TRAVEL TIME INVERSION OF MULTI-OFFSET VERTICAL SEISMIC PROFILES

by

Carl B. Godkin

S.B., Massachusetts Institute of Technology
(1984)

SUBMITTED TO THE DEPARTMENT OF
EARTH, ATMOSPHERIC, AND PLANETARY SCIENCES
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

at the

© MASSACHUSETTS INSTITUTE OF TECHNOLOGY

July 30, 1985

Signature of Author.....
Department of Earth, Atmospheric, and Planetary Sciences
July 1985

Certified by .....
M. Nafi Toksöz
Thesis Advisor

Accepted by .....
Theodore R. Madden
Chairman
Departmental Committee on Graduate Students

Travel Time Inversion of Multi-Offset Vertical Seismic Profiles

by

Carl B. Godkin

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
on July 30, 1985 in partial fulfillment of the
requirements for the Degree of Master of Science in
Earth and Planetary Sciences

Abstract

A robust method of determining one-dimensional velocity structure from vertical seismic profile travel times is formulated by the linearized least squares technique known as Gauss' method. Since this algorithm uses point to point ray tracing and considers headwave refractions, it is versatile enough for even wide offset surveys. The technique is tested with synthetic data under a variety of conditions and is found to produce good results in the presence of noise and when the horizontal layering is improperly modeled.

Three sets of field data, from Mounds, Oklahoma, the Gulf Coast, and the Michigan Basin, were studied with this technique. The Michigan Basin data set was especially interesting as it consisted of data from eight source offsets ranging from 110 to 1940 m with sources azimuthally spread about the borehole for the purpose of seismic imaging of a potential reservoir. Good agreement was found between the zero-offset VSP inversion results and velocities from well logs. An attempt made to image the reservoir in question, a Silurian pinnacle reef, by means of travel time residuals calculated from an average model of the region was successful at least from a qualitative point of view.

Thesis Advisor: M. Nafi Toksöz
Title: Professor of Geophysics

Acknowledgments

I would like to thank everyone who has helped me with this work. Considering that I started this project by asking Nafi what a VSP was, I have come a long way baby. I never would have gotten out of the starting gate without the encouragement and advice of my advisor, Nafi Toksöz.

I am very grateful to Bob Nowack, Roger Turpening, Carol Blackway, and Stephen Gildea for all of their help with my thesis. The four of them together answered 116,962 questions. Bob has a first-principles way of thinking about geophysics that makes him an excellent source of information and a very good teacher. Roger and Carol actually *understand* vertical seismic profiling and their doors were always open. Stephen has the Touch with computers. He has convinced me that computer science is a worthwhile profession after all. I spent many enjoyable hours discussing the what-if world of artificial intelligence with Stephen, Roger, and Gilles Garcia, who has stolen my entire *Far Side* collection.

A host of others have been influential in my development as (dare I say it?) a geophysicist. My very good friend Jay Pulli can consider himself responsible for getting me into this business in the first place. The entire seismic network crew over the years, lead always by the king of the late risers, Mike Guenette, are stars. Together we appeared on television, enjoyed the Eric experience, and chased aftershocks in weather only the hard-drinking population of Plaster Rock, New Brunswick could tolerate.

Thanks go to Greg Duckworth for reading the part of the thesis where I use words like bandlimited and sampling theorem. Elsie Hanson proofread the whole thesis, a feat I appreciate very much. If you smell a Macintosh in some of the figures, Jay Pulli is the man behind the mouse. Thanks, boss. Tim Keho gave me data from Alabama so tamed that it practically inverted itself. Michael Prange helped me learn seismology above and beyond the call of even a TA's duty.

I value highly my friendship with many members of that polyglot community that is the Dept. of Earth, Atmospheric, and Planetary Science. Seeing at least two of three of Christophe Gonguet, Wafik Beydoun, and Cengiz Esmersoy always logged in gave my world a sense of continuity. Someday I will take João and José Rosa up on their offer and visit Brazil and its two leading scientists. George Blumberg and Jim Mendelson never invited me up to their place in New Hampshire last winter which probably saved me from hurting myself on the slopes. Ted Charrette kept me on my toes last fall when I TA'd 12.521. I hope someone returns him the favor someday. I can't think of anything clever to say about Shane Ingate except that he talks funny and we all knew that already. Roy Wilkens leads a great expedition, especially if you like places with names like Blow Me Down. Dave Lineman, Richard Herrmann, and Danny Fain are okay, no matter what anyone says about them. Mike Stevens brings his lunch too much (Excuse me?). Those four and a host of others made the Zone what it is today.

This work was supported by the M.I.T. Vertical Seismic Profiling Consortium.

Table of Contents

CHAPTER 1

Vertical Seismic Profiling

1.1. Introduction

Vertical seismic profiling (VSP) is an important technique in geophysical exploration because it provides a direct record of the seismic wave field within the Earth. Each geologic horizon within the structure may be studied directly. The source waveform can be observed on both sides of the horizon, as incident, reflected and transmitted waves. The changing response of the medium to the same signal at many depths may be recorded in situ. Rock properties such as velocity, impedance, and attenuation for both compressional and shear waves may be recovered (Gal'perin, 1974; Hardage, 1983; Toksöz and Stewart, 1984). In addition, the VSP may be used as a densely sampled check-shot survey to relate sonic logs and surface seismic surveys. The more common surface seismic survey measures the reflected seismic wave field at the Earth's surface. VSP records complement more conventional surface seismic records by enabling us to resolve some of the inherent nonuniqueness problems of reflection seismic data.

1.2. The VSP Geometry

VSP data is collected in a fashion similar to the surface seismic method except that there are usually far less receivers per shot. Instead of recording, say, hundreds of channels per shot as in reflection experiments, only a few channels (usually only one) per shot are recorded. Often three component VSP data are recorded to capture the vector wavefield.

Another major difference is between the horizontal reflection seismic geophone arrays and the vertical VSP geophone arrays. Because of the modified geometry, the downgoing (transmitted) seismic wave field dominates at least the early part of the VSP record while the surface seismic experiment records only upgoing (reflected) waves. Good signal to noise ratios can be achieved with only a few traces or even a single trace; the large amount of stacking usually required for surface reflection data is unnecessary. The ray tracing geometry needed to model each experiment is thus quite different. In fact, the VSP geometry is just like that associated with local earthquake seismology except that the source and receiver positions are reversed.

Figure 1.1 shows a schematic view of an idealized VSP experiment. It is idealized in some of the same ways that the theory presented in this thesis is idealized: the geologic layering is horizontal and the layers themselves are homogeneous. Other idealizations include the straight, apparently cylindrical borehole and the point source which produces only compressional waves. Deviated holes and nonideal sources are not problems for the methods presented in subsequent chapters as long as the assumptions regarding the simplicity of the structure are valid.

1.3. The Utility of Travel Time Inversion

Travel time studies play a fundamental role in seismic interpretation. Full waveform inversion schemes (e.g. Stewart, 1983; Beydoun, 1985) make use of amplitude information contained in the entire seismic trace to solve for velocity, attenuation and density from the upgoing and downgoing waves. Why, then, would anyone be interested in travel time inversion which can provide only the velocity structure? The more complete results of full waveform inversion come at a price. It is much more expensive, in terms of computer time, than travel time inversion. Existing full waveform methods cannot handle refracted arrivals or the joint inversion of data from multiple sources. Many properties of the survey which are of little interest to

someone using travel time inversion become critically important when the full waveform problem is considered.

Full waveform inversion also has difficulties dealing with headwave refractions due to problems calculating amplitudes. The forward travel time problem is actually easier to solve for refracted waves than for direct waves so that headwaves do not pose difficulties for travel time inversion.

1.4. Overview

The problem of travel time inversion of multi-offset VSP data is addressed in detail in subsequent chapters. Chapter 2 describes the forward problem, one-dimensional ray tracing. The special problem of root-finding near an asymptote is also treated. Chapter 3 presents a review of travel time inversion for VSP geometries and describes two different formulations of the inverse problem. Potential problems associated with both methods are discussed. The fourth chapter treats data acquisition for this problem. Measuring travel times from a Vibroseis® source by manual and pseudo-manual methods are discussed. Chapter 5 presents the results of the travel time inversion of three sets of field data. The emphasis is on the interpretation of multi-offset data collected during the MIT/CGG Experimental Group Shoot in the Michigan Basin in 1983. Two appendices discuss the inversion formulation presented in Chapter 3 in greater detail and compare several variations on the method. Appendix C studies data error surfaces for this inverse problem. Appendix D investigates the validity of cubic spline interpolation of sampled data.

Figure Caption

Figure 1.1. Schematic view of an idealized VSP experiment geometry with simple horizontal geologic layering.

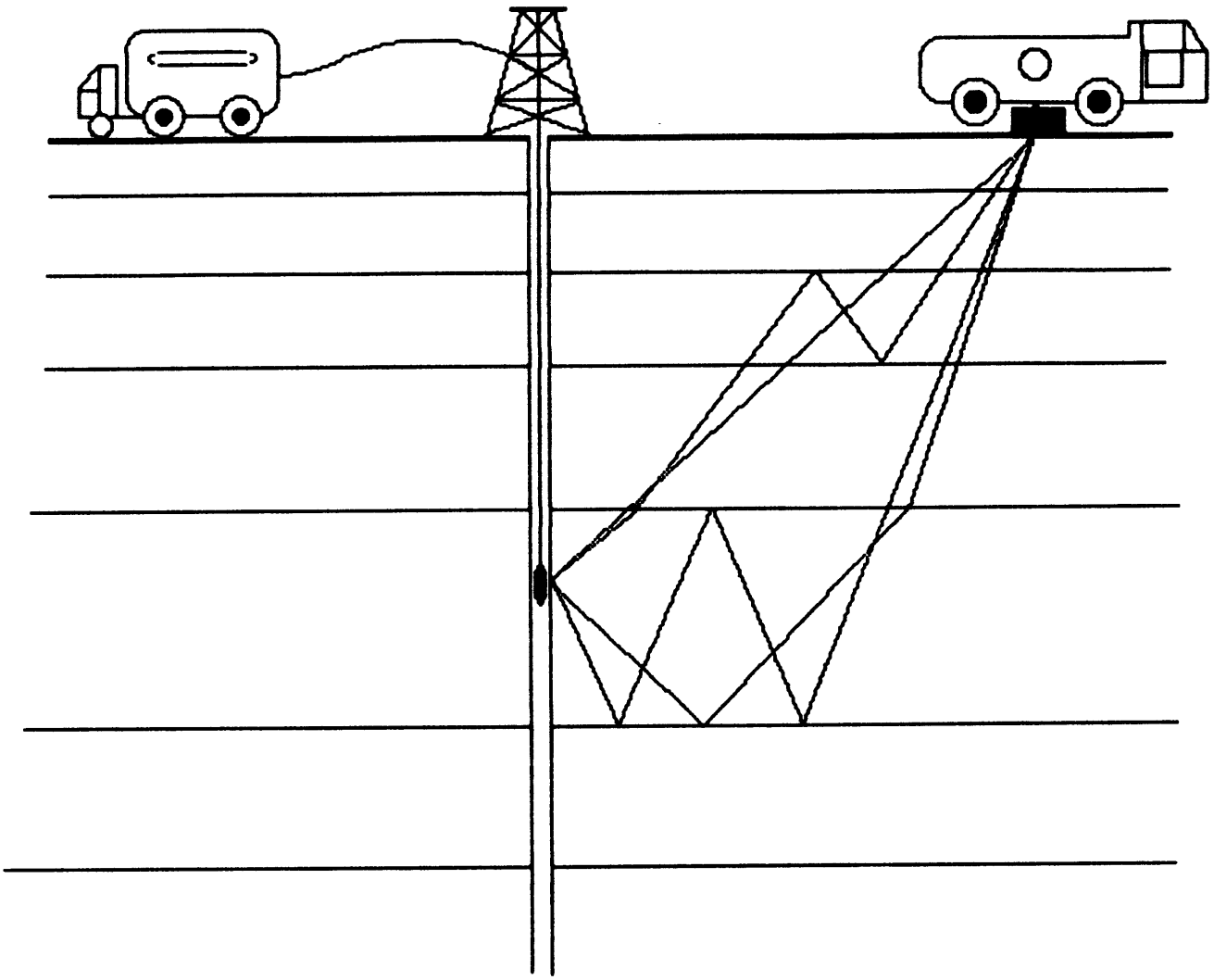


Figure 1.1

CHAPTER 2

One-Dimensional Ray Tracing

2.1. Introduction

In a medium with velocity variations in only one dimension, it is a straightforward matter, at least in principle, to find the minimum time path between a source and receiver combination. This is the forward problem.

In order to solve the inverse problem, it is necessary to solve the forward problem repeatedly for trial models and compare these results with actual observed data. This chapter presents methods for the one-dimensional ray tracing of direct and critically refracted rays. One section is devoted to numerical problems associated with the ray shooting process which is used to find the appropriate ray parameter.

2.2. The Direct Wave

For a given one-dimensional velocity distribution $V(z)$ we can formulate the travel time and range equations for a ray which does not reach the bottom of its trajectory (i.e. for which the angle of incidence never exceeds 90°).

From Figure 2.1, we see that

$$\frac{dx}{ds} = \sin\vartheta \tag{2.1}$$

and

$$\frac{dz}{ds} = \cos\vartheta. \tag{2.2}$$

The range of the ray is then given by

$$x = \int ds \sin\vartheta = \int dz \tan\vartheta \tag{2.3}$$

and the travel time by

$$t = \int \frac{ds}{V} = \int \frac{dz}{V\cos\vartheta}. \tag{2.4}$$

A simplifying approximation that is often made is that the Earth is made up of a series of homogeneous layers so that discrete analogs of the range and travel time equations are used. Consider the VSP geometry shown in Figure 2.2. The range and travel time equations for a ray from a source at the surface to a receiver at depth d from the top of layer m are

$$x = \sum_{i=1}^{m-1} h_i \tan\vartheta_i + d\tan\vartheta_m \tag{2.5}$$

and

$$t = \sum_{i=1}^{m-1} \frac{h_i}{V_i\cos\vartheta_i} + \frac{d}{V_m\cos\vartheta_m}, \tag{2.6}$$

where h_i is the thickness of the ith layer. Values of ϑ_i can be found from Snell's Law:

$$p = \frac{\sin\vartheta_i}{V_i} = \frac{\sin\vartheta_{i+1}}{V_{i+1}} \tag{2.7}$$

where p is the ray parameter. For computational ease, the travel time and range equations are usually reformulated in terms of the ray parameter so that equations (2.5) and (2.6), respectively, become

$$x = \sum_{i=1}^{m-1} \frac{h_i p V_i}{\sqrt{1-p^2V_i^2}} + \frac{dpV_m}{\sqrt{1-p^2V_m^2}} \tag{2.8}$$

and

$$t = \sum_{i=1}^{m-1} \frac{h_i}{V_i\sqrt{1-p^2V_i^2}} + \frac{d}{V_m\sqrt{1-p^2V_m^2}}. \tag{2.9}$$

It is necessary to start from a given value of ϑ at either the source or the receiver in order to get the travel-time-depth calculation going. In some cases the initial value of ϑ (or equivalently the ray parameter p) may be found analytically but very often it must be approximately determined with numerical methods.

2.3. Finding the Ray Parameter

In order to solve the forward problem, the value of the ray parameter p must be found for each ray. In the direct ray case, p cannot be found analytically in general so that some form of root-finding will be necessary. It will turn out that the most stable algorithm for this particular root-finding problem is also the simplest method.

Given a discretized velocity vs. depth model, the ray parameter is implicitly determined by equation (2.8). Because of the nonlinear form of the equation numerical methods must be used. The ray parameter is bounded above and below for physical reasons:

$$0 < p < \frac{1}{V_{max}}. \tag{2.10}$$

Since $p = \frac{\sin \vartheta_i}{V_i}$, a negative ray parameter implies either a negative velocity or a negative angle of incidence. The first is of course physically unrealistic while the second implies a ray heading away from the receiver. The ray parameter is bounded above by $1/V_{max}$, which is the reciprocal of the maximum velocity above the receiver. If $pV_{max} \geq 1$, then the denominator of at least one of the terms on the right hand side of equation (2.8) will be imaginary which does not make sense because the range is a real quantity. The locus of $x = x(p)$ is thus bounded as p increases by an asymptote at $1/V_{max}$ as shown in Figure 2.3. For larger values of x, small changes in p mean enormous changes in x.

To determine $p(x)$, a two-point straddle method must be used to keep p within bounds. This criterion rules out many root-finding algorithms such as Newton's

method or the secant method. The method of false position (Acton, 1970) is such a two-point straddle method. To solve $x(p) = x$ between p_1 and p_2, one would simply compute $x(p_1)$ and $x(p_2)$ and linearly interpolate between them to find p_3. Then throw away either p_1 or p_2 so that the root is surrounded by the survivor and p_3 and repeat this procedure until

$$|x(p_n) - x| < \varepsilon. \tag{2.11}$$

The method of false position always converges unless the two original endpoints straddle a discontinuity which is unfortunately almost the case in this problem since one endpoint is *at* a discontinuity. One suggestion might then be to use $p_2 = 1/V_{max} - \delta$ as the upper bound instead but this has the effect of limiting the offset to $x(1/V_{max} - \delta)$ which may be undesirable in VSP inversion.

It is necessary, therefore, to fall back on the bisection method which always works, no matter how simple-minded it may seem. Given endpoints p_1 and p_2, choose $p_3 = \frac{1}{2}(p_1 + p_2)$ and, as in false position, keep p_3 and either p_1 or p_2 so that the root is surrounded. Lather, rinse, and repeat.

Convergence to within $\varepsilon = 0.3$ meters (my criterion) for a structure in which $V_{max} = 3$ km/sec is achieved after only 10 iterations. Depending upon the offset (corresponding to the height of x on the asymptote), the method of false position may require many more.

2.4. The Critically Refracted Wave

The upper traces on a VSP record section often show first arrivals which have been critically refracted along layer interfaces. Since these headwave refractions are first arrivals, it is necessary to include them in the inversion.

There are two conditions which must be met if headwaves are to be observed. They both constitute geometrical constraints resulting from Snell's Law. For the

horizontally layered case, headwave refractions occur only along the top of layers having higher velocities than any above. Figure 2.4 shows a hypothetical velocity model in which there are only two possible critically refracted rays: one bottoming at 2.1 km/s and the other at 2.4 km/s. Layers 3 and 4 in the figure define a low velocity zone and there is no ray with a real takeoff angle which would be critically refracted along the 2/3 or the 3/4 interface.

Another existence criterion for the headwave refraction is that the source offset must exceed what is known as the critical distance. The critical distance is a constant "overhead" which is the horizontal projection of the path down to the refractor and up to the geophone (Fig. 2.4).

The range and travel time equations describing headwaves are in general much simpler than the corresponding equations for direct waves because the ray parameter is not an unknown but is the reciprocal of the velocity of the refractor. This can be seen by noting that the angle of incidence at the refractor is 90° so that from (2.7)

$$p = \frac{\sin 90^\circ}{V_r} = \frac{1}{V_r} \tag{2.12}$$

A ray bottoming on the top of the rth layer and reaching a geophone at depth d from the top of the mth layer ($m < r$ as in Figure 2.4) is governed by the critical distance x_{crit} which is given by

$$x_{crit} = \sum_{i=1}^{r-1} h_i \tan\vartheta_i + \sum_{i=m+1}^{r-1} h_i \tan\vartheta_i + (h_m - d)\tan\vartheta_m, \tag{2.13}$$

where, by Snell's Law, $\sin\vartheta_i = \frac{V_i}{V_r}$. It is apparent that if ϑ is to remain real, then all V_i should be less than or equal to the refractor velocity V_r, which verifies the first existence criterion mentioned above. The overhead x_{crit} is always positive since all of its terms are positive.

The equation for travel time is

$$t = \sum_{i=1}^{r-1} \frac{h_i \cos\vartheta_i}{V_i} + \sum_{i=m+1}^{r-1} \frac{h_i \cos\vartheta_i}{V_i} + \frac{(h_m - d)\cos\vartheta_m}{V_m} + \frac{x}{V_r}. \tag{2.14}$$

Equation (2.14) is often written in terms of the ray parameter as

$$t = \sum_{i=1}^{r-1} h_i \sqrt{\frac{1}{V_i^2} - p^2} + \sum_{i=m+1}^{r-1} h_i \sqrt{\frac{1}{V_i^2} - p^2} +$$

$$(h_m - d)\sqrt{\frac{1}{V_i^2} - p^2} + \frac{x}{V_r}. \tag{2.15}$$

Telford et al. (1976) show that the first arrival in a one-dimensional layered geometry is either a direct or a critically refracted wave. In order to calculate the travel time of the first arrival, find the minimum of the travel times of the direct and all possible refracted waves.

Figure Captions

Figure 2.1. Offset VSP geometry with one-dimensional velocity function.

Figure 2.2. Offset VSP geometry with discrete one-dimensional velocity layering.

Figure 2.3. Plot of source offset $x(p)$ vs. p for receiver depth 310 m in a simple model. Simple bisection is the most efficient root-finding technique for large offsets due to the effects of the asymptote.

Figure 2.4. Critically refracted raypaths for a one-dimensional model. Due to the low velocity zone, only two headwave paths are possible.

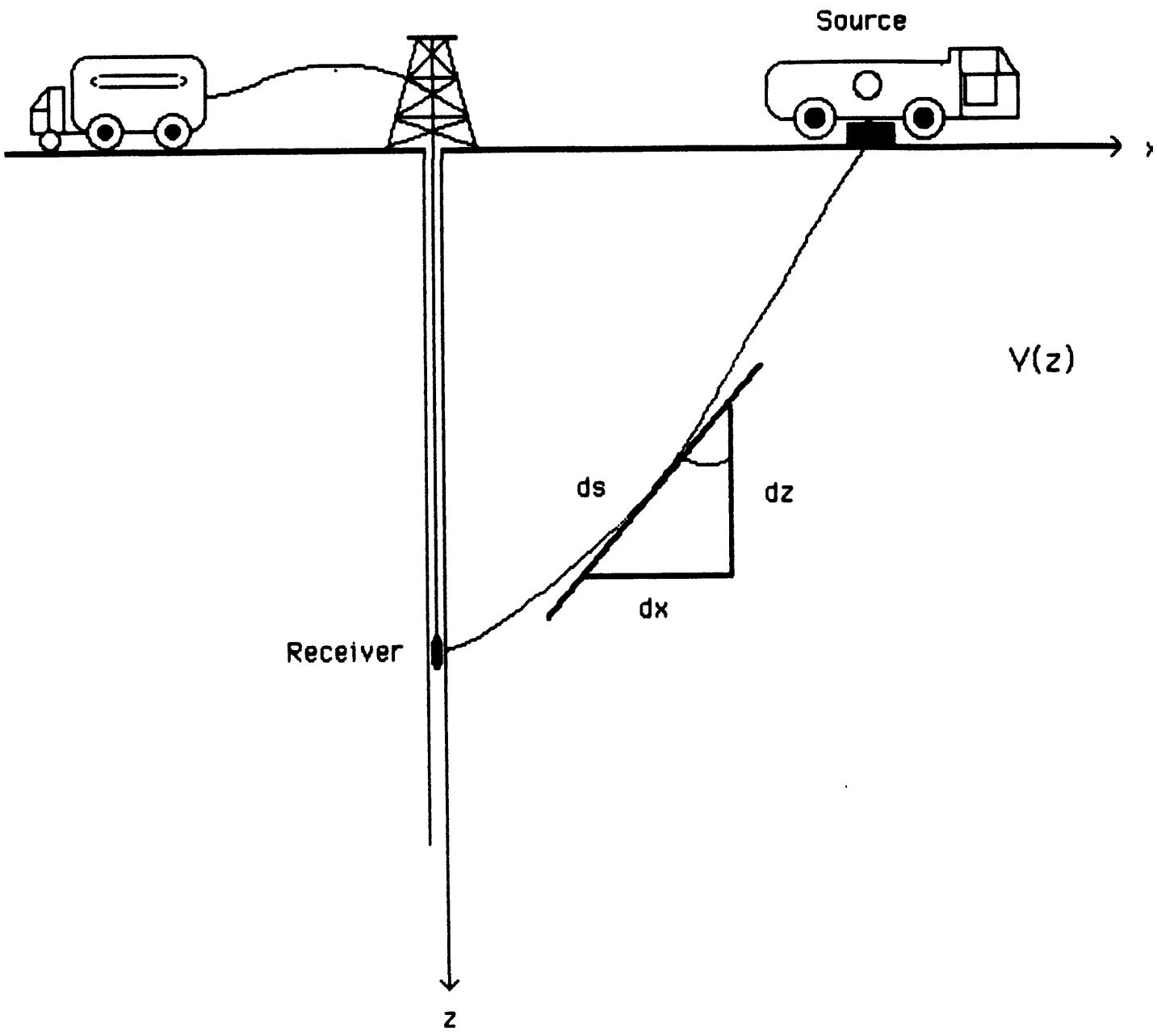


Figure 2.1

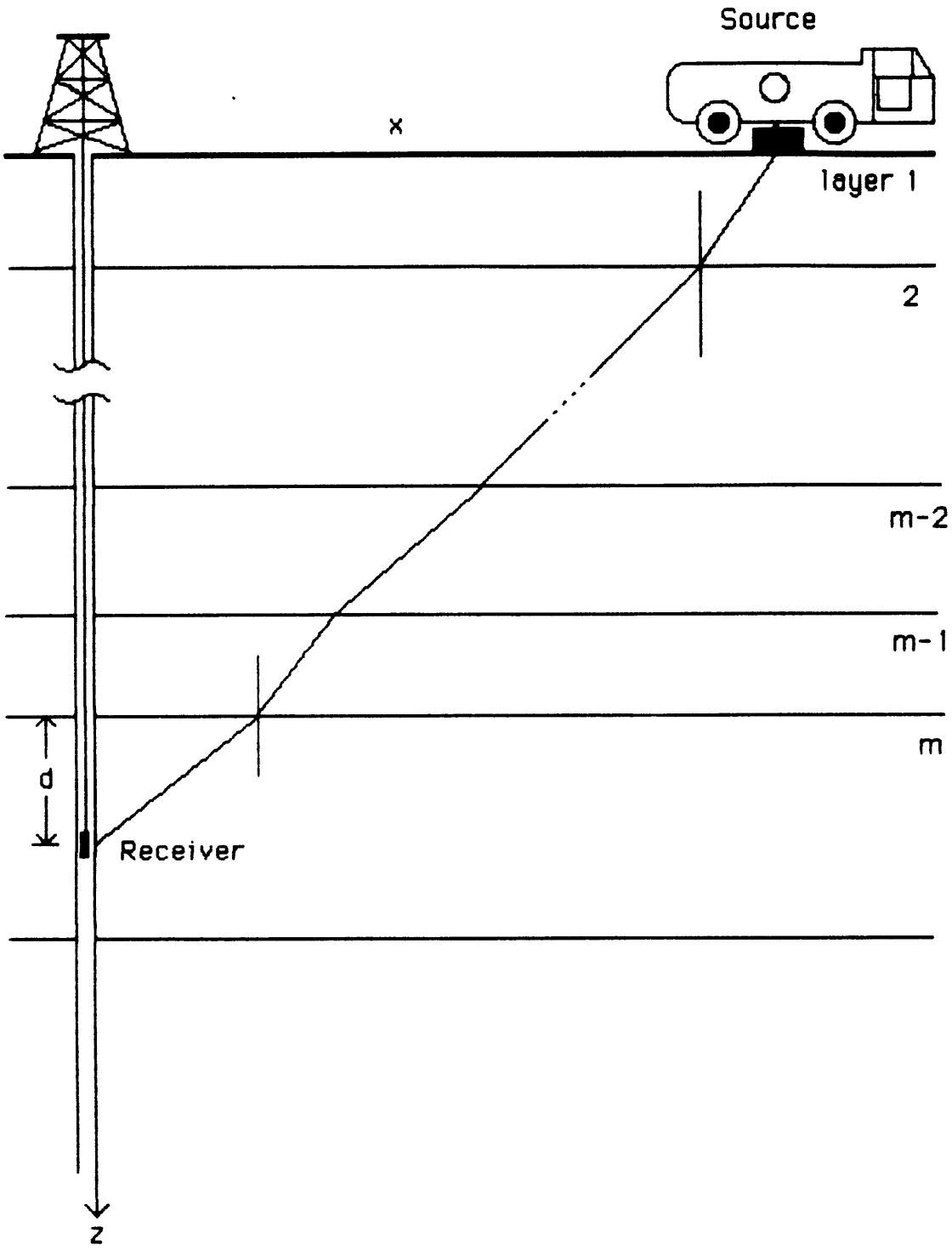


Figure 2.2

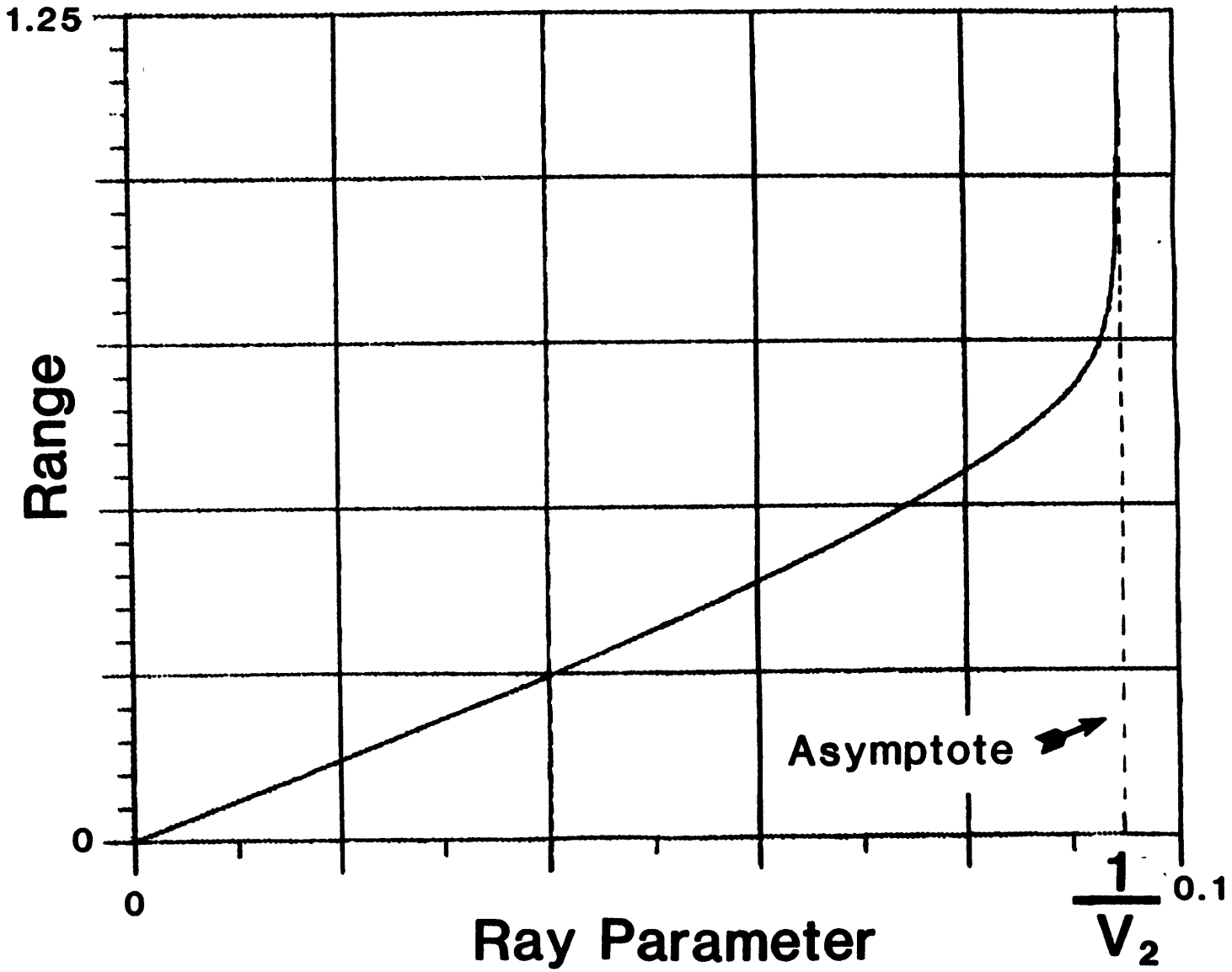


Figure 2.3

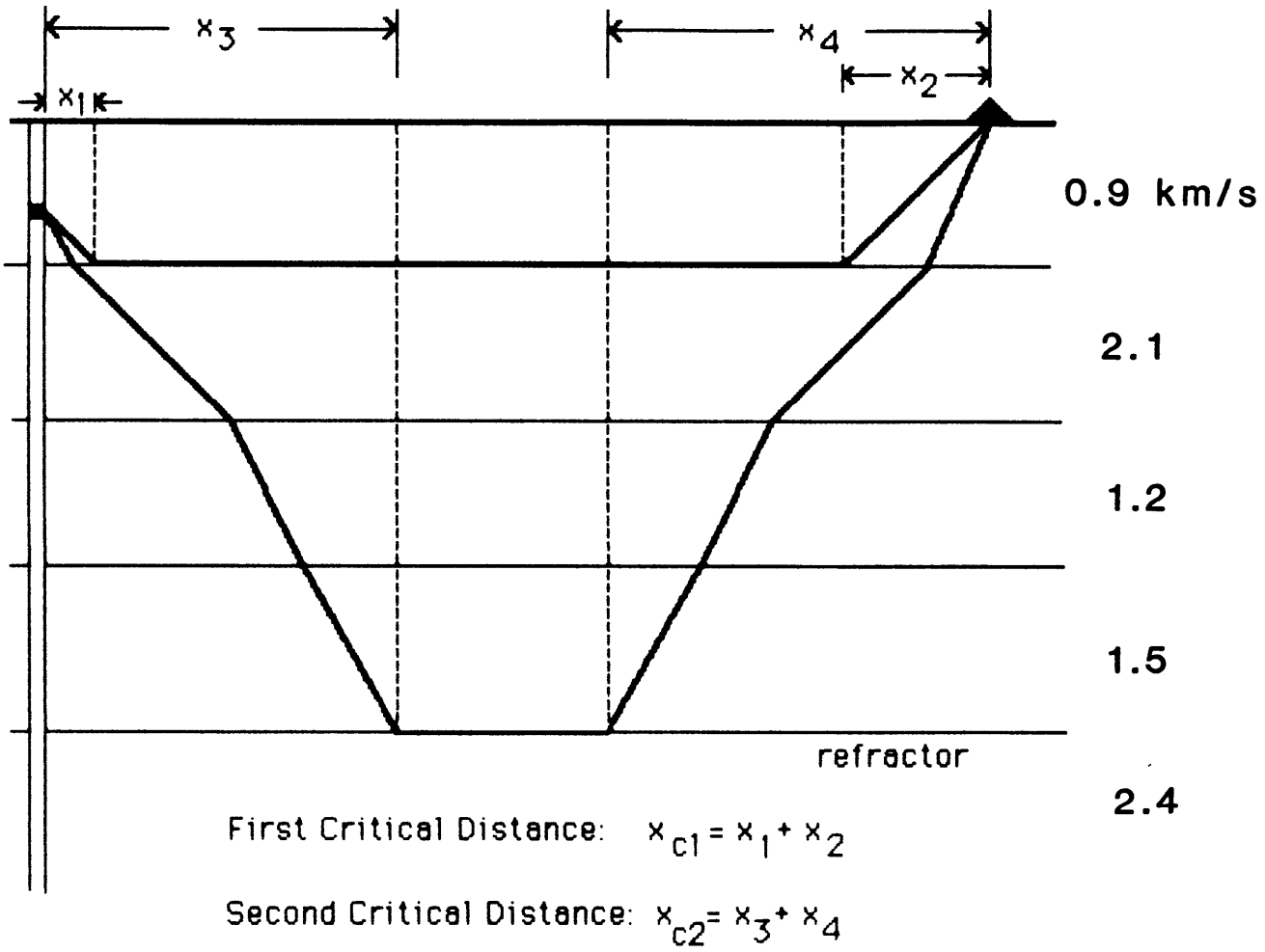

x3
x4
x2
x1
0.9 km/s
2.1
1.2
1.5
refractor
2.4
First Critical Distance: xc1 = x1 + x2
Second Critical Distance: xc2 = x3 + x4

Figure 2.4

CHAPTER 3

Travel Time Inversion

3.1. Introduction

The process of recovering model parameters from data is called inversion or parameter estimation. The relationship of the data to the model is very often nonlinear so that the inverse problem may not be solvable in general by simple means such as linear regression according to the least squares criterion. In fact, there is no simple method of solving nonlinear parameter estimation problems that guarantees results. Nonlinear problems may have all of the problems associated with linear inverse problems (such as eigenvalues of vastly different magnitudes in their linearized versions) plus their own special problems too (such as multiple error minima).

The traditional way of approaching nonlinear inverse problems involves a linearizing approach called Gauss' method. There are various approaches (e.g. Beck and Arnold, 1977; Aki and Richards, 1980), some of which are more useful than others but all of which are useful on some problems. They all seek to minimize the error associated with the parameter choices by systematically (or randomly) searching the model space for the minimum error according to some error criterion, usually, but not always, least squares.

