INFORMATION TO USERS

The most advanced technology has been used to photo-graph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are re-produced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17" x 23" black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Seismic tomography in Western Washington

Lees, Jonathan Matthew, Ph.D.

University of Washington, 1989

SEISMIC TOMOGRAPHY IN WESTERN WASHINGTON

by

Jonathan Matthew Lees

A dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

University of Washington

1989
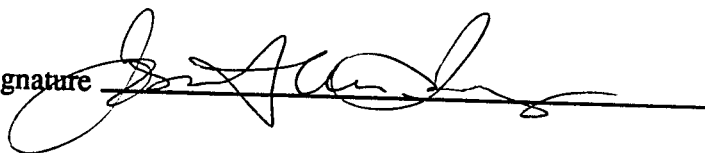
Approved by _____
(Chairperson of Supervisory Committee)

Program Authorized
to Offer Degree _____ Geophysics _____

Date _____ 12 - 13 - 88 _____

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 300 North Zeeb Road, Ann Arbor, Michigan 48016, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature

Date _____ 12 - 13 - 88

University of Washington

Abstract

SEISMIC TOMOGRAPHY IN WESTERN WASHINGTON

by Jonathan M. Lees

Chairperson of the Supervisory Committee: Professor Robert S. Crosson
Graduate Program in Geophysics

Tomographic inversion techniques have been applied to local earthquake travel times to delineate lateral crustal velocity variations in Western Washington. Shallow earthquakes from the early 1970's to the present, ranging to depths of 40.0 km, are used as sources. Two methods for inverting large sparse matrix systems are compared: an iterative back-projection method, ART, using relaxation and smoothing to attain regularization versus a conjugate gradient method, LSQR, which uses constraints to regularize. These techniques are compared with synthetic examples that simulate the characteristics of real data inversion. The resolution is approximated by calculating impulse responses at blocks of interest and estimates of standard errors are calculated by the jackknife. Initial reference models are one dimensional layered velocity structures derived by least squares analysis.

In the Puget Sound region the target consists of a 150x250 km area divided into blocks of 5 km per side. In depth the model extends to 40 km divided into 10 layers. A total of 4387 earthquakes gave rise to 36,865 raypaths which had good resolving power in the 2-16 km depth range. High correlations with known features are apparent in the shallow structures and evidence for accretionary underplating dipping to the east appear at depth beneath the low velocity sediments of the Puget Sound. In the Mt. St. Helens region, a separate inversion was performed with 17,659 rays using an 80x80 km grid divided into cells 2.0 km per side. Major structural features such as the Spirit Lake and Spud Mountain plutons are evident in the shallow layers. The Saint Helens Seismic zone is characterized by a prominent low velocity feature. A low velocity anomaly beneath the crater at depths of 4-9 km may indicate the presence of magma accumulations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

v

ACKNOWLEDGEMENTS

DEDICATION

For my mother

CHAPTER 1
INTRODUCTION and THEORY

1. INTRODUCTION

The objective of this research is to determine a detailed 3-dimensional structure of the P-wave velocity in the upper 40 kilometers beneath western Washington. A refined knowledge of the velocity structure has small scale and large scale implications. On the local scale we can gain insight into the geologic structure and its relationship to surface features. Also, better knowledge of velocity variations may improve our ability to locate earthquakes. On the broader scale it is hoped that a detailed knowledge of the crustal structure will shed light on the large scale structure of the Cascadia subduction zone, first by shedding light on the accretionary nature of large geologic features present on the surface and second by allowing the removal of the distorting shallow structure when studying deeper structures. Finally, by refining the information we have on the geologic structure we may ultimately be able to improve our estimation of earthquake hazard in this region.

The origins of tomography date to the early 1960's when researchers in the medical fields discovered that high quality 2-dimensional images of internal anatomy can be obtained by applying inverse techniques to x-ray projections. The application of these techniques to geophysical data did not begin until the late 1970's, due in part to the large amounts of high quality data required for this kind of study. With the advent of such data sets and powerful computing facilities, tomography has become a viable means of studying the earth's interior. The extensive shallow seismicity of western Washington and the comparatively dense seismic network established since the early 1970's offer an opportunity to develop and apply the tomographic techniques to velocity inversion on a regional or local

scale.

This dissertation addresses several issues related to the application of tomographic techniques to local earthquake data. First, an introduction to the theory and initial analysis of the data is presented. Next, a detailed comparison of the two major computational techniques commonly used in geophysical tomography is presented. Special attention is paid to the way each method employs regularization. The methods are each tested using a set of synthetic data where the distribution of rays and noise is known. Additionally, several approaches to estimating resolution are suggested and a new way to estimate standard errors is proposed. Finally the techniques are applied to real data in two target areas: the Puget Sound and Mt. St. Helens. The results indicate a high correlation with known structures at shallow depth lending credence to inferred structure at greater depth.

2. LINEARIZATION

Under the approximations of ray theory, we assume that the time a seismic signal takes to travel from point A to point B (Figure 1.1), in a given medium, is a function of the seismic velocity of the intervening material and the path, called the ray-path, that the wave traverses. Determining the travel time, given the velocity and the ray-path, is called the forward problem, and is written mathematically as,

$$T = \int_{ray} \frac{1}{v(\mathbf{x})} dr \qquad (1.2.1)$$

where T is the travel time, $\mathbf{x}$ is the spatial position vector, and dr is a differential line element along the path from A to B. The travel time is thus the line integral of the inverse of the velocity along the ray-path. Since the travel time T does not depend linearly on the velocity, $v(\mathbf{x})$, it is convenient to introduce the 'slowness', $s(\mathbf{x})$, where $s(\mathbf{x}) = 1/v(\mathbf{x})$. Then our functional relationship between the travel

time and the model becomes,

$$T = \int_{ray} s(\mathbf{x})dr \tag{1.2.2}$$

Because the ray-path itself depends on the slowness (the velocity) the problem is still non-linear. We consider the following perturbation approach. Suppose the model, $s(\mathbf{x})$, that we are seeking can be assumed to be a reference model, $s_0(\mathbf{x})$, plus a small perturbation $\delta s(\mathbf{x})$, i.e.

$$s(\mathbf{x}) = s_0(\mathbf{x}) + \delta s(\mathbf{x}) \tag{1.2.3}$$

Then,

$$T = \int_{ray} [\ s_0(\mathbf{x}) + \delta s(\mathbf{x})\ ]dr \tag{1.2.4}$$

Fermat's principle states that, to first order, the travel time is stationary with respect to small perturbations in the ray-path [*Aki et al*, 1977]. We can then integrate over the ray-path in the unperturbed model to determine the right hand side of Equation (1.2.4). If T_0 is the travel time for the unperturbed model then,

$$T = \int_{ray_0} s_0(\mathbf{x})dr + \int_{ray_0} \delta s_0(\mathbf{x})dr \tag{1.2.5}$$

or, rearranging,

$$\delta T = T - T_0 = \int_{ray_0} \delta s(\mathbf{x})dr \tag{1.2.6}$$

where δT is called the travel time residual. We now have a linear relationship between the travel time residual, δT, and the slowness perturbation, δs. Our data set, or observations, consists of a set of travel time residuals (observed minus computed arrival times) and we wish to determine the slowness perturbations by linear inversion. For simplicity in notation we will drop the spatial dependence $\mathbf{x}$ and from our equations and refer to δs as simply s and δT as simply t, remembering that these are the slowness perturbations and residuals respectively. Using

techniques borrowed from medical tomography we parameterize the structure by partitioning it into small cells within which the slowness perturbation is considered constant. The real slowness perturbation field s_{TRUE} is thus approximated by the discrete version s, where it is assumed the blocks are chosen small enough such that $s \approx s_{TRUE}$, and there is no aliasing of structure. The inverse problem is then discretized by considering sums instead of integrals in Equation (1.2.6). The travel time residual will be the sum of the slowness perturbation in each cell times the length of the ray within that cell. For many such observations we have,

$$t_n = \sum_{j=1}^{m} a_{nj} s_j \qquad (1.2.7)$$

where t_n is the travel time residual associated with the n-th ray, and a_{nj} is the length of the n-th ray in the j-th cell. In matrix notation this can be expressed as,

$$\mathbf{t} = \mathbf{A}s \qquad (1.2.8)$$

where $\mathbf{t}$ is an n-row column vector of observations, $\mathbf{A}$ is an nxm matrix of coefficients, called the information matrix, describing the lengths of each ray in each cell, and s is the slowness perturbation vector of length m. The "inverse problem" involves finding a solution $\tilde{s}$ which satisfies (1.2.8).

3. LINEAR SYSTEMS

We begin by considering the linear system of equations described in (1.2.8). This system is generally nxm where n represents the number of equations in the system and m is the number of model parameters in s. Typically $n>m$ and the system is apparently overdetermined. However, the rank of the matrix $\mathbf{A}^T\mathbf{A}$ is often less than m and the system will be thus underconstrained. Such a system is often inconsistent or ill-conditioned, as will be the case when 2 equations in the system have the same coefficients on the right hand side of (1.2.8) but different travel times are observed. No exact solution can be found for an inconsistent system,

while an infinite number of solutions are possible for an underconstrained system [*Menke*, 1984].

As a first consideration in solving the system of equations in (1.2.8) we desire a solution which will predict the data. To insure that this is the case we will want some measure of the difference between the observed travel times and the predicted travel times, namely the residuals, to be small. This can be expressed mathematically as

$$\min\left\{\phi(t - As)\right\} \qquad (1.3.1)$$

where ϕ is some suitable distance measure. In classical least squares the function ϕ is the sum of the squares of the residuals:

$$\phi(t - As) = (t - As)^T(t - As) = \|t - As\|^2 \qquad (1.3.2)$$

If we consider the data, t, and the model, s, to be samples of multivariate Gaussian distributions then this choice of ϕ is the maximum likelihood estimator for problem (1.2.8). Because the system of equations in (1.2.8) is likely to be underconstrained, the solution, ŝ, which satisfies the condition that ϕ be minimized, will in general be non-unique and unstable [*Titterington*, 1985]. For this reason we must impose an additional constraint on the model that produces a unique solution which also predicts the data. One such constraint is that the variance of the model be small, i.e. find a model s that predicts the data, such that $s^Ts = \|s\|^2$ is simultaneously minimized. Our goal is to then find the maximum likelihood minimum variance model and this can be expressed as finding the minimum of the following functional:

$$\|t - As\|^2 + \lambda\|s\|^2 \qquad (1.3.3)$$

where λ is the trade-off parameter that regulates the relative importance we assign to models that predict the data versus models that have small variance [*Herman*,

1980]. It should be noted that other constraints or regularizations are also possible.

The quality of the data depends on a variety of factors, one of which is the ability of the analyst to pick the time of arrival on the seismogram. For signals of high frequency this can be done fairly consistently, but much of the data will have lower frequency components and the data will have a higher degree of uncertainty. Since we have more confidence in picks that come from better quality data we will weight these data more heavily in the inversion process. The analyst who picks the data estimates a confidence interval that represents the standard error of the pick about the mean. Using these estimates we multiply each equation in (1.2.8) by $1/\sigma_i$, where σ_i is the estimated uncertainty in the i-th datum. In addition to simple row weighting we may anticipate covariances between data points, as would be the case for several observations obtained from the same earthquake. If the covariance matrix of the data is $\mathbf{C}_{data}$, then we form the weighting matrix

$$\mathbf{W} = \mathbf{C}_{data}^{-1/2}$$

and Equation (1.2.8) is transformed to:

$$\mathbf{Wt} = \mathbf{WAs}. \tag{1.3.4}$$

This is the standard approach taken in weighted least squares [Bevington, 1969]

In a similar fashion we may have a priori knowledge of the covariance of the model parameters. First, if we know in advance what the ray coverage is, we may weight blocks differently depending on the configuration of rays in the vicinity of the block. For instance, blocks that have heavy coverage may be weighted in such a way as to reflect the fact that we have more confidence in them as opposed to blocks that are lightly sampled [Spakman and Nolet, 1988]. Second, we may have a priori knowledge of the geological structures in the subsurface, like the location of faults or the presence of specific buried features which we obtain from

external sources. In either case we can require that the solution have the desired covariance C_{model} by considering a change of variables,

$$x=C^{-1/2}s$$

Then the system is replaced by

$$Wt = WAC^{1/2}x \tag{1.3.5}$$

[*Van der Vorst and Van der Sluis*, 1987].

At this point we return to the question of regularization alluded to earlier. Regularization is the mathematical term used for damping least squares problems, referred to as ridge regression in statistics [*Lawson and Hanson*, 1974] and sometimes called smoothing. Regularization can be accomplished by augmenting the system of equations in (1.2.8) with an artificial set of equations, one for each model parameter:

$$\begin{bmatrix} A \\ \lambda F \end{bmatrix} s = \begin{bmatrix} t \\ 0 \end{bmatrix} \tag{1.3.6}$$

As before, λ, is a trade-off parameter that regulates the relative importance one puts in minimization of the prediction error versus minimization of the variance of the model. If $F = I$, this expression leads to the minimization of the functional ϕ from Equation (1.3.3) [*Lawson and Hanson*, 1974; *Herman*, 1980].

There may be situations, though, where we have a priori information about the smoothness of the model. In such cases we would require that the solution have a certain degree of smoothness. To do so we can take two approaches: either we find a model that maximizes a smoothing constraint or one that minimizes a roughening constraint. Smoothing constraints are the inverses of roughening constraints: in filter theory a smoothing operator is a low pass filter and a roughening operator is a high pass filter.

4. THE DATA

The data used in this study are derived from seismograms recorded on the western Washington seismic network in operation from the early 1970's. The short period stations record vertical motion and telemeter the signals over various communication channels to the data processing center at the University of Washington. If the signals are found to be from earthquake events, the system is triggered and the data are saved in digital form on magnetic tapes. An example of one such event is provided in Figure 1.2.

A seismic analyst examines the events and determines the onset of the first arrivals of the P-waves and the S-waves if these are observable. Portions of several seismic traces have been presented in Figure 1.2. The actual picks for these data and the limits of uncertainty have been plotted on each seismogram. In the first four traces the signal to noise ratio is very high and picks of first arrival times are made quite accurately. In the last trace the signal to noise ratio is considerably lower and there is a high amount of uncertainty in determination of event onset. The wide error bars bounding the pick reflect this fact. In general, when there is a moderate amount of noise in the data and the event is emergent (i.e. has low frequency components), there is a consistent tendency for the analyst to pick the first arrival late. This introduces a bias in the data set of observations. Presently we do not have a way of estimating and removing this bias, but research is being conducted towards this goal. These picks are then used to independently locate each event in space and time using Gieger's method, which is basically a linearized least squares inversion [*Lee and Stewart*, 1981].

4.1. 1-D REFERENCE MODEL

To determine a 1-dimensional reference model from which perturbations will be calculated, a small subset of high quality data are selected and a joint hypocenter-velocity parameter inversion is performed [*Crosson*, 1976]. Since the number of parameters in this inversion is small, classical damped least squares inversion is used. The resulting velocity model is presented in Table (1.1). Since for the tomographic inversions a finer vertical spacing was desired, new depth boundaries were calculated by averaging between layers. Figure 1.4 shows the final dependence of velocity with depth that was used to locate earthquakes as well as the reference model for the tomographic inversions.

4.2. RAY TRACING

Both for the purposes of earthquake location and tomographic inversion (solution of Equation 1.2.8) it is necessary to determine the locus of points in the model that the waves traverse. In the presence of some simple velocity structures the determination of raypaths can be found quite easily [*Cerveny*, 1987]. For the layered models used in this study, for example, refracted rays can be calculated analytically and direct raypaths can be found using the shooting method [*Lee and Stewart*, 1981]. However, a more sophisticated approach to ray tracing, using 3-dimensional velocity structure, would be required to perform a true non-linear velocity inversion. The use of the above simplification is primarily due to the speed of calculation in computing the raypaths. *Cerveny* [1987] reviews techniques for calculating raypaths in more complicated models, but these were not used in this research. Future work on this data should include some kind of iterative method that incorporates laterally varying velocity fields, either performing true 3-D raytracing or applying an approximation more appropriate than the 1-D horizontal model used here.

4.3. STATION CORRECTIONS

If we collect all the residuals for a given station in the network and plot a histogram showing the distribution of the residuals we would expect, under ideal conditions, the character of the distribution to be gaussian and have approximately zero mean. An examination of the data shows this not to be the case, and we therefore modify our model of travel time residuals to include a term for this bias in the distribution. We assume that the bias is due to local structure in the immediate vicinity of the station, structure due to the peculiarities of a weathered layer or instrument variability. The travel time residuals thus have the following form:

$$t = t_0 + \delta t + \phi_k \qquad (1.4.1)$$

where ϕ_k is the station correction for the k-th station.

The question we must address is how to determine the station correction, ϕ_k. One simple approach would be to calculate the arithmetic mean of the residuals for all residuals at a particular station. However, examination of the distributions of the residuals for the many of the stations indicates that the mean would be a poor choice of indicator for distributions that are highly skewed and occasionally bimodal. On the left of Figure 1.3 histograms of residuals at four stations are displayed with no station corrections applied in the location procedure. Note the bimodality and non-normal distributions, especially for station SPW.

The alternative we've chosen in this study is to examine the individual histograms of each station and attempt to choose a representative average that characterizes the center of the distribution. The procedure is coupled to the earthquake locations, as the addition of the station correction term will modify the relative travel times for a given event. Iterating between station correction determination and event location is required to converged to a stable choice of station corrections. These are then considered fixed and the travel time observations of

Equations (1.2.6) and (1.2.8) are modified by adding the appropriate station correction. The histograms on the right side of Figure 1.3 show the distribution of residuals after the station corrections have been applied and the earthquakes relocated. Notice the significant improvement of the shape of the distributions for stations MBW and SPW.

A different approach to calculating the station corrections has been studied extensively by *Pavlis* [1982]. This approach, called joint hypocenter-station correction inversion attempts to determine the location of the multiple events and the value of the station corrections simultaneously. However, it can be shown [*Pavlis*, personal communication] that in the limit of large numbers of events the resulting station corrections are equivalent to the arithmetic means described earlier with the same drawbacks. For this reason we chose not to pursue this method.
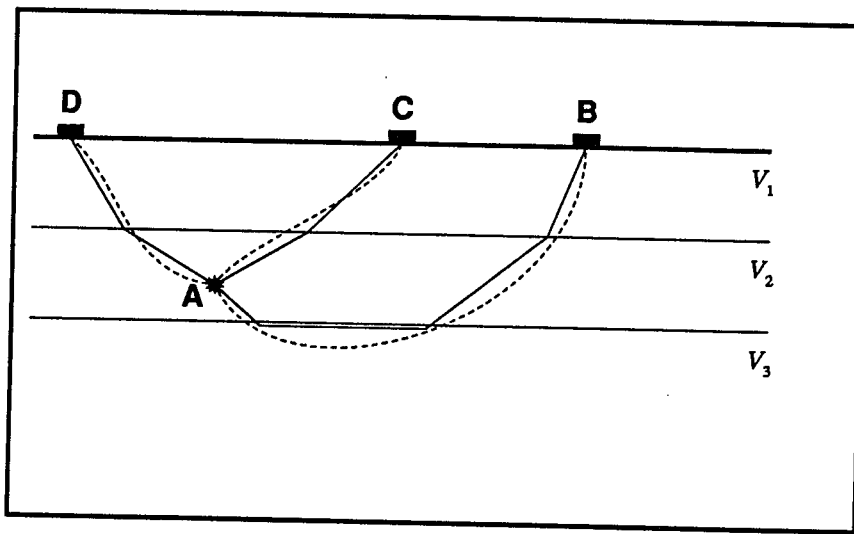
Figure 1.1: Hypothetical raypaths in a heterogeneous model. The dotted lines represent the true raypaths and the solid lines represent the raypath approximation in a 1-D layered model. Two kinds of rays are present for this approximation: **AD** and **AC** are direct arrivals and **AB** is a refracted wave.
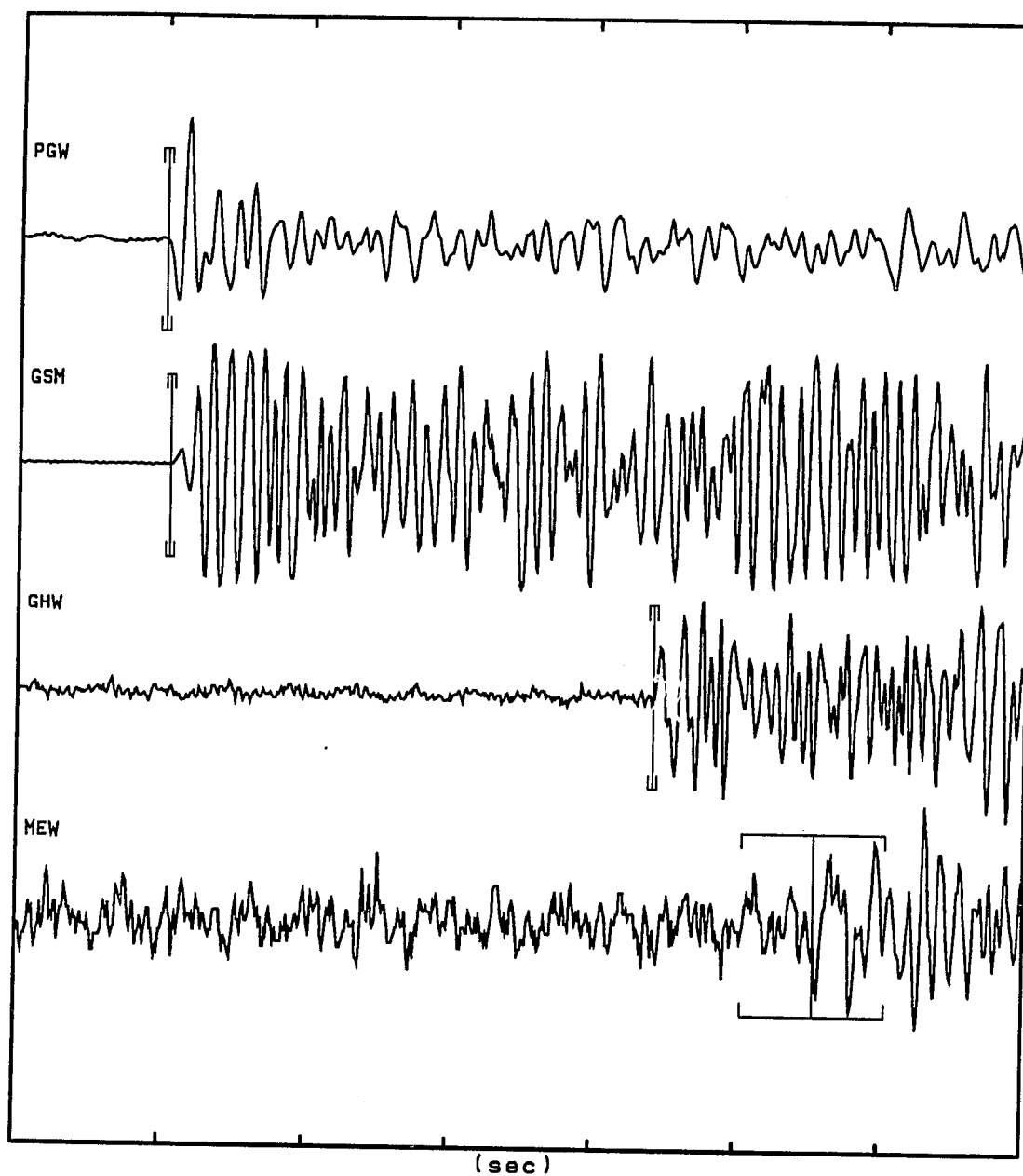
Figure 1.2: Selection of four traces showing actual seismic data recorded on the western Washington seismic network. Vertical marks on the traces indicate first arrival picks and horizontal bars show estimated uncertainty in pick. Station names are displayed to the left.
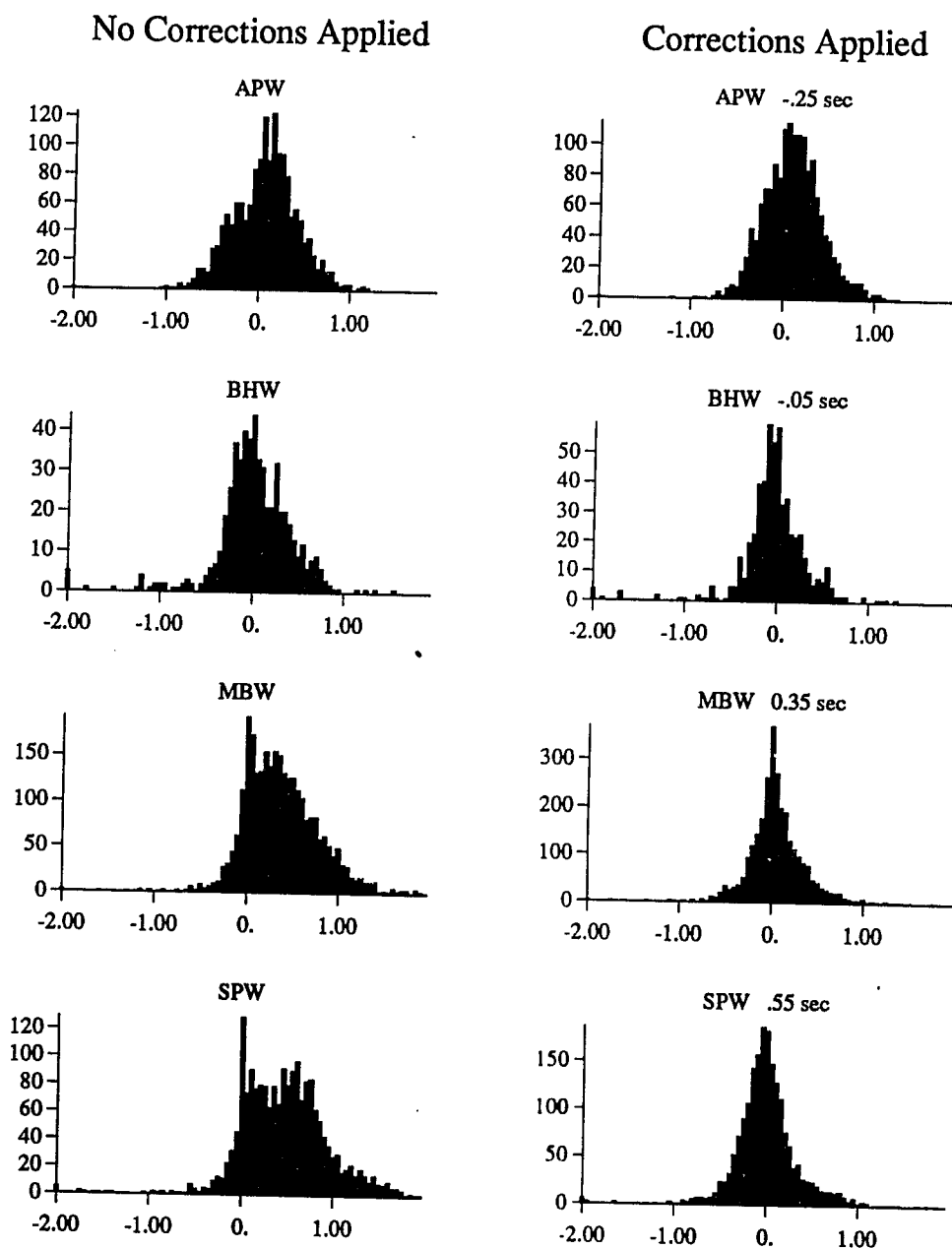
Figure 1.3: Histograms of station residuals. The horizontal axis is in seconds. On the left are the residuals for earthquakes located with no station corrections and on the right are the same residuals with corrections applied (the delay is indicated for each station). Notice the improvement after delays are incorporated.

15
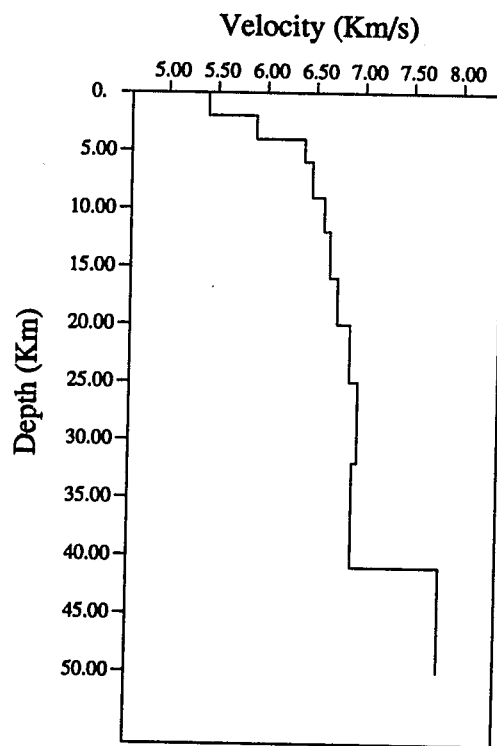
Figure 1.4: 1-D model used for inversion.

Table 1.1: P-wave velocity model based on least squares model developed by *Crosson* [1976]. The velocities and boundaries used in the inversion routine were interpolated from this model.

Velocity Model		
Depth	P-velocity	P-slowness
0.0	5.4	0.185
4.0	6.38	0.157
9.0	6.59	0.152
16.0	6.73	0.149
20.0	6.86	0.146
25.0	6.95	0.144
32.0	6.90	0.145
41.0	7.80	0.128

CHAPTER 2

INVERSION TECHNIQUES

1. LEAST SQUARES

Least squares analysis has been used to solve the problem of Equation (1.2.8) for nearly two hundred years [*Plackett*, 1972]. The approach is quite simple and extensive discussions can be found in standard textbooks [*Strang*, 1980; *Draper and Smith*, 1966; *Lawson and Hanson*, 1974]. The method is motivated in the following manner. Consider the problem described in Equation (1.2.8):

$$Ax = b \tag{2.1.1}$$

We wish to find an x such that

$$\phi = \|Ax - b\| = (Ax - b)^T(Ax - b) \tag{2.1.2}$$

is minimized. To do so we consider ϕ as a function of x, differentiate ϕ with respect to x and set the derivative to zero to determine the extrema of ϕ. Doing this we arrive at the so called normal equations:

$$A^TAx = A^Tb. \tag{2.1.3}$$

2. Singular Value Decomposition

At this point we introduce the ideas behind the singular value decomposition, a method of solving the least squares problem that illustrates several important features common to all methods of solution and gives insight into the nature of regularization [*Golub and Van Loan*, 1983; *Lawson and Hanson*, 1974]. The singular value decomposition is based on the fact that any matrix $A \in \mathbb{R}^{n \times m}$ can be decomposed according to:

$$A = U\Sigma V^T \tag{2.2.1}$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{m \times m}$ are orthogonal matrices spanning the data and model spaces respectively. $\Sigma = diag(\xi_1, \xi_2, \ldots, \xi_k)$ where the ξ_i are known as the singular values of the matrix $\mathbf{A}$.

One can see the immediate usefulness of this decomposition to the understanding of the least squares problem. Substituting (2.2.1) into (2.1.1) gives

$$\mathbf{U}\Sigma\mathbf{V}^T\mathbf{x} = \mathbf{b} \tag{2.2.2}$$

or by premultiplying by $\mathbf{U}^T$ and substituting $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ we get

$$\Sigma\mathbf{y} = \mathbf{U}^T\mathbf{b}. \tag{2.2.3}$$

Since Σ is a diagonal matrix, calculating it's inverse is trivial,

$$\Sigma^{-1} = \begin{bmatrix} \frac{1}{\xi_1} & & & 0 \\ & \frac{1}{\xi_2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\xi_k} \end{bmatrix}. \tag{2.2.4}$$

This naturally leads to a solution of the least squares problem. Since

$$\mathbf{y} = \Sigma^{-1}\mathbf{U}^T\mathbf{b}$$

and $\mathbf{y} = \mathbf{V}^T\mathbf{x}$, substitution and premultiplication by $\mathbf{V}$ gives,

$$\hat{\mathbf{x}} = \mathbf{V}\Sigma^{-1}\mathbf{U}^T\mathbf{b}.$$

The inverse relationship between the singular values of $\mathbf{A}$ and $\hat{\mathbf{x}}$, the estimated solution, indicates that small singular values project into large estimates of the model parameters. Since small singular values may arise from errors (either from measurement or numerical) these would then translate into large and unstable solutions [*Crosson*, 1976]. Regularization techniques are implemented precisely to dampen the effects of small singular values.

The singular values of the normal equations (2.1.3) are the values ξ_i^2 where ξ_i are the singular values of **A**. Subsequently the difficulty encountered with small singular values is compounded, and the normal equations are singular (or nearly singular). This is one of the major reasons the method of normal equations is not used directly for near singular problems. On the other hand if a singular value decomposition is calculated one could examine the singular values and eliminate components that lie below a specified threshold. This will have a regularization effect, effectively smoothing the model.

Alternatively, one could use the Levenberg-Marquandt approach [*Crosson*, 1976] and augment the system by the $m{\times}m$ matrix $\lambda \mathbf{I}$ as in (1.3.6). If we form the normal equations for this system, we get

$$(\mathbf{A}^T\mathbf{A} + \lambda^2\mathbf{I})\mathbf{x} = \mathbf{A}^T\mathbf{b}.$$

By substituting the singular value decomposition of $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ and noting that $\mathbf{A}^T = \mathbf{V}\Sigma\mathbf{U}^T$, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}$ and that

$$\mathbf{V}\Sigma\mathbf{V}^T + \lambda^2\mathbf{I} = \mathbf{V}(\Sigma^2 + \lambda^2\mathbf{I})\mathbf{V}^T$$

we get a solution,

$$\mathbf{x} = \mathbf{V}(\Sigma^2 + \lambda^2\mathbf{I})^{-1}\Sigma\mathbf{U}^T\mathbf{b}.$$

The important point is to note is that the eigenvalues of this system are

$$\frac{\xi_i}{\xi_i^2 + \lambda^2}.$$

The regularization parameter, λ, can thus be adjusted to insure that the effect of very small ξ_i does not have a large effect on the solution **x**.

The main difficulty in using both the normal equations approach or the singular value decomposition is a practical one. In the case of the normal equations one is required to form the matrix $\mathbf{A}^T\mathbf{A}$ and then find its inverse $(\mathbf{A}^T\mathbf{A})^{-1}$. For

very large, sparse matrices the demands on computer memory and time using this approach are prohibitive. Similarly, computer limitations make the singular value decomposition calculation and storage of the orthogonal transformations U and V generally inefficient.

We are thus lead to methods of solving the least squares that avoid the computation and storage of whole matrices such as $A^T A$ or even A and their inverses. The methods used in this study have the property that they require only a small fraction of the storage required to solve the normal equations explicitly. They operate on each row of A sequentially and they exploit the sparsity of A. Procedures of this type are called row action techniques. The row action methods most commonly used in geophysical tomography fall into into two categories: ART (Algebraic Reconstruction Techniques), with its various implementations, and CG (Conjugate Gradient) methods. In the next sections I discuss and implement versions of these techniques.

3. ART - ITERATIVE RECONSTRUCTION

The difficulties that arise in solving the least squares problems can be avoided if a strategy can be developed that avoids the explicit formation and solution of the normal equations. Two approaches will be discussed in this study: the first is an iterative approach that is called in the literature ART, for Algebraic Reconstruction Techniques. The second method is referred to as a CG technique and will be described in the next section.

The ART method originated with *Kaczmarz* [1937], who proposed the following scheme for solving the LS problem. Letting a_i represent the *i–th* row of the A matrix, and given an initial solution s_0 choose an $\tilde{s}$ such that it satisfies the equation $a_i^T s = t_i$ exactly, with the condition that the k-norm of the change $|s_0 - \tilde{s}|_k$ be small. Using the L_2 norm leads us to the solution whose change $(s_0 - \tilde{s})$ is

minimal in euclidian length (minimum energy) [*Van der Vorst and Van der Sluis*, 1987]. The result is to locate the orthogonal projection of a given vector s_0 onto the subspace spanned by a_i and to use this as a new solution $\tilde{s}$ which satisfies the *i–th* equation.

In terms of the tomographic problem we can think of this process in a slightly different way. Now we consider a raypath traversing the intervening medium. Since the total travel time is the sum of the travel times in each section (block) the ray passes (1.2.7), weighted by the length the ray in each block it is natural to take the residual for each ray and distribute it back as a slowness correction into the model, weighted by the proportion of the ray spent in that each block. This can be summarized in the following algorithm:

Algorithm 1 : Kaczmarz Method (ART)

$$s^{(0)} = 0$$

$$s^{(k+1)} = s^{(k)} + \rho_k \frac{(t_i - a_i^T s^{(k)})\, a_i}{\|a_i\|^2} \qquad (2.3.1)$$

Here the ρ_k is a scaler multiplicative factor called the relaxation factor. Each equation of the system is used sequentially to update the present model and the update version the of the model is used to calculate the residual at each step. After all the equations have been used we say that one iteration is complete and iterations are repeated until convergence is achieved. If the system is consistent, and

$$0 \leq \rho_k \leq 2,$$

then this algorithm will converge to the minimum norm solution [*Herman*, 1980]. *Trummer* [1982] has further shown that if ρ_k satisfies the following criteria:

(*a*) $\rho_k \geq 0,$

(*b*) $\sum\limits_{k=0}^{\infty} \rho_k^2 < \infty$

(*c*) $\sum\limits_{k=0}^{\infty} \rho_k = \infty$

then the sequence of vectors $s^{(k)}$ will converge to the minimum norm solution of (2.1.1), provided the rows of **A** are normalized, $\|a_i\| = 1$. This scaling of the rows is discussed in more detail below.

ART-Example

To illustrate how ART operates we consider an example in 2 dimensions following *Herman* [1980]. The problem is to find the solution of 2 equations in 2 unknowns in the 2D plane. The exact solution is the point of intersection of the two lines which can be found analytically by solving the system of linear equations. The example demonstrates the basic ideas behind ART and shows the benefit of using relaxation.

Consider the system of equations:

$$\begin{bmatrix} 4x + y = 24 \\ 2x + 5y = 25 \end{bmatrix}. \tag{2.3.2}$$

To begin we pick an arbitrary starting point, say, (8.0,9.0). We iterate through each equation in turn and find an updated solution by applying the ART algorithm outlined above. To arrive at the first point we apply ART to the first equation

$$4x + y = 24$$

and calculate the perpendicular projection onto the line,

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 8.0 \\ 9.0 \end{bmatrix} + \begin{bmatrix} \dfrac{24.0 - (4.0 \times 8.0 + 1.0 \times 9.0)}{4.0^2 + 1.0^2} \end{bmatrix} \begin{bmatrix} 4.0 \\ 1.0 \end{bmatrix} = \begin{bmatrix} 4.0 \\ 8.0 \end{bmatrix}$$

From (4.0,8.0) the next point is (2.41,4.03) and consecutive points are calculated similarly. When all the equations have been used we call that one iteration and continue by starting at equation 1 for the next iteration. Each successive application of ART is equivalent to dropping a perpendicular down to the subspace spanned by that equation and moving the solution point to that position. This is illustrated quite well in Figure 2.1a where the consecutive points are represented by triangle symbols connected by lines to show the path of convergence. As is evident in the Figure the intersection of the two equations is found in relatively few steps as each updated solution gets closer to the correct solution calculated analytically. In this case it took less than 10 steps to reach the solution.

Next consider what happens when the effect of relaxation is incorporated, as illustrated in Figure 2.1b. Here we note first that the exact solution has not been attained after the given number of steps are passed. The path, though, seems more direct in its approach towards the solution. While the unrelaxed path bounced from equation to equation as it proceeded to the solution, the relaxed version takes a shorter route and never quite arrives at the exact answer. In this synthetic problem where we have a simple exact solution it is clear that the unrelaxed "correct" solution is desirable. But in a situation involving real data with errors we are more likely to have an overdetermined set linear equations to which there may not be a unique solution.

Consider the example in Figures 2.2a and b. Here we've added an equation to the system so that now we are attempting to solve 3 equations and 2 unknowns:

$$\begin{bmatrix} 4x + y = 24 \\ 2x + 5y = 25 \\ x + 5y = 50 \end{bmatrix}.$$

Now compare the solution paths for the unrelaxed ART in Figure 2.2a and for the

relaxed ART in Figure 2.2b. The unrelaxed solution will oscillate between the various lines without ever "converging". The relaxed version chooses a point enclosed by the three lines and converges accordingly. In situations where there are many parallel lines it would be highly undesirable for the solutions to oscillate wildly from one line to the next. In this sense the relaxation helps stabilize the solution for the overdetermined least squares inversion in the presence of noise. The ideas illustrated here in two dimensions can be extended to multidimensional vector spaces. Then each row of the matrix system represents a subspace embedded in the vector space spanned by all possible models. The unrelaxed solutions will oscillate between these subspaces.

Bayesian ART

When the system is underconstrained the solutions to the least squares problem will be non-unique and an additional criteria is needed to choose among those solutions that satisfy the data. We now describe such an algorithm suggested by *Herman et al* [1979] and *Herman* [1980]. In this case we consider the system of equations

$$As = t + r \tag{2.3.3}$$

which is a consistent system since for any solution $\tilde{s}$ we may set $r = As - t$. Therefore a solution to the system will involve m unknowns for the model s and n unknowns for the residuals, r. We consider then a new unknown vector quantity,

$$\begin{bmatrix} r \\ s \end{bmatrix} \tag{2.3.4}$$

and the matrix

$$\begin{bmatrix} \lambda I & A \end{bmatrix}. \tag{2.3.5}$$

We then form the following consistent system of equations,

$$\left[\lambda \mathbf{I} \; \mathbf{A}\right] \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} = \mathbf{t} \, . \tag{2.3.6}$$

Since this is a consistent set of equations by definition we may apply Kaczmarz's ART algorithm to find its least squares solution. It can be shown [*Herman*, 1980] that the solution to this system minimizes

$$\phi = \|\mathbf{A}\mathbf{s} - \mathbf{t}\|^2 + \lambda^2\|\mathbf{s}\|^2 \tag{2.3.7}$$

which is the same as the ϕ discussed in the damped least squares case above. The iterative step for this system of equations becomes:

$$\begin{bmatrix} r^{(k+1)} \\ s^{(k+1)} \end{bmatrix} = \begin{bmatrix} r^{(k)} \\ s^{(k)} \end{bmatrix} + \gamma_k \begin{bmatrix} \lambda \hat{e}^{(k)} \\ a_i^{(k)} \end{bmatrix} \tag{2.3.8}$$

where

$$\gamma_k = \rho_k \frac{t_i - \lambda(r_i) - (a_i^T s)}{\lambda^2 + \|a_i\|^2} \tag{2.3.9}$$

and ρ_k is the relaxation parameter as above. This leads to the following algorithm which is a modification of Kaczmarz's Algorithm:

Algorithm 2 : Bayesian ART

$$r^0 = 0$$

$$s^0 = 0$$

$$\gamma_k = \rho_k \frac{t_i - \lambda(r_i) - (a_i^T s^{(k)})}{\lambda^2 + \|a_i\|^2}$$

$$r^{(k+1)} = r^{(k)} + \gamma_k \lambda \hat{e}_i \tag{2.3.10}$$

$$s^{(k+1)} = s^{(k)} + \gamma_k a_i \tag{2.3.11}$$

Compared to Algorithm 1, Algorithm 2 requires the storage of an additional $N\times1$ vector r of residuals and a few extra multiplications. Otherwise the time and storage efficiencies of Algorithm 1 apply here as well.

Row Weighting

As mentioned in the first chapter the relative importance one places on a particular observation can be controlled by weighting the rows of the system by the weighting matrix W. For the time being we will assume that W is a diagonal matrix, implying zero pair-wise covariance in the data. As a first estimate of the relative importance of the data we assume that travel time picks that are more precise should be weighted more heavily than those that have large uncertainties. Since an estimate for the standard error of each pick is supplied by the analyst this can be done easily by weighting each observation inversely according to the uncertainty in the pick. If the noise in the data has an independent, gaussian distribution applying this weighting scheme leads to the maximum likelihood estimate [*Bevington*, 1969]. Of course, real data are known not to be distributed with such convenient properties, and this model serves as an approximation at best.

We know, for example, that due to possible errors in station corrections there will be correlated noise added to all data recorded at a particular site. Similarly, all data from a particular event will have a common error introduced due to earthquake mislocation.

An important difference between simple ART, represented by Kaczmarz' method, and the Bayesian ART method arises when considering row waiting. Suppose the rows of the system (1.2.8) are scaled according to (1.3.4). The back-projected residual (right hand term) of (2.3.1) becomes,

$$\rho_k \frac{(w_i t_i - w_i \mathbf{a}_i^T \mathbf{s}^{(k)_i}) \, w_i \mathbf{a}_i}{\|w_i \mathbf{a}_i\|^2}$$

and the w_i are seen to cancel. An important advantage of the Bayesian approach is that the incorporation of the regularization parameter λ allows for retention of the proper row waiting since the w_i will not cancel in (2.3.9). The implicit row scaling in simple ART is a serious drawback and is discussed at length by *Van der Vorst and Van der Sluis* [1987]. The implementation of the Bayesian algorithm avoids this problem altogether and allows for explicit waiting.

In order to alleviate some of the error introduced by correlated error along heavily traversed portions of the model we introduce the following weighting procedure. Consider a block in the model that has a large number of earthquakes located within it, like an earthquake swarm. It is quite likely that most of the earthquakes in the swarm will be recorded at the same stations, thus traveling along identical (or nearly identical) raypaths. If the rays that travel along such a bundle have a correlated component of noise that is unaccounted for in the uncertainty estimates for the individual data, an anomalous structure along the blocks penetrated by the bundle will result in the reconstructed image. To compensate for this we propose to 'debundle' the data by weighting all the rows of a bundle according to $1/n$, where n represents the total number of rays in the bundle. In

this study raypaths having the same source and receiver blocks were considered members of a bundle. This is a rather conservative bundle definition and more inclusive (broader) definitions may also be appropriate. Debundling has the effect of evening out the spatial distribution of the rays, giving an equal weight for all raypaths.

Least squares inversions are particularly sensitive to data points that are outliers (large residuals). One solution to this problem is to not use least squares but use some other norm to determine misfit, like, for instance the absolute value (the L_1 norm). Alternatively, if we could dampen out the effect of large outliers we could develop a robust least squares procedure. One way to accomplish this is to weight the equations according to the size of the residual that is being back-projected. The following scheme can be used to determine the additional scaling ($w(r)$) for each equation:

$$w(r) = \begin{cases} 1 & |r| \leq r_1 \\ r_1 \,/\, |r| & |r| \geq r_1 \\ 0 & |r| > r_2 \end{cases} \qquad (2.3.12)$$

where r is the residual for that equation. In the synthetic example given below robust weighting was introduced in every iteration, however this scheme could be introduced in a programmed manner, perhaps depending on iteration number. This would insure that the large outliers do not bias the solution without physically removing data from the problem.

Smoothing

The result of applying the Bayesian ART technique is equivalent to a damped least squares solution. However, we may wish to incorporate additional *a priori* information into the solution if we have knowledge that such constraints are appropriate. An example of this would be the constraint that the model be smooth

over a certain distance measures. This kind of constraint is particularly appropriate for the geological situation explored in this study where we assume that the slowness of adjacent blocks be close to one another on the average. This kind of constraint is easily implemented in the iterative method: after each iteration the resultant model is smoothed with a simple low pass filter. In our case the low pass filter is a 9-point filter for 2-D or a 27-point filter for 3-D where example coefficients are illustrated in Figure 2.3.

There may be situations where we wish to have the level of smoothing be dependent on the iteration number or on the model itself. This can be done simply by forming a linear combination of the smoothed and unsmoothed versions of (k–th) model to form the (k+1–th) iterate. This can be denoted mathematically as,

$$s^{(k+1)}_{smoothed} = \eta s^{(k)} + (1-\eta)F(s^{(k)}) \qquad (2.3.13)$$

where $F(s^{(k)})$ is the filtered version of s and $0 \le \eta \le 1$ regulates the amount of smoothing to be applied. Note that η can be a function of k, the iteration number or anything else, e.g. the variance of s. In this study we allowed η to vary only as a function of k, where typically we reduced the smoothing as the solution approached the final value.

Constraints other than linear filtering may also be applied easily with the ART approach. If we have reason to believe that the model parameters must lie within certain bounds we can constrain them by forcing outlying values to be within the limits [Herman, 1980]. For example, it is unphysical for real material to possess a negative velocity. In this case the model parameters should be constrained to be positive and any model parameter that deviated from this constraint would be set to a minimum value. (Since in this study we are solving for slowness perturbations, this particular type of constraint does not apply). Likewise, constraints such as maximum entropy [Titterington, 1985] or other non-linear smoothing may be conveniently applied if that is considered appropriate.

ART versus SIRT

In Algorithm 1 and 2 the model s is updated after each consecutive equation is used. This is common to iterative methods known as ART. If no relaxation is used this could lead to images that have a streaked or noisy appearance [Herman, 1980]. Introduction of relaxation helps suppress this. An alternative is to sum the incremental updates in a separate vector keeping the model fixed during the calculation of residuals and, after one cycle through the equations, to update the model with the summed perturbations:

$$s^{(k+1)} = s^{(k)} + \Delta s^{(k+1)} \tag{2.3.14}$$

Methods based on this approach are known as the SIRT, for Simultaneous Iterative Reconstruction Techniques. This method has been applied to seismic data to invert travel times for borehole tomography by *Dines and Lytle* [1979], teleseismic tomography by *Humphreys* [1984], and local earthquake tomography by *Nakanishi* [1986]. Since the model is updated after all the equations are used the ordering of the equations has no effect on the solution, as opposed to ART methods which depend on the sequential analysis of the equations. I have found, though, that if the number of equations is large and the order of the equations has little structure (i.e. the order is random) the ordering has little, if no effect on the results of the inversion. Several experiments were conducted with different ordering of the data and the results were virtually the same. I concluded that in the case of local earthquake tomography where sources are distributed quasi-randomly in the target area the application of ART techniques is justified. For controlled experiments, as in medical tomography or bore-hole tomography special care must be taken to avoid artifacts introduced due to ordering. *Herman* [1980] suggests using data that differ by widely varying angles (60°) to help reduce this effect. Based on experiments I conducted, randomization of the input data will have a similar effect.

An extensive discussion of SIRT techniques can be found in Van der Sluis and Van der Vorst [1987] where they compare the SIRT approach to CG techniques which are discussed in the next section. They pointed out the inherent row scaling discussed above and further showed that SIRT methods also have an inherent column scaling concluding that the SIRT approach solves the problem of finding the minimum of

$$\phi' = \|B(As - t)\|^2 + \lambda^2\|Cs\|^2 \qquad (2.3.16)$$

where B and C are the inherent row and column scaling built into the SIRT algorithm, as opposed to minimizing ϕ in (2.3.7). This implicit scaling is a major drawback of the SIRT methods. Note, however, that the scaling problem is absent from the Bayesian Algorithm 2. In this case the scaling factors do not cancel and the only scaling introduced is explicit. This is one major factor in choosing Algorithm (2) over the SIRT methods.

The other reason for using ART over SIRT methods is a practical one. It can be shown that for consistent systems ART methods converge significantly faster than SIRT methods and require less computer storage [*Herman*, 1980]. However, using the Bayesian Algorithm where extra storage and time is required to calculate the error vector r will in general reduce, to some extent, the efficiency of ART over SIRT.

4. CG METHODS

In the past few years alternative approaches to the iterative techniques for solving the least squares problem has emerged as strong contender. These have been dubbed "projection methods" by *Van der Sluis and Van der Vorst* [1987]. The algorithm that I have used in this study is derived from tridiagonalization procedure known generally as the Lanczos Process. In summary this process produces an orthonormal set of bases vectors which are used to transform the system into

tridiagonal form. The transformed system can then be solved simply using a straight forward QR decomposition (see Appendix A). While this procedure is not specifically a conjugate gradient approach, it can be shown that the solutions obtained in this manner are identical to the conjugate gradient solutions [*Paige and Saunders*, 1982].

The CG methods generally involve several matrix-vector multiplications which can be done on a row by row basis as in the row-action ART methods. These techniques are not strictly iterative, in that they arrive at the desired solution in a finite number of steps, but because of computer limitations in accuracy convergence is normally not complete in the prescribed number of iterations. Since the result is an approximation to the final step the method is said to be iterative [*Scales*, 1987].

In this study I have implemented the Algorithm LSQR to solve the tomographic problem of (1.2.8). Regularization has been implemented as described in Equation (1.3.6) where the roughening matrix, **F**, is derived from a second order differential operator applied in 2-dimensions, i.e. the Laplacian defined as follows [*Young*, 1971; *Menke*, 1984]. If u is a continuous function of two variables x,y defined at discrete lattice points separated by a distance h then the Laplacian is,

$$\nabla^2 u(x,y) = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} . \qquad (2.4.1)$$

If we make the approximations

$$\frac{\partial^2 u}{\partial x^2} \approx [u(x + h,y) + u(x - h,y) - 2u(x,y)]/h^2$$

$$\frac{\partial^2 u}{\partial y^2} \approx [u(x,y + h) + u(x,y - h) - 2u(x,y)]/h^2 \qquad (2.4.2)$$

and substitute into (2.4.28) we get the difference equation:

$$4u(x,y) - u(x + h,y) - u(x - h,y) - u(x,y + h) - u(x,y - h)$$

$$= -h^2 \nabla^2 u(x,y) \qquad (2.4.3)$$

If, as a constraint on the model parameters, we set the Laplacian to zero we get constraint equations for each model parameter:

$$4*(j\text{--}th \text{ block slowness})-\textstyle\sum(\text{adjacent slowness}) =$$

$$4 \, u_j - (u_{j+1} + u_{j-1} + u_{j+n} + u_{j-n}) = 0$$

Where blocks are ordered consecutively and n is the number of discrete elements per side of the model. (For a detailed description model numbering and storage see *Scales* [1987]) Applying this to each of the model parameters, and taking care to avoid edge effects, the L matrix (for 2-D) is formed as:

$$L = \begin{bmatrix} 4 & -1 & 0 & \cdots & 0 & -1 & 0 & \cdots \\ -1 & 4 & -1 & 0 & \cdots & 0 & -1 & 0 \\ 0 & -1 & 4 & -1 & 0 & \cdots & 0 & -1 \\ & & \cdot & \cdot & \cdot & & & \\ & & & \cdot & \cdot & \cdot & & \\ -1 & 0 & \cdots & & -1 & 4 & -1 & \cdots \\ & -1 & 0 & \cdots & 0 & -1 & 4 & -1 \\ & & -1 & & \cdots & 0 & -1 & 4 \end{bmatrix}$$

Notice that this roughening matrix shares the sparseness properties of the coverage matrix A. Indeed, the values of L are easily calculated and need not be stored in computer memory at all. The constraint of zero Laplacian produces a model which is smoothed in a fashion similar to a low pass filter.

