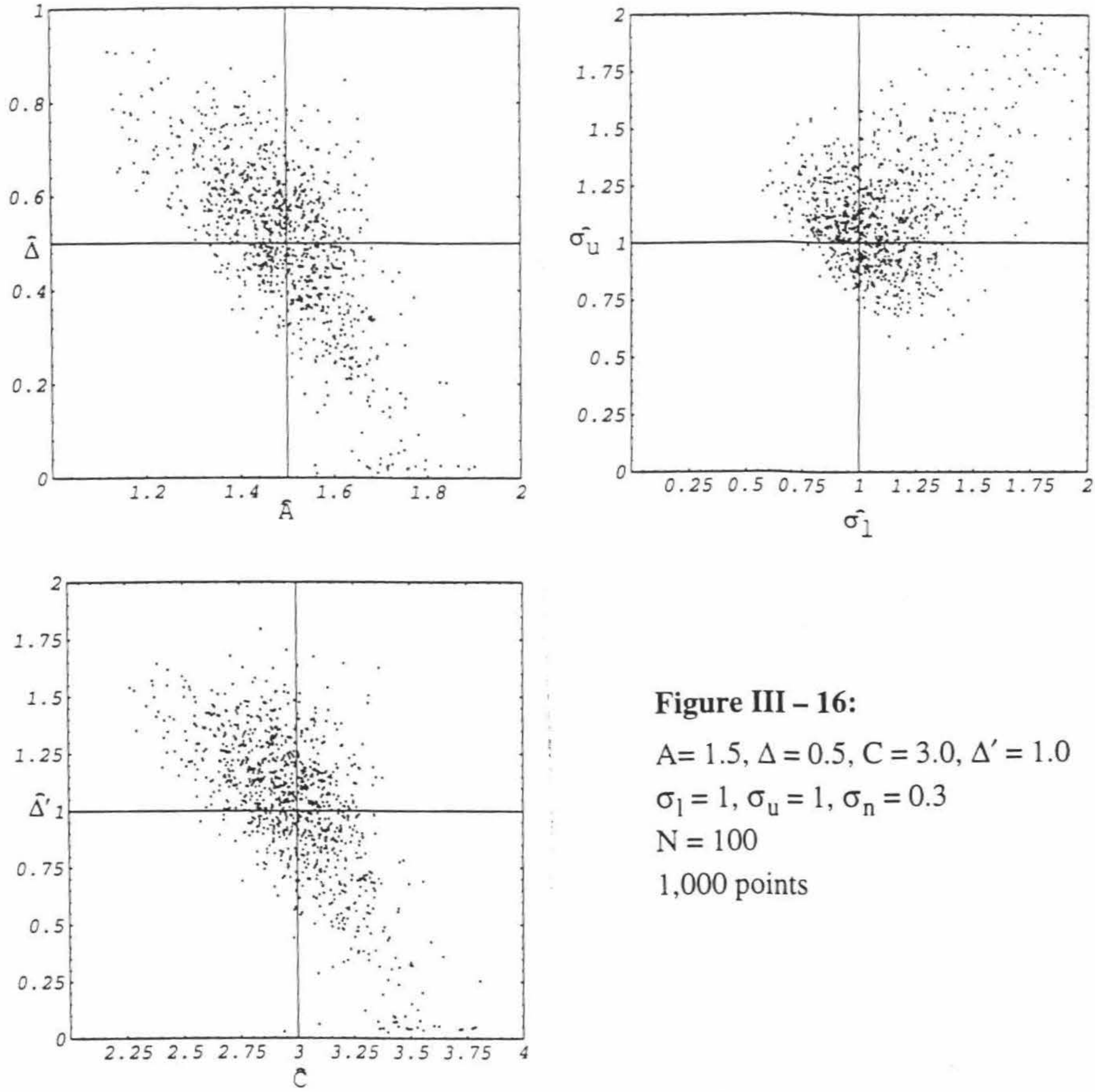


Figure III – 16:

$A = 1.5$, $\Delta = 0.5$, $C = 3.0$, $\Delta' = 1.0$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.3$

$N = 100$

1,000 points

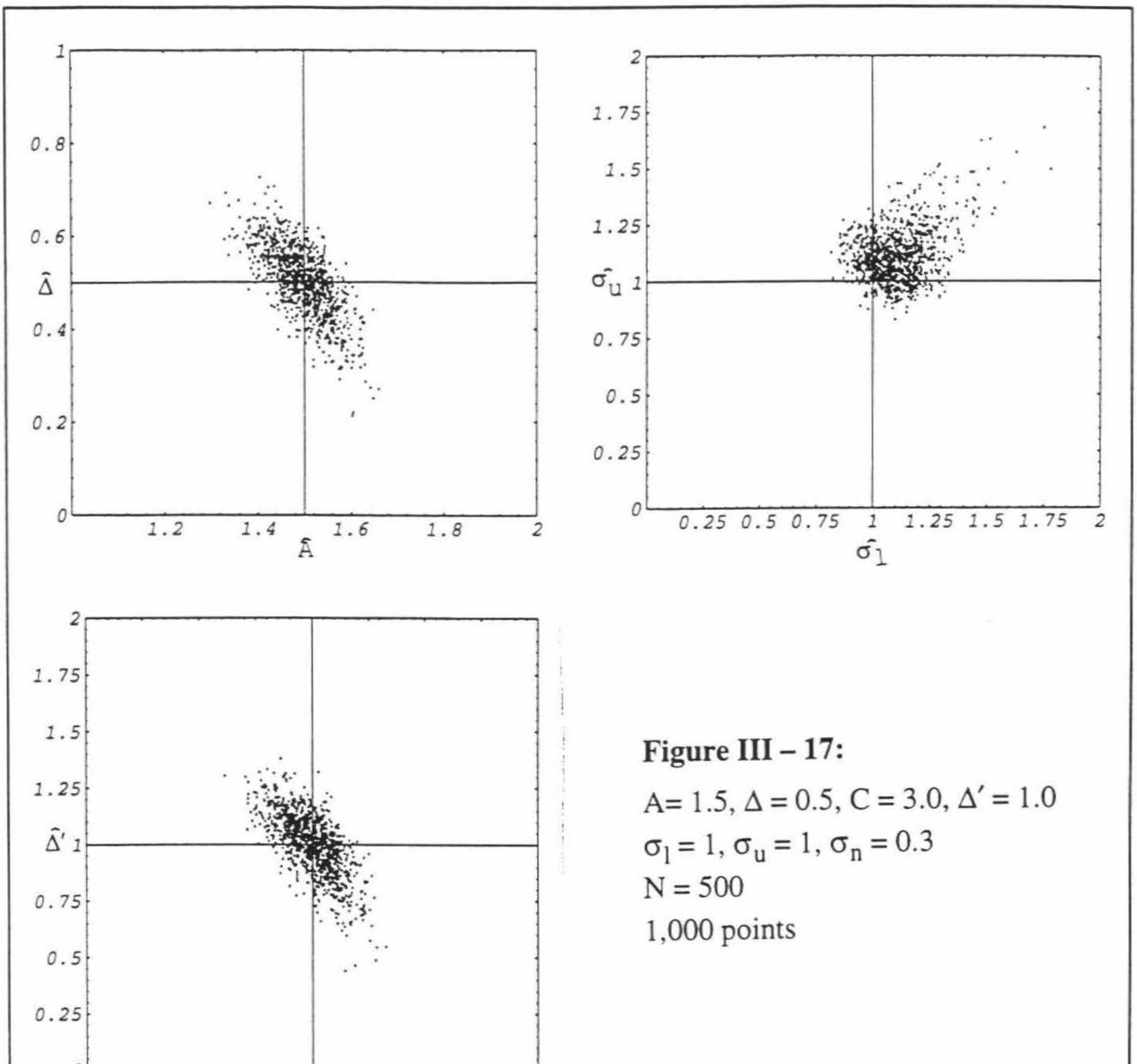


Figure III – 17:

A= 1.5, $\Delta = 0.5$, C = 3.0, $\Delta' = 1.0$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.3$

N = 500

1,000 points

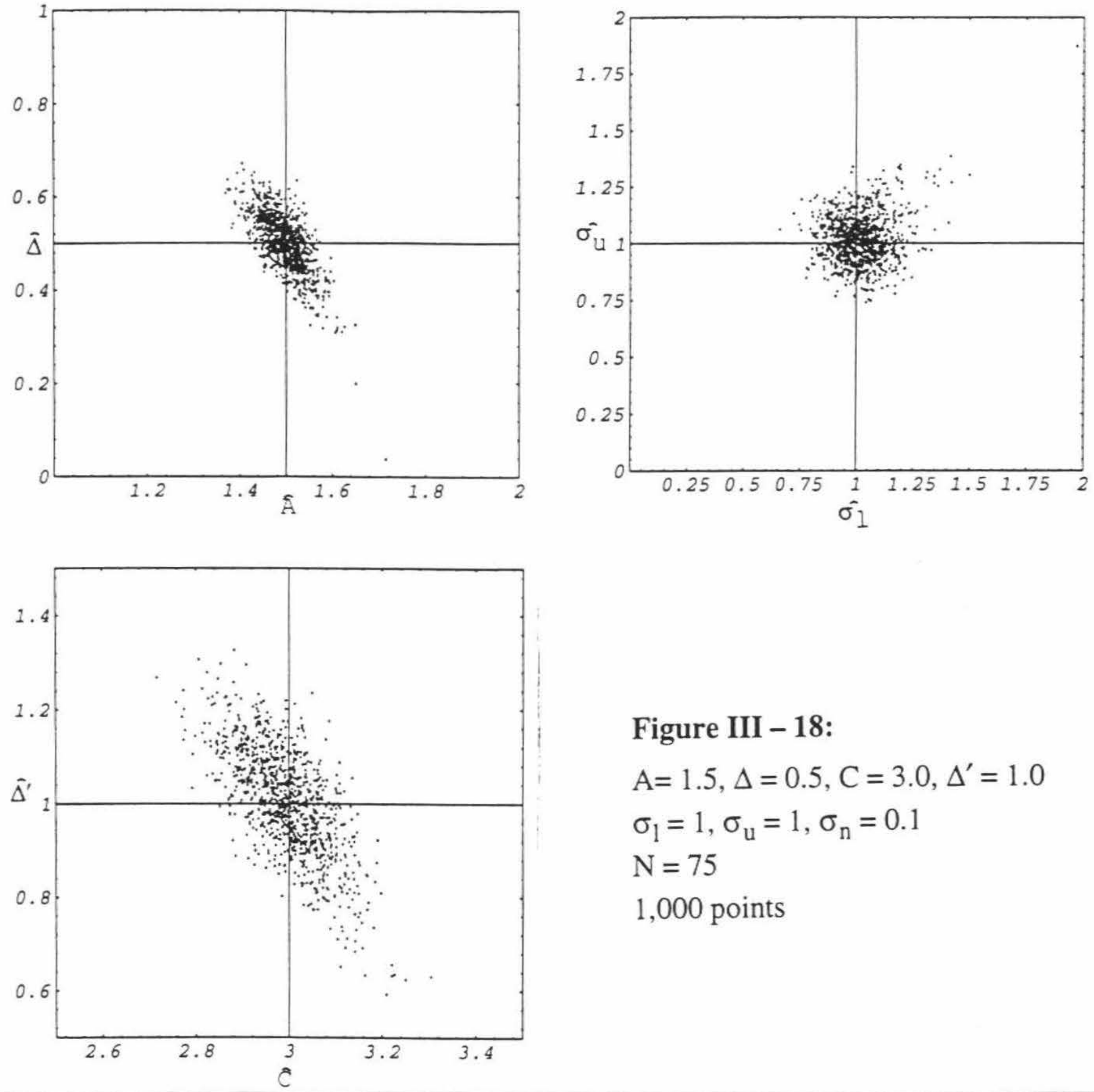


Figure III – 18:

A= 1.5, $\Delta = 0.5$, C = 3.0, $\Delta' = 1.0$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

N = 75

1,000 points

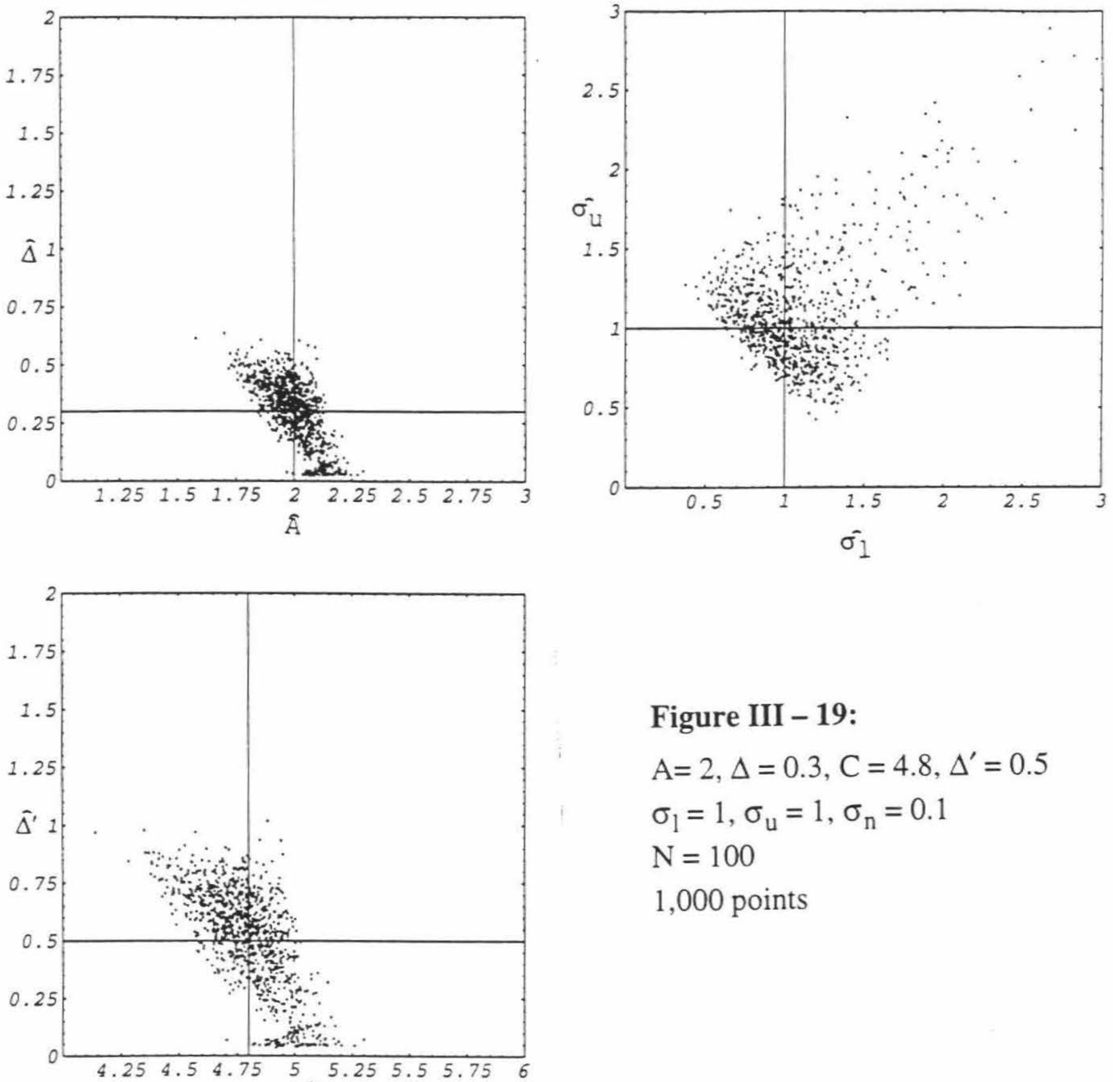


Figure III – 19:

$A = 2$, $\Delta = 0.3$, $C = 4.8$, $\Delta' = 0.5$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

$N = 100$

1,000 points

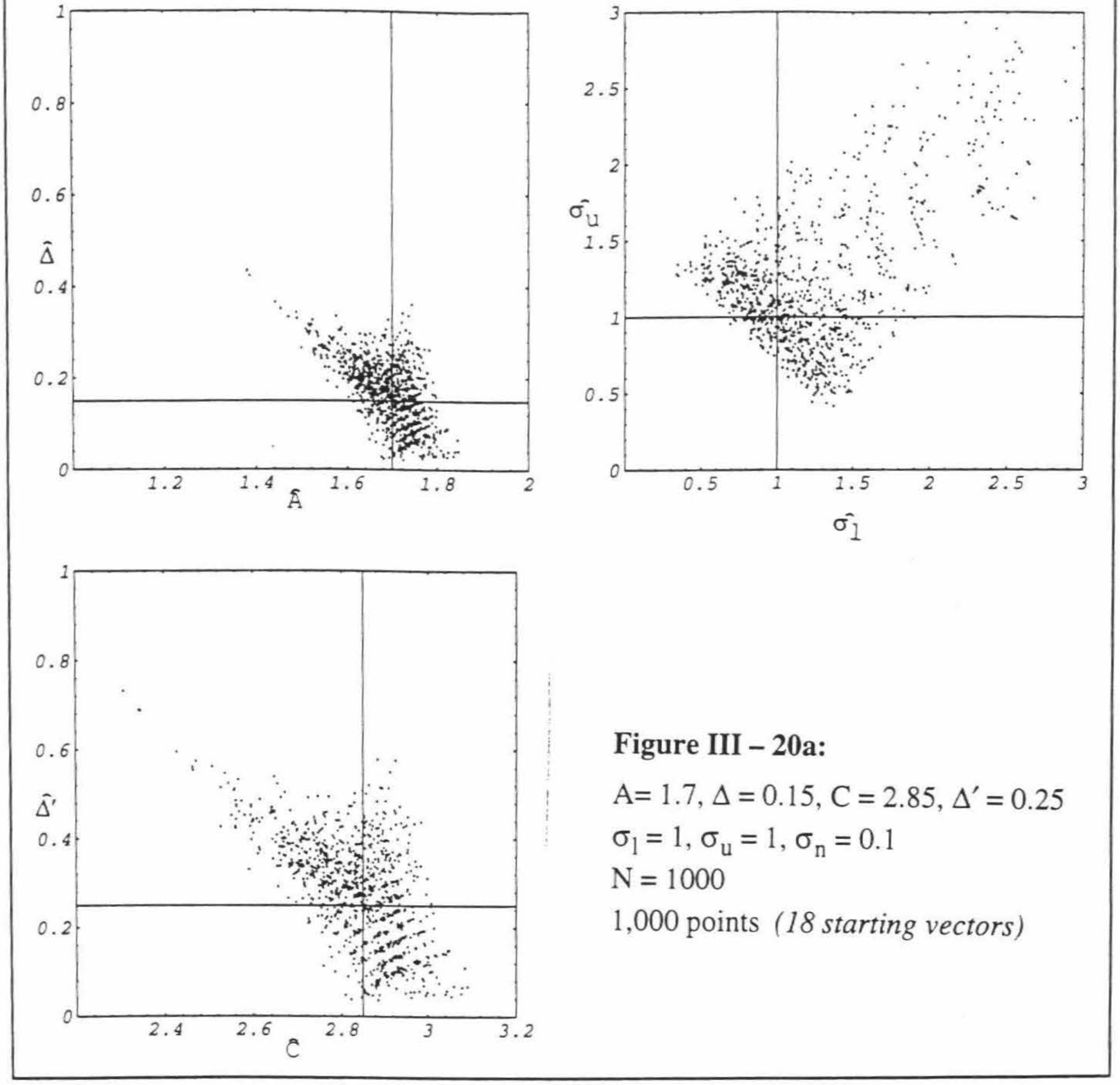


Figure III – 20a:

A= 1.7, $\Delta = 0.15$, C = 2.85, $\Delta' = 0.25$

$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$

N = 1000

1,000 points *(18 starting vectors)*

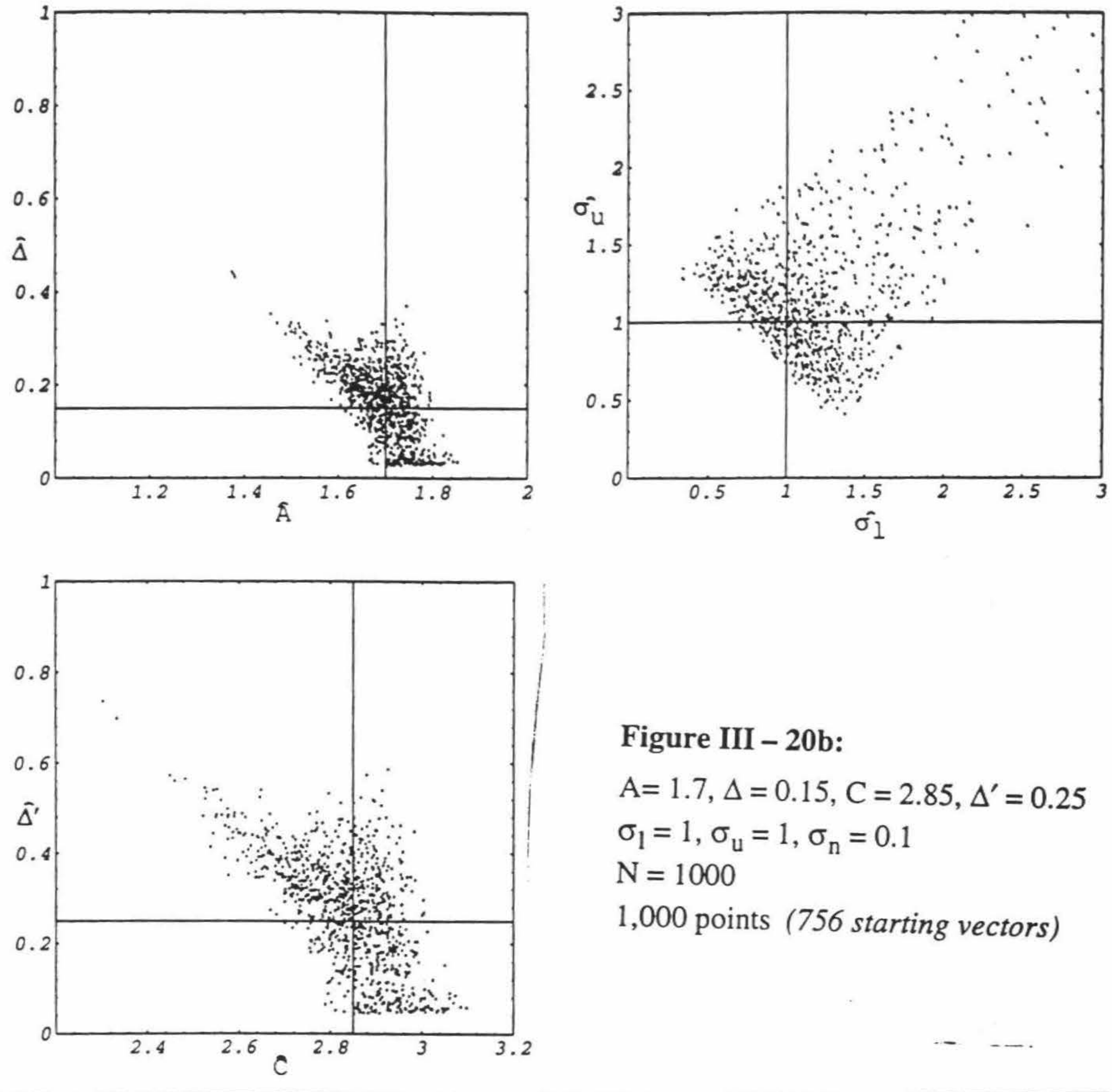


Figure III – 20b:

$A = 1.7$, $\Delta = 0.15$, $C = 2.85$, $\Delta' = 0.25$
$\sigma_l = 1$, $\sigma_u = 1$, $\sigma_n = 0.1$
$N = 1000$
1,000 points *(756 starting vectors)*

III.10 An Experimental Test of the Solution

Having derived the solution theoretically and confirmed its validity by "numerical experiment", we are ready to test the solution by actual physical experiment. For our experimental test, we are limited to using synthetic aperture radar interferometric data that has been previously acquired for other purposes. Unfortunately, the best available such data showing subsurface imaging, as described in the next section, suffers from too much temporal decorrelation between observations to allow the method to detect a layer even 40 meters thick.

This temporal decorrelation, between observations acquired on successive days, is due to some combination of physical changes in the scene, changes in the spacecraft attitude and errors in the processing by NASA of the raw radar echoes into the synthetic aperture radar images. Assuming the elimination of the last factor, successful application of this method in the future may or may not require simultaneous observations to avoid the effects of physical changes in the scene. In section III.17 I propose a test to determine whether or not simultaneous observations are required, and then detail the radar system requirements for successful application of the method for both possible outcomes of the test.

III.11 The Best Previously Acquired Data Available for an Experimental Test of the Solution

The best available data is 3 C-Band (5.7 cm wavelength) images acquired from parallel orbital paths by the Shuttle Imaging Radar (SIR-C) over the eastern edge of the Selima Sand Sheet on the Egypt/Sudan border (Figure III-21). These data were acquired on 3 consecutive days, within 50 minutes of the same local time each day, beginning October 8, 1994. I consider this data to be the best available because these images, in addition to being acquired from parallel paths the requisite minimum of three times, are of the one area in the world where subsurface imaging is known to occur over a wide area, as first observed by SIR-A in November 1981 [*McCauley et al.*, 1982; *Blom et al.*, 1984].

The braided stream channels below the oasis Bir Safsaf (lat 22°36'N, long 29°30'E) were originally imaged by SIR-A at L-band (24 cm wavelength), and confirmed by subsequent field expeditions to be buried up to 2 meters deep in sand [*McCauley et al.*, 1982]. These channels, referred to in Fig 2 of *Elachi et al.* [1984] and visible in Fig 7a of *McCauley et al.*, 1982, are visible in the C-band images of this region acquired by SIR-C (Figure III-22 and *Schaber et al.* [1997]). This is the other reason why I consider the SIR-C data to be the best available, because subsurface imaging is achieved through possibly meters of sand at the shorter C-Band wavelength. This is significant because from the numerical simulations presented in the previous section we know that the greater Δ and Δ' are the better the method works at separating the echoes and determining the depth of burial. We know from equation (III.28) that Δ is proportional to the depth of burial L and inversely proportional to the wavelength. Hence the combination of L possibly being of the order of meters even while the wavelength is relatively short (subsurface imaging is normally attempted and achieved at L-band, which is 4.2 times longer than C-Band) makes this data have the thickest possible layers in terms of phase of any data set available.

III.12 Images, Interferometric Fringes and Correlation Maps Formed from the Data

In Figure III-22 I show the C-band image of the SIR-C data of the region near Bir Safsaf, acquired on October 8, 1994, the first day of the three successive orbits. Figure III-21 is a map which shows the geographical location and orientation of this image, and also shows the path of the L-band SIR-A images of this region acquired in November 1981. This map demonstrates the existence of an area of overlap between the two sets of data, allowing us to test the method in a region where subsurface imaging is known to occur.

The C-Band image is actually in the form of a strip of 4 smaller images, because I have divided the data into these 4 segments. I do this mainly because the 3 images acquired from parallel orbital paths on consecutive days cover the same area, but the pixels in each image are in general offset from image to image by a non integer amount. In order to align the images so that at each set of pixel coordinates the pixel values on consecutive days represent the echo from the same physical location, I determine the offset of one image from another in the four corners of the image, and then interpolate this offset everywhere within the image. A smaller image allows a better approximation of the interpolation to the true offset, so dividing the full image into 4 smaller components allows each component to be better aligned with the components from other days.

In Figure III-23 I show a Landsat TM-7 (Thermatic Mapper, a 7 band visible-to-near-infrared orbital imager) image of the same region shown in Figure III-22, but acquired on 1/13/86. The dark channels visible in the radar image are not visible in the Landsat image. In Figure III-27 I show the non-multilooked fringes between day 1 and day 2, day 2 and day 3 and day 1 and day 3, respectively, for a small 500*250 pixel region centered at x=1250 pixels, y=7275 pixels. Unlike the radar image shown in Figure III-22, which is multilooked 5 pixels in each direction (25 pixels total multilooking), I do not have to multilook the fringes to get a physically meaningful quantity, since the speckle for each pixel is the same for successive observations.

The different spacing of the fringes for the different pairs of observations is apparent from Figure III-27; the component of the baseline perpendicular to the line of sight is larger for the x-z pair than for the y-z or x-y pairs, resulting in a finer fringe spacing. The correlation between the observations, much lower in the dark channel, is indicated by the noise in the fringes, and the scale over which the fringes are relatively straight can be seen to be of the order of at least 50 pixels in any direction. The pixel width on the ground is 4.4 m square, so the absence of any appreciable kinks in the fringes on this scale indicates the topography is flat on a scale of about 200 m.

In Figures III-24, III-25 and III-26 I show the correlation between the day 1 - day 2 observations, day 2 - day 3 observations and day 1 - day 3 observations, respectively. To calculate the correlation I use the definition

$$correlation = \left| \frac{\langle x^* y \rangle}{\sqrt{\langle |x|^2 \rangle \langle |y|^2 \rangle}} \right| = \left| \frac{\sum_{i=1}^{N} x_i^* y_i}{\sqrt{\sum_{i=1}^{N} |x_i|^2 \sum_{i=1}^{N} |y_i|^2}} \right| \qquad \text{(III.52)}$$

The correlation between two observations is the product of the spatial, thermal and temporal correlations [*Zebker and Villasenor*, 1994]. The spatial correlation is given by

$$spatial\ correlation\ = 1 - \#\ fringes/\ pixe \qquad (III.53)$$

and the thermal correlation is

$$thermal\ correlation\ = \frac{1}{1 + \frac{1}{SNR}} \qquad (III.54)$$

where SNR is the signal power to thermal noise power ratio. The fringe frequency in # fringes/pixel is largest for the day 1 - day 3 pair of observations, and is about 30 pixels per fringe, so the fringe rate is 1/30 fringes per pixel. Hence the spatial correlation is at worst 0.967 for the day 1 - day 3 pair of observations. The signal power to thermal noise power is at worst around 10 dB in the bright regions, so SNR=10 and the thermal correlation in the bright regions is at worst 0.90. For 20 dB signal to noise, which represents a very good signal to (thermal) noise ratio, SNR=100 and the thermal correlation is 99%. Hence the product of the spatial and thermal correlation in the bright regions is at worst about 0.87.

One can see from the correlation maps that in the bright regions the correlation between day 1 and day 3 is typically around 0.5, so the temporal correlation in these regions is only around 0.57. Hence the predominant cause of decorrelation between the observations is temporal decorrelation - that is, physical changes in the scene between observations.

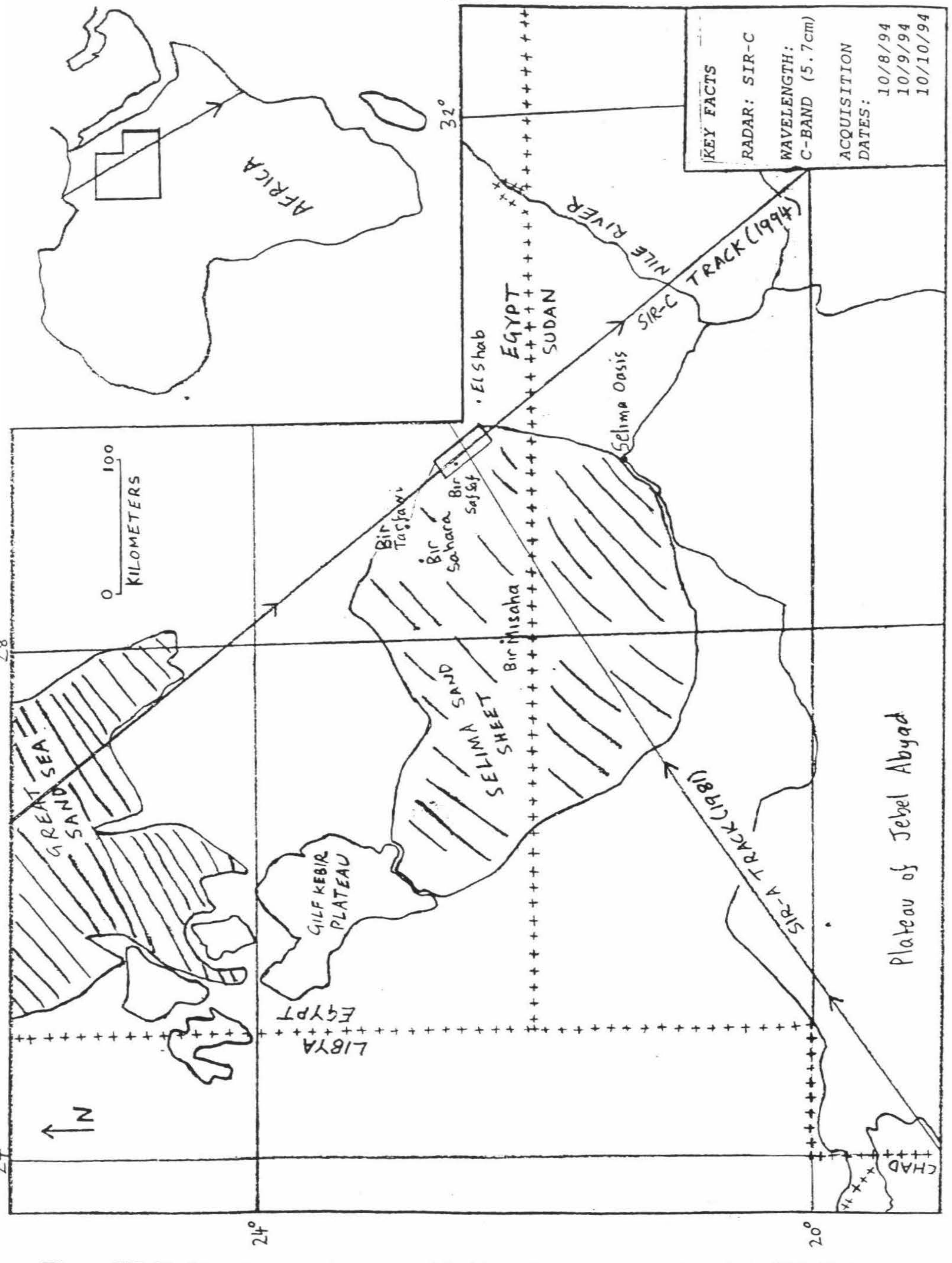


Figure III-21: Map showing the geographical location and orientation of the SIR-C data. This map demonstrates the existence of an area of overlap between the SIR-C data and the original SIR-A observations of the region that first demonstrated subsurface imaging here.

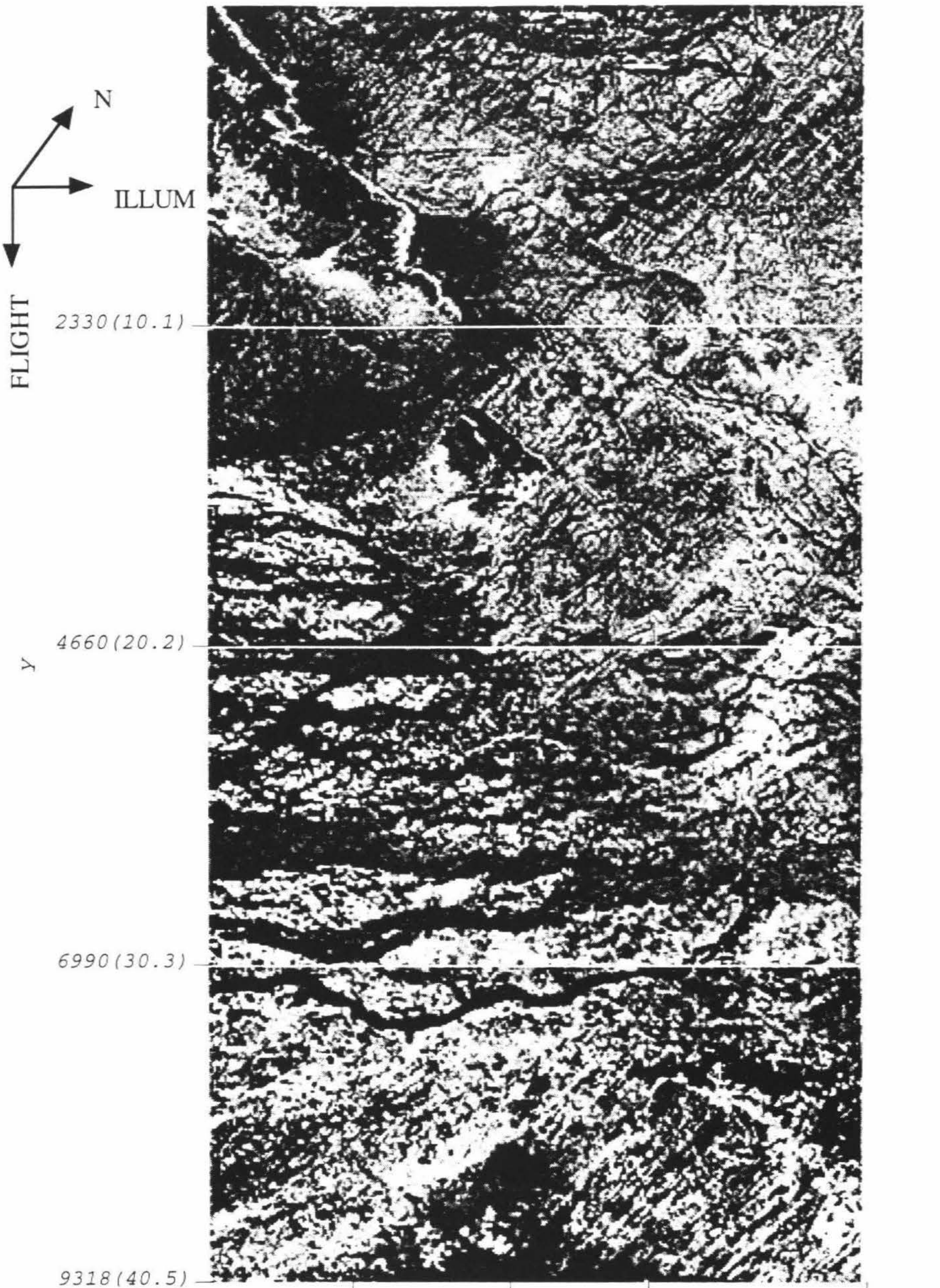


Figure III-22: C-band image acquired by SIR-C; pixels (km). Dark=low radar reflectivity. This image is multilooked 5 pixels in each direction (25 pixels total multilooking) in order to reduce the speckle. The amplitude of the radar echo is displayed.