Consider the problem of estimating the P wave velocities of two homogeneous layers given travel time data for rays passing through them in a VSP geometry. Figure 3.1 shows a contoured plot of data variance for part of (V_1, V_2) model space. A minimum is found in the middle of the plot. Without explicitly calculating the data variance at every point in (V_1, V_2) space, how should one find this minimum? In this two dimensional case, it would be feasible to exhaustively search (V_1, V_2) space, but as

more and more layers are added, both exhaustive and Monte Carlo type searches become impractical.

VSP travel time inversion is a refinement of techniques used to estimate seismic velocities from check-shot surveys. The travel time difference between two receiver locations can be used to calculate an average interval velocity at depth (Grant and West, 1965) with

$$\bar{V} = \frac{\Delta z}{\Delta t}. \tag{3.1}$$

A better approximation for small source offsets is the slanted straight ray velocity:

$$\bar{V} = \frac{\Delta z}{\Delta t}\cos\vartheta \tag{3.2}$$

Grant and West (1965) introduced an integral equation method of inverting a travel time vs. depth curve to obtain a function $V(z)$. The operative word here is *curve* since their method relies on values of the travel time derivative with respect to depth and it can be treacherous to estimate derivatives from sampled data. The ray trace integral equation they present is

$$t(z) = \int_0^z \left\{1 - V^2(\zeta)\left[\frac{1}{V^2(z)} - \left[\frac{dt}{dz}\right]^2\right]\right\}^{-\frac{1}{2}} \frac{d\zeta}{V(\zeta)}. \tag{3.3}$$

Stewart (1983) investigated the straight ray methods and the integral equation technique and found that, as expected, the successive refinements produce better results. He also noted that the ray trace integral behaves poorly in the presence of noise due to the boot-strapping nature of the numerical integration required to solve (3.3) for $V(z)$.

Least squares travel time inversion has been used successfully in similar studies in the past, particularly in earthquake seismology. Crosson (1976), for example, performed a joint inversion of earthquake arrival time data for hypocentral parameters

and one-dimensional velocity structure. Stewart (1983) presents a VSP travel time inversion similar to the method presented in this work for calculating layer velocities from zero-offset VSP data. Stewart's algorithm differs from the present method in many details but also in two significant ways. First, this algorithm is able to deal with headwave refractions in situations where they are first arrivals so that the extension to offset VSP inversion is made. Second, the computer implementation of this method should be considerably faster than Stewart's technique since analytic expressions for the elements of the partial derivative (Jacobian) matrix are used instead of finite differencing.

A slightly different tact was taken by Pujol et al. (1985) in their generalized inversion of offset VSP travel time data. They take both travel times and the source offset distance as data and thereby manage to avoid root-finding (see Chapter 2) in their ray tracing. The ray parameters become free parameters like the velocities in this scheme. A problem with the method of Pujol et al. arises since p and V vary independently. Because of this, their product can grow larger than one which renders equation (2.9) meaningless. This problem, they claim, is overcome by careful manual adjustment of the damping parameter. Pujol et al. also ignore headwave refractions because the near surface traces in their study areas show very complicated first arrivals, perhaps due to complex geology (J. Pujol, pers. comm.).

Lines et al. (1984) use travel time data from offset VSP experiments to estimate two-dimensional dip near the borehole. They use velocities obtained from sonic logs but suggest that a simultaneous inversion for layer velocities and dips is possible although subject to ambiguity.

3.2. The Linearized Inverse Problem: Gauss' Method

The formulation of Gauss' method involves setting the partial derivatives of the data error function with respect to all model parameters to zero and then finding the

roots of linearized versions of the resulting simultaneous equations (Menke, 1984; Beck and Arnold, 1977). Gauss' method is a least squares technique since the error function that we attempt to minimize is the sum of the squares of the data residuals.

The formulation, as presented by Beck and Arnold (1977), proceeds as follows. Assume that the data error function is the sum of squares of the residual travel times, i.e.

$$S = \sum_{i=1}^{m} (t_{obs} - t_{calc})^2, \tag{3.4}$$

where the observations t_{obs} form the data vector $\mathbf{d}$ and the computed travel times t_{calc} form the vector $\mathbf{g(m)}$. In vector notation the sum of squares is

$$S = [\mathbf{d} - \mathbf{g(m)}]^{\mathrm{T}} [\mathbf{d} - \mathbf{g(m)}]. \tag{3.5}$$

This is a simplified form of the maximum likelihood expression

$$S_{\mathrm{ML}} = [\mathbf{d} - \mathbf{g(m)}]^{\mathrm{T}} \mathbf{C}_d^{-1} [\mathbf{d} - \mathbf{g(m)}] + (\mathbf{m} - \mathbf{m}_{ap})^{\mathrm{T}} \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_{ap}) \tag{3.6}$$

(Tarantola and Valette, 1982; Aki and Richards, 1980) which seeks to use *a priori* estimates of the model parameters (the vector $\mathbf{m}_{ap}$) to impose stability and uniqueness on the problem. An equivalent technique, damping, was used in this study to ensure stability in the inversion while nonuniqueness has not proved to be a problem. The matrices $\mathbf{C}_d$ and $\mathbf{C}_m$ are data and model covariances, respectively. The explicit use of these covariances in the formulation of this problem was avoided for the sake of simplicity and because the data and model errors are not known well enough to assign covariances. The covariance matrices can also be used to scale the data or model parameters if either group is made up of quantities with different physical dimensions.

The gradient of S with respect to the model parameters is

$$\nabla_m S = 2[- \nabla_m \mathbf{g}^{\mathrm{T}}(\mathbf{m})][\mathbf{d} - \mathbf{g(m)}]. \tag{3.7}$$

Let

$$\mathbf{G}(\mathbf{m}) \equiv [\nabla_m \mathbf{g}(\mathbf{m})]^T.$$

If $\mathbf{m} = \hat{\mathbf{m}}$ is the solution to $\mathbf{g}(\mathbf{m}) = \mathbf{d}$, then $\nabla_m S$ is zero at $\hat{\mathbf{m}}$ since S is minimized so that

$$\mathbf{G}^T(\hat{\mathbf{m}})[\mathbf{d} - \mathbf{g}(\hat{\mathbf{m}})] = 0. \tag{3.8}$$

If $\mathbf{g}(\mathbf{m})$ has continuous first and bounded higher derivatives with respect to $\mathbf{m}$ near $\hat{\mathbf{m}}$, then steps may be taken to approximate (3.8) as a linear function. Since the solution $\hat{\mathbf{m}}$ is not known, replace it in the equation with an estimate $\mathbf{m}_0$ near $\hat{\mathbf{m}}$ and then expand (3.8) to first order:

$$\mathbf{G}^T(\mathbf{m}_0)[\mathbf{d} - \mathbf{g}(\mathbf{m}_0) - \mathbf{G}(\mathbf{m}_0)(\hat{\mathbf{m}} - \mathbf{m}_0)] \approx 0. \tag{3.9}$$

An approximation to the solution $\hat{\mathbf{m}}$ is obtained from (3.9) and is

$$\hat{\mathbf{m}} \approx \mathbf{m}_0 + [\mathbf{G}^T(\mathbf{m}_0)\mathbf{G}(\mathbf{m}_0)]^{-1}\mathbf{G}^T(\mathbf{m}_0)[\mathbf{d} - \mathbf{g}(\mathbf{m}_0)]. \tag{3.10}$$

Due to the assumptions and approximations made, equation (3.10) will not generally find the actual minimum of S but usually a better estimate than $\mathbf{m}_0$, say $\mathbf{m}_1$, so that a more precise statement is

$$\mathbf{m}_1 = \mathbf{m}_0 + [\mathbf{G}^T(\mathbf{m}_0)\mathbf{G}(\mathbf{m}_0)]^{-1}\mathbf{G}^T(\mathbf{m}_0)[\mathbf{d} - \mathbf{g}(\mathbf{m}_0)]. \tag{3.11}$$

Usually a number of iterations will be required to minimize S. Use the new model as the next "initial" model and repeat the process. A more general statement of equation (3.11) is then

$$\mathbf{m}_{k+1} = \mathbf{m}_k + [\mathbf{G}^T(\mathbf{m}_k)\mathbf{G}(\mathbf{m}_k)]^{-1}\mathbf{G}^T(\mathbf{m}_k)[\mathbf{d} - \mathbf{g}(\mathbf{m}_k)] \tag{3.12}$$

which is known as Gauss' method.†

If the assumptions are stretched (e.g. if $\mathbf{m}_k$ is far from $\hat{\mathbf{m}}$), the estimate $\mathbf{m}_{k+1}$ may be worse or much worse than $\mathbf{m}_k$ was. Stability problems such as these are often ameliorated by the addition of damping to (3.12) which changes the solution $\mathbf{m}_{k+1}$

† Gauss' method is known by many names. Other popular names are the Gauss-Newton method, the Newton-Gauss method, and the linearization method.

(Beck and Arnold, 1977). The general formulation is due to Levenberg (1944) but has been modified by many others:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + [\mathbf{G}^T(\mathbf{m}_k)\mathbf{G}(\mathbf{m}_k) + \lambda\Omega]^{-1}\mathbf{G}^T(\mathbf{m}_k)[\mathbf{d} - \mathbf{g}(\mathbf{m}_k)]. \tag{3.13}$$

Although the intent of damping is to add stability to the inversion by preventing $\mathbf{G}^T\mathbf{G}$ from becoming singular, a connection can be made with the maximum likelihood inverse discussed above. Menke (1984) shows that damping small eigenvalues of $\mathbf{G}^T\mathbf{G}$ is equivalent to introducing the *a priori* information that the model parameters are small. Indeed, $\nabla_m S_{ML} = 0$ leads directly to equation (3.13) if the covariance matrices are both expressible in the form $\sigma^2\mathbf{I}$.

Among the simplest modifications to Levenberg's formula (equation 3.13) is an algorithm called the *method of steepest descent*‡ which is obtained from (3.13) by setting $\Omega = \mathbf{I}$ and letting λ become very large so that $\lambda\Omega$ overwhelms $\mathbf{G}^T\mathbf{G}$ and produces

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \lambda^{-1}\,\mathbf{G}^T[\mathbf{d} - \mathbf{g}(\mathbf{m}_k)]. \tag{3.14}$$

Gauss' method and the method of steepest descent are essentially endpoints of a virtual continuum of nonlinear inversion techniques. Steepest descent tends to be quite stable even when very far from the solution in model space but is not particularly accurate when the solution is neared. Gauss' method, on the other hand, may be unstable away from the solution but approaches a final solution very quickly as that solution is neared and the linearizing approximations get better and better (Brown and Dennis, 1972). Another point of view is taken by Beck and Arnold (1977) who admit that the method of steepest descent is valuable for finding the *direction* of descent but point out that the *magnitude* of the step size is arbitrarily chosen.

‡ The method of steepest descent means different things to different people. The definition in the text is from that of Beck and Arnold (1977). Some readers may be more familiar with an iterative one-dimensional line search algorithm which is also called the method of steepest descent (e.g. by Luenberger, 1973).

A compromise approach is advocated by most researchers which uses the global convergence properties of the descent method while far from the solution, but gives more and more weight to Gauss' method contribution as the iterations proceed. Therefore, the important question now is, how does one optimally update the damping factor $\lambda\Omega$? For the $\Omega = \mathbf{I}$ case, the damping parameter λ should be decreased as the iterations proceed, but the best choice of updating algorithms is not generally agreed upon, nor perhaps should it be. Differences in problems and in the shapes of their sum of squares surfaces are so large that one "optimum" method probably does not exist. Because the damping in equation (3.13) changes the final solution (by introducing *a priori* information), it may be desirable to use the smallest value of λ possible as convergence is reached. The details of the inversion formulation including analytical expressions for the partial derivatives are given in Appendix A. Several variations on Gauss' method are presented in Appendix B.

3.3. Pitfalls in Linearized Inversion

There are a number of possible problems associated with Gauss' method. There is no guarantee, for example, that a solution found in this way will be a minimum, let alone a global minimum.

Linear least squares problems always have paraboloidal model error surfaces (Fig. 3.2a) while the error surface associated with linearized problems may have multiple minima as shown in Figure 3.2b (Menke, 1984). For lower dimensional models, graphical methods such as the contour plot shown in Figure 3.1 may be helpful in searching the model space for the global minimum. This becomes all but impossible in higher dimensions so that even if it is known that there is a global minimum to the model error function, there is in general no way of determining solely from the data whether it has been found. This is because the shape of the error surface away from the estimate $\mathbf{m}_k$ cannot be predicted as it can in the linear problem (Menke, 1984). As the

error surface for the linear problem is quadratic, linearizing a nonlinear problem has the effect of fitting a paraboloid through $\mathbf{m}_k$ and returning the minimum of that paraboloid as the revised estimate $\mathbf{m}_{k+1}$ (Fig. 3.3). If the initial estimate $\mathbf{m}_0$ is close enough to the solution $\hat{\mathbf{m}}$ (again, hard to predict), Gauss's method eventually converges because any minimum with continuous derivatives is quadratic locally (Menke, 1984).

Sources of error in the travel times will be discussed in Chapter 4. The covariance matrix $\mathbf{C}_m$ relates these errors in data to model errors. It is defined to be

$$\mathbf{C}_m = \Delta\mathbf{m}\Delta\mathbf{m}^T \equiv \sigma_d^2(\mathbf{G}^T(\mathbf{m}_k)\mathbf{G}(\mathbf{m}_k))^{-1} \tag{3.15}$$

(Aki and Richards, 1980) where $\Delta\mathbf{m}$ is the model error vector and σ_d^2 is the variance of the data. The variance is defined to be

$$\sigma_d^2 = \frac{S}{l-n} \tag{3.16}$$

where l is the number of observations (i.e. travel times) and n is the number of model parameters (velocities). The difference $l-n$ is the number of degrees of freedom of the system. If $\mathbf{C}_m$ is a diagonal matrix, then the model parameters are independent and their individual errors can be estimated by

$$(\Delta V_j)^2 = \sigma_d^2(\mathbf{G}^T(\mathbf{m}_k)\mathbf{G}(\mathbf{m}_k))^{-1}_{jj} \tag{3.17}$$

where the subscripts refer to matrix elements. These error estimates are only valid, strictly speaking, for the linear problem and are usually underestimates of the errors for the nonlinear problem.

The assumption that the covariance matrix is diagonal is generally not valid. Returning to the contoured data variance plot shown in Figure 3.1, it is clear that there is some interdependence between the two velocities. Appendix C describes some of the other information available from these contoured data variance plots.

3.4. Layer Stripping Inversion

The linearized inversion technique presented above requires a reasonable starting model. A much simpler exact inversion, called ***layer stripping***, can be used in many cases to provide such a starting model. It is called layer stripping because one layer velocity at a time is found using only the travel time data and the velocity profile above. As soon as a new velocity is found, it becomes part of the model, one layer having been stripped off the problem.

Pujol et al. (1985) also present a layer stripping velocity inversion which is very similar to a procedure used in the present work. The layer stripping attack works only for waves arriving at the geophone from above so that headwave refractions cannot be included. For zero-offset VSP surveys, however, headwave refractions are usually not seen and inversions of this sort get good results (Stewart, 1983; Pujol et al., 1985).

As outlined in previous sections, the travel time for a ray from the surface to a geophone a distance d from the top of the mth layer is given by

$$t = \sum_{i=1}^{m-1} \frac{h_i}{V_i\sqrt{1-p^2V_i^2}} + \frac{d}{V_m\sqrt{1-p^2V_m^2}}, \tag{3.18}$$

and the range is given by

$$x = \sum_{i=1}^{m-1} \frac{ph_iV_i}{\sqrt{1-p^2V_i^2}} + \frac{pdV_m}{\sqrt{1-p^2V_m^2}}. \tag{3.19}$$

Equations (3.18) and (3.19) may be manipulated to produce an equation for the geophone layer velocity in terms of the ray parameter p and all of the layer velocities above the geophone layer:

$$V_m = \left\{ \frac{d^2}{\left[\frac{x}{p} - \sum_{i=1}^{m-1}\left[\frac{h_iV_i}{\sqrt{1-p^2V_i^2}}\right]\right]^2} + p^2 \right\}^{-\frac{1}{2}}. \tag{3.20}$$

If (3.20) is inserted into (3.18), then p may be found numerically given an observed t.

Equation (3.20) may then be used to obtain V_m which is of course the goal.

If geologic layer thicknesses are unknown *a priori*, then each geophone may be considered to mark the bottom of a new layer. If geologic reasoning or well log results suggest particular layer thicknesses, then all of the velocities found from equation (3.20) for a given layer may be averaged to determine layer velocities.

It is desirable to compute both p and V_m to high precision in order to avoid excessive error accumulation because of the deterministic nature of the layer stripping inversion. Even in noise-free data, effects like round-off errors can adversely affect the velocities. Deeper layer velocities are often poor estimates even if the data are fairly noise-free. This algorithm is so susceptible to noise in the data that one wild data point can sometimes ruin a model, especially if it is shallow in the section and the model layering is based on the geophone spacing.

As mentioned above, the layer stripping method does not require any sort of initial guess at the solution so that it can be used to generate the initial guess required by Gauss' method or one of the related algorithms. This tactic is especially promising if headwave refractions constitute only a small number of the first arrivals. Layer stripping is certainly better for this application than calculating interval velocities since the latter method usually neglects source offset and always assumes straight rays.

3.5. Numerical Experiments with Synthetic Data

A number of synthetic data sets were employed to investigate the properties of the inversion methods presented in this study. Table 3.1 presents a summary of the experiments. All of the algorithms described in Appendix B were implemented on a Vax 11/780 computer. Table 3.2 presents a comparison of these methods for some of the present experiments in terms of the number of iterations required for convergence to an arbitrary minimum.

The best algorithms seem to be the straight Gauss' method and a modification of the damping technique given by Brown and Dennis (1972). The statistics recorded in Table 3.2 are misleading since the finite difference Levenberg-Marquardt algorithm uses, as the name implies, finite difference approximations to the partial derivatives. It is much less efficient than using analytic partial derivatives as is discussed in Appendix B. The choice of the best algorithm appears to depend on the amount of noise in the travel time data and the number of free parameters in the problem.

The first travel time data set (Experiment A) was calculated from a simple five layer model with increasing velocity with depth. Figure 3.4 shows 50 theoretical ray paths from a source offset of 91.4 m while the velocity model and time vs. depth curve are shown in Figure 3.5. Both the layer stripping and Gauss' method inversion techniques were easily able to recover the velocities to high precision. The effect of adding increasing amounts of noise to the travel times is shown in Figures 3.6 (B) and 3.7 (C) The normally distributed random noise had a zero mean and standard deviations of 2 ms (henceforth to be denoted $N(0,4\ \text{ms}^2)$) and 5 ms, respectively. As expected, the models obtained from inversion look less and less like the true model as more noise is added. The velocities obtained from the inversion tend to oscillate about the true model velocities, apparently attempting to compensate for velocity errors above. This effect is an argument for smoothing results obtained from noisy field data if the noise is unsystematic.

In the next experiment (D) the same model was overparameterized by subdividing each layer into three. That is, each of the 183 m thick layers was divided into three 61 m thick layers with the same velocity. The same travel time data were used to invert for fifteen layers instead of five. The results are shown in Figure 3.8. Figures 3.9 (E) and 3.10 (F) show the inversion results in the presence of increasing noise. Once again the compensating velocity oscillation is visible. The degradation of the inversion

results due to noise is much more severe in this experiment since the redundancy of the data set has been reduced. The 50 travel times were sufficient to easily resolve five layer velocities but were not capable of resolving fifteen layer velocities in the presence of noise. More travel times would improve the results.

The source was moved to 914 m away from the borehole for the next experiments. Figure 3.11 shows the theoretical ray paths for this trial. Travel time data were generated from the same five layer model for the same receiver depths and inverted for five and fifteen layer velocities as before. The results are shown in Figures 3.12 (G) and 3.13 (H). Both inversions successfully recovered the true model although the fifteen layer model shows small oscillations in velocity about the true value in the top three layers. The source of error in the upper layer may result from contributions from travel times to deeper receivers. Whereas the deeper layers are affected by only a few travel times and the associated partial derivatives, the shallower layer velocities are affected by all of the data and subject to errors throughout the model. This problem will reappear in similar configurations during subsequent trials.

A new model, with ten layers and low velocity zones, was used to calculate travel times for the next experiment. The ray paths for source offsets of 91.4 m and 914 m are shown in Figures 3.14 and 3.15. The models obtained via inversion are nearly exact fits as can be seen from Figures 3.16 (I) and 3.17 (J). This experiment indicates that low velocity zones are not a problem for this algorithm.

The next numerical experiment looked at the effect of using multi-offset travel time data in a single inversion. This experiment was based on the same ten-parameter model used in I and J. This time, the 100 travel times were from three source offsets: 91.4 m, 305 m, and 914 m. Experiment K (no noise) converged very quickly to the true model (shown in Fig. 3.16a) while Experiments L ($N(0,4\ ms^2)$) and M ($N(0,25\ ms^2)$) converged rapidly to degraded versions of the true model. The results show that the solu-

tion to the inverse problem is about the same whether there are 100 receiver positions and one shot point or 33 receiver positions at each of three shot points. However, the model errors are often smaller in the multi-offset case.

Frequently the layer thicknesses are not known *a priori* or the true velocity structure is not parameterizable as flat layers. One strategy that may be used in such cases is to assign arbitrary layer thickness to the model and proceed as usual from there. The response of the inversion technique to this situation was tested by calculating travel times from a model with uneven layer thicknesses and inverting for a model with constant thickness layers. The effect of normally distributed random noise was investigated for these travel time data. Figures 3.18 through 3.20 (Expts. N – P) show the inversion result and the travel time vs. depth curve for standard deviations of 0 ms (i.e. no noise), 2 ms, and 5 ms, respectively. For the noise-free case (Fig. 3.18), the inversion result is about as close as possible considering that a perfect fit is impossible. Only at the very top, where the first arrivals were headwaves, did the inversion fail in some sense. It is similar to a problem discussed earlier. The addition of noise to travel times seems to affect the deeper velocity structure more than the near surface velocities.

A good starting model of course speeds convergence. In order not to bias these experiments by providing unrealistically good or bad initial models, the layer stripping inversion, which does not require a starting model, was used to obtain the first approximation. For simple structures and noise-free data, layer stripping is sufficient. In slightly more complex situations, it provides a starting model often just a few iterations away from the best fit. Figure 3.21 shows an example of the output of the layer stripping algorithm for the case of a simple five layer model and travel time data with $N(0,4\ \mathrm{ms}^2)$ added noise (Experiment B). The final inversion result is shown in Figure 3.6. An unfortunate aspect of layer stripping in the presence of noise is that compen-

sating oscillations in velocity increase with depth due to the single pass nature of the algorithm. The model shown in Figure 3.21 is only slightly affected by this phenomenon; it is easy to see that the velocity oscillation could get out of control if too many more layers are added. Figure 3.22 shows an example of this unbounded velocity oscillation for Experiment 0. To prevent such wild (and unphysical) velocity oscillations, the layer stripping program was written to constrain velocities to be in the range $1.2 \leq V \leq 8.2$ km/s. Still, an initial model like that shown in Figure 3.22 (for Expt. 0) will cause the ultimate inversion to require many iterations. In this situation and others like this, a constant velocity initial model ($V = 3.1$ km/s) was used with more satisfactory results. Whether to use the layer stripping results or a constant velocity model as the trial model for the full inversion depends on the model determined from layer stripping. Wild oscillations in velocity usually indicate a poor fit caused perhaps by the presence of noise in the travel times.

Experiment	Number of Parameters	Number of Data Points	Low Velocity Zone	Source Offset (m)	Added Noise (ms)
A	5	50	No	91.4	0
B	5	50	No	91.4	2
C	5	50	No	91.4	5
D	15	50	No	91.4	0
E	15	50	No	91.4	2
F	15	50	No	91.4	5
G	5	50	No	914	0
H	15	50	No	914	0
I	10	100	Yes	91.4	0
J	10	100	Yes	914	0
K	10	34 33 33	Yes	91.4 305 914	0
L	10	34 33 33	Yes	91.4 305 914	2
M	10	34 33 33	Yes	91.4 305 914	5
N	36	120	Yes	610	0
O	36	120	Yes	610	2
P	36	120	Yes	610	5

Table 3.1. Summary of numerical experiments used to test travel time inversion algorithm.

Algorithm	Experiment					
	D	G	H	I	J	N
Gauss	3	38	79	3	68	†
Scaled Levenberg Damping	3	*	59	3	84	85
Modified Brown & Dennis	3	38	72	3	52	45
Modified Box-Kanemasu	7	77	†	7	*	†
Finite Difference Levenberg-Marquardt	7	5	17	5	8	11
Convergence Criterion (Iterations continued until data variance falls below value)	10^{-10}	10^{-8}	10^{-8}	10^{-10}	10^{-8}	3×10^{-7}

† Computations stopped when model parameters diverged.
* Computations stopped when convergence was not attained after 100 iterations.

Table 3.2. Comparison of algorithms presented in Appendix B for various numerical experiments summarized in Table 3.1.

Figure Captions

Figure 3.1. a) Velocity vs. depth curve for a simple two layer velocity model. b) Travel time vs. depth curve for the model in (a). c) Contoured data variance for travel times shown in (b). The data variance is $\sum(t_{obs} - t_{calc})^2/f$ where f is the number of degrees of freedom in the system. f is defined as the number of data points (t_{obs} in this case) less the number of parameters. In a properly parameterized system which is free from noise, the minimum data variance is zero.

Figure 3.2. a) The model error surface for a one-dimensional linear problem. The error surface for a linear problem is always a paraboloid. b) The model error surface for a nonlinear problem may have multiple local extrema (after Menke, 1984).

Figure 3.3. The error surface for a nonlinear problem (solid line) and for the linearized version of the same problem (dashed line; after Menke, 1984).

Figure 3.4. Theoretical ray paths for Experiment A. Note exaggerated horizontal scale.

Figure 3.5. a) Velocity vs. depth profile for Expt. A. Solid line is true model; dashed line is inversion result. b) Travel time vs. depth curve for Expt. A.

Figure 3.6. a) Velocity vs. depth profile for Expt. B. Solid line is true model; dashed line is inversion result. b) Travel time vs. depth curves. Solid line is curve for Expt. A; dashed curve for Expt. B ($N(0,4\ \text{ms}^2)$ added).

Figure 3.7. a) Velocity vs. depth profile for Expt. C. Solid line is true model; dashed line is inversion result. b) Travel time vs. depth curves. Solid line is curve for Expt. A; dashed curve for Expt. C ($N(0,25\ \text{ms}^2)$ added).

Figure 3.8. a) $V(z)$ vs. z profile for Expt. D. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curve for Expt. D.

Figure 3.9. a) $V(z)$ vs. z profile for Expt. E. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curves. Solid curve is for Expt. D; dashed curve for Expt. E ($N(0,4\ \text{ms}^2)$ added).

Figure 3.10. a) $V(z)$ vs. z profile for Expt. F. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curves. Solid curve is for Expt. D; dashed curve for Expt. F ($N(0,25\ \text{ms}^2)$ added).

Figure 3.11. Theoretical ray paths for Expts. G and H.

Figure 3.12. a) $V(z)$ vs. z profile for Expt. G. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curve for Expt G.

Figure 3.13. a) $V(z)$ vs. z profile for Expt. H. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curve for Expt H.

Figure 3.14. Theoretical ray paths for Expt. I.

Figure 3.15. Theoretical ray paths for Expt. J.

Figure 3.16. a) $V(z)$ vs. z profile for Expt. I. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curve for Expt I.

Figure 3.17. a) $V(z)$ vs. z profile for Expt. J. Solid line is true model; dashed line is inversion result. b) $t(z)$ vs. z curve for Expt J.

Figure 3.18. a) Inversion result for Expt. N (dashed line) drawn over true model (solid line) Note that a perfect fit is impossible. b) $t(z)$ vs. z curve for Expt. N.

Figure 3.19. a) Inversion result for Expt. O (dashed line) drawn over true model (solid line) b) $t(z)$ vs. z curves. Solid line is curve for Expt. N; dashed line is curve for Expt O ($N(0,4\ \text{ms}^2)$ added).

Figure 3.20. a) Inversion result for Expt. P (dashed line) drawn over true model (solid line) b) $t(z)$ vs. z curves. Solid line is curve for Expt. N; dashed line is curve for Expt P ($N(0,25\ \text{ms}^2)$ added).

Figure 3.21. a) Dashed line is layer stripping result ($V(z)$ vs. z model) used for trial model in Expt. B; solid line is true model. b) Travel time vs. depth curves. Solid line is curve calculated from true model; dashed line reflects added $N(0,4\ \text{ms}^2)$ noise.

Figure 3.22. a) Dashed line is layer stripping result ($V(z)$ vs. z model) used for trial model in Expt. O; solid line is true model. b) Travel time vs. depth curves. Solid line is curve calculated from true model; dashed line reflects added $N(0,4\ \text{ms}^2)$ noise.

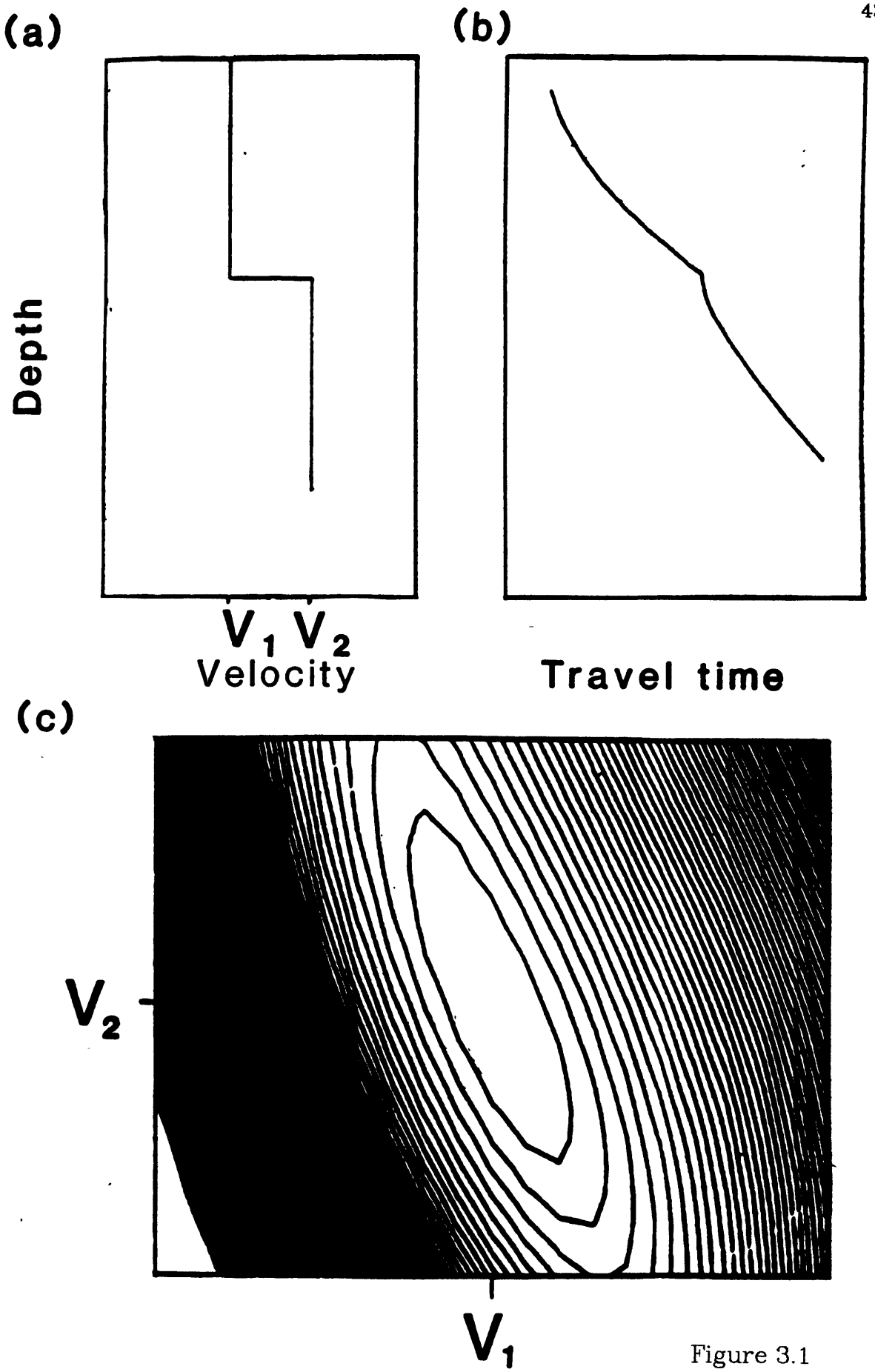
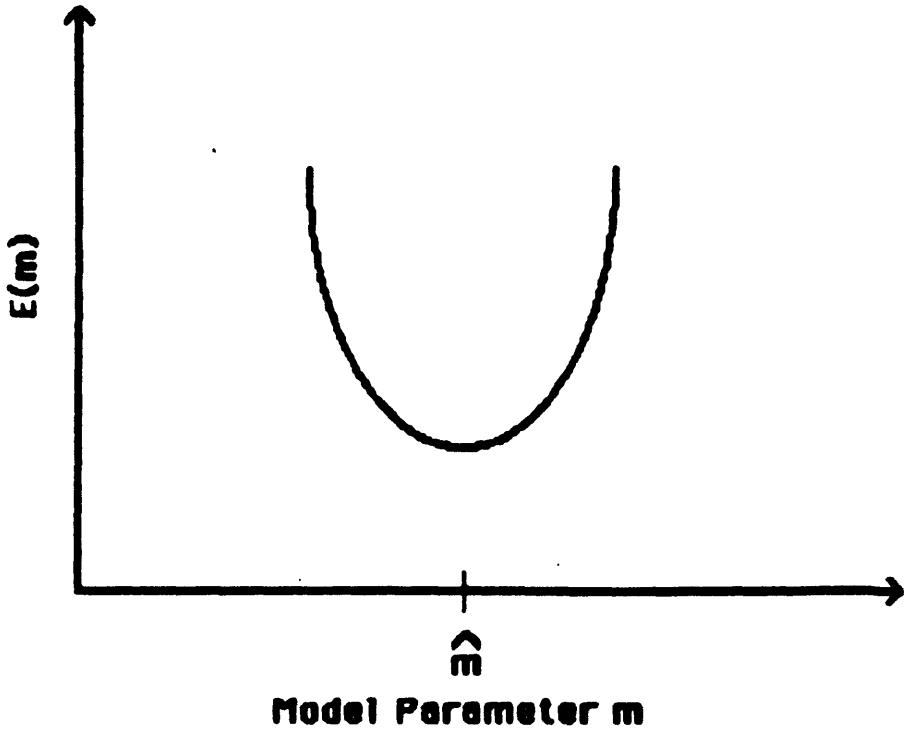
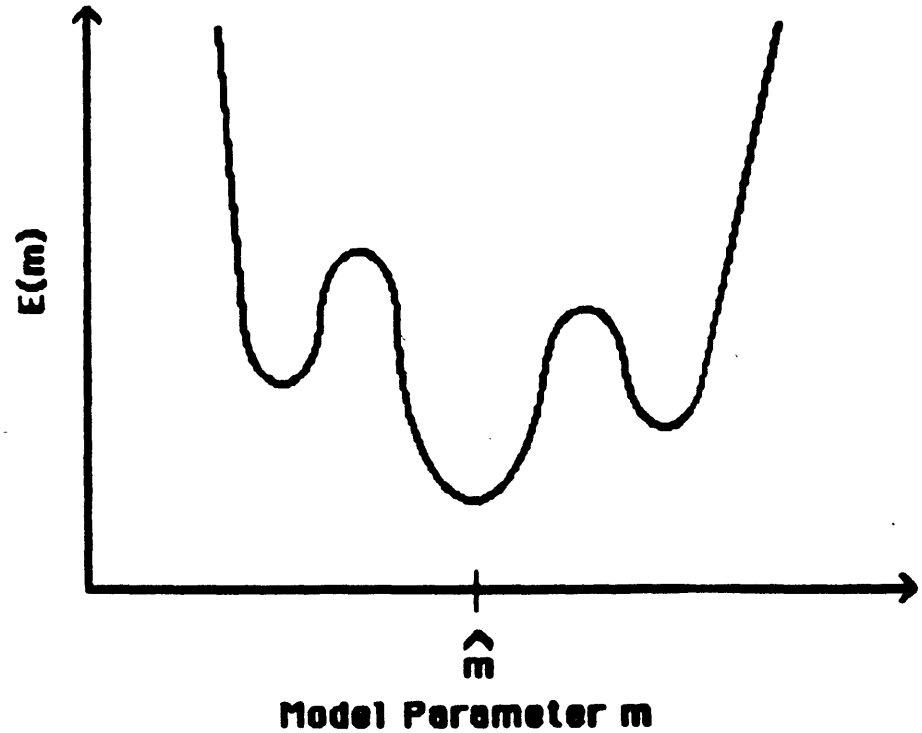