A better understanding of the smoothing and roughening operations can be achieved by transforming the regularization matrices into the wavenumber domain. Representing the filter by $h(n_1,n_2)$ its 2-dimensional discrete fourier transform [*Dudgeon and Mersereau*, 1984] is formed:

$$H(\omega_1,\omega_2) = \sum_{n = -\infty}^{\infty} \sum_{n = -\infty}^{\infty} h(n_1,n_2) \exp(-i\omega_1 n_1 - i\omega_2 n_2) \qquad (2.4.4)$$

Using this we can derive a general formula for a symmetric 9-point filter whose coefficients defined by are:

$$h(n_1, n_2) = \begin{cases} a, & n_1, n_2 = 0 \\ b, & n_1 = \pm 1, n_2 = 0 \\ b, & n_1 = 0, n_2 = \pm 1 \\ c, & n_1 = \pm 1, n_2 = \pm 1 \\ 0, & \text{otherwise} \end{cases}$$

or, written in the form of a filter,

$$\begin{bmatrix} c & b & c \\ b & a & b \\ c & b & c \end{bmatrix}$$

which we will represent simply as [a,b,c]. The wavenumber response is derived by substituting these filter coefficients into (2.4.4),

$$H(\omega_1, \omega_2) = a + 2b[\cos(\omega_1) + \cos(\omega_2)] + 4c[\cos(\omega_1)\cos(\omega_2)] .$$

From this formula we can generate the frequency response of the smoothing filter we used in the previous chapter for regularization between ART iterations or any other 3-point filter. For example a typical smoothing filter (in 2-dimensions) is [4.0, 1.0, 0.0] whose wavenumber response is displayed in Figure 2.4, where the low pass nature of the filter is illustrated.

For the situation where we constrain the Laplacian to be zero, let the roughening filter be [4.0, -1.0, 0.0], which results in a typical high pass filter illustrated on the right of Figure 2.4. This filter is the complement of the smoothing filter [4.0,1.0,0.0] presented above which can be produced in the space domain or the wavenumber domain: by subtracting a spike (all-pass filter) in the space or the frequency domain from our filter we get the complement filter. The wavenumber domain illustrations in Figure 2.4 display the complementary nature of the two

approaches: in one case the smoothness is maximized and in the other the rough-ness is minimized.

This suggests a general approach one may take to the regularization problem presented here. One could devise a smoothing (or roughening) filter in the 2 or 3-dimensional spatial frequency domain that encorporates *a priori* knowledge one possesses regarding the nature spectral components inherent in the structure that is being imaged. The filter is then transformed back to the space domain and sam-pled for application in the inversion routine. Since it is crucial to keep the number of points in the filter to a manageable number care must be taken in speci-fying the ideal filter to avoid large sidelobes in the transformed domain. An example of this procedure is provided by *Dudgeon and Mersereau* [1984] where they produce an 11×11 point smoothing filter derived from an ideal, cylindrically symmetric, low pass filter defined in the frequency domain. The use of window-ing, e.g. hanning windowing, can be incorporated to reduce the effects of side lobes and edge effects.

The main power of this approach is realized when we consider incorporation of *a priori* information other than simple smoothing criteria used for regulariza-tion. Consider a situation where we know of lineations, stratifications or other organized patterns inherent in the geologic structures such as may appear in sedi-mentary basins or folded belts. In these cases a description of the periodicity of the pattern is accomplished intuitively in the (k_x, k_y, k_z) domain, where filters and constraints may be devised prior to inversion, transformed to *(x,y,z)* space, and used as additional constraints over those parts of the model that share these features. With this approach the full power of multidimensional signal analysis comes to bear upon the problem.
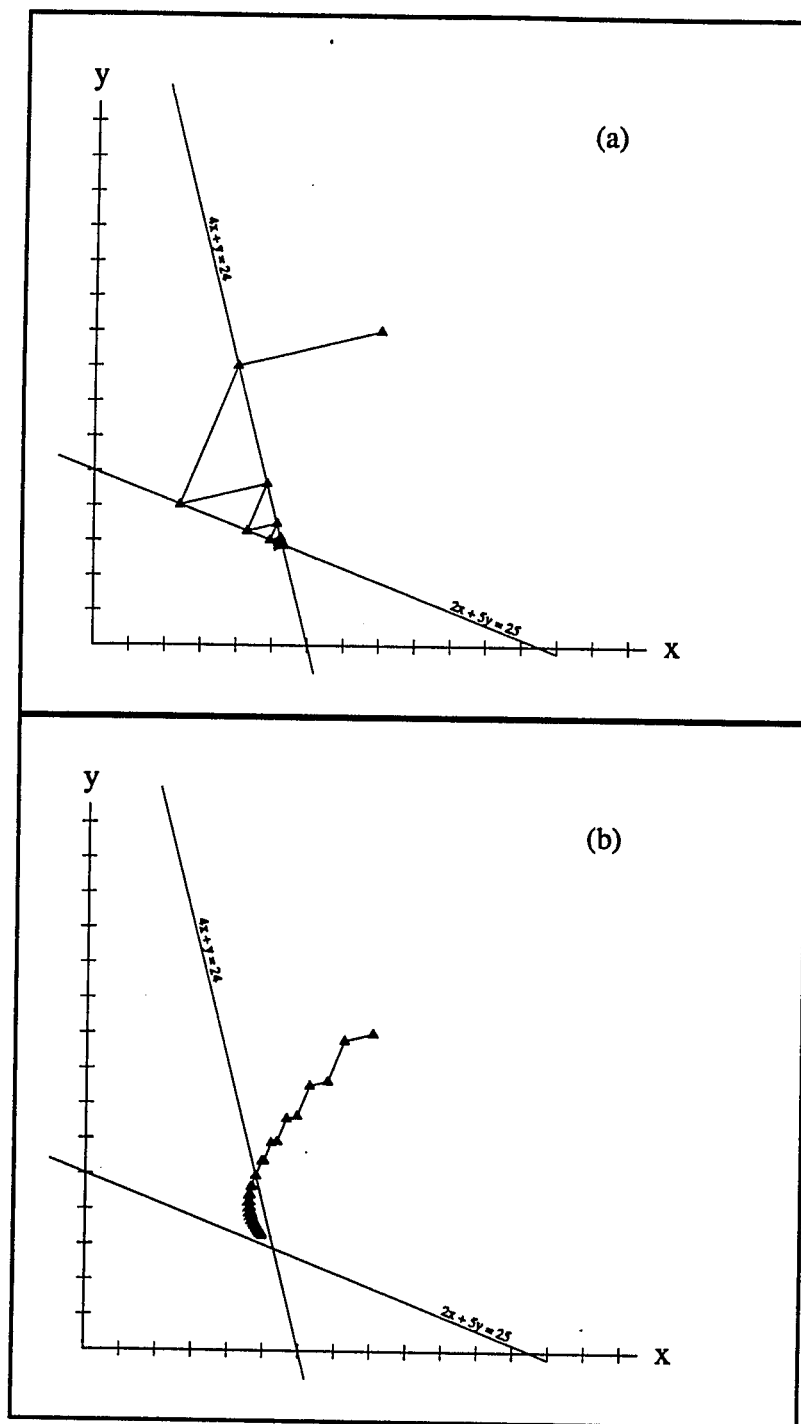
Figure 2.1: (*a*) Solution of 2 equations in 2 unknowns using ART with no relaxation. This is a consistent set of equations and the exact solution is found after 8 iterations. (*b*) Same as (*a*) except a relaxation of 0.2 has been incorporated. The path towards the solution is much smoother.
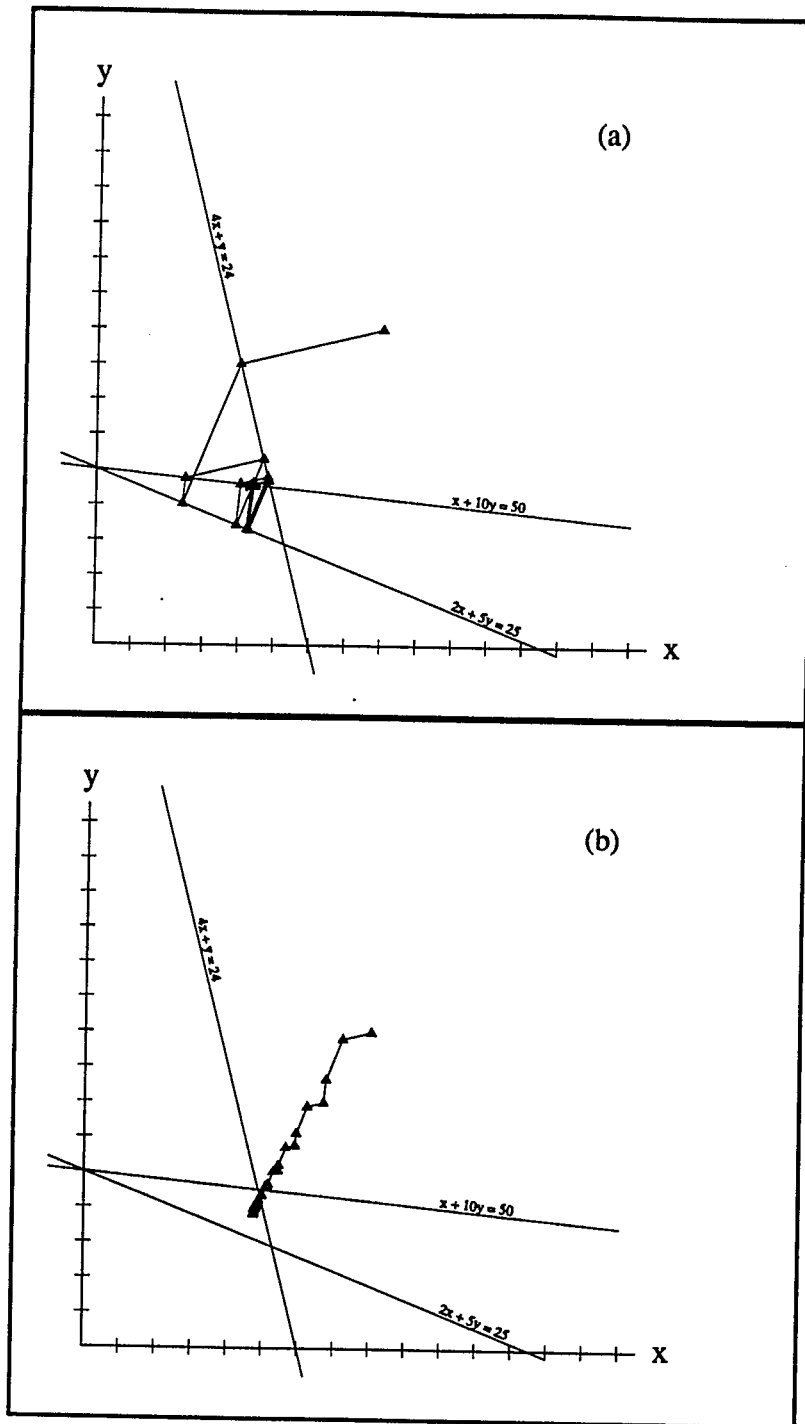
Figure 2.2: (a) For 3 equations and 2 unknowns the system is inconsistent and the solution oscilates between the two lines that are nearly parallel. (b) After relaxation is applied the solution is stabalized and converges to a point near the assumed intersection of the three equations.
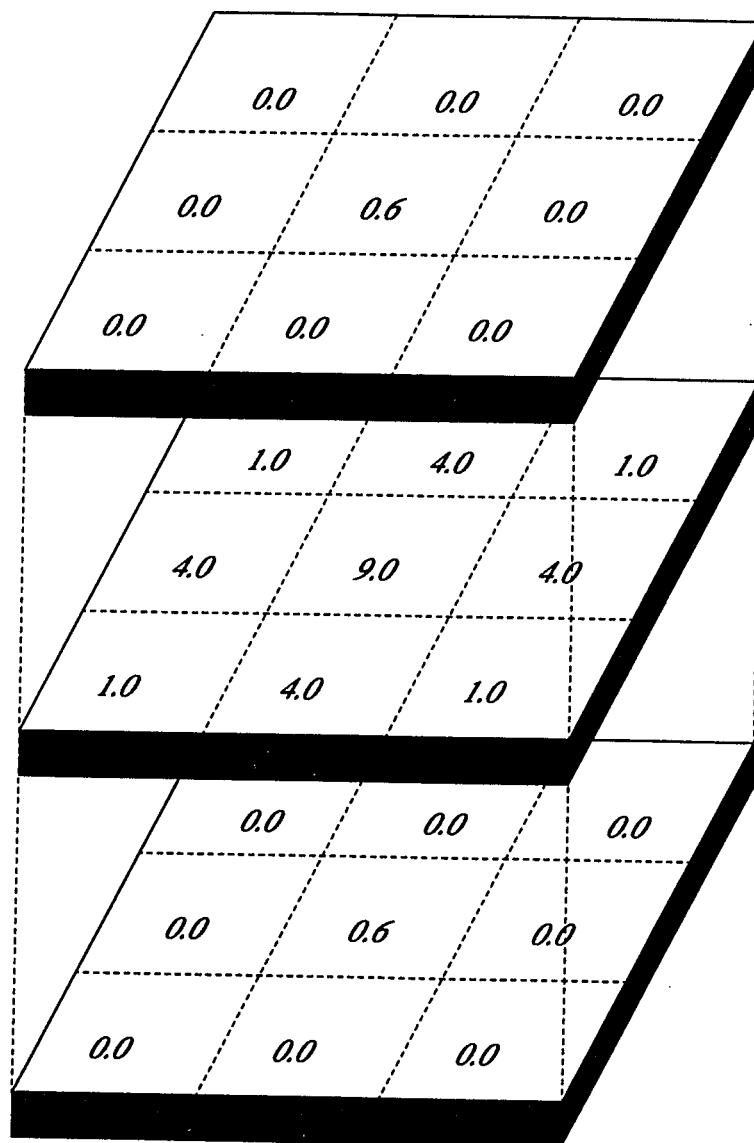
Figure 2.3: Three dimensional low pass filter. Filters of this kind are used in the ART inversion between iterations to introduce regularization and impose *a priori* knowledge of the structure in the solution.
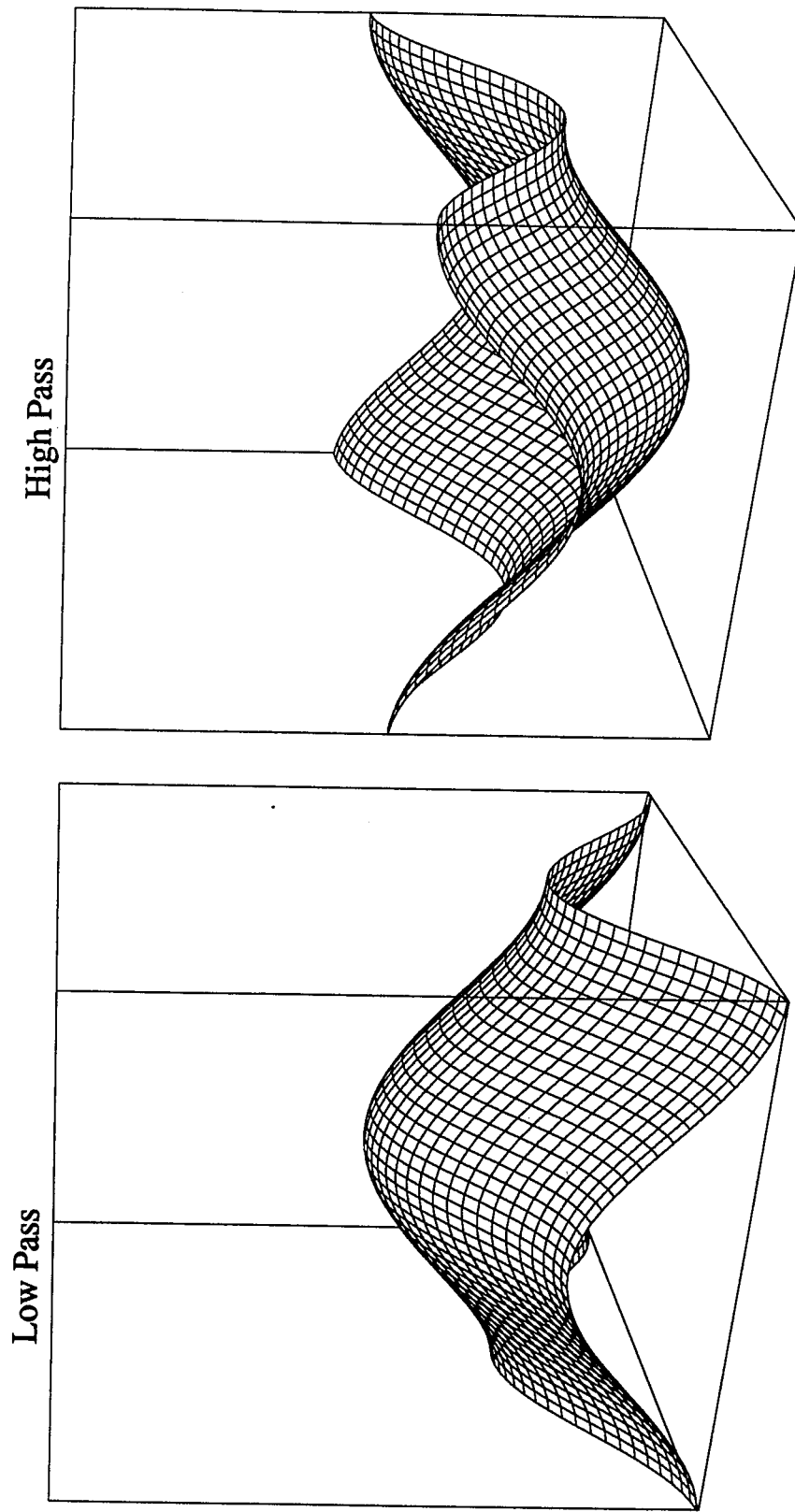
High Pass

Low Pass

Figure 2.4: Wavenumber domain response for low pass filter [4,1,0] and high pass Laplacian filter [-4,-1,0]

CHAPTER 3
SYNTHETIC EXAMPLES

SYNTHETIC MODEL

To illustrate the implementation of the inversion techniques in a tomographic setting that simulates a real situation, we construct a model and a synthetic set of raypaths distributed over the model. The travel times are calculated by summing the velocity through the blocks the raypath has sampled. Noise is added to reflect the uncertainties that are expected in real situations and the data set is used as input to the inversion programs.

For simplicity, we will use a 2-dimensional synthetic model, commonly called a phantom, illustrated in Figure 3.1. It consists of a background slowness with two major perturbation objects: a cross consisting of a positive anomaly (+4.0%) and a torus with a negative anomaly (-4.0%) whose center is a +1.0% perturbation from the background. The phantom is parameterized by dividing it into 1600 equal area blocks, 40 per side, where each block is 2 by 2 km. For synthetic station locations we randomly distribute stations throughout the target and these remain fixed throughout as in Figure 3.2. To simulate a situation where earthquakes are distributed in the earth, we likewise randomly generated 236 hypocenter locations (Figure 3.3). A random size for the event was generated and then a portion of all the stations that fell within a specific radius of the event were considered triggered stations and a datum for each was generated. This resulted in 3000 hypocenter-station pairs, or raypaths. Since this is a 2-dimensional model the rays are straight line segments from source to receivers. The total ray coverage is illustrated in Figure 3.4. The raycoverage pattern produced in this manner reflects the characteristics of the real data coverage.

Each datum consists of the sum of the slowness perturbations along the ray-path weighted by the distance each ray spent in the sampled blocks. To these values we introduce noise. Thus the input data become

$$\delta t_i = \sum_{ray_i} \delta s_k (dr)_k + \eta_i$$

where η is the noise component. In this synthetic example the noise is a random gaussian variable with RMS level 80% of the RMS residual. The noise has possibly three components: 1) random, independent reading error due to mispicking of first arrivals, 2) noise introduced due to mislocation of the event which may be correlated for each event, and 3) bias introduced from error in determination of the appropriate station correction. In the following examples we have introduced the first kind of noise in all the examples and in several cases we have also incorporated noise simulating the second and third types. Each component of noise has a gaussian distribution with varying standard deviations.

Several experiments and comparisons between ART and LSQR were performed. For reference these are listed in Table 3.1. In each example a gray shade plot showing the percent perturbation from the background model, whose slowness is set at the arbitrary value of 0.18519 sec/km (slowness of the first layer of the real model), is presented. Supplementing the gray shade plots, a cross section of the values of the phantom and the reconstructions along a line marked A-A' shown explicitly on Figure 3.1 (all subsequent reconstructions show just the points A-A') is displayed. This allows for a block by block comparison of inversion results to the true phantom.

It is sometimes useful to provide an overall quantitative measure of distance between the phantom and a reconstructed image. Following *Herman* [1980] we calculate three measures of distance for each inversion:

$$d_1 = \left[\frac{\sum\limits_{i=1}^{M} (s_i - \tilde{s}_i)^2}{\sum\limits_{i=1}^{M} (\tilde{s}_i - \overline{s})^2} \right]^{1/2}$$

$$d_2 = \frac{\sum\limits_{i=1}^{M} |s_i - \tilde{s}_i|}{\sum\limits_{i=1}^{M} |\tilde{s}_i|}$$

$$d_3 = \max |s_i - \tilde{s}_i|$$

where $\tilde{s}$ is the reconstructed image, s is the phantom and $\overline{s}$ is the arithmetic mean of $\tilde{s}$. The measure d_1 is the root mean square normalized distance which will be large for images that have a few blocks that differ by a large amount from the phantom. On the other hand, d_2 is the average absolute value difference which emphasizes the importance of many small errors distributed over the whole model. Finally d_3 is a measure of the worst case. The distance summaries for all inversions are reported in Table 3.2. A cautionary note: These quantitative comparisons should be used primarily as a guide: visual pattern recognition may be more important for determining the quality of one inversion over another. To monitor the convergence behavior we have plotted the sum of the squared weighted prediction error,

$$\chi^2 \equiv \sum_{i=1}^{n} [w_i (t_i - \sum_j a_{ij} s_j)]^2 \tag{3.1}$$

versus iteration number, k.

Comparison

As a first comparison we show the results of inversion (labeled ART1 and LSQR1) when no noise is added to the data. In this case we need neither regularization nor relaxation. Note in Figure 3.5(a) and (b) that the reconstruction is nearly perfect for both ART1 and LSQR1, but ART1 has outperformed LSQR1 (for a fixed number of iterations) in the lower right corner where the data coverage is more sparse. All three distance measures of Table 3.2 also indicate that ART1 is a better reconstruction than LSQR1. In Figure 3.6(a) and (b) noise is introduced and damping is applied but no smoothness constraints are implemented (ART2 and LSQR2). Here, using the same damping, the resulting images are identical. The characteristic "salt and pepper" degradation appears as noise from the data is projected into the image but the phantom is still evident beneath the noise.

Next, smoothing is implemented by constraining the Laplacian in the LSQR3 inversion and by applying a lowpass filter between iterations in ART3 (Figure 3.7(a) and (b)). In the ART3 case a relaxation of $\rho^{(k)} = 0.02$ is used with regularization parameter adjusted to $\lambda = 65$. Here, even though the solutions predict the data to the same χ^2 accuracy, they appear slightly different, due to the different way the smoothing is performed in each method. Table 3.2 indicates that ART3 is over all slightly closer to the phantom than LSQR3, but has a higher worst case measure (d_3). If we apply this same filter but keep $\rho^{(k)} = 1.0$ with $\lambda = 400.0$ as in the previous example, the technique does not converge in 30 iterations. After the first few iterations, the smoothing dominates the inversion process, producing successively smoother models which converges to a different χ^2. Figure 3.8(a) shows ART4 after 30 iterations. Notice in Table 3.2 that the d_1 measure is smaller for ART4 than ART3 but d_2 is larger. In ART5, the relaxation parameter is allowed to decrease between iterations with a suitable reduction in smoothing. This results

in an image (Figure 3.8(b)) very close to that of ART3 in Figure 3.7(a) and is indicated as such in Table 3.2.

In Figure 3.9 present a plot of the $\%\chi^2$ reduction versus iteration number for each of the synthetic inversions is presented. For the case of perfect data, ART1 reduces the χ^2 faster than LSQR1 until the 8–th iteration where the respective χ^2 converges to nearly 100% reduction. In the damped case ART2 out performs LSQR2 until 4 iterations when the two methods converge. When smoothing is applied in ART3 and constraining in LSQR3 the situation is reversed. Here LSQR3 appears to have reduced the χ^2 faster for the first 3 iterations where the χ^2 for each merge. The χ^2 reduction for ART4, however, is better than that of LSQR3 for the first 3 iterations, after which the smoothing degrades the data predictive nature of the solution. If the relaxation is allowed to decay with appropriate decay of smoothing, the χ^2 of ART5 is made to follow that of ART3.

Figure 3.10(a) and (b) displays the standard error for the inversions in ART3 and LSQR3 of Figure 3.7 respectively. Since the methods of regularization are different in each case, the mini-inversions and pseudo values calculated will vary accordingly giving rise to different distributions of standard errors. Overall, though, the errors are of the same magnitude. In both cases the estimated errors are generally larger in the lower right corner where coverage is sparse.

Robust Example

To illustrate the effect of incorporating a robust approach in dealing with outlier residuals, the original data set is extended by adding in a set of 500 equations whose errors are large compared to the original artificial data such that the distribution of residuals has heavy tails compared to a normal distribution. The data set is inverted with residual weighting as described in equation (2.3.12) where $r_1 = 1.0$ sec and $r_2 = 2$ sec. The results are shown in Figure (3.11) where the first

case is with no residual weighting and the second incorporates the above scheme. Notice in Figure 3.11(*a*), where no robustness is used, artifacts from data with large residuals appears in the upper left corner with a 12.3% anomaly. By implementing a robust inversion we have eliminated these artifacts at the price of not having a strictly least squares inversion.

Debundling Example

As a last example, the value of debundling is illustrated. Here 500 additional rays were added to the data set. The epicenters of each were constrained to lie within a small area of 4×4 km, and all were recorded at a single station, as shown in the ray diagram in Figure 3.12. A correlated additional component of noise was added to each of these rays and they were then mixed in randomly with the original data. Algorithm LSQR is used to invert the data with and without special weighting applied to remove the effects of large bundles. When no "debundling" has been applied, we see the inversion has a prominent linear feature along the path of the large bundle (Figure 3.13(*a*)). The averaging introduced by debundling has improved the situation by suppressing the effects of the correlated noise in the bundle (Figure 3.13(*b*)).
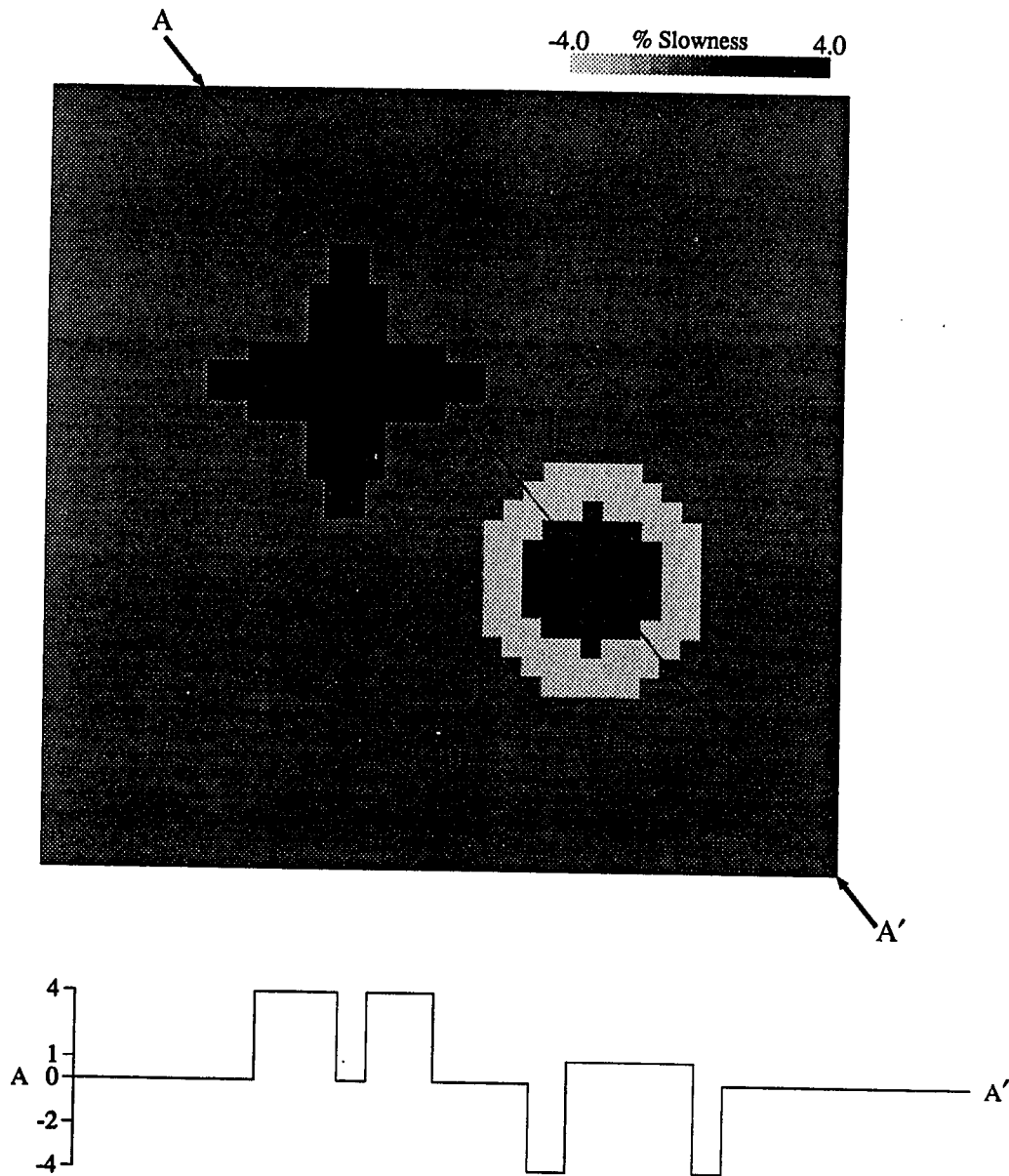
Figure 3.1: Synthetic phantom model. The model consists of a positive 4.0% cross anomaly and a negative -4.0% torus anomaly. Cross section A-A′ is provided to aid in comparing reconstructions.
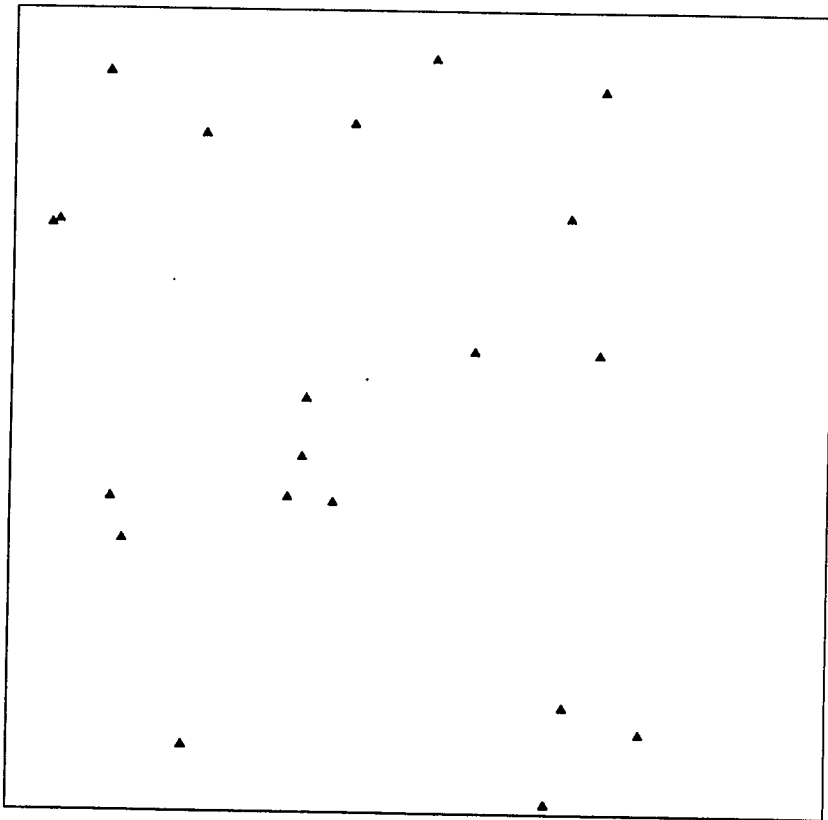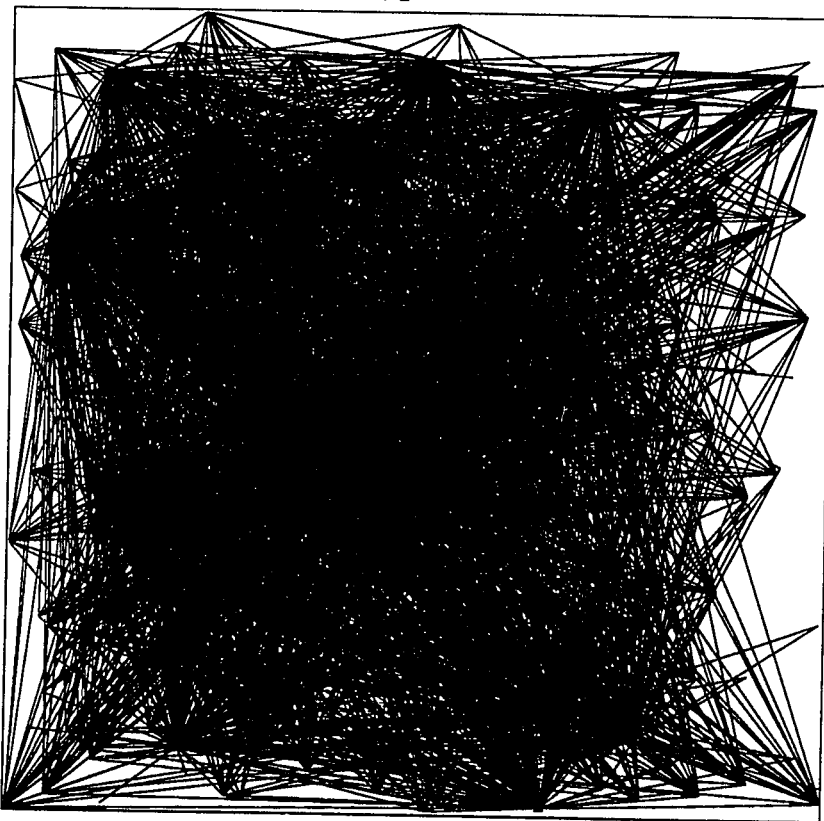
Figure 3.2: Synthetic station distribution. 20 stations are located randomly over the phantom.

Synthetic Source Distribution

0 ⌊⌁⌁⌁⌁⌁⌁⌁⌁⌁⌁⌋ 20.0 km
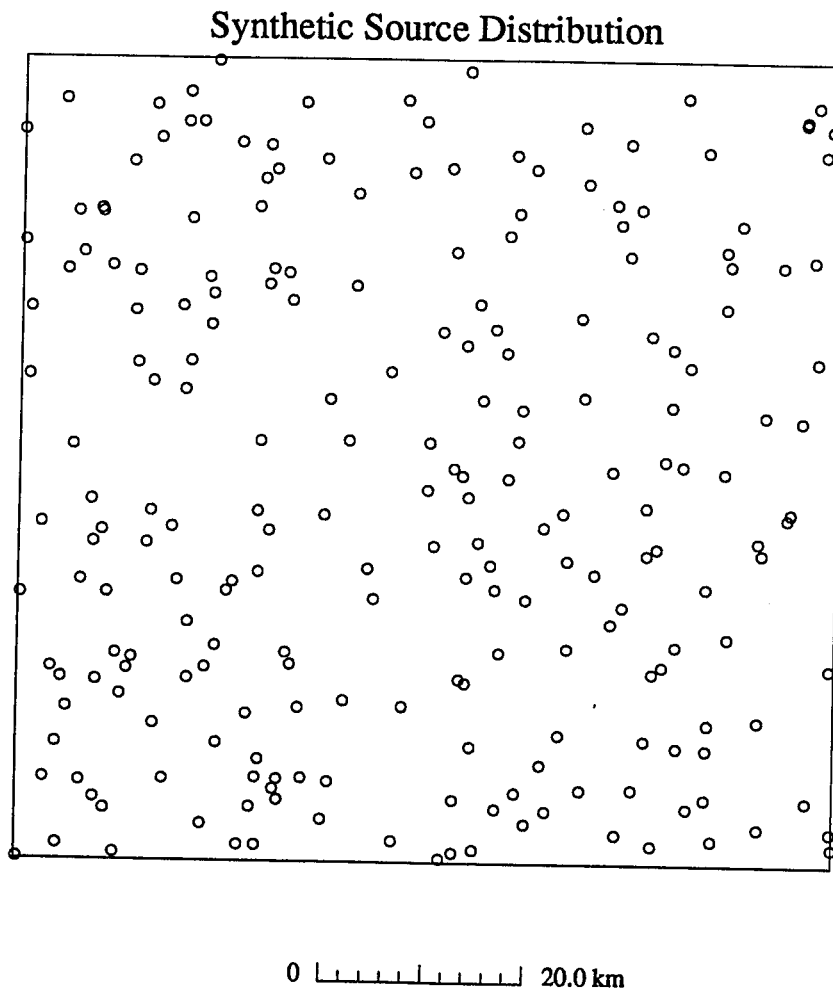
Figure 3.3: Synthetic source distribution. 236 sources are located randomly over the phantom.

Synthetic Raypath Distribution

Figure 3.4: Synthetic raypath distribution. To simulate earthquake data sources and stations are connected by choosing only a select number of stations that fall within a certain radius of the source. 3000 raypaths were generated and used in inversion experiments.
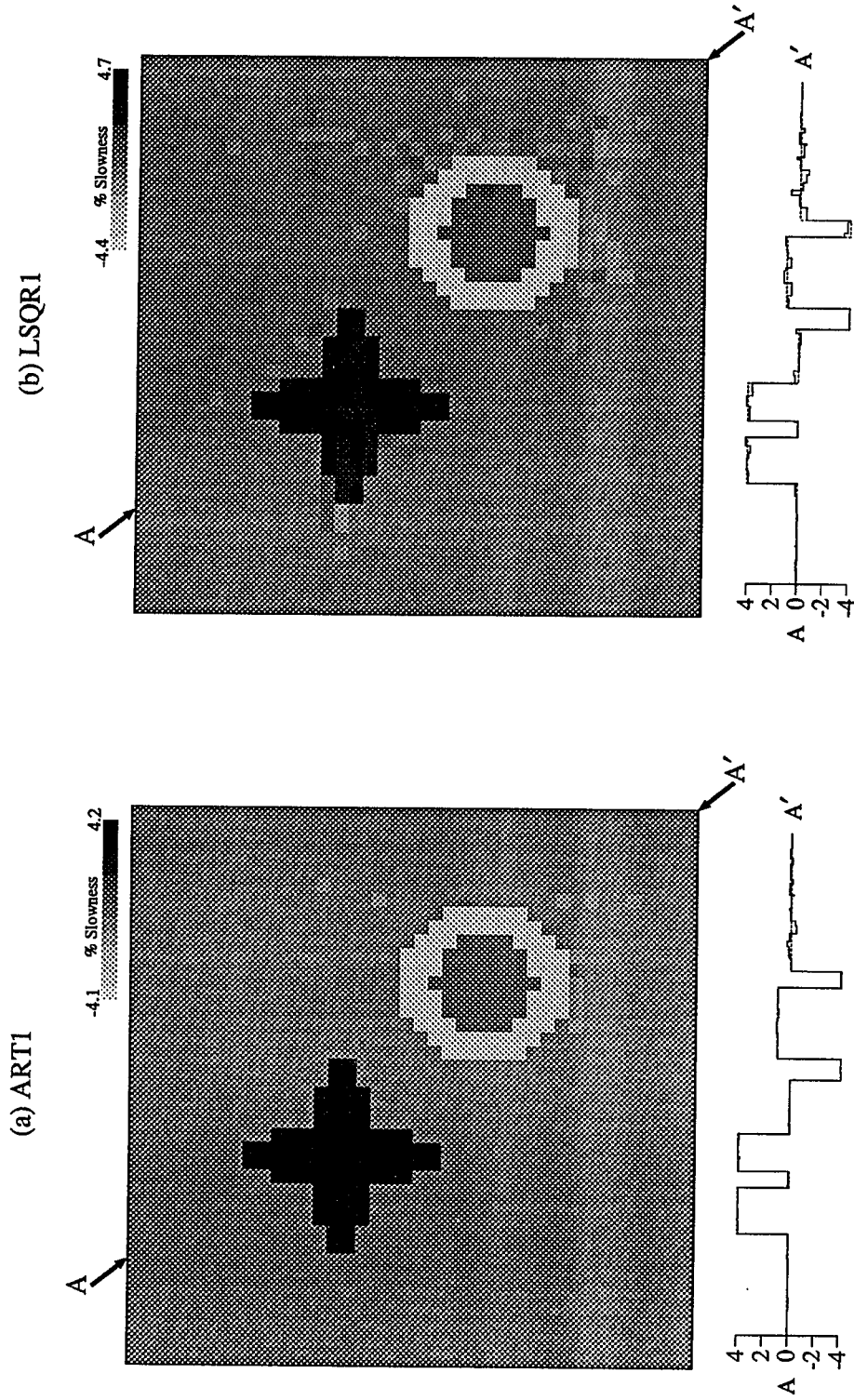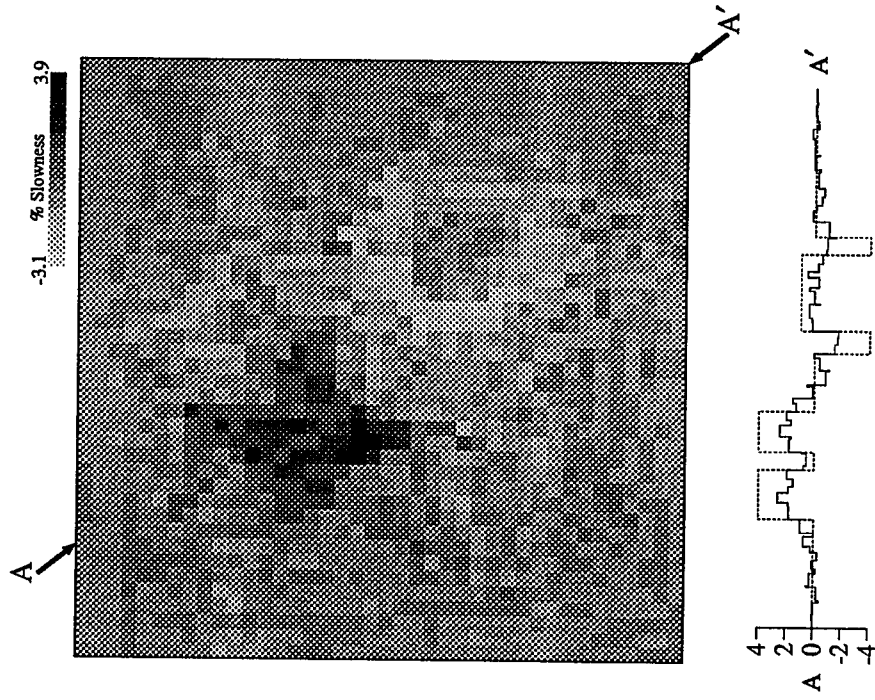
(a) ART1

(b) LSQR1

Figure 3.5: Synthetic inversions with perfect data. Both ART (*a*) and LSQR (*b*) were allowed 30 iteratations. When no noise is present reconstructions are nearly perfect with this ray coverage.
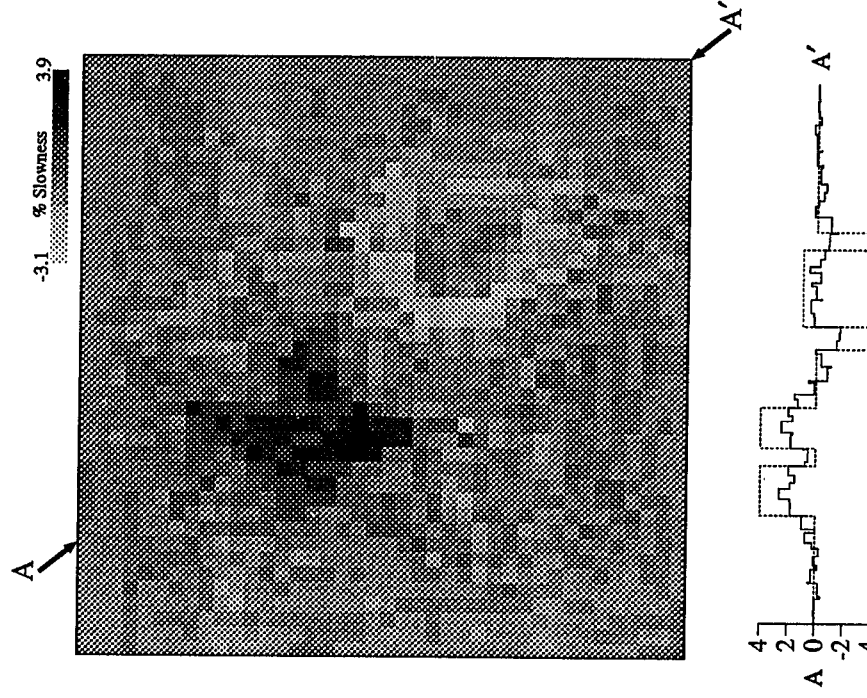
(a) ART2

(b) LSQR2

Figure 3.6: Synthetic inversions with noisy data and damping. Both techniques result in identical reconstructions.
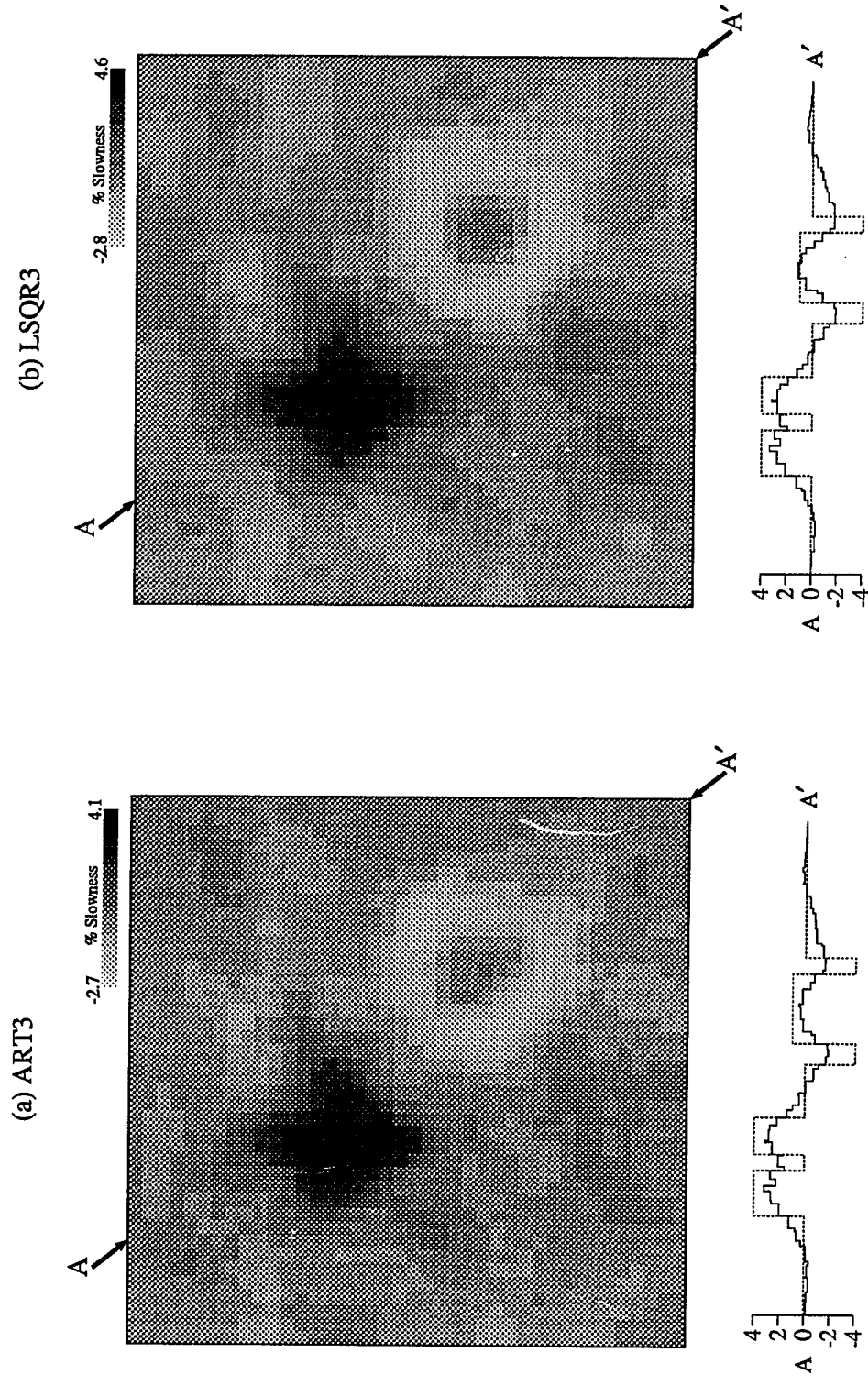
Figure 3.7: Synthetic inversions with regularization. ART (a) has been smoothed between iterations and LSQR (b) has been constrained to minimize the Laplacian.
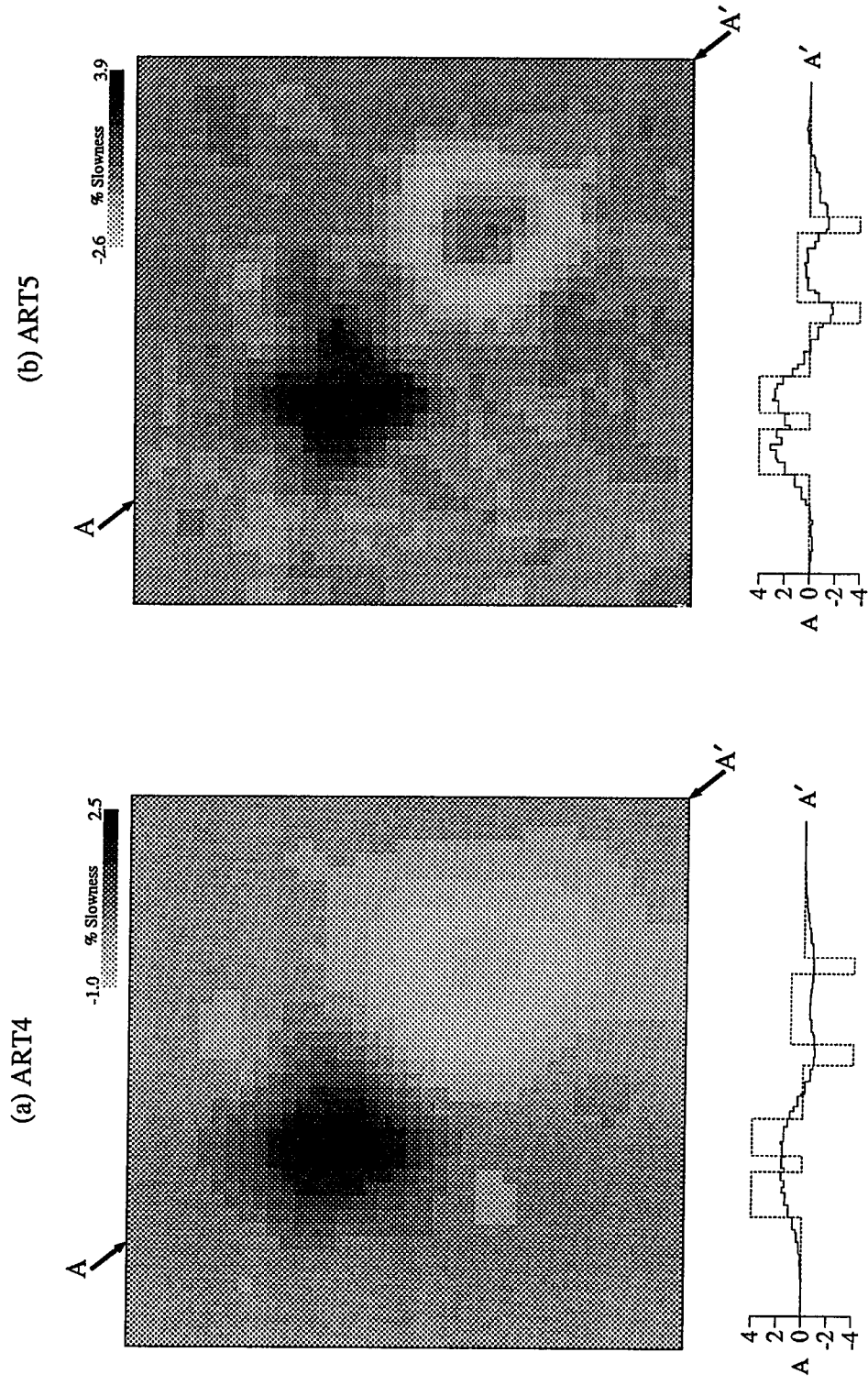
Figure 3.8: Synthetic inversions with ART. ART4 (*a*) been smoothed with no reduction of relaxation whereas ART5 (*b*) is smoothed with a corresponding reduction in relaxation at each iteration.