Figure III-23: False color Landsat TM image of the same area. Band 1 (0.45 μ) = blue, Band 2 (1 μ) = green, Band 3 (2.45 μ) = red. Image was acquired on 1/13/86. The braided stream channels visible in the previous image acquired by SIR-C are not visible in this visible - near-infrared image, which samples only the top few microns of the surface.

Figure III-24: The correlation between the day 1 and day 2 observations, which is determined over 8*8 pixel patches. Comparison with Figure III-22 shows that the lowest correlation regions are the radar dark areas, and vice versa, as expected. However, within the radar bright areas the correlation is much lower than expected from the combined effects of spatial(baseline) and thermal decorrelation (assuming a conservative signal to thermal noise ratio of 10 dB within the radar bright areas). The cause must be temporal decorrelation, which just means that the scene changed physically between observations. The key for the display of correlation as a color is shown below.

Correlation between day 1 and day 2 - 8*8 averaging

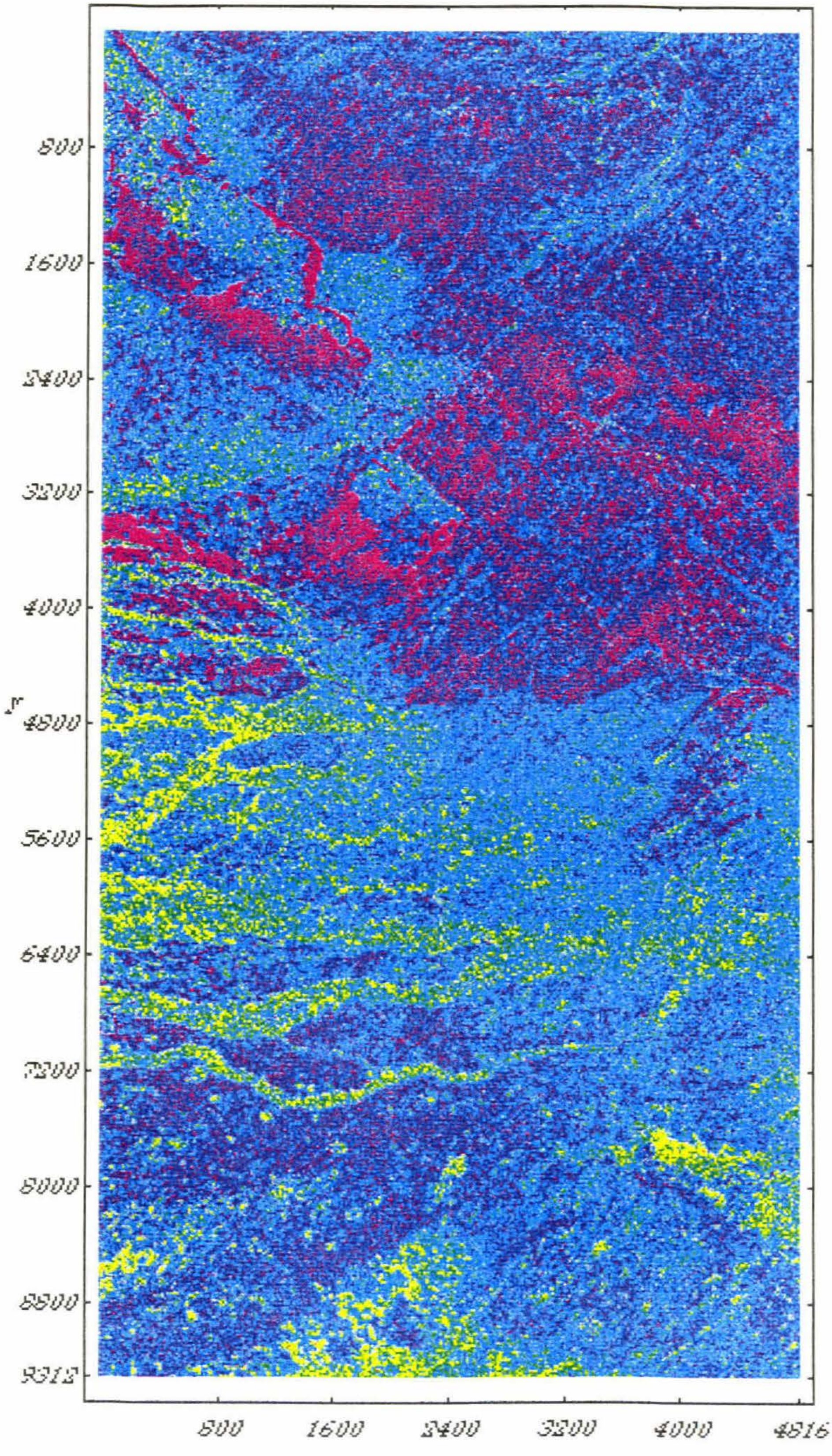


Figure III-25: The correlation between the day 2 and day 3 observations, which is determined over 8*8 pixel patches. Comparison with Figure III-22 shows that the lowest correlation regions are the radar dark areas, and vice versa, as expected. However, within the radar bright areas the correlation is much lower than expected from the combined effects of spatial(baseline) and thermal decorrelation (assuming a conservative signal to thermal noise ratio of 10 dB within the radar bright areas). The cause must be temporal decorrelation, which just means that the scene changed physically between observations. In general the correlation is higher than between day 1 and day 2. The day 2 - day 3 baseline is shorter than the day 1 - day 2 baseline, as can be seen from the greater spacing of the fringes between day 2 and day 3 than between day 1 and day 2 (Figure III-27). This causes less spatial (baseline) decorrelation, but not enough to account for the significantly higher correlation between day 2 and day 3 in most bright regions. Therefore, there was less physical change in the scene between day 2 and day 3 than between day 1 and day 2, perhaps due to less wind. The key for the display of correlation as a color is shown below.

Correlation between day 2 and day 3 - 8*8 averaging

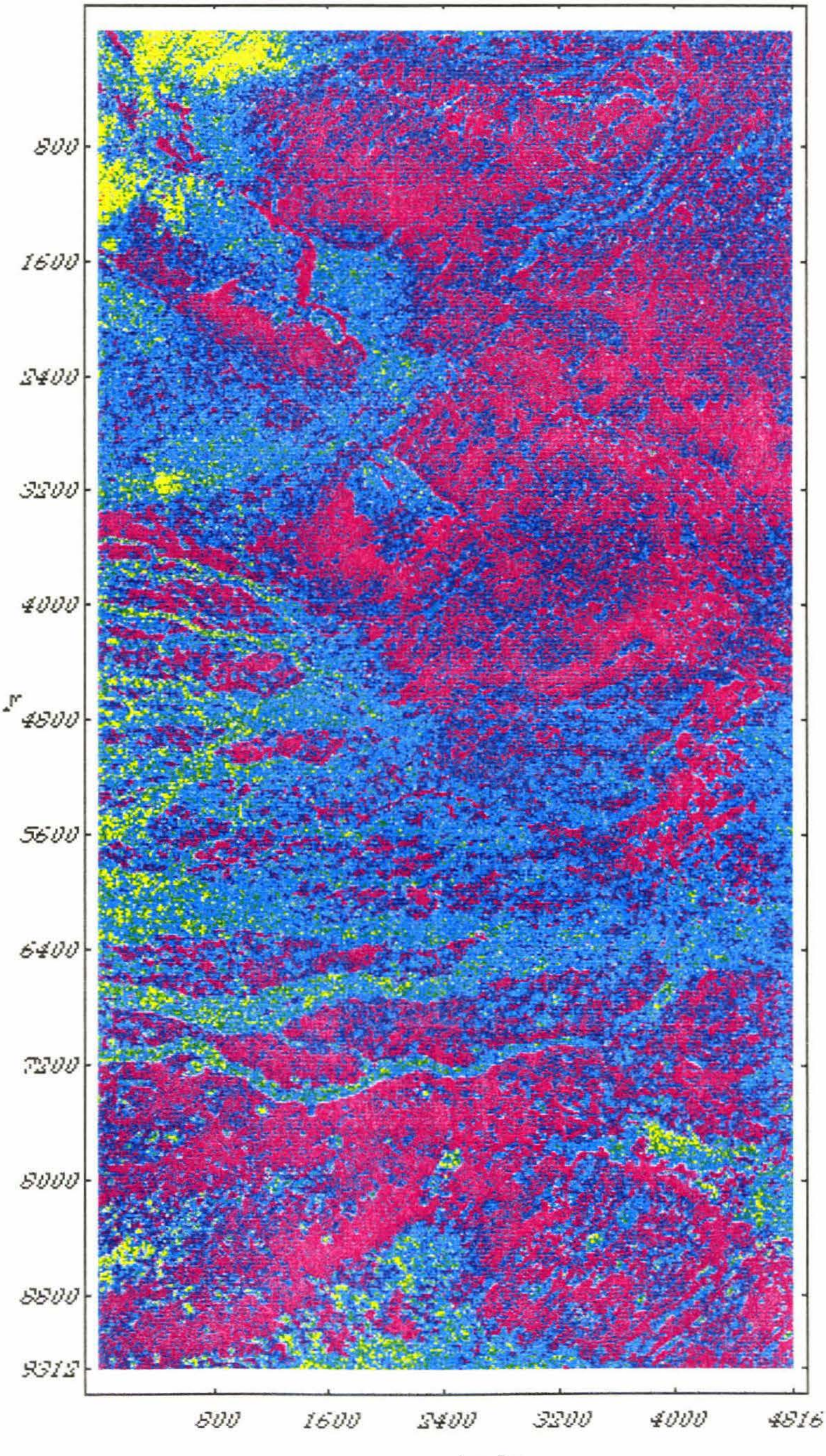
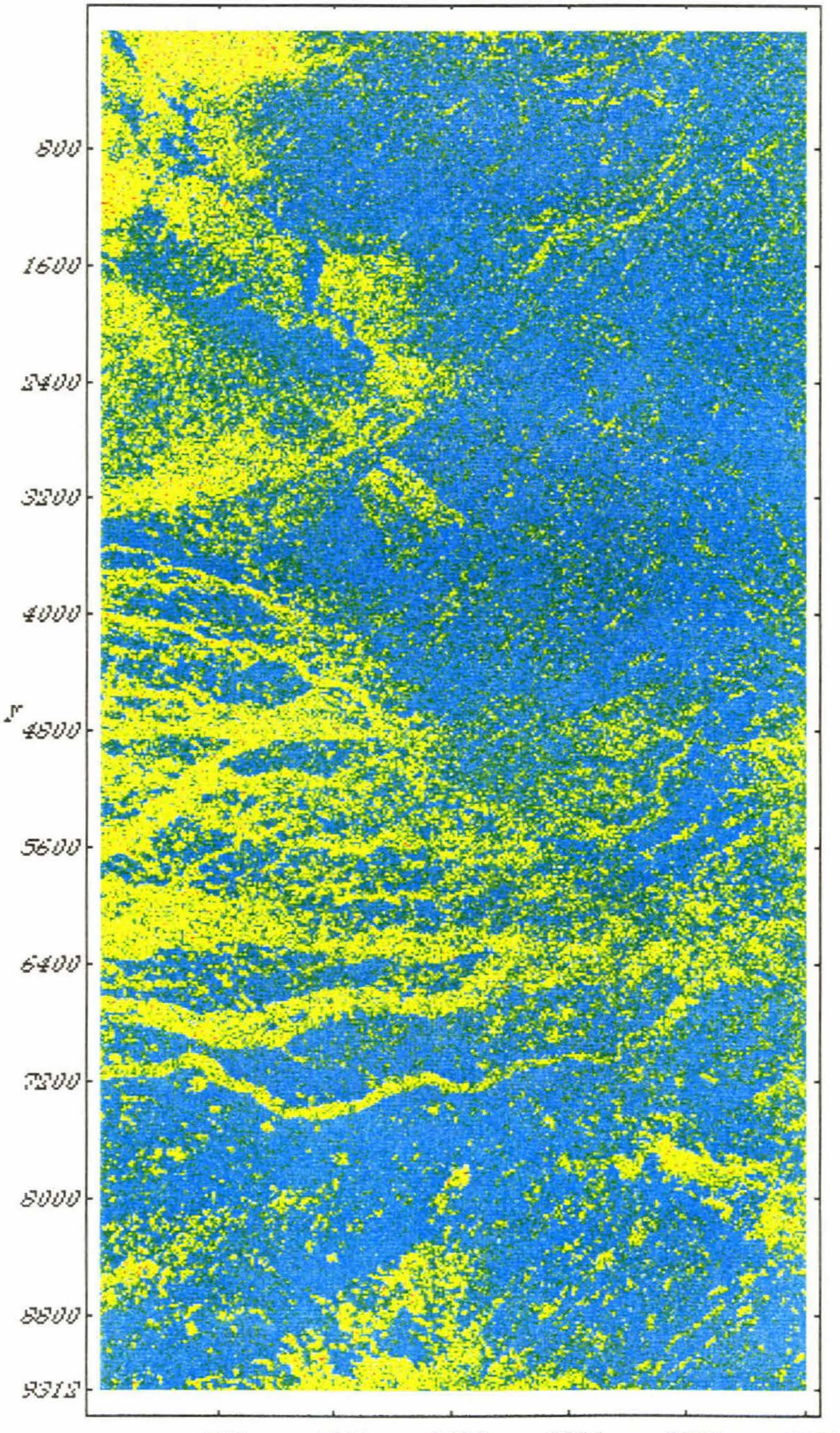
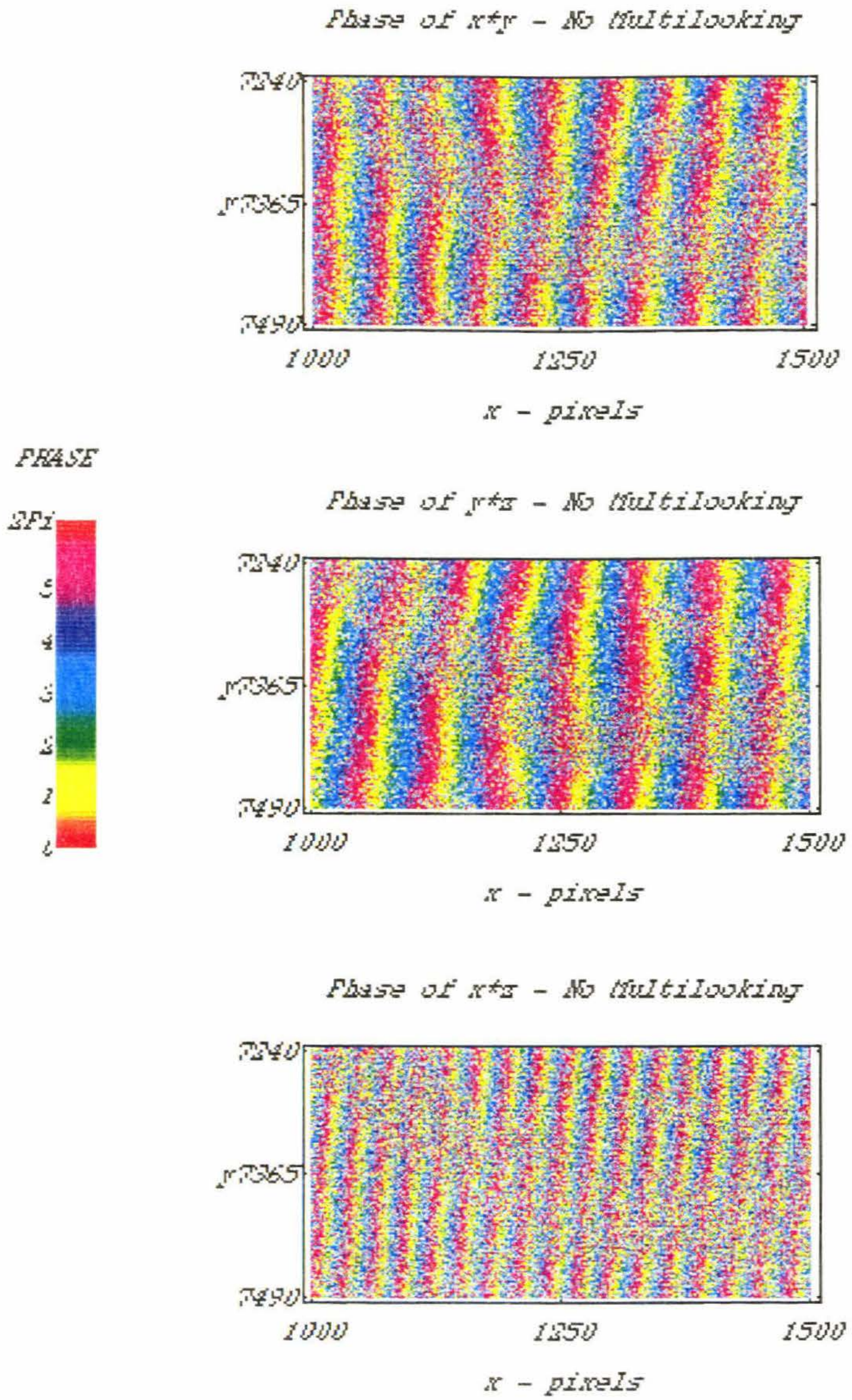


Figure III-26: The correlation between the day 1 and day 3 observations, which is determined over 8*8 pixel patches. Comparison with Figure III-22 shows that the lowest correlation regions are the radar dark areas, and vice versa, as expected. However, within the radar bright areas the correlation is much lower than expected from the combined effects of spatial(baseline) and thermal decorrelation (assuming a conservative signal to thermal noise ratio of 10 dB within the radar bright areas). The cause must be temporal decorrelation, which just means that the scene changed physically between observations. The correlation between day 1 and day 3 represents the combined effects of temporal changes from day 1 to day 2 and day 2 to day 3, and therefore has the lowest correlation of any pair. The day 1 - day 3 baseline is the longest, as can be seen from the finer spacing of the fringes between day 1 and day 3 than between any other pair of observations (Figure III-27). This causes more spatial (baseline) decorrelation, but not enough to account for the low correlation between day 1 and day 3 in the bright regions, which is only around 0.5, compared with a correlation of 0.87 that would be expected from spatial and thermal decorrelation. The key for the display of correlation as a color is shown below.

Correlation between day 1 and day 3 - 8*8 averaging





Phase of x*y - No Multilooking

Phase of y*z - No Multilooking

Phase of x*z - No Multilooking

Figure III-27: An enlargement of segment 4 showing the fringes in a 250*500 pixel region around a bend in the radar dark channel, centered at x=1250 pixels, y=375 pixels. The different spacing of the fringes for different pairs of observations is apparent; the component of the baseline perpendicular to the line of sight is larger for the x-z pair than for the y-z or x-y pairs, resulting in a finer fringe spacing. The correlation between the observations, much lower in the dark channel, is indicated by the noise in the fringes.

III.13 Simulated Results Expected from Highly Correlated, One Surface Data

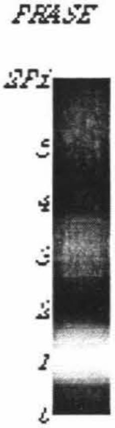
I simulated the radar echoes from a single surface (Δ and $\Delta' = 0$) by using equations (III.35) - (III.37) to generate the data for each pixel within a patch for each of three simulated observations. I assumed that the data remained highly correlated between observations, so that only thermal noise affected the observations, and I assumed that this noise was equal in size, on average, for all three observations of each pixel. This would be true if all three observations are made near the same local time, so that the reflected sunlight in the region of the solar blackbody spectrum near the radar wavelength and emitted thermal radiation from the ground due to the ground temperature are approximately the same. I also assumed that the signal to noise ratio was equal, on average, throughout the patch of pixels, which would be true for a patch covering an area of approximately constant radar reflectivity (cross section).