Figure 3.1



(a)



(b)

Figure 3.2

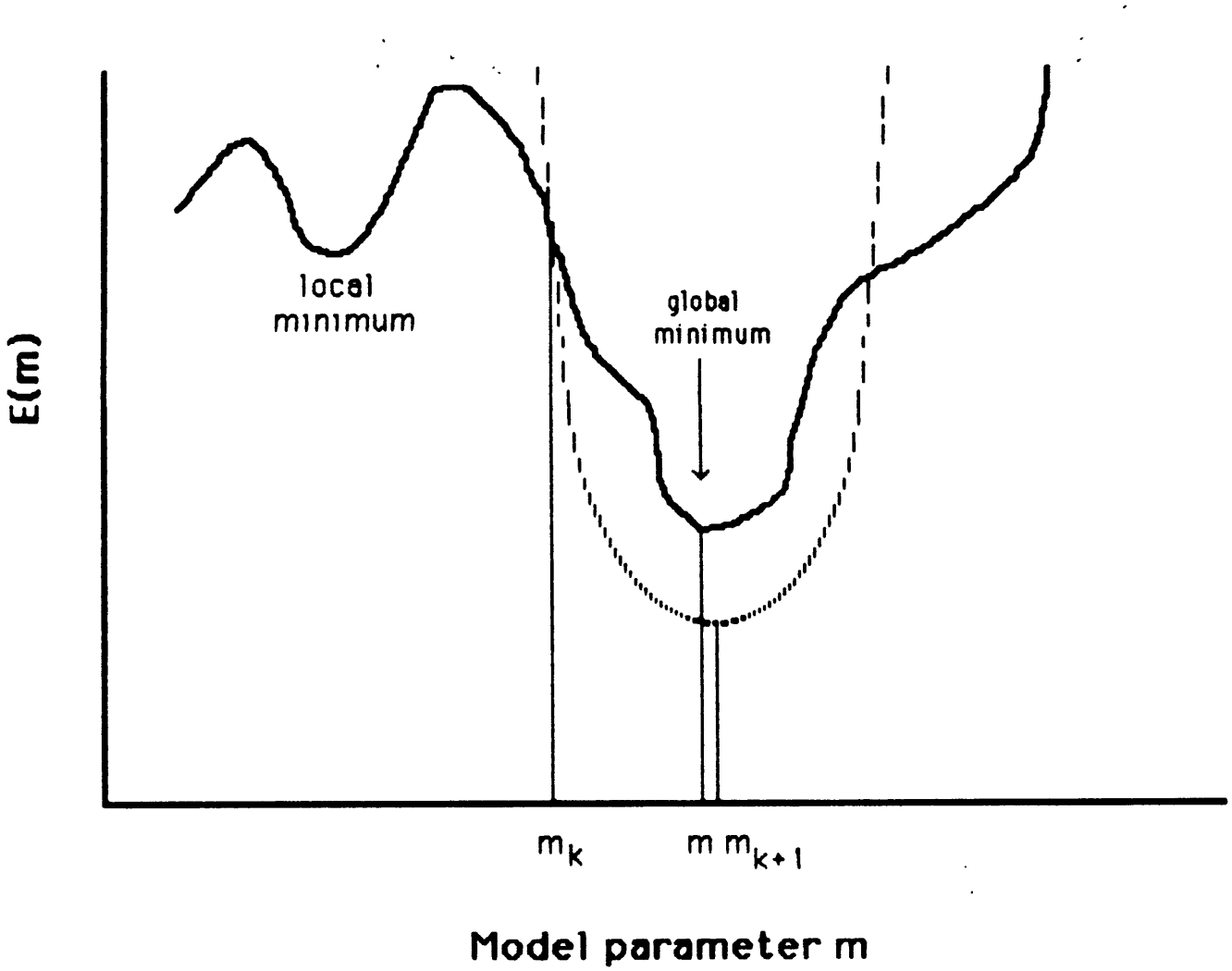


Figure 3.3

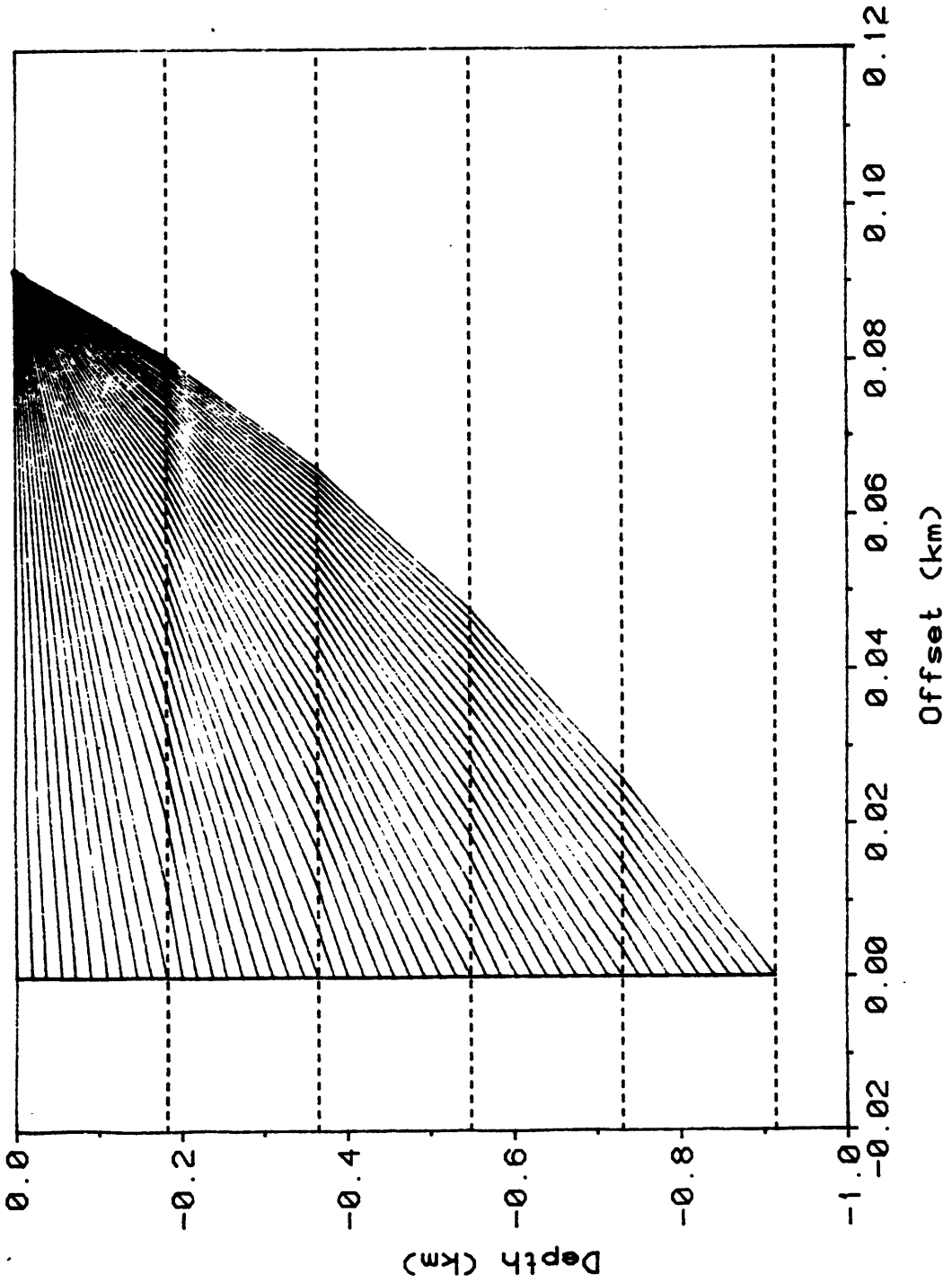

Depth (km)
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Offset (km)
-0.02
0.00
0.02
0.04
0.06
0.08
0.10
0.12

Figure 3.4

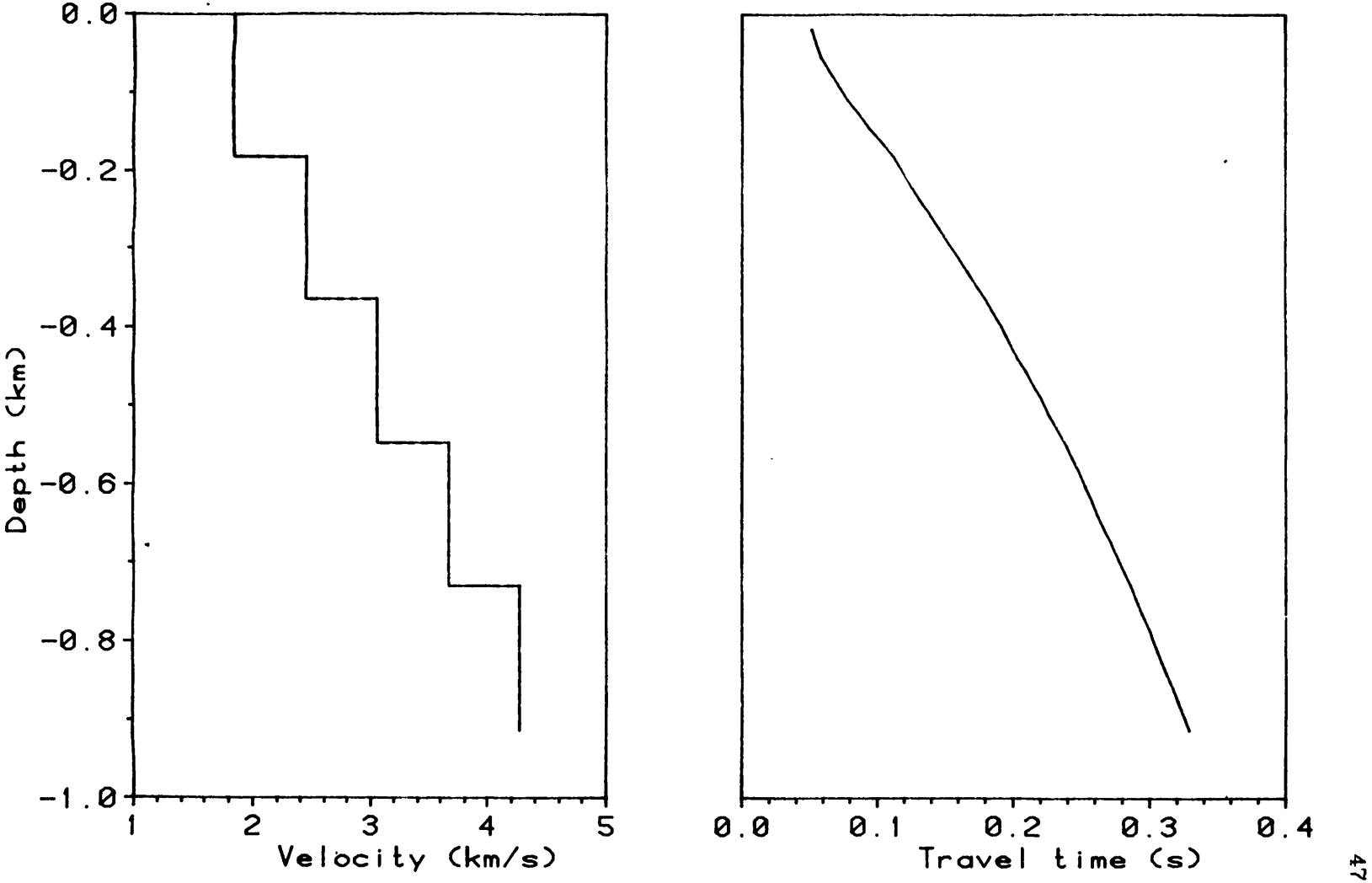
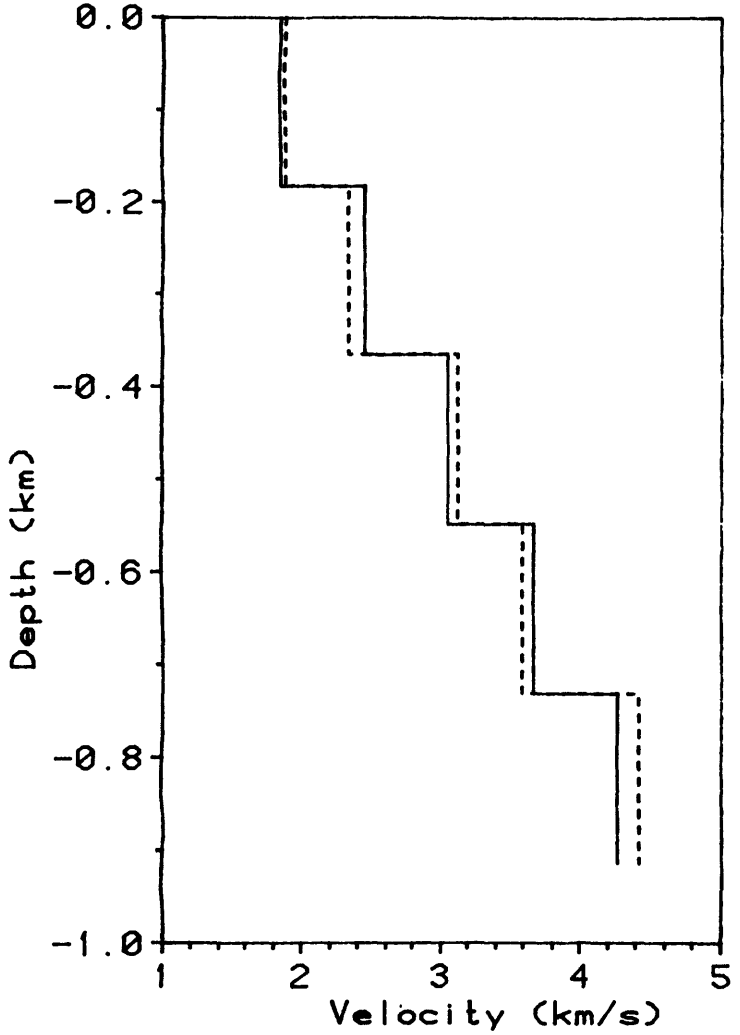
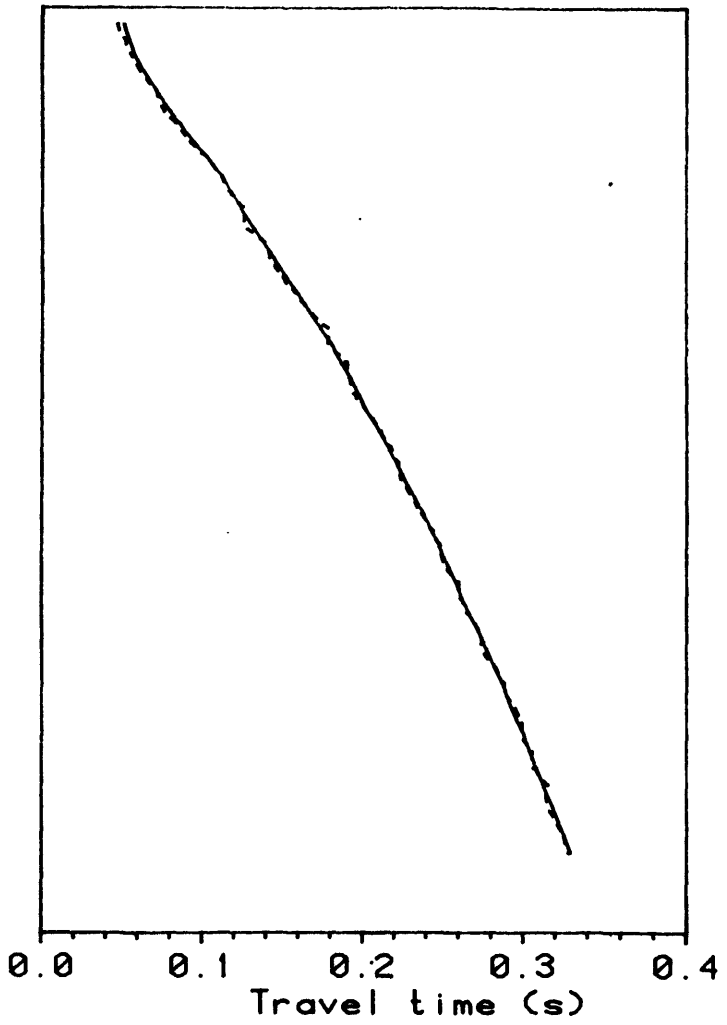

0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)
1
2
3
4
5
Velocity (km/s)
(a)
0.0
0.1
0.2
0.3
0.4
Travel time (s)
(b)

Figure 3.5


0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)
1
2
3
4
5
Velocity (km/s)

(a)


0.0
0.1
0.2
0.3
0.4
Travel time (s)

(b)

Figure 3.6

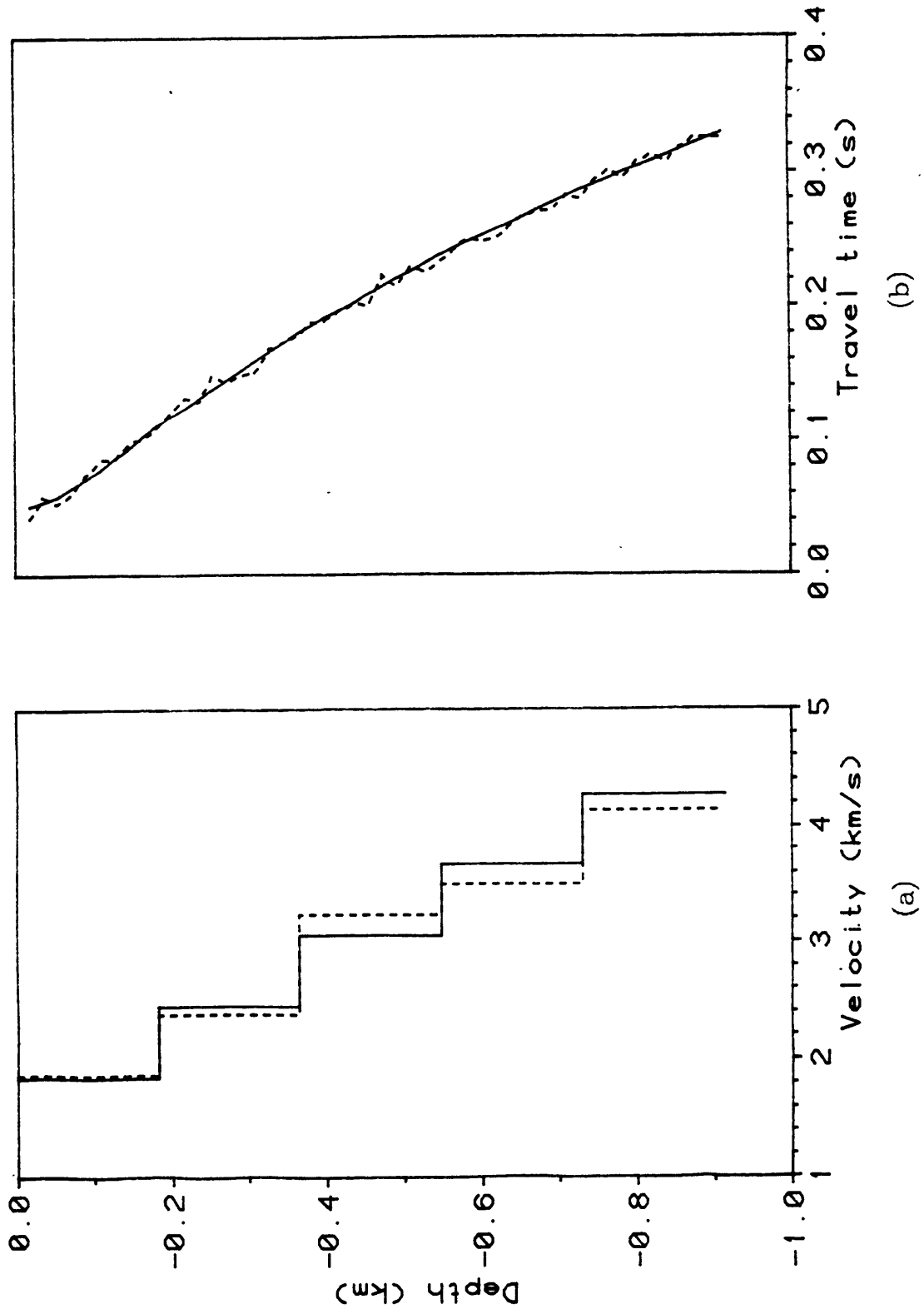
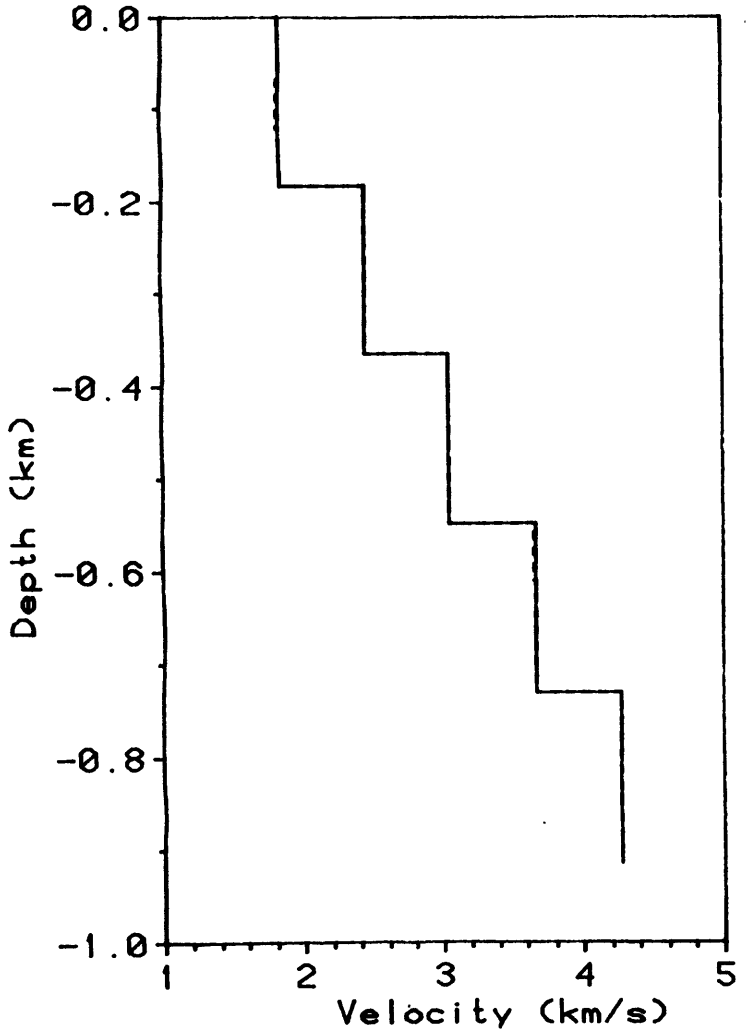
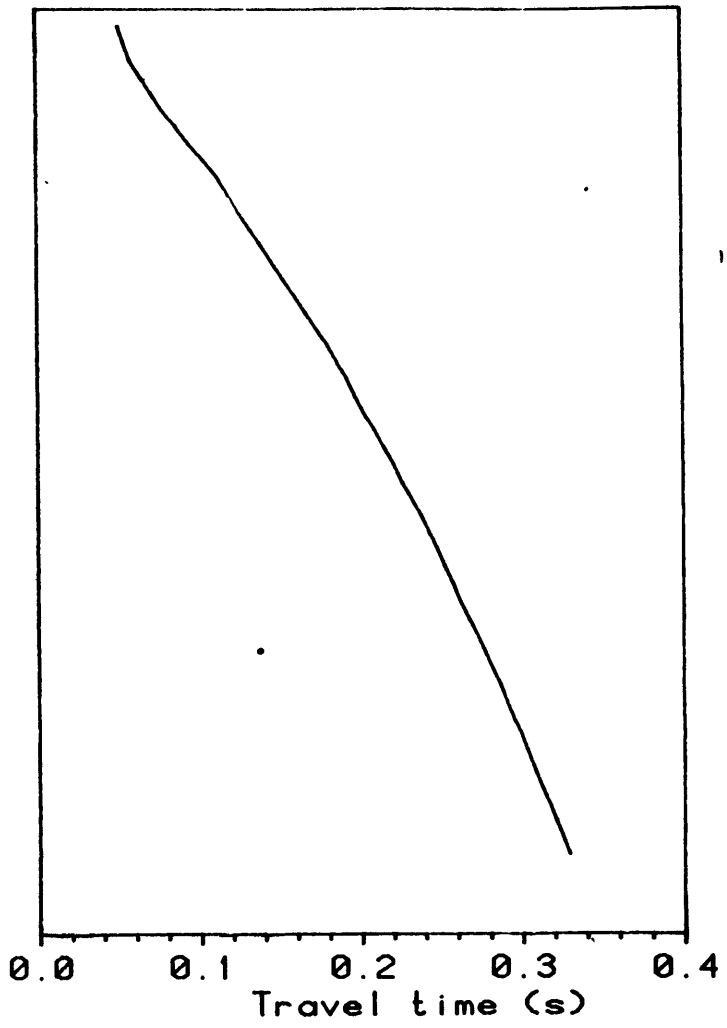

0.0
0.1
0.2
0.3
0.4
Travel time (s)
(b)
1
2
3
4
5
Velocity (km/s)
(a)
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)

Figure 3.7


0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)
1
2
3
4
5
Velocity (km/s)

(a)


0.0
0.1
0.2
0.3
0.4
Travel time (s)

(b)

Figure 3.8

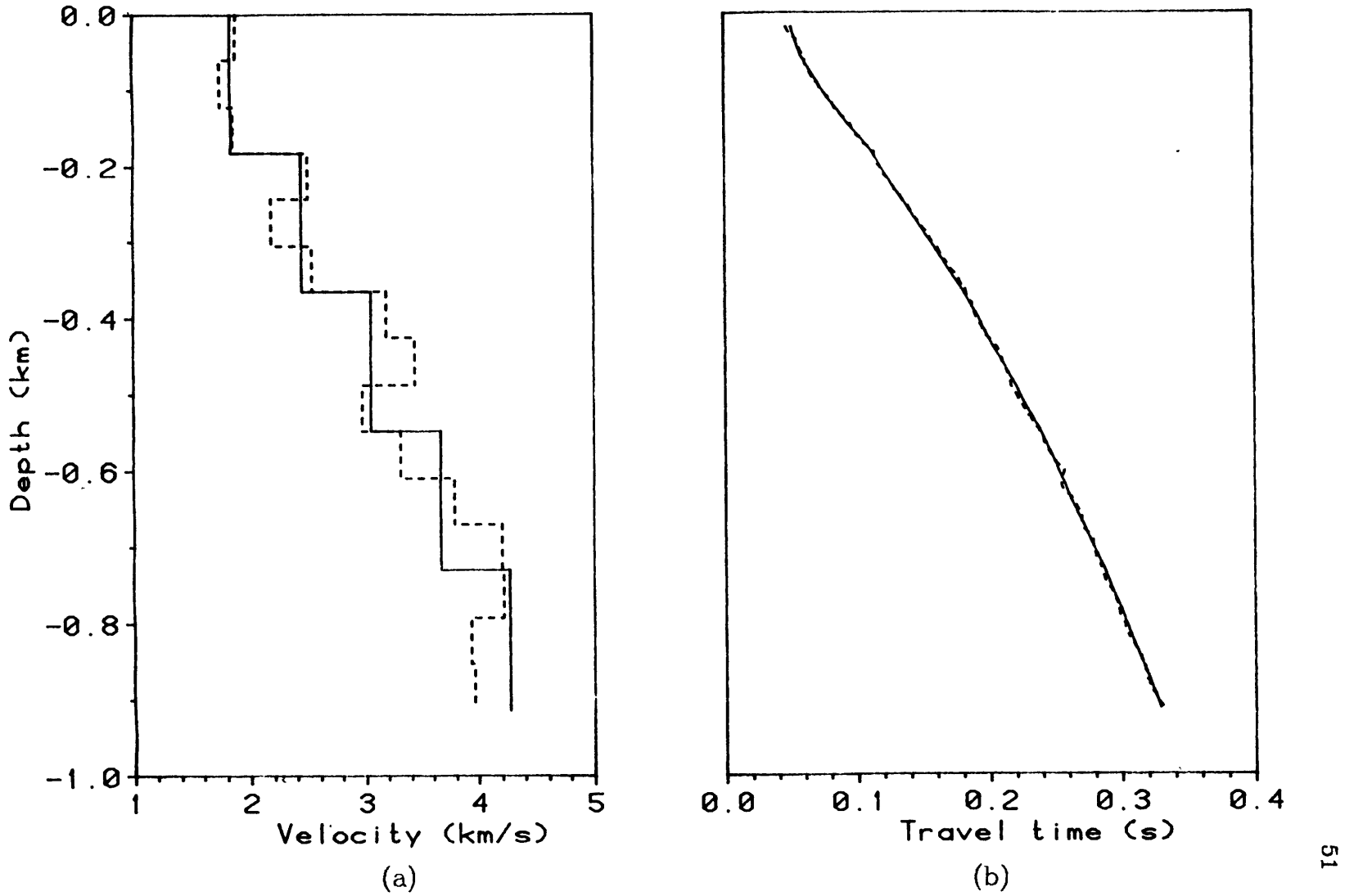

0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)
1
2
3
4
5
Velocity (km/s)
(a)
0.0
0.1
0.2
0.3
0.4
Travel time (s)
(b)

Figure 3.9

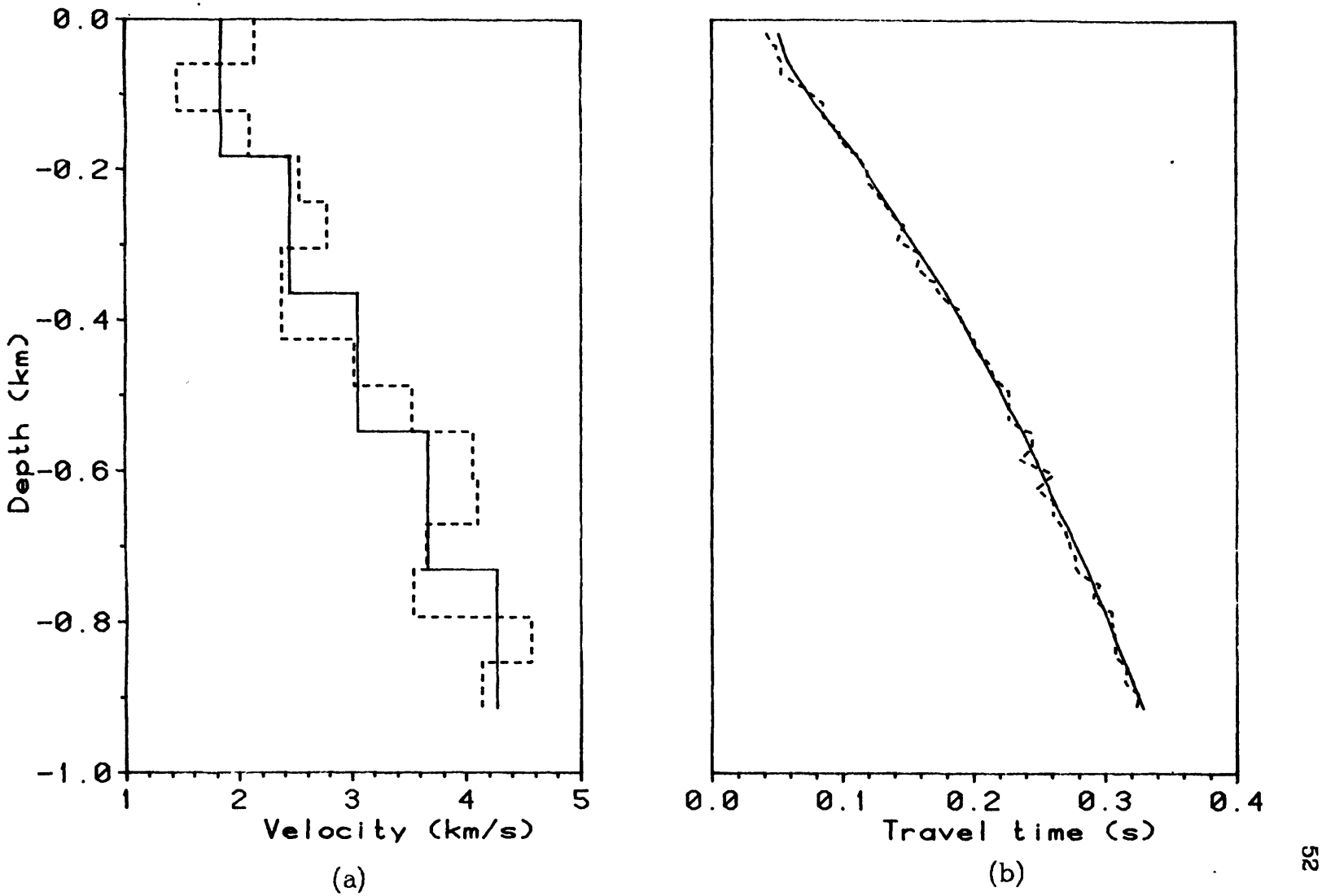

0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)
1
2
3
4
5
Velocity (km/s)
(a)
0.0
0.1
0.2
0.3
0.4
Travel time (s)
(b)