Residual Reduction

Legend

———————— ART1:No-Noise

················· LSQR1:No-Noise

––––––––. ART2:Noise,damping

––––––––– LSQR2:Noise,damping

–··–··–··. ART3:Noise,smoothed

–···–····· LSQR3:Noise,constra

– – – . ART4:7a

–··–··· ART5:7b
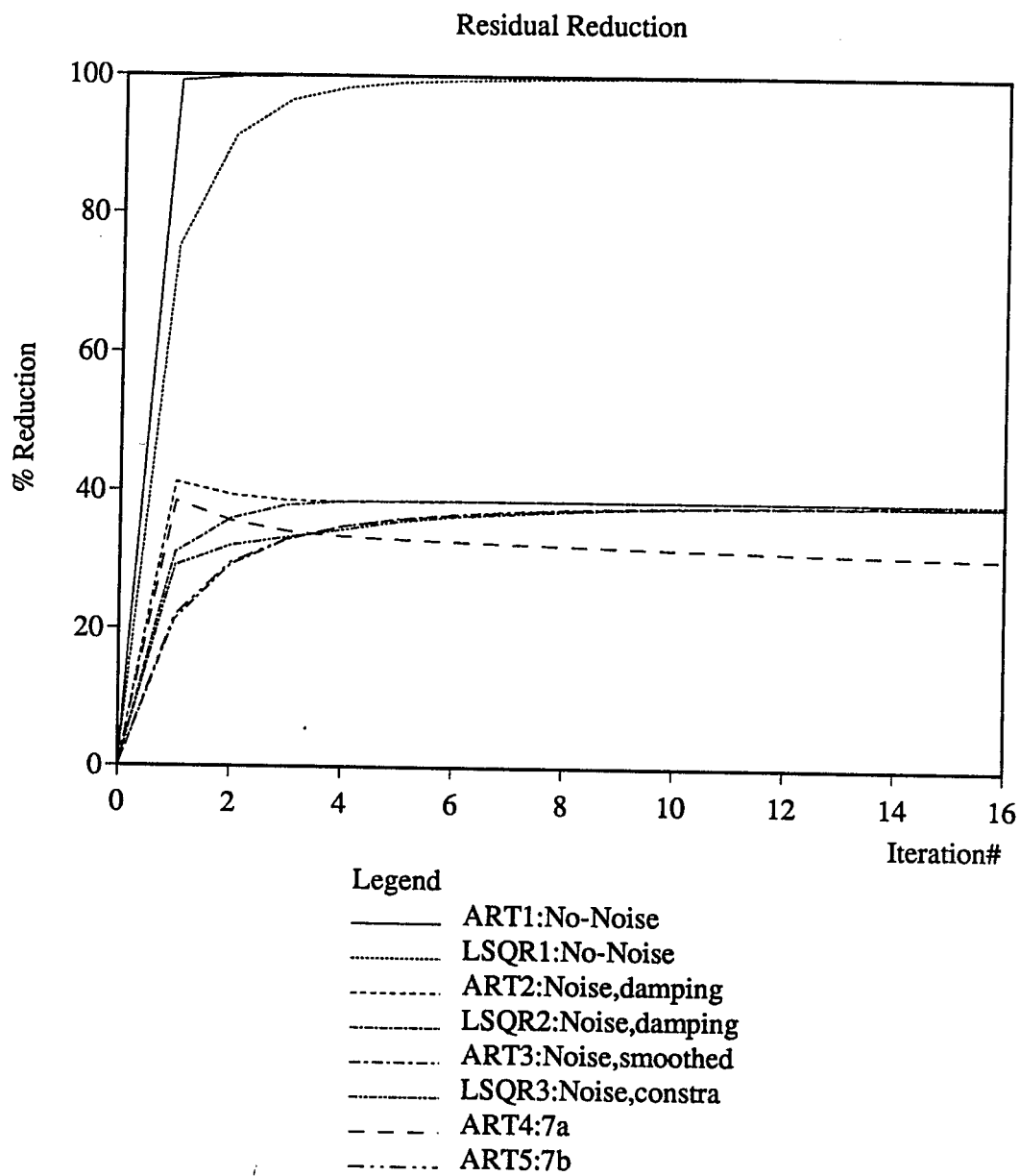
Figure 3.9: Reduction of χ^2 misfit for synthetic inversions. The vertical axis is in % reduction from the initial χ^2.
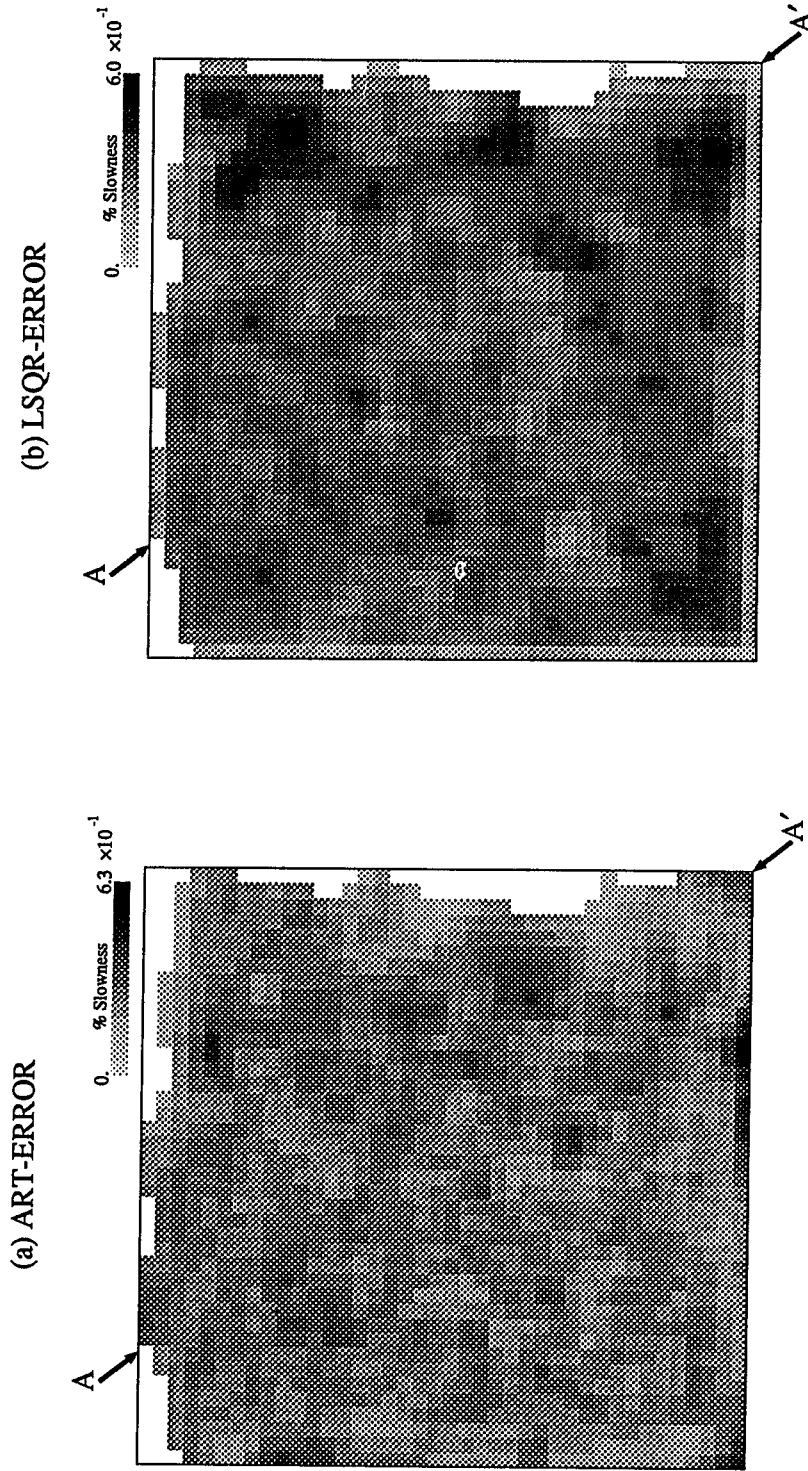
Figure 3.10: Error estimates of model parameters using the jackknife for ART and LSQR. The jackknife was performed with smoothing as in examples ART3 and LSQR3.

(a) Non-Robust (b) Robust
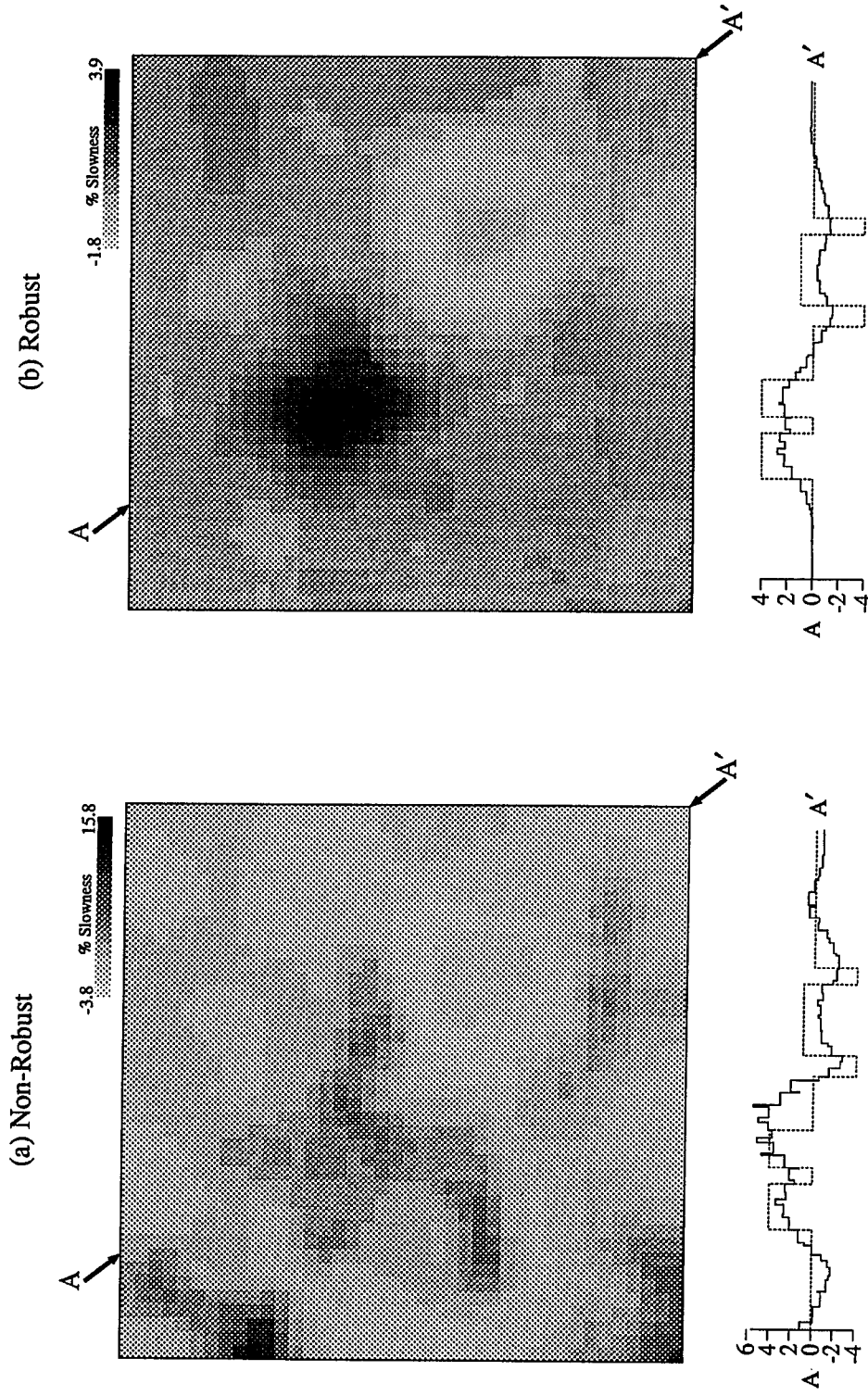
Figure 3.11: Example of the effect of robust inversion. 200 additional noisy rays have been added to the data set with large amounts of error. (*a*) No special consideration was taken to guard against outliers. (*b*) Robust inversion where large residuals were down weighted.
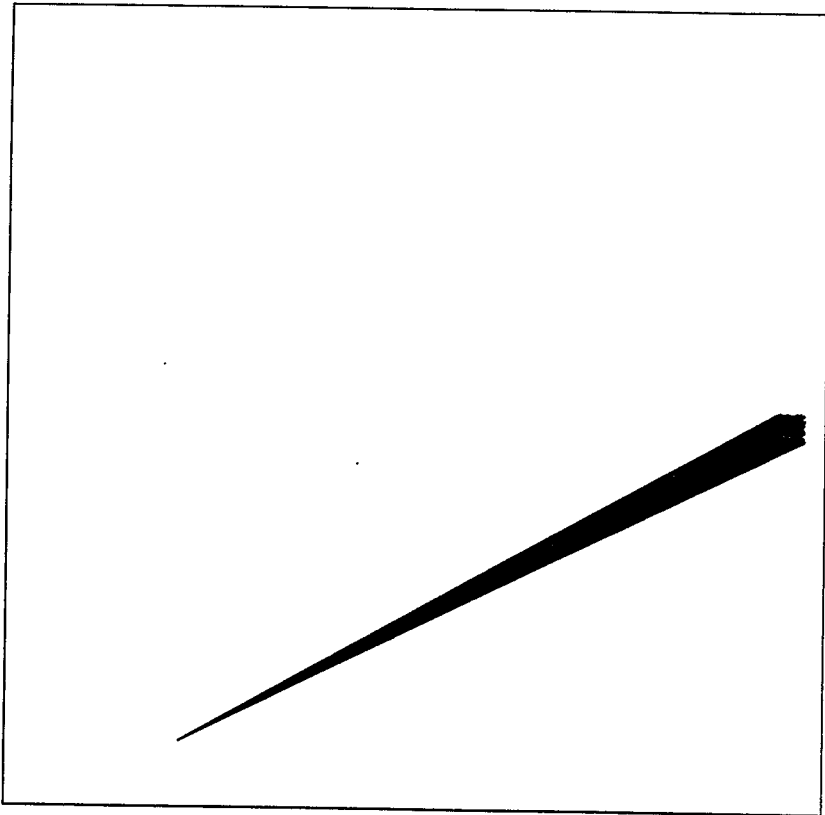
Figure 3.12: Ray bundle added to data to illustrate the effects of debundling (500 rays).

(a) LSQR no debundling
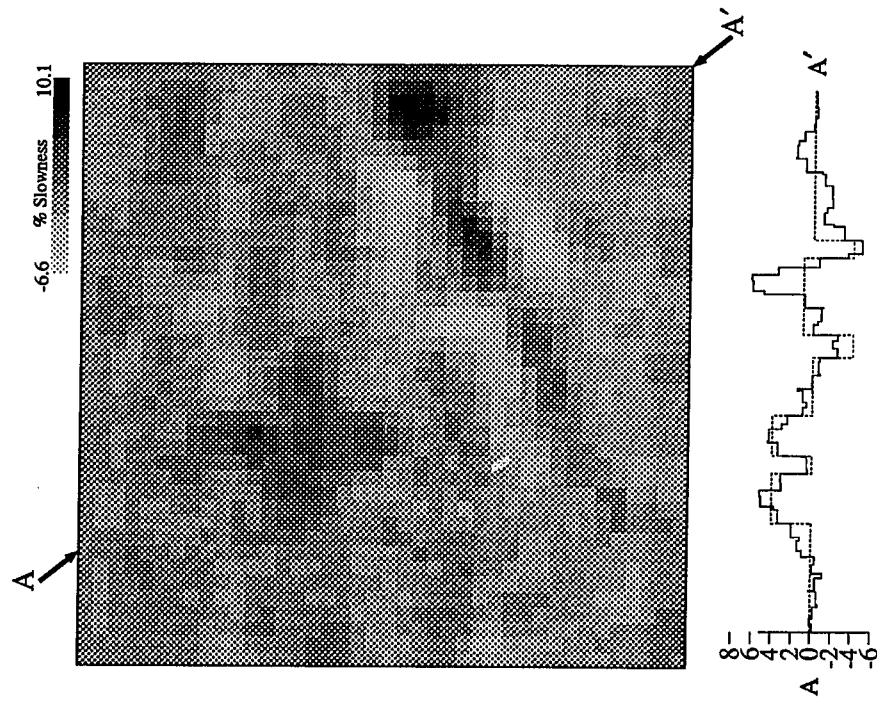
(b) LSQR with debundling
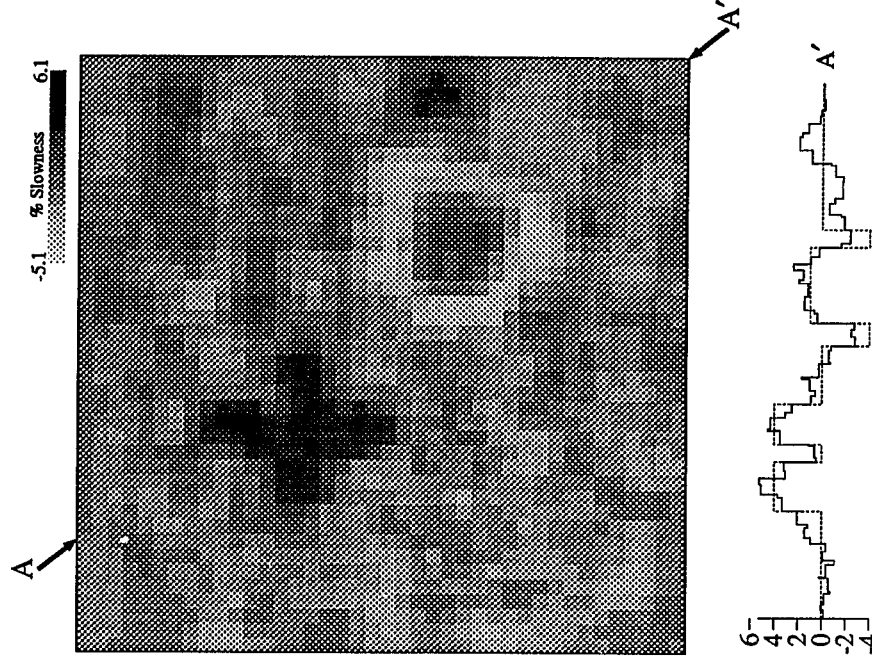
Figure 3.13: Example illustrating the effects of debundling. (*a*) No debundling has been applied. (*b*) The rays inside the bundle have been down weighted to equalize their influence on the inversion.

Table 3.1: Summary of tests performed on synthetic inversions.

Model	Process
ART1	Perfect data, no relaxation, no smoothing
LSQR1	Perfect data, no relaxation, no smoothing
ART2	Noisy data, damped least squares
LSQR2	Noisy data, damped least squares
ART3	Noisy data, relaxed and smoothed
LSQR3	Noisy data, Laplacian constrained
ART4	Noisy data, relaxation reduced, smoothing constant
ART5	Noisy data, relaxation reduced with smoothing reduced
Robust1	Extra Noisy data, no compensation for this
Robust2	Extra noisy data, robust inversion
Debund1	Bundle of rays introduced, no compensation
Debund2	Bundle of rays introduced, debundling applied

Table 3.2: Distance measures of inversion images to phantom model. Here d_1 is the root mean square normalized distance, d_2 is the average absolute value difference, and d_3 is a measure of worst case.

Distance Measures to Phantom			
Model	d_1	d_2	d_3
ART1	0.816994	0.065758	0.001112
LSQR1	1.938234	0.144696	0.002443
ART2	9.232765	0.691058	0.007067
LSQR2	9.232765	0.691058	0.007067
ART3	8.686169	0.613092	0.006573
LSQR3	8.985689	0.615608	0.006460
ART4	8.234601	0.752227	0.006503
ART5	8.353343	0.608981	0.006614
Robust1	21.6796	1.50888	2.27742E-02
Robust2	9.04130	0.674702	6.06702E-03
Debund1	18.8831	1.24606	1.86905E-02
Debund2	14.6347	0.841030	7.85360E-03

CHAPTER 4

RESOLUTION

Resolution

The resolution of a method of inversion is a way of quantifying the ability to distinguish a part of a model from its surrounding. Suppose there exists a slowness field s^* which can be sampled at a discrete set of loci (blocks, for example) to form a discretized version s. The data are related to the discrete model by

$$As = t .$$ (4.1.1)

If there exists a generalized inverse A^t of A then an estimate of the model is given by

$$\hat{s} = A^t t$$ (4.1.2)

and substituting from (4.1.1) gives

$$\hat{s} = A^t As$$ (4.1.3)

Thus $\hat{s}$ can be thought of as a version of s that has been modified by the operator $A^t A$. In Backus-Gilbert theory one says that $\hat{s}$ is the discretized truth s seen through the window (filter) $A^t A$ [*Backus and Gilbert*, 1968; *Jackson*, 1972]. The $M \times M$ matrix $R = A^t A$, called the resolution matrix, describes how each element of the model $\hat{s}$ is related to the elements of s. If $R = I$ then we say we have perfect resolution implying $\hat{s} = s$. Since we have a finite amount of data our resolution will always be less than perfect. The *i–th* row of R can be thought of as a weighting function where the coefficient for the *j–th* column indicates the influence the *j–th* block has in determining the *i–th* parameter. The *i–th* row of R is called the resolution kernel for that model parameter and if the parameters have a natural ordering then plotting of the rows of R conveys information about the

resolving power of the inversion system. The closer the plots of the resolution kernels resemble delta functions (spikes) the better the resolution in the vicinity of the given block.

In the case where the matrix A is very large computing R directly is virtually impossible. This is because the matrix A is often too large to store in the computer, and calculation of A^t explicitly would require inordinate amounts of time making making resolution analysis unfeasible. The inversion techniques proposed in this study are iterative by nature and are useful particularly because they are space and time efficient. For this reason we are forced to compromise and present an alternative to a full computation of the resolution matrix. This approach entails approximating the resolution for a few selected locations where resolution information is desired. It is reasonable to assume that if the raycoverage is fairly homogeneous over a region the resolution kernels for the adjacent blocks in the region will be nearly similar [*Humphreys and Clayton*, 1988].

Since the resolution depends heavily on the spatial distribution of the ray coverage, and we do not have access to the resolution matrix, it is valuable to display the ray coverage as an aid in determining qualitatively the extent of resolution. Regions that have rays traversing in narrow bands will have poor resolution along the direction of the rays. Similarly, sparsely sampled regions are likely to have poor resolution compared to regions of heavy, isotropic coverage. Inspection of ray coverage diagrams help in determining some of these factors. The ray coverage diagram for the synthetic data set was presented in Figure 3.3. Notice that where the coverage is sparse (as in the lower left corner) it is easy to see individual rays and note anisotropies, but where the coverage is dense it is impossible to see what the distribution is.

To determine a quantitative measure of the distribution of coverage through particular cells while allowing us to easily survey a large number of blocks we

have adopted a method of *Kissling et al* [1984]. In this approach we calculate the moment of inertia tensor for all the rays traversing all the blocks of the model [*Mardia*, 1972]. For *j-th* ray, the moment of inertia tensor in the *k-th* block is

$$\mu_{jk} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \times \begin{bmatrix} x,y,z \end{bmatrix} = \begin{bmatrix} xx & xy & xz \\ xy & yy & yz \\ zx & zy & zz \end{bmatrix}$$

where (x,y,z) corresponds to the projected length of that portion of the *j-th* ray that transects the *k-th* block onto the each corresponding coordinate axis. The total moment of inertia tensor for the *k-th* block is the sum over all the rays in the data set,

$$\mu_k = \sum_{j=1}^{N} \mu_{jk} \, .$$

For each moment tensor the eigenvalues and eigenvectors are calculated, which determine a moment of inertia ellipsoid. The shape and orientation of the ellipsoid (Figure 4.2) provides information on the average directionality of rays traversing a block. Ellipsoids that are nearly spherical imply isotropic coverage in all directions and conversely, elongated ellipsoids indicate a preferred ray direction. However, representing the 3-dimensional ellipsoids on flat paper entails loss of information if our plots are to remain simple enough to be useful. To convey information about the relative sizes of the three eigenvalues, two orthogonal ellipses, corresponding to the eigenvectors, are plotted by rotating the axis such that the ellipses are parallel with the horizontal plane. The direction of the major axis relative to north and east is conveyed by orienting the largest axis along the horizontal azimuth (angle β in Figure 4.2). The dip of the largest eigenvector is plotted on an adjoining plot where the direction of dip (angle γ from Figure 4.2) for each ellipsoid is represented by a slanting line whose length is proportional to $\cos(\gamma)$. In addition the number of rays penetrating a block is presented as a circle

whose size is proportional to the logarithm of hit counts for each block. For the synthetic data set used in Chapter 3 the ellipsoids are presented in Figure 4.3. (Here all the dip directions in Figure 4.3(b) are zero because the data are 2-dimensional.)
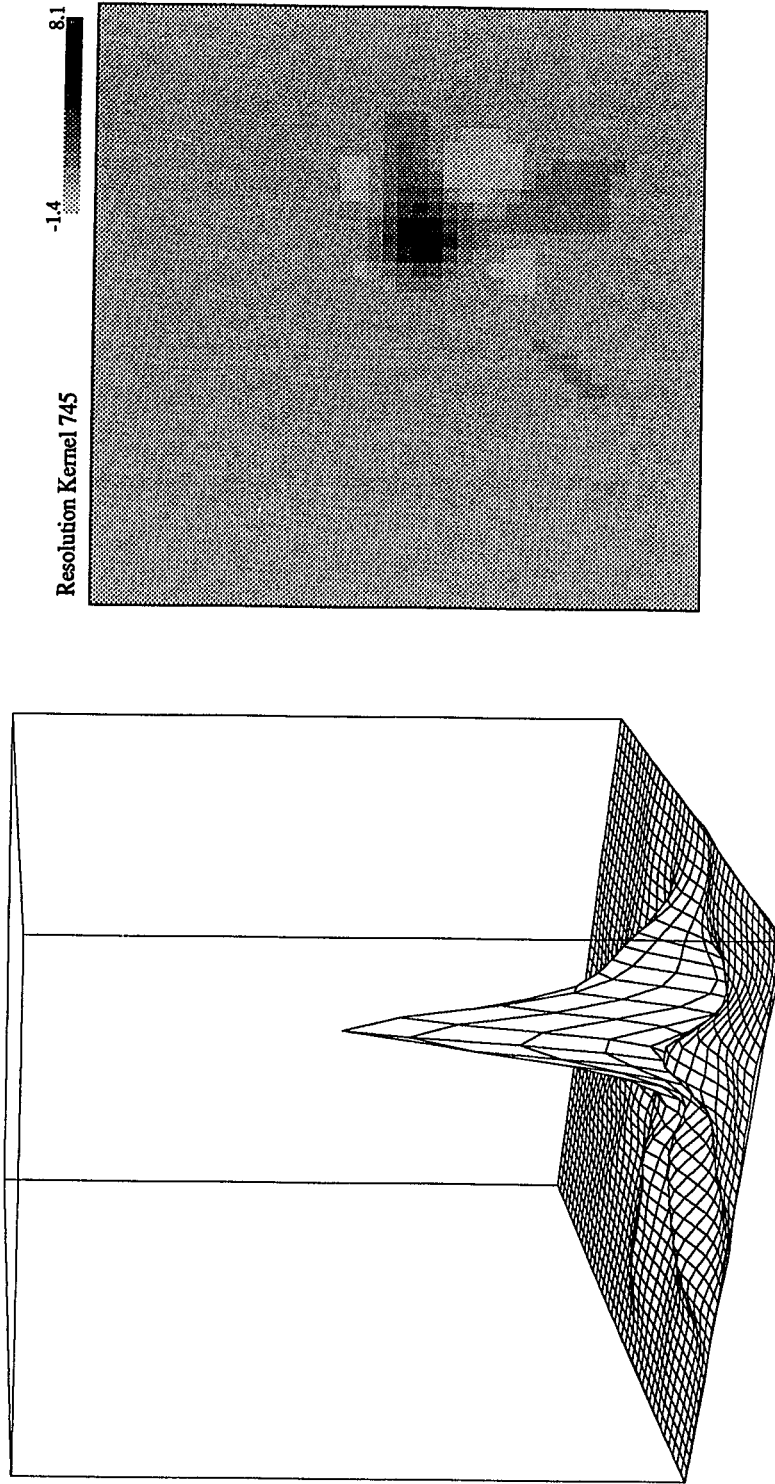
Figure 4.1: Resolution kernel for synthetic data. The left is a perspective plot and the right is the gray shade representation of the impulse response of block 745.
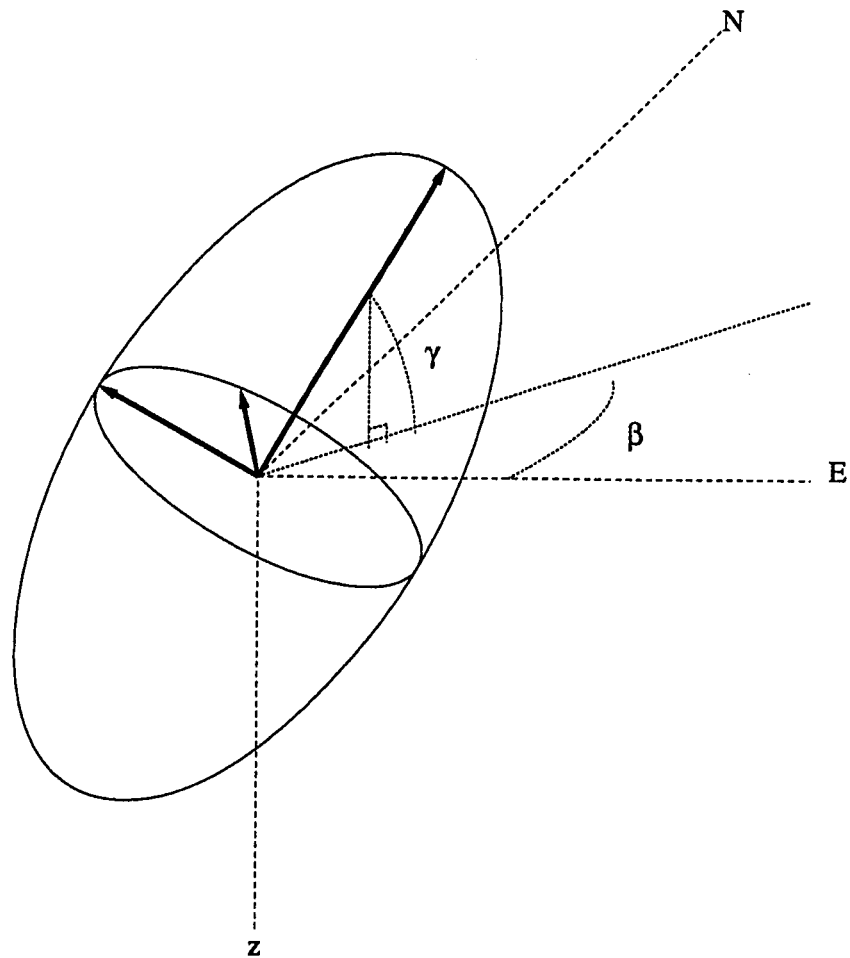
Figure 4.2: Sample ray ellipsoid in 3-D. The three arrows represent the eigenvectors for the moment of inertia ellipsoid of all rays passing through a given box. β is the azimuth of the major axis in the horizontal x-y plane and γ is the angle of dip with respect to the horizontal. In the ellipsoid figures the major ellips and minor ellipse are rotated and plotted flat so their relative sizes can be compared. The dip angles are presented on separate plots.
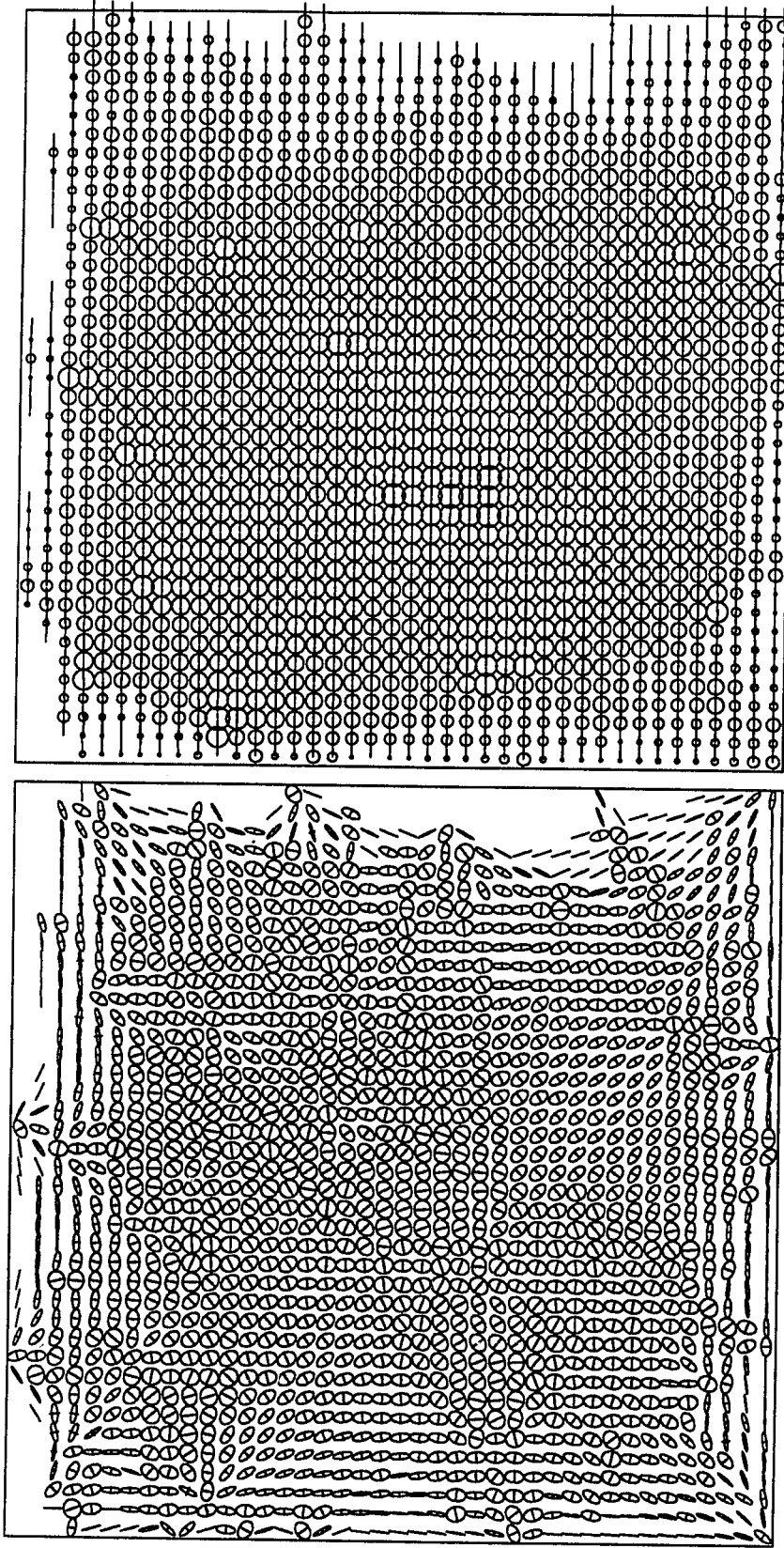
Figure 4.3: Ray ellipsoids for synthetic data. The left displays the relative sizes of the eigenvalues and the azimuthal direction and the right displays the dip and the relative hit count for each block.

CHAPTER 5

ERROR ANALYSIS

THEORY - THE JACKKNIFE

In the estimated model, errors arise from inaccurate measurements, incomplete coverage of rays over the target area, over simplified parameterization and mislocation of the earthquakes. In classical least squares discussed above, one can show that an estimate for the error in the model is found in the covariance matrix, $cov(s) = ss^T$. If the estimated variances of the weighted observations is a constant value σ^2, then it can be shown formally that $cov(s) = \sigma^2(A^TA)^{-1}$ [*Menke*, 1984]. As before, the matrix $(A^TA)^{-1}$ is generally not available due to limitations in computer storage and space, thus leaving us with no estimate for the uncertainties inherent in the slowness structures we derive.

The approach used in this research to estimate the uncertainties in the inversion images is derived from techniques developed for statistical applications and is commonly known as the jackknife, named so for its rough and ready usefulness [*Schucany et al*, 1971]. The method is described thus: suppose we are given n identically independent random variables $\{X_1, X_2, \ldots, X_n\}$ and we wish to estimate the value of a parameter, $\hat{\Theta} = \hat{\Theta}(X_1, X_2, \ldots, X_n)$, e.g. a statistic on the data such as the mean, median or correlation coefficient ($\hat{\Theta}$ represents an estimate of the true value Θ). We form a set of intermediate estimates by considering the values of $\hat{\Theta}^{(i)}$ calculated by leaving out the $i-th$ datum, i.e.

$$\hat{\Theta}^{(i)} = \Theta(X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n) \tag{5.1.1}$$

A set of 'pseudo-values' is created by considering the weighted sum,

$$\tilde{\Theta}^{(i)} = n\hat{\Theta} - (n-1)\hat{\Theta}^{(i)} \tag{5.1.2}$$

and the jackknife estimate of the parameter Θ is the arithmetic mean of the

pseudo-values:

$$\Theta_{JACK} = \frac{1}{n} \sum_{i=1}^{N} \tilde{\Theta}^{(i)}$$ (5.1.3)

The historical motivation for this formulation stems from an attempt to estimate the bias for many common statistics, particularly quadratic functionals [*Quenouille*, 1949; *Efron*, 1982]. Only later did researchers [*Tukey*, 1958] realize the more important use of the jackknife in estimating variability of model parameters. Heuristically, the form of (5.1.2) stems from noting that each pseudo-value measures the influence a particular datum has on estimating the model parameter by forming the linear combination of the estimates with and without the particular datum.

In the tomographic setting the jackknife involves partitioning the set of observations into subsets, performing inversions on the subsets, and calculating a standard error from the set of image vectors that result. Instead of leaving out one datum per inversion, we divide the data into k subsets with each group containing $n - \frac{n}{k}$ data values where the $\frac{n}{k}$ values omitted are chosen without replacement from the original data set. Each subset then has a different portion of the full data set missing. We then perform an inversion for each of the k subsets and denote the slowness image derived from such an inversion $\hat{s}_j$. From these 'mini-inversions' a 'pseudo-inversion' is formed following equation (5.1.2):

$$\tilde{s}_j = k\,\hat{s}_{all} - (k-1)\,\hat{s}_j \ .$$ (5.1.4)

The jackknifed estimate of the slowness is simply the average of the pseudo-inversions:

$$\tilde{s} = \frac{1}{k} \sum_{j=1}^{k} \tilde{s}_j$$

which has variance,

$$v = \frac{\sum \hat{s}_j^2 - \frac{1}{k}(\sum \hat{s}_j)^2}{k(k-1)} \quad . \tag{5.1.5}$$

Given the variance, the standard error is $E_\sigma = \sqrt{v}$. This will be an estimate of the variability of the model due to the variability and distribution of the data. Presently there is no clear cut way to determine the optimal choice of k, the number of mini-inversions to perform. One would guess the larger k is the better, but very large k implies performing large numbers of inversions, which would be extremely time consuming and provide little advantage over computing errors in the classical fashion. A compromise can be struck if we assume that the variability in the pseudo-inversions will be represented in far fewer partitions of the data. In this study I have found $k = 30$ to be a reasonable for estimation of errors on the real data. In order to avoid the influence of any special ordering, the partitions are chosen randomly for each subset.

Jackknife Examples

Let's consider an example with a miniature model where all the matrices are small enough to manipulate and display easily. We start with a model of slowness perturbations generated randomly on a small 6×6 grid. The model is represented as a matrix:

$$
\begin{bmatrix}
2.855 & -1.715 & 5.265 & 3.901 & 2.749 & 4.661 \\
1.540 & -5.464 & -4.254 & -4.262 & 4.572 & 3.475 \\
-1.927 & 2.888 & -3.371 & -5.175 & -4.814 & -4.213 \\
2.163 & -1.112 & 2.311 & 5.440 & -1.181 & 4.097 \\
-4.334 & 5.316 & -1.771 & -1.350 & -2.413 & 1.773 \\
-2.455 & 3.179 & 3.165 & 2.755 & 2.650 & 4.850
\end{bmatrix}
$$

Here each number represents the percentage of slowness perturbation for a block located in the corresponding x-y position of the model. A ray pattern is devised to synthetically calculate the travel times through this model and the data are

generated. A diagram of the ray coverage is presented in Figure 5.1. For this example error was introduced by adding in 1 % of the mean travel time multiplied by a random gaussian distribution. These data were then inverted by application of the normal equations and by the ART algorithm. The classical least squares inverse (via the normal equations) is:

$$\begin{bmatrix} 3.742 & -2.019 & 4.382 & 6.232 & 2.596 & 3.548 \\ 0.060 & -5.883 & -3.014 & -5.071 & 5.214 & 4.352 \\ -2.630 & 5.907 & -3.848 & -4.514 & -4.549 & -3.446 \\ 1.308 & -1.814 & 3.002 & 5.490 & -0.396 & 4.646 \\ -3.148 & 5.064 & -2.237 & -0.855 & -1.007 & 1.282 \\ -2.003 & 4.526 & 2.654 & 1.626 & 1.363 & 3.712 \end{bmatrix}$$

Note that the solution is very close to the truth, but there are some discrepancies due to the errors introduced. The least squares estimate of the errors, as discussed above is given in the following matrix:

$$\begin{bmatrix} 0.662 & 0.708 & 0.797 & 1.018 & 1.029 & 0.907 \\ 0.696 & 0.736 & 0.708 & 0.862 & 0.794 & 0.741 \\ 0.800 & 0.771 & 0.531 & 0.471 & 0.671 & 0.678 \\ 0.795 & 0.761 & 0.529 & 0.467 & 0.665 & 0.674 \\ 0.696 & 0.737 & 0.713 & 0.877 & 0.795 & 0.744 \\ 0.667 & 0.716 & 0.814 & 1.053 & 1.055 & 0.930 \end{bmatrix}$$

Next, the same data set is inverted using the LSQR technique. The program is allowed to iterate until a 99.5% rms reduction in misfit is reached. The answer is:

$$\begin{bmatrix} 3.635 & -2.265 & 4.066 & 5.915 & 2.350 & 3.442 \\ -0.192 & -6.094 & -3.226 & -5.282 & 5.000 & 4.105 \\ -2.947 & 5.694 & -3.993 & -4.659 & -4.759 & -3.767 \\ 0.992 & -2.026 & 2.857 & 5.344 & -0.606 & 4.326 \\ -3.398 & 4.852 & -2.452 & -1.068 & -1.221 & 1.035 \\ -2.106 & 4.281 & 2.337 & 1.307 & 1.115 & 3.609 \end{bmatrix}$$

This solution has a jackknife estimate of standard errors:

$$\begin{bmatrix} 0.737 & 0.809 & 1.111 & 1.527 & 1.663 & 1.341 \\ 1.017 & 1.320 & 1.143 & 1.326 & 1.601 & 1.273 \\ 1.155 & 1.385 & 1.029 & 0.922 & 1.380 & 0.998 \\ 1.329 & 1.668 & 1.037 & 1.048 & 1.266 & 0.908 \\ 1.346 & 1.744 & 1.451 & 1.356 & 1.462 & 1.086 \\ 1.189 & 1.084 & 1.199 & 1.371 & 1.703 & 1.083 \end{bmatrix}$$

One can see by comparison of the two error estimates that the jackknife estimates of error are slightly higher and thus more conservative than the least squares estimate. In this case the difference is very slight. However the example illustrates that the error estimates derived from the jackknife are reasonable in that they give results similar to least squares estimates and for this example are conservative estimates. Indeed, *Efron and Stein* [1981] have shown that jackknife estimates of variance will in general be conservative.

As a second example we offer an estimate of the errors for one of the inversions illustrated in Chapter 3. Here we are using the LSQR Algorithm with regularization ($\lambda = 400$) and $k = 30$ subsets for jackknifing. The error estimates are displayed in Figure 5.2. First, notice that the errors in general are very small, implying that overall we should have high confidence in our model. Second, notice that the errors are largest near the edges of the model where data coverage is sparse. (Blocks that not sampled are excluded from consideration because we have no information about them.) Of course, presence of data with varying amounts of noise will also influence the pattern of the model errors.

Since we have the estimate of the model, the error estimate and the true phantom it is possible to see if our estimates are reasonable. To do so we must compare the estimate of the model parameter to the truth as seen through the resolution kernel at that block. The parameter estimate should then fall within the error bounds. In Figure 5.3, the percent slowness perturbation of block 745 for each of

the 30 pseudo-inversions has been plotted as a function of jackknife number. On the right side of the diagram the mean pseudo-value ("Mean"), the value of s(745) for an inversion using all the rays ("All"), the phantom value ("Truth") and the phantom convolved with the resolution kernel window at 745 ("Kern") are plotted to show the variability of the pseudo values and their relationship to the summary values. The jackknife error estimates are represented as dashed lines plotted about the mean pseudo-value. Notice that the point marked "Kern" lies within the error estimates predicted by the jackknife. This illustrates that, at least at this point, the jackknife estimates of error bounds are reasonable. In Figure 5.4 the cross section A-A' is displayed where the jackknife error estimates are plotted about the mean pseudo-values (s_m) along the cross section. The second dashed line is the LSQR estimate using all the data (s_a). In general, s_m and s_a are comparable and we can use the jackknife estimates of error for either. The magnitude of the errors indicates that we should have a high degree of confidence in the structure, not a surprising result considering we know the inversion has produced a reasonable image.

A closely related alternative to the jackknife is the so called 'bootstrap' approach [*Efron*, 1982]. Here one chooses a random selection of data from the pool of rays *with replacement* and inversions are calculated. As statistics are accumulated an estimate of the covariance matrix is attained. The advantage of the bootstrap over the jackknife is that the basic assumptions are less demanding [*Efron*, 1982]. However, since the subsampling with the bootstrap method is done with replacement, one may be required to analyze a large number of subsets of the data to insure that the variability of all the data is incorporated in the estimate. An *a priori* estimate of how many partitions are sufficient to assess variability is unavailable. On the other hand, the jackknife is guaranteed to use all of the data in providing estimates. For this reason the jackknife was chosen over the

bootstrap for determining model variance estimates.

An alternative method of investigating the effects of noise on model parameters is to create an artificial data set consisting of pure, random, gaussian noise as travel time residuals along with the same raypath coverage from the observed data. This pure noise data set is inverted with the identical procedure used for the real data. In doing this, the level of noise is chosen to have a variance at least as high as that of the estimated errors of the observed data. If the level of the resulting reconstruction with the pure noise is well below the size of the anomalies found with the real data, we can say that the inversion process is relatively insensitive to noise. In that case we would be relatively confident that the anomalies represented true structure and not mere artifacts of noise. This was the approach taken by *Humphreys* [1985] and *Hearn and Clayton* [1985]. The advantage of the data oriented jackknife estimation is that we are using the data themselves to estimate uncertainties without making assumptions regarding the nature of the distribution of errors in the data.
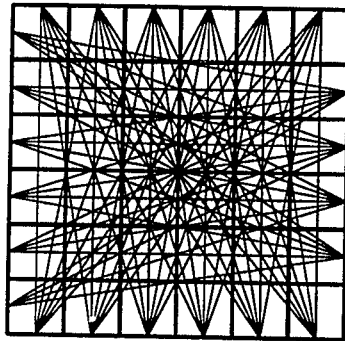
Figure 5.1: Synthetic ray distribution for small example used in jackknife example. The model is a 6 by 6 matrix consisting of random slowness perturbations and there are 60 rays covering the image.
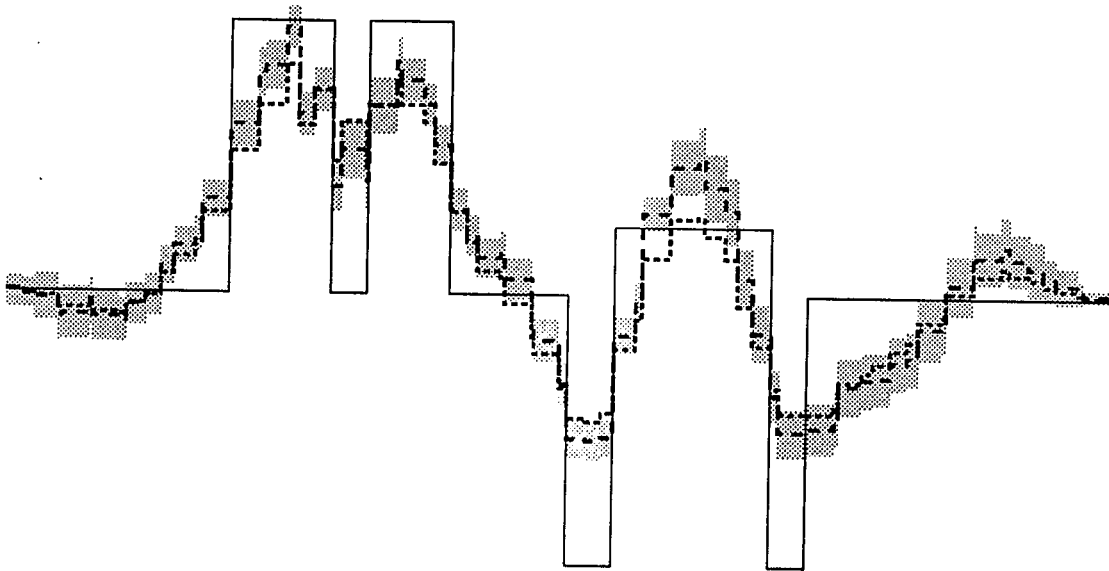
Jackknife Errors
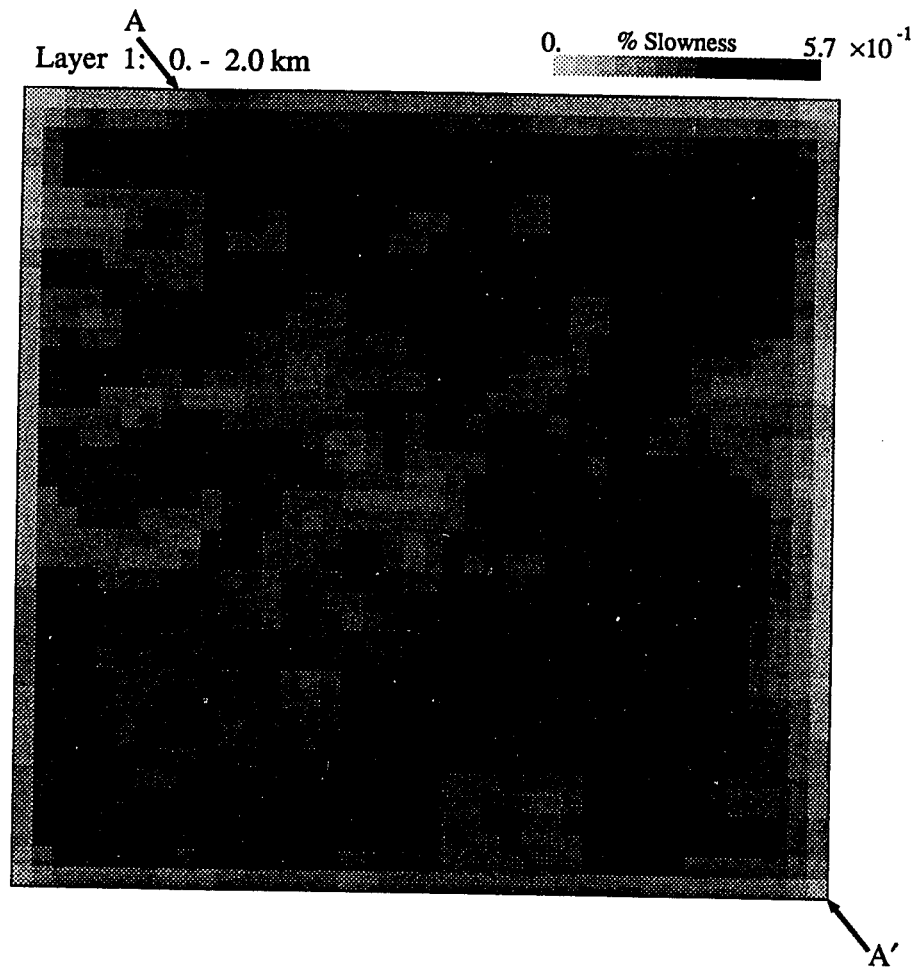
Figure 5.2: Estimated jackknife errors for the LSQR3 inversion of Chapter 3. 30 partitions were inverted to assess variability. On the bottom and right edges the errors are higher because of poor ray coverage.
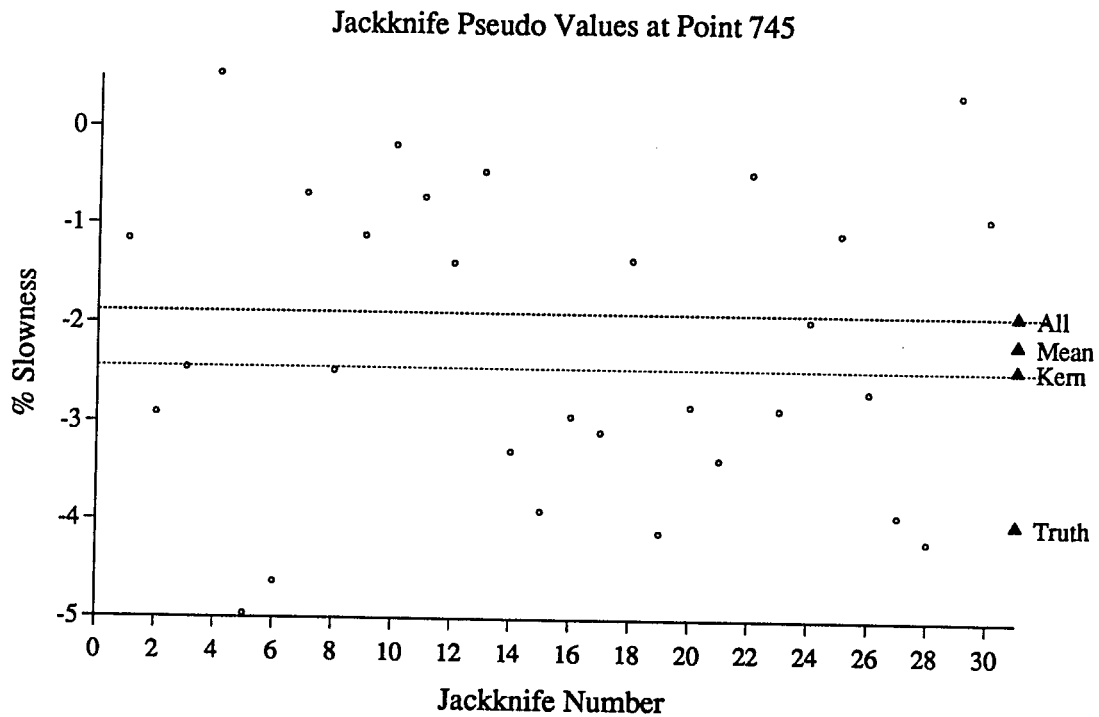
Figure 5.3: Plot of the 30 jackknife estimates for point 745. The triangles represent the mean value (Mean), the inversion using all the data (ALL), the true slowness anomaly (Truth) and the truth convolved with the resolution kernel at point 745 (Kern). The dotted lines are the jackknife estimates of error plotted from the mean. Notice that (ALL) and (Kern) lie within the error estimates.