I generated this simulated data assuming a signal to noise ratio of 10 dB, where $\mathrm{SNR\,(dB)} = 10\log\left({}^{\text{signal power}}/_{\text{noise power}}\right)$, and applied the method of solution for separating upper and lower surface echoes (determining l and u) and determining the depth of burial (via solving for Δ and Δ'). This analogous to what I did in section III.10, except there I generated simulated data that did contain echoes from an upper and lower surface, not a single surface as I do here. Hence I obviously don't expect to be able to separate upper and lower echoes and determine the depth of burial, since these don't exist for a single surface. Instead, I intuitively expect the method to find solutions for, say, the phases

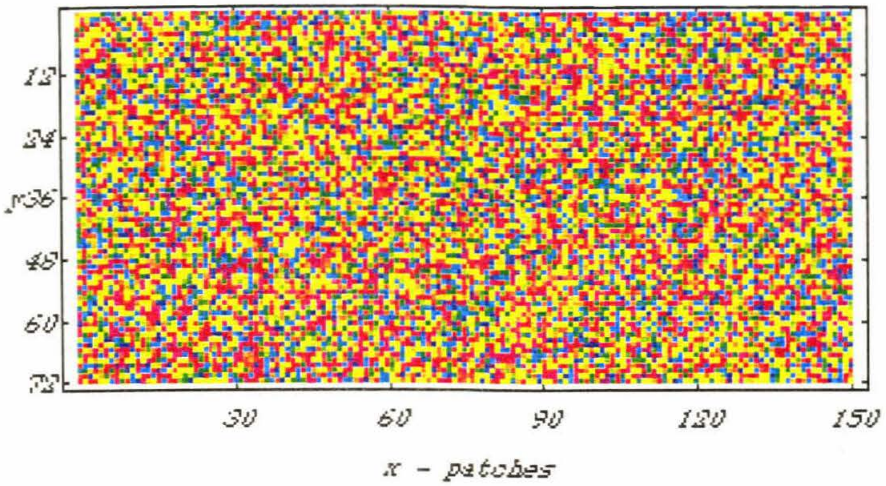
Δ and Δ' that vary randomly as I solve for each new patch of pixels, the thermal noise causing the solutions to jump randomly around.

In Figure III-28 I show the solutions for Δ and Δ', respectively, for a total of 10,800 patches of pixels, each patch composed of 1,024 pixels. These 10,800 patches are displayed in a 150*72 array, and the value of the phase, lying in the range 0 to 2π, is shown as a color, the key for which is given in the Figure caption on the next page. A glance at the Figures confirms our intuition about the expected solutions, for the color (and hence the phase) varies randomly over the array of patches for both Δ and Δ'. This random display of color is then the qualitative solution expected when no layers are present in the data, provided that the correlation between any two days' data is approximately constant (requiring that there be no significant temporal or spatial decorrelation between observations, so that the only source of decorrelation is thermal noise). On the other hand, when layers are present in data suffering from little spatial or temporal decorrelation, the scatter plots of section III.10, which show the solutions for the phases Δ and Δ' scattering around their true values, imply that the solutions for Δ and Δ', when displayed as a color rather than a scatter plot, will show an approximately constant color.

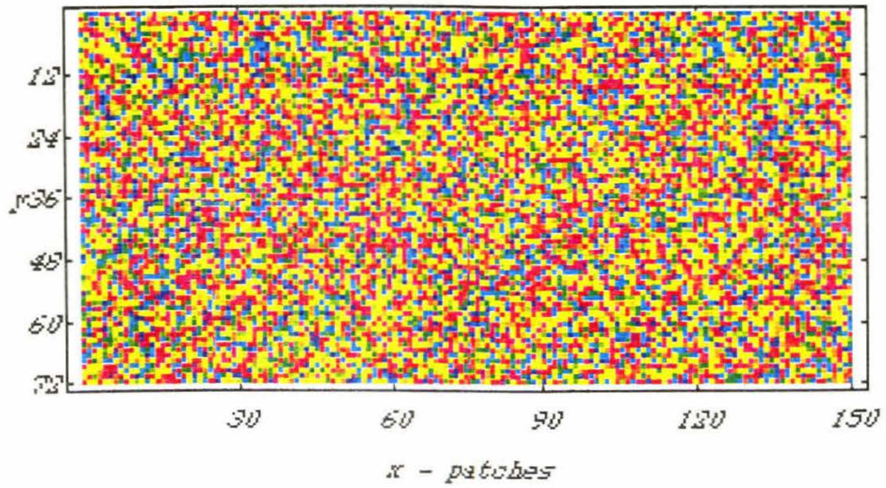
Figure 28: The solutions for Δ and Δ', respectively, for a total of 10,800 patches of pixels, each patch composed of 1,024 pixels. These 10,800 patches are displayed in a 150*72 array, and the value of the phase, lying in the range 0 to 2π, is shown as a color, the key for which is given below. A glance at the Figures confirms our intuition about the expected solutions, for the color (and hence the phase) varies randomly over the array of patches for both Δ and Δ'. This random display of color is then the qualitative solution expected when no layers are present in the data, provided that the correlation between any two days' data is approximately constant (requiring that there be no significant temporal or spatial decorrelation between observations, so that the only source of decorrelation is thermal noise). On the other hand, when layers are present in data suffering from little spatial or temporal decorrelation, the scatter plots of section III.10, which show the solutions for the phases Δ and Δ' scattering around their true values, imply that the solutions for Δ and Δ', when displayed as a color rather than a scatter plot, will show an approximately constant color.



Solution for delta for one surface (simulated)



Solution for delta' for one surface (simulated)



III.14 Results

I applied the method to all 4 segments of the SIR-C data, with the process illustrated by the flowchart in Figure III-29. I averaged over 32*32=1,024 pixel patches, over which I assume the ground is flat. This can be seen to be a good assumption by looking at the scale of the kinks in the interferometric fringes between observations shown in Figure III-27. The kinks indicate topography, and it can be seen that on the scale of a 32*32 pixel box the fringes are generally straight.

Even on a flat plane there will be fringes, however, since as one moves across track the phase difference between two observations will rise (or fall, depending on the orientation of the two observation points) at a constant rate with increasing distance across track (in the x direction). As the phase rises (or falls) through multiples of 2π, fringes result. Because one of our assumptions is that the phases A and C are constant over the patch of pixels being averaged over, I remove the variation in A and C due to geometry prior to processing by the method. Any remaining variation, minimal over the scale of the patches, is then due solely to topography.

Averaging over 1,000 pixels would allow a layer 0.25 radians deep in phase to be detected for about a 20 dB SNR. This detection is achieved through observation of the solution for Δ' scattering on either side of 0.25 radians with a scatter of about 0.2 radians (Figure III-20 and III-21). I can estimate the size of the phase thickness of a layer in the SIR-C data from the spacing of the interferometric fringes between observations, along with the knowledge that the maximum physical depth of a layer is about 2 m [*McCauley et* *al.*, 1982].

By geometry, if one excludes the effect of refraction, the change in the look angle θ between the lower and upper surfaces (separated by a depth L) is L/Dtanθ times the change of the look angle associated with moving one pixel along the ground across track (a distance D). As discussed in section III.4, the effect of ground refraction is to increase the change in the look angle between the lower and upper surfaces by around 30% compared with that for no refraction, for values of the relevant parameters characteristic of the SIR-C imaging of this region.

Including this effect of refraction, and using the SIR-C look angle of $\theta=40^{\circ}$ and pixel width D=4.4 m, the change in the look angle between the lower and upper surfaces of a layer 2 m deep (the maximum depth expected from ground observations of this region) is 0.70 times the change in the look angle associated with moving one pixel along the ground across track. But this latter quantity is not an abstract invention, for a change in the look angle leads to a change in the phase difference between observations of the echo from that look angle made from parallel but displaced locations. Equation (III.40) expresses the change in the phase difference in terms of the change in the look angle, where $B_{\perp_l}$ is the component of the baseline between observation locations perpendicular to the line of sight. One can see that the change in the phase difference is proportional to the change in the look angle, so the change in the phase difference between echoes recorded from parallel but displaced locations, when recording those echoes from the upper surface compared with recording the echoes from the lower surface, is 0.70 times the change in the phase difference between echoes that occurs when moving one pixel over across track.

This latter quantity is just the interferometric spatial fringe frequency, in radians per pixel. For example, one fringe cycled through every 60 pixels would correspond to a change in phase difference of about 0.10 radians per pixel. In fact, for the day 1 - day 2 and day 1 - day 3 baselines the fringe frequencies are 0.11 and 0.20 radians per pixel, respectively. Therefore, the expected maximum phase differences between the phase difference between observations of the lower echo and the phase difference between observations of the upper echo, which I have defined as Δ for the day 1 - day 2 baseline and Δ' for the day 1 - day 3 baseline, are 0.08 and 0.15, respectively.

From the numerical simulations performed in section III.10 we know that halving the largest of Δ and Δ' does not significantly affect the accuracy of the solution for a given number of pixels averaged over. Therefore, averaging over 1,000 pixels should allow the deepest 0.15 radian (2 meters) layers to be detected by the method in radar bright regions, *if the correlation between any two days data were approximately constant.* This would be the case for data with no temporal or spatial decorrelation between observations, just thermal noise - assuming the data were acquired at similar local times. Unfortunately, the correlation maps shown previously demonstrate that in fact the correlation between any two days data does vary greatly depending on which two days are used. As I explain in the next section, this, and not the detection of a buried layer, causes the solutions for Δ and Δ' to be approximately constant.

In Figure III-30 I show the day 1 image, now multilooked to 32*32 pixels, to show one of the “input” images; the day 2 and day 3 input images look the same when multilooked to 32*32 pixels and displayed as an amplitude image, so it is not necessary to show all three input images. For the output images I show the putative lower and upper echoes in Figures III-31 and III-32, respectively, also multilooked to 32*32 pixels. In Figures III-33 and III-34 I show the output phases Δ and Δ', which have physical relevance through being proportional to the physical depth of the putative subsurface layer. The output phases A and C only contain information about the topography of the surfaces, which can be extracted with a priori knowledge about the scene through the use of tie points. This is an established application of radar interferometry, and since this thesis is concerned with new applications, I do not show the output phases A and C.

Figure III-29: A flowchart which illustrates schematically the inputs and outputs of the method developed here for using multiple, interferometric images (the inputs day 1 image, day 2 image and day 3 image, respectively) to separate the lower and upper echoes (outputs lower image and upper image, respectively) and determine the depth of burial (which is proportional to the outputs Δ and Δ'). The output phases A and C only contain information about the topography of the surfaces, which can be extracted with a priori knowledge about the scene through the use of tie points. This is an established application of radar interferometry, and since this thesis is concerned with new applications, I do not show the output phases A and C.

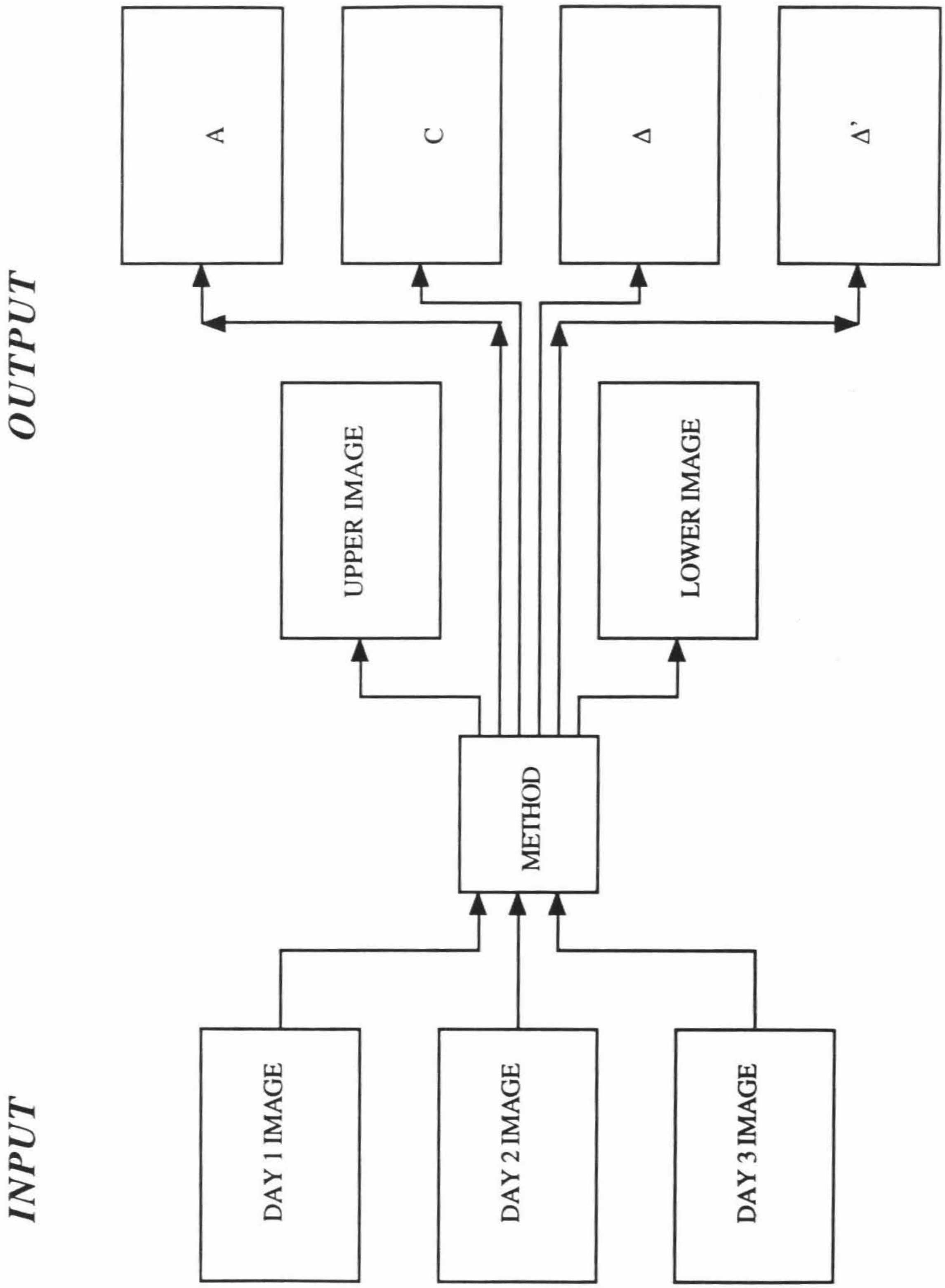

INPUT
OUTPUT
DAY 1 IMAGE
DAY 2 IMAGE
DAY 3 IMAGE
METHOD
UPPER IMAGE
LOWER IMAGE
A
C
Δ
Δ'

Figure 30: The day 1 image of the SIR-C data, multilooked to 32*32 pixels and representing the magnitude of the amplitude. The multilooking reduces the spatial resolution of image to the size of the patches (32*32 pixels in size) that the topography is assumed constant over and that the method averages the data over.The bright regions represent regions with higher radar reflectivity, and vice versa. The braided stream channels are radar dark because the ancient stream beds were eroded smooth by running water, compared with the surrounding terrain. At the boundaries of the 4 segments visible in Figure III-22 one can notice discontinuities in the appearance of this image; this is because, as discussed in the text, each segment is actually processed by the method individually and so I created the output images of each segment individually, and combined the 4 images into one mosaic. When creating an image for display I scale the display of the amplitude of a pixel from one standard deviation below the mean pixel amplitude to one standard deviation above the mean. Each pixel is then assigned one of 256 different levels within this range. Because the mean and standard deviation of each segment is, in general, different, the assigned pixel values for a region of a constant brightness may in general be different for different segments, so at the boundary of the segments there may be a discontinuity in the brightness.

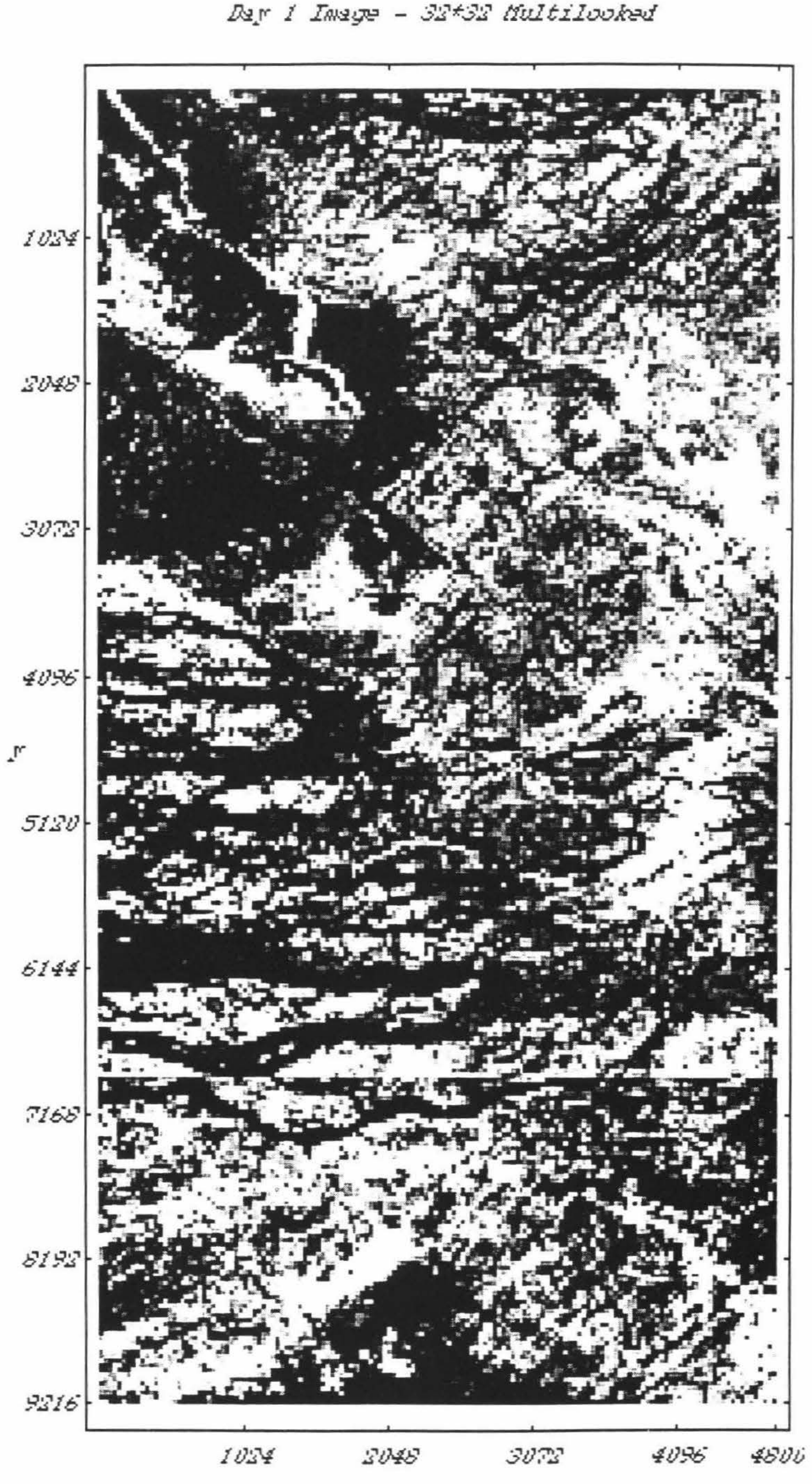

Day 1 Image - 32*32 Multilooked
1024
2048
3072
4096
y
5120
6144
7168
8192
9216
1024
2048
3072
4096
4800
x - pixels

Figure 31: The putative lower echo output by the method, multilooked to 32*32 pixels and representing the magnitude of the amplitude. The multilooking reduces the spatial resolution of image to the size of the patches (32*32 pixels in size) that the topography is assumed constant over and that the method averages the data over.The bright regions represent regions with higher radar reflectivity, and vice versa. At the boundaries of the 4 segments visible in Figure III-22 one can notice discontinuities in the appearance of this image; this is because, as discussed in the text, each segment is actually processed by the method individually and so I created the output images of each segment individually, and combined the 4 images into one mosaic. When creating an image for display I scale the display of the amplitude of a pixel from one standard deviation below the mean pixel amplitude to one standard deviation above the mean. Each pixel is then assigned one of 256 different levels within this range. Because the mean and standard deviation of each segment is, in general, different, the assigned pixel values for a region of a constant brightness may in general be different for different segments, so at the boundary of the segments there may be a discontinuity in the brightness.