Figure 3.10

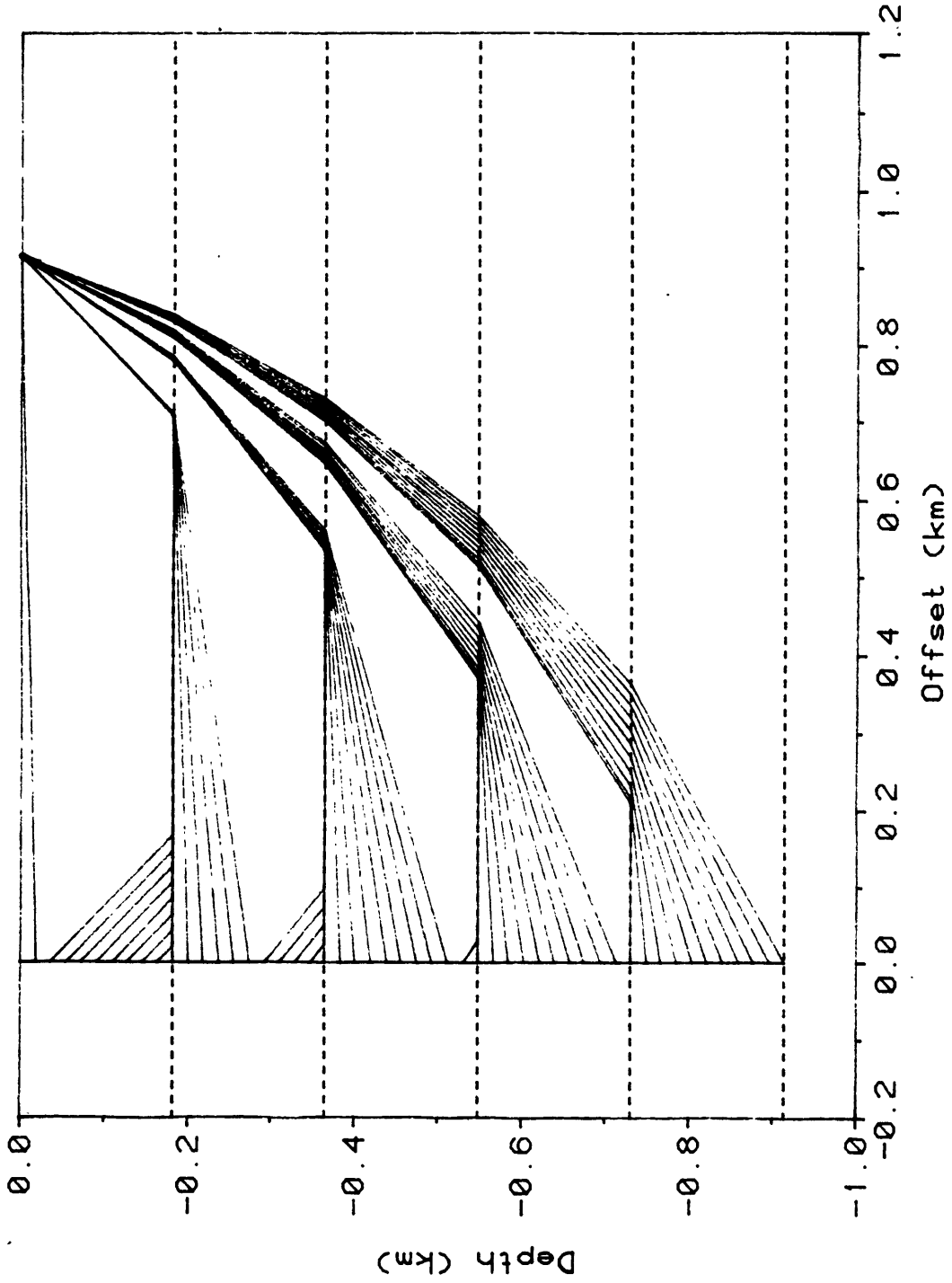


Figure 3.11

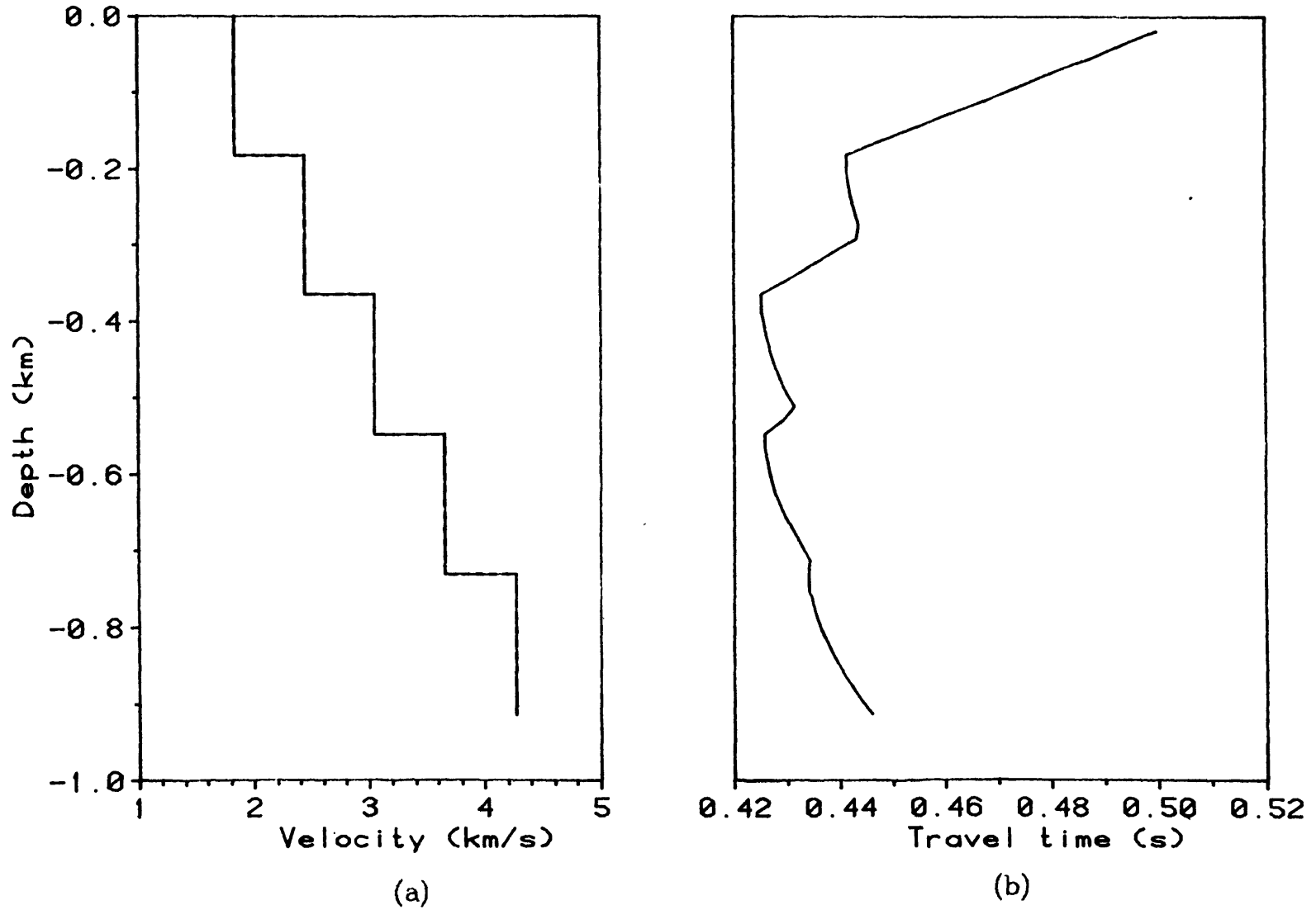
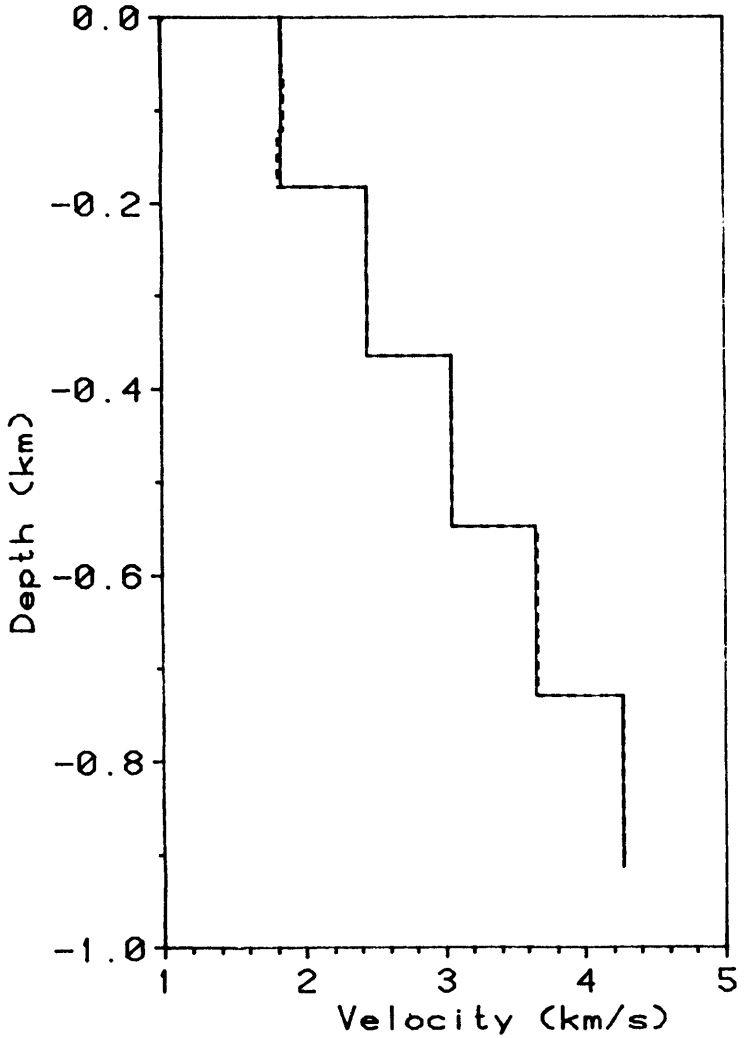
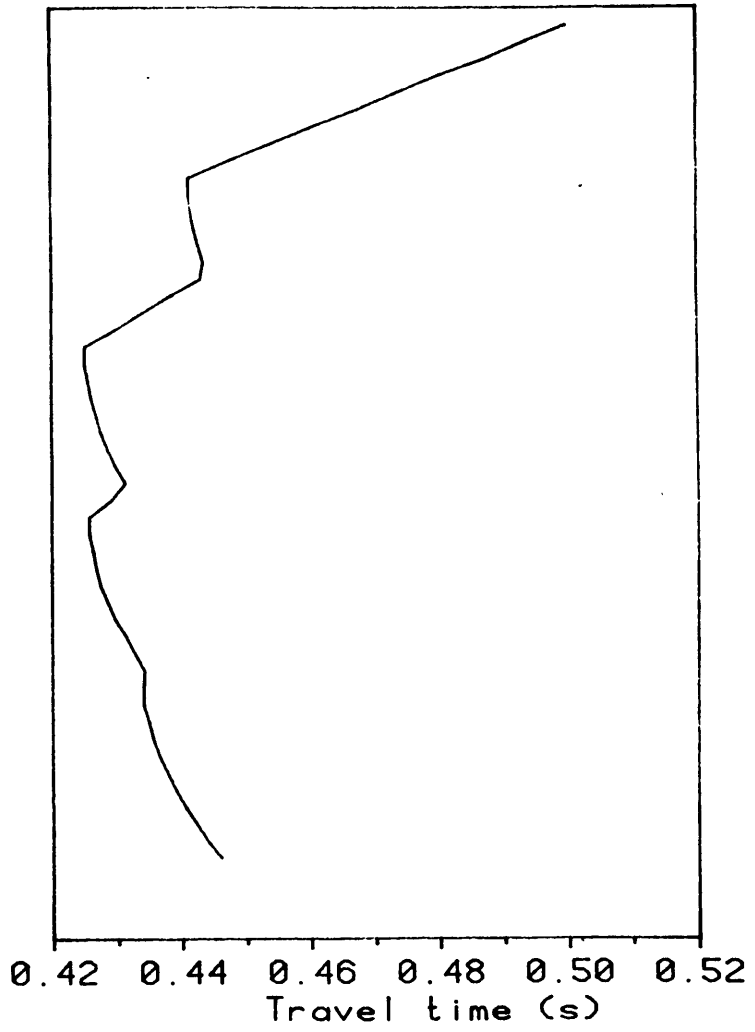

0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)
1
2
3
4
5
Velocity (km/s)
(a)
0.42
0.44
0.46
0.48
0.50
0.52
Travel time (s)
(b)

Figure 3.12



(a)



(b)

Figure 3.13

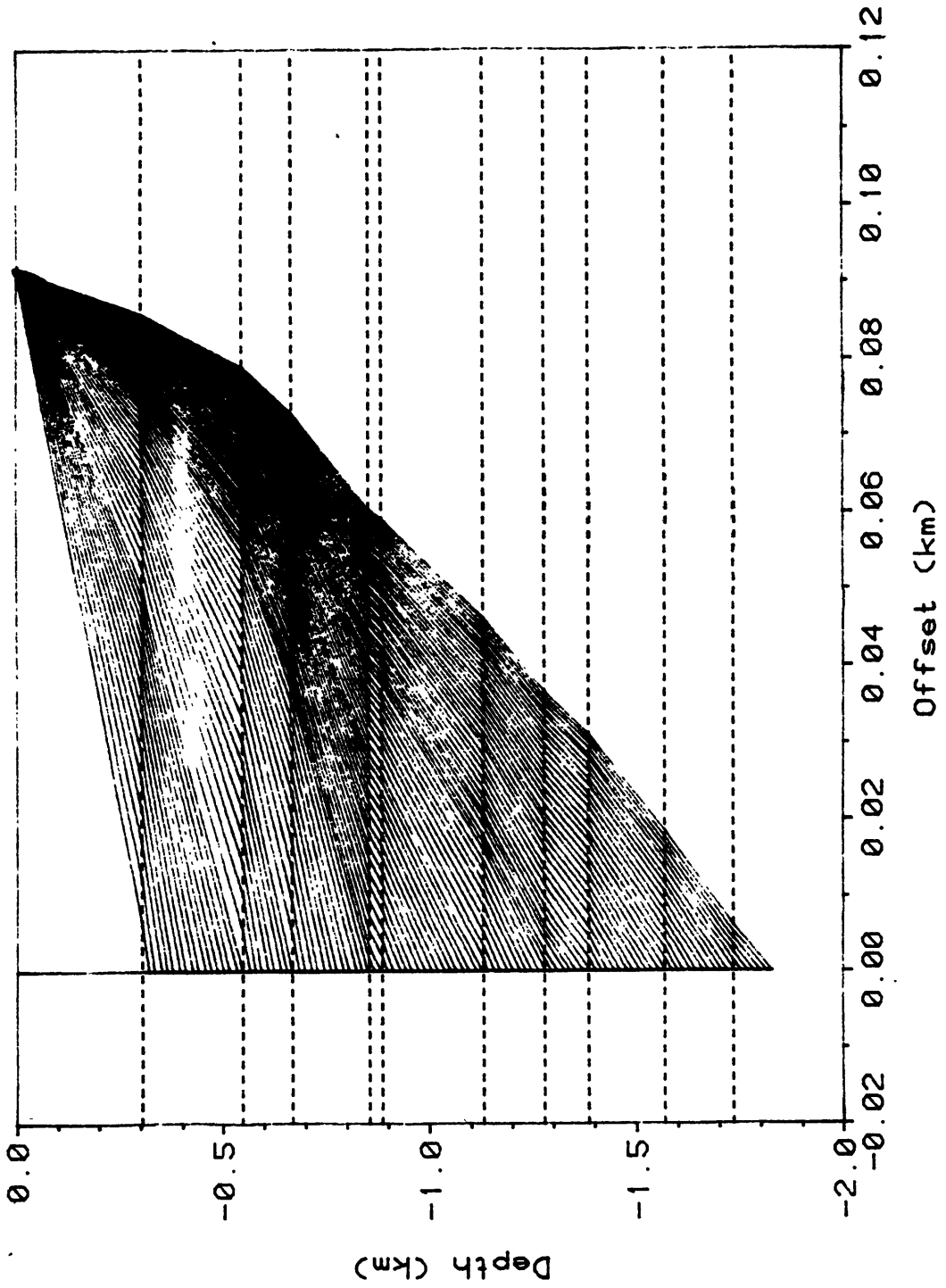


Figure 3.14

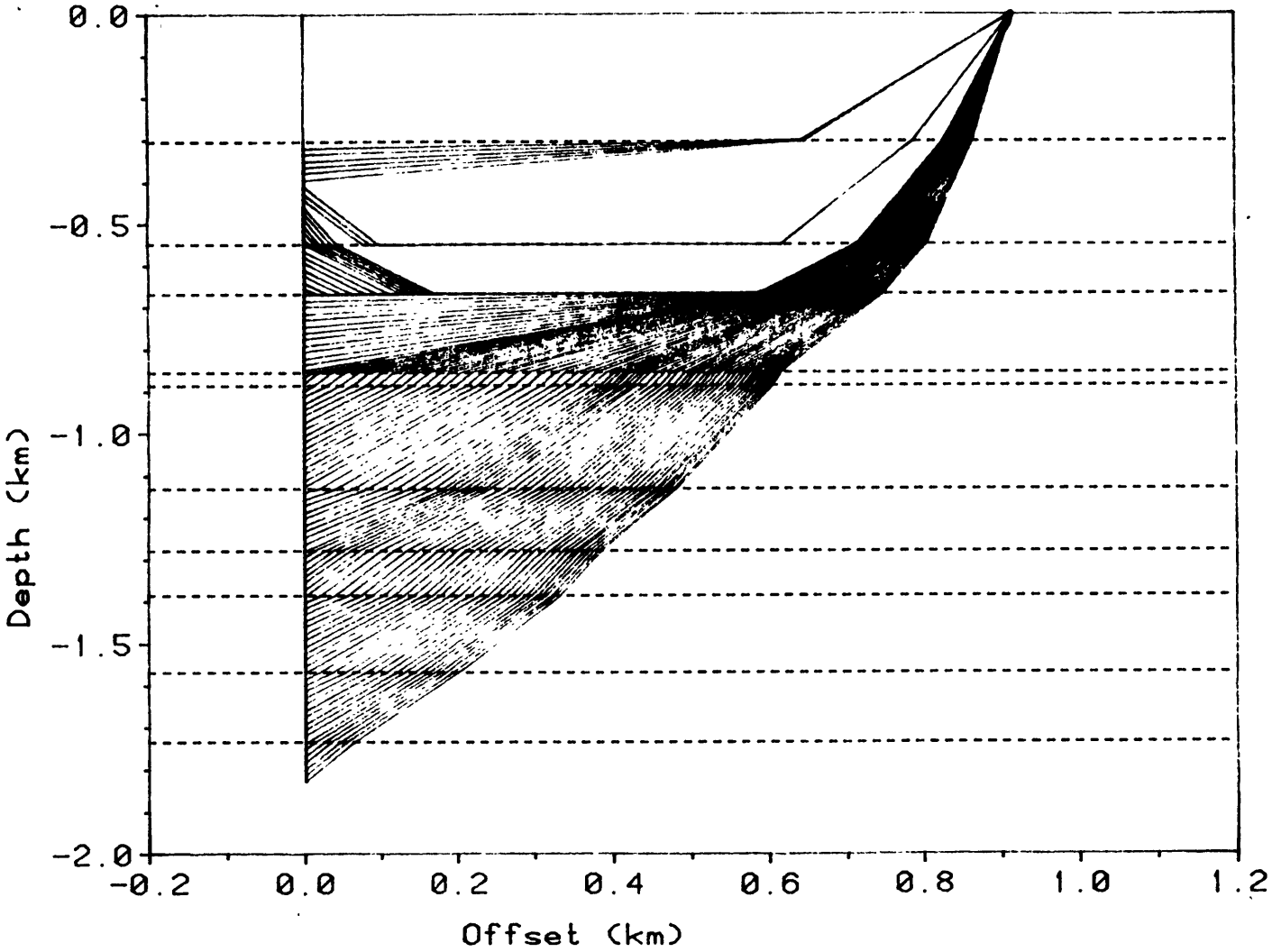


Figure 3.15

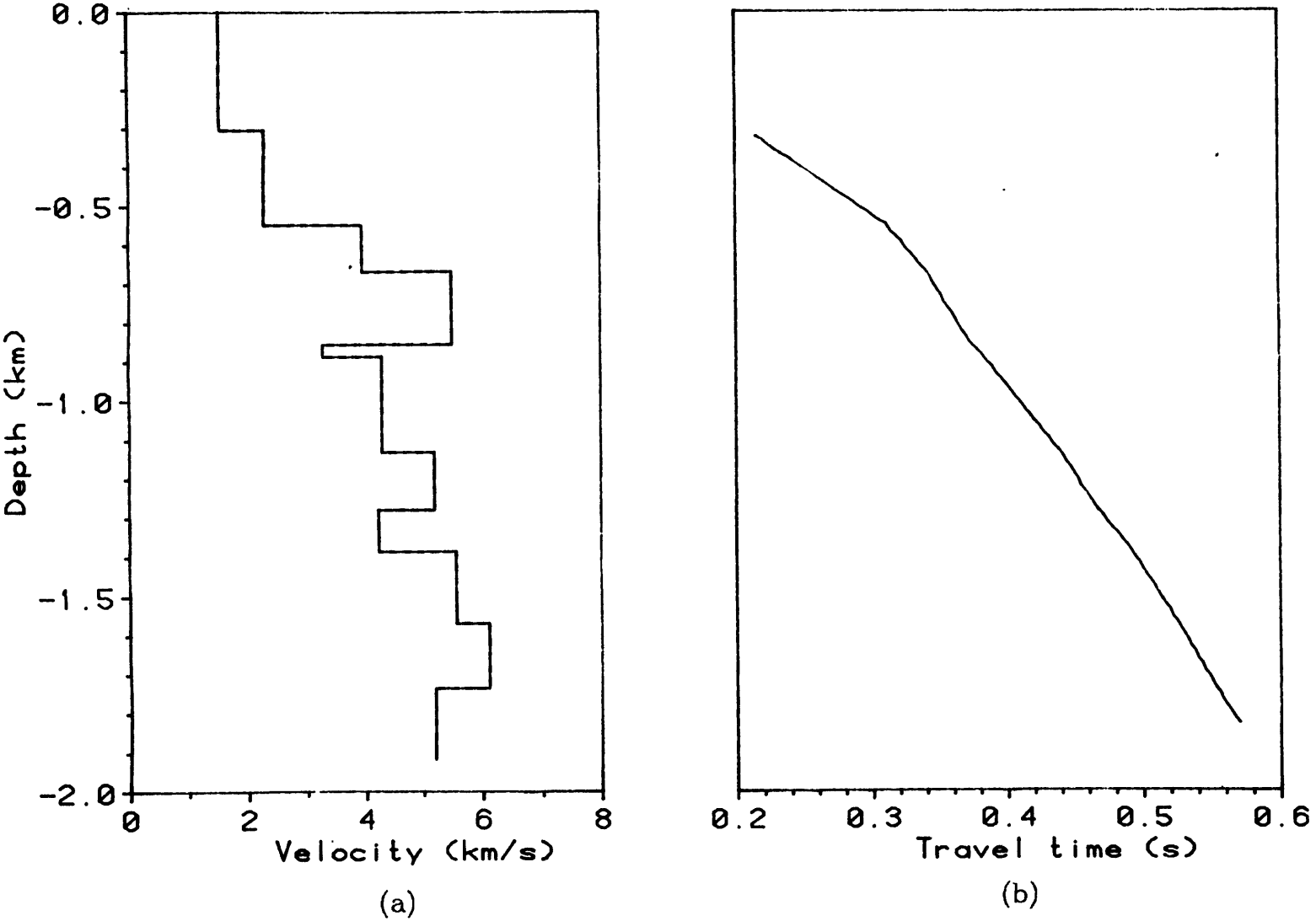

0.0
-0.5
-1.0
-1.5
-2.0
Depth (km)
0
2
4
6
8
Velocity (km/s)
(a)
0.2
0.3
0.4
0.5
0.6
Travel time (s)
(b)

Figure 3.16

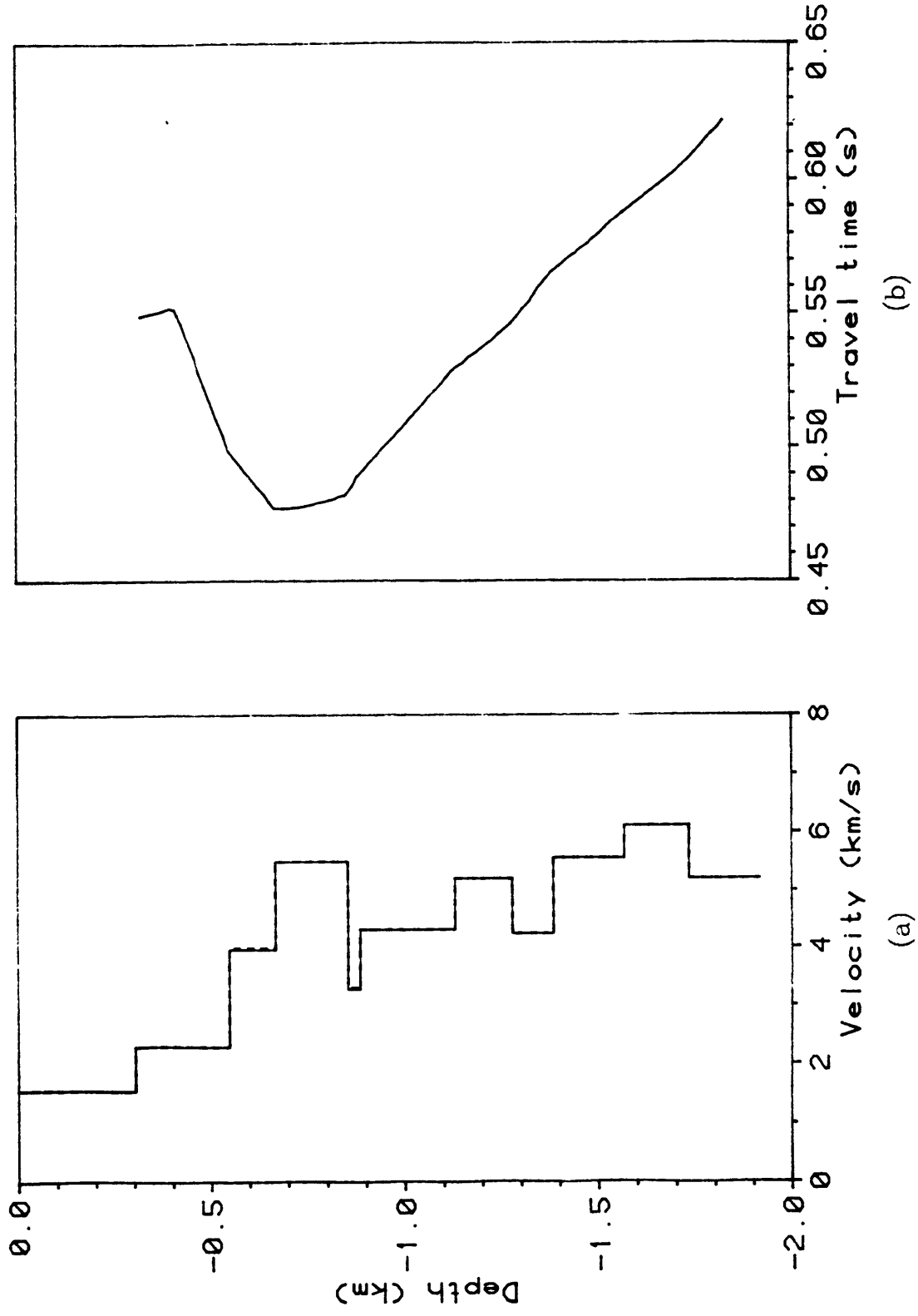
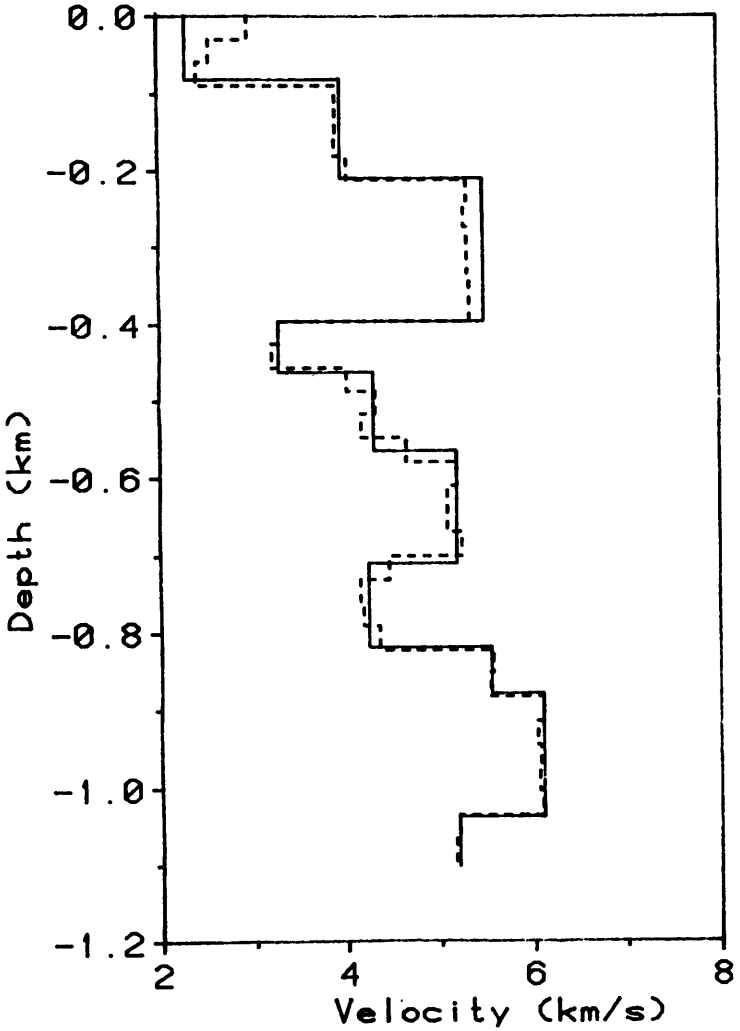
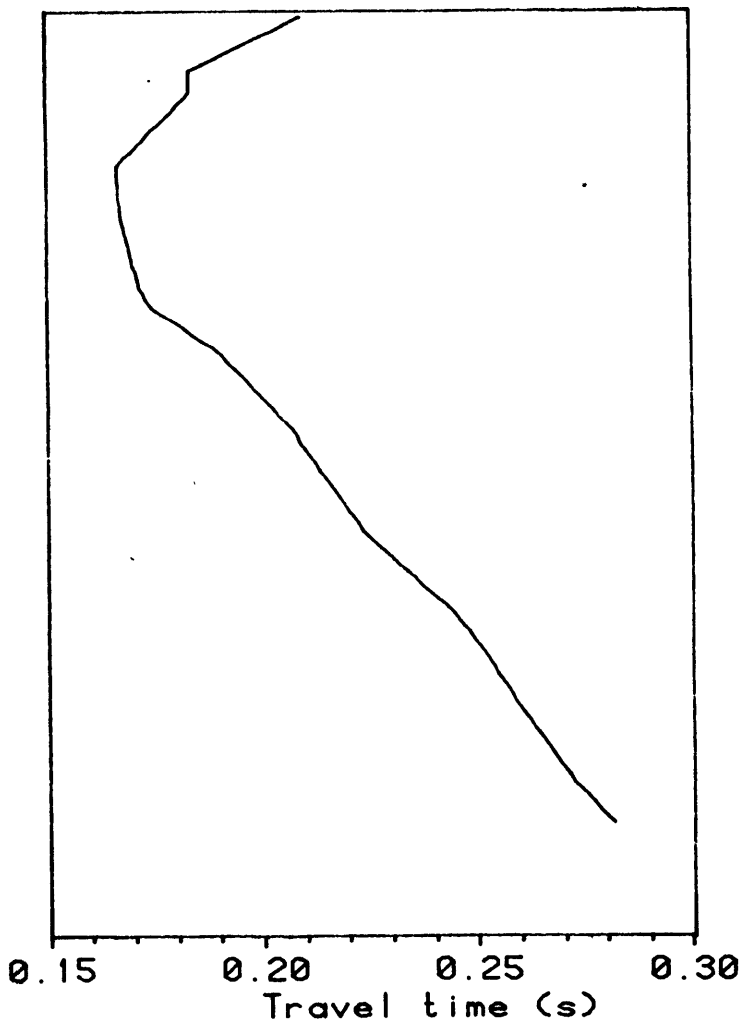

0.45
0.50
0.55
0.60
0.65
Travel time (s)
(b)
0
2
4
6
8
Velocity (km/s)
(a)
0.0
-0.5
-1.0
-1.5
-2.0
Depth (km)

Figure 3.17

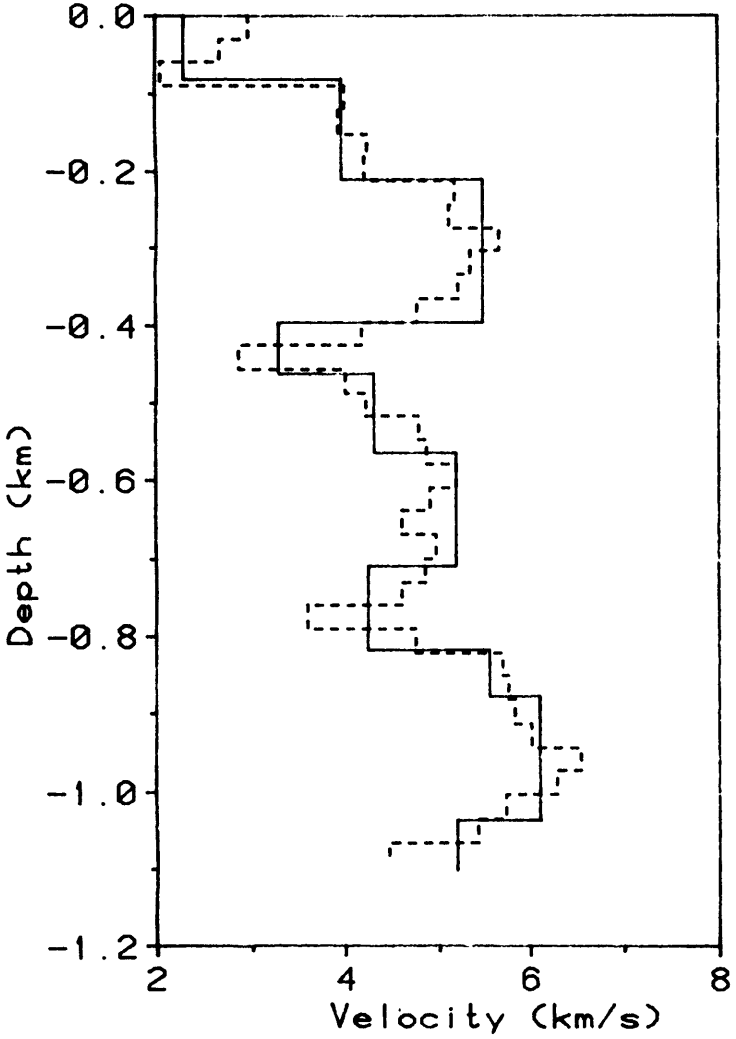


(a)

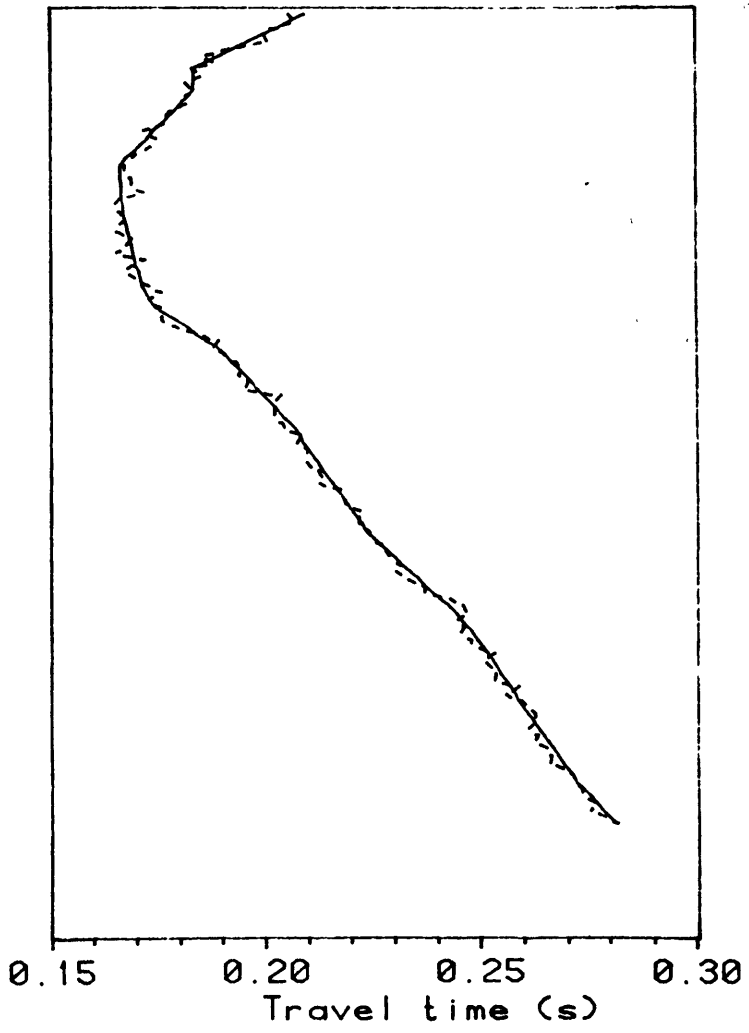


(b)

Figure 3.18

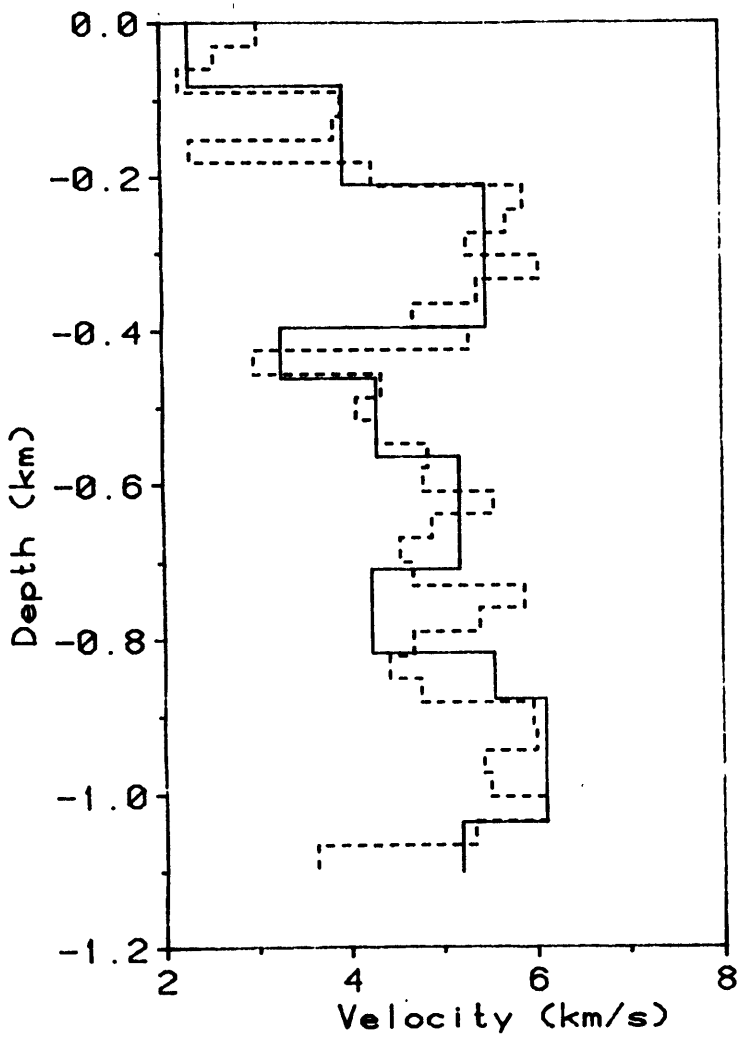


(a)