Figure 5.4: Plot of errors along cross section A-A'. Thick dashes are the mean jackknife estimates (Mean) and thin dashes are the estimates using all the data (All). The stippled area represents the jackknife errors plotted from the mean.

CHAPTER 6

TOMOGRAPHIC INVERSIONS IN THE PUGET SOUND

Introduction and Geology

The geology of western Washington is dominated by tectonics related to the convergence of the Farallon plate with the North American plate since the early Cretaceous. The general nature of the convergence of these plates was first described by *Atwater* [1970] and *Armstrong* [1978] and details of the convergence related to the Tertiary accretions to the North American plate are outlined in *Duncan* [1982] and *Wells et al* [1984]. The features relevant to this study can be divided into four major tectonic units: 1) pre-Tertiary sedimentary and volcanic rocks overlying Precambrian to Mesozoic basement. 2) Mesozoic marine sedimentary and basaltic rocks known as the Leech River schist in southern Vancouver Island. 3) Lower Eocene Basalts of the Crescent formation in Washington and the Metchosin on Vancouver Island. 4) The Core rocks of the Olympic Peninsula, comprised primarily of cenozoic marine sediments [*Clowes*, 1987; *Johnson*, 1984; *Taber*, 1983].

A detailed discussion of the evolution of the Cascade volcanic system was provided by *McBirney* [1978]. Extensive volcanism began in the second half of the Mesozoic and has been continuing episodically through to the present. Major spurts of intense activity are known to have occurred in late Eocene to early Miocene, mid to late Miocene, Pliocene and in the Quaternary. The present lineation of prominent volcanism began approximately 1 million years ago, and, as evidenced by the recent eruptions of Mt. St. Helens, is quite active today. The large strato-volcanos that dominate the landscape of the high Cascades are predominantly andesitic in composition, however periods of basaltic and rhyolitic volcanism are evident in the geologic record. Basalt flows from fissures are also present

and large portions of the range show traces of the Columbia River basalts of the middle to late Miocene. Plutonic bodies are present along the Cascades from Canada to Oregon [*McBirney*, 1978] and several are exposed near Mt. Rainier [*Cowan*, 1984] and Mt St. Helens [*Evarts*, 1987]. The tertiary volcanic sediments are believed to be overlying Pre-Cambrain to mesozoic basement which outcrops in northern Washington and Canada [*Clowes et al*, 1987].

On the west flank of the Cascades there are several exposed sedimentary basins, particularly the Chuckanut Formation to the north and the Puget Group to the southeast of Puget Sound. These are predominantly nonmarine Eocene sediments [*Johnson*, 1984].

The Puget Sound lowlands are covered by a thick layer of Quaternary glacial sediments. Below this extensive region of unconsolidated sediments it is believed there is a thick (down to 20 km) accretionary package of sedimentary rocks of marine origin. These may correlate with the Leach River schist exposed on the southern end of Vancouver Island [*Clowes et al*, 1987]. Because of the thickness of the Quaternary sediments little is known about the geology at depth.

The east flank of the Olympic Mountains is dominated by a large unit of Eocene Basalts known as the Crescent formation in Washington and correlated with the Metchosin formation of southern Vancouver Island [*Clowes et al*, 1987]. This formation appears to be a continuous linear feature along the northwest Pacific coast from Canada down to central Oregon and has been the subject of several recent investigations [*Taber*, 1983; *Duncan*, 1982; *Wells et al*; 1985] in which it is interpreted as a series of seamounts originating in a marine setting and subsequently accreted against the North American plate [*Duncan*, 1982]. *Johnson* [1984] has suggested that a major transcurrent fault system exists along the Crescent formation in southern Vancouver Island and trends southward, perhaps under the Puget Sound.

West of the Crescent formation in the Olympic Peninsula is a large unit of sedimentary rocks comprising the so called "Olympic Core rocks" [*Tabor and Cady*, 1978a; *Taber*, 1984]. This accretionary complex extends out below the continental margins off of Vancouver and Washington [*Clowes et al*, 1987]. The eastern Core rocks that are adjacent to the Crescent formation exhibit a high degree of deformation in contrast to those lying to the west [*Taber*, 1983], indicating extensive compression as the accretion progressed.

Related Geophysical Research

A complete Bouguer gravity anomaly map of Washington was published by *Bonini et al* [1974]. The gravity anomalies in the Puget Sound region have extremely large variations, and in places include gradients that are steeper than anywhere in North America [*Gower et al*, 1985]. The Olympic Mountains are characterized by a gravity high which terminates at the eastern edge of the Crescent formation. In the Puget Sound lowlands three major gravity lows are present: southeast of Olympia, over lake Washington in Seattle and in the Everett region. These are separated by gravity highs. A steep gravity gradient trends south on the western border of the Puget Sound and takes a sharp bend to the southeast north of Olympia. *Gower et al* [1985] have interpreted these large gravity gradients as being evidence for faults below the unconsolidated sediments.

The seismicity beneath the Puget Sound region can generally be divided into two major depth zones. The great majority of earthquakes occur in the shallow depths down to 40 km. These are generally small in magnitude (< 2) and are spatially dispersed through out the region west of the Cascades and east of the Olympics. While some lineations of seismicity are apparent, there are no known surface faults that coincide with specific trends in the spatial distribution of these shallow earthquakes. The deeper events appear to be associated with the subducting

oceanic crust and are interpreted as intraplate earthquakes. A discussion of the separation of seismicity can be found in *Crosson and Owens* [1987].

Reflection and refraction studies have been used on southern Vancouver Island to help define the accretionary geometry of the underthrusting terranes [*Clowes et al*, 1987]. The high impedance contrast across such boundaries as the Crescent basalts and adjacent marine sedimentary wedges allowed the authors to make a detailed interpretation of the features in the subsurface. In the Olympics, *Taber* [1983] made a detailed study of two refraction lines that crossed the Olympics, one in a northwest to southeast direction and one east to west. While the main emphasis of this work was to determine the dip and geometry of the oceanic subducting slab Taber's velocity model indicates low velocity for the Olympic Core (4.7-5.5 km/s) and high velocity for the Crescent Basalt formation (5.5-6.2 km/s). The model for the regions outside of the Olympic Peninsula do not differ significantly from the flat one dimensional model used in this study as the reference model.

A large conductivity anomaly has been observed by *Stanley et al* [1987] in the region roughly bounded by Mt. Rainier, Mt. St. Helens and Mt. Adams. They interpret this low conductivity feature to be evidence for marine forearc basin sediments accreted onto North America during the Eocene. The conductivity anomaly is flanked on the east, west and south by low aeromagnetic anomalies which appear as lineations.

Several aeromagnetic surveys have been performed over parts of western Washington [*Manson*, 1984]. These are generally available individually in map form and have not been compiled and published for the state as a whole, or for the broad areas discussed in this research, and are thus difficult to use. Some work, though, has been done to analyze the aeromagnetic data in a more detailed fashion [*Stanley et al*, 1987; *Finn and Williams*, 1987]. Broad high aeromagnetic

anomalies are present in the south end of the Puget Sound and southwest of Mt. St. Helens [Stanley et al,1987]. Several lineations of low aeromagnetic anomalies, as mentioned above, exist in the southern Washington Cascade region. A detailed report on the analysis of aeromagnetic anomalies in the Mt. St. Helens area was provided by *Finn and Williams* [1987] and is discussed further below.

Puget Sound Inversion

The target for the Puget Sound inversion is displayed in Figure 6.1 along with the western Washington seismic network stations represented as triangles on the map. Major strato-volcanos are designated on the map by stars. The lateral dimensions of the target are 150 km in the east-west direction and 250 km north-south ranging from 46.4° to 48.65° latitude north and 121.47° to 123.42° longitude west. This target was chosen primarily for the available concentration of seismicity. The target area was divided laterally into blocks 5 by 5 km square and in depth according to the reference model described in Table 1.1.

Earthquakes were chosen from the western Washington seismic network data-base including data from 1972 to the present. The selection parameters were fairly liberal: earthquakes with a azimuthal gap $\leq 180°$, number of stations ≥ 5, minimum distance ≤ 50 km, RMS residual $\leq .5$ sec and magnitude ≥ 0.0 were selected. Of these earthquakes, only those rays (station-hypocenter pairs) that remained entirely within the target were used for inversion. There were 4387 earthquakes that passed the above criteria resulting in 36,865 raypaths used for the inversion. The epicenters are plotted on a map view of the target area in Figure 6.2. There is relatively good distribution except for an apparent seismic gap in the southern Puget Sound basin.

Prior to inversion, station corrections are determined and the earthquakes are located according the methods outlined in Chapter 1. A table of the station

locations, elevations and correction delays used in this study is provided in Table 6.1. The station corrections are determined using the full data-set from the whole network, not just the data selected within the target area. Histograms of the residuals for each station are provided in Figure 6.3 using only the data selected for inversion. On the left are the histograms using the 1-dimensional model and in the right hand column are the residuals with the 3-dimensional perturbations added in. Notice that there is an overall tightening of the station residual distributions, as is expected.

For the Puget Sound inversion the LSQR technique of Section 2.4 was used with a damping parameter $\lambda = 1000$ and regularization is achieved by constraining the 2-D horizontal Laplacian to be zero as described previously. After 30 iterations a 40% reduction in residual mean square error was achieved. A plot of the reduction in residual travel time is shown in Figure 6.4. It is clear that the majority of residual reduction is accomplished within the first 10 iterations.

To display the results, greyshade plots are provided for each layer individually. Since the perturbations are from a reference model that is changing with depth, the level of perturbation may be different between layers. For this reason a different grey scale is calculated for each level. The scales are chosen so that the maximum and minimum grey-shades are 2 standard deviations from the mean for that layer. The absolute maximum and minimum percent perturbations for each layer are noted under the scales indicating the extreme range of values for that layer. Percent perturbations of high slowness (low velocity) are represented by dark shades and low slowness (high velocity) by lighter shades. The actual greyshade values range from light to dark smoothly following a shifted cosine curve. Important geographic outlines are included on each plot along with stations plotted as triangles for reference between layers.

It is important to emphasize that the greyshade plots represent percent perturbation from the reference model and care should be taken when comparing anomalies across layers. The actual velocity structures can be obtained from these plots by adding the slowness perturbation to the background slowness in each layer and taking the reciprocal. Plots such as these were found to have nearly all the essential features that are evident in the percent perturbation plots, and are not presented here.

Discussion

Results of the Puget Sound velocity inversion are displayed in Figure 6.5(a-i). A layer by layer description of these results follow:

Layer 1, 0 - 2 km: The shallowest layer of the inversion does not have very good lateral resolution because rays are clustered directly below the stations as they emerge from below. Nevertheless a pattern of slower velocity ranging over the Puget Sound area with higher velocity towards the western flanks of the Cascade Range seems apparent. Since the structures here are concentrated around the stations one can interpret these anomalies as being adjustments to be made to station corrections.

Layer 2, 2 - 4 km: At this level the anomalies begin to exhibit some of the general characteristics expected from the geological data: The east flank of the Olympic Peninsula appears as a high velocity anomaly correlating with the Crescent Formation of Eocene Basalts. The southern Puget Sound exhibits a large low velocity. There is a high velocity anomaly slightly south of the Seward Park (SPW) station in Seattle.

Layer 3, 4-6 km: At this level a large low velocity anomaly in the southern part of the Puget Sound is evident. This low velocity structure appears to extend to the southeast in a linear trend following a correlated low gravity apparent in this region [*Bonini et al*, 1974; *Gower et al*, 1985]. Under Seattle (SEA) and to the north are two additional low velocity anomalies which also correlate with gravity lows. Note the high anomaly at SPW in contrast to the low at SEA. This reflects the large gravity gradient beneath this area interpreted as a fault by Gower et al [1985]. To illustrate this point a reproduction of the gravity anomalies with interpreted faults is supplied in Figure 6.6. For comparison a close up view of Layer 3 is plotted in Figure 6.7 at the same scale as Figure 6.6. Note the strong correlation between the low velocity structures and the low gravity anomalies.

Layer 4, 6-9 km: Here the Crescent Basalts appear quite prominently as a high velocity anomaly and the Puget Sound low velocity anomaly appears to wrap around in the characteristic crescent shape. The large low velocities of the southern Puget Sound are greatly diminished in magnitude and extent.

Layer 5, 9-12 km: The Crescent basalts again appear here butted against the north-south low velocity trend which appears to be dipping eastward. On the western edge of the Crescent Formation a low velocity anomaly appears, representing the underplated rocks of the Olympic Core.

Layer 6, 12-16 km: The trend described for layer 5 continues here. Several prominent anomalies appear in the north east sector of the target area that have not been stressed so far. These are due to the complicated juxtaposition of allocthanous terranes, sedimentary basins and volcanics that

characterizes the northern portion of the target. These complications make interpretations difficult.

Layers 7 through 9, 16-32 km: The basic trend of low velocity anomaly stepping eastward under the high velocity basalts continues.

To aid the interpretation of these velocity anomalies several cross sections through the model have been prepared. The map location of the cross sections are presented in Figure 6.8. The cross sections are created by stacking the results of all the layers and taking a vertical slice through the 3-D block. For each layer the same greyshade scale that was used in the horizontal plots above is used for the cross sectional plots. These plots (Figure 6.9*a-d*) are presented primarily to help delineate the broad spatial patterns with respect to depth.

Cross Section XA′: This section runs east-west through the northern part of the Puget Sound. It shows the dipping nature of the Crescent basalt formation (high velocity) as it is underthrust beneath the oceanic sedimentary units deposited in a backarc basin. Below the Crescent basalts are the Olympic Core rocks appearing as a low velocity marine sedimentary terrane.

Cross Section XB′: This section runs from the Olympic Peninsula through Mount Rainier in the southeast. The Crescent formation is again seen to be dipping to the east above the Core rocks. At Mount Rainier a low velocity anomaly is present between 9 and 12 km depth. Immediately to the southeast of Mt. Rainier is a small region of high velocity material. The Puget Sound lowland is characterized by a low velocity zone down to 6 km.

Cross Section XC′: The trend of the dipping basalts is most marked here. An apparent dip of 20-28° is measured from this cross section. Again, the same pattern of low velocity Core rocks underlying the Crescent is evident as is the low velocity unconsolidated sediments of the Puget Sound.

Cross Section DD′: This section corresponds to Cowan's [1984] transect B-3. The same structures discussed above appear here. In the vicinity of Mt. Rainier two high velocity anomalies appear to flank the volcano. These are interpreted as evidence for pre-tertiary plutons which have surface expression in various localities north and south of Mount Rainier.

Resolution and Error

As discussed in Chapter 4, the resolution for an inversion of this size is difficult to calculate. To approximate the resolution, several plots are supplied indicating an estimate of the resolving power of the data based on the distribution of ray coverage over the model. Two impulse responses are illustrated in Figures 6.10-12. The first one (Figure 6.10a-c) was placed in the center of the model (depth of 9-12 km), an area where the resolution is presumed to be very good. The impulse response displays show that indeed the lateral resolution in layer 5 (9-12 km depth) is approximately 15 km in width. In depth the resolution appears to be about the same, the drop in amplitude in layers 4 and 6 is about one third of the amplitude in the block under investigation. An east-west cross section is supplied in Figure 6.11 to show the vertical resolution. While the amplitude of dispersed energy is significantly diminished a pattern dipping to the east is observed. This implies that some of the structure alluded to above is due to the effect of smearing and raypath effects.

The second impulse response is located in the vicinity of Mt. Rainier at a depth of 6-9 km (Figure 6.12a-c). While the resolution here is not quite as good as in the example above it is still in the range of 15-20 km (3-4 blocks). Again there is smearing in the layers above and below this position as indicated by the side lobes of the resolution kernel in layers 3 and 5.

To get an overall picture of the resolving power throughout the ellipsoids describing the ray distribution are presented in Figure 6.13. Adjacent to the ellipsoid plots are plots of the dip of the major axis and hit counts represented on a logarithmic scale by circles of varying radii. These two illustrations thus indicate the density of distribution in a particular area and the relative spread of the data in three mutually orthogonal directions. The plots generally show that in the center of the model along the Puget Sound lowlands the ray distribution is dense and fairly isotropic. On the edges the the distribution degrades and the resolution is likewise degraded. In the center of the model, from depths 6-16 km the rays appear to be primarily oriented horizontally with an occasional block having dominant vertical ray orientation. Of course at the surface and on the edges there is more smearing in the vertical direction due to rays leaving and arriving at the end points of raypaths.

The standard errors, as estimated using the jackknife with 30 partitions, is presented in Figure 6.14. Overall the errors are relatively small, less than 2.7% over the whole model. In layer 3 (4-6 km) there is a high estimated standard error (2.7%) near the center of the target. The dominant anomaly in this region is approximately 8.2% so the ±2.7% is not large enough to wipe the anomaly out, i.e. the sign of the anomaly is well constrained. This is true for all the blocks that have high ray coverage density over them. This indicates that errors inherent in the data, due to mispicking or mislocation, are not introducing spurious anomalies into the model.

Summary

In summary both gross features of the surface geology and several details surmised from geophysical studies are apparent as anomalous velocity structures imaged by the tomographic technique. The Crescent Basalts on the eastern flanks of the Olympics appear as a high velocity anomaly as opposed to the sedimentary deposits in the Puget Sound lowlands which are predominantly low velocity. Three large low velocity anomalies appear in the central to southern Puget Sound region which correlate well with large gravity anomalies in this area [Bonini et al, 1977]. In the vicinity of Mt. Rainier small shallow high velocity structures appear to correlate with plutons observed at the surface and inferred at depth [Cowan, 1984]. A large low velocity anomaly is evident below Mt. Rainier which may correspond to the large gradients in conductivity observed by Stanley [1987] which was interpreted as a an accretionary prism of sedimentary deposits. On a broad scale there appears to be evidence for large scale underthrusting of the Crescent formation and the Olympic Core formation beneath the Puget Sound. The geometry and apparent dip of the boundaries of these accretionary units corresponds well with independent measurements made north of the Puget Sound on Vancouver Island [Clowes et al, 1987]. These correlations indicate that the velocity structures derived using the tomographic method indeed reflect important geologic structures in the subsurface.

Western Washington Target Area
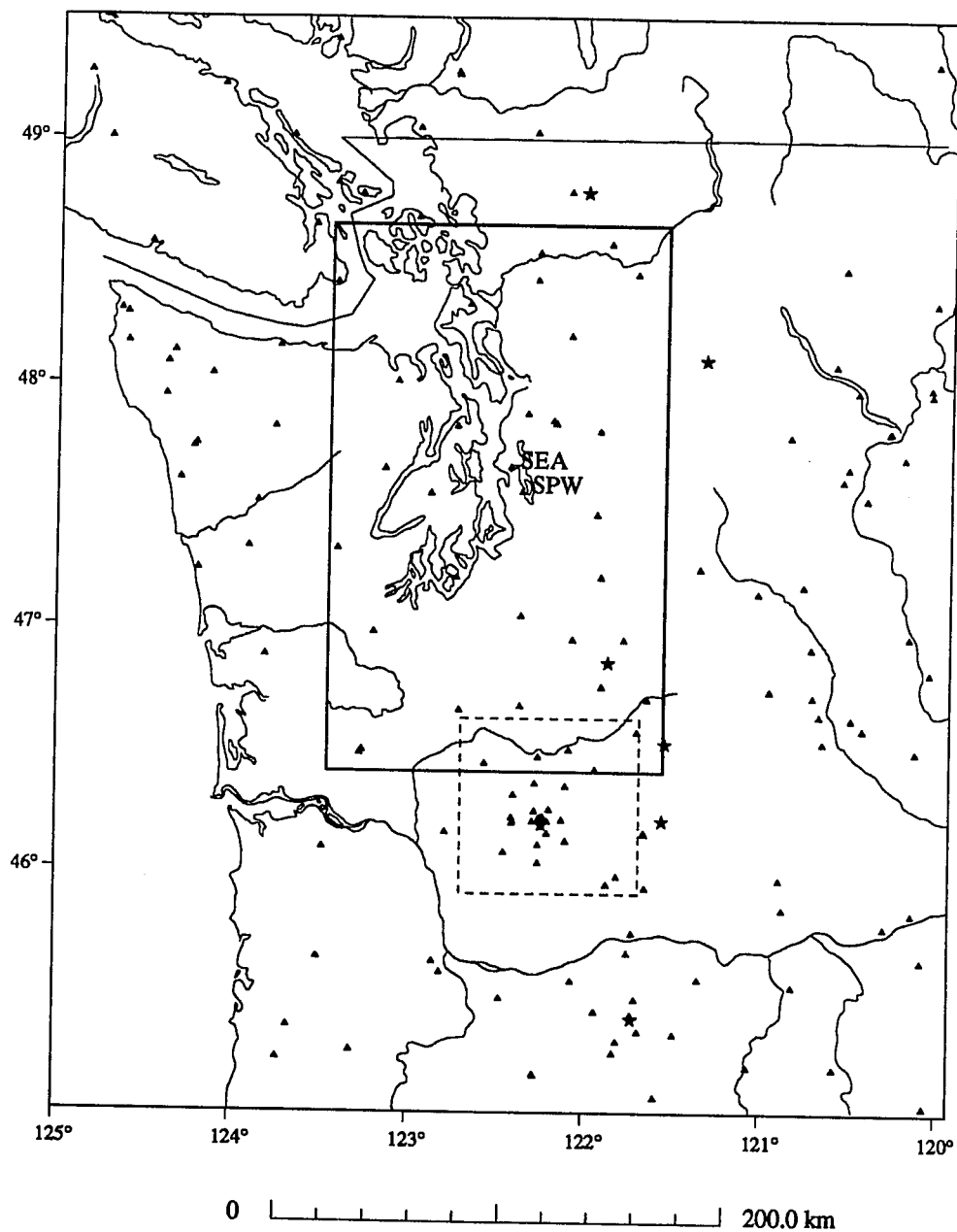
Figure 6.1: Map of target areas and station distribution of the western Washington seismic network. Triangles are station locations and stars are major strato-volcanos in the region. The hashed target is the Mt. St. Helens target area discussed in Chapter 7. Stations SEA and SPW are labelled for reference.

Table 6.1. Station corrections determined by inspection of histograms of residuals for each station using all the data. The S-wave corrections are calculated from the P-wave corrections.

Station Corrections					
Station	Latitude	Longitude	Elevation	P-delay	S-delay
APW	46N39.1000	122W38.8500	0.457	-0.2500	-0.4280
AUG	45N44.1667	121W40.8333	0.865	0.1500	0.2570
BFW	46N29.2000	123W12.8900	0.902	-0.3500	-0.5990
BHW	47N50.2100	122W 1.9300	0.198	-0.0500	-0.0860
BLN	48N 0.4417	122W58.3107	0.585	-0.4500	-0.7700
BOW	46N28.5000	123W13.6833	0.870	-0.3500	-0.5990
BRV	46N29.1200	119W59.4900	0.925	0.5000	0.8550
CBW	47N48.4250	120W 1.9600	1.160	0.6000	1.0260
CDF	46N 6.9700	122W 2.8500	0.780	0.0040	0.0070
CMM	46N26.1167	122W30.3500	0.62	-0.1480	-0.2530
COW	46N29.4600	122W 0.7267	0.305	-0.0980	-0.1680
CPW	46N58.4300	123W 8.1800	0.792	-0.2480	-0.4240
CRF	46N49.5100	119W23.3000	0.260	0.8000	1.3680
DY2	47N59.1152	119W46.2172	0.884	0.8000	1.3680
DYH	47N57.6300	119W46.1600	0.820	0.8000	1.3680
EDM	46N11.8400	122W 9.0000	1.609	0.2000	0.3420
ELK	46N18.3333	122W20.4500	1.270	-0.1960	-0.3350
ELL	46N54.5833	120W34.1000	0.805	0.8500	1.4540
EPH	47N21.2133	119W35.7700	0.628	0.7000	1.1970
EST	47N14.2800	121W12.3633	0.756	0.0540	0.0920
ETT	47N39.3000	120W17.6000	0.439	0.4500	0.7700
FL2	46N11.7833	122W21.0167	1.378	-0.1500	-0.2570
FLT	46N11.3550	122W21.3750	1.387	-0.0460	-0.0790
FMW	46N56.5167	121W40.1900	1.890	-0.1500	-0.2570
FOR	45N58.2333	121W45.5000	1.152	0.1000	0.1710
FPW	47N58.1500	120W12.7750	0.352	0.5000	0.8550
GBL	46N35.8600	119W27.5900	0.330	0.8500	1.4540
GHW	47N 2.5000	122W16.3500	0.268	-0.3000	-0.5130
GLD	45N50.2167	120W48.7667	0.610	0.6500	1.1120
GLK	46N33.8367	121W36.5117	1.320	0.0500	0.0860
GMW	47N32.8750	122W47.1800	0.506	-0.5000	-0.8550

Table 6.1: (continued)

Station	Latitude	Longitude	Elevation	P-delay	S-delay
GSM	47N12.1900	121W47.6700	1.305	-0.0500	-0.0860
HDW	47N38.9100	123W 3.2533	1.006	-0.3980	-0.6810
HNB	49N16.4698	122W34.7517	0.183	-0.0500	-0.0860
HTW	47N48.2083	121W46.1442	0.829	-0.1500	-0.2570
JCW	48N11.7117	121W55.5178	0.792	-0.1480	-0.2530
JUN	46N 8.8000	122W 9.1800	1.049	0.0040	0.0070
KOS	46N27.6800	122W11.4300	0.828	-0.1500	-0.2570
LMW	46N40.0800	122W17.4800	1.195	-0.0960	-0.1640
LON	46N45.0000	121W48.6000	0.853	-0.1980	-0.3390
LVP	46N 4.1000	122W24.5000	1.170	-0.0460	-0.0790
MAS	46N 8.6833	121W35.5117	1.370	0.1000	0.1710
MBW	48N47.0400	121W53.9800	1.676	0.3500	0.5990
MCW	48N40.7800	122W49.9400	0.693	-0.1460	-0.2500
MDW	46N36.8000	119W45.6500	0.330	0.5000	0.8550
MOW	47N50.7817	122W 2.8817	0.180	-0.0460	-0.0790
MOX	46N34.6333	120W17.5833	0.540	0.6000	1.0260
MTM	46N 1.5300	122W12.7000	1.121	0.0540	0.0920
NAB	49N13.3500	124W 0.2157	0.256	-0.1000	-0.1710
NAC	46N44.0633	120W49.5533	0.738	0.5500	0.9410
NEL	48N04.6967	120W20.2950	1.490	0.3000	0.5130
NLO	46N 5.3000	123W27.0000	0.900	0.1500	0.2570
OBH	47N19.5750	123W51.9500	0.383	-0.1000	-0.1710
OHW	48N19.4000	122W31.9100	0.054	-0.3500	-0.5990
OMK	48N28.8200	119W33.6500	0.421	0.8000	1.3680
ONR	46N52.6250	123W46.2750	0.257	0.1000	0.1710
OSD	47N49.2500	123W42.1000	2.010	0.1000	0.1710
OTH	46N44.3400	119W12.9900	0.260	0.3000	0.5130
PAT	45N52.8350	119W45.6683	0.300	0.5000	0.8550
PEN	45N36.7200	118W45.7750	0.430	-0.6000	-1.0260
PFB	48N34.4998	124W26.6638	0.465	-0.3000	-0.5130
PGC	48N38.9998	123W27.0295	0.005	-0.2500	-0.4280
PGO	45N28.0000	122W27.1667	0.237	0.3500	0.5990
PHO	45N37.1300	122W49.8367	0.299	0.0500	0.0860
PLN	47N47.0800	120W37.9803	0.700	0.3500	0.5990
PRO	46N12.7600	119W41.1500	0.552	0.5000	0.8550
RAN	46N24.5000	121W51.8167	1.620	0.0540	0.0920
RED	45N56.2200	121W49.1800	1.510	0.1500	0.2570

Table 6.1: (continued)

Station	Latitude	Longitude	Elevation	P-delay	S-delay
RMW	47N27.5825	121W48.3200	1.024	0.0040	0.0070
RPW	48N26.9000	121W30.8167	0.850	0.1000	0.1710
RSW	46N23.4700	119W35.3200	1.037	0.6000	1.0260
RVC	46N56.5750	121W58.2883	1.000	-0.0960	-0.1640
RVW	46N 8.9700	122W44.6200	0.460	-0.3500	-0.5990
SAW	47N42.1000	119W24.0600	0.690	0.5000	0.8550
SBL	46N20.4200	122W 2.3300	1.665	-0.0960	-0.1640
SHW	46N11.8433	122W14.1400	1.399	0.1000	0.1710
SMW	47N19.1700	123W20.5000	0.840	-0.1460	-0.2500
SNB	48N46.5598	123W10.2718	0.408	0.2000	0.3420
SOS	46N14.6417	122W 8.2000	1.270	-0.0960	-0.1640
SPW	47N33.2217	122W14.7517	0.008	0.5500	0.9410
STD	46N14.2667	122W13.3650	1.268	-0.0980	-0.1680
STW	48N 9.0483	123W40.2183	0.308	-0.1460	-0.2500
SUG	46N12.8700	122W10.4900	1.859	0.2500	0.4280
SYR	46N51.7800	119W37.0700	0.267	0.7500	1.2830
TBM	47N10.1690	120W35.9000	1.064	0.7500	1.2830
TDH	45N17.3900	121W47.4200	1.541	0.1500	0.2570
TDL	46N21.0500	122W12.9500	1.400	-0.1980	-0.3390
VBE	45N 3.6200	121W35.2100	1.544	0.4000	0.6840
VGZ	48N24.8338	123W19.4637	0.067	-0.4000	-0.6840
VLL	45N27.8000	121W40.7500	1.195	0.4500	0.7700
VLM	45N32.3100	122W 2.3500	1.150	0.1500	0.2570
VTG	46N57.4800	119W59.2400	0.208	0.8000	1.3680
VTH	45N10.8700	120W33.6800	0.773	-0.2000	-0.3420
WA2	46N45.4033	119W33.7583	0.230	0.7500	1.2830
WAT	47N41.9167	119W57.2500	0.900	0.6500	1.1120
WEN	47N31.7700	120W11.6500	1.061	0.4000	0.6840
WNS	46N42.6167	120W34.5000	1.000	0.7500	1.2830
WPW	46N41.8900	121W32.8000	1.250	0.0040	0.0070
WRD	46N58.1900	119W 8.6000	0.378	0.2500	0.4280
YAK	46N31.2633	120W31.7533	0.619	0.5500	0.9410
YEL	46N12.5833	122W11.2667	1.750	0.	0.

Western Washington Seismicity
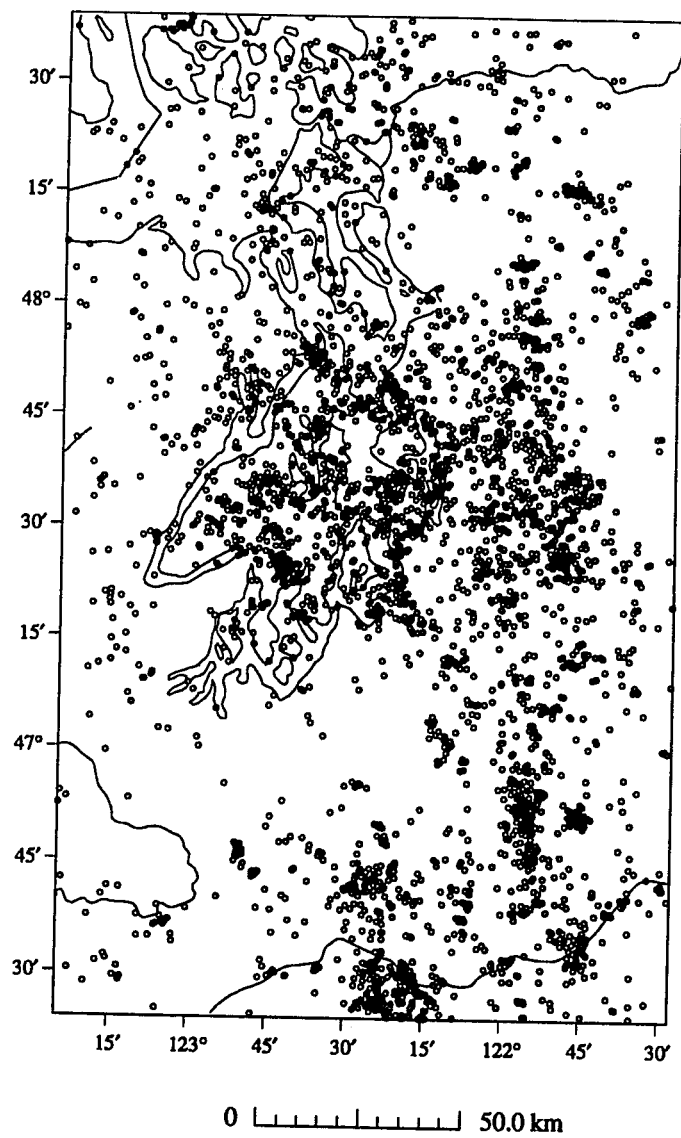
Figure 6.2: Map of earthquake epicenters for the Puget Sound target. Over 4000 earthquakes were selected and used in the inversions.

Figure 6.3: Plots of station residual histograms. The left plot for each station represents the station residuals with the 1-D reference model and the right plot is the histogram of station residuals using the 3-D model derived in this study. Only stations with large numbers of total counts are plotted. The first pair of histograms include the residuals for the whole dataset. The horizontal axes are in units of seconds and the vertical axes are counts.

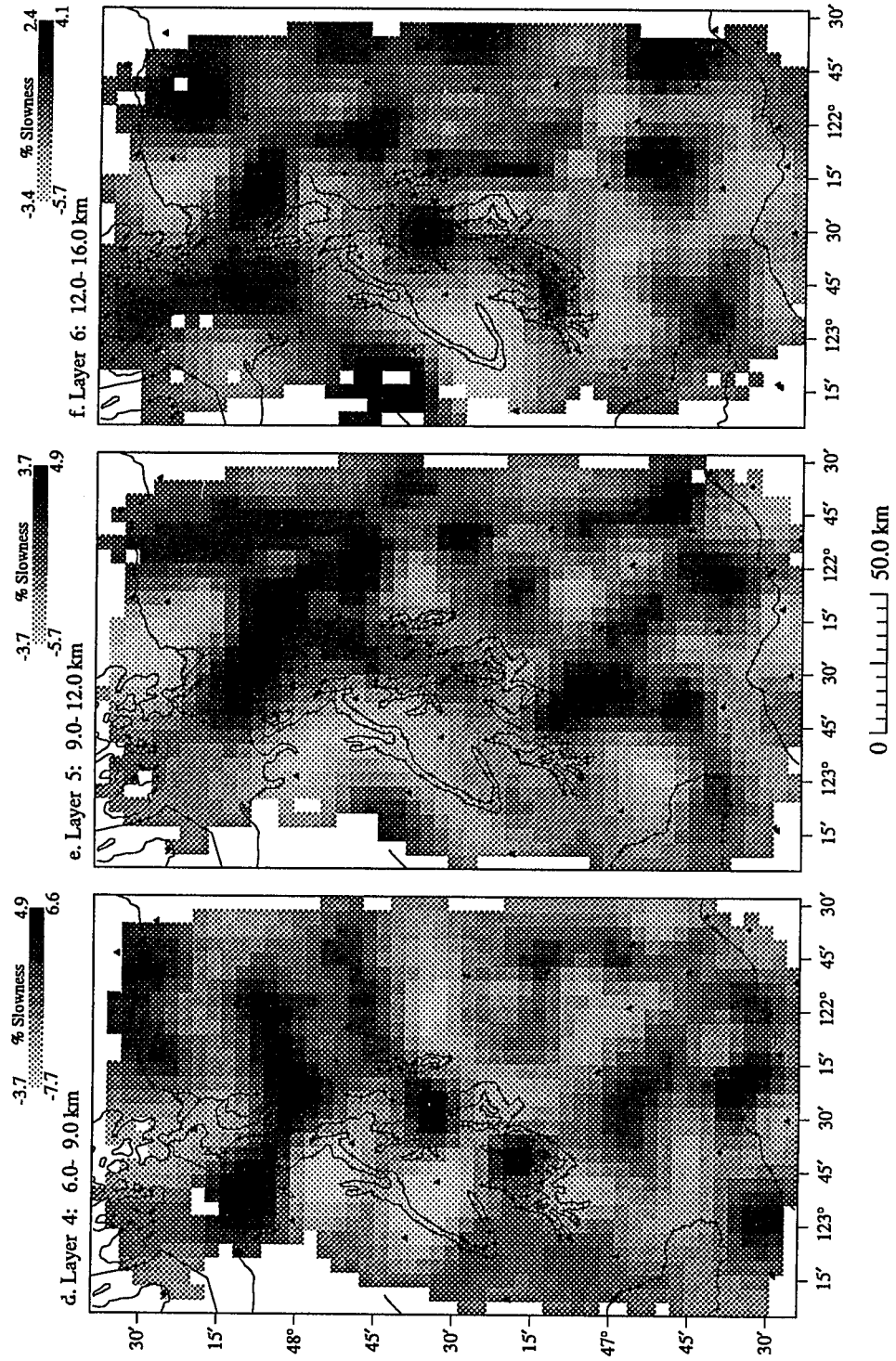Figure 6.3 (continued)   97
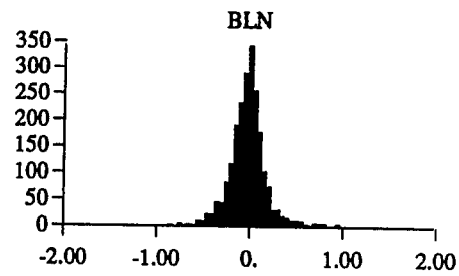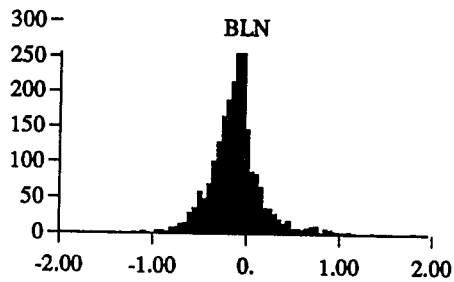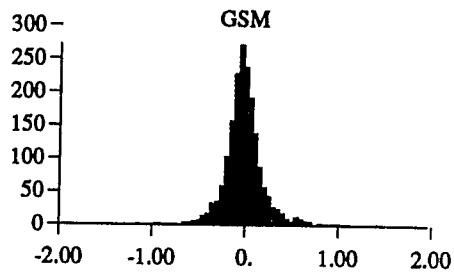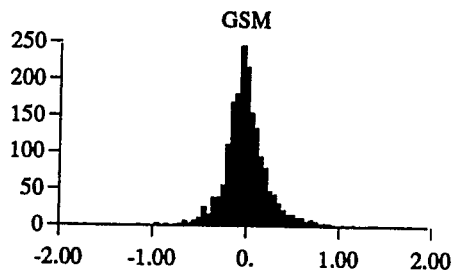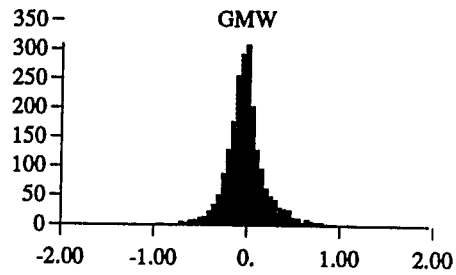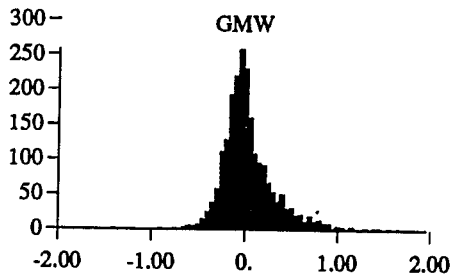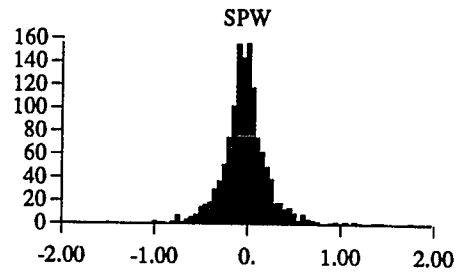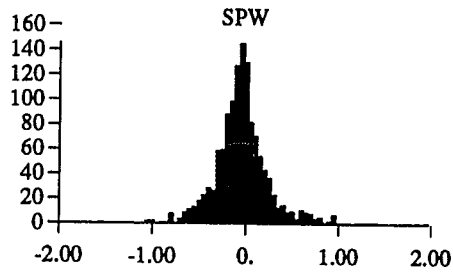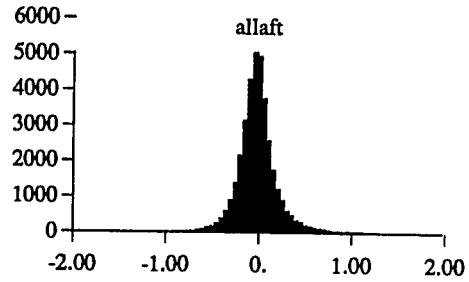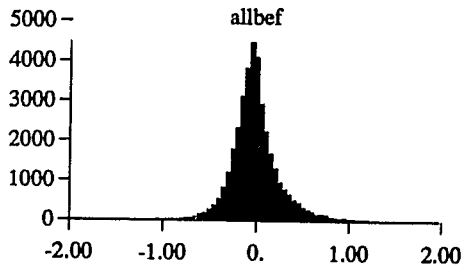
Figure 6.3 (continued)   98
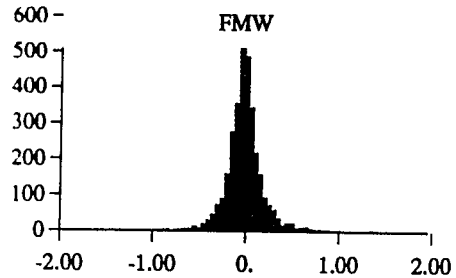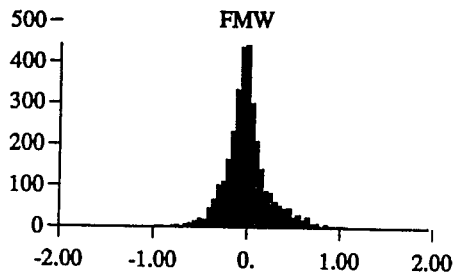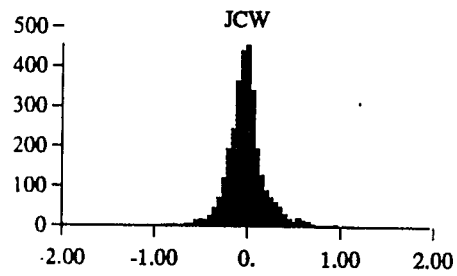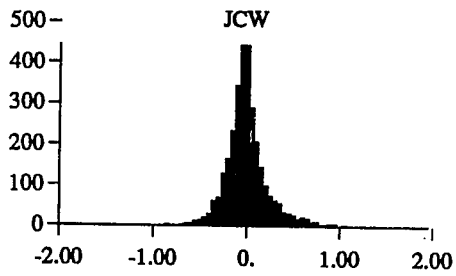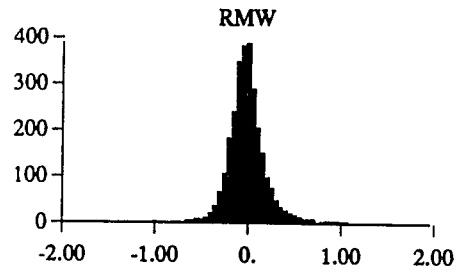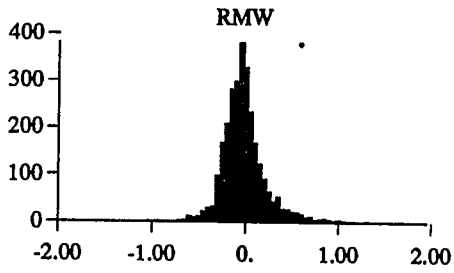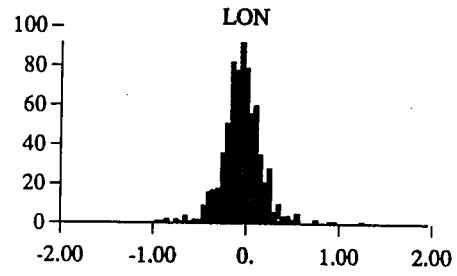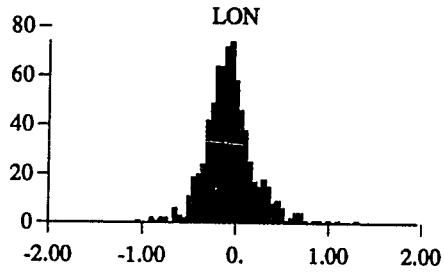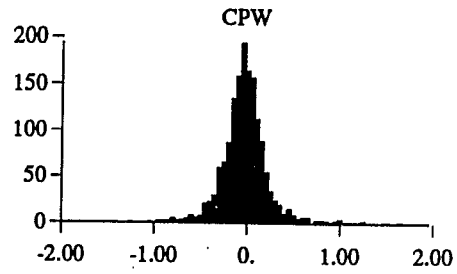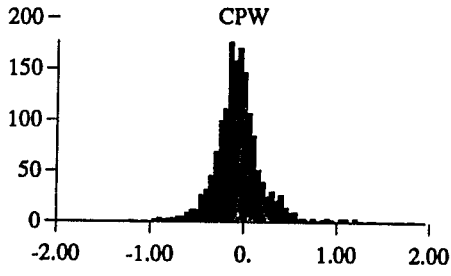
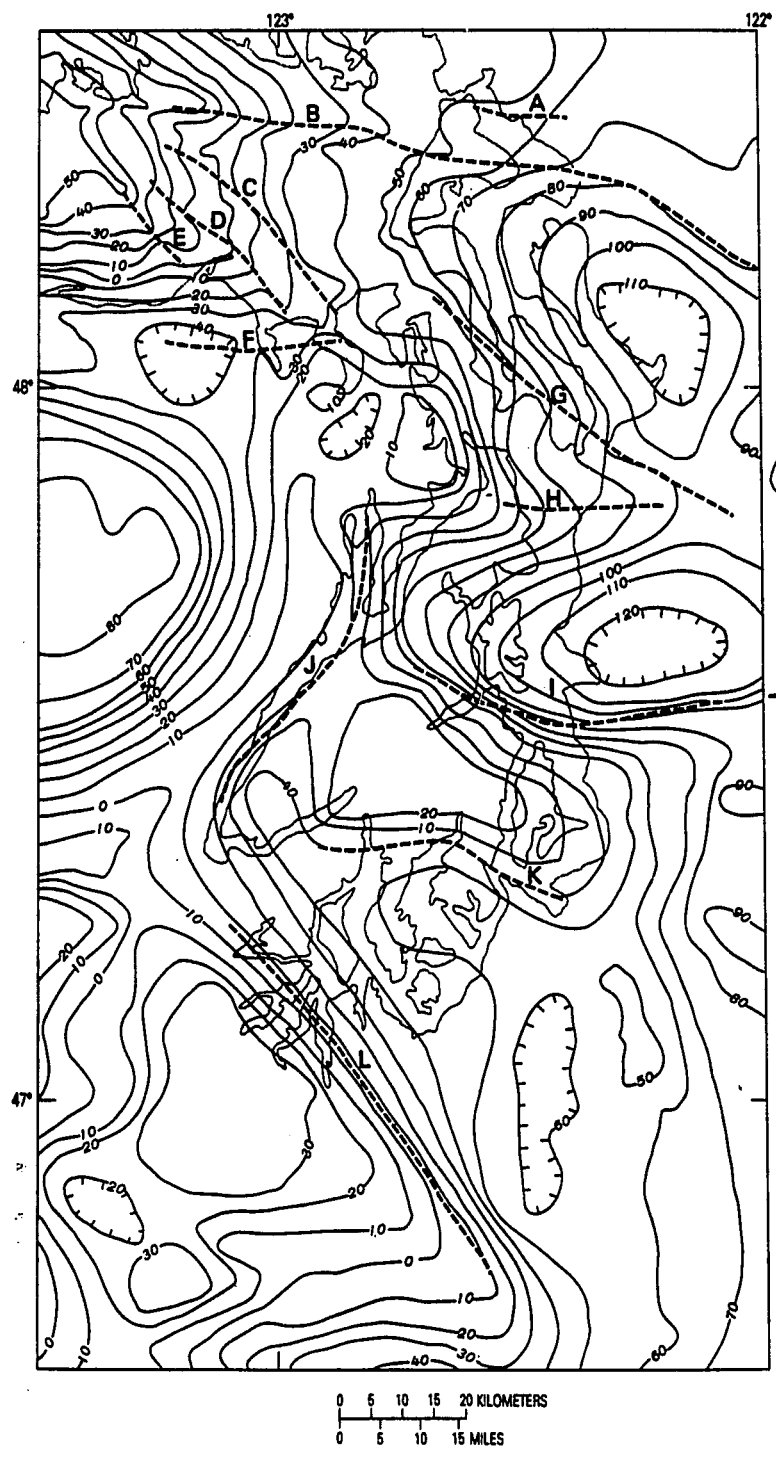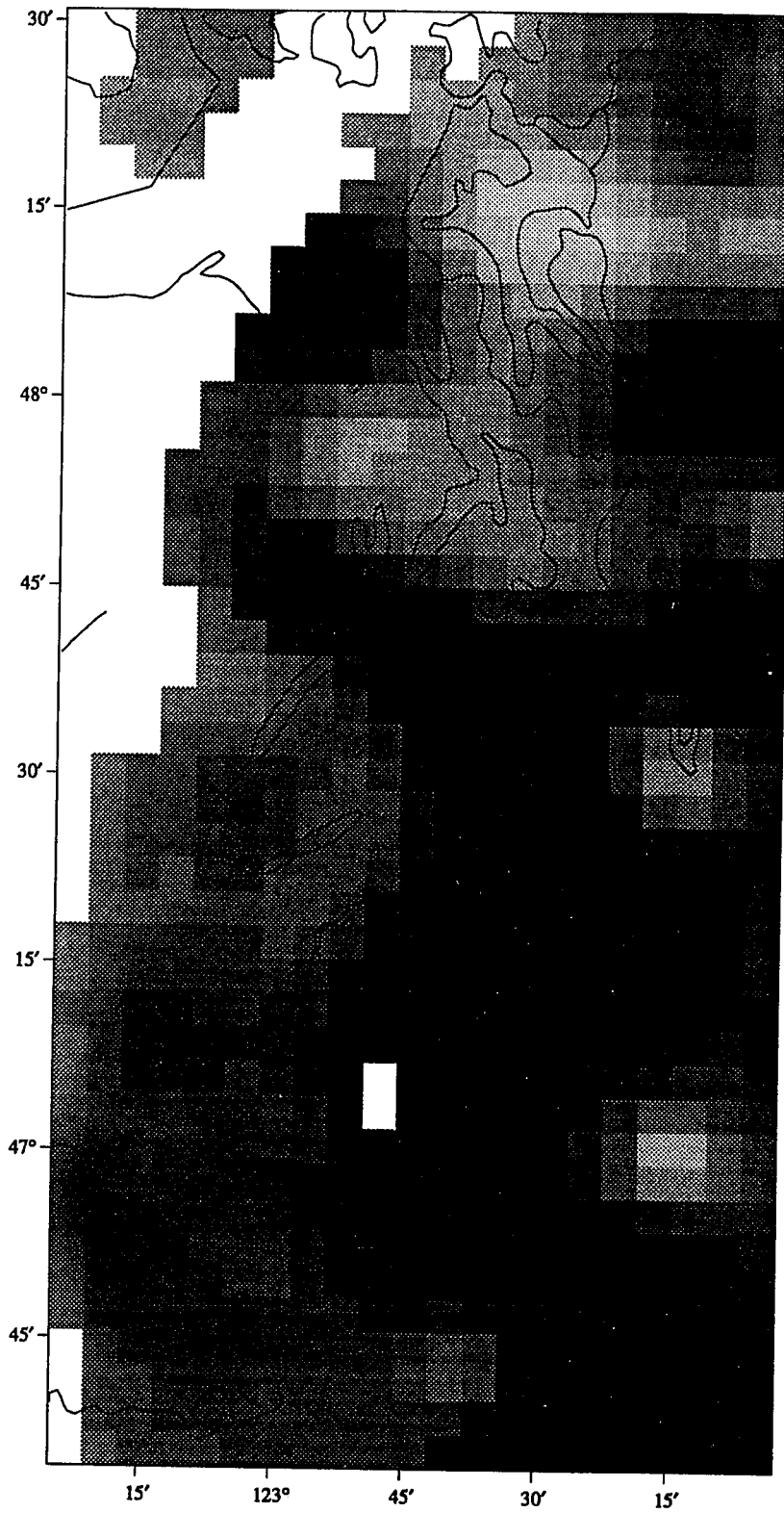Figure 6.3 (continued)   99
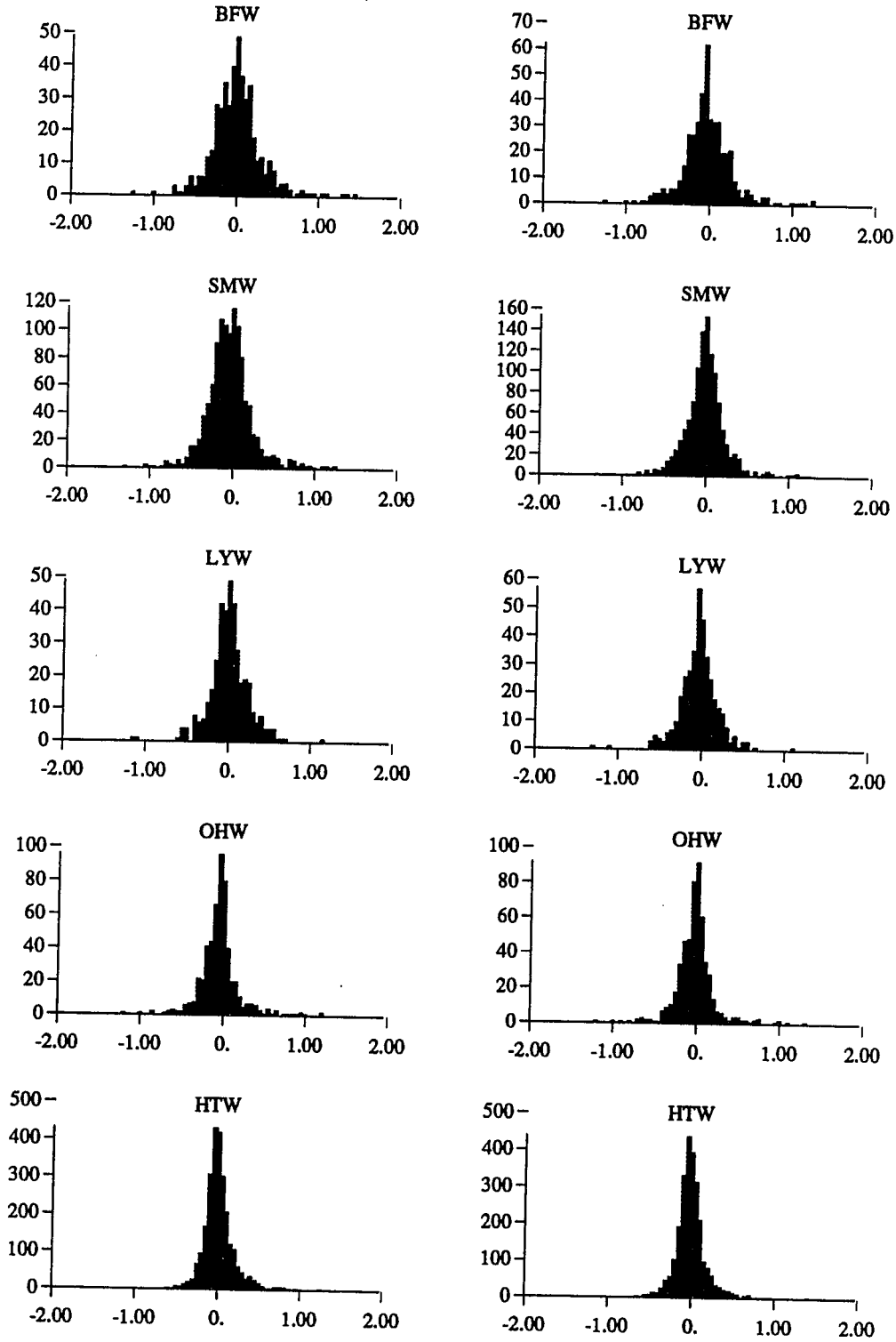
Percent Reduction in Misfit
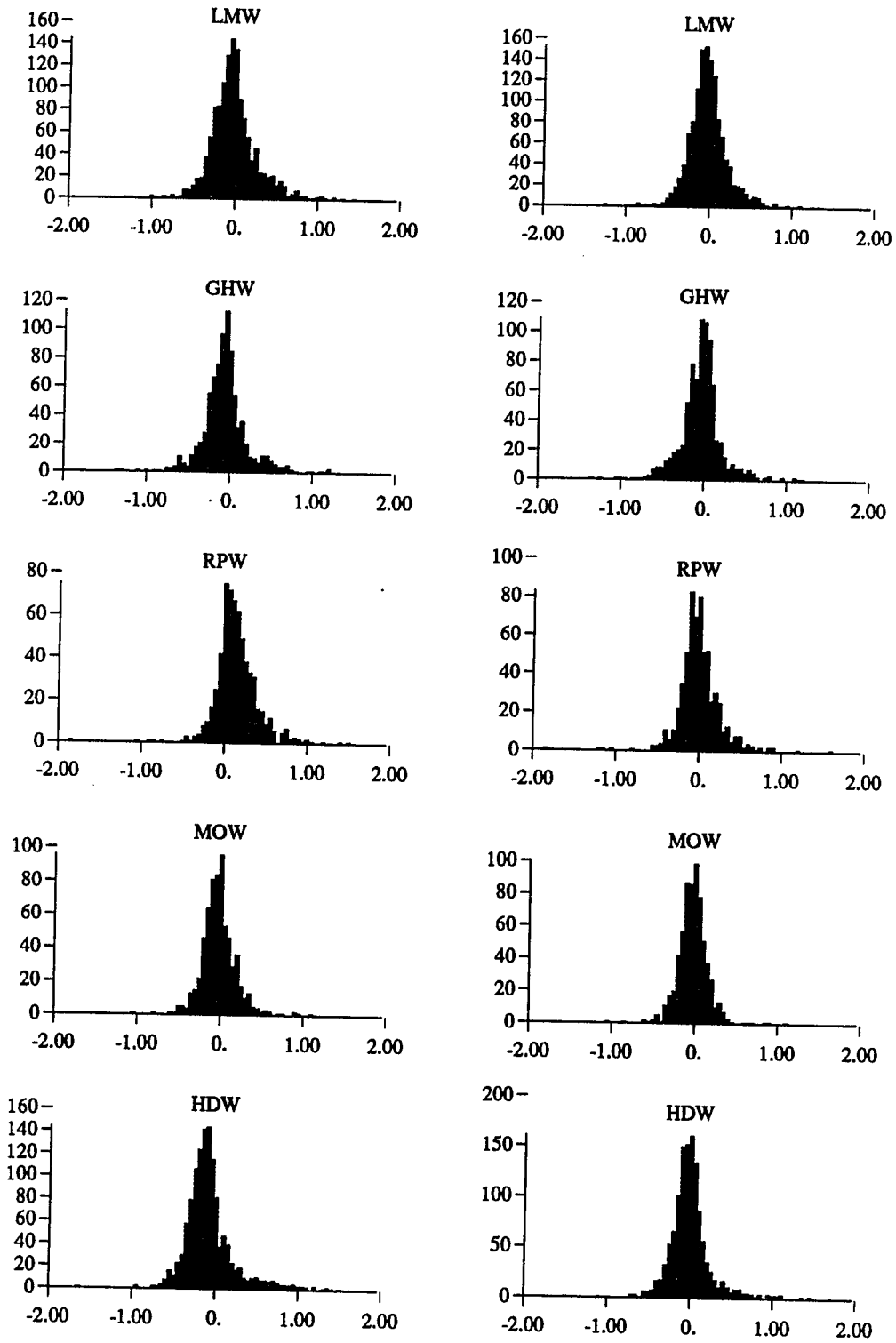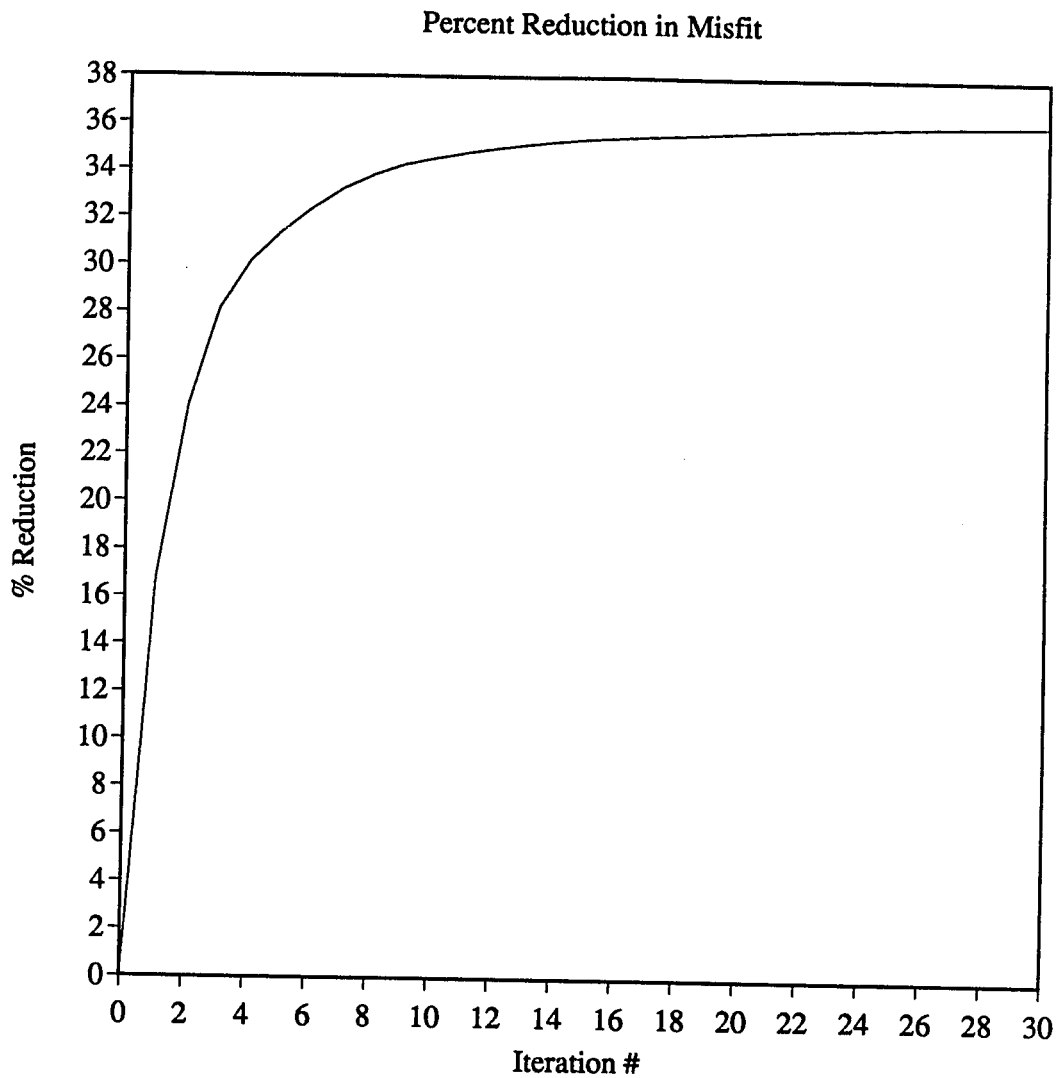
Figure 6.4: Percent reduction of χ^2 versus iteration number. The inversion was done on Puget Sound data using the LSQR algorithm with regularization implemented by constraining the Laplacian to be zero in horizontal layers.