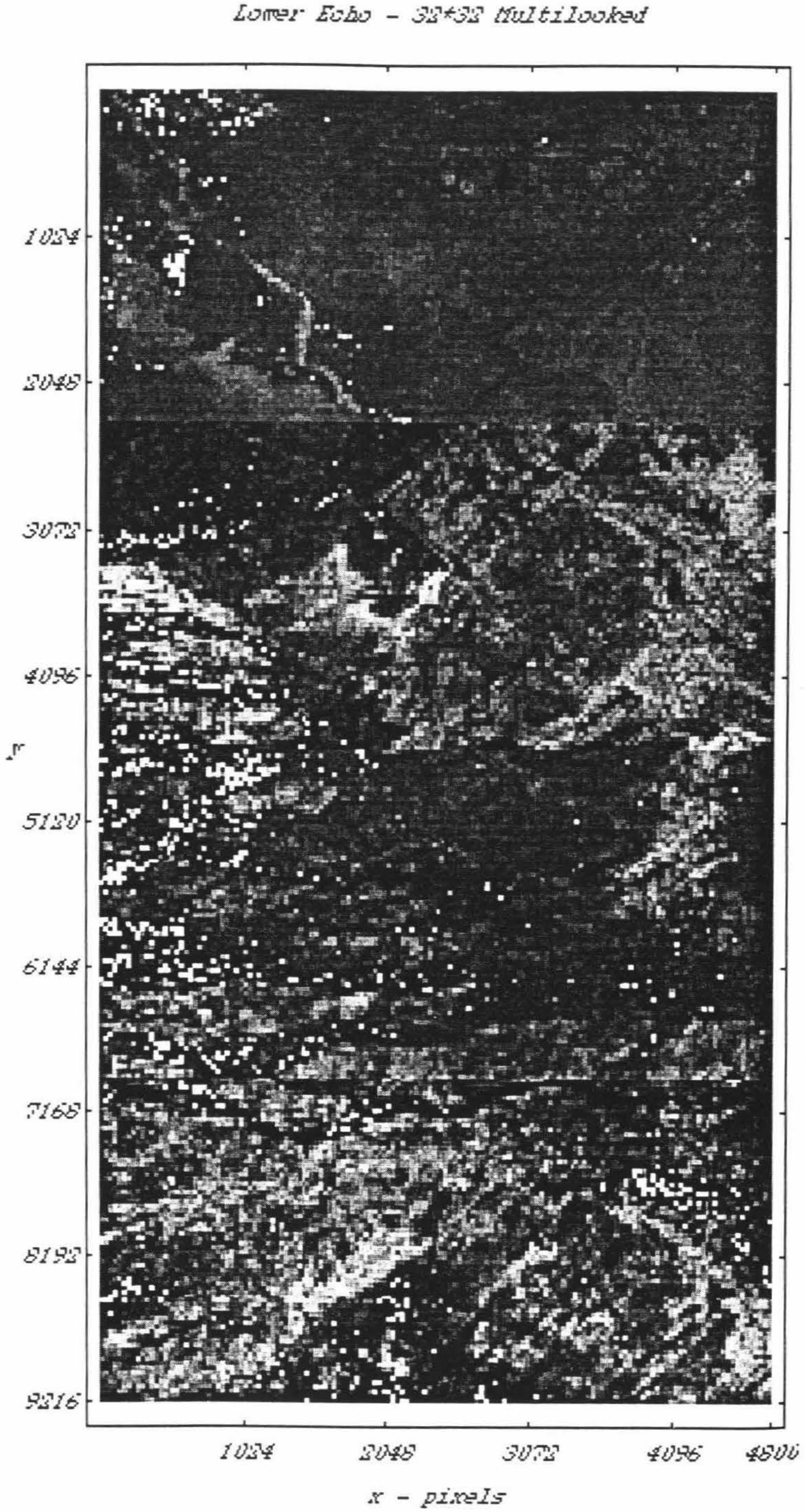

Lower Echo - 32*32 Multilooked
1024
2048
3072
4096
y
5120
6144
7168
8192
9216
1024
2048
3072
4096
4800
x - pixels

Figure 32: The putative upper echo output by the method, multilooked to 32*32 pixels and representing the magnitude of the amplitude. The multilooking reduces the spatial resolution of image to the size of the patches (32*32 pixels in size) that the topography is assumed constant over and that the method averages the data over.The bright regions represent regions with higher radar reflectivity, and vice versa. At the boundaries of the 4 segments visible in Figure III-22 one can notice discontinuities in the appearance of this image; this is because, as discussed in the text, each segment is actually processed by the method individually and so I created the output images of each segment individually, and combined the 4 images into one mosaic. When creating an image for display I scale the display of the amplitude of a pixel from one standard deviation below the mean pixel amplitude to one standard deviation above the mean. Each pixel is then assigned one of 256 different levels within this range. Because the mean and standard deviation of each segment is, in general, different, the assigned pixel values for a region of a constant brightness may in general be different for different segments, so at the boundary of the segments there may be a discontinuity in the brightness.

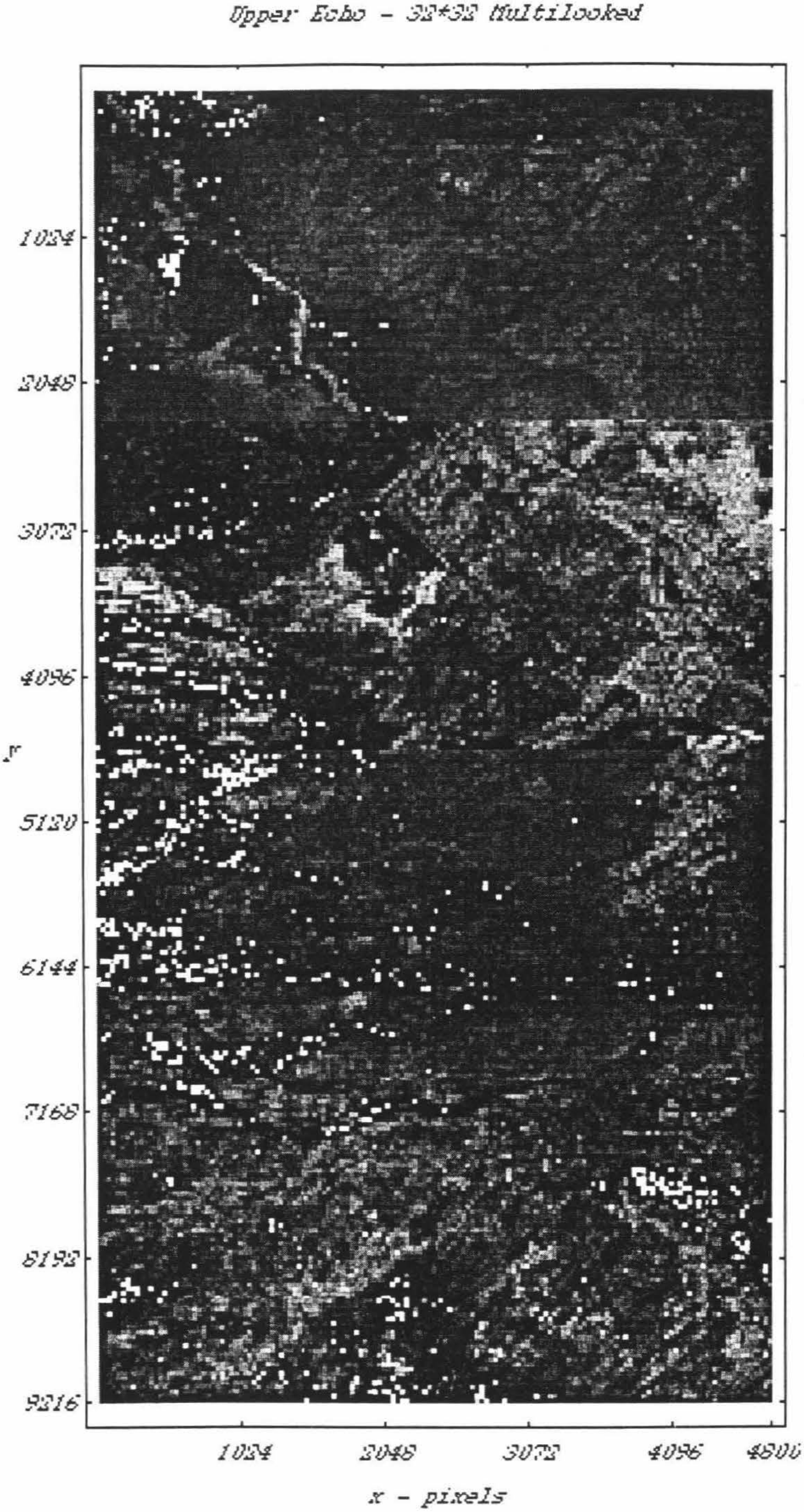

Upper Echo - 32*32 Multilooked
1024
2048
3072
4096
y
5120
6144
7168
8192
9216
1024
2048
3072
4096
4600
x - pixels

Figure 33: Image of Δ output by the method, with the key for the phase represented as a color shown below. The stability in the solution for Δ is at first glance encouraging, for stability is the criterion I intended to use for successful detection of a subsurface layer. However, the solution for Δ hovers around 2 radians. As previously discussed, the spacing of the interferometric fringes suggests a maximum size for Δ of 0.08 radians for a 2 m deep layer. The indication, therefore, of a 40 m deep layer is obviously an artifact. I argue that the stability in this case is not an indication of the detection of a layer, but rather the effect of the varying decorrelation between different days' observations.

Delta - 32*32 Averaging

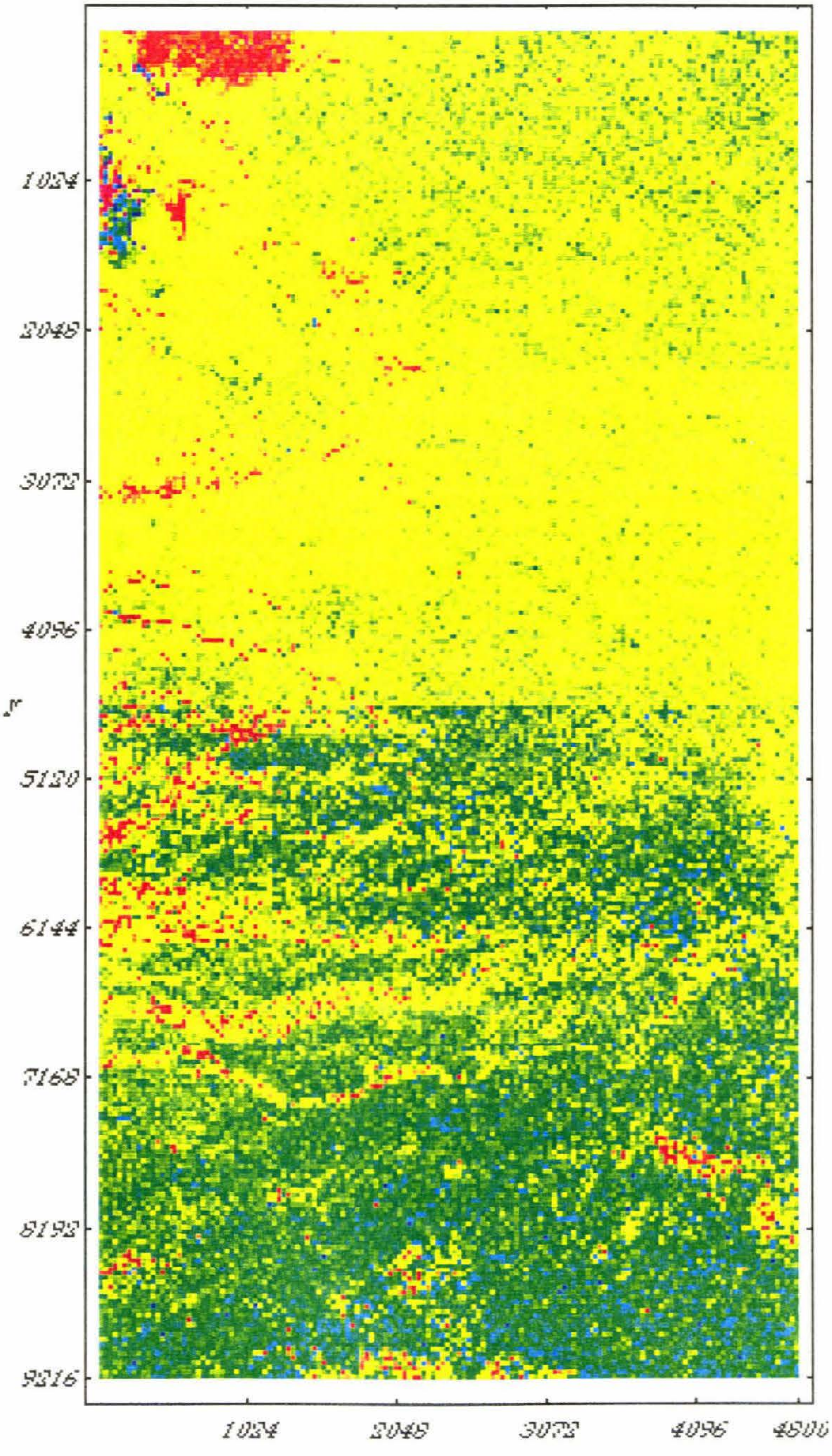
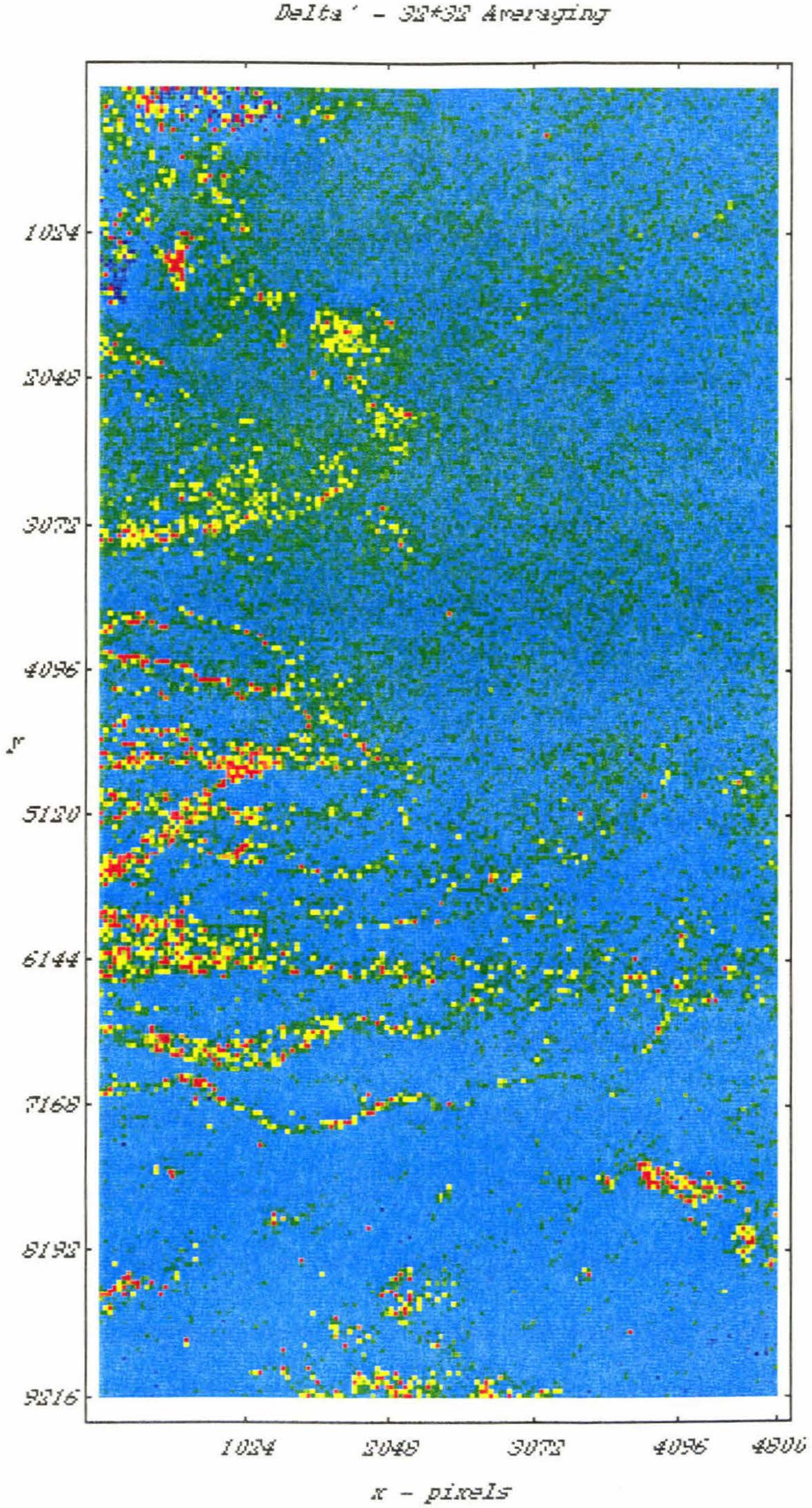


Figure 34: Image of Δ' output by the method, with the key for the phase represented as a color shown below. The stability in the solution for Δ' is at first glance encouraging, for stability is the criterion I intended to use for successful detection of a subsurface layer. However, the solution for Δ' hovers around 3 radians. As previously discussed, the spacing of the interferometric fringes suggests a maximum size for Δ' of 0.15 radians for a 2 m deep layer. The indication, therefore, of a 40 m deep layer is obviously an artifact. I argue that the stability in this case is not an indication of the detection of a layer, but rather the effect of the varying decorrelation between different days' observations.