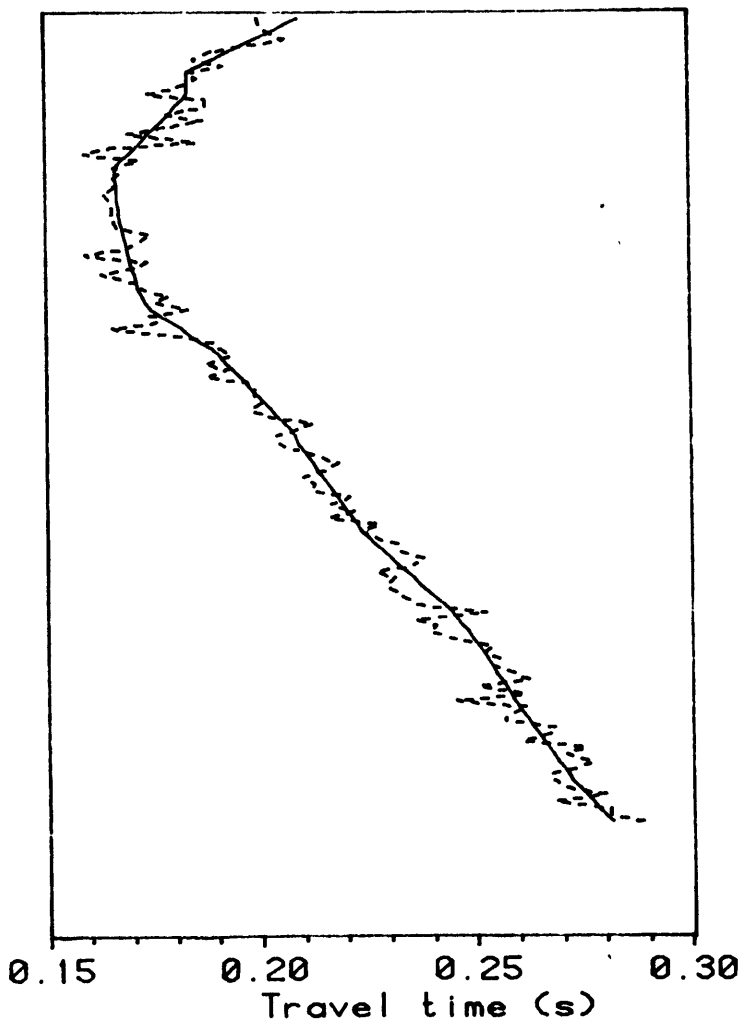


(b)

Figure 3.19



(a)



(b)

Figure 3.20

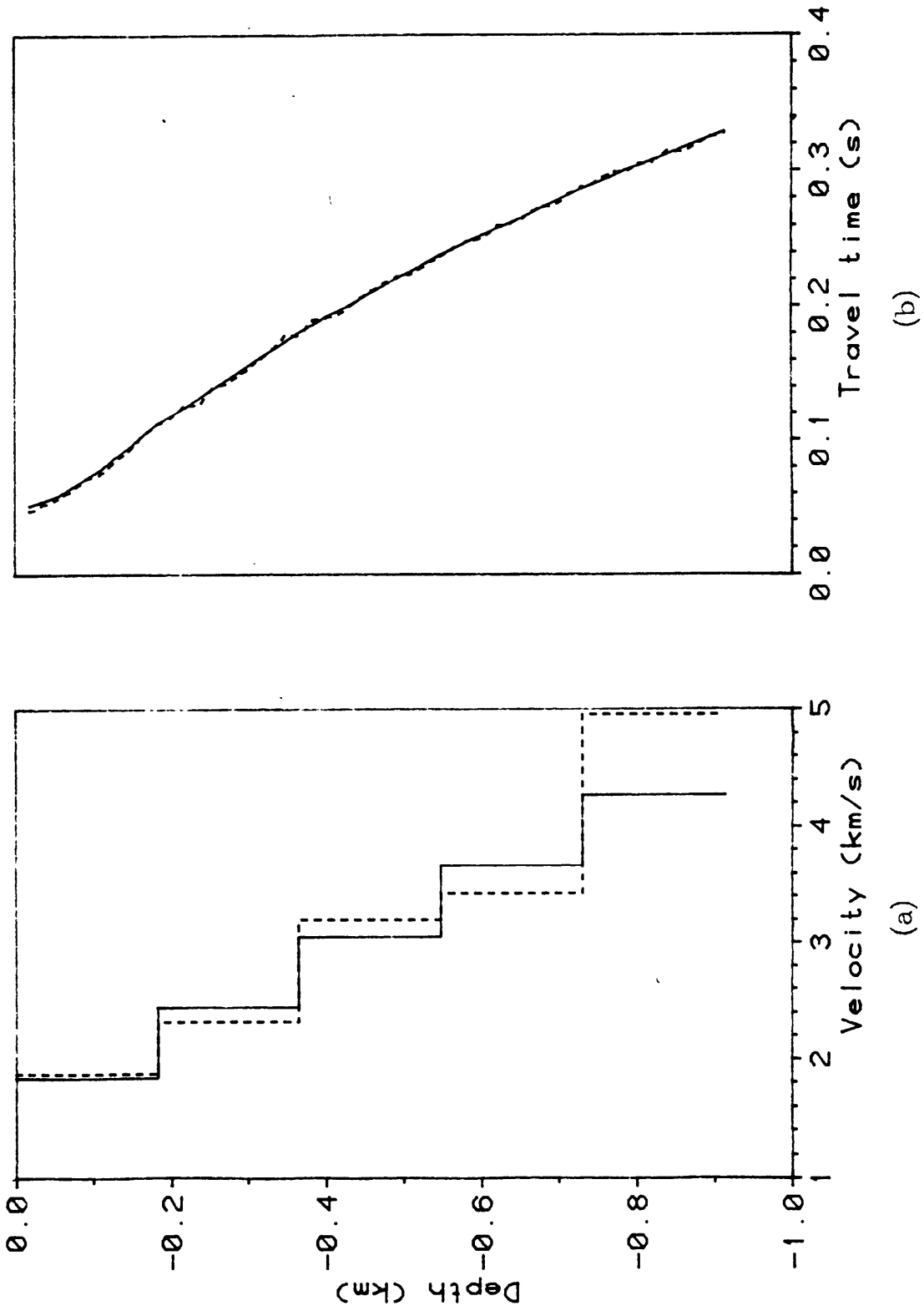
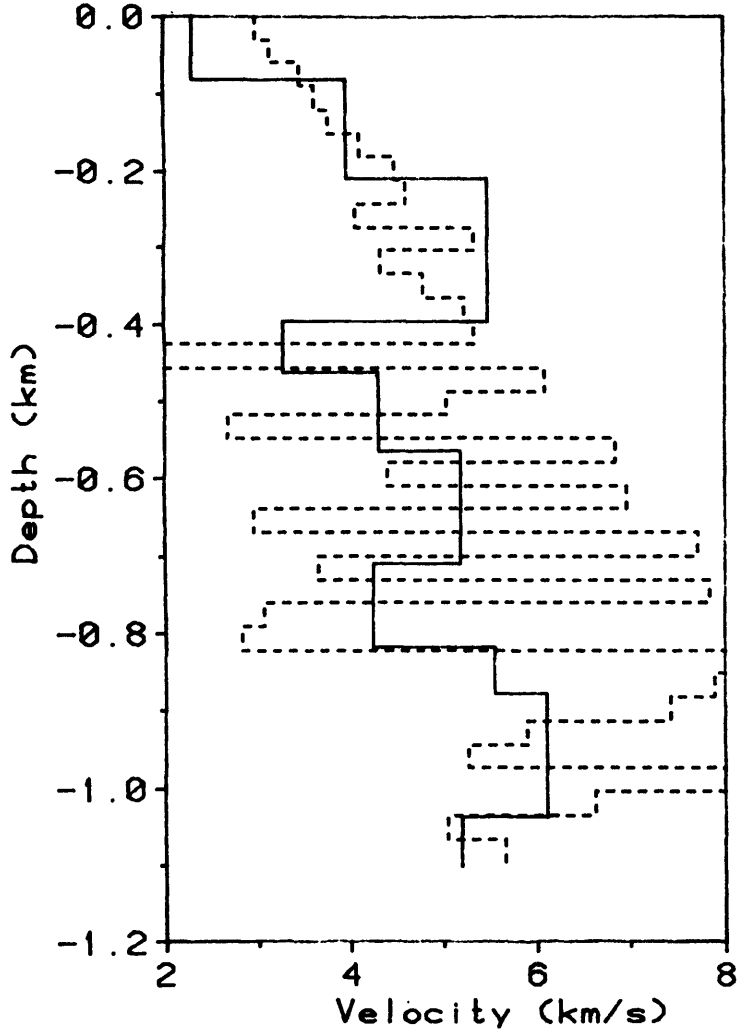
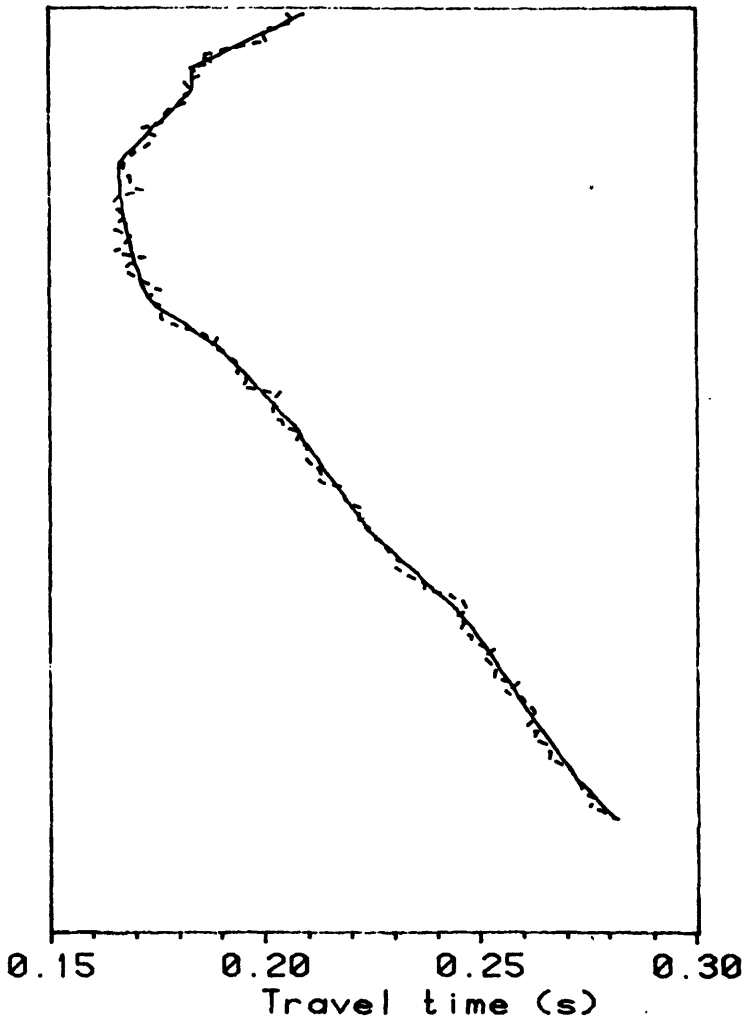

0.0
0.1
0.2
0.3
0.4
Travel time (s)
(b)
1
2
3
4
5
Velocity (km/s)
(a)
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
Depth (km)

Figure 3.21



(a)



(b)

Figure 3.22

CHAPTER 4

Measuring Travel Times

4.1. Introduction

This chapter describes methods that might be useful for accurately and precisely measuring travel times from digitized VSP record sections. Methods for automatically picking times are investigated and the utility of such methods is discussed.

Picking arrival times is of course a very crucial step in travel time inversion because the travel times are the data. The problem of picking arrival times is familiar to earthquake seismologists who must carefully determine when first breaks occurred in order to locate earthquakes. If an impulsive energy source is used in seismic reflection or VSP experiments, the analyst is faced with the same problem: When *exactly* does the first arrival come in? Even small errors in these picks can result in large model errors. Fortunately for some analysts, the author included, most seismic experiments today are conducted with a Vibroseis® source for which the first arrival, after a preprocessing step to be described below, is actually a peak or trough and thus usually very easy to pick.

Another potential problem is avoided by the recording equipment which begins recording at the shot time so that there is never any ambiguity about travel time versus arrival time. To the analyst, they are the same.

Vibroseis is a registered trademark of Conoco, Inc.

4.2. The Vibroseis Source

The Vibroseis source is a large ground vibrator which sweeps across some preselected frequency range. The sweep duration, terminal frequencies, and peak force are all variables depending upon noise conditions, target characteristics, and even cultural considerations. Typical sweeps last 12 to 16 seconds and cover from two to four octaves from, say, 10 to 80 Hz (Fig. 4.1a). Sweeps which are nonlinear in instantaneous frequency (Fig. 4.1b) are often used to concentrate more energy in the higher frequencies as these frequencies are attenuated more rapidly by the Earth (Hoover et al., 1984). Modern vibrators are able to produce a peak force of 200 000 N but much less force is normally used to avoid damaging roads and other structures. Multiple vibrators are often used to increase the imparted energy.

The input waveform $F_1(t)$, sometimes called the pilot, is crosscorrelated with the output signal $F_2(t)$ recorded at the geophone. The result (Fig. 4.1c) is a zero-phase waveform which is amplified and compressed in time (Waters, 1978). If the recorded waveform is defined by

$$F_2(t) = \sum_i a_i F_1(t - \tau_i) \tag{4.1}$$

where a_i and τ_i are the amplitudes and delays of the various arrivals, then crosscorrelation with the pilot gives

$$\varphi_{12}(t) = F_2(t) * F_1(-t) = \sum_i a_i F_1(t - \tau_i) * F_1(-t) = \sum_i a_i \varphi_{11}(\tau_i). \tag{4.2}$$

The output signal $\varphi_{12}(t)$ is thus a weighted and delayed sequence of autocorrelations $\varphi_{11}(\tau_i)$ with peaks at all τ_i. The first autocorrelation peak occurs at delay τ_1 and corresponds to the first arrival. That the final output is inherently noncausal should not be a concern; the preprocessing described here is delayed by the operator length. A typical autocorrelation function for a linear sweep is of the form

$$\varphi_{11}(\tau) = \frac{\sin \pi k \tau (T - \tau)}{\pi k \tau} \cos 2\pi f_o \tau, \qquad -T \leq \tau \leq T \tag{4.3}$$

where

f_1, f_2 are the terminal frequencies,

$f_o \equiv \frac{(f_1 + f_2)}{2}$ is the carrier frequency,

T is the duration of the input signal, and

$k \equiv \frac{f_2 - f_1}{T}$.

This autocorrelation function is called the Klauder Wavelet (Geyer, 1972).

4.3. Interpolation

As the first autocorrelation peak may usually be easily identified on VSP records, it would seem to be a straightforward matter to pick the appropriate local maximum for every trace on the section. Since this is digital data one might think to pick the local maximum amplitude sample. This procedure fails, however, to take advantage of the additional precision contained in the sampled waveform. The continuous wavefield is adequately sampled since there is usually analog anti-aliasing circuitry in the recording equipment. Because neither the recorded waveform nor the crosscorrelation is aliased, the sampling theorem says that the continuous waveform can be exactly reconstructed (Oppenheim et al., 1983). The continuous time function $x(t)$ is given by

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{\sin 2\pi W(t - \frac{n}{2W})}{2\pi W(t - \frac{n}{2W})} \tag{4.4}$$

where the discrete time sequence values $x[n] \equiv x(\frac{n}{2W})$ and W is the bandwidth of $x(t)$. This equation can be used as a formula for bandlimited interpolation. A number of problems can arise, however, since one seldom has sample values recorded for all

time. Another drawback of sampling theorem interpolation is that a large number of computations is required to convolve a sequence with an infinite length filter. One might argue that $x[n] \equiv 0$ before recording was started and that $x[n] \to 0$ as the coda decays. But the Klauder Wavelet is not time-limited to the recording window. Typically $\varphi_{12}(t)$ is only two seconds long while the full Klauder Wavelet lasts twice the sweep length which may be 12 to 16 seconds so that there is surely some error introduced. Practical application of the sampling theorem therefore turns out to be difficult, so theoretical soundness was temporarily dispensed with in favor of the computational ease of cubic spline interpolation; the validity of the cubic spline interpolation is discussed in Appendix D.

It is clear that, at least from an engineering point of view, it is better to make use of some sort of interpolation scheme than not to interpolate at all as can be seen from the following example. Consider wide offset (2 000 m) VSP data recorded digitally at 500 samples per second at 10 m intervals in a 5 000 m/s homogeneous medium (Fig. 4.2). The travel time difference between the direct waves to 1 000 m and 1 010 m is only 0.9̇ ms which is less than half the sampling interval. If no interpolation were used, no moveout at all would be seen between these traces. The situation may be much worse when dealing with complex structures or headwave refractions.

Cubic spline interpolation fits a third order polynomial between adjacent samples so that the interpolated time series is continuous through the second derivative (deBoor, 1978). The particular algorithm used in this study treats the endpoints in the "not a knot" sense described by deBoor (1978).

The method used to pick all of the travel times used in this study uses cubic spline interpolation and the Golden Section search algorithm to find the peak of the appropriate autocorrelation feature.

The Golden Section search technique is a rapid method for minimizing a unimodal function without using derivative information. Adding a negative sign in the appropriate place turns it into a rapid method for maximizing a unimodal function which is the problem under consideration here. This algorithm is described in detail in Foulds (1981). Given a search interval $[a,b]$, Brent (1973) shows that the number of iterations required to achieve convergence to within δ is about

$$\frac{\log(b-a) - \log\delta}{\log\gamma} \tag{4.5}$$

where $\gamma = \frac{1+\sqrt{5}}{2}$ which is Euclid's golden section. For example, finding the maximum value of the crosscorrelation in a 2 ms interval to within 0.01 ms requires only 11 function evaluations with the Golden Section technique whereas a brute-force approach would require 100.

An alternate approach to the maximization problem involves using the derivative information given by the cubic spline coefficients. One could find the roots of the polynomial derivative and choose the choose the abscissa with the larger function value as the maximum of the polynomial. This method, turns out to be computationally trickier to implement since the spline coefficients from the correct gap must be identified and used in the root-finding. (But then nearly everything is trickier than using an existing IMSL subroutine.)

The computer implementation of this method displays the seismogram, allows the analyst to point at the autocorrelation peak, and then searches for the maximum in the manner described. The program that does all of this is quite robust and very fast because the Golden Section search is quite efficient. The Michigan data set was digitized at 500 samples per second and the program seeks a peak in the interpolating spline with a tolerance of 0.01 ms. This is admittedly ambitious but is far better than no interpolation and certainly no worse than eye-balling linearly interpolated data.

4.4. Automatic Methods

Partial automation of the above process is quite feasible and has been implemented. There are certain situations for which partially automated picking might actually improve the accuracy of the travel time measurements. If the feature representing the first arrival is not clearly identifiable on every trace or if the shape of the feature is slowly changing, then a trace-to-trace crosscorrelation technique may be able to follow the feature more consistently than an analyst. Vibrators at wide offsets or impulsive energy sources are examples of potentially troublesome waveforms.

Completely automatic time picking is always a lively issue in seismology. Several authors (e.g. Allen, 1982) report limited success with P-picking algorithms. For VSP processing, Stewart (1983) reports poor results from a scheme to pick first breaks (for an impulsive source) based on second derivative values. He suggests the use of sampling theorem interpolation to find the point of maximum amplitude for zero-phase waveforms (from Vibroseis-like sources) which sounds hopeful but fails to consider some of the noise problems which may be encountered. A free-running P wave picker is bound to make seriously bad picks on data which show other large features or contain, for example, only a low frequency waveform due to problems clamping the geophone to the borehole wall. Since perhaps 5% of the traces in the data sets to be treated here are bad for one reason or another, automating the process would not be very helpful unless it could be made sufficiently sophisticated to be able to detect and discard faulty picks or poor quality waveforms. Otherwise, the user will have to frequently overrule the machine's pick. It is arguable whether the time spent monitoring the program's performance and modifying its picks would be significantly smaller than the time spent hand-picking the travel times. There is also something to be said for staying close enough to the dataset to understand its particular quirks and shortcom-

ings.

A trace-to-trace crosscorrelator was implemented for this study which ably follows a given feature until it is lost in noise. At that point the analyst must manually reidentify the feature before letting the program continue. This program also makes use of the Golden Section search algorithm and cubic spline interpolation to pick the successive peaks of the trace-to-trace crosscorrelation. To minimize problems due to round-off errors, it is better to try to follow the features of a single trace down the section as far as possible rather than to use a leap-frog technique.

4.5. Errors in Travel Time Measurements

The errors in the travel time measurements are due to a variety of sources. Stewart (1983) lists time-picking errors, inexact zero times, instrument imprecision, and depth errors.

The techniques listed in this chapter seek to minimize time-picking errors. For zero-phase data (e.g. Vibroseis source), the error in the actual time pick from the record should be quite small assuming that the right feature is picked, perhaps only 0.2 ms. Aki and Richards (1980) give an expression for travel time uncertainty in the presence of noise:

$$\Delta t = \frac{1}{W \log_2 \left[1 + \frac{S^2}{N^2} \right]} \tag{4.6}$$

where S^2 and N^2 are the power in the signal and the noise, respectively, and W is the signal bandwidth. Given VSP data digitized at 500 Hz, the effective value of W is somewhat less than 250 Hz, probably closer to the high end of the vibrator sweep frequency which in the case of the Michigan VSP data (see Chapter 5) was 114 Hz. The signal to noise ratio for the Michigan VSP data averages about 5:1 so that Δt is about 2 ms. Experience suggests that this is a reasonable value although, as Stewart (1983) points

out, it may be a little high since the VSP analyst is aided by visual correlation with neighboring traces.

Errors from bad time bases and other instrumental errors are added to the data during the experiment and are difficult, if not impossible, to correct. One could imagine inverting for the time origin as is done in joint earthquake hypocenter and velocity inversion (e.g. Crosson, 1976) but this would undoubtedly add more error than it would alleviate. Sources cited by Stewart (1983) give estimated total timing errors between 0.5 ms and 1.0 ms. Pujol (pers. comm.) reports time picking errors alone of less than 1.0 ms.

Another source of error in the travel time data is the receiver depth. The wireline is usually slackened once the geophone is locked at depth to diminish cable wave noise (Stewart, 1983) so that tool slippage may occur undetected. Even if the tool is clamped tightly, the accuracy of placement depends on a number of factors such as the depth reference (usually the floor of the drilling platform), cable stretching, and winch operator enthusiasm.* Stewart (1983) estimates the total depth error may be as high as several feet and introduces an additional 0.5 ms (or more) in timing error.

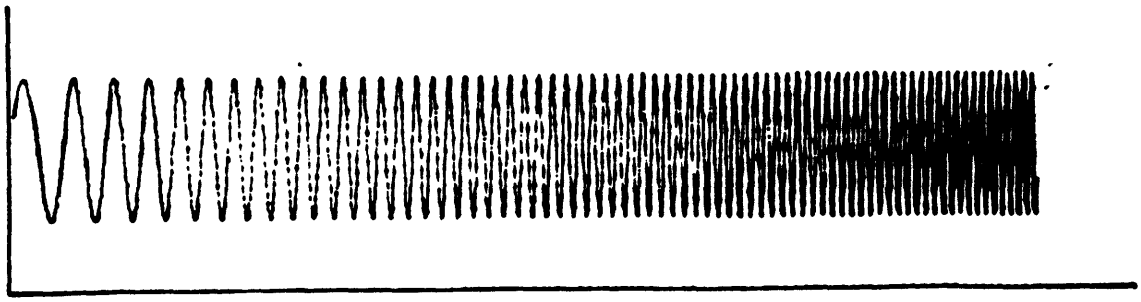
The total travel time error is therefore expected to be about 1 ms.

* I witnessed a VSP survey in the Gulf of Mexico last winter. "I thought the company man said 50 foot intervals," I said when the depth meter showed a value 48.1 feet above the previous measurement. The operator, who had been up for almost two days, said, "Yeah. We try to get it pretty close." So much for quantifying operator error.

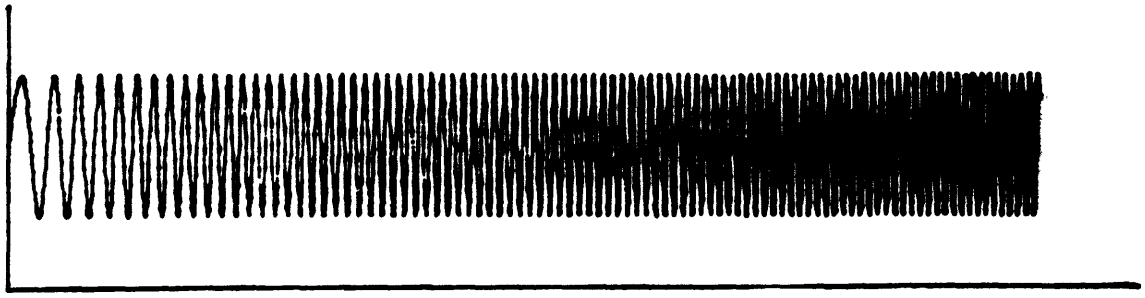
Figure Captions

Figure 4.1. (a) Vibrator pilot waveform which linear in frequency with time. (b) Logarithmic vibrator pilot waveform. (c) Klauder wavelet.

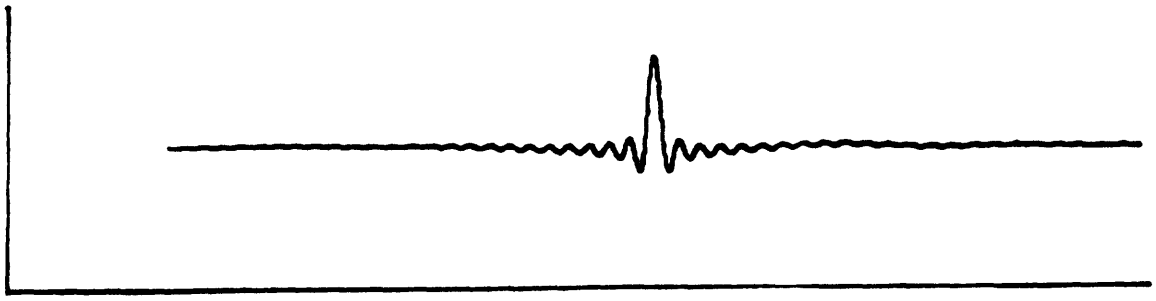
Figure 4.2. Typical VSP geometry showing raypaths to two receivers showing the travel times to both. The difference, 0.9 ms, is less than the resolution of uninterpolated data.



(a) **Linear sweep**



(b) **Logarithmic sweep**



(c) **Klauder wavelet**

Figure 4.1

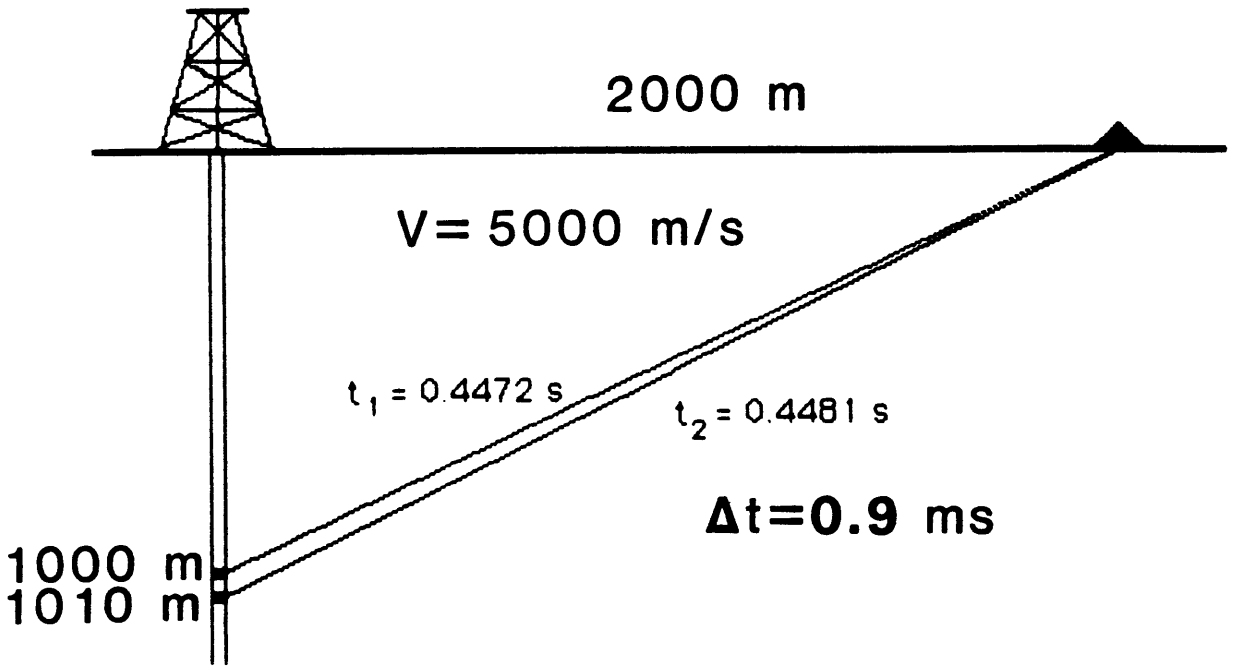


Figure 4.2

CHAPTER 5

Examples With Field Data

5.1. Introduction

The utility of an analysis technique is best tested with real data. Three sets of field data were inverted using the travel time inversion techniques presented here. Two of the data sets are rather simple examples with nearly zero offset sources and a fairly simple structure. The third set, however, came from a multi-offset experiment from which well over one thousand travel times were measured.

5.2. Mounds, Oklahoma VSP

The first data set analyzed came from a VSP experiment run prior to the Gas Research Institute's Hydrofrac Monitor Project near Mounds, Oklahoma to obtain velocity estimates for modeling the area before and after hydrofracture. The P wave data studied here were collected from a near offset VSP with a Vibrator source 46 m from the top of the well. The well, which is only 350 m deep, is straight but slightly deviated from vertical so that the projection of T.D. on the surface is 18 m closer to the source than the well-head. This 3° deviation was taken into account in the inversion.

The model chosen for the inversion consists of twelve 30.5 m thick layers. Thinner layers were not used since there were only 43 travel times available. The resulting model is shown in Figure 5.1 along with a travel time versus depth curve. The only other model of the area available was a preliminary model used by the Amoco Production Co. (R. Heiser, pers. comm.). It is shown in Figure 5.1 as a dotted line. To say that the models agree is overstating the case since the Amoco model is very

simple. They do not disagree, however.

The absolute values of the velocity errors, obtained from the covariance matrix, are small, averaging 43 m/s. As discussed in Chapter 3, these values are probably underestimates. The RMS travel time error is 0.2 ms. These small errors, underestimates that they may be, suggest a very good fit.

5.3. Gulf Coast VSP

The velocity structure near a particular well on Alabama's gulf coast has been investigated previously by Wingo (1981), Stewart (1983), and Keho et al. (1984). It provides a good opportunity to compare results. Wingo's (1981) results were obtained by repeated forward modeling while the latter two papers used travel time inversion. In addition, Keho et al. (1984) studied spreading, transmission coefficients, attenuation, and radiation patterns. The experiment was done by the Amoco Production Co. near Gulf Shores, Alabama, in May 1980. A slanted weight drop device, offset 80 m from the well-head, was used as both a P and S wave source. SH waves with opposite polarities were generated by reversing the direction of the weight drop from right to left perpendicular to a line from the source to the borehole. Wingo (1981) picked travel times from the vertical P wave component and picked SH wave times by overlaying the traces recorded from the left- and right-slanted source configurations and identifying the onset of reversed polarity energy.

Receivers were spaced 3.05 m apart from 3.05 m to a depth of 488 m for both the P and S wave experiments. Because of the dense geophone spacing in the hole, the model to be obtained from travel time inversion could have quite thin layers. Stewart (1983) used 12.2 m thick layers in the hopes of detecting thin layers of gas sands. Forty layers, each 12.2 m thick were also in this study. The P and S wave travel times were inverted separately for velocity. Modified Brown and Dennis damping (see Appendix B) was used to maintain stability and the initial models in both cases were

obtained using the layer stripping technique developed in Chapter 3.

The inversion results are shown in Figures 5.2 and 5.3. The travel time vs. depth curves are also shown. The results are close to the velocities presented by Stewart (1983) and Keho et al. (1984) and agree reasonably well with the lithology presented in the latter paper. The average velocity error in the P wave model is about 107 m/s. The P wave RMS travel time error is 0.5 ms; the S wave RMS error is 1 ms.

The shallow well (490 m) penetrated only unconsolidated sands, silts, and shales. Two thin layers of gas sands were found at 405 m and 457 m. The low P wave velocity at those depths hints at their presence.

5.4. Michigan Basin VSP

The third data set considered, from the Michigan Basin, was much larger than the first two in that the borehole was about 1900 m deep and data were used from sources at eight different offsets. The structure of the study area in this experiment is also more complicated.

These VSP experiments were undertaken as part of the M.I.T./C.G.G. Experimental Group Shoot in Manistee County, Michigan in September and October, 1983. A primary goal of the study was to evaluate the effectiveness of the VSP technique as a tool for reservoir delineation (Compagnie Général de Géophysique, 1985). The borehole used for this study, State Burch 1-20, was a dry well less than 300 m away from a producing well (State Springdale 1-20). The source rock for the producing well is a Silurian pinnacle reef 1370 m deep.

Before presenting the findings of the current study, it is worthwhile to think about the task at hand. Can a small three-dimensional structure like the reef be successfully delineated solely on the basis of a one-dimensional model? No, of course not, but there is some hope. If a reasonable one-dimensional model can be found which, on

the average, represents the velocity structure for the subsurface around the well, then higher order effects such as the variation of travel time residuals for the different source offsets may perhaps be used to identify velocity perturbations. Tomographic techniques might then possibly be used to understand these anomalies. Because there is no good reason to conduct multi-offset VSPs unless a two- or three-dimensional structure is expected, a purely one-dimensional interpretation is inadequate. A qualitative interpretation will therefore be presented of the travel time residuals from around the Burch 1-20 well.

The eight VSPs analyzed in this study are a subset of the 13 VSP experiments run during the group shoot. Besides the VSPs, a 3D seismic reflection survey and a transposed VSP (with the source in the borehole and receivers along the surface) were also conducted to compare the utility of these three techniques in reservoir delineation (C. G. G., 1985). A map of the study area is shown in Figure 5.4. Table 5.1 lists the source positions and depth ranges covered from each offset. The receiver spacing used for most of the survey was 9.1 m.

The data were treated in three groups. The first group was made up of two zero-offset experiments run at different times at offset A. The first zero-offset VSP run on Burch 1-20 was performed by Schlumberger Well Services for the Marathon Oil Company. The second was undertaken during the Group Shoot by the Compagnie Général de Géophysique. A record section from offset A is shown in Figure 5.5. In addition, the shallower C. G. G. data were collected later than the deeper traces with some overlap in the middle. Efforts were made during processing to tie the two sets of C. G. G. data. That the travel times measured at some of the depths common to the two VSPs were different indicates that at least one of the experiments had timing problems or perhaps depth control problems. Figure 5.6 shows the difference between times obtained from the Schlumberger and C. G. G. data at common depths. Below the sec-

tion of the hole which was cased, C. G. G. times are usually earlier while above the casing Schlumberger-derived travel times are systematically earlier. Both data sets were, however, used in this study. The second group consisted of four VSPs with approximately 600 m source offsets. Offsets B, D, E, and G were all shot on the same trip up the well as the lower section of A offset data. Multiple vibrators were used so that data from all five offsets could be taken while geophone remained clamped at each station. The final group of three VSPs (F, J, and L) were collected together in a similar fashion. All three had source offsets of about 1 800 m.

Travel times were picked from all of the traces used in this study with the methods described in Chapter 4. The eight travel time vs. depth curves are shown in Figures 5.7 through 5.14. The source elevations, although only slightly different from the well-head elevation in most cases (Offset L is an exception), were considered in all of the subsequent work.

The A offset data set was inverted for two kinds of models which differ in their layering strategy. Since various well logs were available, C. G. G. (1985) produced a velocity model with layer boundaries taken from the logs. These layer boundaries are a form of *a priori* information. Minor modifications to this thickness model were made based on information from a refaction survey by the Marathon Oil Co. and P velocities picked from a full waveform acoustic log. A strong reflection seen in the middle of the Detroit River Salt suggested a subdivision. Mud logs from that level indicate that the Detroit River layer is actually as many as 20 sublayers of alternating anhydrite (5.5–6 km/s) and salt (5.5 km/s). The result of the inversion for 36 layer velocities (with the top layer velocity held fixed) is shown in Figure 5.15; the velocities themselves may be of interest in this case and are given in Table 5.2. P velocities obtained from the full waveform acoustic log are shown in Figure 5.16. There is a very good fit. The RMS travel time error is 1.5 ms. When *a priori* layering information is not used, the results

are similar but do not appear to fit the log-derived P wave velocities as well. Figures 5.17 and 5.18 show the inversion for models with constant thickness layers 30.5 and 61 m thick, respectively. Although the RMS travel time errors, 0.8 ms for the 30.5 m model and 1.0 ms for the 61 m layer model, are quite low, a tradeoff between model variance and model resolution (e.g. Menke, 1984) is very apparent here. The width of the velocity curve in Figures 5.17 and 5.18 represent the 95% confidence interval. The 61 m thick layers are clearly better resolved than the 62 layers of the 30.5 m model.