Figure 6.5(*a-i*): Tomographic velocity inversion of the Puget Sound. Each figure represents a horizontal slab in the earth. Greyshades represent percent slowness perturbations from the reference model. Light colored regions are high velocity and dark regions are low velocity. Dark triangles are station locations plotted for reference.

Puget Sound Velocity Perturbations

a. Layer 1: 0.-2.0 km

b. Layer 2: 2.0-4.0 km

c. Layer 3: 4.0-6.0 km

Figure 6.5: (continued)

d. Layer 4: 6.0 - 9.0 km

e. Layer 5: 9.0 - 12.0 km

f. Layer 6: 12.0 - 16.0 km

Figure 6.5: (continued)

g. Layer 7: 16.0- 20.0 km

h. Layer 8: 20.0- 25.0 km

i. Layer 9: 25.0- 32.0 km
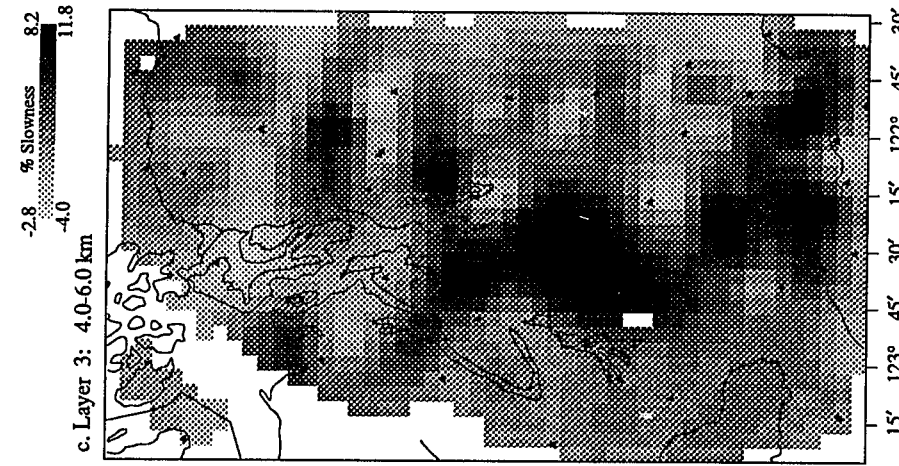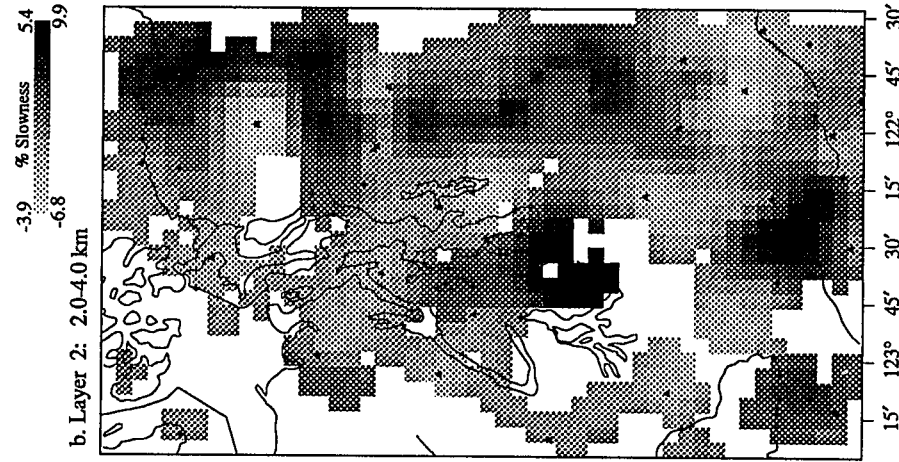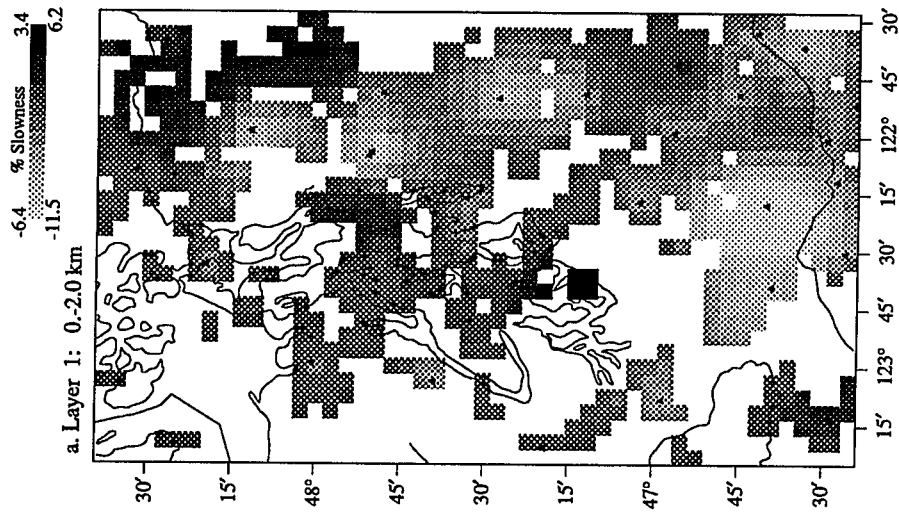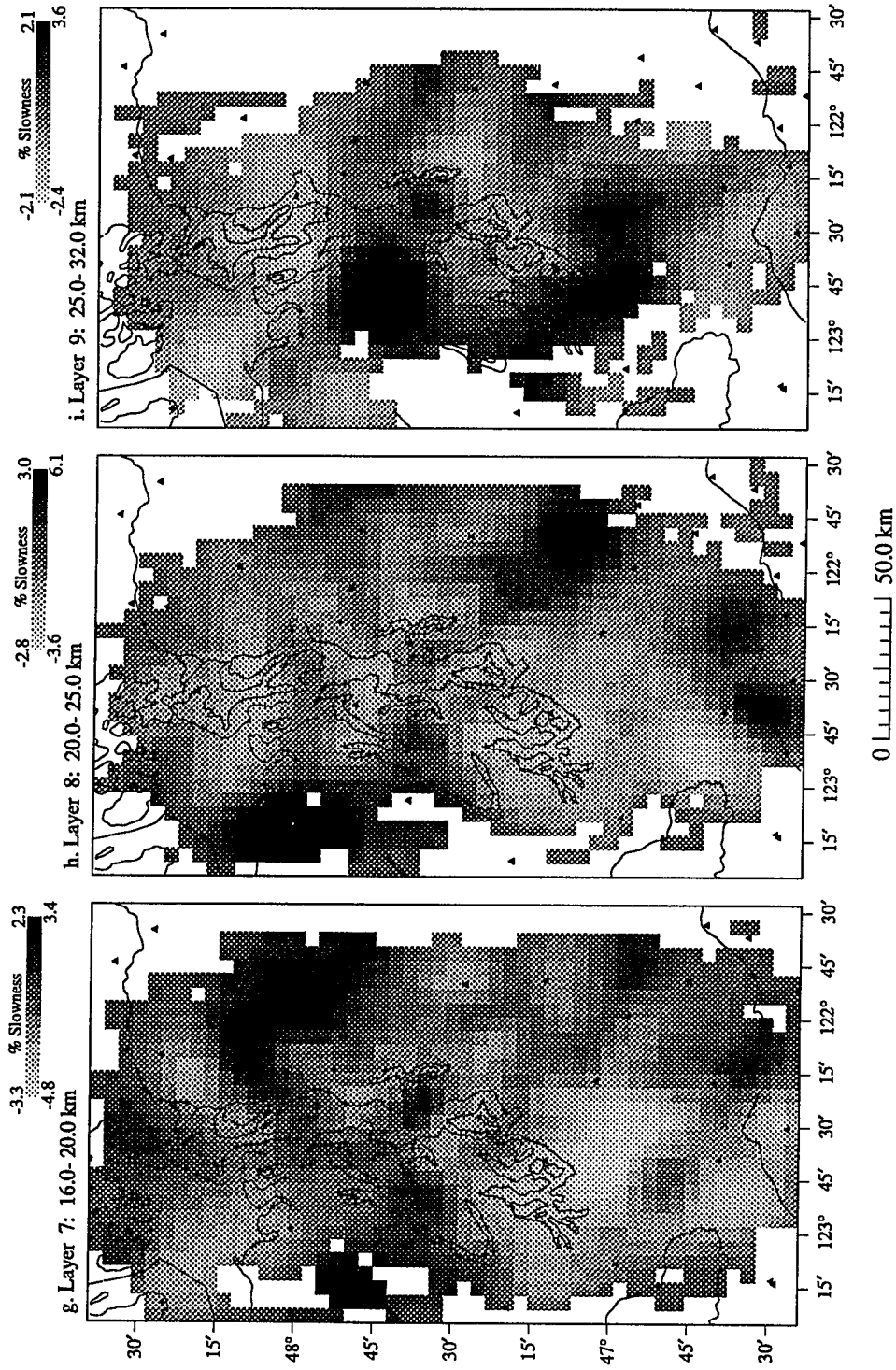
Figure 6.6: Regional Bouguer gravity map in milligals. From *Gower et al* [1985].

Figure 6.7: Close up of Layer 3 from Figure 6.5. To compare velocity structure to Bouguer gravity structures this is plotted at same scale as Figure 6.6.
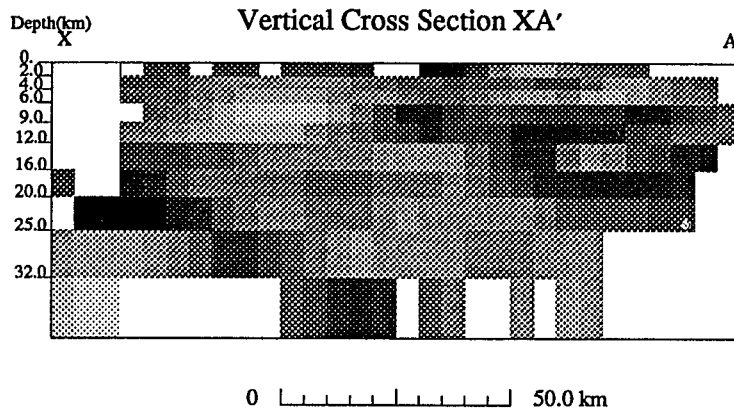
Layer 3: 4.0 to 6.0 km
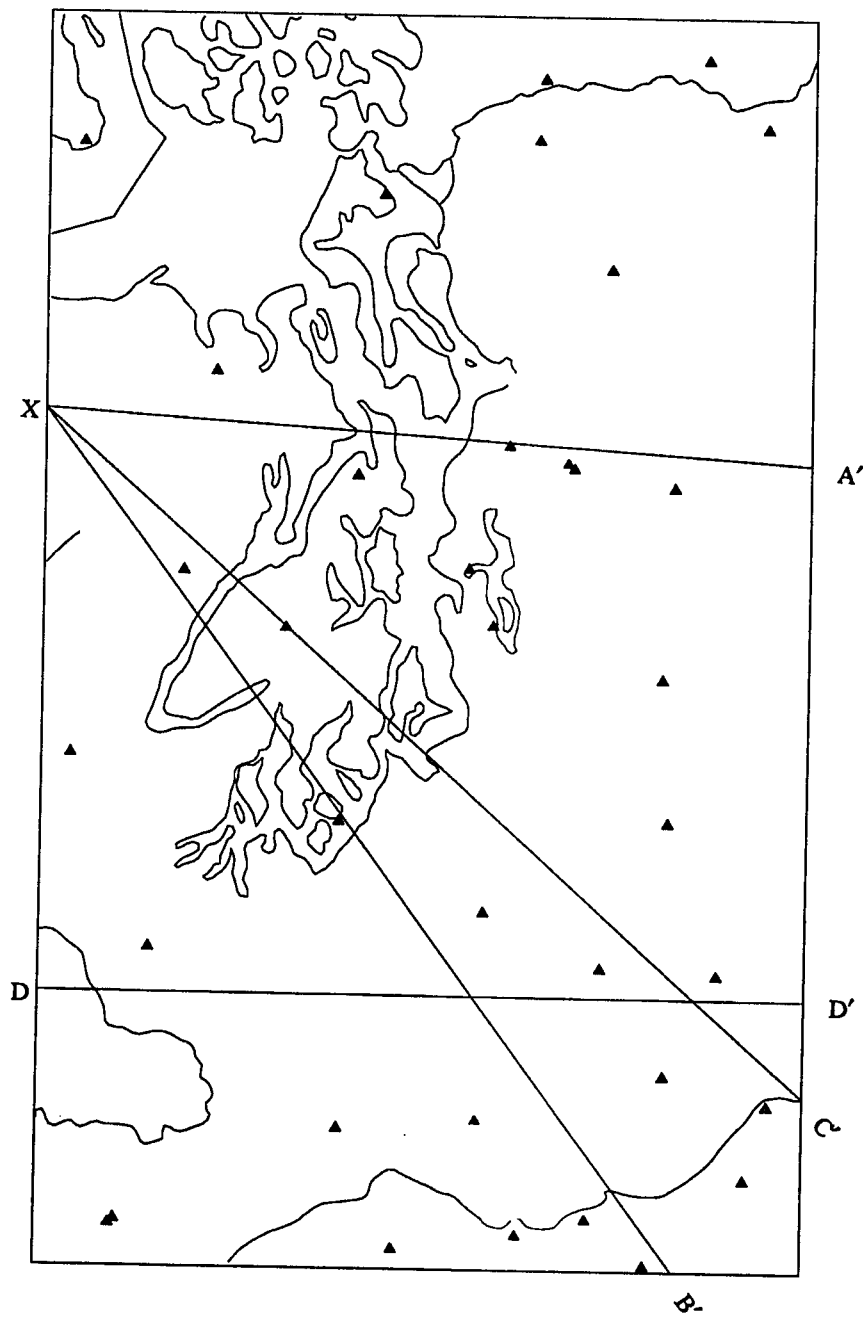
Figure 6.8: Map of cross sections through model.

Depth(km)
X

Vertical Cross Section XA'

A'

0.
2.0
4.0
6.0
9.0
12.0
16.0
20.0
25.0
32.0

0 └─┴─┴─┴─┴─┴─┴─┴─┴─┘ 50.0 km

Depth(km)
X

Vertical Cross Section XB'

B'

0.
2.0
4.0
6.0
9.0
12.0
16.0
20.0
25.0
32.0

0 └─┴─┴─┴─┴─┴─┴─┴─┴─┘ 50.0 km
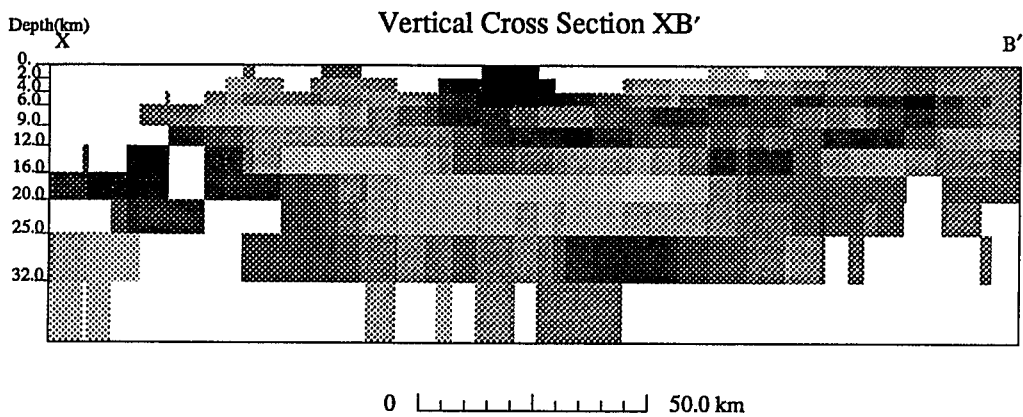
Figure 6.9: Vertical cross sections through the Puget Sound model. The 3-D model has been sliced along line illustrated in Figure 6.8 and the layers are displayed with the greyshade scale adjusted within each layer according to Figure 6.5. The structure of the Crescent formation can be seen dipping to the southeast in section XB. Cross section DD corresponds to transect B3 [*Cowan*, 1984]

Figure 6.9: (continued)

Depth(km)

Vertical Cross Section XC′

0 |⌐⌐⌐⌐⌐|⌐⌐⌐⌐⌐| 50.0 km

Depth(km)

Vertical Cross Section DD′
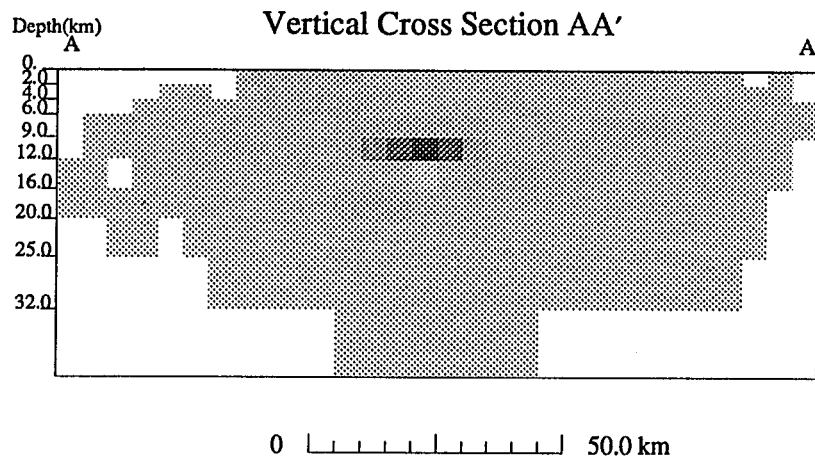

0 |⌐⌐⌐⌐⌐|⌐⌐⌐⌐⌐| 50.0 km
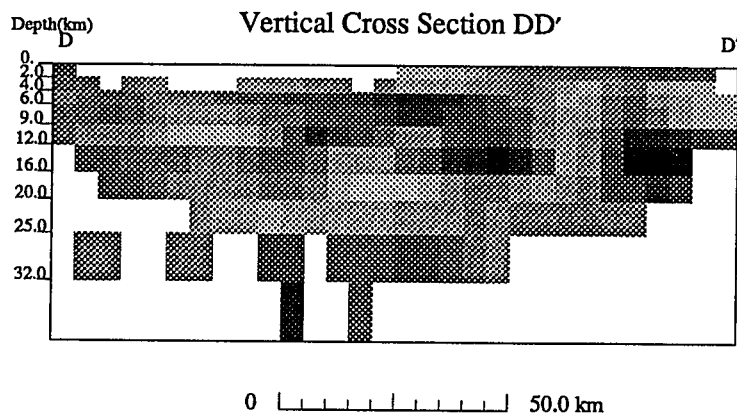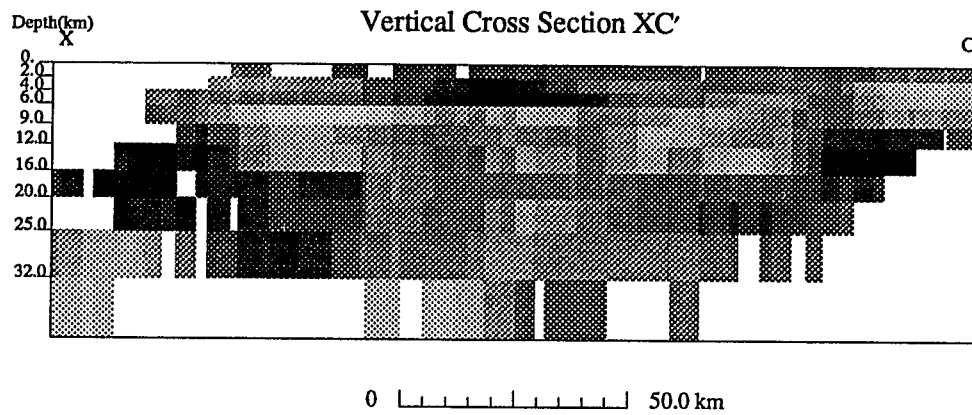
Figure 6.10(*a-c*): Resolution kernel located in the center of layer 5, depth 9-16 km. The lateral resolution in this area is approximately 3 blocks, or 15 km. While there is smearing in the vertical the amplitude of the side lobes is down by a third.

Resolution Kernel for Center of Layer 5

a. Layer 4: 6.0-9.0 km

b. Layer 5: 9.0-12.0 km

c. Layer 6: 12.0-16.0 km
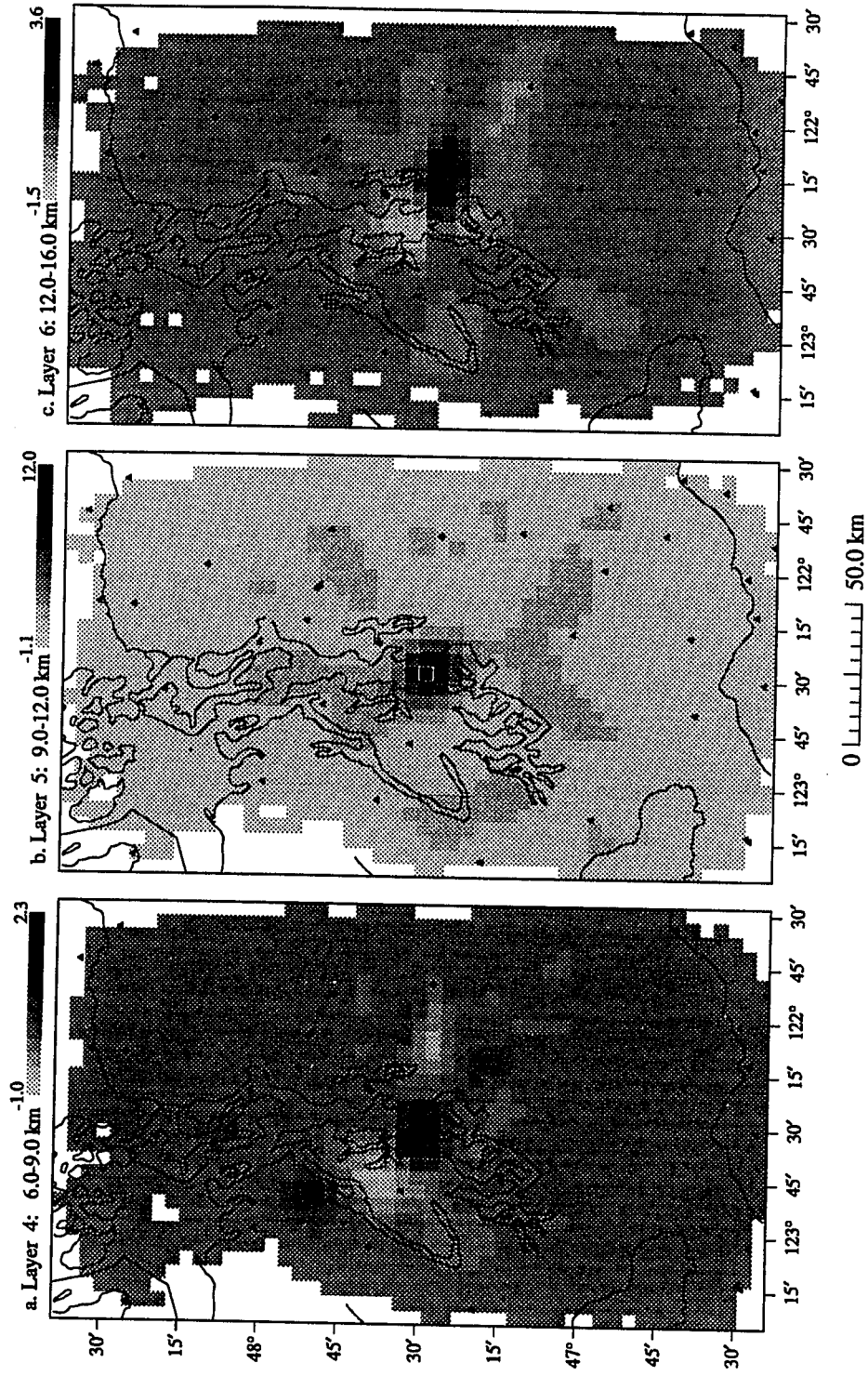
Figure 6.11: East-west slice through the resolution kernel in Figure 6.10. Vertical smearing is evident here but the amplitude drops off significantly. For ease in comparison the greyshades are plotted with the same scale for all levels.
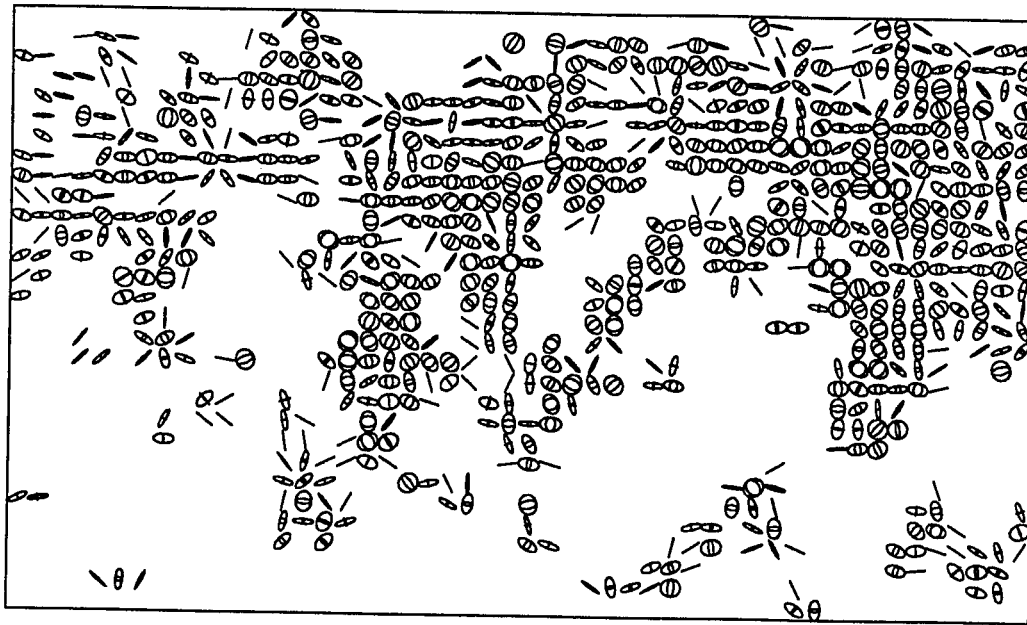
Figure 6.12(*a-c*): Resolution kernel located near Mt. Rainier in layer 4, depth 6-9 km. The lateral resolution in this area is approximately 3 blocks, or 15 km. While there is smearing in the vertical the amplitude of the side lobes is down by a third.

Resolution Kernel at Mt. Rainier

a. Layer 3: 4.0- 6.0 km

b. Layer 4: 6.0- 9.0 km

c. Layer 5: 9.0- 12.0 km

0   50.0 km

112
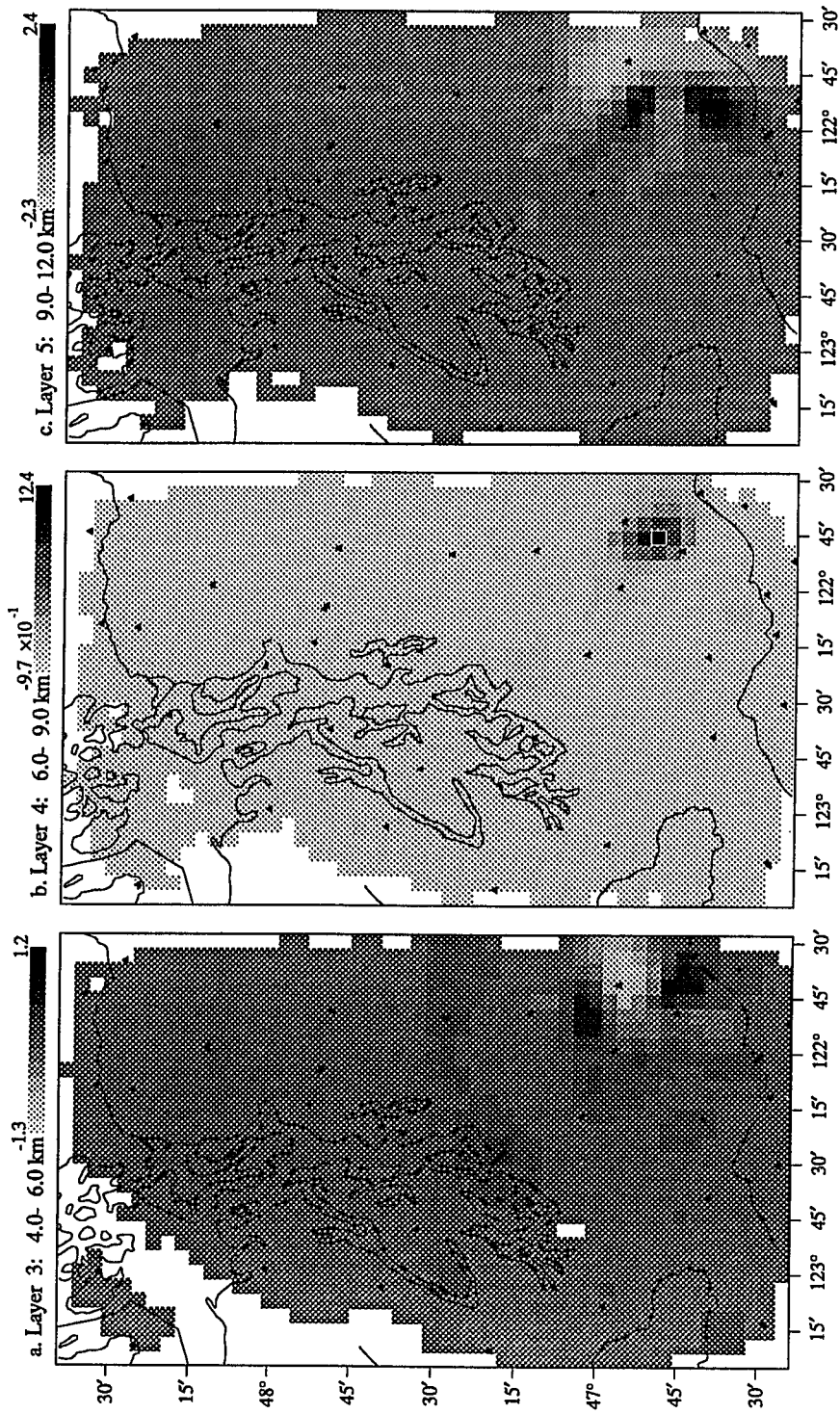
Figure 6.13(*a-i*): Ray ellipsoids and hit count density plots for the Puget Sound inversion. For each layer, the left side shows the relative size of eigenvalues for each ellipsoid and its azimuthal direction. The right side shows the vertical dip of the ellipsoid and the logarithm of the hit count plotted as scaled circles. The number Max refers to the maximum hit count for the layer.
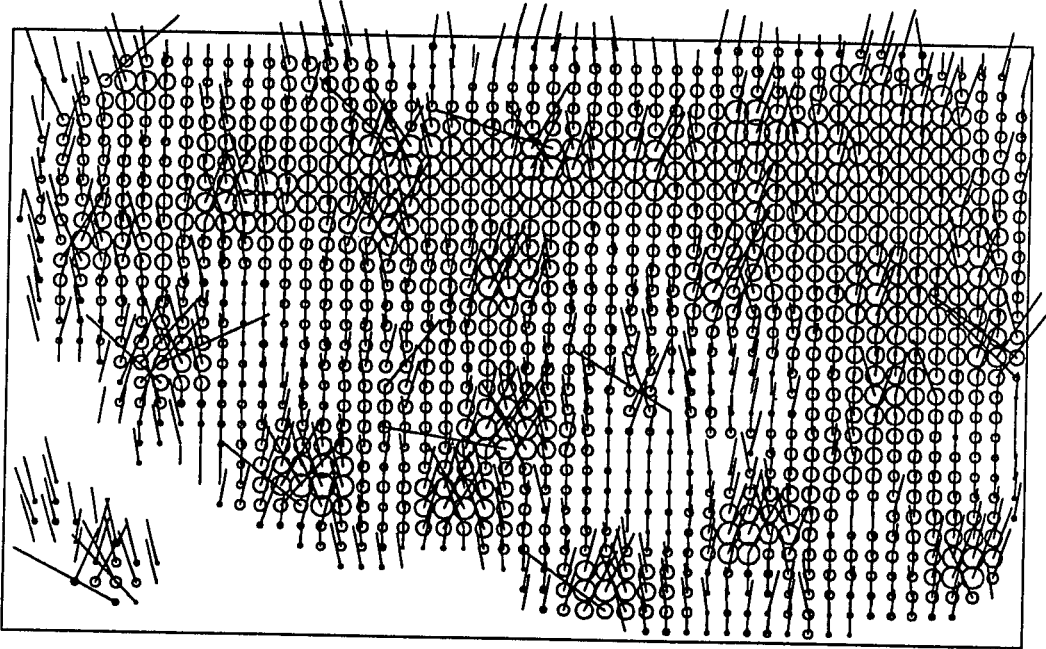
Vertical Projections: Max Count= 3199

a. Layer 1: 0. - 2.0 km Ray Ellipsoids

Figure 6.13: (continued)

Vertical Projections: Max Count= 2276

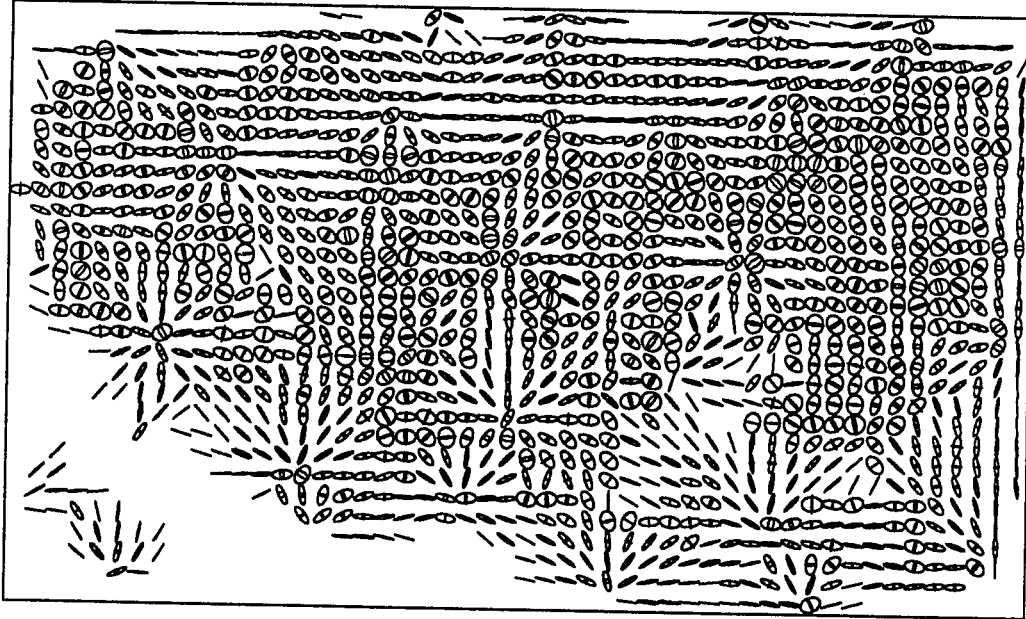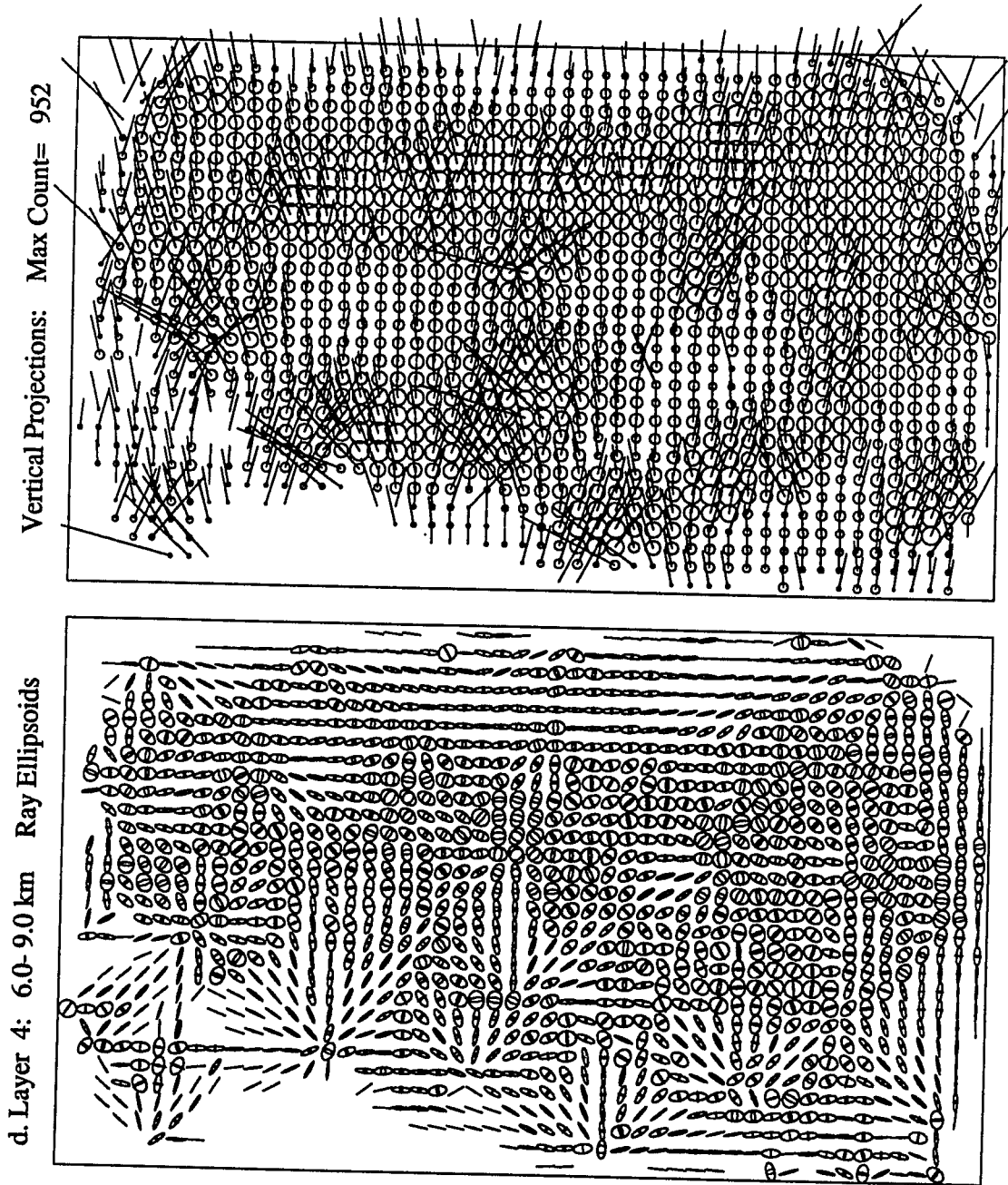b. Layer 2:   2.0- 4.0 km   Ray Ellipsoids
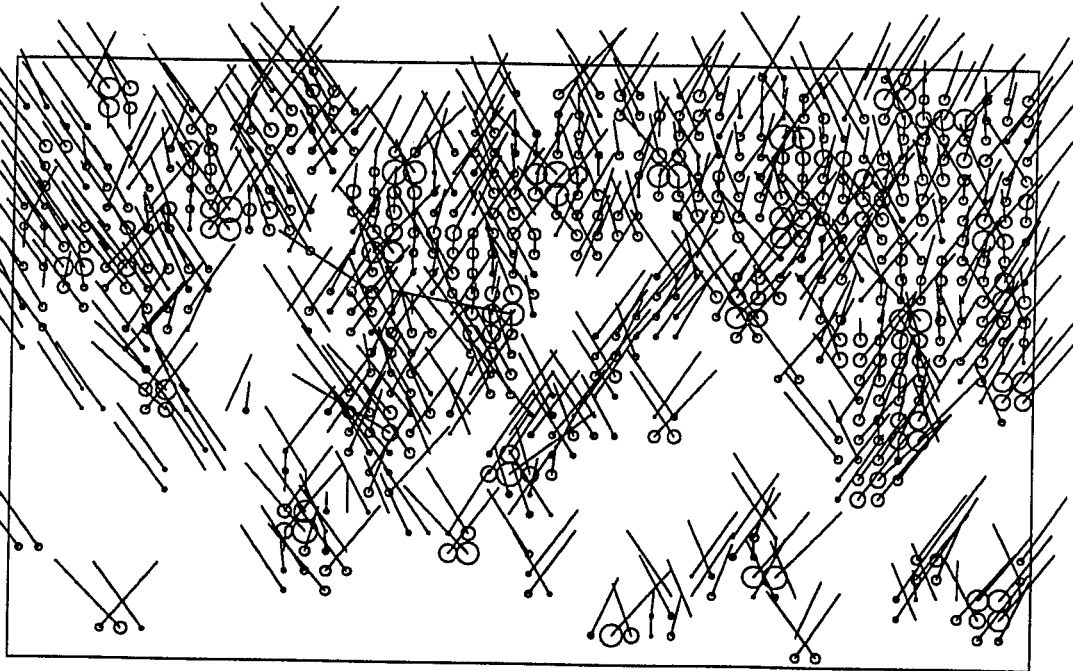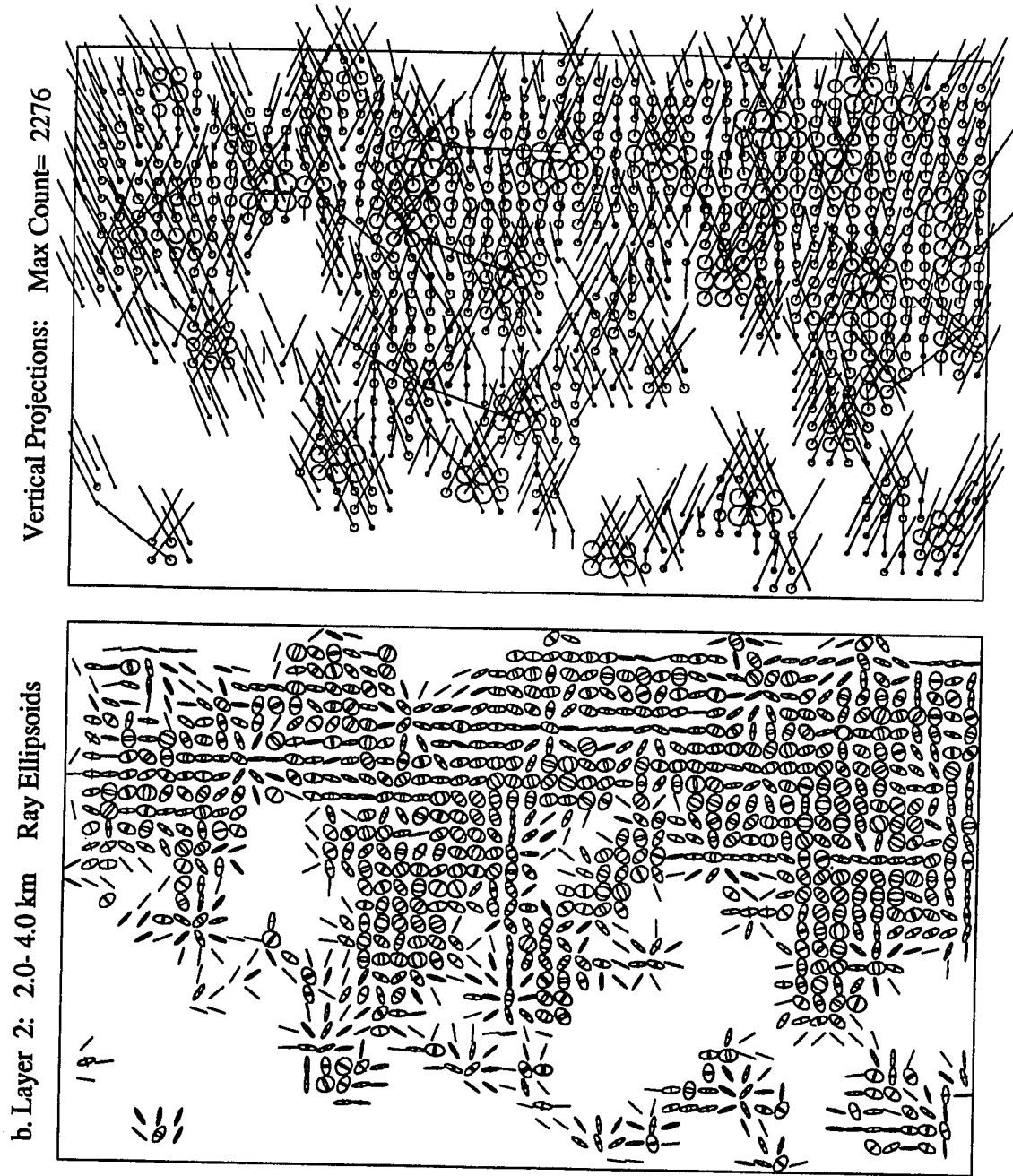
Vertical Projections: Max Count= 1467

c. Layer 3: 4.0- 6.0 km   Ray Ellipsoids

Vertical Projections: Max Count= 952

d. Layer 4: 6.0-9.0 km Ray Ellipsoids

Vertical Projections: Max Count= 617

e. Layer 5: 9.0-12.0 km   Ray Ellipsoids

Figure 6.13: (continued)