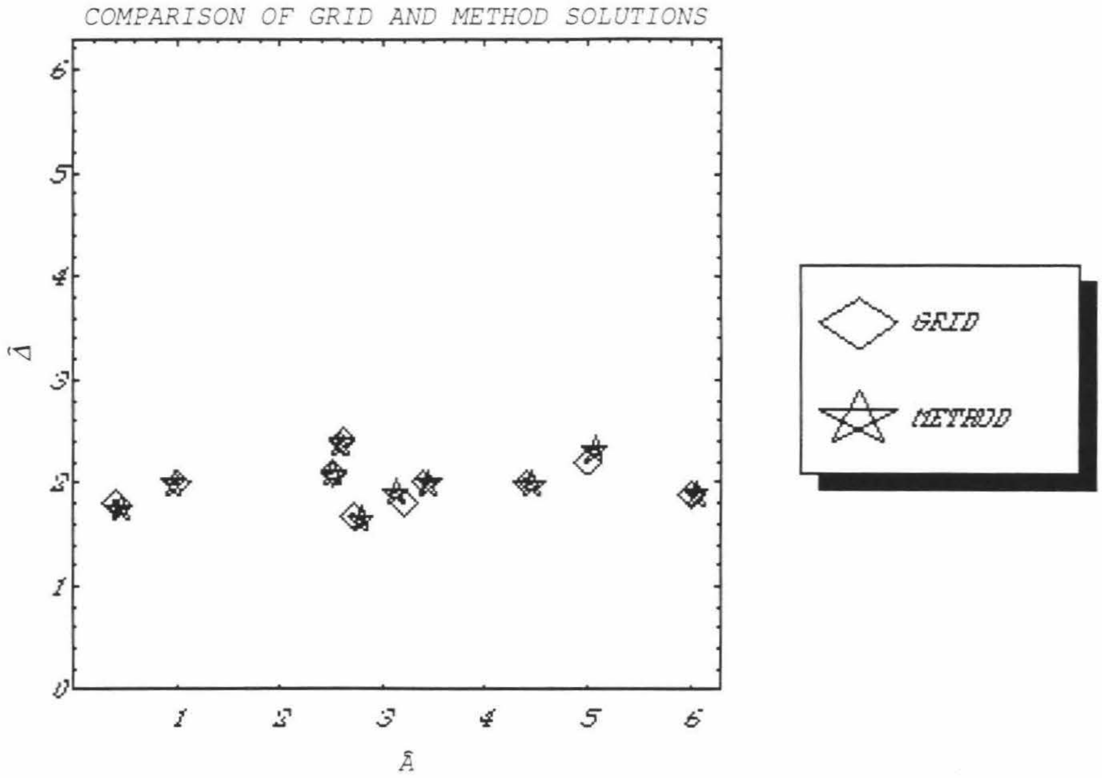

Delta' - 32*32 Averaging
1024
2048
3072
4096
y
5120
6144
7168
8192
9216
1024
2048
3072
4096
4800
x - pixels

III.15 Test of Results

I performed a test of the validity of the results, specifically testing whether or not the method I use to find the minimum of the χ^2 (the Fletcher-Reeves-Polak-Ribiere multidimensional conjugate gradient method) is in fact finding the minimum of the χ^2. I have confidence in the method based on the numerical simulations performed in section III.10, but I have thought of an additional test which confirms the validity of the χ^2 minimum finding method.

For a string of 10 patches from segment 4, arbitrarily chosen to extend from (x=1600,y=7680) to (x=1888,y=7680) (each patch is an average of 32 pixels in each direction), I display the solutions obtained using the method for A, C, Δ and Δ', purportedly to be those values of A, C, Δ and Δ' for which the χ^2 defined by equation (III.48) is a minimum, as star shaped symbols (labeled "METHOD") in Figure III-35. For each of these patches I also determined the values of A, C, Δ and Δ' that minimize the χ^2 by simply dividing the allowed range for the angles (0 to 2π) into a finite number of points (31, corresponding to an interval between points of 0.2) and calculating the value of χ^2 at all 31^4=923,521 grid points in the 4 dimensional space, allowing the grid point at which χ^2 is a minimum to be determined. This is a foolproof, if time consuming, way of determining the values of A, C, Δ and Δ' at which χ^2 is a minimum. I refer to it as the grid method and display the results for the 10 patches as diamond shaped symbols (labeled "GRID") in Figure III-35.

As can be seen from the Figure, the results for this randomly chosen string of patches from the method closely match those from the foolproof grid method, confirming that the method is reliably finding the minimum and that the stability in the solutions for Δ and Δ' is not an artifact of the idiosyncratic method I use to minimize the χ^2. The fact that it would take 5 years (on a workstation class computer circa 1997) to process all 4 segments of data using a grid based approach for minimizing χ^2 illustrates why I was forced into creating such an idiosyncratic (yet reliable) method for minimizing χ^2. Using this method it still took 10 days to process all four segments.



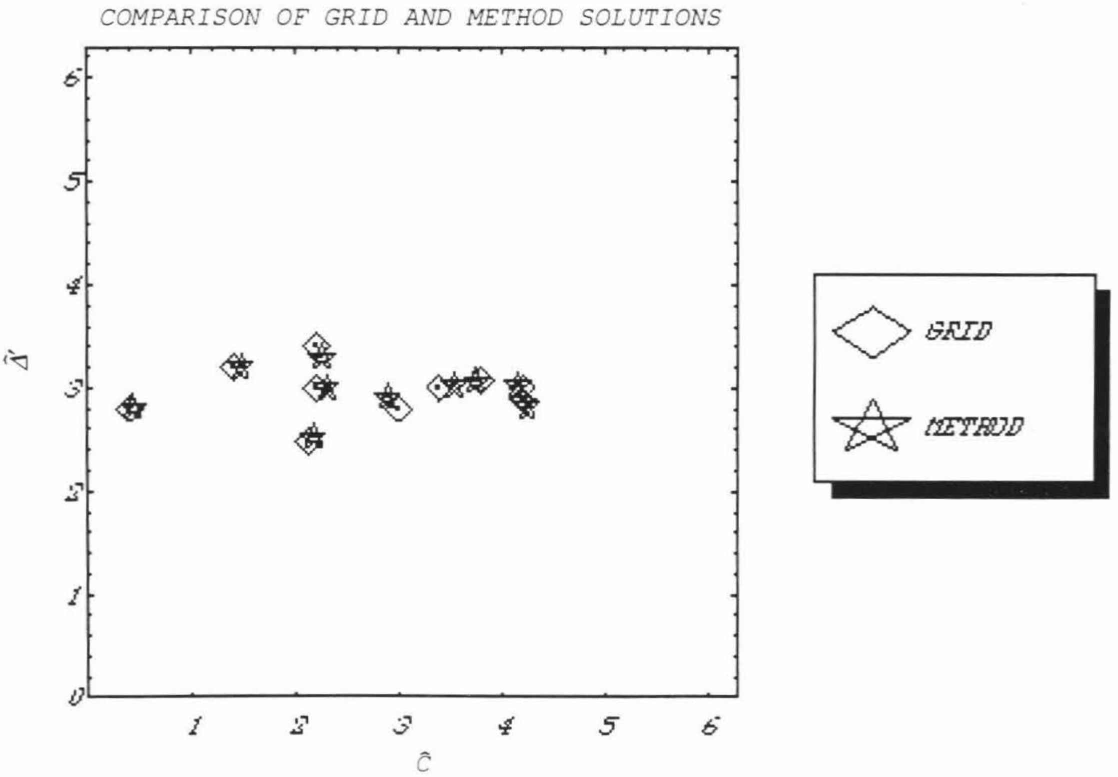


Figure III-35: Comparison of the solutions for A, C, Δ and Δ' for which the χ^2 defined by equation (III.48) is a minimum, using both the minimization method used in this research ("METHOD") and a foolproof but slow grid based search for the minimum ("GRID"). The solutions match closely, confirming the validity of the method.

III.16 Explanation of Results

At first glance the stability in the solutions for Δ and Δ' is encouraging, for stability is the criterion I intended to use for successful detection of layers. However, with the actual data the solutions for the phases Δ and Δ' hover around 2 and 3 radians, respectively. As previously discussed, the spacing of the interferometric fringes suggests maximum sizes for Δ and Δ' of 0.08 and 0.15, respectively, for a 2 m deep layer. The indication, therefore, of a 40 m deep layer is obviously an artifact. I argue that the stability in this case is not an indication of the detection of a layer, but rather the effect of the varying decorrelation between different days' observations. A surprising, important, and disappointing result of the analysis is that while the method can work with the thermal noise and baselines of SIR-C, it requires much higher temporal correlation than had been anticipated.

When presented with data with the property that different pairs of data have significantly different correlations (for example, the day 1 - day 2 pair may have a significantly higher correlation than the day 1 - day 3 pair), the solutions for Δ and Δ' will be approximately constant over any region that the amount of decorrelation between day 1 and day 2 and day 1 and day 3, respectively, is approximately constant. Furthermore, for these regions the size of Δ and Δ' will reflect the amount of decorrelation from day 1 to day 2 and day 1 to day 3, respectively, increasing with decreasing correlation.

This behavior follows from the fact that the method of solution is predicated upon the assumption that the noises associated with each observation are all drawn from the same standard deviation zero mean Gaussian distribution - that is, I assume that all the noises are equal in size, on average. If the data suffered no decorrelation and the observations were made at approximately the same time of day so that the magnitude of the thermal noise was approximately constant, the assumption of equal noises should be valid. Similarly, if the data were decorrelated, but any pair of observations had approximately the same correlation, our assumption would also be valid, and in both cases the method would return approximately constant solutions for Δ and Δ' only when two surfaces really exist. It would return values of Δ and Δ' randomly distributed from 0 to 2π when there is only one surface.

However, the actual SIR-C data from Egypt/Sudan has the property that the correlation between any two days' data depends on which two days are considered; for example, the correlation between day 2 and day 3 is much higher than the correlation between day 1 and day 3. With just a single surface contributing to the echo, it is not possible to model such data with noises of equal magnitude (on average), for in this case the correlation of any pair will be approximately the same.

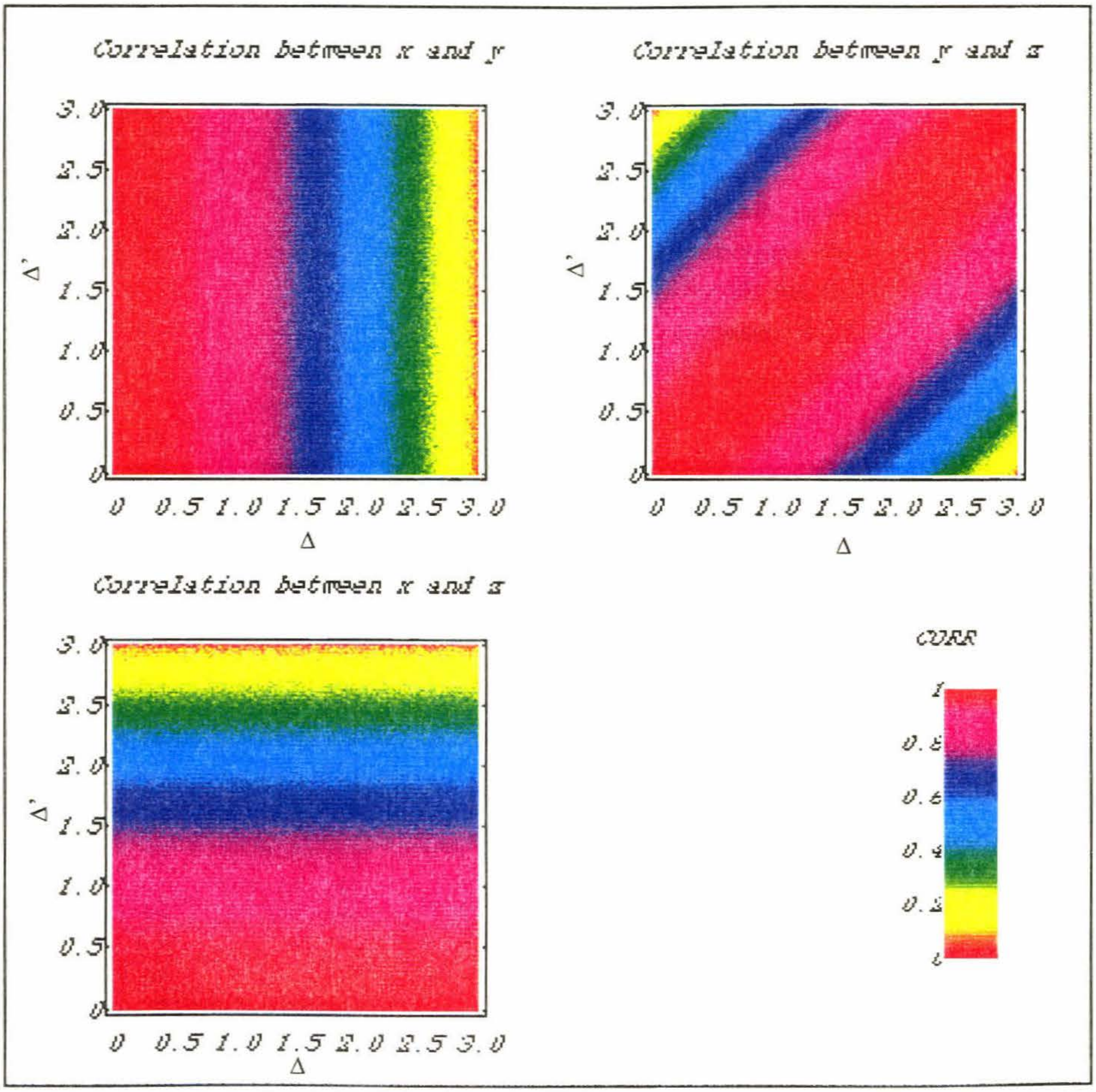
However, a two surface model has the property that the correlation between different pairs of data depends on the values of Δ and Δ'; the correlation of the day 1 - day 2 pair decreases as Δ increases, the correlation of the day 1 - day 3 pair decreases as Δ' increases and the correlation of the day 2 - day 3 pair decreases as the difference between Δ and Δ' increases (see Figure).

Therefore, *in order to account for the different decorrelations of different pairs of* *data*, the method will fit a two surface model to data which has the property that different pairs of data have significantly different correlations, and the phase thicknesses of the layer, Δ and Δ', will be determined by the amount of decorrelation between different pairs of data. These values are approximately constant over regions of the data of approximately constant correlation because the same values of Δ and Δ' can account for the observed decorrelation of different pairs of data while that decorrelation is approximately constant. Larger or smaller values of the phase differences would lead to larger or smaller decorrelations, respectively, for the model and would not fit the data as well.

I confirm this intuitive argument by introducing the concept of the *equivalent delta*. By simulating the data from 2 surfaces according to equations (III.35) - (III.37), with the assumption that the upper and lower echoes are equal in size, on average (a valid assumption considering the observed output from the method for the putative lower and upper images - they are approximately equal), I am able to determine the correlation between simulated observations of 2 surfaces with varying phase thicknesses Δ and Δ', for a signal to noise ratio of 20 dB. The correlation between the three different pairs (x-y, y-z and x-z) of simulated observations is shown in Figure III-36, plotted as a color as a function of Δ and Δ'. 1,000 simulated pixels are used to determine the correlation for a given pair of values of Δ and Δ'. By averaging the correlation observed for a given Δ for all the different Δ' values simulated, I can generate a look up table that gives the x-y correlation for a given Δ. By reading this look up table, Table III.1, backwards I can get the phase thickness Δ, which I call the equivalent delta, that would reproduce a given correlation.

In this way I am able to generate the equivalent deltas of the correlations between the day 1 - day 2 and day 1 - day 3 observations and compare them with the actual deltas, Δ and Δ' respectively, solved for by the method. In Figures III-37 and III-38 I give the percentage difference between the Δ and Δ' solved for by the method and the equivalent Δ and Δ' from the correlations. The close match between the two sets of quantities confirms that the method is solving for Δ and Δ' such that they account for the observed correlations between observations.

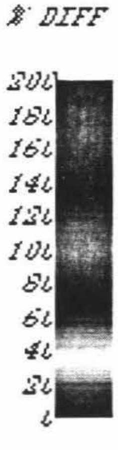
Figure III-36: The correlation between simulated observations of 2 surfaces with varying phase thicknesses Δ and Δ', for a signal to noise ratio of 20 dB. The upper and lower echoes are assumed to be equal in size, on average (a valid assumption considering the observed output from the method for the putative lower and upper images - they are approximately equal). The correlation between the three different pairs (x-y, y-z and x-z) of simulated observations is shown, plotted as a color as a function of Δ and Δ'. The key giving the correlation as a function of color is given in the Figure. 1,000 simulated pixels are used to determine the correlation for a given pair of values of Δ and Δ'.


Correlation between x and y
Correlation between y and z
Correlation between x and z
Δ'
Δ
0
0.5
1.0
1.5
2.0
2.5
3.0
CORR
1
0.8
0.6
0.4
0.2
0

DELTA	CORRELATION
0	0.9901076
0.05	0.9897963
0.1	0.9888697
0.15	0.9873074
0.2	0.9851897
0.25	0.9823325
0.3	0.9790478
0.35	0.9748722
0.4	0.9702846
0.45	0.9650953
0.5	0.9592941
0.55	0.9528806
0.6	0.9458709
0.65	0.9381697
0.7	0.9304091
0.75	0.9211999
0.8	0.9121228
0.85	0.9023329
0.9	0.8914208
0.95	0.8806915
1	0.8690355
1.05	0.8571105
1.1	0.8440507
1.15	0.8309709
1.2	0.8173378
1.25	0.8031592
1.3	0.7880887
1.35	0.7729184
1.4	0.7568122
1.45	0.7408752
1.5	0.7242196
1.55	0.7064513
1.6	0.6883126
1.65	0.6710449
1.7	0.6524412
1.75	0.6340641
1.8	0.6148759
1.85	0.5959794
1.9	0.5774811
1.95	0.5554637
2	0.535207
2.05	0.5151898
2.1	0.4912219
2.15	0.4692003
2.2	0.4499159
2.25	0.4262012
2.3	0.4052626
2.35	0.3829548
2.4	0.3605759
2.45	0.3358445
2.5	0.3119958
2.55	0.2891931
2.6	0.2663847
2.65	0.240659
2.7	0.2183204
2.75	0.1934225
2.8	0.1716396
2.85	0.144639
2.9	0.1208761
2.95	0.099427454
3	0.071358211

Table III.1: *Delta equivalent* look up table created by averaging the correlation for a given Δ over all the different Δ' values simulated in Figure III-36.

Figure III-37: The percentage difference between the Δ solved for by the method and the equivalent Δ from the correlation between the day 1 and day 2 observations. The close match between the two sets of quantities confirms that the method is solving for Δ such that they account for the observed correlations between observations. The key for the percentage difference as a function of color is given below.