Before the travel times picked from the offset VSPs were added to the inversion data set, a static correction was made to the remove the effect of the glacial till's varying properties throughout the study area. The source of the static shift seen in the offset VSPs is thought to be either shallow velocity anisotropy, lateral velocity variations in the glacial till, variations in the thickness of the till, or some combination of these factors (C. G. G., 1985). 180 m monitor holes drilled near most of the vibrator stations indicate that the glacial till does not vary wildly in thickness. Support for the anisotropy theory comes from the apparent azimuthal independence of the negative (i.e. observed before calculated) static shifts, the increase in shift with distance, and the fact that the sign of the time shift is correct. The Bell Shale, at about 625 m, is perhaps the responsible anisotropic material. The static shift at offset G, however, is 10.6 ms which may be too large to be explained by a single anisotropic layer. On the basis of evidence from the transposed VSPs, C. G. G. (1985) believes that variations in the velocity of the glacial till is a better explanation.

The A offset model was used to calculate travel time vs. depth curves for the seven other offsets and the average travel time residual for each offset was subtracted from each of the observations. Data from the 600 m offsets were then added to the set and inverted as before with the log-derived layering. The resulting model is shown in Figure 5.19. Finally, the three data sets from 1 800 m offset sources were added and the

travel time inversion was repeated with close to a thousand data points. The computed model is shown in Figure 5.20.

5.5. Analysis of Travel Time Residuals

As suggested earlier, travel time residuals from multiple source offsets might enable qualitative statements to be made about three-dimensional structure based on only one-dimensional models. This is similar to a technique used in earthquake seismology. Most computer programs for earthquake location rely on a one-dimensional crustal model. Perturbations in the velocity structure local to seismic stations are modeled by constant shifts applied to observed arrival times. These time shifts are known as "station corrections." The residual static corrections already made are also first order models of lateral structural or velocity variations.

The model obtained from inverting all of the data together represents an average velocity structure in the vicinity of the borehole. Travel time residuals were calculated for each offset at all of the receiver depths based on this average model. Plots of travel time residual vs. depth are shown in Figures 5.21 through 5.28.

The travel time residuals can be divided into two broad categories: those that are structurally significant and those that are not. The latter category includes residuals originating from bad travel time picks, timing errors, and depth errors. Figure 5.29 shows a well profile derived from a caliper log of Burch 1-20. Several washed out zones are visible in the figure, notably in the F Salt, the B Salt, and the A2 Evaporite. Even though a special probe with longer arms was used in these sections, it is prudent to regard travel time anomalies at those levels with suspicion. The fact that data were collected at the same time from offsets A, B, D, E, and G makes it very likely that the similar residual signatures seen in the washed out zones on these five plots (Figs. 5.21–5.24 and 5.26) are from the same source, possibly depth errors or bad time picks due to poor geophone coupling with the geophone wall. The largest anomalies occur

at the same depths on all five curves. There are similar wild variations in the wash-outs in the plots for offsets F, J, and L (Figs. 5.25, 5.27, and 5.28). The residuals for the 1 800 m offset experiments are quite noisy to begin with due principally, it would appear, to poor time picks which are in turn due to a lower signal to noise ratio on the seismograms. It is also plausible that the one-dimensional assumption begins to break down in this region at such large offsets.

There is still some information left in the travel time residual plots despite all of the above caveats. The curves for offsets B (Fig. 5.22) and D (Fig. 5.23) are almost identical as are those for offsets F (Fig. 5.25) and J (Fig. 5.27) if the noise is ignored. The shapes of the curves for offsets E (Fig. 5.24) and G (Fig. 5.26) are also similar. Do the differences between curves from the reef side (B, D, F, and J) and the off-reef side (E, G, and L) of the well point to the presence of the reef? Figure 5.30 and 5.31 show theoretical raypaths for offsets D and J assuming a one-dimensional geometry and the velocity structure derived from all eight offsets. The geometry of this experiment was fortuitous in that waves transmitted through the reservoir could be recorded since Burch 1-20 was so much deeper than the Springdale reef. The approximate position of the pinnacle reef is sketched on these figures although of course the rays shown are not affected by it in this one-dimensional model. Nevertheless, it appears that the travel times from offset D (and B) to receivers at depths from about 1 450 m to T.D. ought to be affected by the reef. Likewise, times from about 1 450 m to T.D. should be affected from offsets J and F. The sign of the travel time anomaly, at least in this first order approximation, is correct given the assumed velocity structure of the reef. The reef, which is composed of the Brown Niagran carbonate, displaces the B Salt which is a much slower material. In addition, the A2 Evaporite (see Table 5.2) is more compacted above the reef than off to the side. The velocity of the A2 Evaporite near Burch 1-20 is about 5.2 km/s whereas it is 6.5 km/s above the reef at Springdale 1-20 (C. G. G., 1985). In fact, this large velocity contrast and accompanying density contrast

makes possible one of the usual techniques for detecting reefs in this region from reflections. Caughlin et al. (1976) report that a weakening of the A2 Carbonate reflection event is characteristic of a reef. This would clearly be the case if the velocity-density contrasts at the A2 Carbonate/A2 Evaporite and A2 Evaporite/A1 Carbonate were small.

The small backwards "S" shape visible on the residual curves from offsets B and D below 1 600 m is probably not related to the reef since a similar feature can be seen on the E and G offset curves.

There is a large anomaly at about 1 520 m on the residual curves from offsets F, J and, in the opposite sense, L. The ray diagram for J offset (Fig. 5.31) shows incidence angles near horizontal for these arrivals. It is likely that the observed arrivals correspond to rays which also traveled through the faster Brown Niagran of the reef and arrived earlier than expected. A possible explanation for the residual of opposite sign seen on the curve for offset L (Fig. 5.28) is that the average velocity structure used to calcualate the residuals is weighted by the excess of data from sources on the reef side and that the anomaly from L is merely compensation in the other direction.

The justification for the above qualitative statements lies in Fermat's Principle. It states that the travel time along a given ray is stationary with respect to the path. Therefore perturbations of travel time due to perturbations in velocity ***along the original raypath*** are the only first order effects; perturbations to the raypath itself are higher order effects on the travel time.

Because of the encouraging qualitative results, an attempt was made to make some rough calculations of the dimensions of the reef. It appears, however, that the assumption of one-dimensionality has been pushed as far as it can be pushed. For example, Figure 5.30 shows that from offset D the first arrivals at depths below about 1 520 m traveled approximately vertically through the reef. The thickness of the

anomalous velocity structure can therefore be computed from

$$h = \Delta t \, \Delta V \tag{5.1}$$

where Δt is the travel time residual and ΔV is the velocity contrast between the reef and the material being sampled by rays not passing through the reef. Such an "unperturbed" velocity model was calculated by inverting travel times from the off-reef side of the well but including offset A since it was the only source position for which travel times could be obtained all the way to the surface. Interpreted dipmeter logs suggest that the layering seen from offset A is not flat but rather that the well nicks the west flank of the carbonate reef (C. G. G., 1985). This is confirmed by full waveform inversion (Beydoun, 1985). Travel time residuals from offsets B and D are only 2 ms which given unreasonably low values for the thickness h given any but a very small velocity contrast. It is important also to remember that the sampling interval of the waveform data is only 2 ms so that there is a fairly large margin of error. Proceeding further, rays from the 1 800 m offset sources on the reef side would travel approximately horizontally through the structure (see Fig. 5.31). The travel time residuals from offsets F and J are about 7 ms. A simple formula for the width of the reef is

$$w = \Delta t \, \Delta V \tag{5.2}$$

which leads to an aspect ratio which agrees with the assumed dimensions of the pinnacle reef structure which is thought to be perhaps 90 m thick and 300 m wide. The velocity contrast in this case would have to be only 45 m/s. The primary difference in velocity structure that would be seen by rays on opposite sides of the reef is the thinning of the B Salt formation (~5 km/s) over the reef and its replacement by the Brown Niagran carbonate (~6 km/s according to a sonic log from the State Springdale 1-20 well) that makes up the reef itself. The A1 and A2 layers also thin over the reef. It is indeed stretching this argument to say that the combined effect of the velocity

changes about the reef do or do not account for the necessary contrast since the structure is so complicated.

These interpretations are admittedly speculative and it is questionable whether the presence of the reef would even be inferred via this method if it were not expected. C. G. G. (1985) bases delineation of the Springdale reservoir on the characteristics of various reflection events seen on the VSPs and on differences in mode conversion on opposite sides of the borehole due to the dip of the reef's flanks. It appears, however, that travel time anomalies may also be useful indicators for reservoir delineation.

Data set	Source Offset (m)	Azimuth from well	Depth Range† (m)
A‡	109	ESE	1902 to 9
B	622	SE	1902 to 887
D	632	ESE	1902 to 887
E	732	WNW	1902 to 887
F	1874	SE	1902 to 732
G	601	NE	1902 to 887
J	1840	ESE	1902 to 732
L	1944	WNW	1902 to 732

† For various reasons (e.g. poor clamping), travel times could not used from every depth in each range. Also, the full survey included geophone positions outside the depth range indicated which were not used in this study.

‡ Includes travel times from two separate surveys. See text.

Table 5.1. Summary of source and receiver positions from Michigan Basin VSP used in this study.

Off Reef Model Near Burch 1-20			
Depth to base of formation (km)	V_p (km/s)	Thickness (km)	Formation
0.027	0.914†	0.027	glacial till
0.175	1.659	0.148	glacial till
0.305	2.822	0.130	Sunbury/Ellsworth Shale
0.348	2.659	0.043	Antrim Shale
0.370	3.555	0.022	Traverse Formation
0.544	5.496	0.175	Traverse Limestone
0.627	4.155	0.083	Bell Shale/Dundee Limestone
0.758	5.288	0.131	Detroit River Anhydrite/Salt
0.864	6.185	0.106	Detroit River Anhydrite/Salt
0.928	6.504	0.064	Amherstburg Dolomite
0.948	4.905	0.020	Bois Blanc
0.981	5.547	0.033	Bois Blanc
1.050	6.158	0.069	Bass Island
1.056	5.075	0.005	Salina
1.060	5.442	0.005	Salina
1.176	4.314	0.115	F Salt
1.185	5.275	0.010	F Salt
1.203	4.743	0.017	F Salt
1.246	4.532	0.043	F Salt
1.277	3.571	0.031	C Shale
1.280	3.843	0.003	C Shale
1.391	4.777	0.110	B Salt
1.429	6.565	0.038	A2 Carbonate
1.447	5.219	0.018	A2 Evaporite
1.502	7.127	0.055	A1 Carbonate
1.505	6.392	0.003	A1 Evaporite
1.517	5.850	0.012	Brown Niagran
1.526	6.449	0.009	Grey Niagran
1.611	6.440	0.085	Grey Niagran
1.620	6.519	0.009	Clinton
1.684	6.341	0.063	Clinton
1.702	4.366	0.018	Cabot Head Shale
1.729	5.851	0.028	Manitoulin
1.756	4.757	0.026	Cincinnatian Shale
1.808	5.269	0.052	Cincinnatian Shale
1.865	3.860	0.057	Utica Shales
1.901	5.721	0.036	Trenton Limestone

† Velocity held constant during inversion.

Table 5.2. Velocity model obtained from travel time inversion given layer boundaries from logs (layering modified from C. G. G., 1985).

Figure Captions

Figure 5.1. a) Velocity vs. depth model obtained from travel time inversion of Mounds, OK. VSP (solid line). The width of the velocity curve represents the 95% confidence limits. The dashed line is a simple model of the area provided by the Amoco Production Co. b) Travel times picked from the VSP record section.

Figure 5.2. a) P wave velocity model from Gulf Coast VSP. Layer thickness is 12.2 m. The width of the velocity curve represents the 95% confidence limits. b) P wave travel times picked by Wingo (1981).

Figure 5.3. a) S wave velocity model from Gulf Coast VSP. Layer thickness is 12.2 m. The width of the velocity curve represents the 95% confidence limits. b) S wave travel times picked by Wingo (1981).

Figure 5.4. Map of the M.I.T./C.G.G. Experimental Group Shoot study area in Manistee Co., Michigan.

Figure 5.5. VSP record section from the zero-offset source (109 m). The geophone spacing is 9.1 m except from 530 m to 878 m, where the spacing is 18.3 m.

Figure 5.6. Plot of travel time difference (C.G.G. minus Schlumberger) at common depths for two VSP surveys conducted from same source location (offset A).

Figure 5.7. Observed travel time vs. depth curve from Michigan Basin VSP for offset A (109 m ESE of borehole). The accompanying line (plotted on top of the observation in this case) is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.8. Observed travel time vs. depth curve from Michigan Basin VSP for offset B (622 m SE). The accompanying line is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.9. Observed travel time vs. depth curve from Michigan Basin VSP for offset D (632 m ESE). The accompanying line is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.10. Observed travel time vs. depth curve from Michigan Basin VSP for offset E (732 m WNW). The accompanying line is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.11. Observed travel time vs. depth curve from Michigan Basin VSP for offset F (1874 m SE). The accompanying line is the line calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.12. Observed travel time vs. depth curve from Michigan Basin VSP for offset G (601 m NE). The accompanying line is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.13. Observed travel time vs. depth curve from Michigan Basin VSP for offset J (1 840 m ESE). The accompanying line is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.14. Observed travel time vs. depth curve from Michigan Basin VSP for offset L (1 944 m WNW). The accompanying line is the curve calculated from the offset A model for this offset and used in the calculation of static corrections.

Figure 5.15. Velocity vs. depth curve from travel time inversion of A offset data. The layering in this model follows boundaries determined from well logs. The width of the velocity curve is the 95% confidence limit.

Figure 5.16. P velocities from a full waveform acoustic log. These data are only available below 600 m.

Figure 5.17. Same as 5.15 with constant thickness layering. Layer thicknesses are 30.5 m.

Figure 5.18. Same as 5.15 with constant thickness layering. Layer thicknesses are 61 m.

Figure 5.19. Velocity vs. depth curve from travel time inversion of A offset data and data from all 600 m offsets (B, D, E, and G). Width of velocity curve repesents 95% confidence interval. Layering determined from well logs.

Figure 5.20. Travel time inversion result with all available data (966 travel times). The 95% confidence interval is shown by the width of the velocity curve. Layering determined from well logs.

Figure 5.21. Travel time residual vs. depth for offset A. Residual is given by $t_{obs} - t_{calc}$ where t_{calc} is computed from average model obtained by inverting travel time data from eight offsets together.

Figure 5.22. Travel time residual vs. depth for offset B.

Figure 5.23. Travel time residual vs. depth for offset D.

Figure 5.24. Travel time residual vs. depth for offset E.

Figure 5.25. Travel time residual vs. depth for offset F.

Figure 5.26. Travel time residual vs. depth for offset G.

Figure 5.27. Travel time residual vs. depth for offset J.

Figure 5.28. Travel time residual vs. depth for offset L.

Figure 5.29. Well profile of Burch 1-20 well. A well profile is obtained by reflecting a caliper log about the borehole axis and reducing the amplitude of each log by a factor of two to adjust for the change from diameter to radius. Wash-outs in salt and other evaporite layers may cause bad travel time picks if the geophone is poorly clamped in those formations.

Figure 5.30. Theoretical ray diagram from offset D (632 m). This plot can also be used for offsets B, E, and G. Horizontal scale is the same as the vertical scale. The

approximate shape and position of the Springdale reef is sketched in for offset D. It is important to realize that statements about wave amplitude cannot be made based solely on the convergence or spreading of rays in these diagrams since these are simply ray tracing results. Ray theory states that wave amplitude increases where rays merge and decreases with spreading. To investigate a strucuture for amplitude response, a set of equally spaced rays should be traced into a given angle. The rays shown in the diagram are not equally spaced but were chosen to reach the receiver positions.

Figure 5.31. Theoretical ray diagram from offset J (1 840 m). This plot is approximately correct for offsets F and L also. The approximate outline of the Springdale reef is shown.

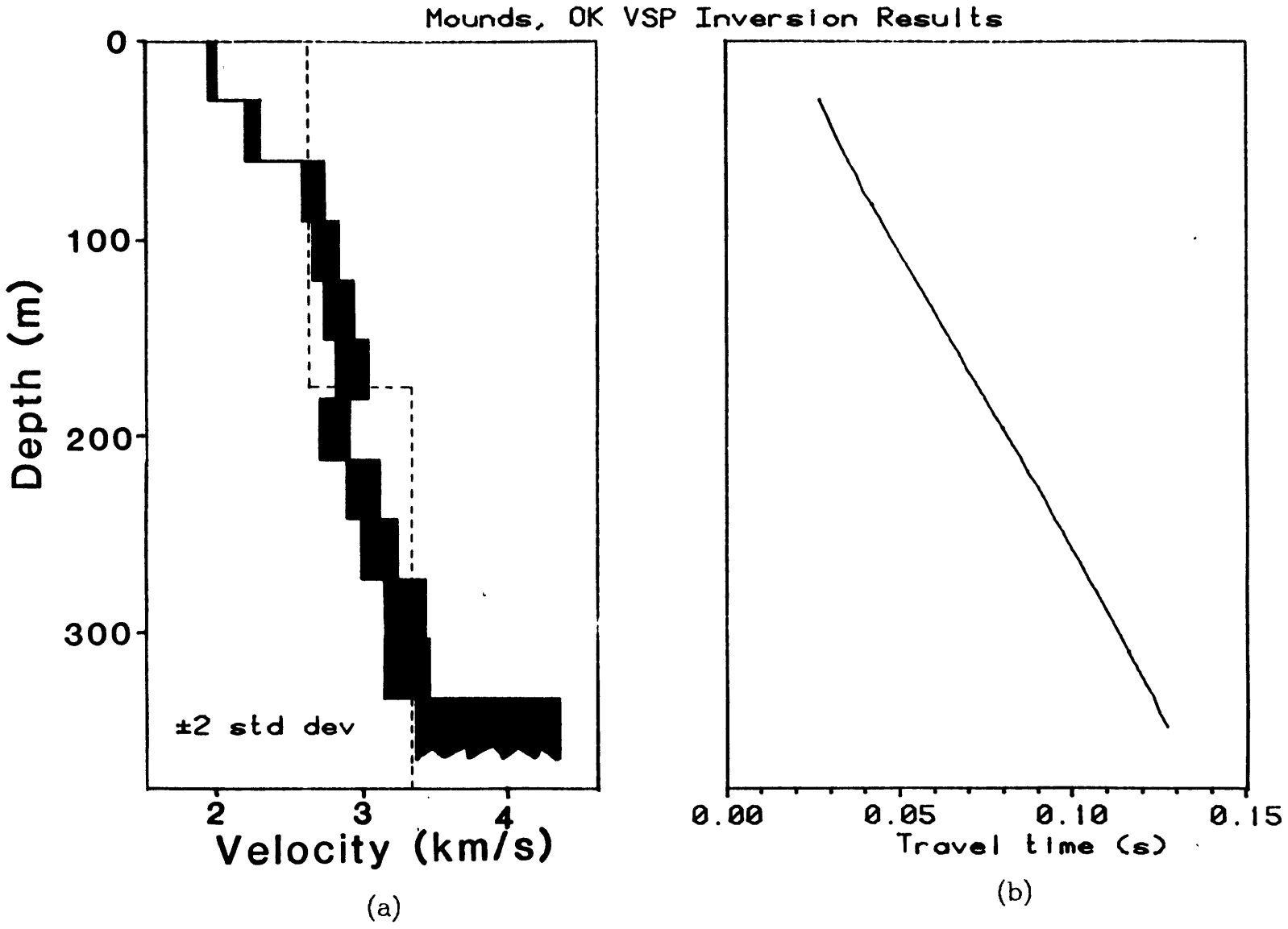

Mounds, OK VSP Inversion Results
0
100
200
300
Depth (m)
±2 std dev
2
3
4
Velocity (km/s)
(a)
0.00
0.05
0.10
0.15
Travel time (s)
(b)

Figure 5.1

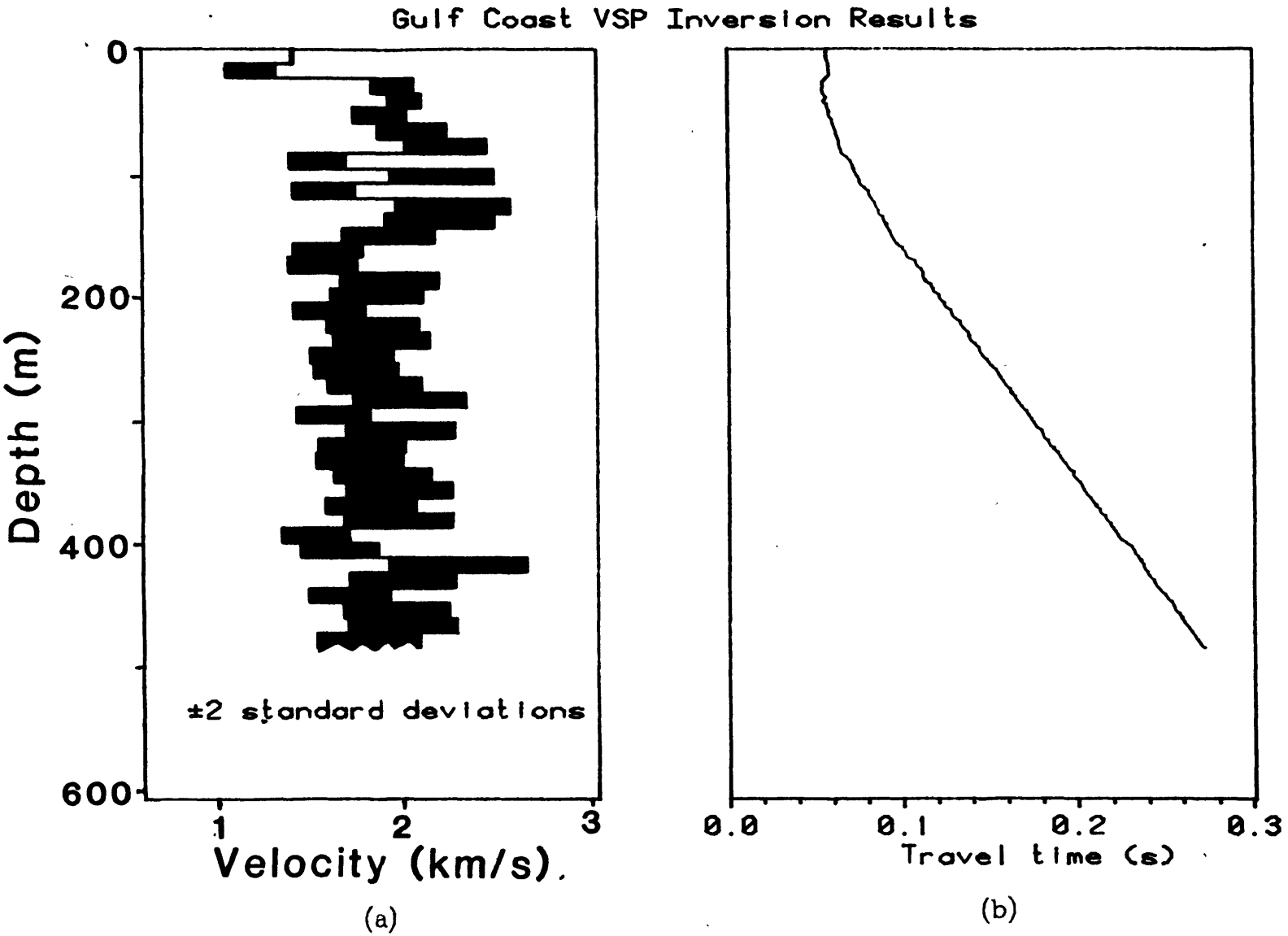

Gulf Coast VSP Inversion Results
Depth (m)
0
200
400
600
±2 standard deviations
1
2
3
Velocity (km/s)
(a)
0.0
0.1
0.2
0.3
Travel time (s)
(b)

Figure 5.2

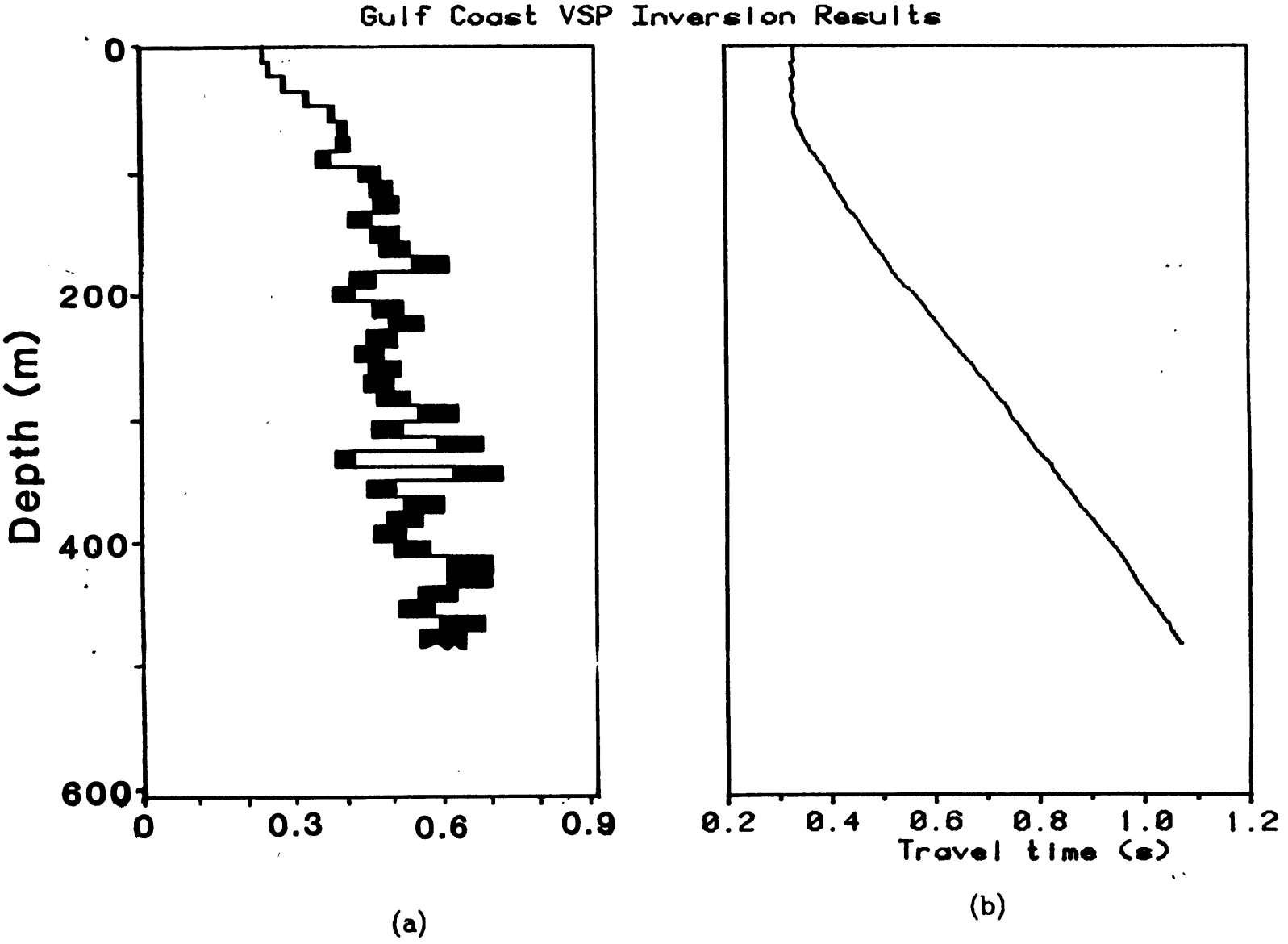

Gulf Coast VSP Inversion Results
Depth (m)
0
200
400
600
0
0.3
0.6
0.9
0.2
0.4
0.6
0.8
1.0
1.2
Travel time (s)
(a)
(b)

Figure 5.3

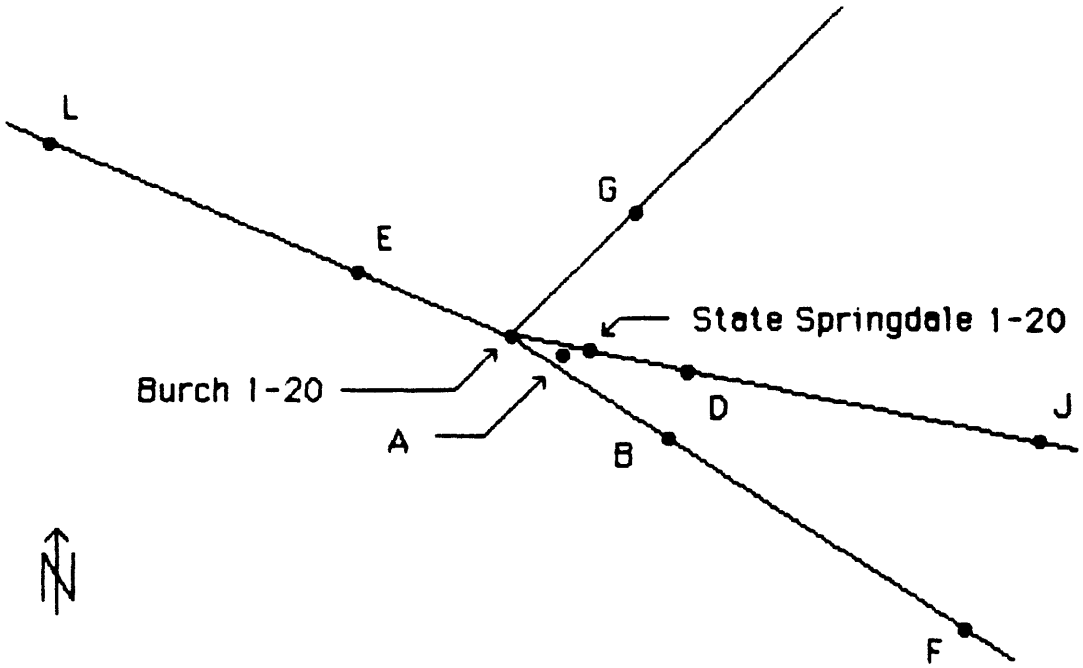


Figure 5.4

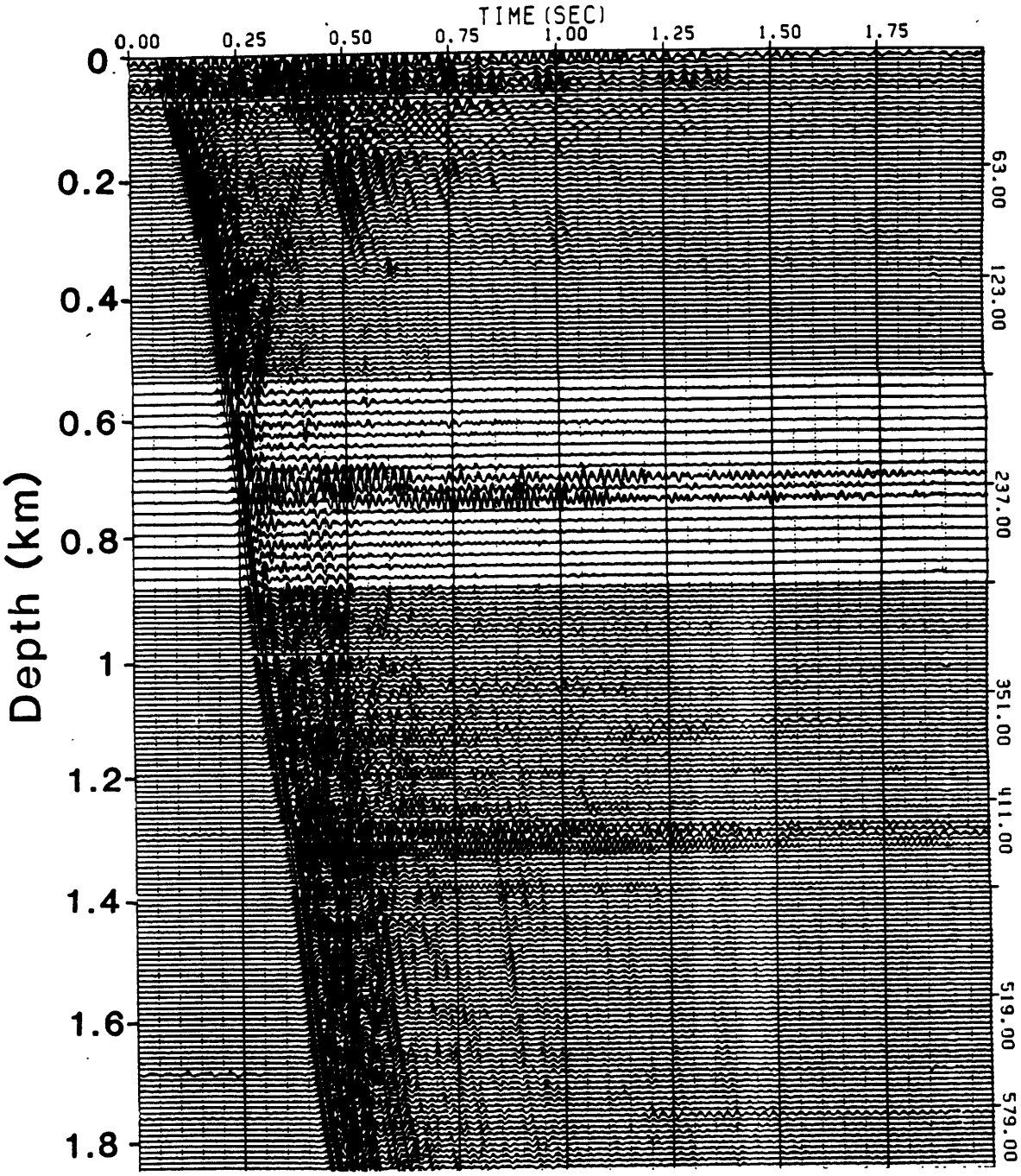

TIME (SEC)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
Depth (km)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
63.00
123.00
237.00
351.00
411.00
519.00
579.00

Figure 5.5

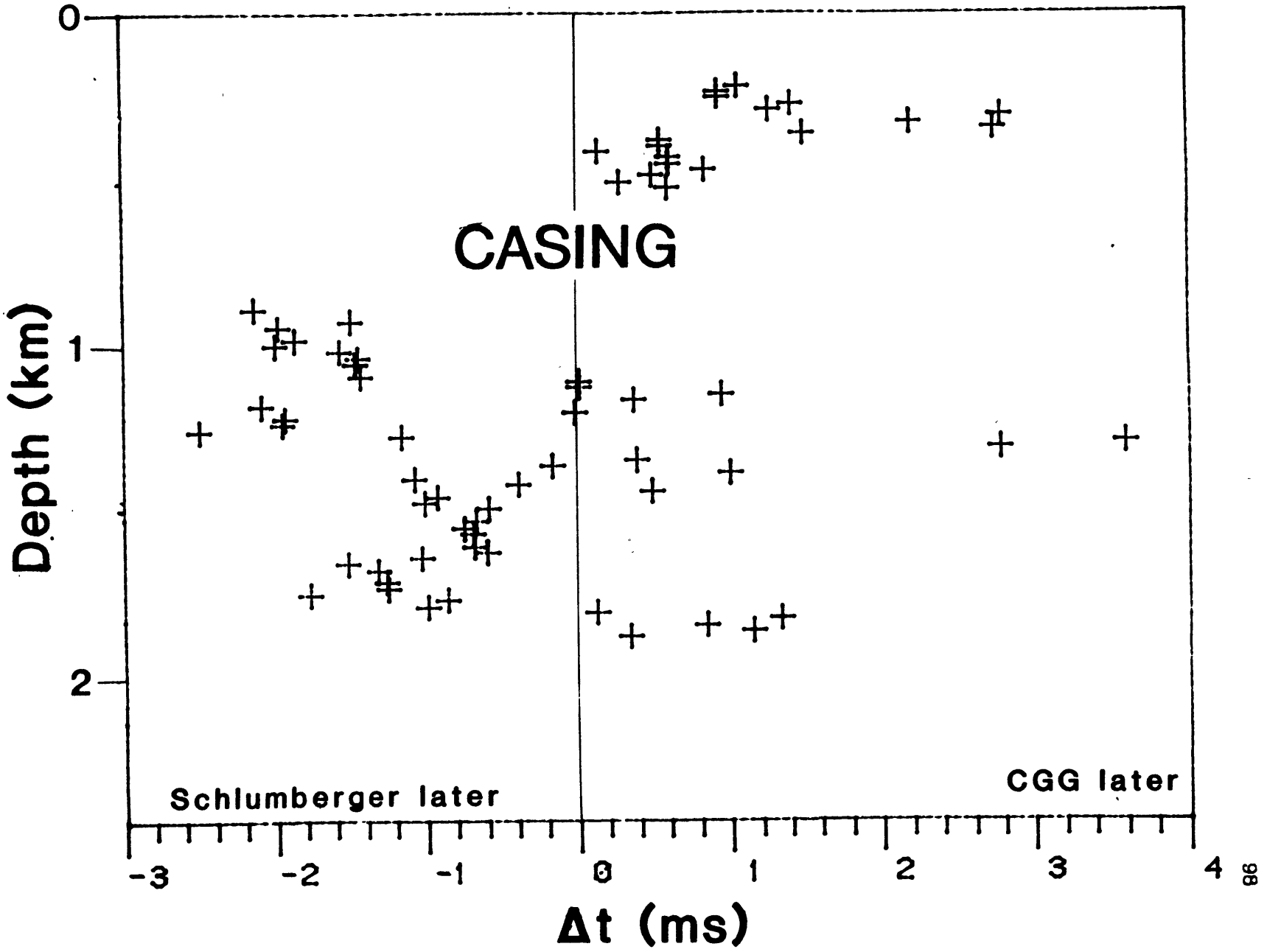

CASING
Depth (km)
0
1
2
Schlumberger later
CGG later
-3
-2
-1
0
1
2
3
4
Δt (ms)