Vertical Projections: Max Count= 433

f. Layer 6: 12.0-16.0 km   Ray Ellipsoids

Vertical Projections: Max Count=   533

g. Layer 7: 16.0-20.0 km   Ray Ellipsoids

Figure 6.13: (continued)
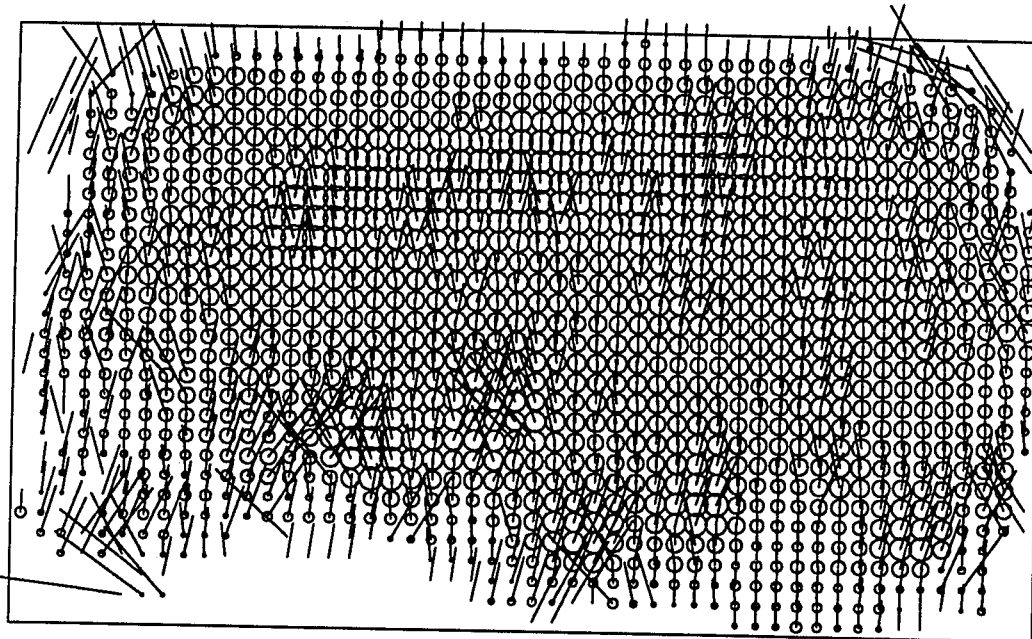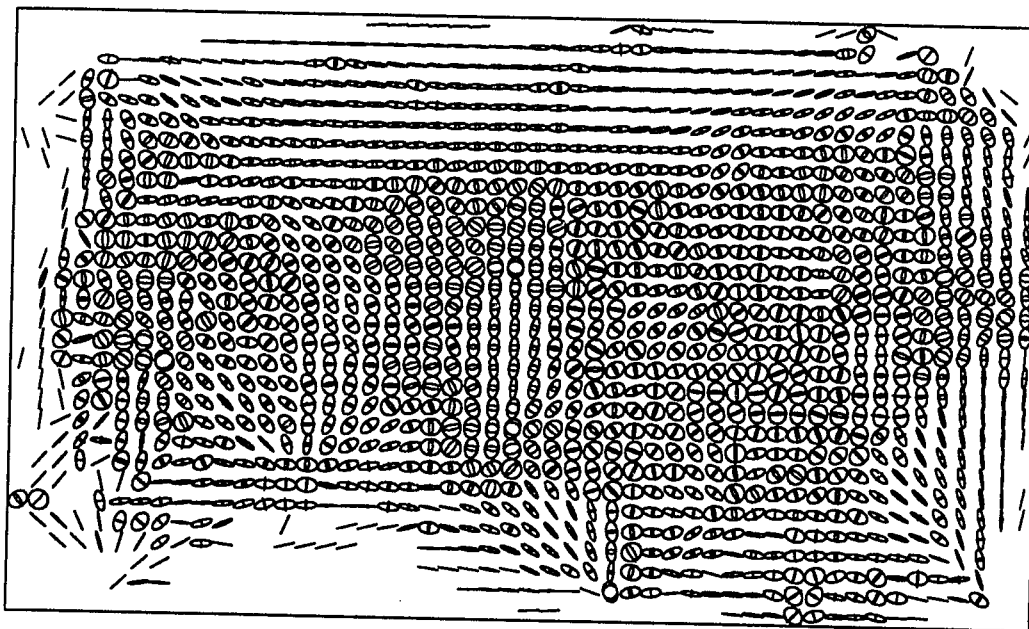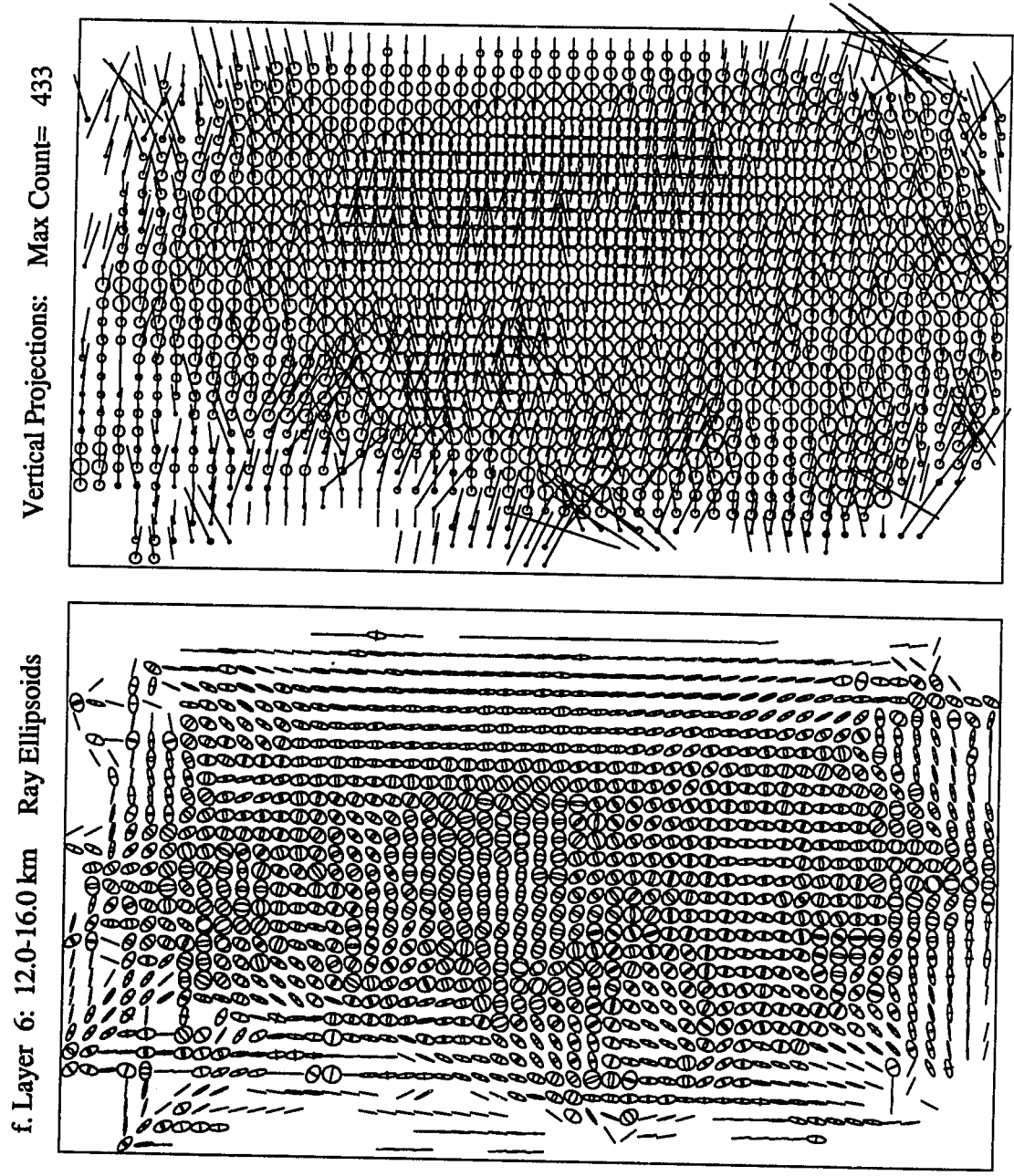
119
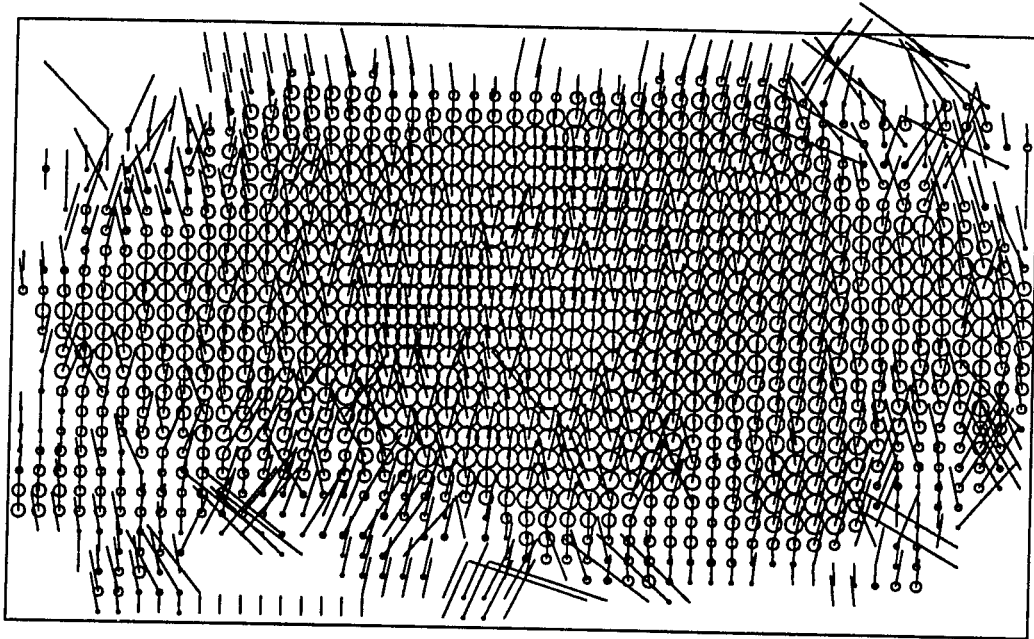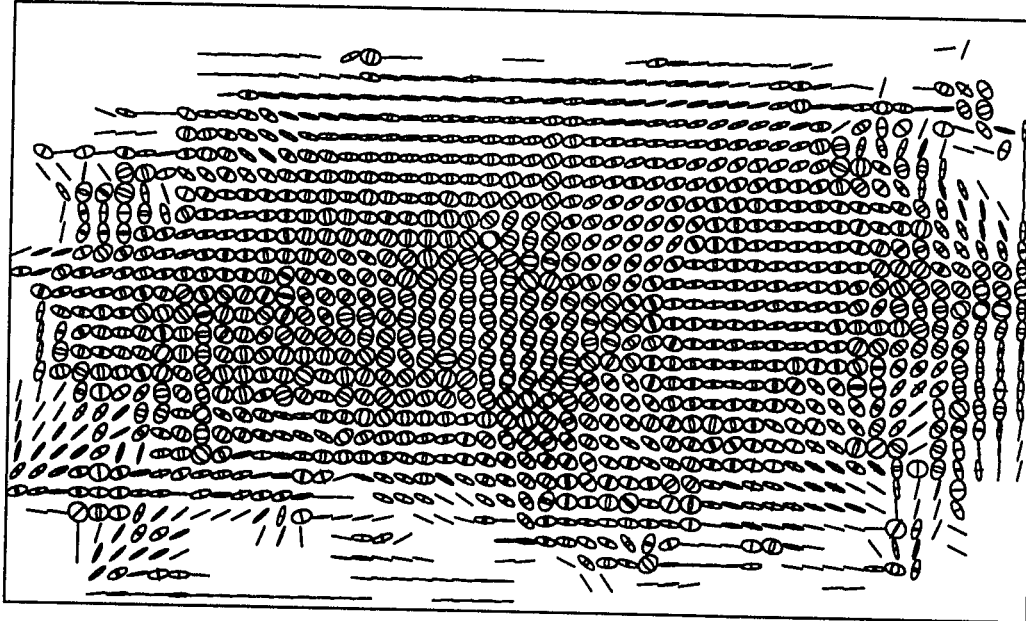
Vertical Projections: Max Count= 573
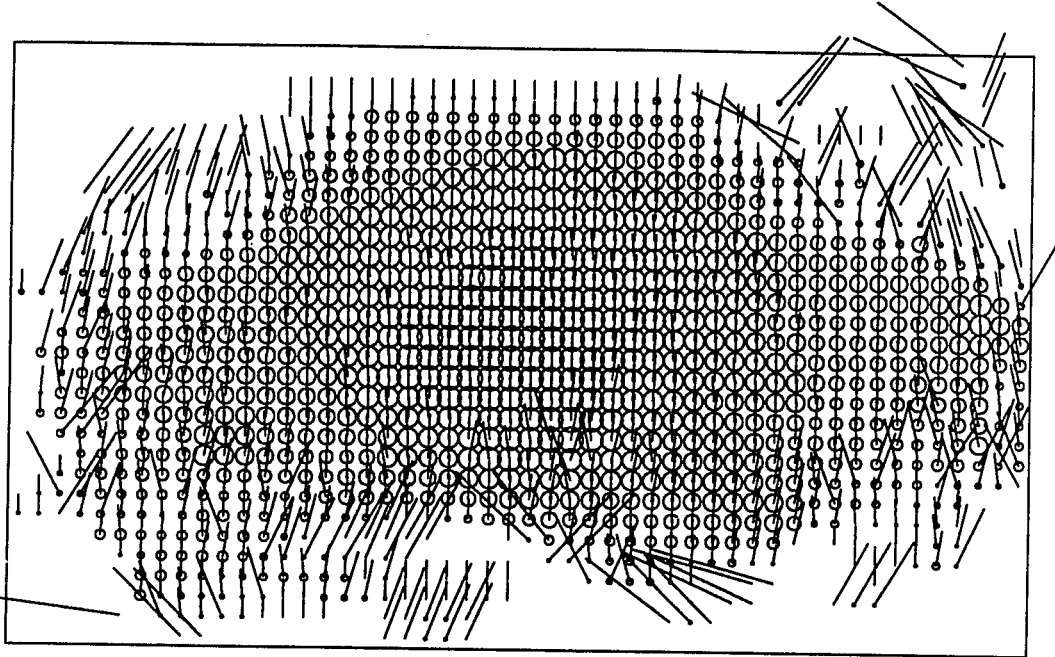
h. Layer 8: 20.0-25.0 km   Ray Ellipsoids
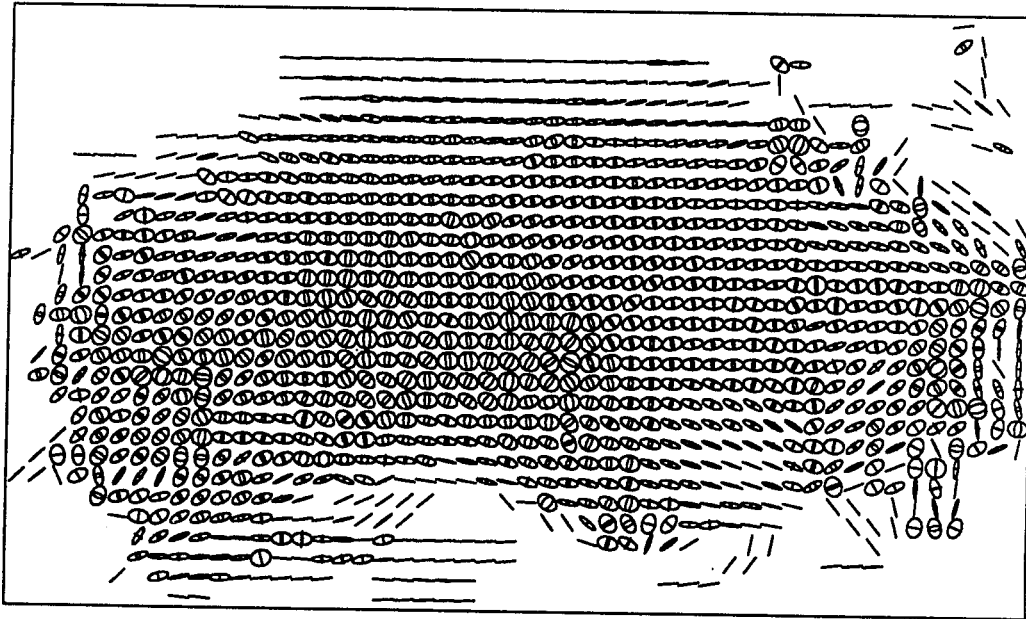
Vertical Projections: Max Count= 586
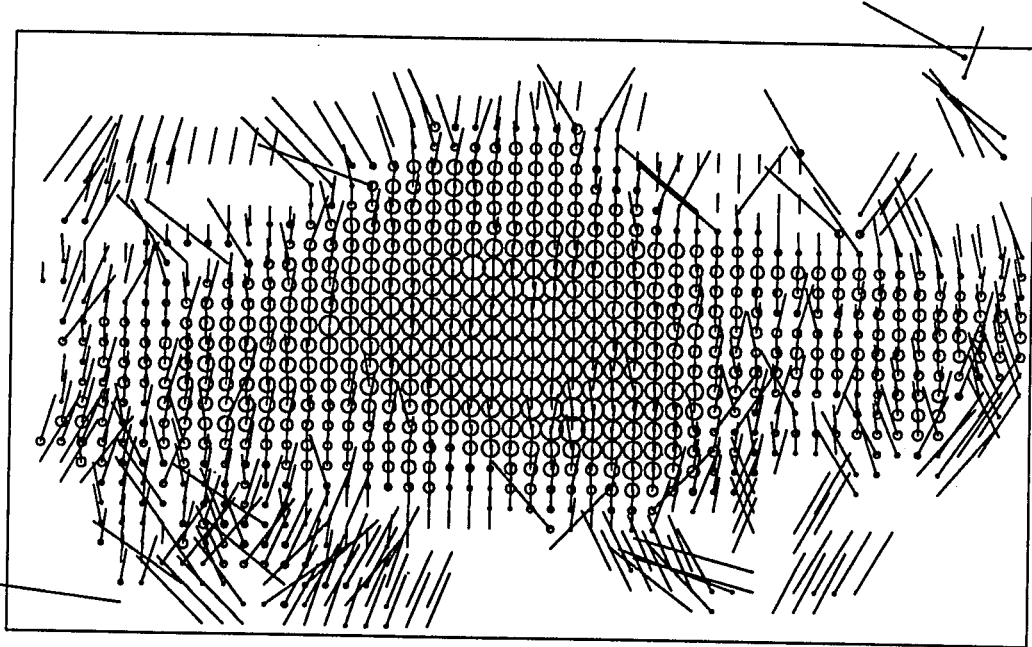
i. Layer 9: 25.0-32.0 km   Ray Ellipsoids
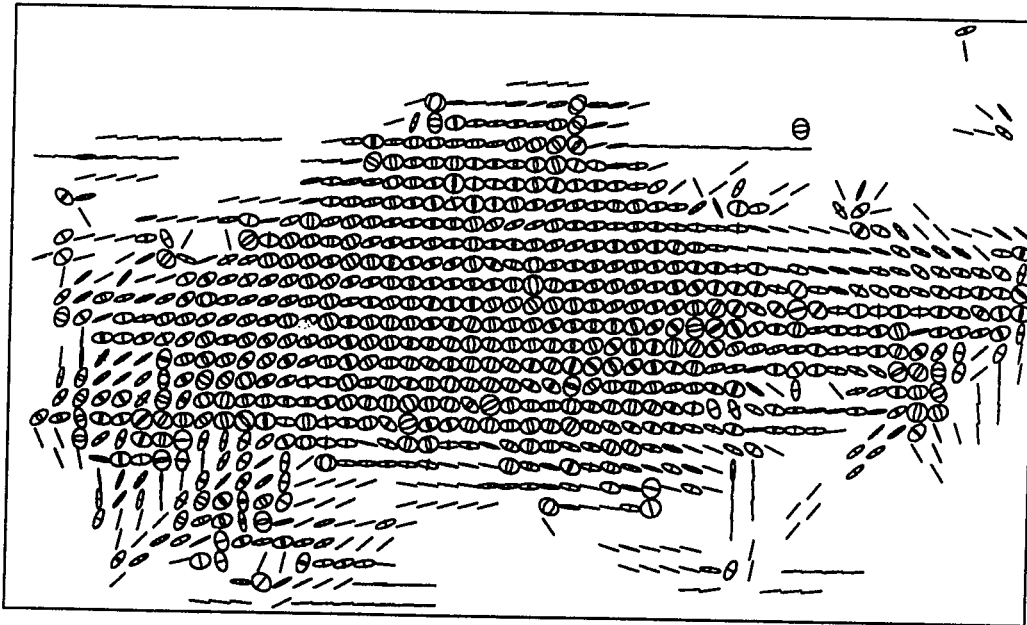
Figure 6.14(a-i): Jackknife estimates of the standard errors for the Puget Sound inversion. The jackknife was performed with 30 partitions of the data inverting with the LSQR algorithm where the Laplacian was constrained for regularization. Error assessment should be used in conjunction with resolution estimation for overall determination of the uncertainties in the inversion results.

Jackknife Errors

a. Layer 1: 0.-2.0 km

b. Layer 2: 2.0-4.0 km

c. Layer 3: 4.0-6.0 km

% Slowness

0. 1.7

0. 2.7

0. 2.9

0 50.0 km

Figure 6.14: (continued)

123

Figure 6.14: (continued)

CHAPTER 7

TOMOGRAPHIC INVERSIONS AT MOUNT ST HELENS

Introduction and Geology

Since the explosive eruptions of May, 1980, the Mount St. Helens region (Figure 7.1) has generated intense interest and speculation regarding the nature of volcanic activity in relationship to subduction in the Pacific Northwest. It is imperative to know the location of magmatic sources, the conduits and transport systems that supply the volcanos, and their relation to adjacent geologic structure if we attempt to determine constraints on the placement of volcanic activity in the Cascade range. While surface geological studies [e.g., *Evarts et al.*, 1987; *Mullineaux and Crandell*, 1981; *Phillips*, 1987; *Swanson and Clayton*, 1983; *Hammond*, 1980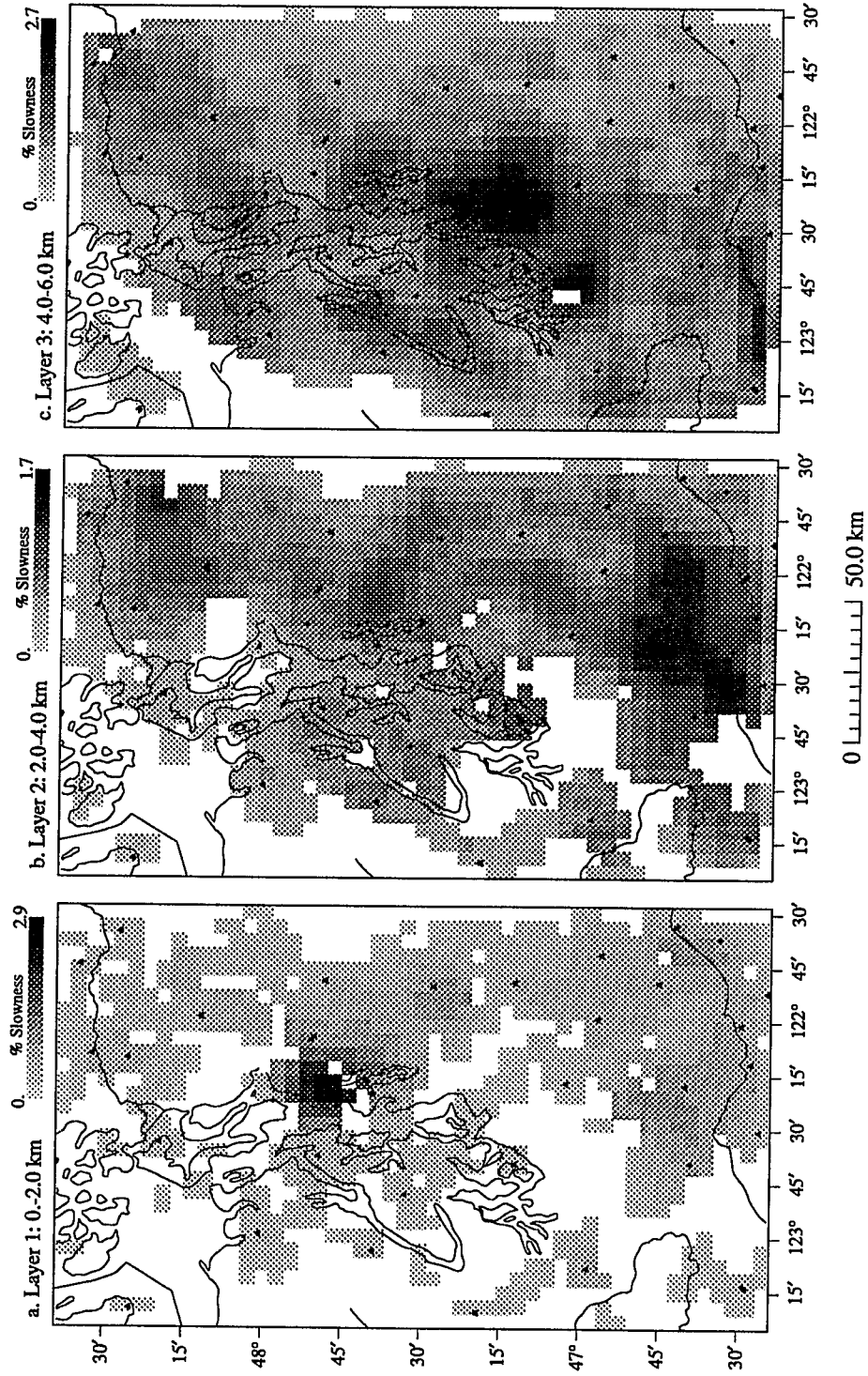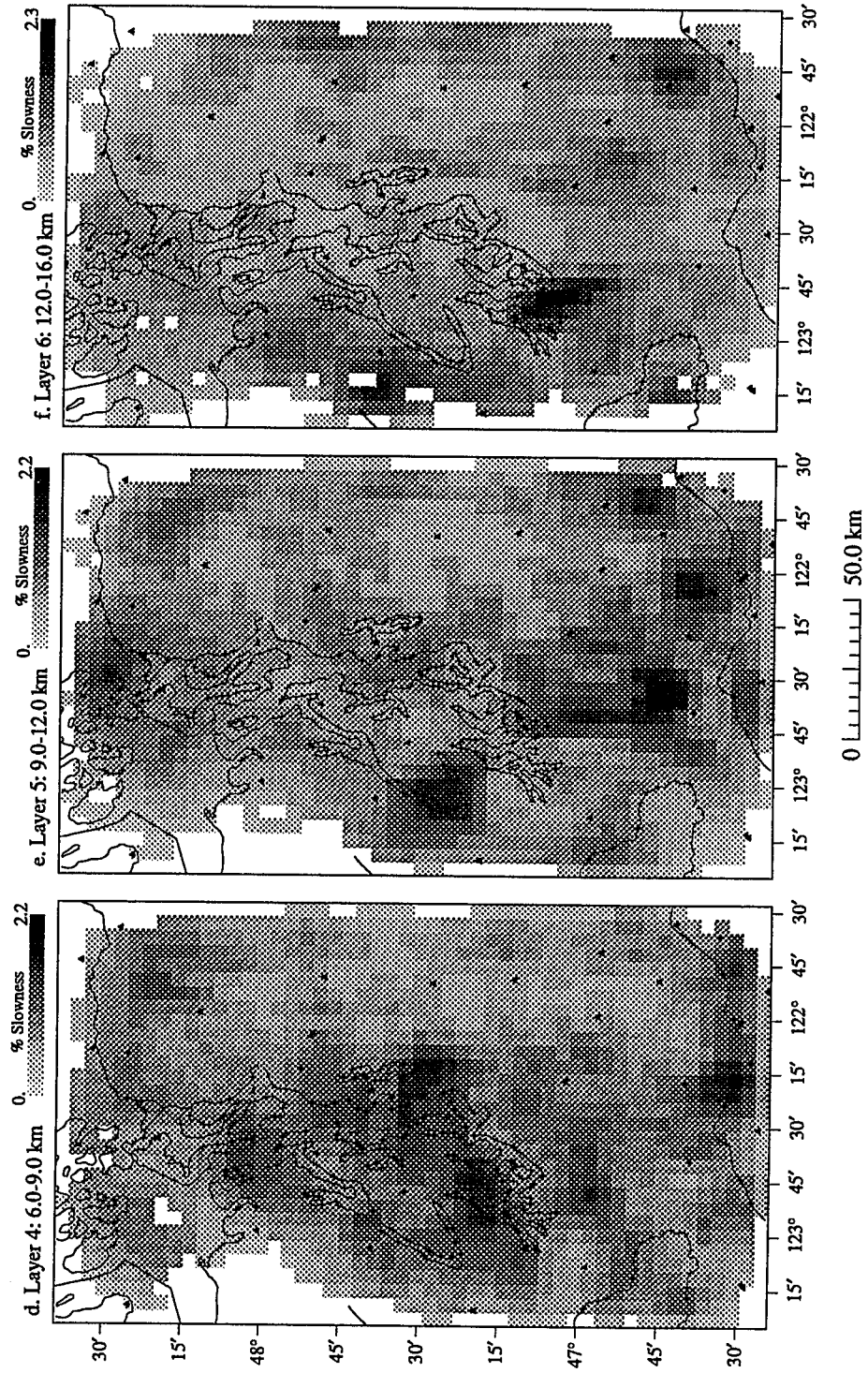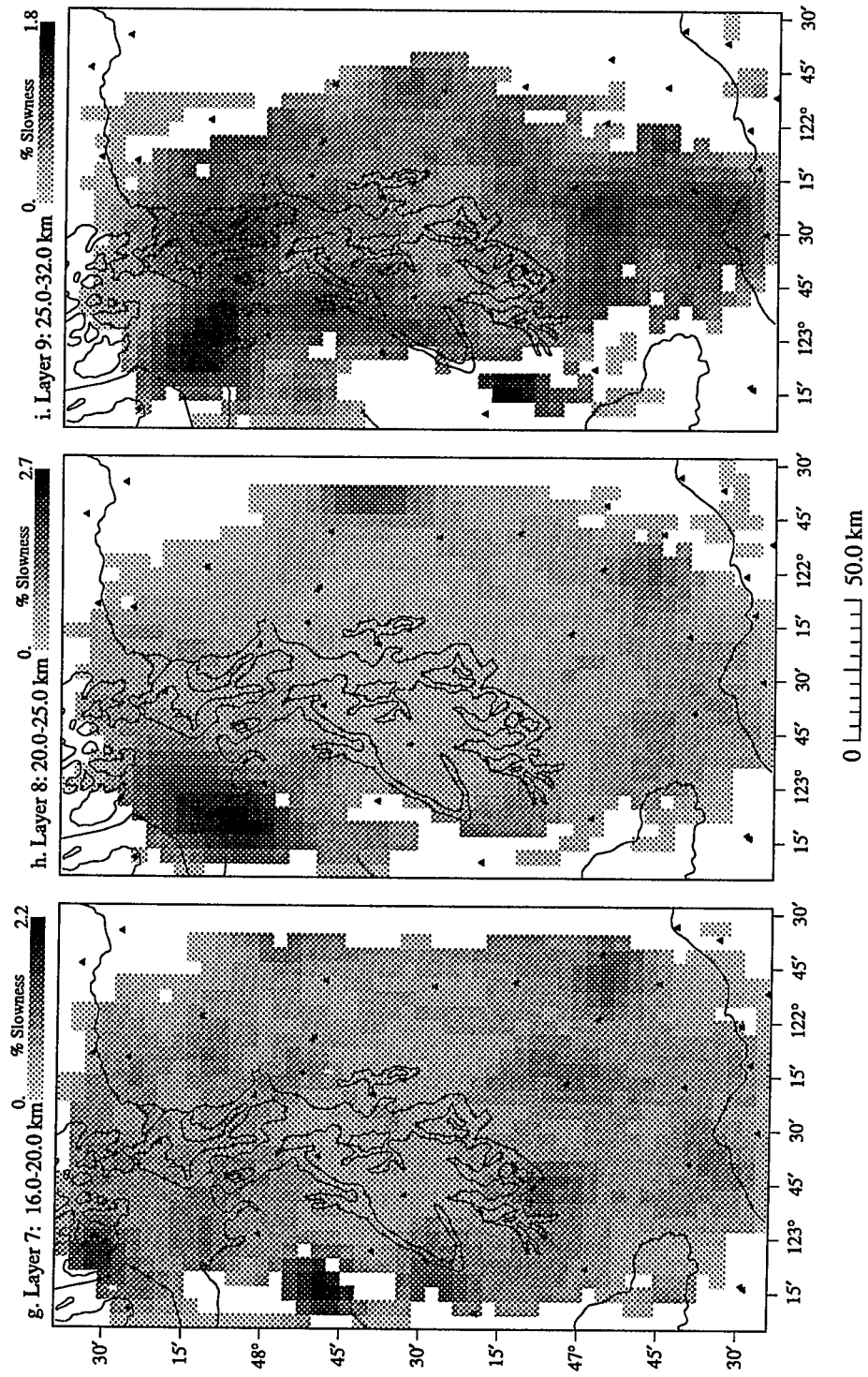] are useful in determining some of the features present, extensive erosion, lava coverage and forestation impedes a clear understanding of the complex structures in this region. Subtle variations in rock properties are measurable by various geophysical means and they remain important avenues for exploring this problem.

The geology of the Mount St. Helens region is summarized in reports by *Evarts et al.* [1987], *Finn and Williams* [1987], *Williams et al.* [1987], and *Stanley et al.* [1987]. The region is dominated by a succession of tectonic, igneous and erosional events that determined the present complex of basalts, andesites and quartz-diorites that are exposed at the surface. Forearc basins were present beginning in the late Jurassic to early Cretaceous and frontal arcs have been compressed against the continental margin since the early Miocene [*Dickinson*, 1976]. During the Eocene, volcanism was intermittent, but in the Oligocene and early Miocene there was a continuous span of volcanic activity [*Evarts et al.*, 1987] resulting by mid-Miocene in a well defined line of composite volcanos. The modern range of

Cascade volcanos dates to approximately 1 Ma and Mount St. Helens itself began producing dacitic lavas about 40,000 years ago [*Mullineaux and Crandell*, 1981]. In the mid Miocene there were periods of extensive plutonism when the Spirit Lake and Spud Mountain plutons of quartz-diorite were intruded north of Mount St. Helens [*McBirney*, 1978]. To the southwest of St. Helens there appears to be a series of dacite plug domes that predate the modern volcano [*Mullineaux and Crandell*, 1981]. Ten kilometers to the SSE lies the extinct Marble Mountain volcano, one of several late Pleistocene volcanos that have erupted in close vicinity to Mount St. Helens [*Williams et al.*, 1987].

Several geophysical studies have recently been published focusing on the Mount St. Helens region [*Finn and Williams*, 1987; *Williams et al.*, 1987; *Stanley et al.*, 1987; *Weaver and Smith*, 1983]. in attempts to clarify the tectonics and to delineate subsurface geologic features. Analysis of aeromagnetic data [*Finn and Williams*, 1987] and gravity data [*Williams et al.*, 1987] indicates the presence of an anomalous, intrusive unit beneath the present volcano. Using conductivity data, *Stanley et al.* [1987] concluded that this anomalous unit overlies a unit of sedimentary rocks described by *Dickinson* [1976]. *Stanley* [1987] further describes a large conductivity anomaly that trends north of St. Helens paralleling the so called St. Helens seismic zone (SHZ) as defined by *Weaver and Smith* [1983].

In this study, the target area for the Mount St. Helens inversion was an 80 by 80 km square between 45.9°-46.62° north latitudes and 121.62°-122.67° west longitudes. The criteria for selecting data was the same as for the Puget Sound inversion. The areal distribution of sources is displayed in Figure 7.2(*a*) and the projection of the hypocenters on an east-west vertical cross section is illustrated in Figure 7.2(*b*). All rays not entirely within the target model were excluded as were those considered to be erroneous outliers (residuals greater than 2 seconds). A large number of aftershocks of the May, 1980 eruption were excluded as were

some stations that had only a small number of rays (<30). The remaining data set included 2,023 hypcenters recorded at 21 stations giving rise to 17,659 rays. The heaviest coverage was in the range of 2 - 10 km depths and the deepest event used was located at nearly 40 km depth. The earthquake coverage was not homogeneous due to the preponderance of events located in the SHZ. This was partially compensated for by weighting as described in chapter 1. The model was partitioned into blocks 2 by 2 km square with depth divisions determined by the initial reference model (Table 1).

Results of Inversion

In Figures 7.3 (*a-h*) we show the results of the tomographic inversion using LSQR for 40 iterations and $\lambda = 1600$ for a 23% reduction in misfit. For comparison results using ART with $\lambda = 260$ for 30 iterations is presented in Figure 7.4(*a-h*). For each case regularization was implemented by either smoothing after iterations as in ART or constraining the Laplacian in LSQR. Each figure represents the slowness perturbation in a horizontal layer of the model. Note that the layers are of slightly different thickness, although each block is 2 km × 2 km horizontally. The grey shading represents levels of percent perturbation from the reference slowness of the layer. Since velocity is the reciprocal of slowness, dark areas (high slowness) represent low velocity and light areas are relatively high velocity. Regions or blocks which are unshaded have not been sampled by any rays, indicating the extent of ray coverage. The outline of the preeruption tree line around Mount St. Helens as well as the stations have been plotted for geographical reference.

Anomalies present in the top layer (Figure 7.3*a*) have larger amplitude and shorter wavelength than we expect are resolvable. This is due to the fact that all the rays must terminate at one of the stations at the surface and large structures

result in the immediate vicinity of the stations. We take these to represent local structure beneath the station that has not been accounted for in the station correction described in the previous section. For this reason the discussion below emphasizes the structure below the surface layer.

Discussion and Conclusions

The prominent anomalies apparent in the 3-D model show a remarkable correlation with the surface geology and geophysical measurements. *Evarts* [1987] has mapped out surface expressions of such features as the Spirit Lake pluton trending NNE from the crater of St. Helens and the Spud Mountain pluton which trends NNW. A lineation of plug domes and extinct vents run in an east-southeast direction from the crater and take a slight bend to the northwest north of St. Helens. These features are apparent on the aeromagnetic map of Finn and Williams (reproduced here as Figure 7.5) where high magnetic anomalies are recorded at Spirit Lake and Spud Mountain and a low trough is evident in between, at the location of the St. Helens Seismic Zone (SHZ) [*Weaver and Smith*, 1983]. The seismic anomalies in the first 3 layers (0-6 km, Figures 7.3*a-c*) of the 3-D model presented here reflect the same structures. A high velocity zone maps directly into the two plutons located north of St. Helens with a prominent low velocity zone at the SHZ separating them. The Goat Mountain complex is evident as a low velocity region. An enlarged version of Figure 7.3*b* is provided in Figure 7.6, showing these features superimposed on the slowness anomalies for clarity. At the crater of St. Helens we see a low amplitude high velocity region with a low velocity directly to the north where the SHZ meets the crater. This anomaly appears to have reversed by a depth of 6-9 km (Figure 7.3*d*) where a strong, slow anomaly is evident under the mountain, slightly displaced to the south. In the deeper layers (9-16 km depth) there is a strong low velocity anomaly beneath the

crater which may be indicative of modern magma accumulations.

The low velocity anomaly that runs along the SHZ is apparent at virtually all depths. In the deeper layers it is the major outstanding feature in the model. This slower zone coincides with high seismicity and is presumed to be the lineation of a crustal zone of weakness characterized by strike-slip faulting extending north-northwest from Mount St. Helens [*Weaver et al.*, 1987]. *Weaver et al.* [1987] suggest that Mount St. Helens is situated at the corner of a bend in the SHZ where it trends to the southeast towards Marble Mountain. The velocity anomalies in our model below the first layer show little sign of such a trend.

There is strong evidence that the boundaries delineated by the seismic velocity inversions represent important subsurface interfaces separating rocks of significantly varying properties. These structures are evident in the surface exposure and in other geophysical studies, such as the aeromagnetic and gravity surveys. If we assume that zones of low velocity reflect the presence of fracturing, magmatic intrusion, fluids, or high heat source, then the SHZ north of Mount St. Helens may be the major tectonic feature of this region. We speculate that this zone is a weak boundary, separating distinct geologic terrains, that is reactivated under regional tectonic stress. This would explain the affinity of both earthquakes and volcanic conduits (e.g. Mt. St. Helens) for this zone. The vents and plug domes of the Goat Mountain complex also appear to exhibit low velocities, however, the Marble Mountain region exhibits low velocity structure only in the shallowest zones (0-2 km). At depth (6-12 km) the low velocity anomaly appears due south of the crater and shows little evidence of bending to the southeast. The high velocity plutons at Spirit Lake and Spud Mountain represent consolidated, brittle material that is significantly more impervious to fluid injection. These plutons are separated by the strike-slip faulting of the SHZ. Because the presence of the quartz-dioritic plutons appears to be consistent in the deeper layers they may

constrain the zone of weakness in the crust that produces the SHZ. In other localities, this technique of using local earthquakes for tomographic imaging may prove to be useful in locating areas of crustal weakness and thus potential seismic hazards. Such information should also prove of great value in studying the processes and structure of volcanic systems.

The comparison of the ART and LSQR inversion shows that with real data the two techniques produce, at least for the broad structures, the same results. The ART inversion appears to have slightly higher frequency content due to the difference in regularization, although both inversions explain the data to approximately the same percent reduction of misfit (24%). For the following discussion the ART methods were used in the analysis of resolution and error.

Resolution and Errors

An impulse response kernel for a block located in layer three directly north of St. Helens is displayed in Figure 7.7. A spike of value 1 was put into the block that is outlined in white (Figure 7.7b) and artificial data were generated using the rays of the real data. The output was then inverted in the same fashion as the actual data. While it is quite evident that the lateral resolution in this layer (Figure 7.7b) is excellent (3 blocks) near the center of the model. Because there are relatively few layers, resolution in the vertical dimension is more difficult to achieve. The anomaly is spread into the layers above and below in a north by north easterly direction. This is due to the high seismicity located in the SHZ where blurring has occurred. Note, however, that the amplitude of the anomaly in the layers above and below the spike are attenuated by about a third of the height of the center block. This resolution is typical of the central portion of the model in layers 2-5. On the edges the resolution is degraded. This can be seen in the plots of ray ellipsoids for the Mt. St. Helens data set (Figure 7.8a-h) The

concentration of seismicity in the SHZ is indicated by the plots of ray density (Figure 7.8*a-h*) indicates that there is likely to be blurring of the images along ray-paths emanating from this zone. Indeed, the plots of the vertical dip angle of ray ellipsoids (Figure 7.8*a-h*) shows that the major axes of the ellipsoids in this region have high dips. Assertions regarding the vertical structure in this region should be qualified by the lack of vertical resolution there.

In general, the errors in this model (Figure 7.9*a-h*) are not large enough to change the sign of the gross anomalies apparent in the inversion. For example, the anomaly at the Spirit Lake Pluton (46°30'N-122°W) is about -6.0%. But the predicted standard error here is very small, about .5%. If the errors are assumed to be uncorrelated, this would indicate that at least the sign of the velocity structure is well constrained by the 3-D inversion. Since the errors are correlated in our case (through, for example, earthquake mislocations) we expect the model errors to be slightly higher.
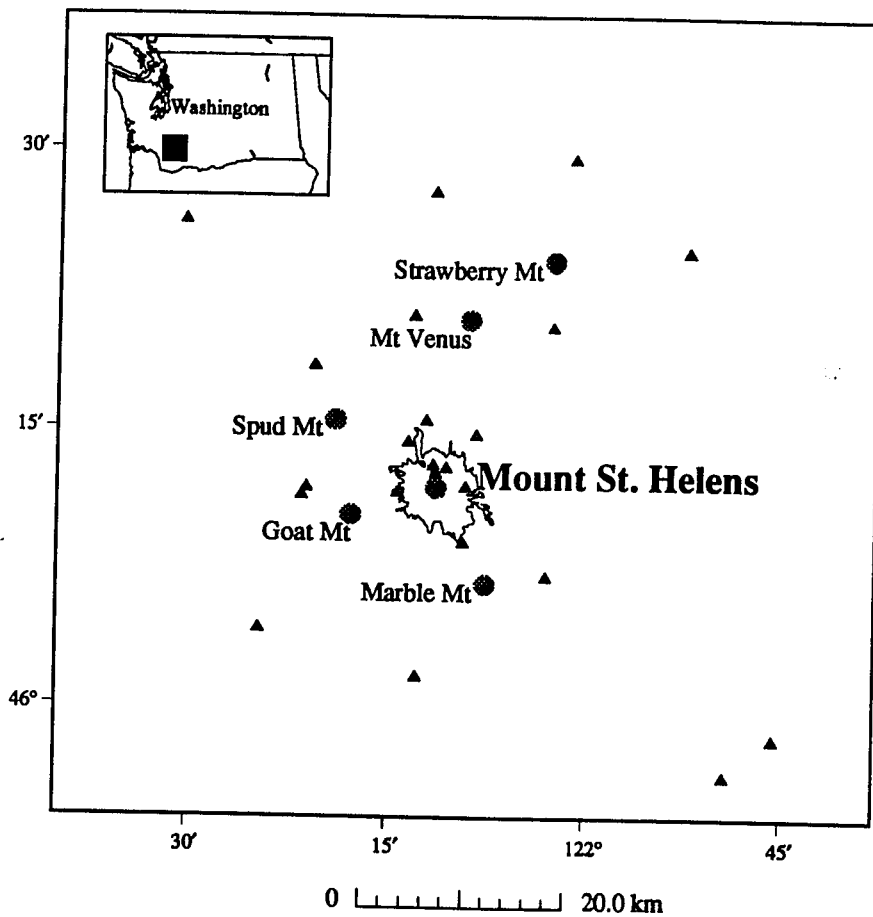
Figure 7.1: Map view of target area showing station distribution. The inset shows the location of target area in Washington. Triangles are station locations and labeled bullets identify major topographic highs in the area. The pre-1980 eruption treeline around Mount St. Helens is plotted for reference.
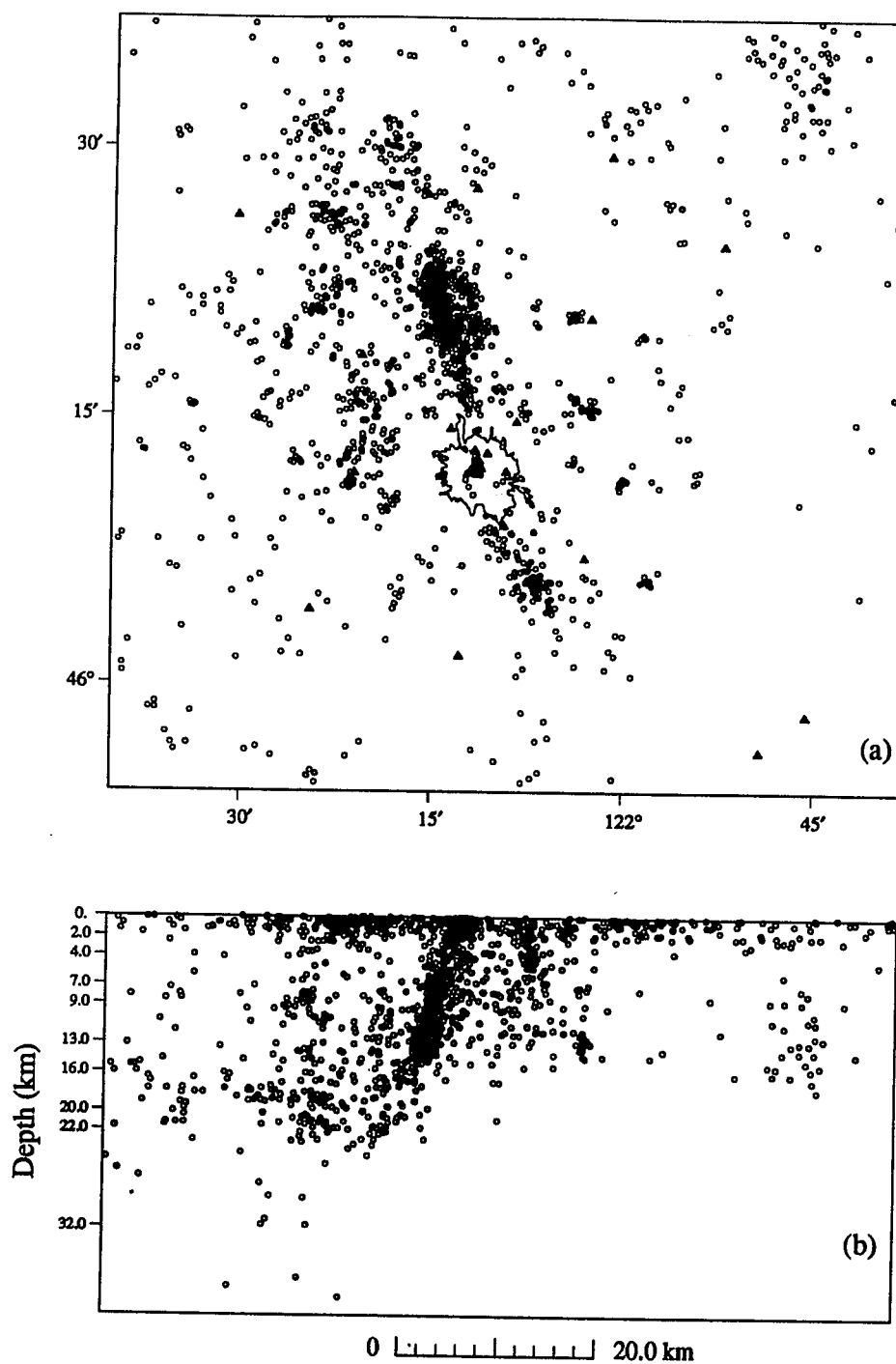
Figure 7.2: (*a*) Epicenter map and (*b*) east-west vertical projection of earthquake sources within target area. The SHZ is clearly evident by the high seismicity lineation extending north of Mount St. Helens.

133

Figure 7.3(*a-i*): 3-D tomographic inversion using LSQR for lateral variations in slowness perturbation. Each figure represents a horizontal slab in the earth. Grayshades represent percent perturbations from the reference model, such that light colored regions reflect high velocity and dark regions are low velocity. Dark triangles are station locations as illustrated in Figure 7.1.