% Diff between real and equivalent delta - 32*32 averaging

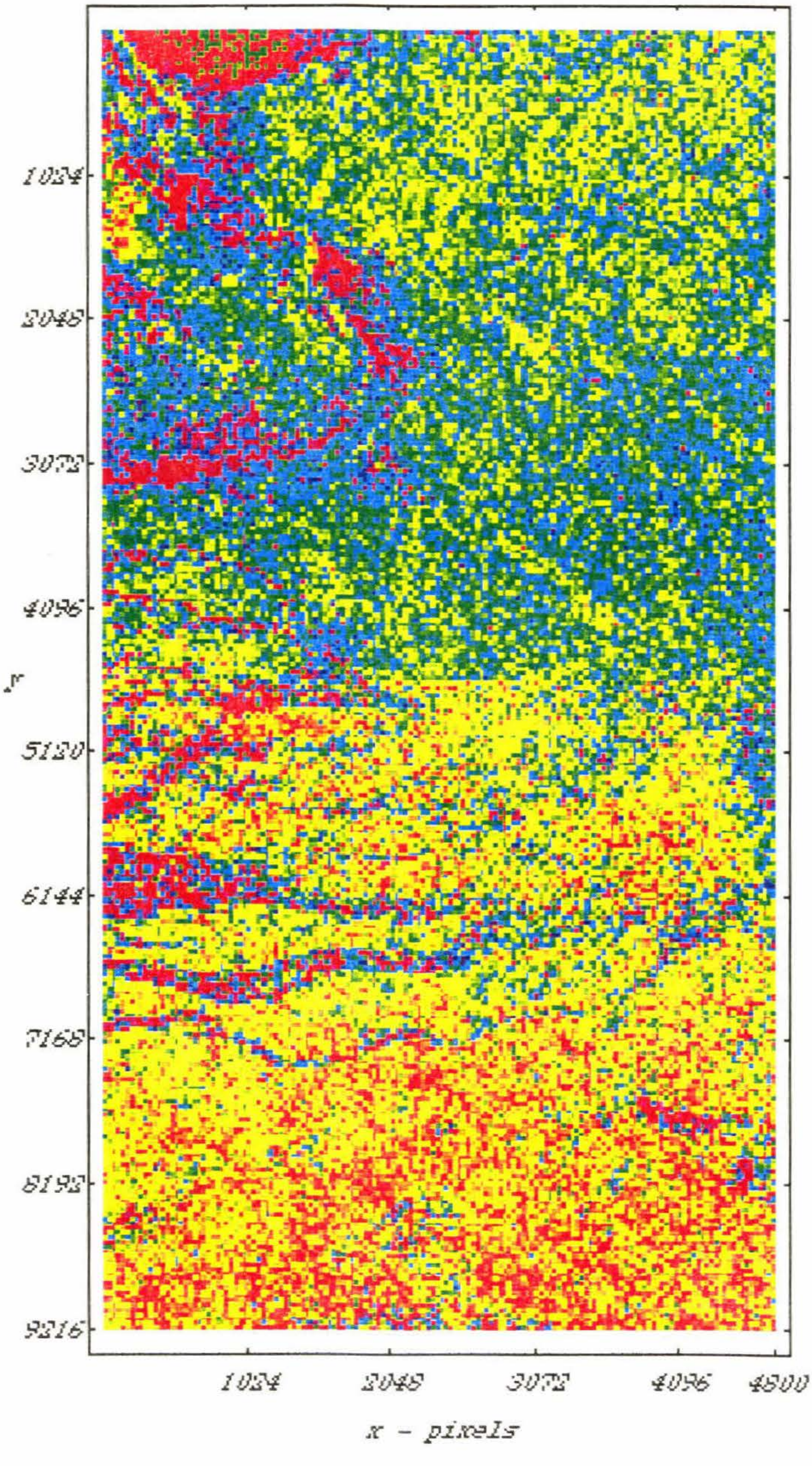
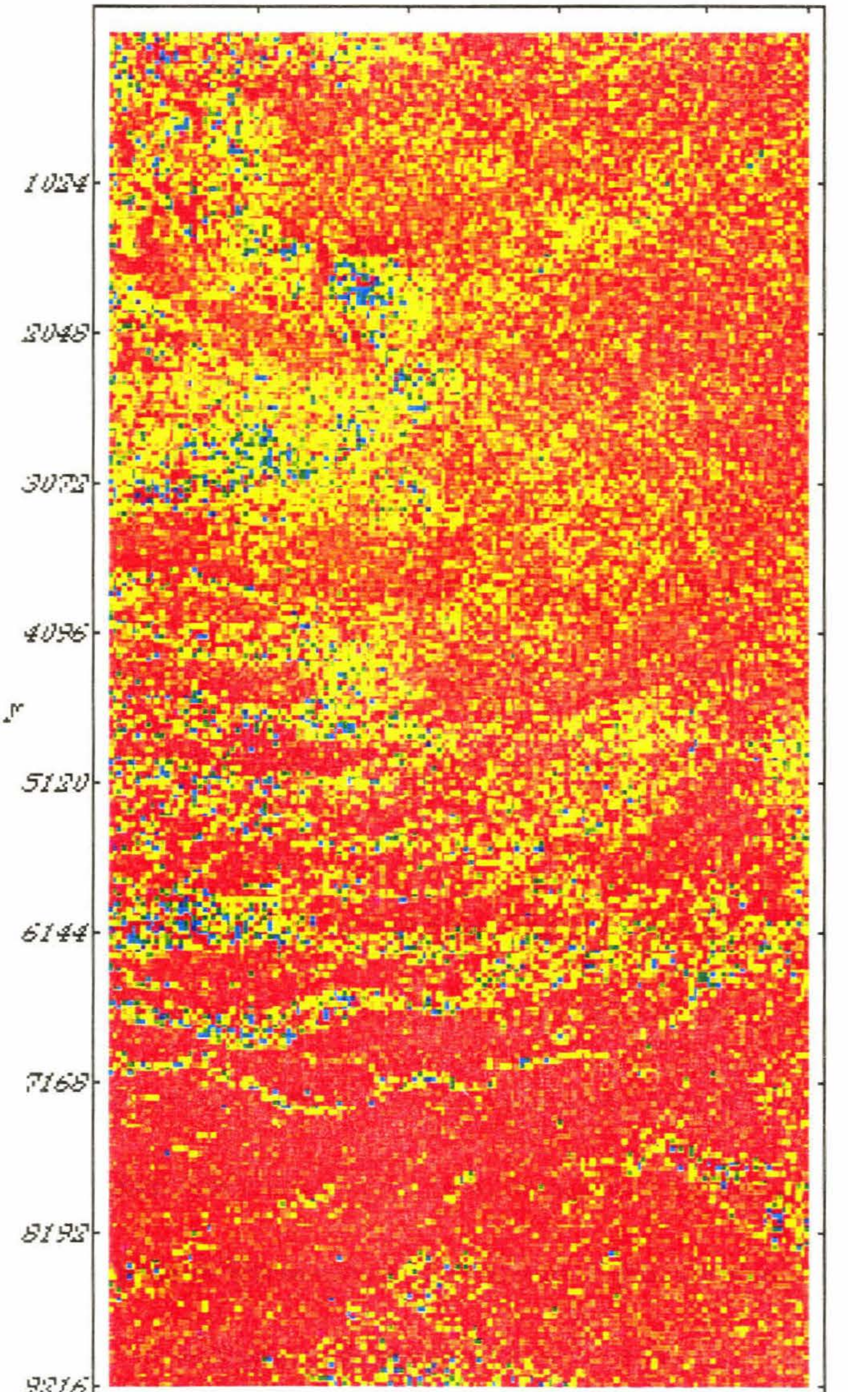


Figure III-38: The percentage difference between the Δ' solved for by the method and the equivalent Δ' from the correlation between the day 1 and day 3 observations. The close match between the two sets of quantities confirms that the method is solving for Δ' such that they account for the observed correlations between observations. The key for the percentage difference as a function of color is given below.



% Diff between real and equivalent delta' - 32*32 averagin

III.17 Requirements for Successful Application of the Method in the Future

Excessive temporal decorrelation between observations prevents the signature from a subsurface layer from being detected. This temporal decorrelation, between observations acquired on successive days, is due to some combination of physical changes in the scene, changes in the spacecraft attitude and errors in the processing by NASA of the raw radar echoes into the synthetic aperture radar images. Assuming the elimination of the last factor, I describe in the next section a test to determine whether or not simultaneous observations will be required for successful application of the method in the future (to eliminate the physical changes in the scene as a source of temporal decorrelation). I then detail the general radar system requirements for successful application of the method and the specific requirements for both possible outcomes of the test.

III.17a Description of a Test to Determine Whether or Not Simultaneous Observations Will be Required

The orbital period at the altitude of the space shuttle (around 200 km) is approximately 2 hours, which suggests the basis for a test of whether or not simultaneous observations must be used. The minimum time between repeat observations from orbit of a certain scene on the ground is one orbital period (assuming orbital corrections are made to compensate for the planet's rotation during the orbit). If physical changes in the scene during this 2 hour interval cause excessive decorrelation then we will know that simultaneous observations are required.

The test is to image a scene where subsurface imaging occurs on successive orbits, performing the required orbital corrections to compensate for the planet's rotation during the orbit and taking care to keep the spacecraft's attitude, specifically yaw and pitch, as similar as possible during the imaging. Keeping the attitude identical on the imaging runs minimizes decorrelation due to differences in viewing geometry. In describing the radar system requirements for the cases of non-simultaneous and simultaneous observations in the following two sections I will quantify the requirements of spacecraft attitude control. For the test I propose here however it is enough to say that the spacecraft's attitude must be kept as identical as possible for the imaging runs, for we are trying to isolate the minimum decorrelation between repeat observations due to physical changes in the scene.

I also assume that the resulting radar echoes are correctly processed by NASA into synthetic aperture radar images. The decorrelation between the observations can now be ascribed to physical changes in a scene where subsurface imaging occurs, and this is the minimum decorrelation between non-simultaneous observations of the scene. If this decorrelation is comparable to or exceeds the signature expected from a subsurface layer, simultaneous observations will be required, otherwise repeat (non-simultaneous) observations may be used.

III.17b General Requirements

The perpendicular component of the baseline must be long enough that the phase thickness of the layer, Δ, is large relative to the signal to thermal noise ratio (SNR). This condition was quantified by the numerical simulations of section III.9, with the result that Δ should be approximately 1 radian in order to deal with a commonplace SNR of 10 dB (Figure III-17).

The physical separation of the antennas necessary to realize a Δ of 1 radian may be quantified. The relationship between the phase thickness of the layer (Δ) and the perpendicular component of the baseline ($B_{\perp_1}$) is (equation (III.28))

$$\Delta = \frac{4\pi B_{\perp_1}}{\lambda} \frac{L}{r_1 \cos\theta} \left(\frac{\varepsilon_1 \sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}} \right) \tag{III.28}$$

where L is the physical depth of the subsurface layer, r_1 is the distance from the antennas to the pixel, λ is the wavelength, ε_1 is the dielectric constant of the material overlying the lower surface and θ is the look angle, which is 90^o-incidence angle. The expression in parentheses is the effect of refraction. Rewriting this expression in terms of the altitude of the antennas, h, using the relationship $r_1 \sin\theta = h$, we get

$$\Delta = \frac{4\pi B_{\perp_1}}{\lambda} \frac{L \tan\theta}{h} \left(\frac{\varepsilon_1 \sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}} \right) \tag{III.55}$$

Using the values of the parameters for the SIR-C data (L=2m, h=222 km, λ=5.7 cm, ε_1=3.5, $\theta=40^o$) we have

$$\Delta = 2.20 \times 10^{-3}\, B_{\perp_1} \tag{III.56}$$

where $B_{\perp_1}$ is in meters. In order that $\Delta = 1$ radian we must have $B_{\perp_1} = 455$ m.

Once the thermal noise is dealt with by using long baselines, we immediately run into a complication; in making the baselines long to deal with the thermal noise we have simultaneously made the spatial decorrelation worse.

These competing effects may be quantified. The spatial decorrelation (1-spatial correlation) is just the spatial fringe frequency in fringes/pixel. The number of fringes per pixel times 2π is the phase change per pixel change across track, which I explained in section III.14 is proportional to the phase thickness of the layer. Therefore, we see that the phase thickness of the subsurface layer, Δ, is proportional to the spatial decorrelation, and writing out the constant of proportionality we have

$$\Delta = \frac{L}{D\tan\theta}\left(\frac{\varepsilon_1 \sin\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) 2\pi \times (\textit{spatial decorrelation}) \qquad \text{(III.57)}$$

Here L is the physical depth of the layer and D is the pixel width across track, and all other variables have been previously defined. We may simplify this expression to

$$\Delta = \frac{L}{D}\left(\frac{\varepsilon_1 \cos\theta}{\sqrt{\varepsilon_1 - \cos^2\theta}}\right) 2\pi \times (\textit{spatial decorrelation}) \qquad \text{(III.58)}$$

Therefore, the phase thickness of the layer, Δ, which we would like to be relatively large to deal with the thermal noise is directly proportional to the spatial decorrelation, which we need to be small otherwise it will introduce a bias into the solution like temporal decorrelation that is larger than the signature from the subsurface layer that we are trying to detect. The only way to resolve these competing requirements is to make the constant of proportionality large.

In (III.58) the fractional term in parentheses has a value of 1.4 for the SIR-C data, with an approximately linear increase to 1.88 for a 90° incidence angle, so there is not much room to increase the constant of proportionality by increasing the incidence angle. The only way to do so is to make the pixel width D narrow compared to the layer depth L.

As I will later quantify, the pixel width should be no larger than 1/4 of the layer depth, which translates to 50 cm wide pixels for the SIR-C data, compared with the actual pixel width of 4.4m, an order of magnitude larger.

In retrospect, it is not surprising that we were unable to confirm the method with the SIR-C data, for SIR-C, which represents the state of the art in spaceborne radars, falls far short of the requirements for successful application of the method. Through some combination of physical changes in the scene, changes in the spacecraft attitude and errors in the processing by NASA of the raw radar echoes into synthetic aperture radar images, the data suffers from too much temporal correlation between observations to allow the signature from a subsurface to be detected. In addition, the pixel width is 4.4 m, compared with a maximum expected subsurface layer depth of 2 m. Substituting these values into equation (III.58) we have Δ=3.95x(spatial decorrelation). A 2 meter deep layer has a phase thickness of 0.15 for the SIR-C x-z baseline, and so the spatial decorrelation is 0.038 (spatial correlation is 0.962). From the look up Table III.1 we can see that this decorrelation would be equivalent to a layer 0.47 radians deep in phase, much less than the 2 - 3 radian equivalent layer phase depth due to temporal decorrelation (it is the dominant effect), but still three times bigger than the signature from the subsurface layer.

While the short baselines of SIR-C are not ideal - the phase thickness of 0.15 corresponds to a perpendicular component of baseline of 68 m - the numerical simulations of section III.9 confirm that, had it not been for the overwhelming effect of temporal decorrelation (and, to a lesser extent, spatial decorrelation), averaging over as many pixels as we did (1,024) would have allowed the signature from the subsurface layer to be detected, at least in the brightest regions with the highest signal to thermal noise ratios.

The fact that the most advanced radar system currently available falls well short of the requirements necessary to successfully apply the method indicates the unfeasibility of using the method with current radar systems. However, as we described in the introduction and will discuss more fully in section III.18, this method has a number of important practical applications that are unique in being based on the way the method gives direct information about the subsurface environment, such as the depth and the strength of the lower and upper echoes, using the geometry of the situation. Normally, complicated models that predict the strength of the radar return at various incidence angles, polarization combinations and wavelengths, and that contain, say, the depth of the subsurface layer as a variable, are used to obtain such information, which is obviously a very indirect way of doing so that is subject to many more assumptions the interferometric method uses.

The applications of the method, unique by not relying on complicated models with many variables, include measuring the depth of burial of ice in the polar regions of Mars, enhancing the visibility of buried features and soil moisture mapping, at the high spatial resolution afforded by synthetic aperture radar, in arid regions of the Earth. These important applications may not be realizable with the method using current radar systems, but I consider them of sufficient importance, and possibly not addressable by any other method, that it is worth describing in detail the requirements of a radar system that could be used to successfully apply the method and so realize these applications.

III.17c Specific Requirements in the Case that Simultaneous Observations Must be Used

In this case the requirements for dealing with temporal decorrelation due to physical changes in the scene and thermal noise are, in a sense, competing, because the necessary simultaneous observations must be acquired over baselines hundreds of meters in length. This rules out the possibility of using a single spacecraft, with antennas spaced out along a fixed boom. Instead, satellites flying in independent orbits, but in "formation" so that the orbital paths are parallel but displaced by hundreds of meters, would be required to realize such long, simultaneous baselines; a much more complicated scenario.

At this point I can demonstrate the generality of the method by pointing out that only *one* of Δ or Δ' needs to be relatively large in order to combat the thermal noise; *the other may in fact be zero*. This follows from realizing that the key point about the information contained in equations (III.35) - (III.37) is that they represent independent observations of the pixel. As long as a separate antenna is used the thermal noise will be different and the measurement of the pixel with provide a constraint, through the least squares method, on the range of possible values of the lower and upper echoes and phases. As long as, say, Δ is relatively large to give the requisite phase sensitivity, the other, Δ', may be zero with no effect on the accuracy of the solutions. This can be further seen from the fact the only place in the method that zero values for Δ or Δ' could possibly be problematic is in equation (III.47), since the first term on the right hand side of the equation is an inverse matrix, which produces a term that is the reciprocal of the determinant of the matrix, which is a function of Δ and Δ'. Specifically, the determinant is

$$det = 9 - \left(1 + e^{-i\Delta} + e^{-i\Delta'}\right)\left(1 + e^{i\Delta} + e^{i\Delta'}\right) \tag{III.59}$$

Expanding the right hand side to second order in Δ and Δ', since the first order terms cancel out, we get

$$det = 2\left(\Delta^2 + \Delta'^2 - \Delta\,\Delta'\right) \tag{III.60}$$

If we put both Δ and Δ' equal to zero we run into problems, because then we are dividing by zero (not just to second order but exactly as can be seen from (III.59)), but if we put just one of Δ or Δ' equal to zero then the determinant is still of order the square of the non zero term.

Therefore, of the three independent, simultaneous observations that are required, two may be acquired from adjacent antennas with hardly any separation; this is obviously easily achievable on a single spacecraft. However, an additional, simultaneous observation is still required with a long baseline; hence a minimum of two satellites flying in formation are required.

A further illustration of the generality of the method that has relevance now that we are quantifying in detail the necessary radar system requirements for successful application of the method is that we may use a single transmitting and receiving antenna and merely receive the echoes at the other two antennas, rather than having all three antennas transmit and receive. This generality follows from considering that the only difference in receiving echoes at, say, ψ and ζ when the pixel is illuminated from ξ, compared with receiving echoes at ψ and ζ when the pixel is illuminated from ψ and ζ, respectively, is that all the expressions for the phases A, C, Δ and Δ' are halved, since there is only a difference in the distance traveled by the echoes on the return path, compared with both the outbound and return paths when all three antennas transmit their own signals.

There is no significant difference in the spatial decorrelation between the observations when only one antenna is used to transmit compared with when each antenna transmits its own signal since the fundamental cause of spatial decorrelation is a change in the relative distance between the receiving antenna and the myriad of sub resolution facets that the pixel can be thought to be made up of. Whether the pixel is illuminated by the receiving antenna or another antenna doesn't alter these relative distances between the receiving antenna and the sub resolution facets of the pixel, so there is no change in the spatial decorrelation. There is a very slight change in the radiation pattern from a facet when it is illuminated by an antenna different from the receiving antenna, but the angular distance between ξ, ψ, and ζ subtended at the pixel is so small that the change in the radiation pattern from a facet is infinitesimal, and so the spatial decorrelation is unchanged. Hence, except for the factor of 2 difference in the expressions for the phases, the method is independent of whether all three antennas transmit their own signals or just one antenna is used to transmit a signal that is received by all three antennas.