Figure 5.6

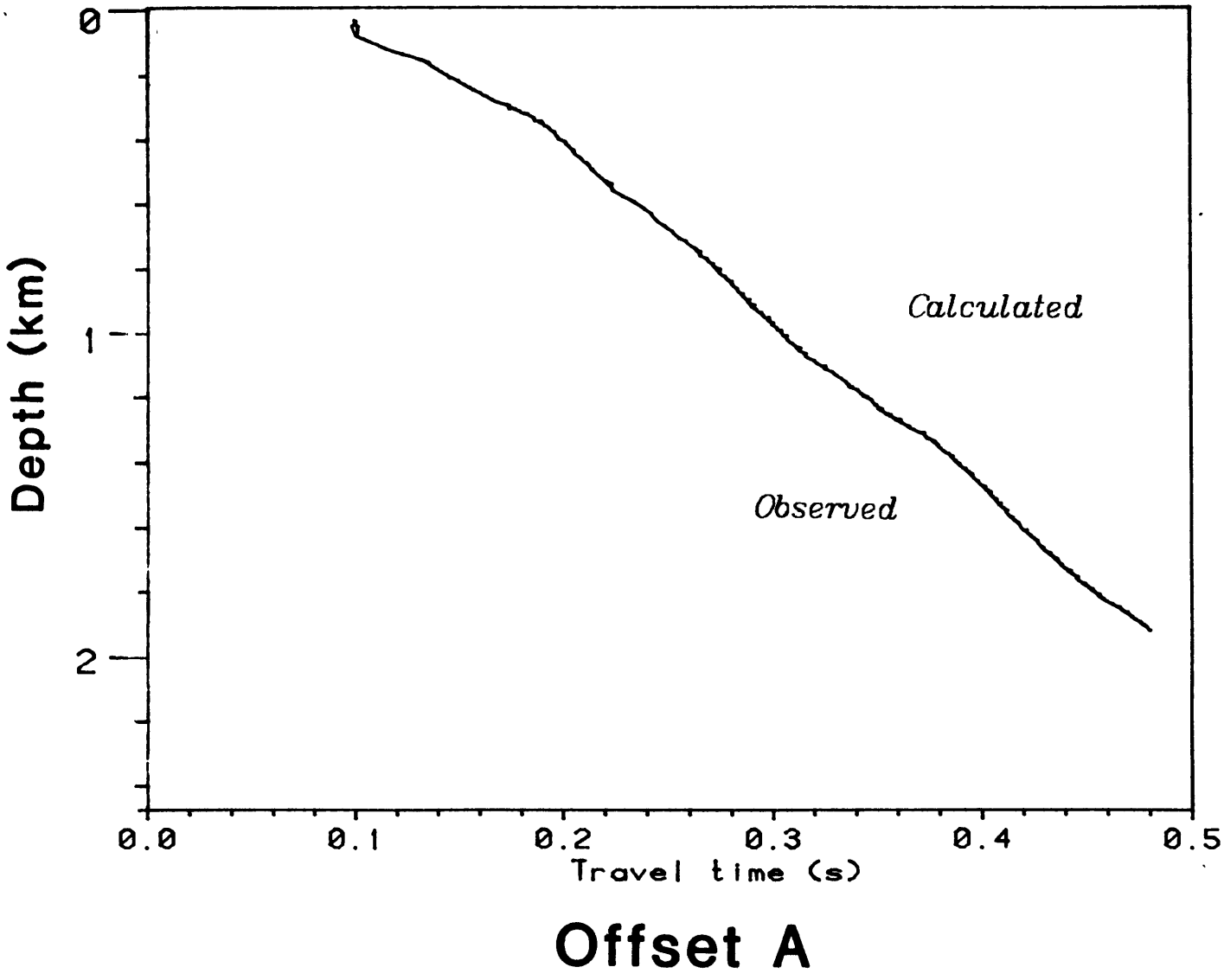

Depth (km)
0
1
2
Calculated
Observed
0.0
0.1
0.2
0.3
0.4
0.5
Travel time (s)
Offset A

Figure 5.7

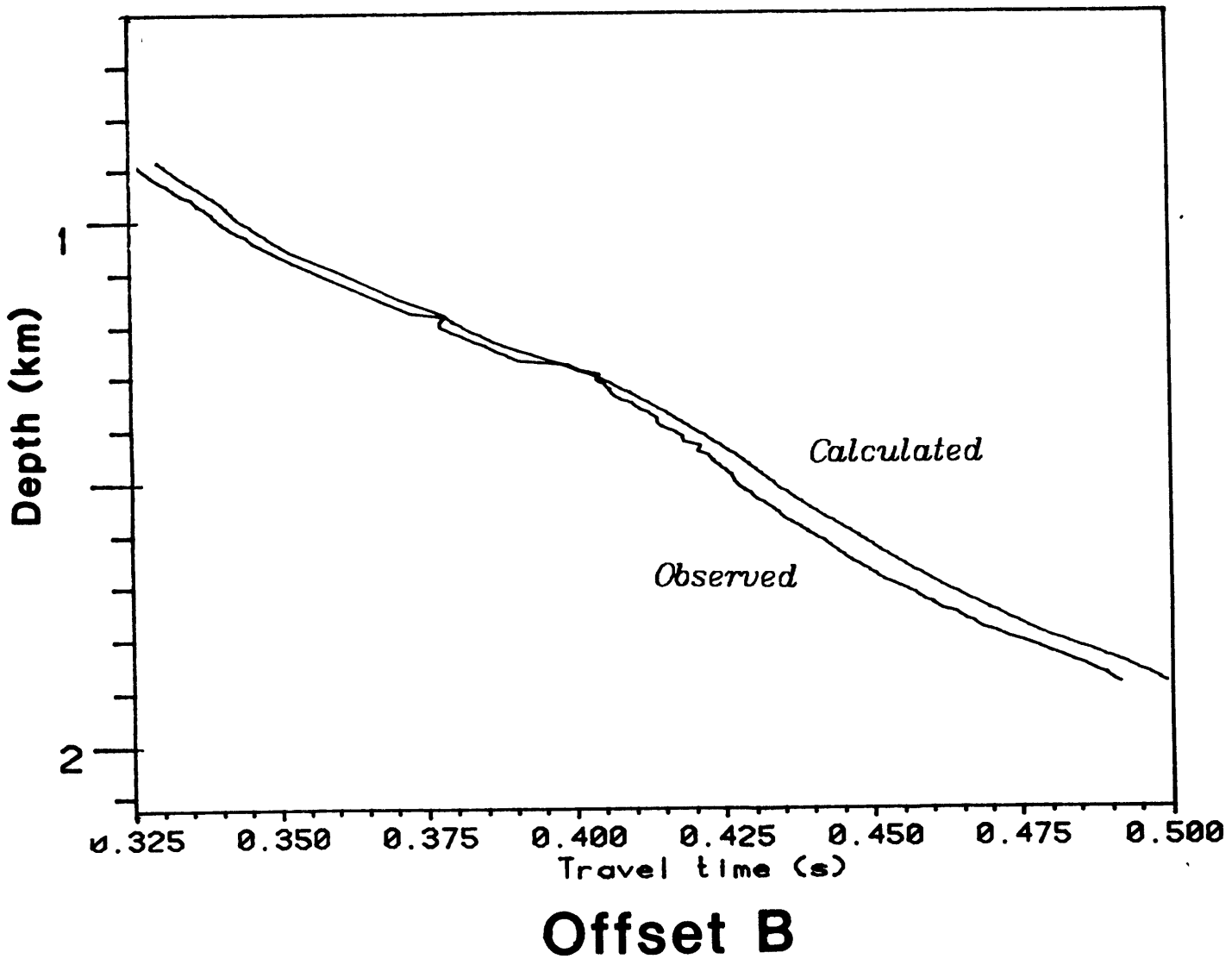

Depth (km)
1
2
Calculated
Observed
0.325
0.350
0.375
0.400
0.425
0.450
0.475
0.500
Travel time (s)
Offset B

Figure 5.8

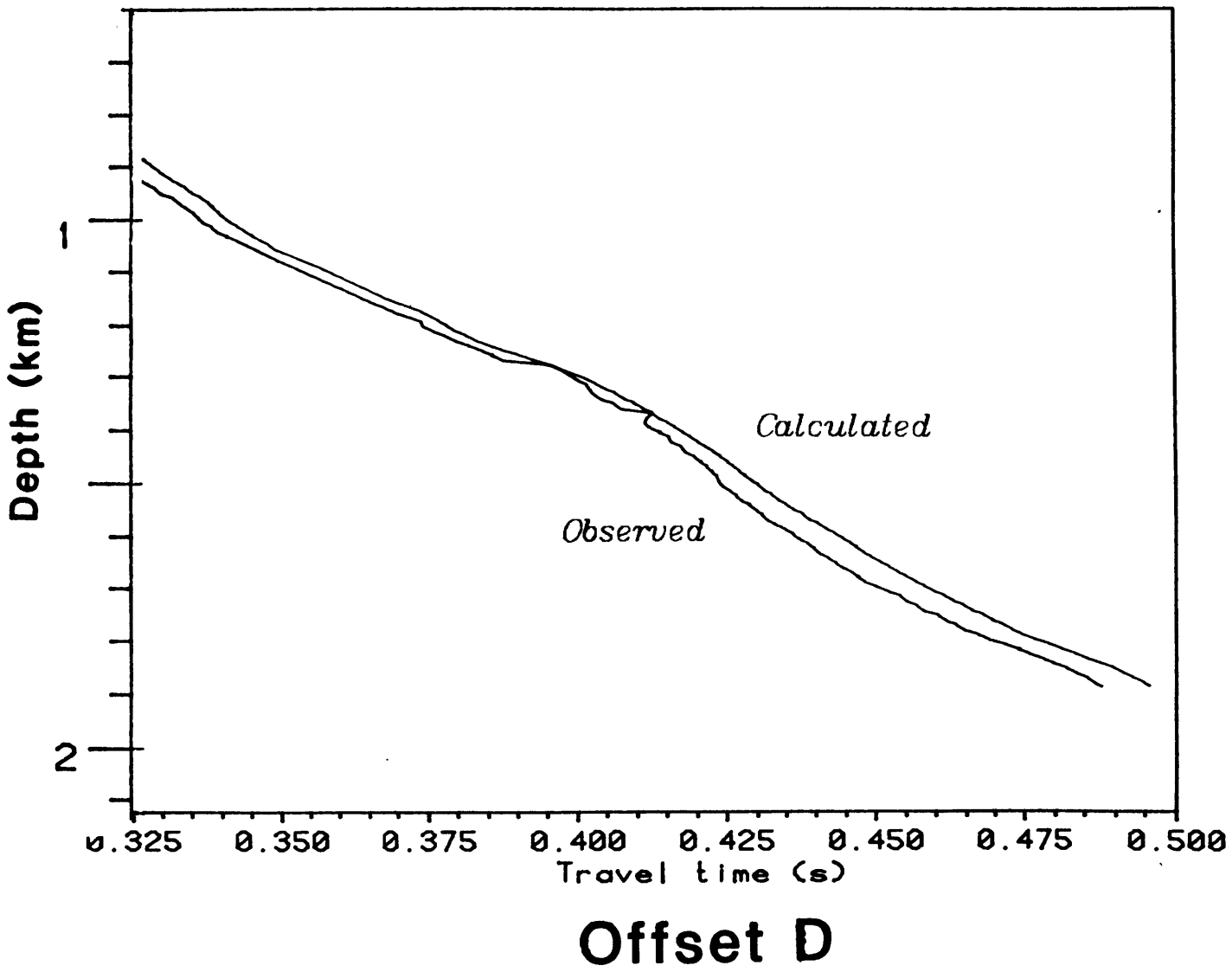

Depth (km)
1
2
Calculated
Observed
0.325
0.350
0.375
0.400
0.425
0.450
0.475
0.500
Travel time (s)
Offset D

Figure 5.9

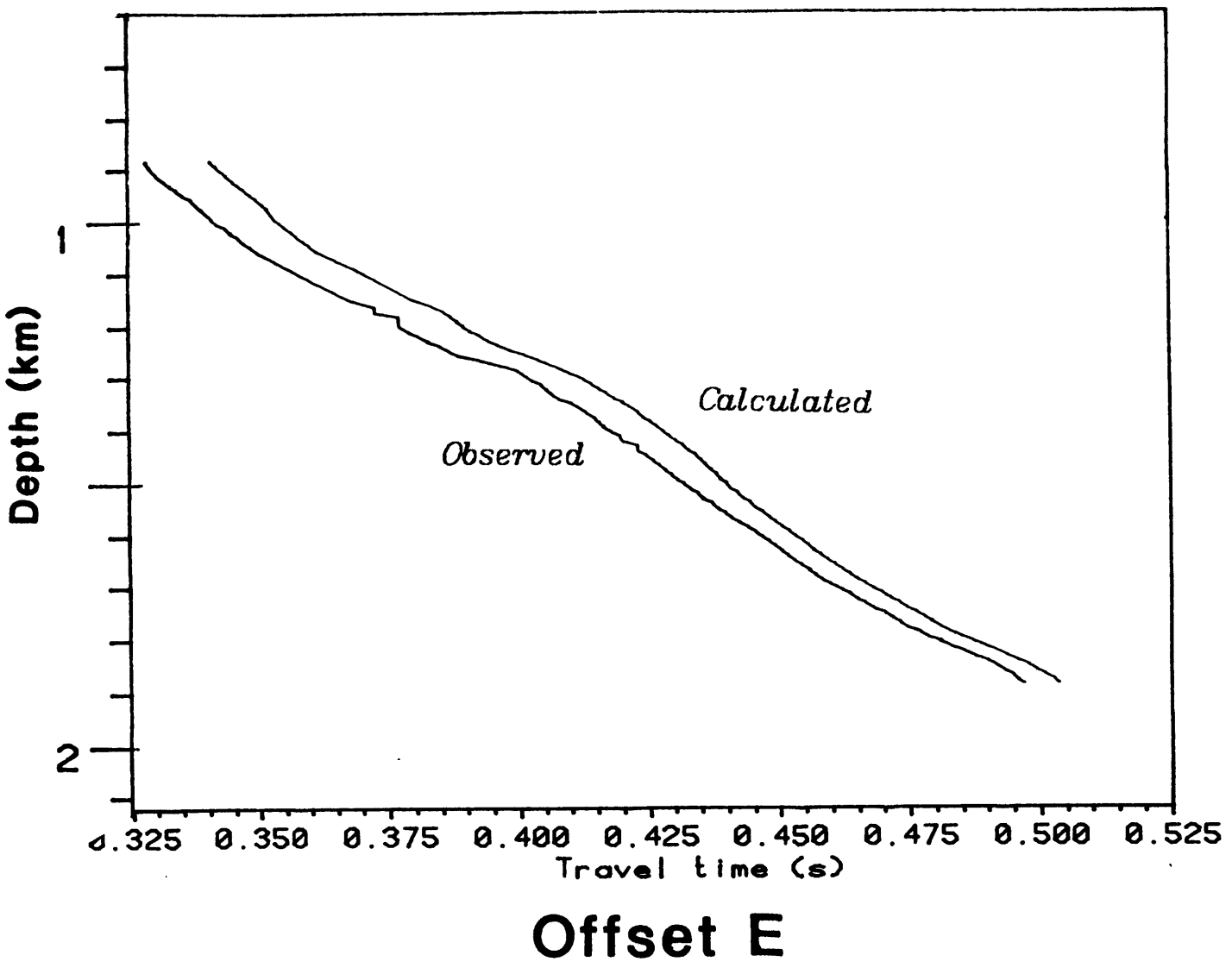

Depth (km)
1
2
Calculated
Observed
0.325
0.350
0.375
0.400
0.425
0.450
0.475
0.500
0.525
Travel time (s)
Offset E

Figure 5.10

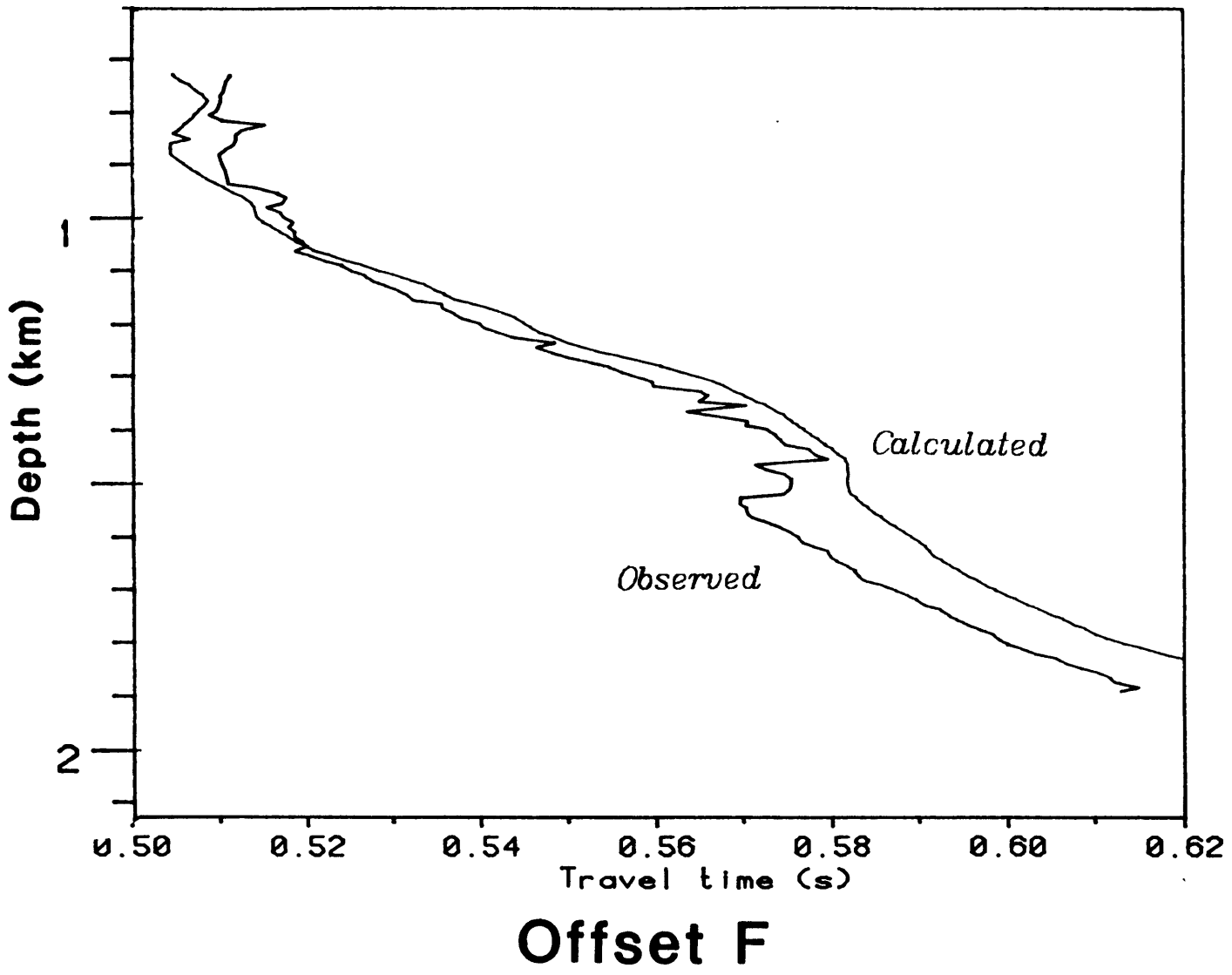

Depth (km)
1
2
Calculated
Observed
0.50
0.52
0.54
0.56
0.58
0.60
0.62
Travel time (s)
Offset F

Figure 5.11

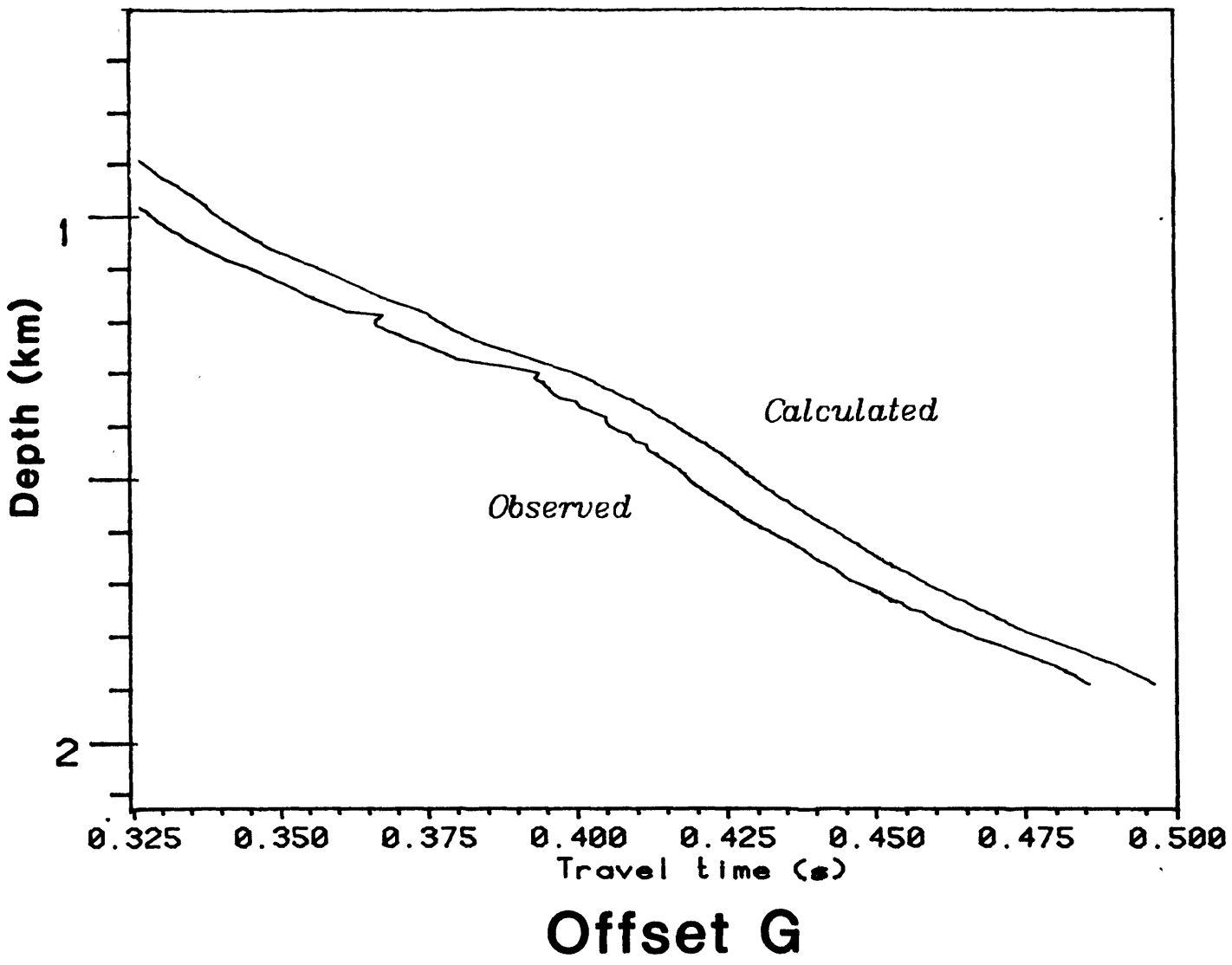

Depth (km)
1
2
Calculated
Observed
0.325
0.350
0.375
0.400
0.425
0.450
0.475
0.500
Travel time (s)
Offset G

Figure 5.12

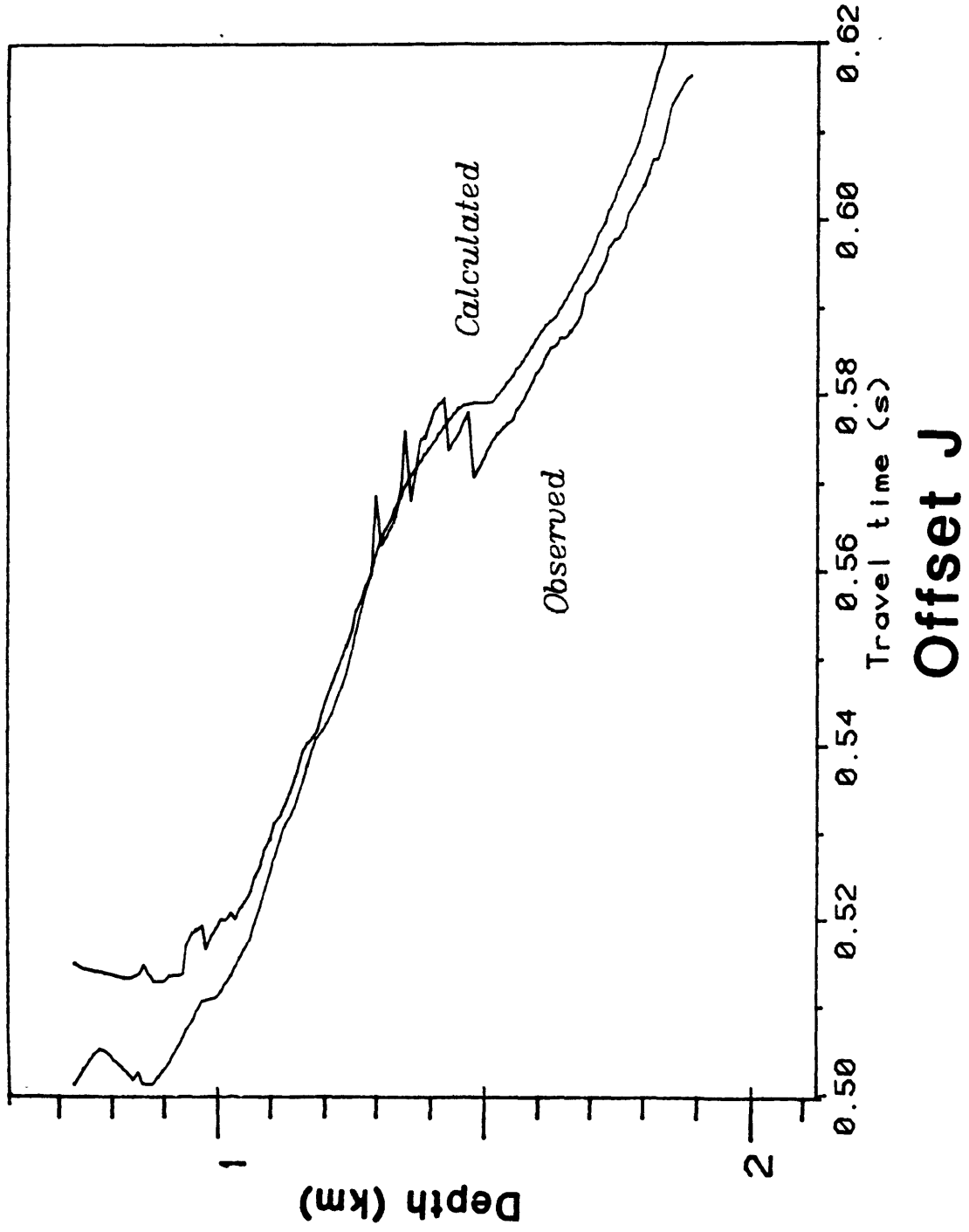

Offset J
Travel time (s)
0.50
0.52
0.54
0.56
0.58
0.60
0.62
Depth (km)
1
2
Calculated
Observed

Figure 5.13

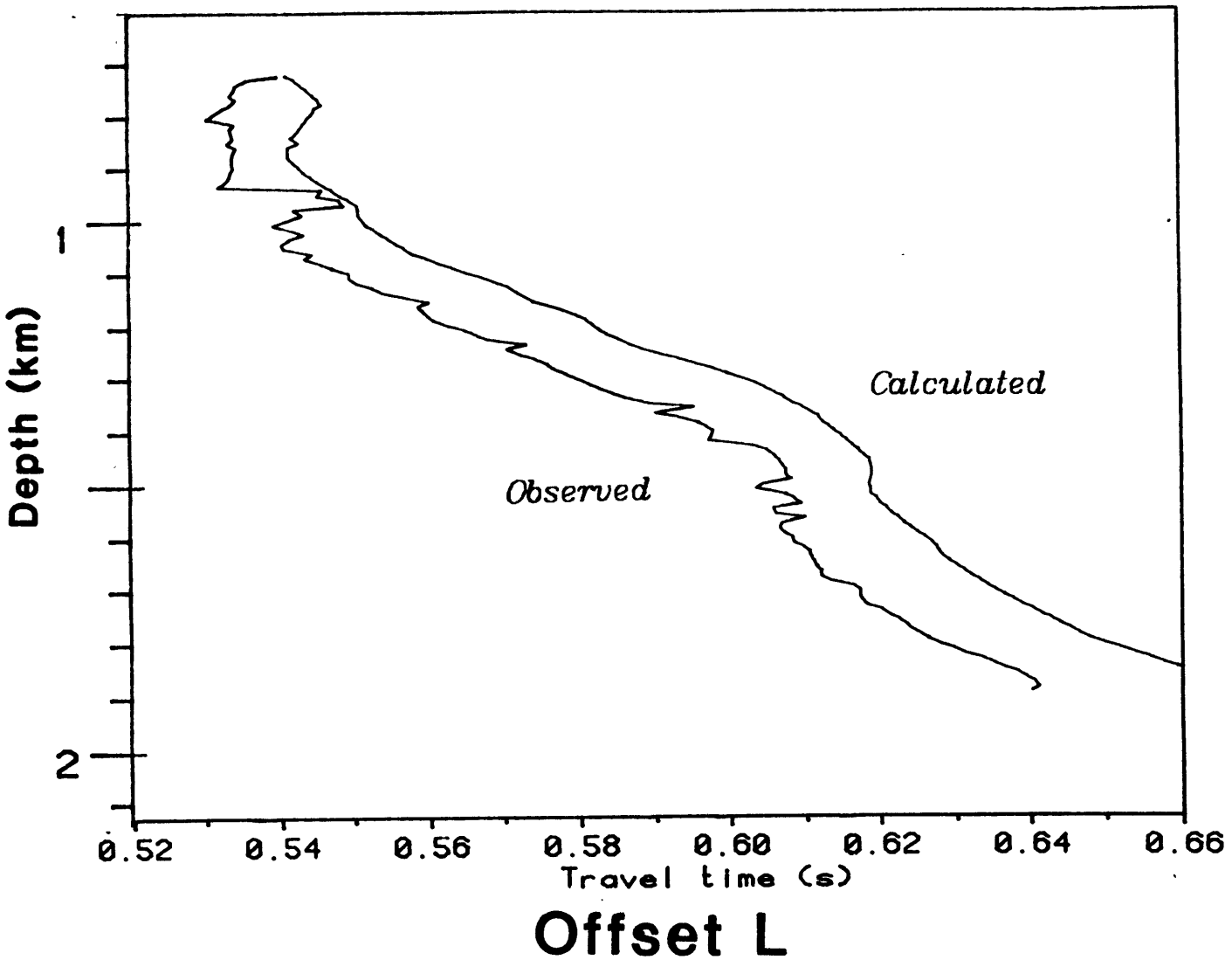

Depth (km)
1
2
Calculated
Observed
0.52
0.54
0.56
0.58
0.60
0.62
0.64
0.66
Travel time (s)
Offset L