Mt. St. Helens Inversion, LSQR

a. Layer 1: 0.-2.0 km

b. Layer 2: 2.0-4.0 km

134

Figure 7.3: (continued)

Figure 7.3: (continued)

e. Layer 5: 9.0-12.0 km

f. Layer 6: 12.0-16.0 km

Figure 7.3: (continued)

g. Layer 7: 16.0-20.0 km

h. Layer 8: 20.0-25.0 km

137
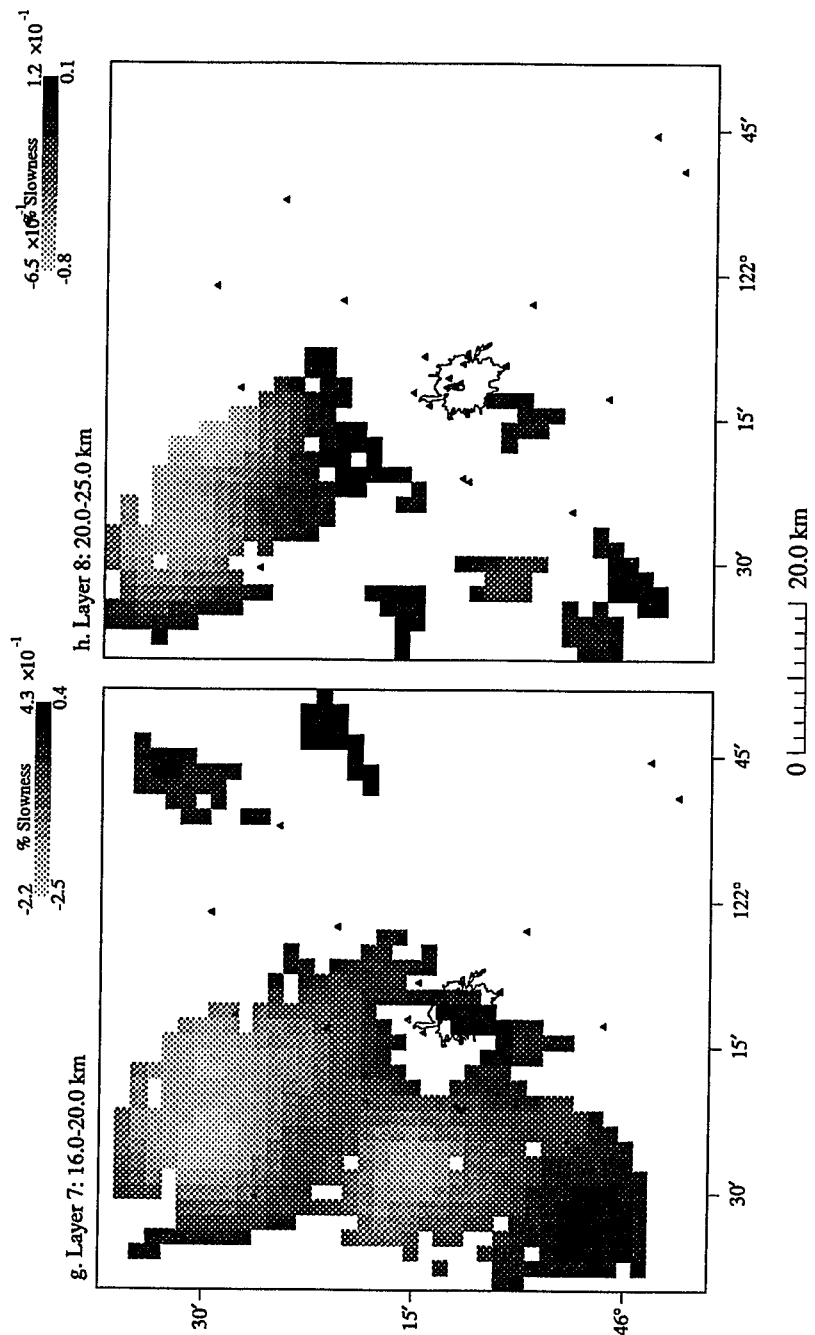
Figure 7.4(*a-l*): 3-D tomographic inversion using ART for lateral variations in slowness perturbation. Each figure represents a horizontal slab in the earth. Grayshades represent percent perturbations from the reference model, such that light colored regions reflect high velocity and dark regions are low velocity. Dark triangles are station locations as illustrated in Figure 7.1.
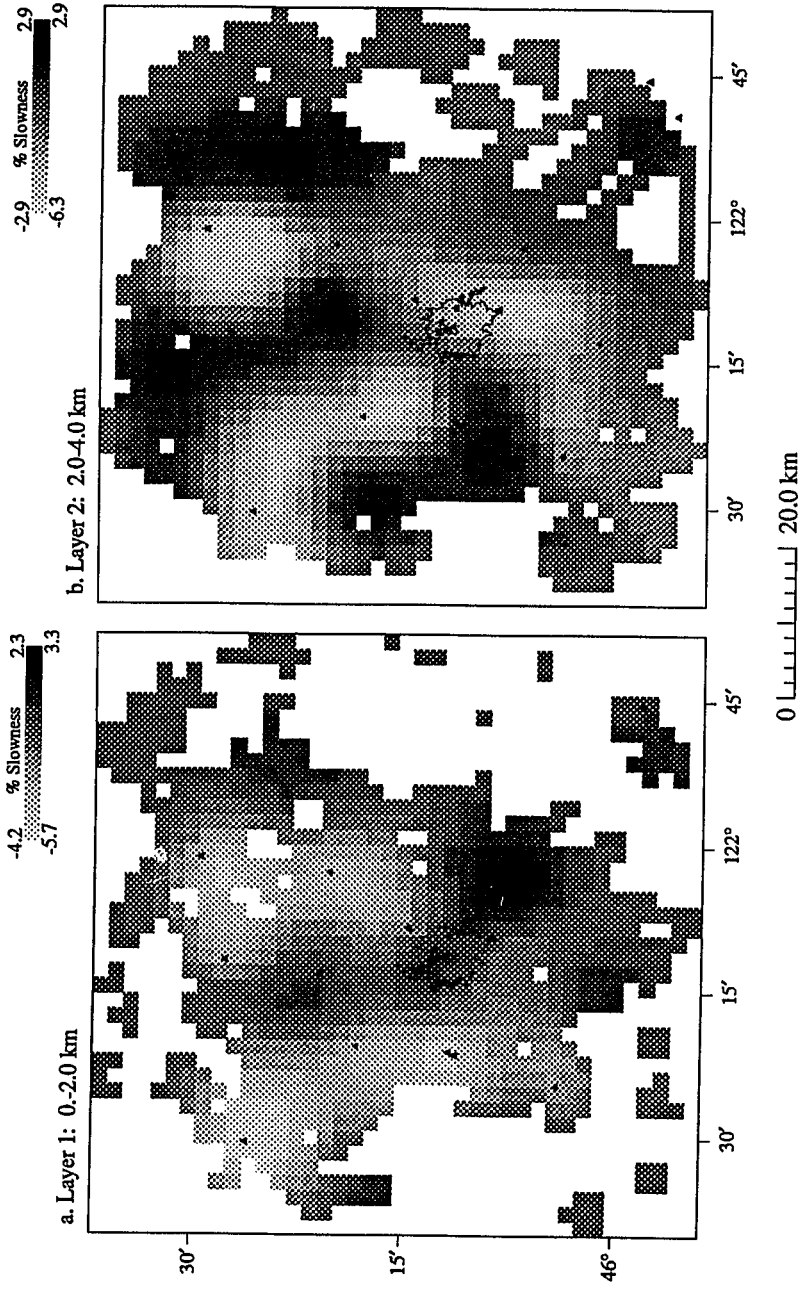
Mt. St. Helens Inversion, ART

a. Layer 1: 0.-2.0 km
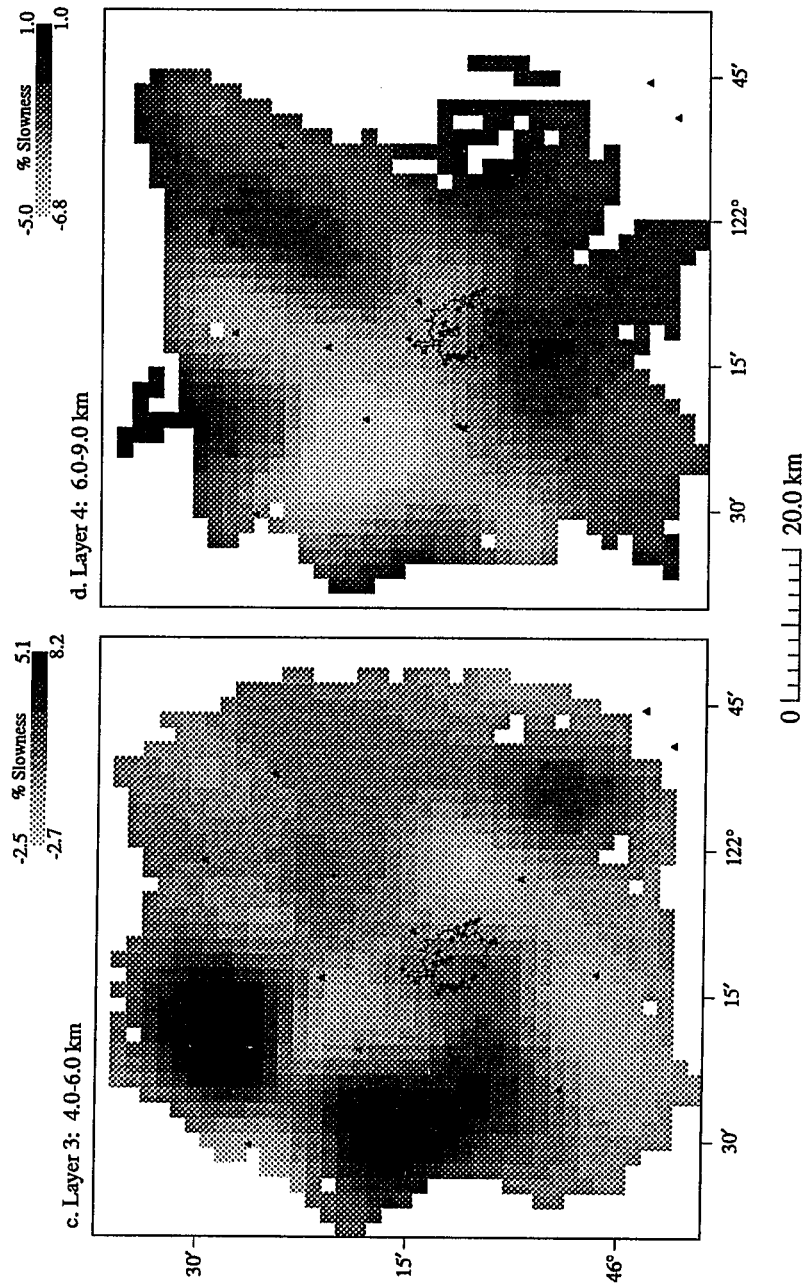
b. Layer 2: 2.0-4.0 km

0   20.0 km

Figure 7.4: (continued)

c. Layer 3: 4.0-6.0 km

d. Layer 4: 6.0-9.0 km

% Slowness

% Slowness

Figure 7.4: (continued)

e. Layer 5: 9.0-12.0 km

f. Layer 6: 12.0-16.0 km

Figure 7.4: (continued)

g. Layer 7: 16.0-20.0 km

h. Layer 8: 20.0-25.0 km
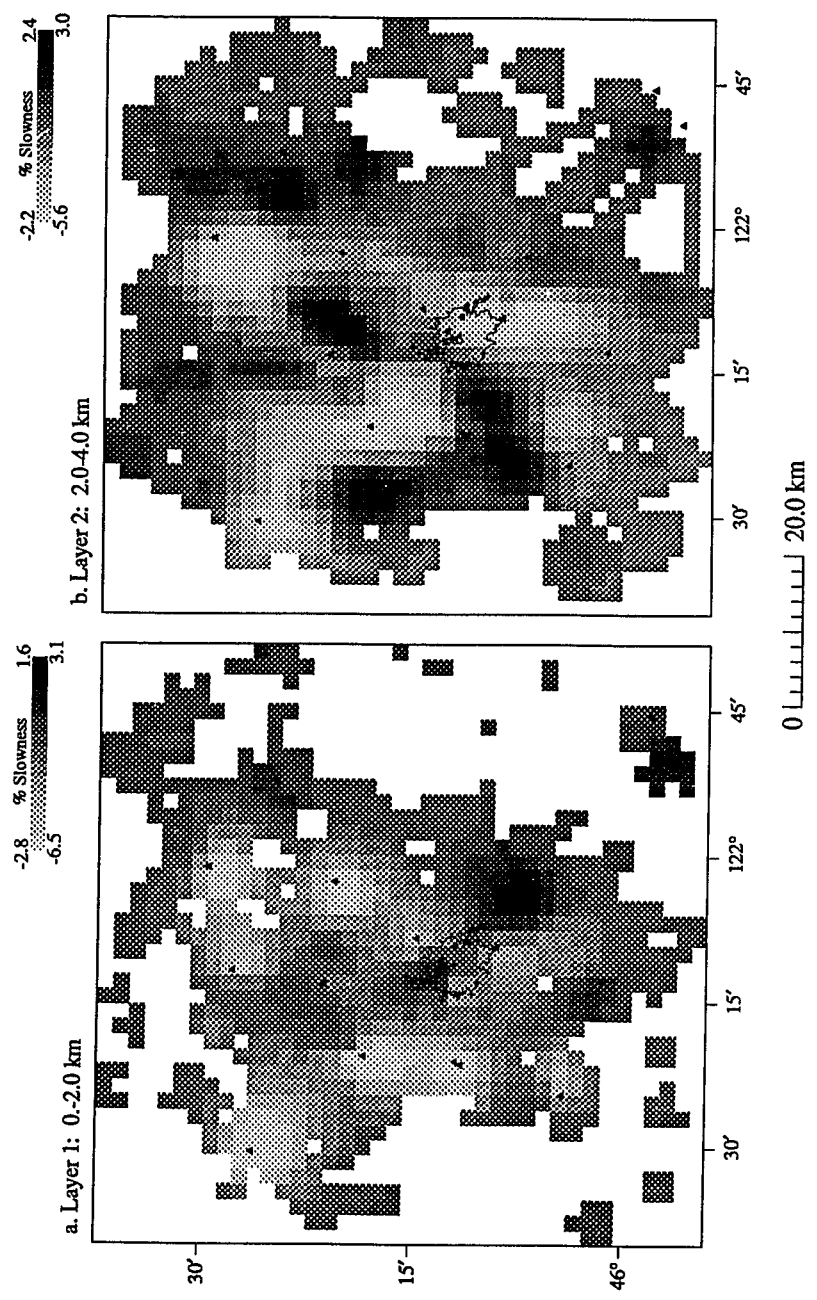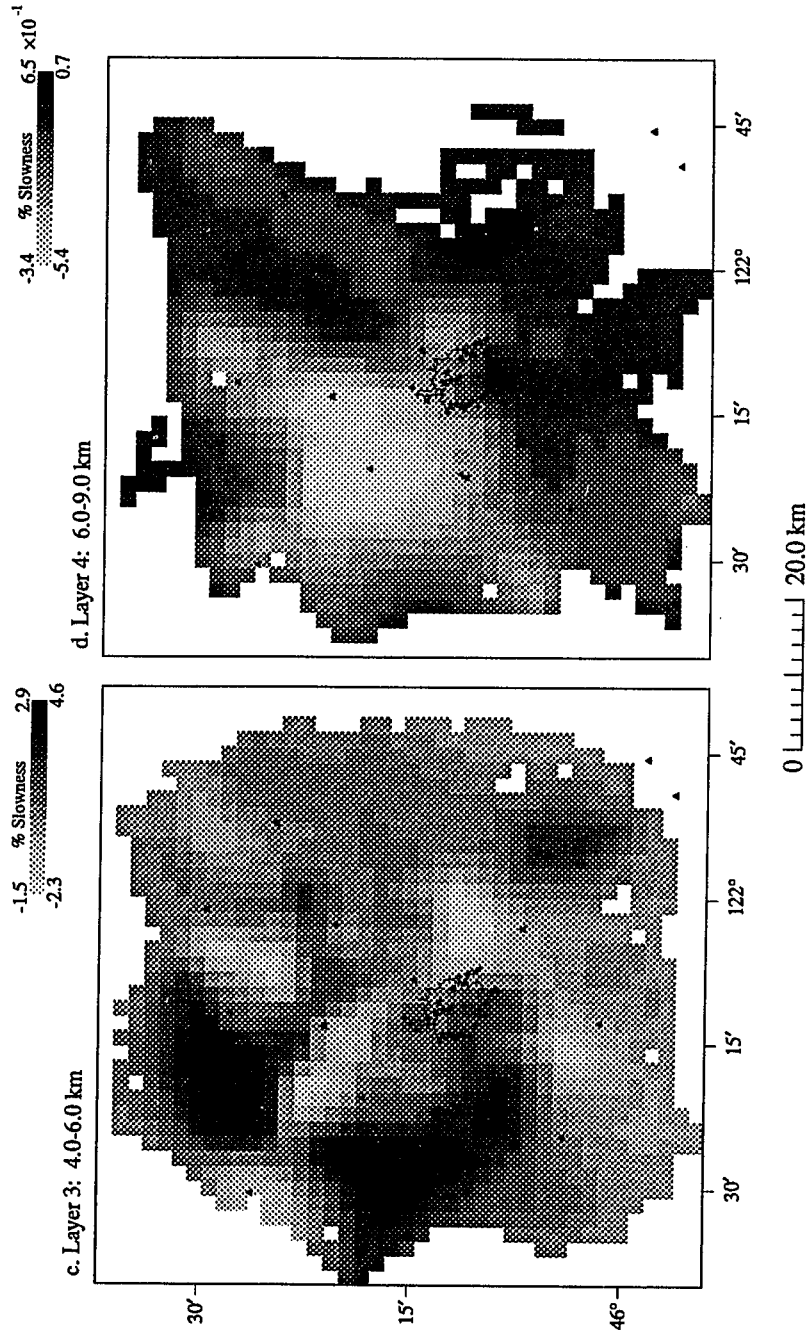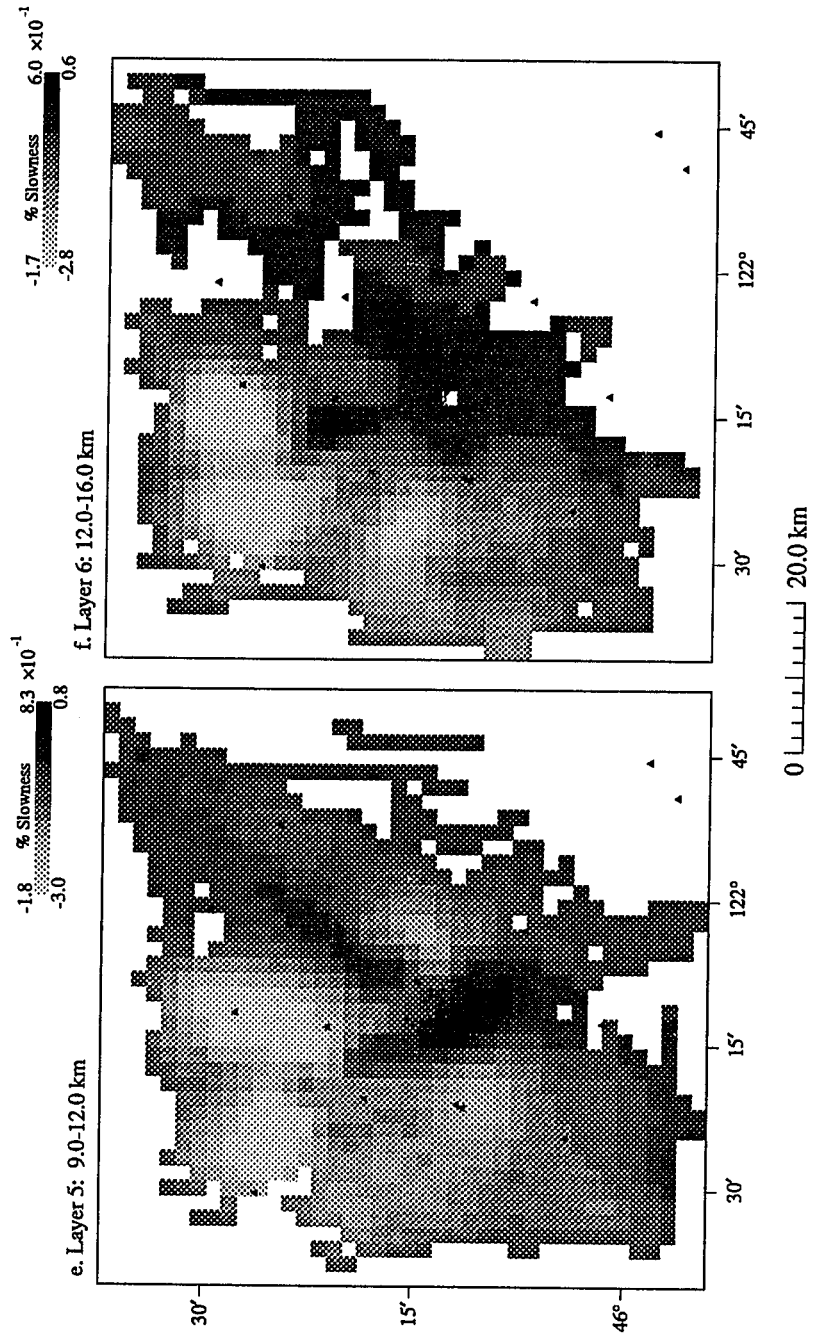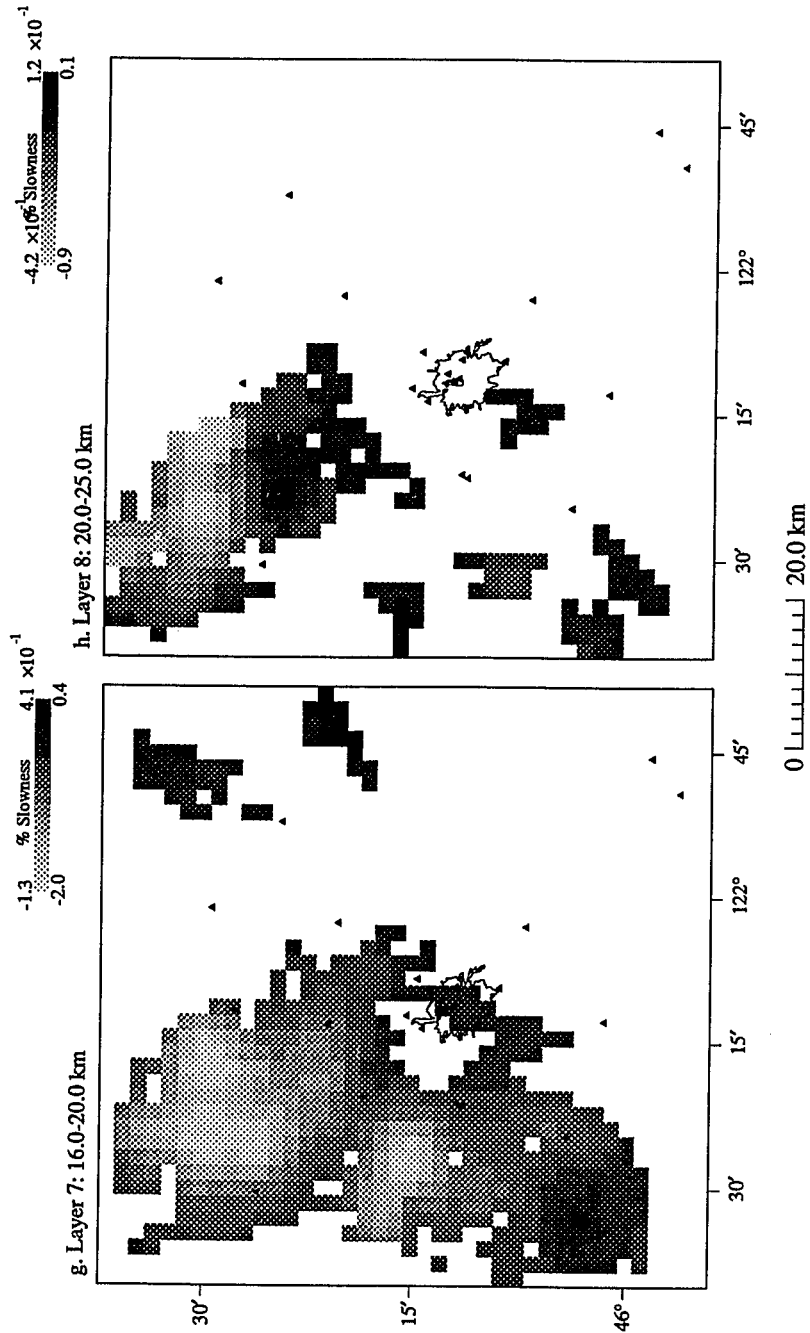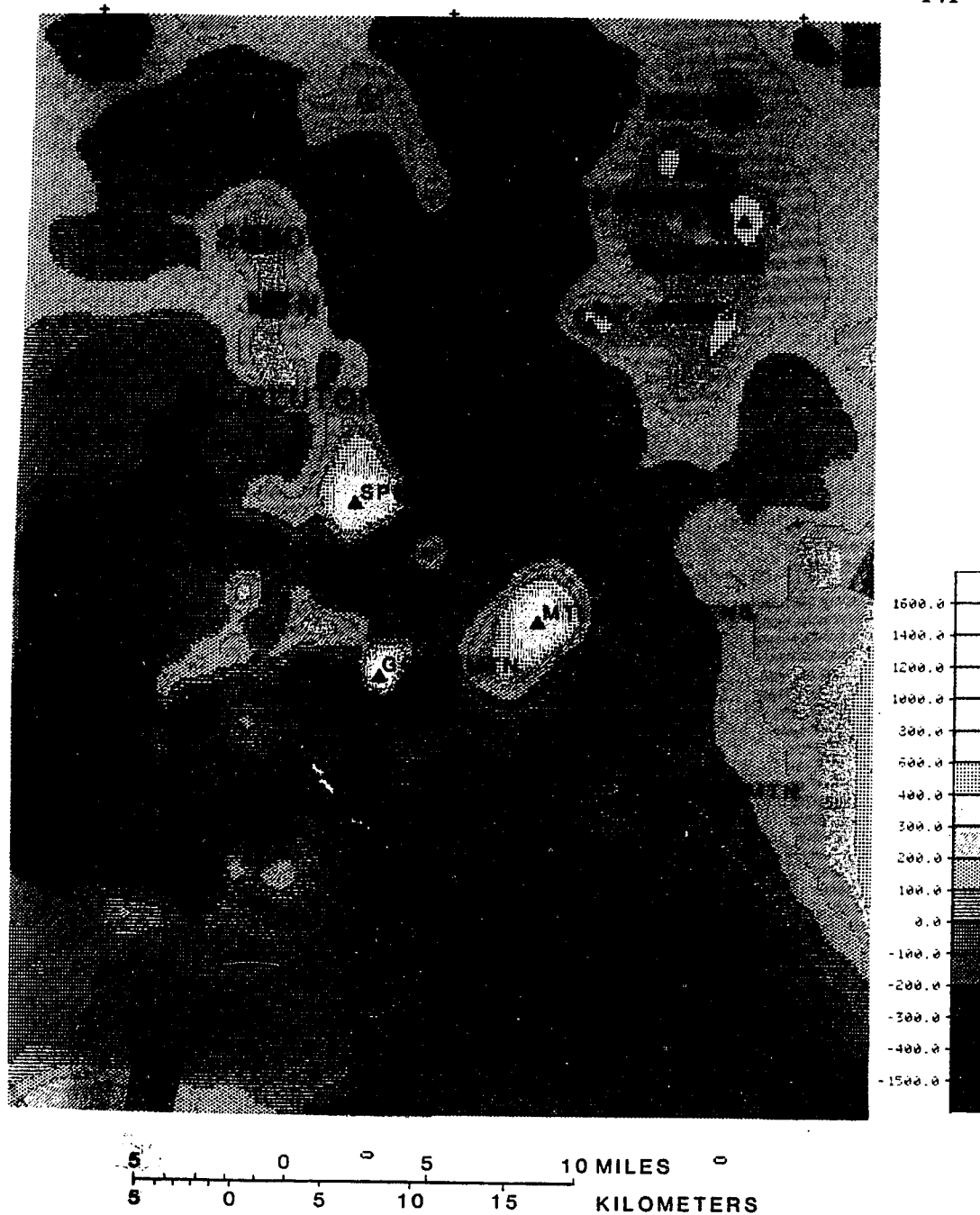
Figure 7.5: Regional aeromagnetic map reduced to pole [*Finn and Williams*, 1987]. Units are nT. The anomalies here correlate with the velocity anomalies at 2-4 km depth (Figure 7.3*b*).

Layer 2: 2.0 to 4.0 km

-8.0   % Slowness   5.9
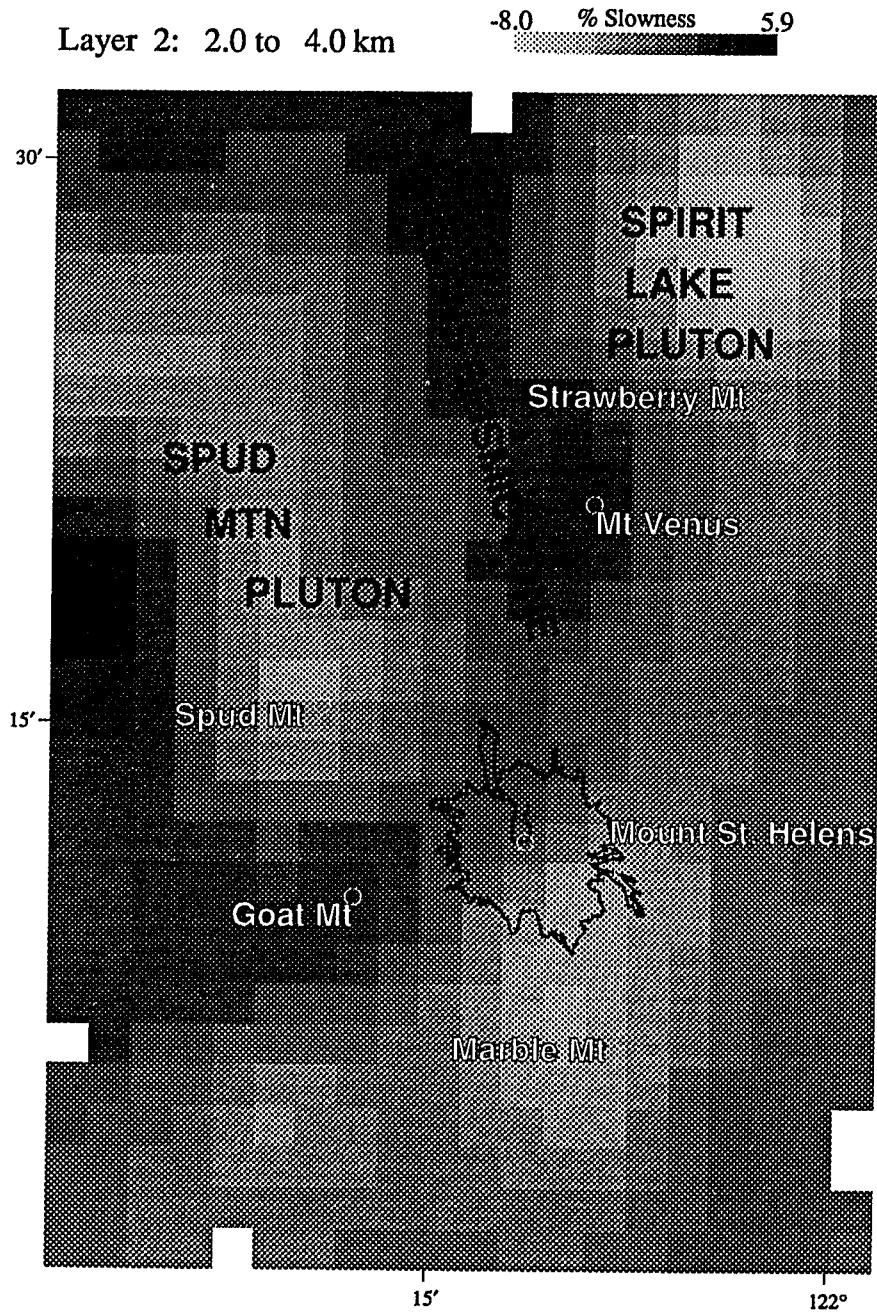
Figure 7.6: A close up of Figure 7.3*b* to match the scale of the aeromagnetic anomalies in Figure 7.5. Important features discussed in the text are labeled for clarity.

143

Figure 7.7(*a-c*): Resolution impulse response for highlighted block located in near the center of layer 2. Perfect resolution would produce 100% in this block, but after forward and inverse modeling only 15% of the amplitude remains with the rest spread throughout the model.

Resolution Kernel at Mt. St. Helens

a. Layer 1: 0.-2.0 km

b. Layer 2: 2.0-4.0 km

144
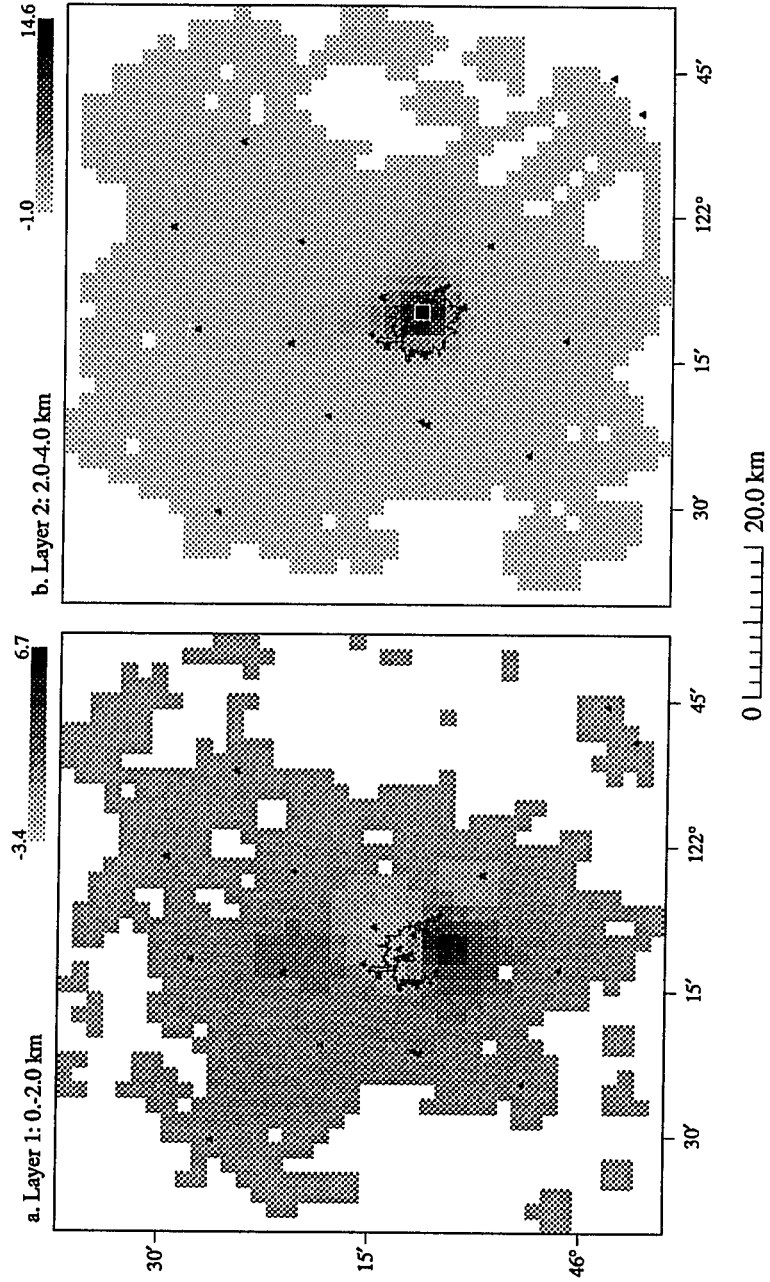
Figure 7.7: (continued)

c. Layer 3: 4.0-6.0 km

d. Layer 4: 6.0-9.0 km

145
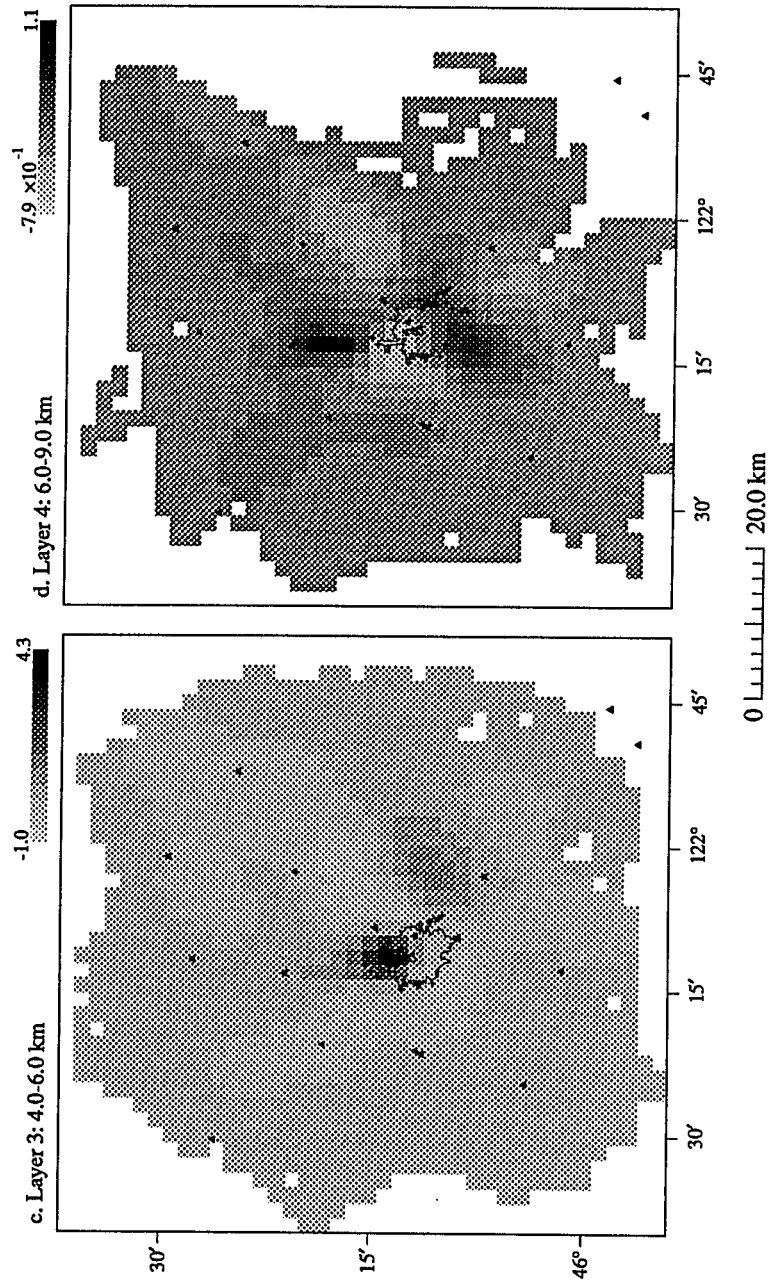
Figure 7.8(*a-h*): Ray ellipsoids and hit count density plots for the Mt. St. Helens inversion. For each layer, the left side shows the relative size of eigenvalues for each ellipsoid and its azimuthal direction. The right side shows the vertical dip of the ellipsoid and the logarithm of the hit count plotted as scaled circles. The number Max refers to the maximum hit count for the layer.

Vertical Projections: Max Count= 1609

Layer 1: 0. - 2.0 km   Ray Ellipsoids

146
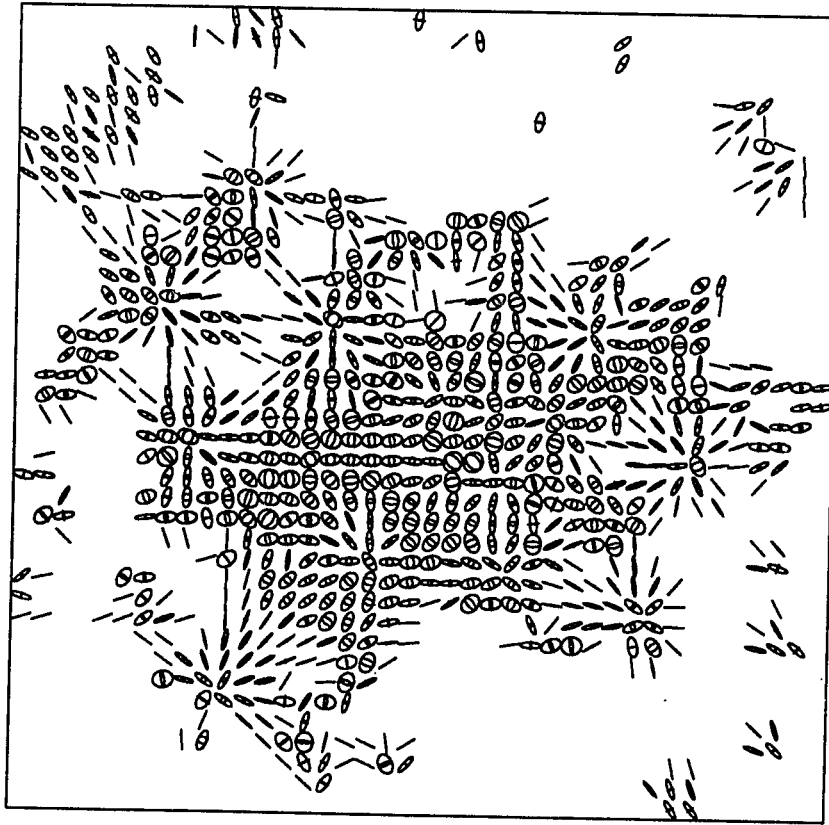
Figure 7.8: (continued)

Vertical Projections: Max Count= 991

b. Layer 2: 2.0-4.0 km Ray Ellipsoids

147
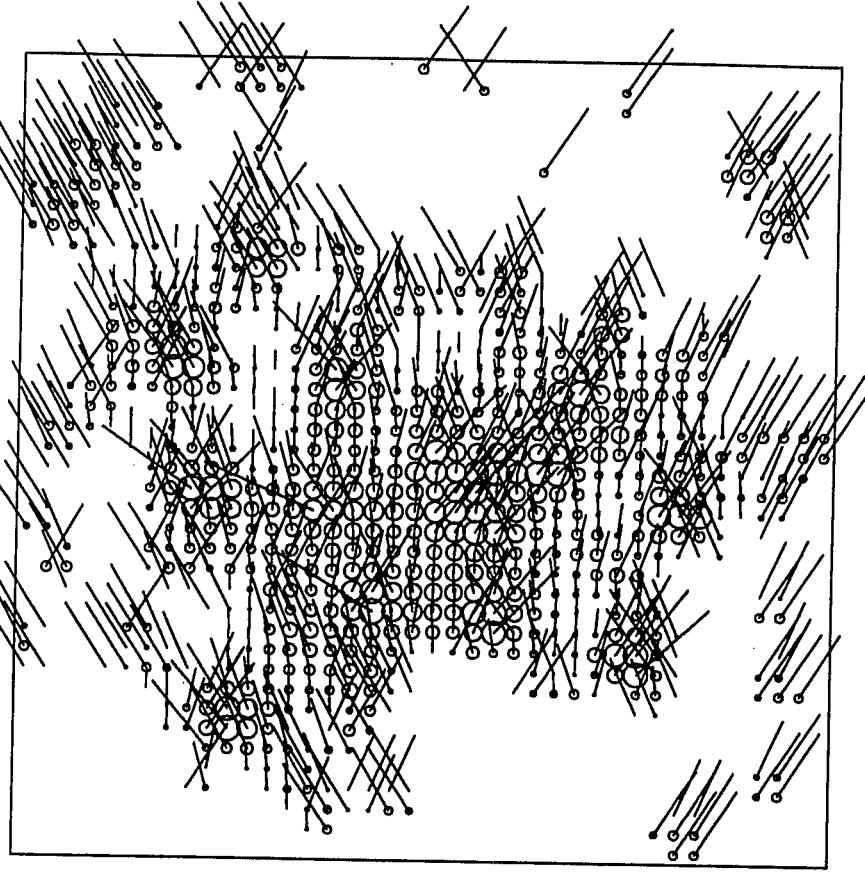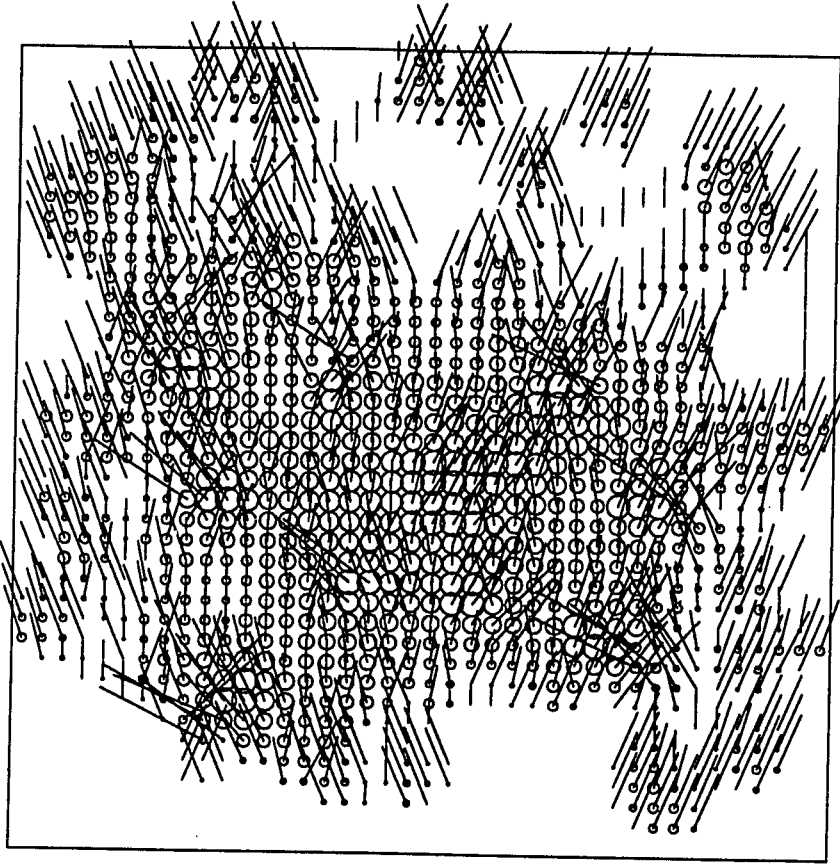
Figure 7.8: (continued)

Vertical Projections: Max Count=   914

c. Layer 3: 4.0-6.0 km Ray Ellipsoids
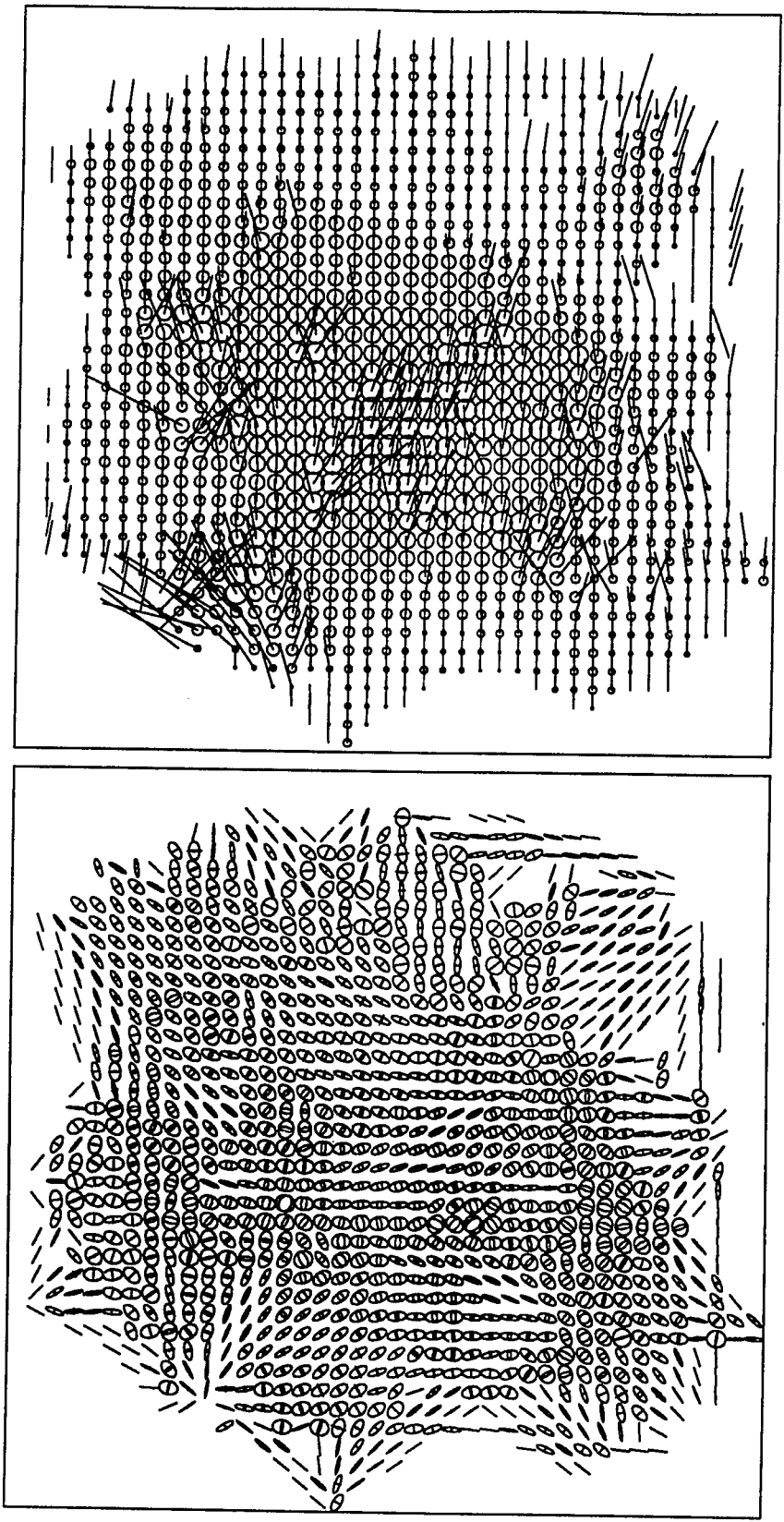
Figure 7.8: (continued)

Vertical Projections: Max Count= 1254

d. Layer 4: 6.0-9.0 km Ray Ellipsoids

148
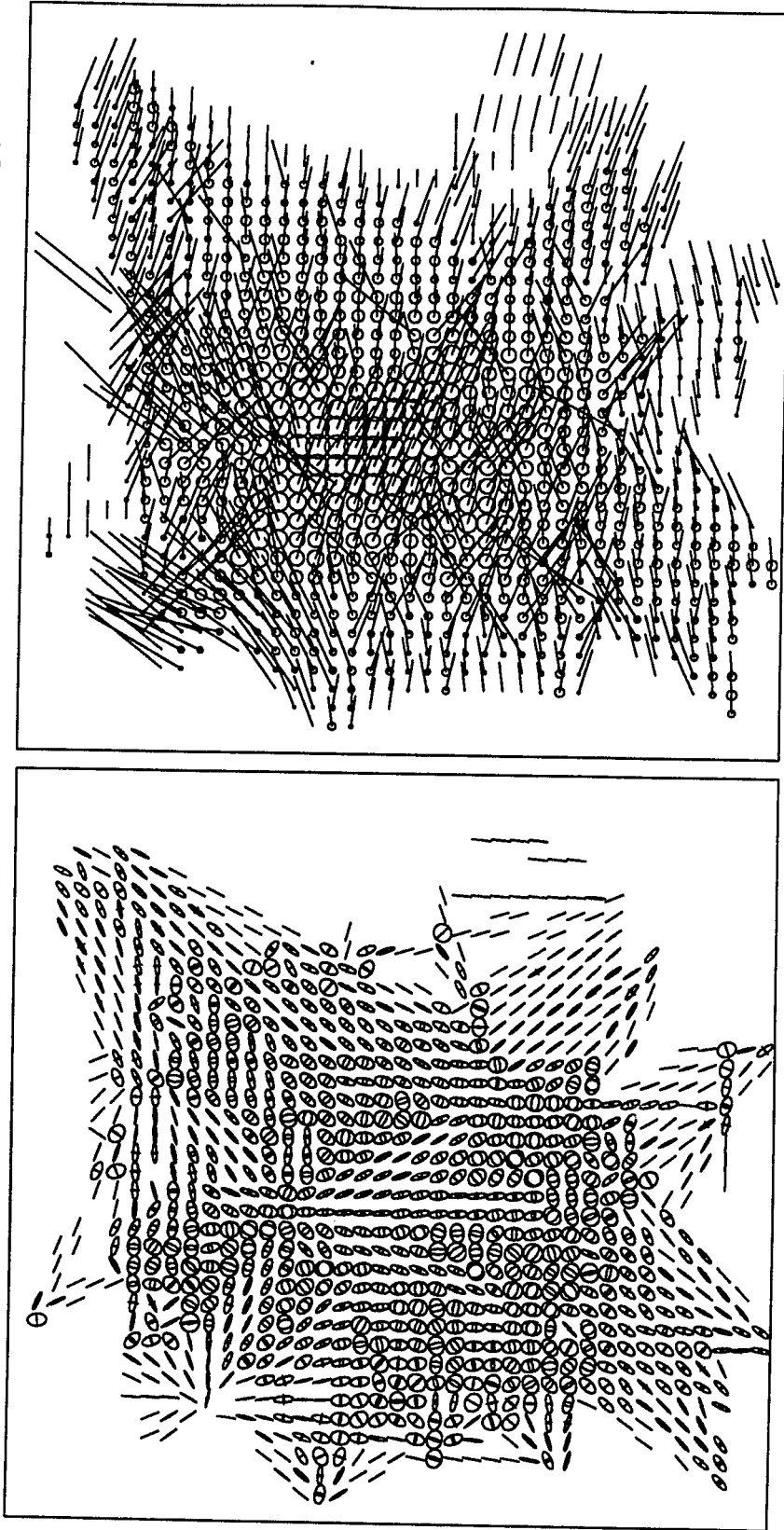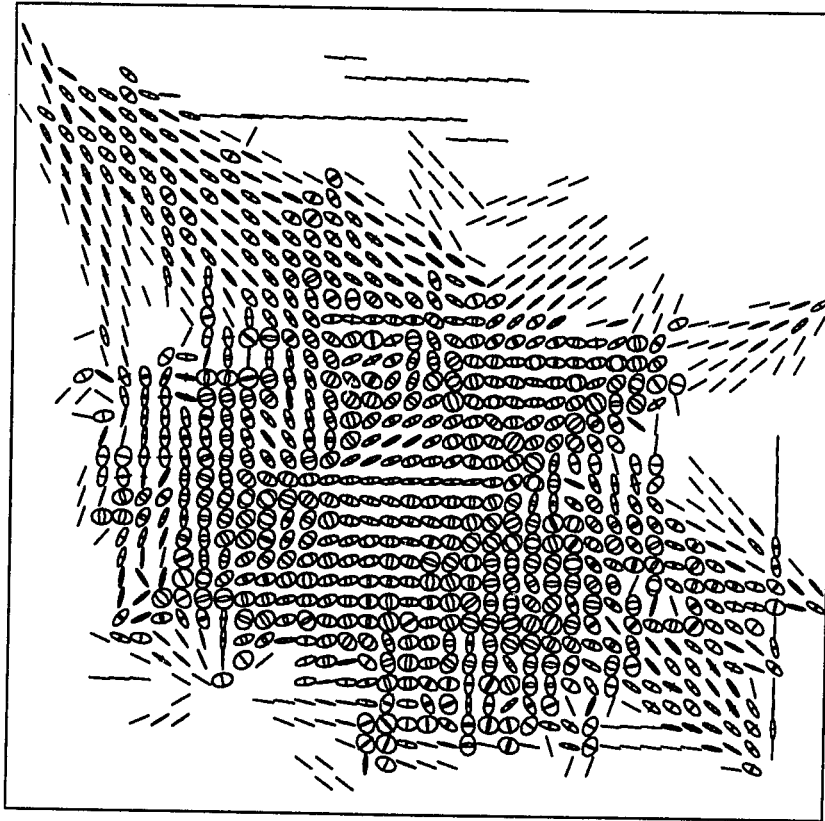
Figure 7.8: (continued)