Hence I propose a radar system comprised of two satellites, one of which contains two antennas and the other a single antenna, and of the three antennas only one needs to transmit the radar signal. All three antennas receive the radar echoes. The satellite with the two antennas can have the antennas displaced by a minimal distance. However, the second satellite orbits parallel to this satellite, displaced by at least 455 m in the direction perpendicular to the line of sight to the region being imaged in order to create a Δ of 1 radian (the displacement parallel to the line of sight is irrelevant), assuming a C-Band wavelength of 5.7 cm and an orbiting altitude of 222 km. In addition, I assume that the antenna size is the same as SIR-C, 12.1 m length and 0.8 m width, and that the radiated power is the same at 3.6 kW. I deliberately keep those variables for which it is possible the same as SIR-C, since this way the required advancements to the radar system are advancements to the state of the art of current spaceborne radars. In addition, I am describing the radar system requirements necessary to successfully apply the method on the same region of Egypt/Sudan as was tested in this thesis with the SIR-C data, since this area is optimal for testing the method as I previously discussed.

A Δ of 1 radian will be sufficient to detect the signature from the subsurface layer even in regions with a SNR (thermal) of 10 dB, which should be common in the brighter regions. Because the observations are simultaneous, temporal decorrelation due to physical changes in the scene between the observations is avoided.

In order to quantify the pixel width, we note from Table III.1 that the spatial decorrelation (1-spatial correlation) can be expressed in terms of the equivalent delta by the empirical relationship (at least for the range of equivalent deltas extending from 0 to 0.5 radians): spatial decorrelation = 0.123x(equivalent delta)^2. Substituting this expression for the spatial decorrelation into equation (III.58) and using the incidence angle and surface dielectric constant characteristic of the SIR-C data we get

$$\Delta = \frac{L}{D} 1.08 \times (\mathit{equivalent\ delta})^2 \qquad \text{(III.61)}$$

If we allow the equivalent delta due to the spatial decorrelation to be at most 0.45, compared to the actual delta of the subsurface layer of 1, which is an error comparable to that introduced by the thermal noise, we get from equation (III.61) that the pixel width D should be at most 22% of the subsurface layer depth. For a subsurface layer depth of 2 m, this translates to a pixel width of 44 cm, which is ten times smaller than the actual SIR-C pixel width of 4.4m.

Since the pixel width is inversely proportional to the bandwidth, a bandwidth of ten times the SIR-C bandwidth of 100 MHz, or 1 GHz (compared with the radar frequency of 5.3 GHz), is required. Increasing the bandwidth by a factor of 10 is not a problem, the problem is that the signal to thermal noise ratio, hereafter called SNR, is inversely proportional to the square of the bandwidth. The bandwidth enters once into the denominator of the expression for SNR from the pixel width - the signal is proportional to the pixel area, which is proportional to the pixel width, which is inversely proportional to the bandwidth. The bandwidth also enters the denominator once because the thermal noise is proportional to the bandwidth. This would imply a factor of 100 increase in transmitted power from SIR-C; however, a common tactic for increasing the bandwidth in order to realize higher spatial resolutions without decreasing the SNR is the use of a linear frequency modulated signal, called an FM chirp signal, with a compression ratio, the ratio of the duration of the chirp signal to the inverse of the bandwidth, that can be over 100 [*Elachi*, 1987]. The SNR is proportional to the compression ratio, so a compression ratio of 100 can offset a factor of 10 increase in the bandwidth. The SIR-C signal already utilizes this method, so additional compression by a factor of 100 would be required to maintain the SNR at current levels without increasing the transmitted power from the current SIR-C value of 3.6 kW (increasing the radiated power of a spaceborne radar is problematic [*Elachi*, 1987]), if the pixel width is reduced by a factor of 10.

The pointing accuracy of the two satellites is defined by the requirement that there be a high degree of overlap of the physical area that is imaged by the two satellites, in order that the same area is observed by both satellites. Because the method utilizes imaging radar, pixels are located based on range and Doppler coordinates. As long as the same physical area is within the swath of the antennas the pixels can be co-located from the range and Doppler coordinates. Therefore it is not necessary that the two satellites have the same pointing angle to within a pixel width on the ground, just that the two swaths overlap significantly. The angle subtended by the swath width at SIR-C is 2.3°, so the pointing accuracy requirement is that the two satellites point to within $\pm 0.5^{\circ}$ of the same direction.

Another significant requirement for the radar system is possibly the accuracy with which the perpendicular component of the baseline between the two satellites is maintained. We can quantify this requirement by noting that the method assumes that the phases A, C, Δ and Δ' are constant within a certain patch of pixels. In order that this assumption not be undermined by drifting of the baseline, because the phases are all proportional to the perpendicular component of the baseline, it is necessary that the baseline be stable to, say, 1 part in 10 along the orbital path for a distance equal to the along track length of a patch of pixels. In this case the phases will be constant to within 10% within the patch assuming the upper and lower surfaces within the patch are physically flat. This produces an error small compared with that introduced by the thermal noise.

If we assume that a patch of 90 pixels is averaged over (because Δ is around 1 radian so not as much averaging is necessary), then assuming a SIR-C pixel length of 4.4 m and pixel width of 0.44 m, as discussed previously, a logical shape for the patch is 3 pixels along track (corresponding to a length of 3x4.4=13.2 m along track) times 30 pixels across track (corresponding to a length of 30x0.44=13.2 m across track), requiring that the lower and upper surfaces be flat over an area 13 m square. The perpendicular component of baseline (which is nominally 455 m) must therefore not change by more than 45 m for an orbital path length of 13 m.

At an altitude of 222 km the orbital velocity is 7.7 kms^{-1}, so the time required for the satellites to travel down the orbital path a distance of 13 m is 1.67 milliseconds, so the magnitude of the relative speed of the two spacecraft in the direction perpendicular to the line of sight to the region being imaged must be below 27 kms^{-1}, which obviously occurs by default since this is above the escape velocity. Therefore, the drifting apart or together of the two orbits is not an immediate issue for the method, because the method does not assume the various orbital parameters are constant across the entire image, just within the patch of pixels being averaged over. However, if the baseline becomes too small, below a few hundred meters, the requisite phase sensitivity vanishes, and if the baseline becomes too large, over 3.5 km for the scenario I have quantified, the images become completely decorrelated and hence useless, so during the imaging the baseline must be kept between 450 m and 3.5 km. Variation within this range is not an issue.

The orbital property that the method is quite sensitive to is what I call the lag/lead distance, and it is the distance along the orbital path that the second (arbitrarily defined as the one with the single antenna) satellite lags or leads the first satellite. If the second satellite transmitted its own signal, then this would not be an issue, as long as the lag/lead distance was not so great that the time between the observations by the first and second satellites (which is the lag/lead distance divided by the orbital velocity of 7.7 kms^{-1}) allowed significant temporal decorrelation between the observations. If the second satellite transmits its own signal, the pixels can be coregistered with the first satellites pixels, using the method that I used to coregister the SIR-C data acquired on successive days. The geometry of the observations is the same in both cases, except of course for the displacement perpendicular to the line of sight.

If, however, the second satellite merely records the echoes from the single transmitter on the first satellite, then any displacement along the orbital path relative to the first antenna creates a different observation viewpoint not only along the intentional baseline but also along the orbital path; the lag/lead distance acts as a second baseline that decorrelates the signal. One can intuitively see that this is fundamentally the same as the spatial decorrelation we have previously discussed, and so is proportional to the product of the pixel *length* along track and the lag/lead distance. If we are to make this decorrelation, say, one third that from the intentional baseline, in order to keep the accumulated effect of spatial decorrelation manageable, then we must have that the product of the pixel length and the lag/lead distance is 1/3 the product of the pixel width (across track) and the perpendicular component of the (intentional) baseline.

The latter product is 0.44 meters times 450 meters, so if the pixel length is 4.4 meters (we do not want to make it any smaller in order to preserve the SNR) then the lag/lead distance must be at most 1/30 that of the intentional baseline, or 15 meters. Therefore, the second satellite must neither lag nor lead the first satellite by more than 15 meters if the second satellite is to avoid carrying its own transmitter. If a second transmitter is used on the second satellite (no matter what the second antenna on the first satellite can merely receive echoes since it is fixed on the same satellite as the transmitting antenna and so there is no lag or lead distance problem) then this requirement is relaxed, and instead the second satellite must simply not observe the same scene either too much earlier or later than the first satellite such that significant temporal decorrelation due to physical changes in the scene occurs between the observations. Presumably a lag/lead time of minutes is acceptable. This compares with an allowable lag/lead time of 1.95 milliseconds (the equivalent time for a lag/lead distance of 15 m) if the second antenna does not carry its own transmitter.

III.17d Specific Requirements in the Case that Repeat Observations Can be Used

This is the preferable scenario, since only one spacecraft is needed, and the minimum of three observations can be acquired from displaced orbits of this single spacecraft. This is in fact how the SIR-C observations were acquired. Parts of the discussion in the previous section about the radar system requirements in the case that simultaneous observations are required also apply here. Specifically, the insight that one of Δ or Δ' may be zero is still true. Also, the requirement that for a subsurface layer depth of 2 m the pixel widths be around 44cm, which is ten times smaller than the actual SIR-C pixel width of 4.4m, still applies. The resulting discussion in the previous section regarding the implications for bandwidth, signal power and signal chirp applies identically in this case of repeat observations.

The major difference between the radar system requirements for the cases of simultaneous and repeat observations, respectively, aside from the number of satellites needed (2 and 1, respectively), is the difference in spacecraft attitude control requirements. Specifically, when simultaneous observations are made from 2 satellites as discussed in the previous section, there is a lag/lead requirement on the second satellite flying in formation but no yaw requirement for the satellites beyond the pointing accuracy requirement that the two satellites point to within $\pm 0.5^{o}$ of the same direction.

In the case of repeat observations of the scene made from a single satellite, there is no such lag/lead requirement, but there is a more stringent requirement for the stability of the yaw of the satellite during the repeat observations. If the yaw of the satellite is different for the next observation, it is as if the ground rotated in the direction opposite to the change in the yaw. This causes the pixels to be viewed from different angles, which is the fundamental cause of spatial (baseline) decorrelation. The maximum tolerable difference in look angles is the same as the angle subtended at the pixel by the maximum tolerable lag/lead distance quantified in the previous section (15 m). Using a range of 345 km from the SIR-C data for the distance between the spacecraft and the imaged region gives a maximum tolerable difference in look angles of 0.043 milliradians (9 arc seconds). The difference in look angles is the change in yaw angle, so the yaw angle of the spacecraft must be maintained on successive observations to within 9" of the yaw angle during the original observation. Changes in the roll and pitch of the spacecraft do not decorrelate the echoes. This is also true for the case of two spacecraft flying in formation acquiring simultaneous observations.

Now that I have described in some detail the radar system requirements necessary for successful application of the method, I can further illustrate the generality of the method by pointing out that the method can easily be extended to the case of each antenna transmitting and receiving HH and VV radiation, rather than just one direct polarization combination. In this case there is a duplicate set of equations (III.35) - (III.37) for each polarization combination, with a different (in general) lower and upper echo for each polarization combination. However, the geometric phases A, C, Δ and Δ' are the same regardless of the polarization combination (but if one attempts to utilize simultaneous observations at different wavelengths, for example, the phases do change, since they depend on the wavelength) and the net result is that for a single pixel we now have 12 known quantities, the measurement of the complex voltages (x,y,z) and (x',y',z') at both

VV and HH polarizations, respectively, and 12 unknown quantities; l and l', the lower echoes at VV and HH, respectively, u and u', the upper echoes at VV and HH, respectively, and the geometric phases A, C, Δ and Δ'. This compares with 6 known quantities and 8 unknown quantities for a single pixel when using a single polarization combination, which required us to average over at least 2 pixels to gain as many known quantities as unknowns. Each additional pixel averaged over using both direct polarizations provides an increase of known quantities over unknowns of 4, compared with an increase of only 2 if just one polarization combination is used. The net result is that about half as many pixels need to be averaged over for the same accuracy if the observations are made simultaneously at both VV and HH polarization combinations, compared with observing with just one polarization combination. This increases the spatial resolution of the method by the square root of 2, or about 30%.

III.18 Applications

The applications of the method that could be realized by a radar system with the previously described specifications include measurement of the depth of burial of ice in the polar regions of Mars, from orbit, at high spatial resolution and over a regional scale.The dielectric discontinuity between the ice and the overlying clastic material forms the lower surface, while the surface of the planet is the upper surface. The absence of liquid water in the regolith of Mars eliminates the major limitation to subsurface radar penetration on Earth, making Mars an ideal target for subsurface radar imaging.

A second application relevant to subsurface imaging on Earth and Mars is that the separation of the lower and upper echoes would allow the visibility of buried features to be enhanced in regions where the surface has a comparable roughness to the lower surface, as discussed in the introduction. This would extend subsurface imaging beyond those regions with a smooth upper surface relative to the lower surface, which intuitively is much rarer than upper surfaces with a roughness up to the roughness of the lower surface.

Based on the method there are also two independent applications for measuring soil moisture in arid regions of the earth. One application is to take advantage of the ability to separate the lower and upper echoes by measuring the lower and upper echoes in arid regions where an impermeable caliche layer exists below the surface. Changes in the moisture of the soil above the caliche layer cause changes in the skin depth of the soil, which changes the strength of the lower echo. The greater the soil moisture the weaker the lower echo. Since the lower echo can be determined using the method, separated from the upper surface echo, changes in the amount of soil moisture could be detected, at the high spatial resolution afforded by synthetic aperture radar, by changes in the strength of the echo from the caliche layer. The lower surface is protected by the overlying material and so would be expected to suffer much less physical change over time than the upper surface. Hence changes in the strength of the echo from the lower surface can be attributed to changes in the skin depth of the overlying material (unlike changes in the upper surface echo which could be due to physical changes in the surface and changes in the surface dielectric constant due to exposure to the elements). A stronger lower echo observed at a later time would correspond to a decrease in soil moisture between the observations, and vice versa.

Another independent application to soil moisture mapping in arid regions of the earth takes advantage of the ability to measure the depth of a subsurface layer using the method. Moisture present below a certain depth but absent above this depth due to evaporation would form a layer (like a caliche layer) if the distance over which the moisture increases at the boundary of the evaporated zone was small compared to the wavelength, which is on the order of cm. As the ground becomes drier the depth of the moisture layer will increase, and so changes in the depth of this layer, detected using the method, could indicate changes in the amount of soil moisture, because a decrease in depth indicates the soil became more moist and vice versa.

The applications to soil moisture mapping require a very high spatial resolution across track, since, as quantified in the previous section, the pixel width must be no more than approximately 1/4 of the subsurface layer depth. In the case of subsurface imaging on the Egypt/Sudan border, a relatively deep subsurface layer up to 2 meters deep can be expected. In arid regions of the Earth the depth of a caliche layer or water table would be significantly smaller than this, so a comparably narrower pixel width must be used (and hence a higher FM chirp compression ratio must be used if the SNR is to be maintained with no increase in the radiated power of the radar).

The application to measuring the depth of burial of ice in the polar regions of Mars is perhaps easier to realize in terms of radar system requirements, albeit the radar system must be in orbit over Mars rather than the earth. The depths of burial here could easily exceed 2 meters, and, as mentioned earlier, the regolith is devoid of liquid water, so only the grain size of the clastic material overlying the ice is a factor in limiting penetration (the grain size should be no larger than 1/10th of the radar wavelength, or no larger than around 5mm if C-Band is used).

The application of enhancing the visibility of buried features could be used on radar data of the Egypt/Sudan border, where I suggest attempts to confirm this method with a future radar system meeting the requirements I have described be concentrated, also *relatively* easily, since the depths of burial here are so large. The applications to soil moisture mapping are, however, the most practical and important, and so I hope that spaceborne radar systems will soon advance to the point that the requirements I have described here can be met to realize the applications to soil moisture mapping at high spatial resolution.

References

Blom, R. G., Crippen, R. E., and C. Elachi, Detection of subsurface features in SEASAT radar images of Means Valley, Mojave Desert, California, *Geology*, 12, 346-349, 1984.

Born, M., and E. Wolf, *Principles of Optics, 5th ed.*, Pergamon Press Ltd., Headington Hill Hall, Oxford, p. 42, 1975.

Campbell, M. J., and J. Ulrichs, Electrical Properties of Rocks and their Significance for Lunar Radar Observations, *J. Geophys. Res.*, 74, 5867-5881, 1969.

Elachi, C., Roth, L. E., and G. G. Schaber, Spaceborne Radar Subsurface Imaging in Hyperarid Regions, *IEEE Trans. Geosci. Remote Sens.*, GE-22, 383-388, 1984.

Elachi, C., *Introduction to the Physics and Techniques of Remote Sensing*, John Wiley & Sons, Inc., pp. 161-166, p. 184, 1987.

Feynman, R. P., Leighton, R. B., and M. Sands, *The Feynman Lectures on Physics, Volume 1*, Addison-Wesley Publishing Co., Inc., Reading, Massachusetts, pp. 6-8 - 6-9, pp. 32-8 - 32-9, 1963.

Goldstein, R. M., H. Engelhardt, B. Kamb, and R. M. Frolich, Satellite Radar Interferometry for Monitoring Ice Sheet Motion: Application to an Antarctic Ice Stream, *Science*, 262, 1525-1530, 1993.

Goldstein, R. M., Atmospheric Limitations to Repeat-Track Radar Interferometry, *GRL*, 22, 2517-2520, 1995.

Kursinski, E. R., Hajj, G. A ., Bertiger, W. I., Leroy, S. S., Meehan, T. K., Romans,

L. J., Schofield, J. T., Mccleese, D. J. , Melbourne, W. G., Initial Results Of Radio Occultation Observations Of Earths Atmosphere Using The Global Positioning System, *Science*, 271, 1107-1110, 1996.

McCauley, J. F., Schaber, G. G., Breed, C. S., Grolier, M. J., Haynes, C. V., Issawi, B., Elachi, C., and R. Blom, Subsurface Valleys and Geoarcheology of the Eastern Sahara revealed by Shuttle Radar, *Science*, 218, 1004-1019, 1982.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN The Art of Scientific Computing, Second Edition*, Cambridge University Press, Cambridge CB2 IRP, pp. 413-418, pp. 650 - 653, 1992.

Roth, L. E., and C. Elachi, Coherent Electromagnetic Losses by Scattering from Volume Inhomogeneities, *IEEE Trans. Antennas Propagat.*, AP-23, 674-675, 1975.

Schaber, G. G., McCauley, J. F., and C. Breed, The Use of Multifrequency and Polarimetric SIR-C/X-SAR Data in Geological Studies of Bir SafSaf, Egypt, *Remote Sens. Environ.*, 59, 337-363, 1997.

Valenzuela, G. R., Depolarization of EM waves by Slightly Rough Surfaces, *IEEE Trans. Antennas Propagat.*, AP-15, 552-557, 1967.

Valenzuela, G. R., Scattering of Electromagnetic Waves from Tilted Slightly Rough Surfaces, *Radio Sci.*, 3, 1057-1066, 1968.

Zebker, H. A., and R. Goldstein, Topographic Mapping From Interferometric Synthetic Aperture Radar Observations, *J. Geophys. Res.*, 91, 4993-4999, 1986.

Zebker, H. A., and J. Villasenor, Decorrelation in Interferometric Radar Echoes, *IEEE*

Trans. Geosci. Remote Sens., 30, 950-959, 1992.