Vertical Projections: Max Count= 2378

e. Layer 5: 9.0-12.0 km Ray Ellipsoids

149

150
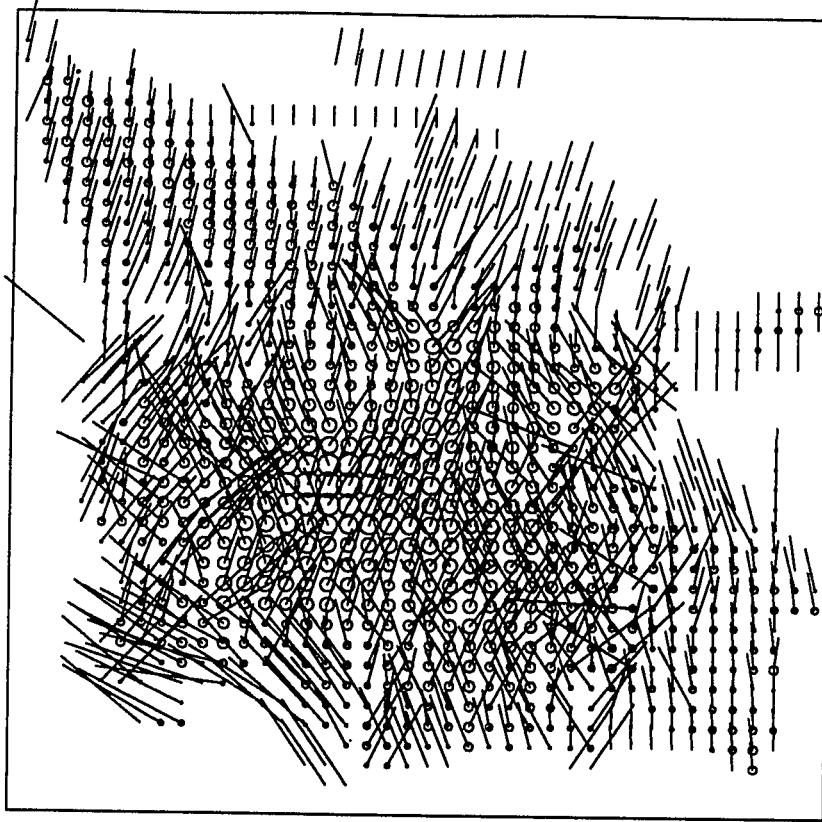
Figure 7.8: (continued)
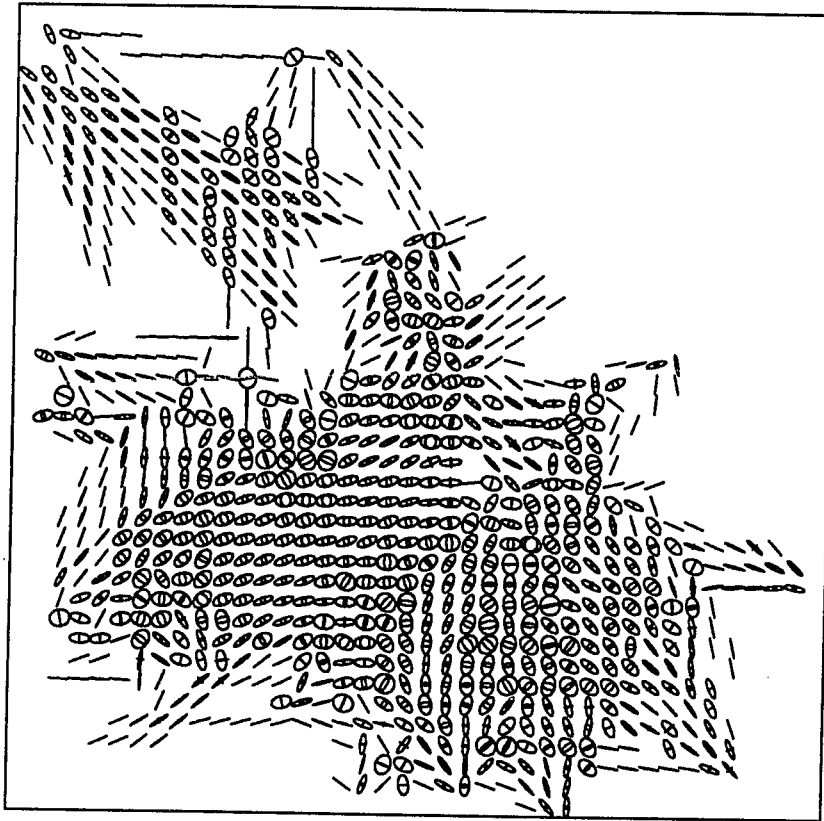
Vertical Projections: Max Count=   866
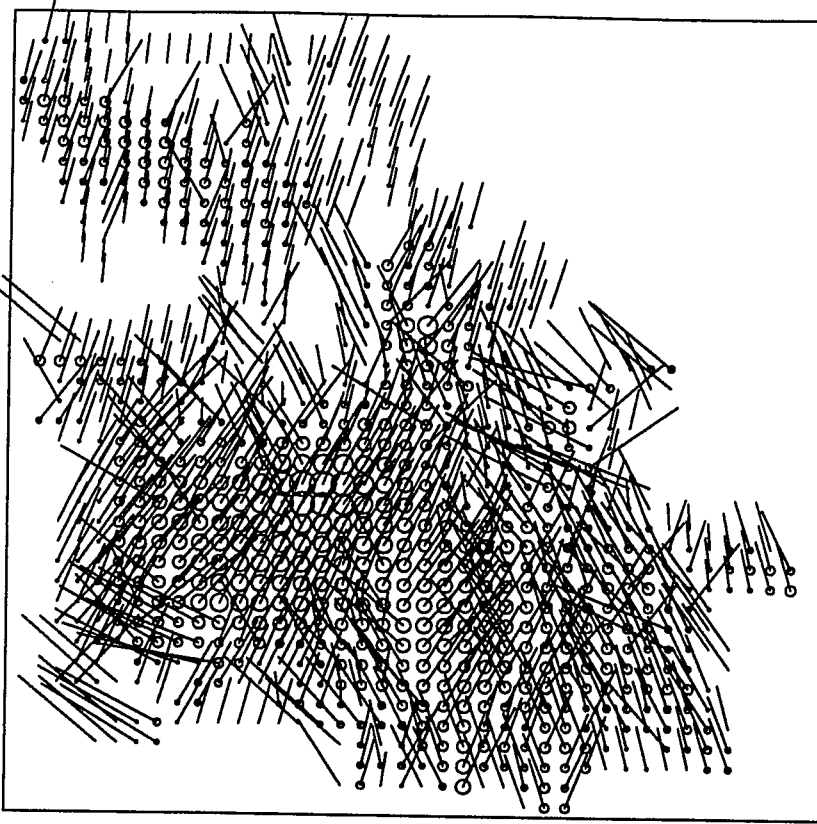
f. Layer 6: 12.0-16.0 km Ray Ellipsoids

Figure 7.8: (continued)

Vertical Projections: Max Count=   97

g. Layer 7: 16.0-20.0 km Ray Ellipsoids
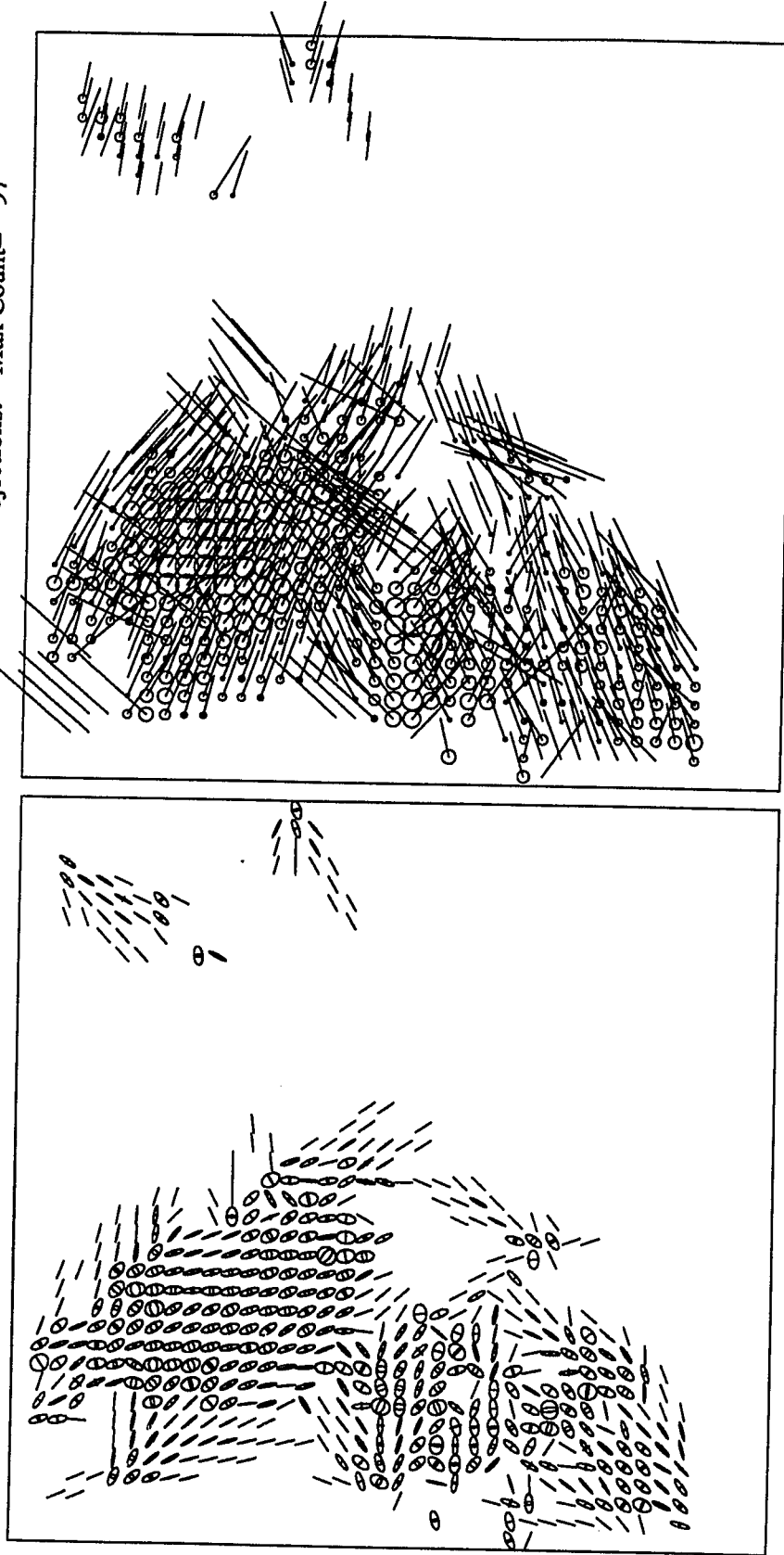
Figure 7.8: (continued)

Vertical Projections: Max Count= 98

h. Layer 8: 20.0-25.0 km Ray Ellipsoids

153
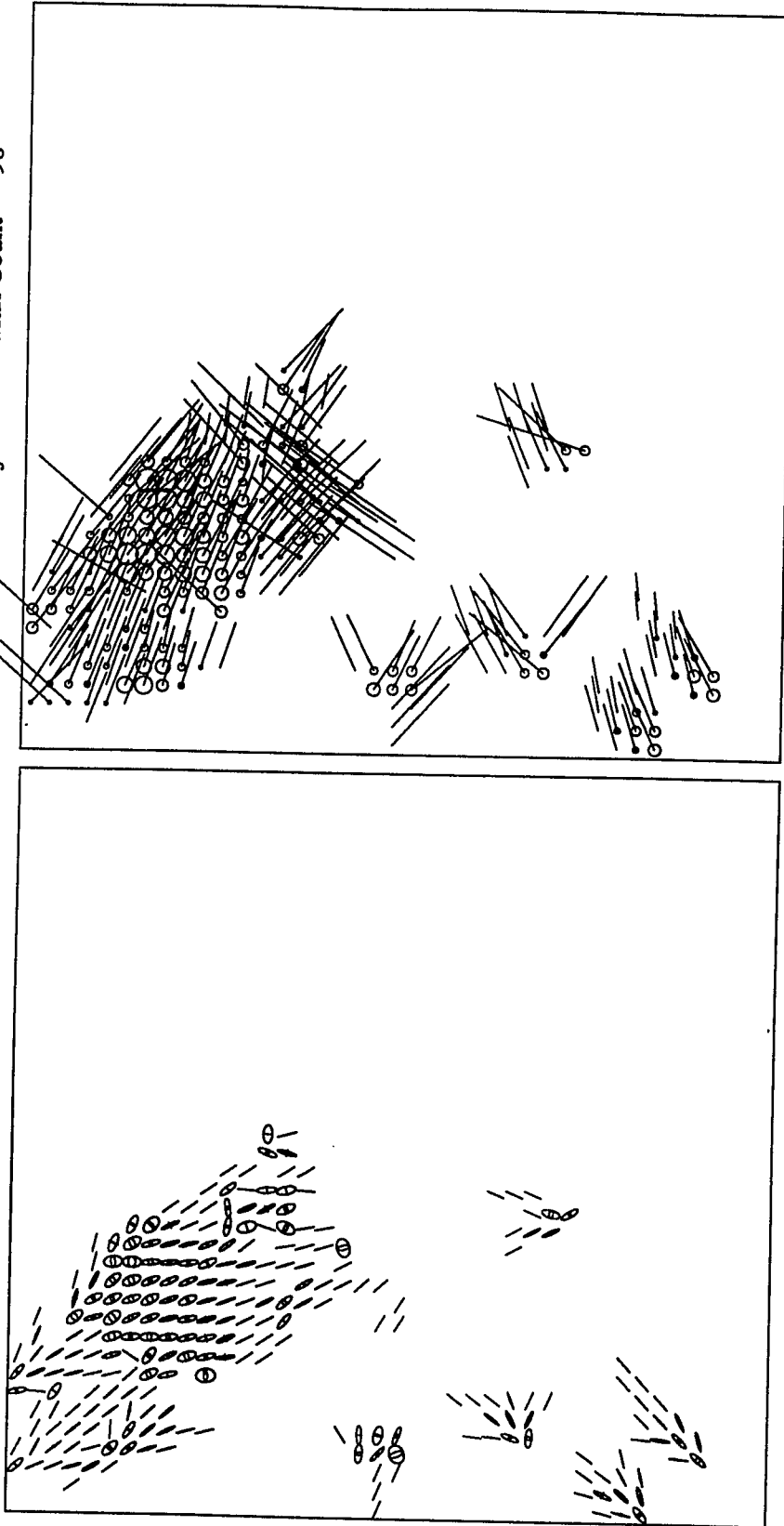
Figure 7.9(*a-h*): Jackknife estimates of the standard errors for the Puget Sound inversion. The jackknife was performed with 30 partitions of the data inverting with the LSQR algorithm where the Laplacian was constrained for regularization. Error assessment should be used in conjunction with resolution estimation for overall determination of the uncertainties in the inversion results.

Jackknife Error Estimates

a. Layer 1: 0.-2.0 km

b. Layer 2: 2.0-4.0 km

154
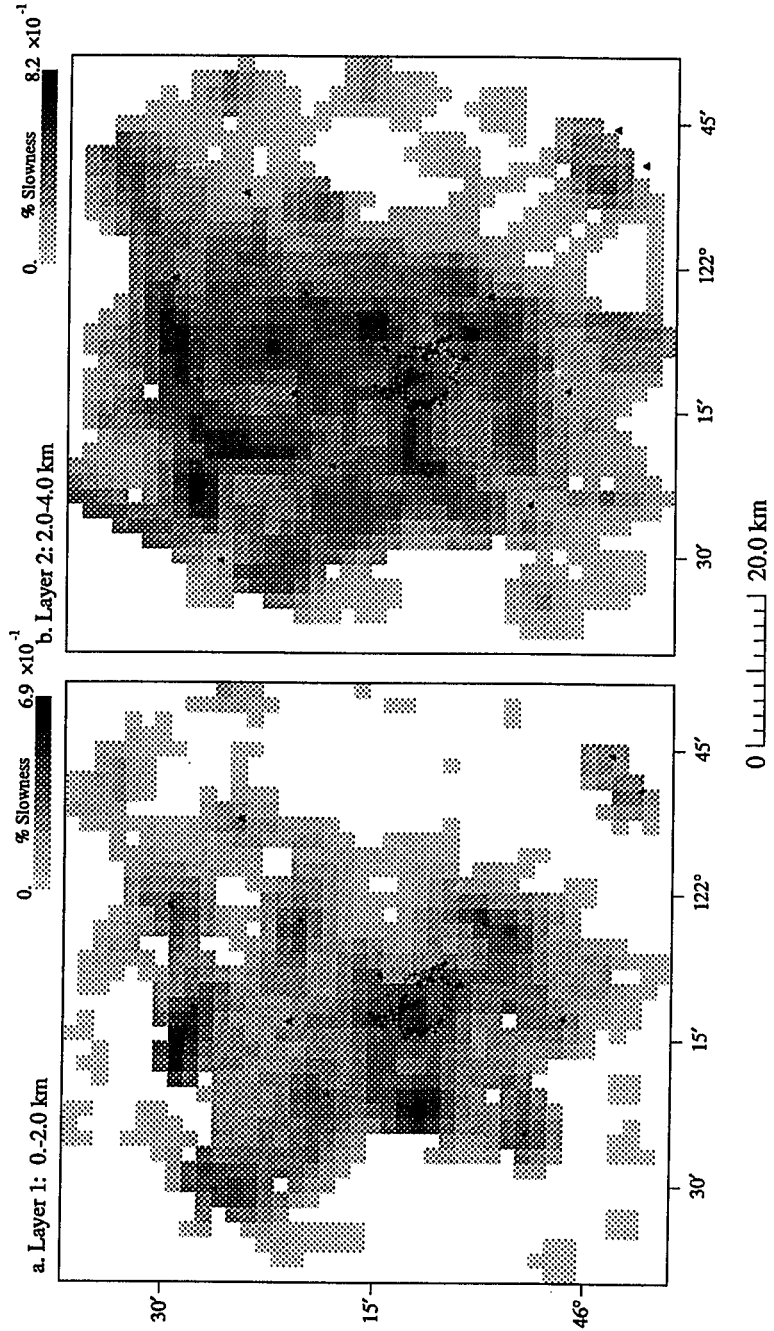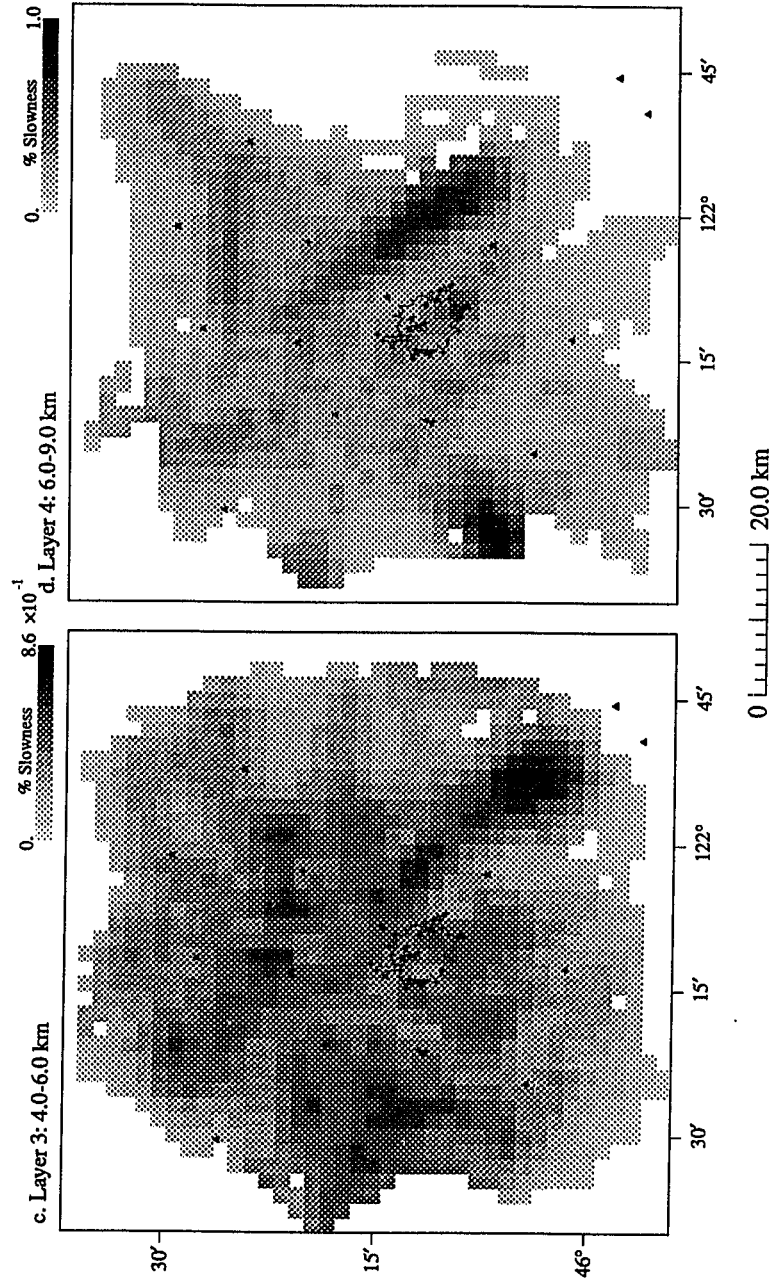
Figure 7.9: (continued)

Figure 7.9: (continued)

e. Layer 5: 9.0-12.0 km

f. Layer 6: 12.0-16.0 km
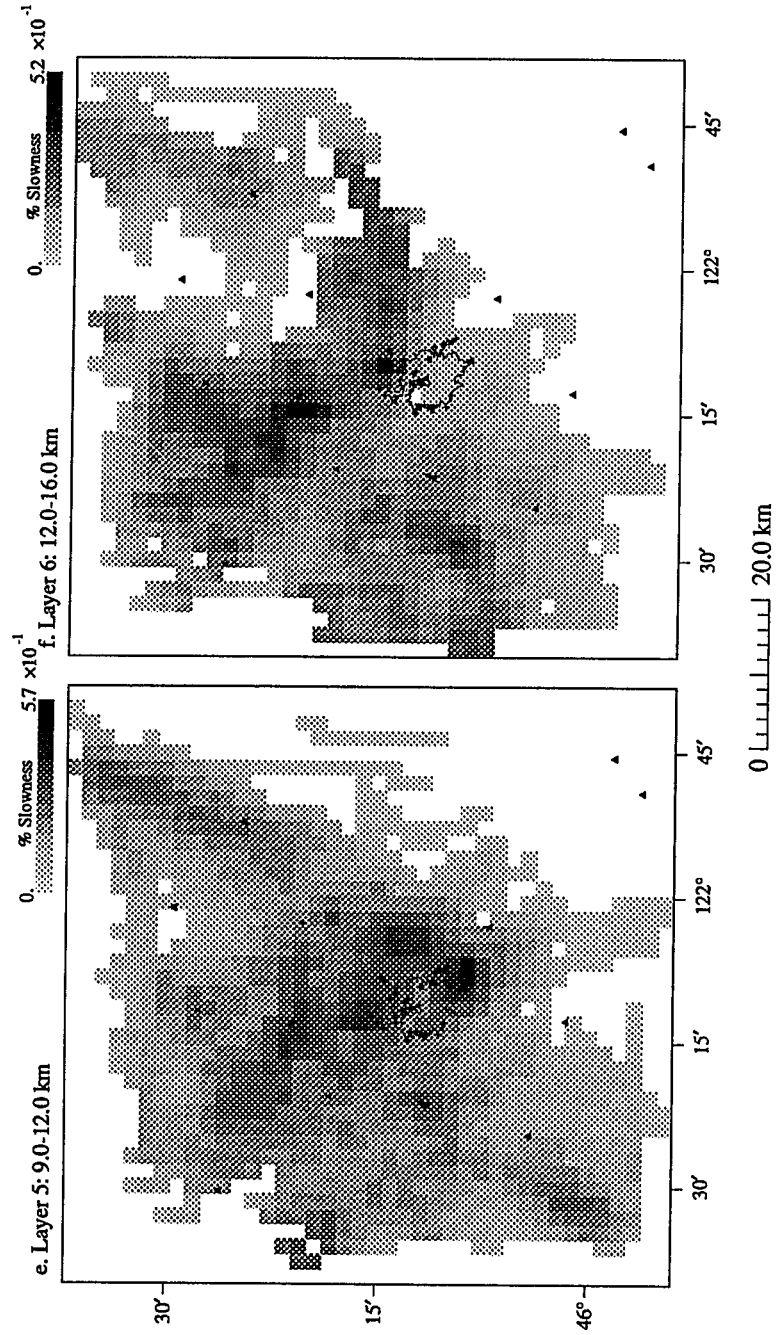
Figure 7.9: (continued)
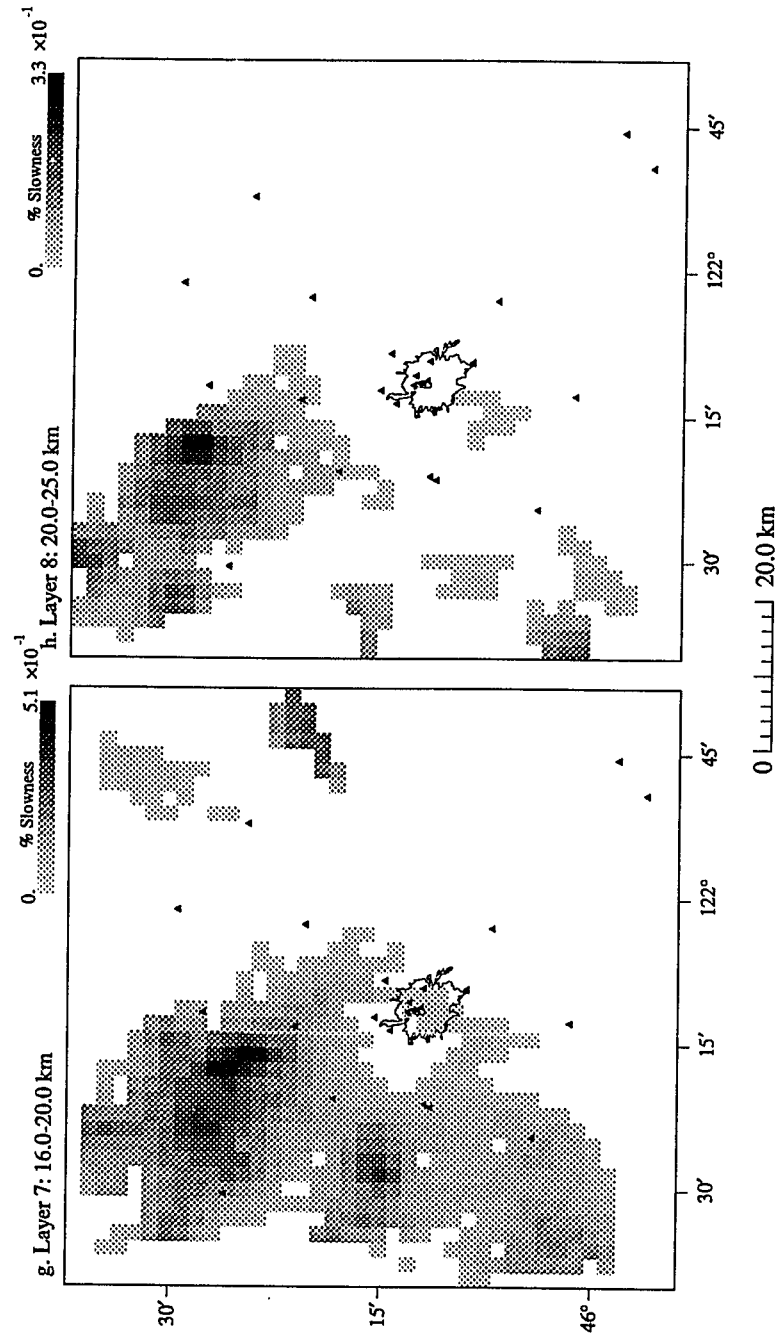
CHAPTER 8

SUMMARY and CONCLUSIONS

The development and implementation of tomographic techniques for P-wave travel time inversion of local earthquake data has resulted in several improvements in both our understanding of the techniques and our knowledge of the structures beneath western Washington. We have seen that by careful use of regularization a Bayesian version of Kaczmarz' technique for solving large sparse matrix systems is comparable to the conjugate gradient techniques represented by LSQR. These techniques produced quite similar results when applied to synthetic data sets with known errors introduced. Furthermore, a new procedure for estimating the model errors using the Jackknife method was introduced. To estimate resolution a more qualitative approach has been adopted by combining several graphical representations of ray statistics providing a better grasp the distribution of rays in the data set. When applied to real data sets in western Washington the resultant images were found to correlate well with known geologic features as well as delineate structures that were previously only inferred.

In the Puget Sound region the Crescent Basalt formation is seen as a prominent, high velocity anomaly on the eastern flanks of the Olympics. This high velocity feature appears to dip beneath Puget Sound towards the east, forming the basement beneath the glacial and marine sediments which have characteristic low velocities. Beneath the accretionary Crescent formation a low velocity anomaly is seen at depth. These are interpreted as contiguous with the low velocity sedimentary rocks of the Olympic Core that have underthrust and accreted adjacent to the Crescent formation. In the Seattle area, several low velocity anomalies correlate well with gravity data and represent distinct, thick sedimentary packages. Near Mt. Rainier small, high velocity anomalies are present near the surface, which correlate with surface expression of pre-tertiary plutonism in this region. At depth a

larger low velocity anomaly is evident, perhaps providing a source for the large gradients in electrical conductivity that are observed here [*Stanley*, 1987].

The Mt. St. Helens inversion correlates as well with known surface geologic information. The Spirit Lake and Spud Mt. plutons appear as high velocity anomalies flanking the low velocity SHZ (St. Helens seismic Zone) which is presumed to be a major fault in this area. Below the crater from 6 km depth a diffuse low velocity anomaly is observed perhaps indicating the presence of unconsolidated, fractured material in the vicinity of the magma accumulations. A more detailed study of the immediate area surrounding the crater could perhaps reveal details of the magma system which supplies the volcano.

Future Work

This research represents a first step in the ongoing effort to determine a detailed 3-dimensional structural image of western Washington. I would like to take this opportunity to suggest several avenues of future work that may contribute to a better understanding of this problem. As mentioned above, a detailed velocity inversion of the Mt. St. Helens area using all the aftershock data from the May, 1980 eruptions may help delineate the magma bodies that feed the crater. This could be augmented with a tomographic inversion for attenuation properties, although practical considerations like the lack of good station calibrations may make this effort difficult. Inversions for S-wave velocity anomalies should be investigated. A 3-D S-wave velocity structure used in conjunction with the P-wave structures developed here would provide us with important specific details of the physical properties of the rocks in the subsurface.

The 3-D tomographic velocity models produced by these methods can be used to improve earthquake locations in this region and help define fault zones from seismicity. In the deep structure, inversion of teleseismic arrivals can be

used to determine the structure of the subducting oceanic crust. Removal of the effects of the surface layers is essential for determining a velocity model at depth. The 3-D model in the top 40 km presented here provides such a model.

On a more theoretical level, with respect to the tomographic method itself, several directions for future work are apparent. The use of the jackknife method for determining errors suggests that statistical methods such as cross validation may be useful to obtain optimum model parameterization. While the use of cross validation appears not to yield useful results for determining regularization, this subject is not fully understood and further work needs to be done to help define a more quantitative method for optimizing regularization. The graphical presentations used in this study to estimate resolution depend heavily on the use of several plots describing ray coverage. In the future, some way of summarizing the information into one or two quantities would help simplify this problem. Finally, a method of non-linear inversion involving the updating of earthquake locations between iterations using efficient 3-D raytracing techniques should be explored and developed.

BIBLIOGRAPHY

Aki, K., A. Christoffersson, and E.S. Husebye, Determination of the three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, *82*, 277-296, 1977.

Aki, K. and P.G. Richards, *Quantitative seismology*, W.H. Freeman and Co., San Francisco, 1980.

Armstrong, R.L., Cenozoic igneous history of the U.S. Cordillera from latitude 42 to 49° north, in *Cenozoic tectonics and regional geophysics of the western Cordillera*, pp. 265-282, Geol. Soc. AM., 1978.

Artzy, E., T. Elfing, and G.T. Herman, Quadratic optimization for image reconstruction, II, *Comp. Graph. and Im. Proc.*, *11*, 242-261, 1983.

Atwater, T., Implications of plate tectonics for the Cenozoic tectonic evolution of western North America, *Geol. Soc. Am. Bull.*, *81*, 3513-3536, 1970.

Backus, G.E. and J.F. Gilbert, Numerical application of a formalism for geophysical inverse problems, *Geophys. J. R. astr. Soc.*, *13*, 247-276, 1967.

Baumeister, J., *Stable solution of inverse problems*, Friedr. Vieweg & Sohn, Braunschweig, 1987.

Ben-Israel, A. and T.N.E. Greville, *Generalized inverses: theory and applications*, John Wiley & Sons, New York, 1974.

Bevington, P.R., *Data reduction and error analysis for the physical sciences*, McGraw-Hill, New York, 1969.

Bonini, W.E., D.W. Hughes, and Z.F. Danes, *Complete Bouguer gravity anomaly map of Washinton*, Washington Division of Geology and Earth Resources, 1974. scale approximately 1:500,000

Carrion, P., *Inverse problems and tomography in acoustics and seismology*, Penn Publishing Co., Atlanta, 1987.

Cerveny, V., Ray tracing algorithms in three-dimensional laterally varying layered structures, in *Seismic Tomography*, D. Reidel Publishing Co., 1987.

Clowes, R.M., M.T. Brandon, A.G. Green, C.J. Yorath, A.S. Sutherland, E.R. Kanesewich, and C. Spencer, LITHOPROBE - southern Vancouver Island: Cenozoic subduction complex imaged by deep seismic reflections, *Can. J. Earth Sci., 24*, 31-51, 1987.

Constable, A.C., R.L. Parker, and C.G. Constable, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics, 52*(3), 289-300, 1987.

Cowan, D.S. and C.J. Potter, *Continent-ocean transect B3: Juan De Fuca spreading ridge to Montana thrust belt*, Geol. Soc. Am., 1986.

Crosson, R.S., Crustal structure modeling of earthquake data 1. Simultaneous least squares estimation of hypocenter and velocity parameters, *J. Geophys. Res., 81*(17), 3036-3046, 1976.

Crosson, R.S., Crustal structure modeling of earthquake data 2. Velocity structure of the Puget Sound region, Washington, *J. Geophys. Res., 81*(17), 3047-3054, 1976.

Crosson, R.S. and T.J. Owens, Slab geometry of the Cascadia subduction zone beneath Washington from earthquake hypocenters and teleseismic converted waves, *Geophys. Res. Letts., 14*, 824-827, 1987.

Dahlquist, G. and A. Bjork, *Numerical Methods*, Prentice-Hall, Inc., New Jersey, 1974.

Deans, S.R., *The radon transform and some of its applications*, John Wiley & Sons, New York, 1983.

Dickinson, W.R., Sedimentary basins developed during evolution of Mesozoic-Cenozoic arc-trench system in western North America, *Can. J. Earth Sci., 13*, 1268-1283, 1976.

Dines, K.A. and R.J. Lytle, Computerized geophysical tomography, *Proc. IEEE*, *67*, 1065-1073, 1979.

Draper, N.R. and H. Smith, *Applied regression analysis*, New York, 1966.

Dudgeon, D.E. and R.M. Mersereau, *Multidimensional digital signal processing*, Prentice-Hall, Inc, New Jersey, 1984.

Duncan, R.A., A captured island chain in the Coast Range of Oregon and Washington , *J. Geophys. Res.*, *87*(B13), 10827-10837, 1982.

Dziewonski, A.M. and D.L. Anderson, Seismic Tomography of the earth's interior, *Am. Scientist*, *72*, 483-494, 1984.

Efron, B. and C. Stein, The Jackknife estimate of variance, *The Annals of Statistics*, *9*(3), 586-596, 1981.

Efron, B., *The Jackknife, the Bootstrap and other Resampling Plans*, Soc. for Ind. and Appl. Math., Philadelphia, 1982.

Evarts, R.C., R.P. Ashley, and J.G. Smith, Geology of the Mount St. Helens area: Record of discontinuous volcanic and plutonic activity in the Cascade arc of southern Washington, *J. Geophys. Res.*, *92*, 10155-10169, 1987.

Fehler, M., Locations and spectral properties of earthquakes accompanying an eruption of Mount St. Helens, *J. Geoph. Res.*, *90*(B14), 12729-12740, 1985.

Finn, C. and D.L. Williams, An aeromagnetic study of Mount St. Helens, *J. Geophys. Res.*, *92*, 10194-10206, 1987.

Golub, G.H. and W. Kahan, Calculating the singular values and pseudoinverse of a matrix, *SIAM J. Numer. Anal.*, *2*, 205-224, 1965.

Golub, G.H. and C.F. Van Loan, *Matrix computations*, The John Hopkins University Press, Baltimore, 1983.

Gower, H.D., J.C. Yount, and R.S. Crosson, *Seismotectonic map of the Puget Sound region, Washington*, U.S. Geol. Surv., 1985.

Hall, J.B. and K.L. Othberg, *Thickness of unconsolidated sediments, Puget Lowland, Washington,* Washington Division of Geology and Earth Resources, 1974. scale approximately 1:316,800

Hammond, P.E., Reconnaissance geologic map and cross sections of southern Washington Cascade Range, latitude 45°30'-47°15'N longitude 120°45'-122°22.5'W, scale 1:125,000, Dep. of Earth Sci., Portland State Univ.,Oreg, 1980.

Hearn, T.M. and R.W. Clayton, Lateral velocity variations in southern California. 1. results for the upper crust from Pg waves, *Bull. Seis. Soc. Am., 76,* 495-509, 1986.

Herman, G.T., H. Hurwitz, A. Lent, and H.P. Lung, On the Bayesian approach to image reconstruction, *Inf. and Control, 42,* 60-71, 1979.

Herman, G.T., *Image Reconstructions from Projections,* Academic Press, New York, 1980.

Hofmann, B., *Regularization for applied inverse and ill-posed problems,* Teubner-Texte, Leipzig, 1986.

Humphreys, E. and R.W. Clayton, Adaptation of back projection tomography to seismic travel time problems, *J. Geoph. Res., 93,* 1073-1085, 1988.

Humphreys, E.D., R.W. Clayton, and B.H. Hager, A tomographic image of mantle structure beneath southern California, *Geophys. Res. Letts., 11,* 625-627, 1984.

Ivansson, S., Remark on an earlier proposed iterative tomographic algorithm, *Geophys. J. R. Astr. Soc, 75,* 855-860, 1983.

Jackson, D.D., Interpretation of inaccurate, insufficient and inconsistent data, *J Geophys. J. R. Astr. Soc., 28,* 97-109, 1972.

Johnson, S.Y., Evidence for a margin-truncating transcurrent fault (pre-late Eocene) in western Washington, *Geology, 12,* 538-541, 1984.

Kissling, E., W.L. Ellsworth, and R.S. Cockerham, Three-dimensional structure of the Long Valley Caldera, California, region by geotomography, *Proc. of Workshop XIX, Active tectonic and Magmatic Processes beneath Long Valley Caldera, Eastern California, 1*, 188-220, 1984.

Lawson, C.L. and R.J. Hanson, *Solving least squares problems*, Prentice-Hall, Englewood Cliffs, N.J., 1974.

Lees, J.M. and R.S. Crosson, Tomographic inversion for 3-D velocity structure at Mount St. Helens using earthquake data, *J. Geophys. Res.*, (submitted, 6/23/88), 1988.

Lees, J.M. and R.S. Crosson, Bayesian ART Versus Conjugate Gradient Methods in Tomographic Seismic Imaging: An Application at Mount St. Helens, Washington, in *Spatial Statistics and Imaging: Proceedings of the 1988 AMS-IMS-SIAM Summer Research Conference*, 1988. submitted (9-30-88)

Manson, C.J., *Index to geologic and geophysical mapping of Washington, 1899-1983*, Washington Division of Geology and Earth Resources, 1984.

Mardia, K.V., *Statistics of direction data*, Academic press, New York, 1972.

McBirney, A.R., Volcanic evolution of the Cascade Range, *Annu. Rev. Earth Planet. Sci., 6*, 437-456, 1978.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Inc., Orlando, 1984.

Mosteller, F. and J.W. Tukey, *Data Analysis and Regression*, Addison-Wesley, Mass., 1977.

Mullineaux, D.R. and D.R. Crandell, The eruptive history of Mount St. Helens, *U.S. Geol. Surv. Prof. Pap., 1250*, 3-16, 1981.

Nakanishi, I., Three-dimensional structure beneath the Hokkaido-Tohoku region as derived from a tomographic inversion of P-arrival times, *J. Phys. Earth, 33*, 241-256, 1985.

Natterer, F., *The mathematics of computerized tomography*, John Wiley & Sons, Teubner, Stuttgart, 1986 .

Neumann-Denzau, G. and J. Behrens, Inversion of seismic data using tomographical reconstruction techniques for investigations of laterally inhomogeneous media, *Geophys. J. R. Astr. Soc., 79*, 305-315, 1984.

O'Sullivan, F., A statistical perspective on ill-posed inverse problems, *Statistical Science, 1*(4), 502-527, 1986.

Olson, A.H., A Chebyshev condition for accelerating convergence of iterative tomographic methods- solving large least squares problems, *Phys. Earth Plan. Int., 47*, 333-345, 1987.

Paige, C.C. and M.A. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Software, 8*, 43-71, 1982.

Phillips, W.M., Compilation geologic map of the Mt. St. Helens 1:100,000 quadrangle, *Wash. Div. Geol. Earth Resour. Open File Rep.*, in press, 1987.

Plackett, R.L., Studies in the history of probability and statistics. XXIX. The discovery of the method of least squares, *Biometrika, 59*, 239-251, 1972.

Pujol, J., Comments on the joint determination of hypocenters and station corrections, *BSSA*, (in preparation), 1988.

Rowland, S.W. and P. Gilbert, Iterative methods for the three-dimensional reconstruction of an object from projections, *J. theo. Biol., 36*, 105-117,, Springer-Verlag 1972.

Scales, J.A., Tomographic inversion via the conjugate gradient method, *Geophysics, 52*, 179-185, 1987.

Schucany, W.R., H.L. Gray, and D.B. Owen, On bias reduction in estimation, *J Amer Stat Assc, 66*(33), 524-533, 1971.

Spakman, W. and G. Nolet, Imaging algorithms, accuracy and resolution in delay time tomography, in *Mathematical Geophysics*, pp. 155-187, D. Reidel Publishing Co., 1988.

Stanley, W.D., C. Finn, and J.L. Plesha, Tectonics and conductivity structures in the southern Washington Cascades, *J. Geophys. Res., 92*, 10179-10193, 1987.

Strang, G., *Linear algebra and its applications*, Harcourt Brace Jovanovich, Inc., San Diego, 1980.

Swanson, D.A. and G.A. Clayton, Generalized geologic map of the Goat Rocks wilderness and roadless areas (6036, Parts A,C, and D), Lewis and Yakima counties, Washington, scale 1:48,000, *U.S. Geol. Surv. Open File Map,*(83-357) 1983.

Taber, J.J., *Crustal structure and seismicity of the Washington continental margin*, University of Washington, 1983. Ph.D. Thesis

Tabor, R.W. and W.M. Cady, Geologic map of the Olympic Peninsula, scale 1:125,000, *Misc. Invest. Map I-994*, U.S. Geol. Surv., Reston, Va., 1978.

Thompson, A.M., J.W. Kay, and D.M. Titterington, A cautionary note about crossvalidatory choice, in preparation, 1988.

Trummer, M.R., Reconstructing pictures from projections: on the convergence of the ART algorithm with relaxation, *Computing, 26*, 189-195, 1981.

Trummer, M.R., A note on the ART of relaxation, *Computing, 33*, 349-352, 1984.

USGS, Aeromagnetic maps for part of southwestern Washington, scale 1:62,500 , *U.S. Geol. Surv. Open File Rep. 82-659*, 1975.

USGS, Aeromagnetic map of the Mount St. Helens area, Washington, scale 1:62,500 , *U.S. Geol. Surv. Open File Rep. 81-659*, 1981.

VanderSluis, A. and H.A. VanderVorst, Numerical solution of large, sparse linear algebraic systems arising from tomographic problems, in *Seismic Tomography*, D. Reidel Publishing Co., 1987.

Weaver, C.S. and S.W. Smith, Regional tectonic and earthquake hazard implications of a crustal fault zone in southwestern Washington, *J. Geophys. Res.*, *88*(B12), 10,371-10,383, 1983.

Weaver, C.S., W.C. Grant, and J.E. Shemeta, Local crustal extension at Mount St. Helens, Washington, *J. Geophys. Res.*, *92*, 10170-10178, 1987.

Wells, R.E., D.C. Engebretson, P.D. Snavely, and R.S. Coe, Cenozoic plate motions and the volcano-tectonic evolution of western Oregon and Washington, *Tectonics*, *3*, 275-294, 1984.

Williams, D.L., G. Abrams, C. Finn, D. Dzurisin, D.J. Johnson, and R. Denlinger, Evidence from gravity data for an intrusive complex beneath Mount St. Helens, *J. Geophys. Res.*, *92*, 10207-10222, 1987.

Willmott, C.J., S.G. Ackleson, R.E. Davis, J.J. Feddema, K.M. Klink, D.R. Legates, J. O'Donnell, and C.M. Rowe, Statistics for the evaluation and comparison of models, *J. Geophys. Res.*, *90*, 8995-9005, 1985.

Young, D.M., *Iterative solution of large linear systems*, Academic Press, New York, 1971.

APPENDIX A

LANCZOS PROCESS: LSQR

Here I state and describe the Lanczos process for tridiagonalization of a symmetric matrix, B, and arbitrary vector $\mathbf{b}$. This material is a condensation of *Paige and Saunders* [1982], *Golub and Van Loan* [1983] and *Van der Sluis and Van der Vorst* [1987]. Choose $\beta_1 = \|\mathbf{b}\|$ such that

$$v^{(1)} = \frac{\mathbf{b}}{\beta_1} \tag{A.1}$$

To get the next vector, $v^{(2)}$, we let

$$w^{(1)} = Bv^{(1)} - \alpha_1 v^{(1)} \tag{A.2}$$

where we choose α_1 such that $w^{(1)}$ and $v^{(1)}$ are mutually orthogonal, i.e. $<w^{(1)}, v^{(1)}> = 0$. This implies that $\alpha_1 = v^{(1)T} B v^{(1)}$, and we set

$$v^{(2)} = \frac{w^{(1)}}{\|w^{(1)}\|} \tag{A.3}$$

Since the $v^{(i)}$ we are constructing will be an orthonormal basis, we proceed by requiring $v^{(3)}$ to be orthogonal to both $v^{(2)}$ and $v^{(1)}$ in the following manner:

$$w^{(2)} = Bv^{(2)} - \alpha_2 v^{(2)} - \beta_2 v^{(1)} \tag{A.4}$$

$$v^{(3)} = \frac{w^{(2)}}{\|w^{(2)}\|} \tag{A.5}$$

This procedure, the repeated orthogonalization and normalization, will result in the desired set of vectors $v^{(i)}$. We summarize the previous discussion by explicitly stating the algorithm in the following form:

Algorithm (3): Lanczos Process (Tridiagonalization)

For symmetric matrix B and starting vector $\mathbf{b}$ let $v^{(0)} \equiv 0$ and $\beta_i \geq 0$ be chosen such that $\|v^{(i)}\| = 1$:

$$\beta_1 v^{(1)} = \mathbf{b} \tag{A.6}$$

$$\left. \begin{array}{l} w^{(i)} = B v^{(i)} - \beta_i v^{(i-1)} \\ \alpha_i = v^{(i)T} w^{(i)} \\ \beta_{i+1} v^{(i+1)} = w^{(i)} - \alpha_i v^{(i)} \end{array} \right\} \quad \text{for } i = 1,2,\ldots, \tag{A.7}$$

The algorithm can be expressed in matrix form by:

$$B\mathbf{V}^{(k)} = \mathbf{V}^{(k)}\mathbf{T}^{(k)} + \beta_{k+1} v^{(k+1)} \hat{\mathbf{e}}_k^T \tag{A.8}$$

Where $\mathbf{V}^{(k)}$ is a matrix consisting of the $v^{(i)}$ as columns and $\mathbf{T}^{(k)}$ is a tridiagonal matrix of the form:

$$\mathbf{T}^{(k)} = \begin{bmatrix} \alpha_1 & \beta_2 & & & & \\ \beta_2 & \alpha_2 & \beta_2 & & \mathbf{0} & \\ & \beta_3 & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & & & \cdot & \cdot & \beta_k \\ \mathbf{0} & & & & \beta_k & \alpha_k \end{bmatrix} \tag{A.9}$$

This can be seen by rearranging the terms in the Lanczos Algorithm such that

$$\beta_{i+1} v^{(i+1)} = w^{(i)} - \alpha_i v^{(i)} \tag{A.10}$$

$$= Bv^{(i)} - \beta_i v^{(i-1)} - \alpha_i v^{(i)} \tag{A.11}$$

which can be regrouped to produce,

$$Bv^{(i)} = \beta_{i+1} v^{(i+1)} + \beta_i v^{(i-1)} + \alpha_i v^{(i)} \tag{A.12}$$

or,

$$Bv^{(i)} = \begin{bmatrix} v^{(i+1)} & v^{(i-1)} & v^{(i)} \end{bmatrix} \begin{bmatrix} \beta_{i+1} \\ \beta_i \\ \alpha_i \end{bmatrix} \tag{A.13}$$

The importance of the above decomposition into tridiagonal form becomes evident when we make the following observation. First, consider the expansion of the k–th approximate solution to the system of equations, $x^{(k)} = V^{(k)} y^{(k)}$. Substituting into the least squares problem

$$Bx = b = \beta_1 v^{(1)} \tag{A.14}$$

$$Bv^{(k)} y^{(k)} = \beta_1 v^{(1)} \tag{A.15}$$

$$V^{(k)} T^{(k)} y^{(k)} = \beta_1 v^{(1)} - y^k \beta_{k+1} v^{(k+1)} \hat{e}_k^T. \tag{A.16}$$

By premultiplying the last equation by $V^{(k+1)\,T}$ and remembering the orthogonality of the columns of V, we get

$$T^{(k)} y^{(k)} = \beta_1 \hat{e}_1. \tag{A.17}$$

This tridiagonal system can then be quite easily solved with standard techniques (QR decomposition) and a solution is found by multiplying by the matrix V. In the case we are dealing with $B = A^T A$ and $b = A^T t$ and the solution x is an approximate solution to the least squares problem, $Ax = t$. The advantage of this technique over the diagonalization of the normal equations is that the calculation of the eigenvectors of the normal equations may require unmanageable amounts of computer memory and time to accomplish. On the other hand, tridiagonalization is simple and efficient.

When applied to the least squares problem,

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{x} \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix} \tag{A.18}$$

the Lanczos process produces a bidiagonalization rather than a tridiagonalization as discussed above. In this case, successive application of Algorithm (2) results in a sequence of alternating vectors,

$$\begin{bmatrix} \mathbf{u}^{(i)} \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ v^{(i)} \end{bmatrix} \tag{A.19}$$

Collecting these in appropriate order in matrices $\mathbf{U}^{(k)}$ and $\mathbf{V}^{(k)}$ the matrix

$$\begin{bmatrix} \mathbf{U}^{(k+1)} & 0 \\ 0 & \mathbf{V}^{(k)} \end{bmatrix} \tag{A.20}$$

will form the basis for (A.18). By setting

$$\begin{bmatrix} r^{(k)} \\ \mathbf{x}^{(k)} \end{bmatrix} = \begin{bmatrix} \mathbf{U}^{(k+1)} & 0 \\ 0 & \mathbf{V}^{(k)} \end{bmatrix} \begin{bmatrix} t^{(k+1)} \\ w^{(k)} \end{bmatrix} \tag{A.21}$$

the Lanczos process leads to the bidiagonal system:

$$\begin{bmatrix} \mathbf{I} & \mathbf{H}^{(k)} \\ \mathbf{H}^{(k)T} & 0 \end{bmatrix} \begin{bmatrix} t^{(k+1)} \\ w^{(k)} \end{bmatrix} = \gamma_0 \hat{e} \tag{A.22}$$

where the $\mathbf{H}$ matrix is lower bidiagonal:

$$\mathbf{H}^k = \begin{bmatrix} \alpha_1 & & & & \\ \gamma_2 & \alpha_2 & & 0 & \\ & \gamma_3 & \cdot & & \\ & & \cdot & \cdot & \cdot \\ & 0 & & \cdot & \cdot \\ & & & \gamma_k & \alpha_k \\ & & & & \gamma_{k+1} \end{bmatrix} \tag{A.23}$$

By comparing (A.18) with (A.22) we see that $w^{(k)}$ is the least squares solution of

$$\mathbf{H}^{(k)}w^{(k)} = \gamma_0\hat{e} \qquad (A.24)$$

and which can be used to get $\mathbf{x}^{(k)}$ from (A.21), since

$$\mathbf{x}^{(k)} = \mathbf{V}^{(k)}w^{(k)} \qquad (A.25)$$

The system in (A.24) can be solved using a QR decomposition as mentioned before. The QR decomposition theorem states that any matrix $\mathbf{H}$ that has linearly independent columns can decomposed according to $\mathbf{H} = \mathbf{QR}$ where $\mathbf{Q}$ is orthonormal and $\mathbf{R}$ is upper triangular and invertible [*Strang*, 1980]. This implies that

$$\mathbf{Q}^T\mathbf{H}^{(k)} = \mathbf{R} \qquad (A.26)$$

and premultiplying (A.24) by $\mathbf{Q}^T$ and substituting from (A.26) we get

$$\mathbf{R}w^{(k)} = \mathbf{Q}^T\gamma_0\hat{e}. \qquad (A.27)$$

Since $\mathbf{R}$ is upper triangular (in this case upper bidiagonal) a solution can readily be found by back-substitution. (Back-substitution is accomplished by determining the solution of an upper triangular matrix by solving the consecutive equations starting from the last and working upward to the first.) This idea, the bidiagonalization of $\mathbf{A}$ and subsequent QR decomposition, is the underlying basis for Algorithm LSQR, developed by *Paige and Saunders* [1982]:

Algorithm (4): LSQR [Paige and Saunders, 1982]

To solve the system $\mathbf{Ax} = \mathbf{b}$ let

$$\beta_1 \mathbf{u}_1 = \mathbf{b}, \quad \alpha_1 v_1 = \mathbf{A}^T \mathbf{u}_1$$

$$w_1 = v_1, \quad \mathbf{x}_0 = 0, \quad \overline{\phi}_1 = \beta_1, \quad \overline{\rho}_1 = \alpha_1$$

For $i = 1, 2, 3,...$

$$\left.\begin{aligned}
\beta_{i+1}\mathbf{u}_{i+1} &= \mathbf{A}v_i - \alpha_i\mathbf{u}_i \\
\alpha_{i+1}v_{i+1} &= \mathbf{A}^T\mathbf{u}_{i+1} - \beta_{i+1}v_i
\end{aligned}\right\} \text{Bidiagonalization}$$

$$\left.\begin{aligned}
\rho_i &= (\overline{\rho}_i^2 + \beta_{i+1}^2)^{1/2} \\
c_i &= \overline{\rho}_i/\rho_i \\
s_i &= \beta_{i+1}/\rho_i \\
\theta_{i+1} &= s_i\alpha_{i+1} \\
\overline{\rho}_{i+1} &= -c_i\alpha_{i+1} \\
\phi_i &= c_i\overline{\phi}_i \\
\overline{\phi}_{i+1} &= s_i\overline{\phi}_i
\end{aligned}\right\} \text{Orthogonal Transformation}$$

$$\left.\begin{aligned}
\mathbf{x}_i &= \mathbf{x}_{i-1} + (\phi_i/\rho_i)w_i \\
w_{i+1} &= v_{i+1} - (\theta_{i+1}/\rho_i)w_i
\end{aligned}\right\} \text{Update Model}$$

Vita

Jonathan Matthew Lees

Born: Syracuse, New York, March 19,1953.

Education: Tichon Hadash High School, Tel-Aviv, Israel, 1971

B.S. Physics, University of Illinois, Urbana, Illinois, 1979.

B.S. Mathematics, University of Illinois, Urbana, Illinois, 1979.

Ph.D. Geophysics, University of Washington, 1989.