Figure 5.14

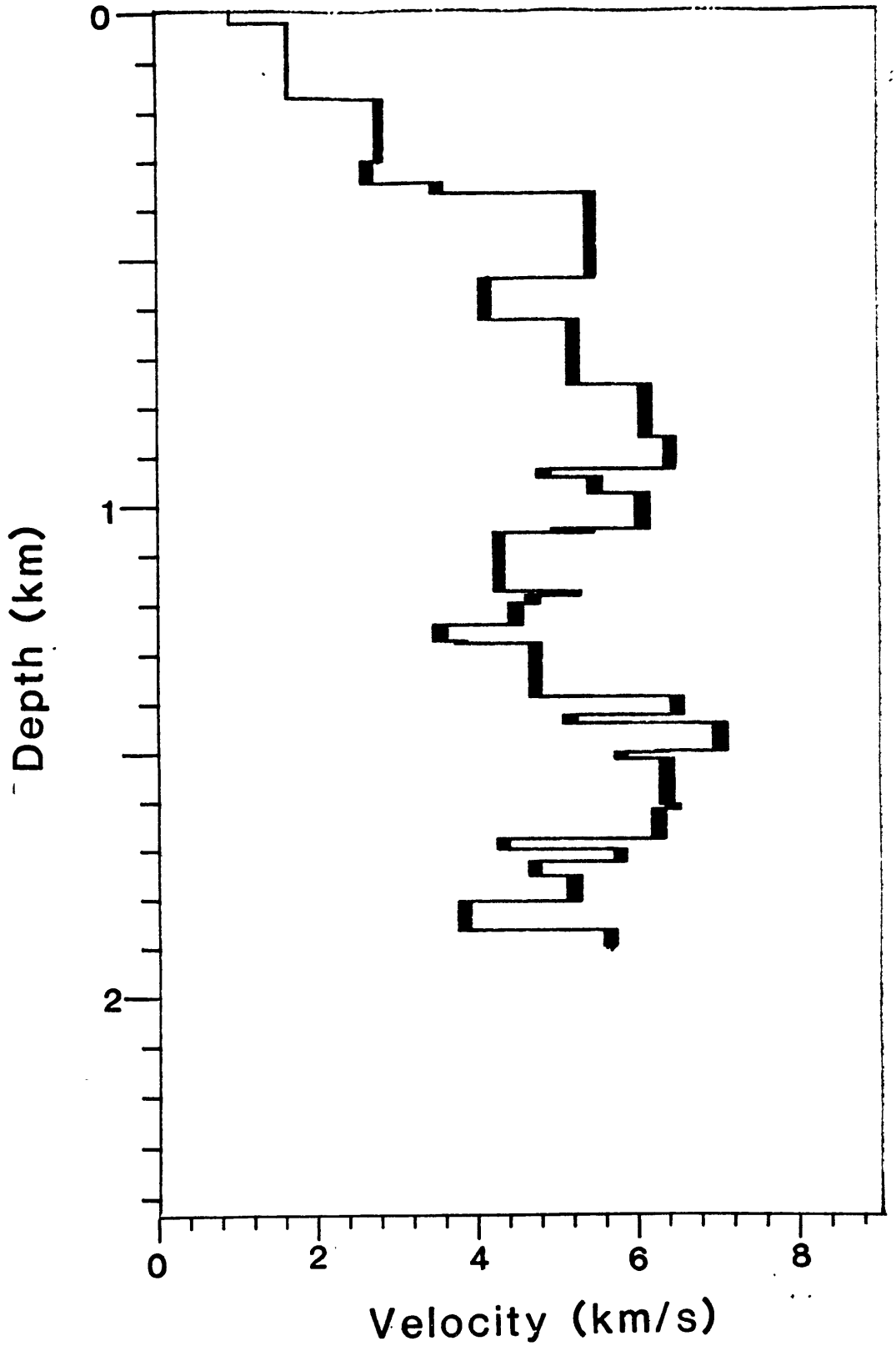


Figure 5.15

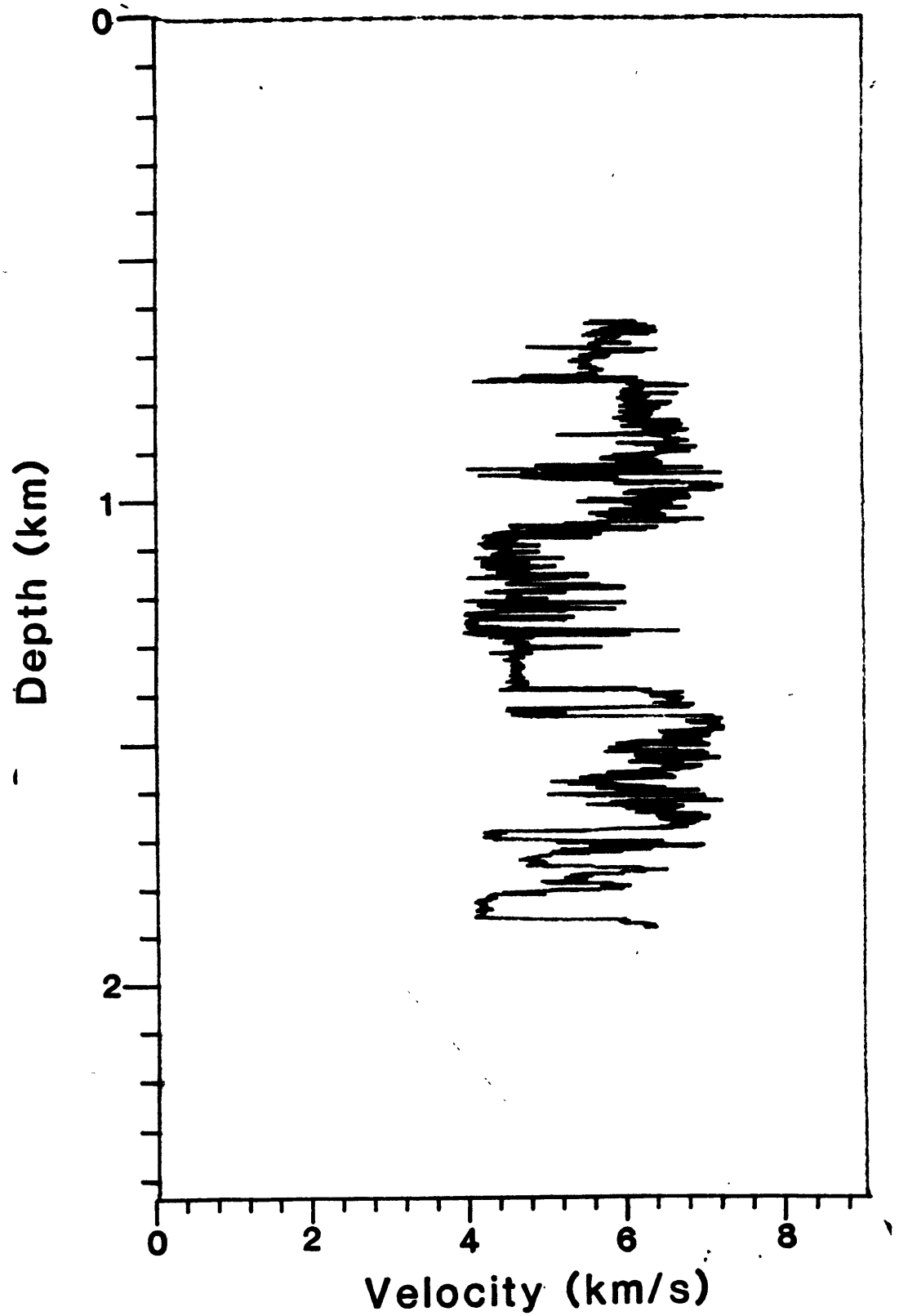


Figure 5.16

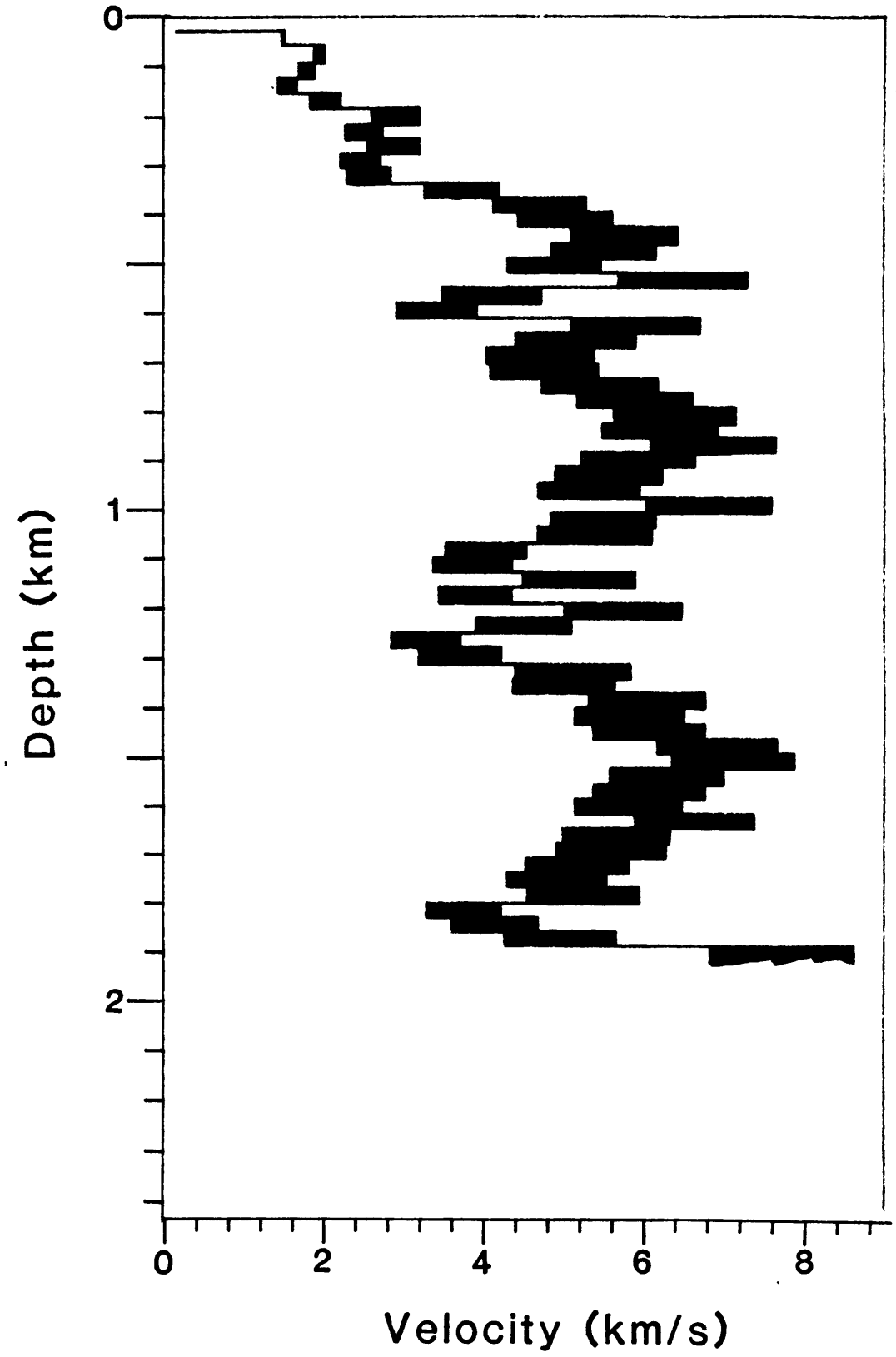


Figure 5.17

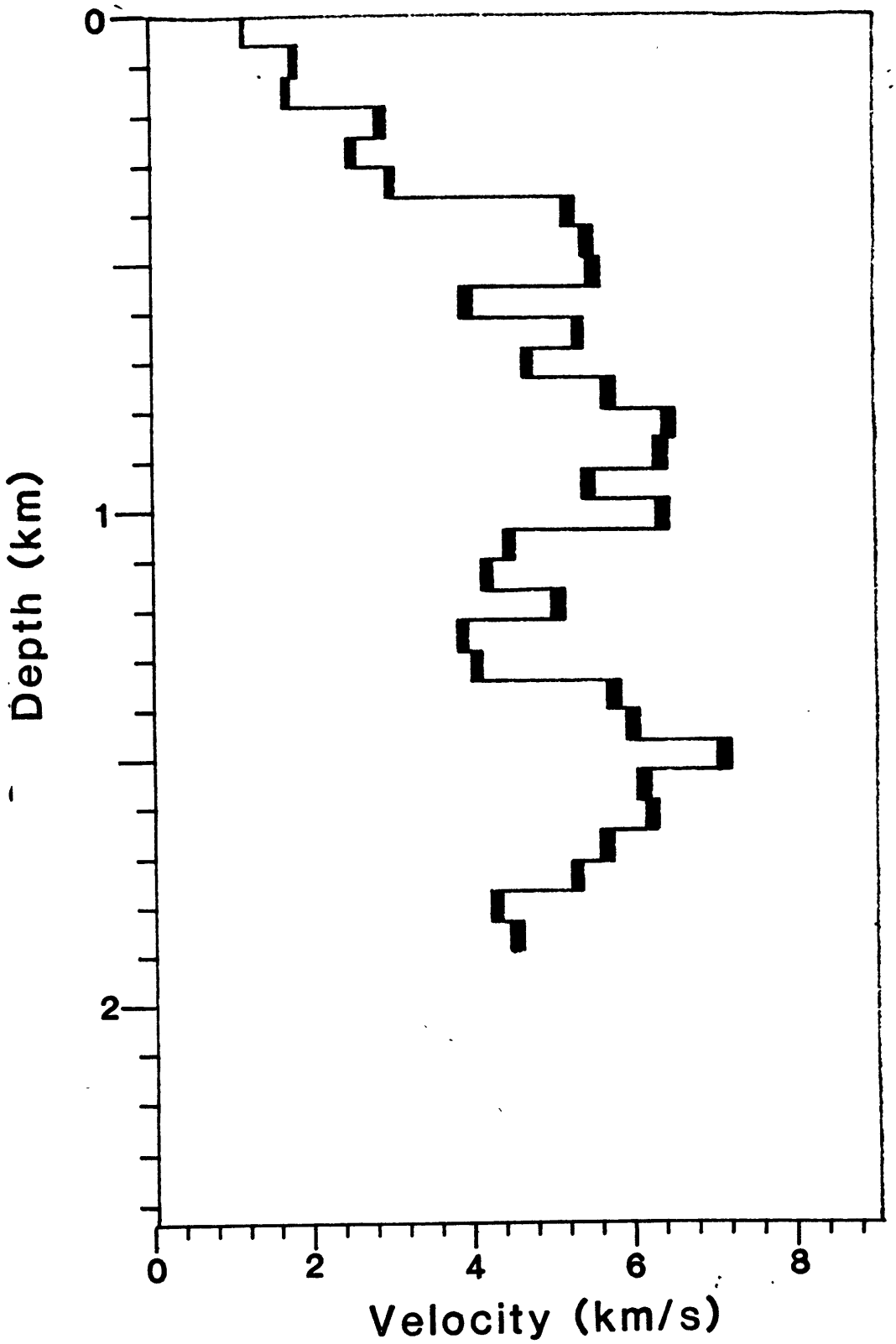


Figure 5.18

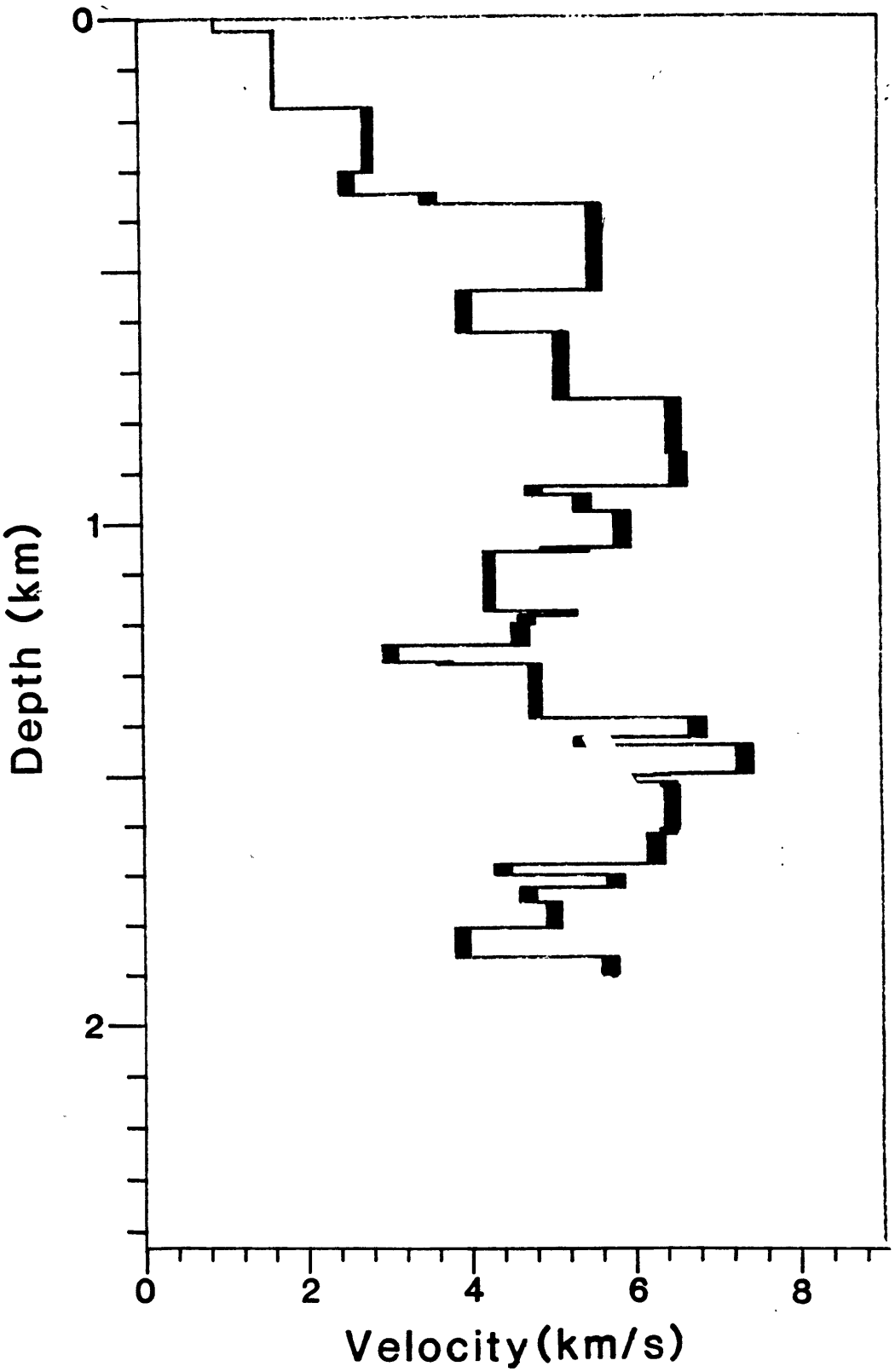


Figure 5.19

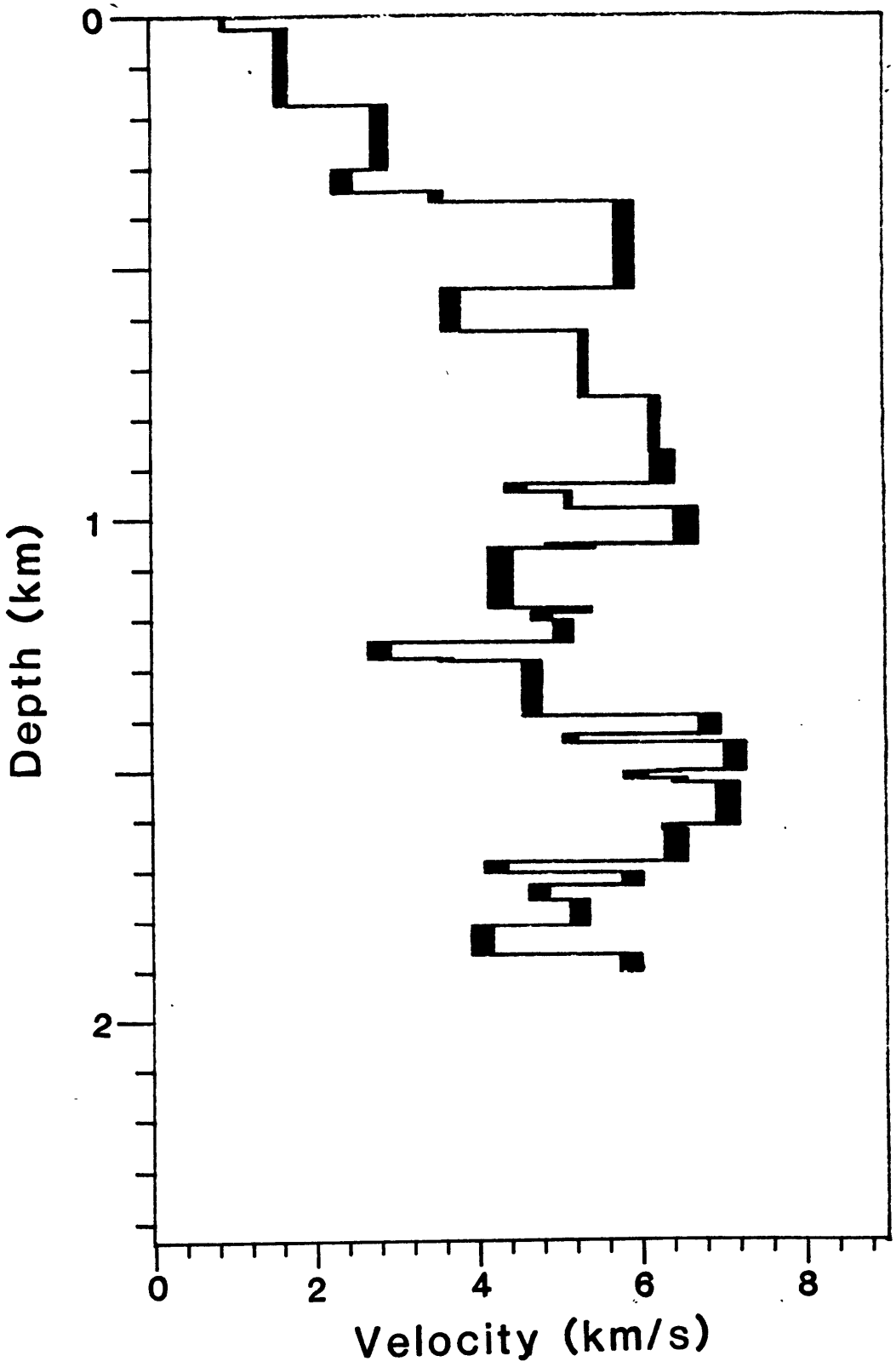
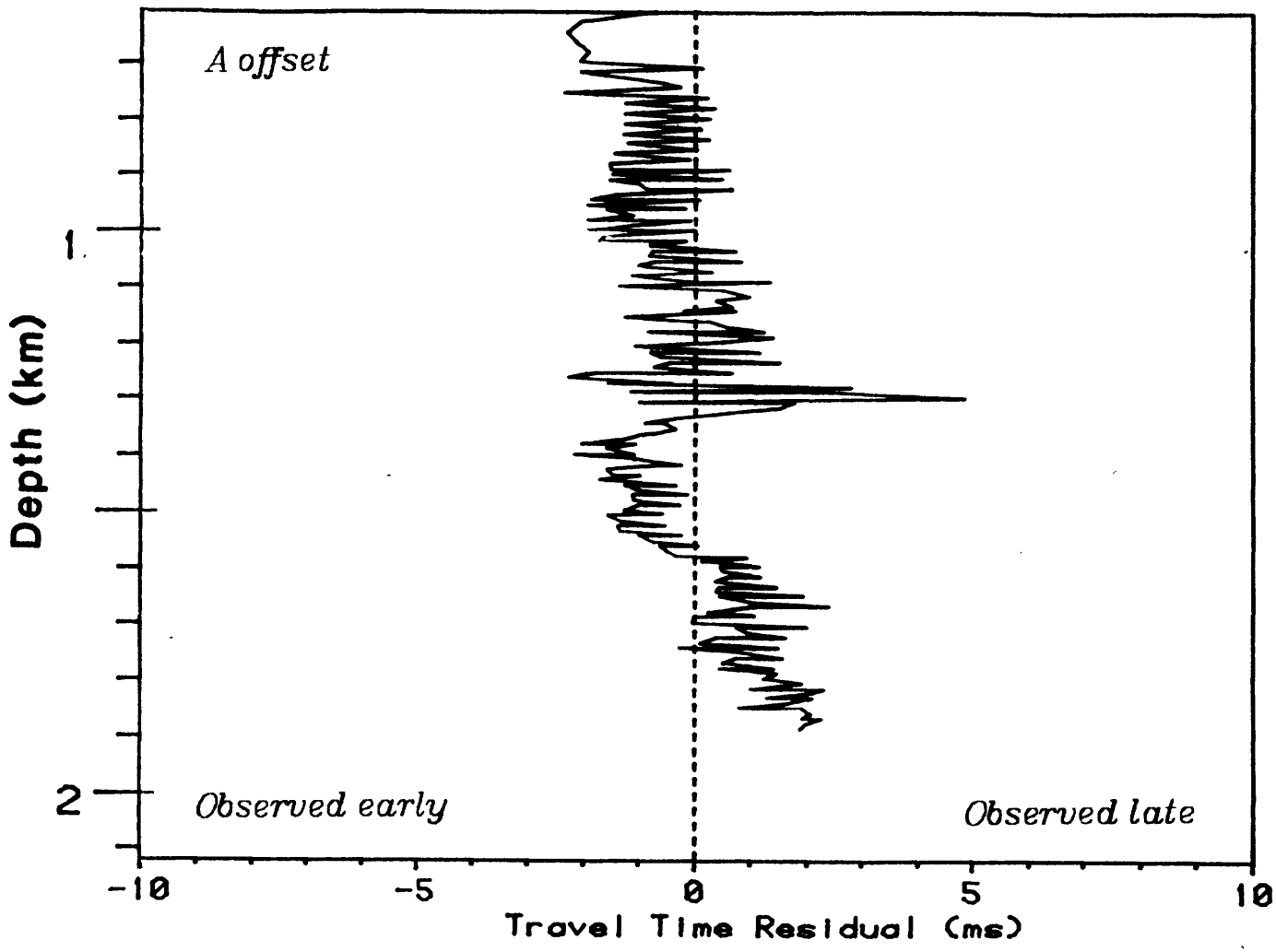


Figure 5.20



109 m ESE of Burch 1-20 well

Figure 5.21

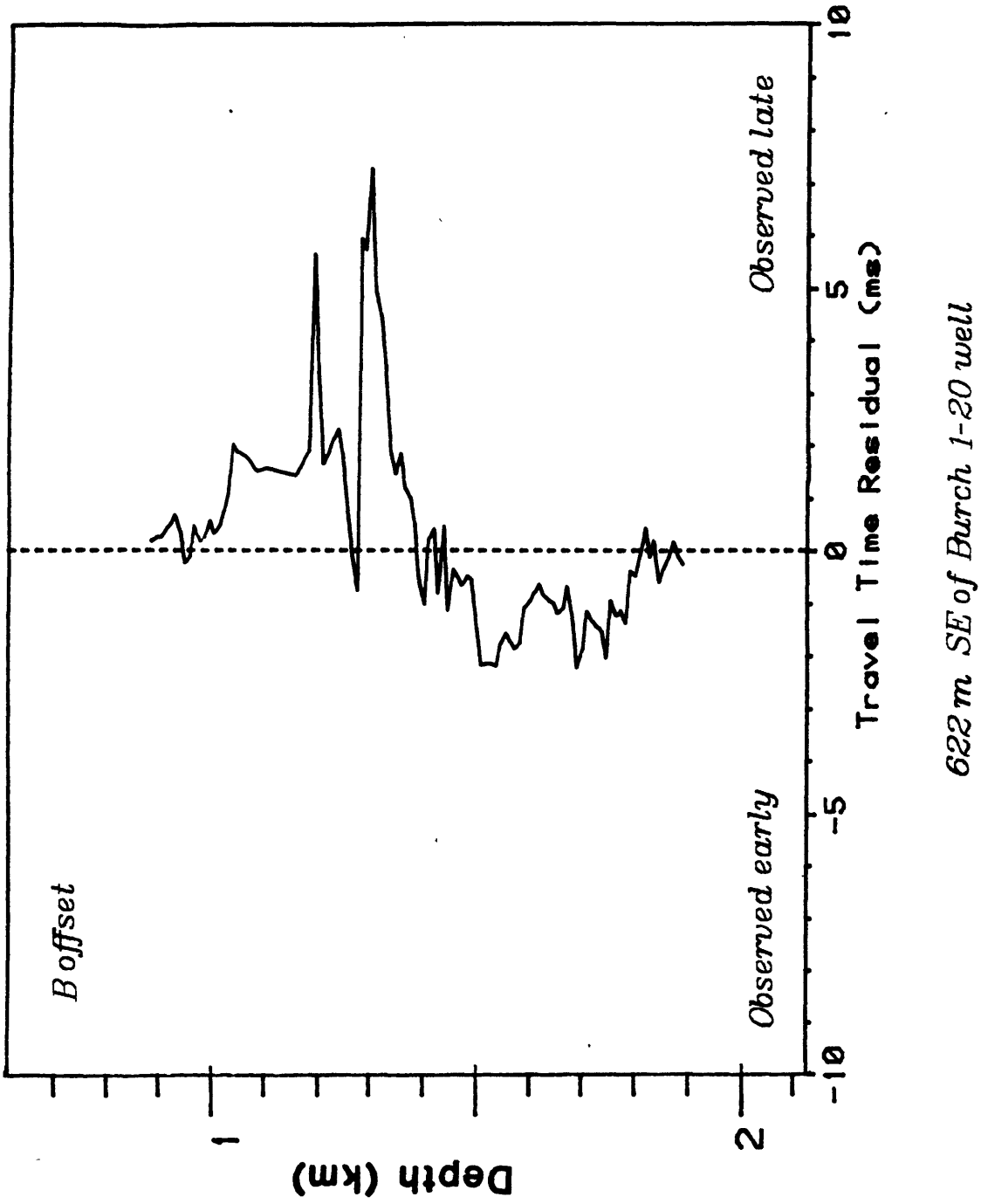

622 m SE of Burch 1-20 well
Travel Time Residual (ms)
10
5
0
-5
-10
Observed late
Observed early
B offset
Depth (km)
1
2

Figure 5.22

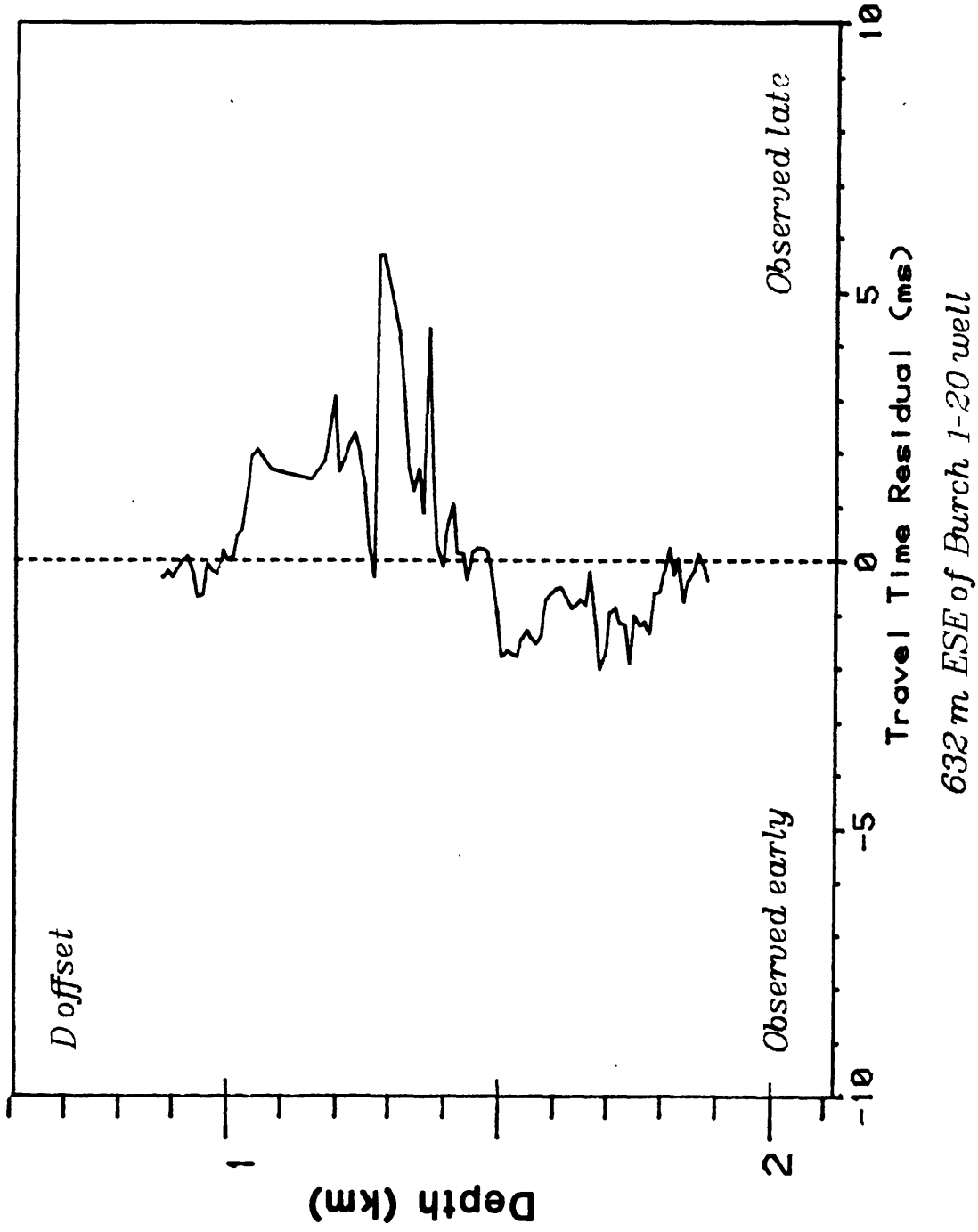
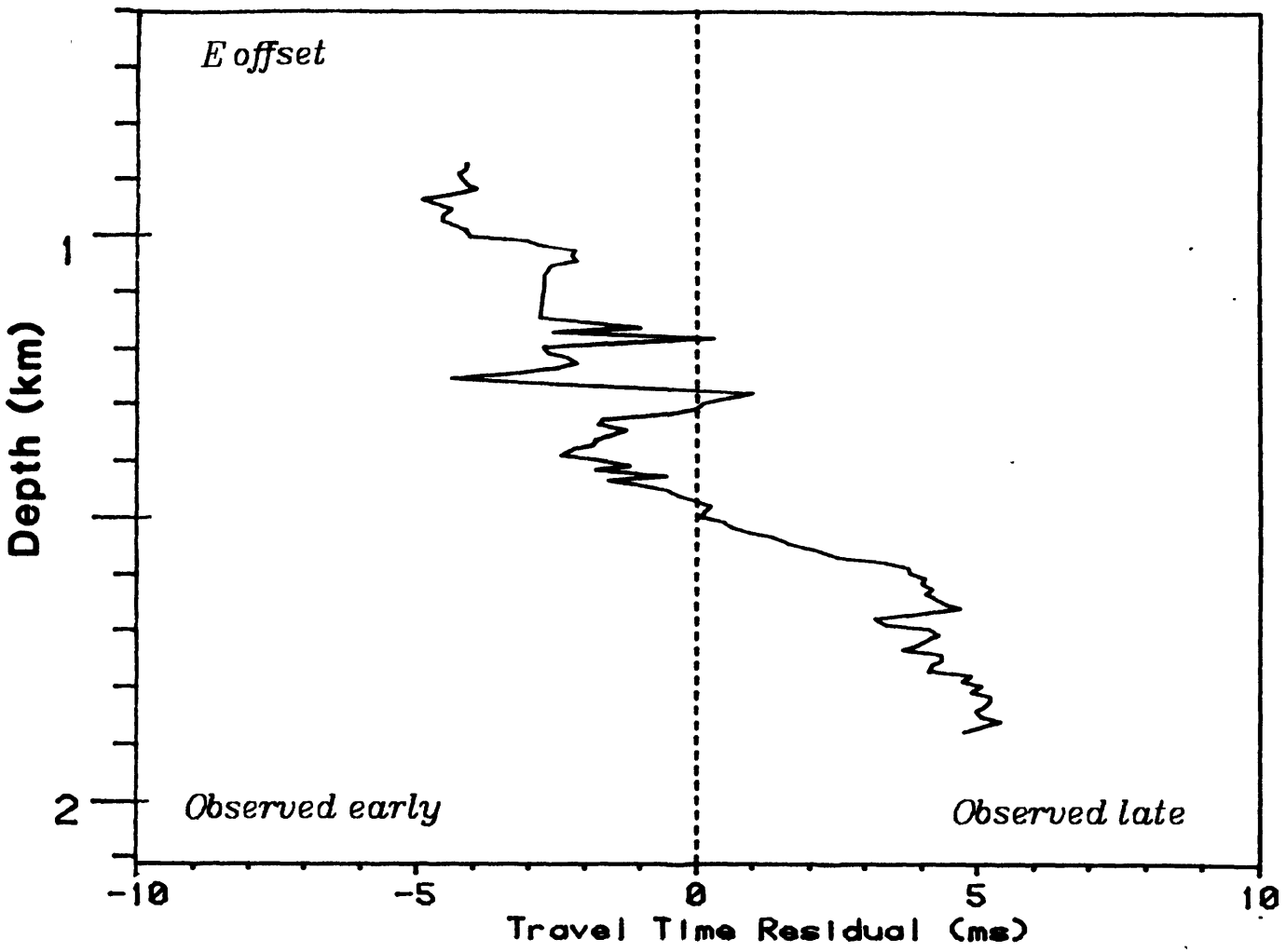
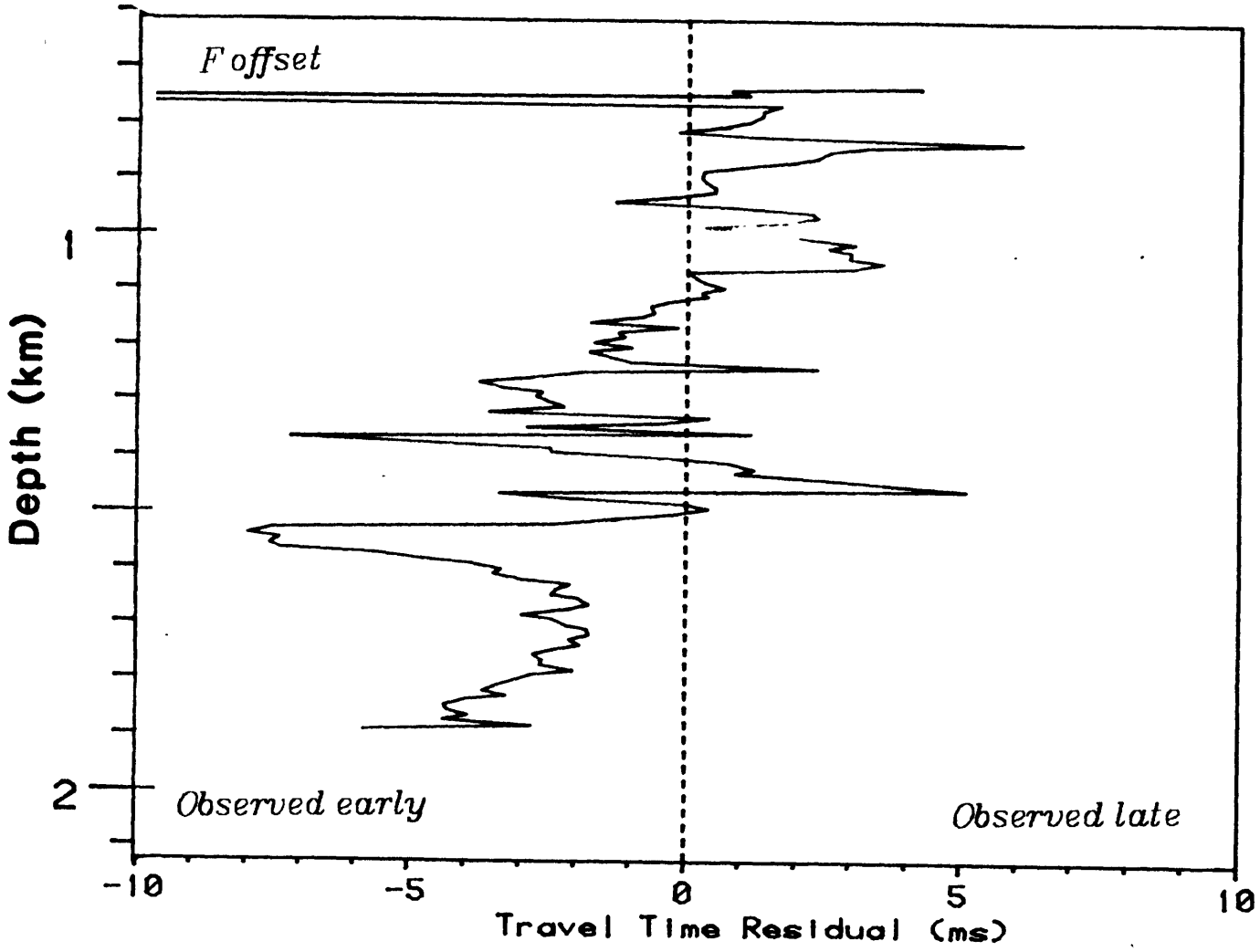

632 m ESE of Burch 1-20 well
Travel Time Residual (ms)
10
5
0
-5
-10
Observed late
Observed early
D offset
Depth (km)
1
2

Figure 5.23



732 m WNW of Burch 1-20 well

Figure 5.24



1 874 m SE of Burch 1-20 well

Figure 5.25

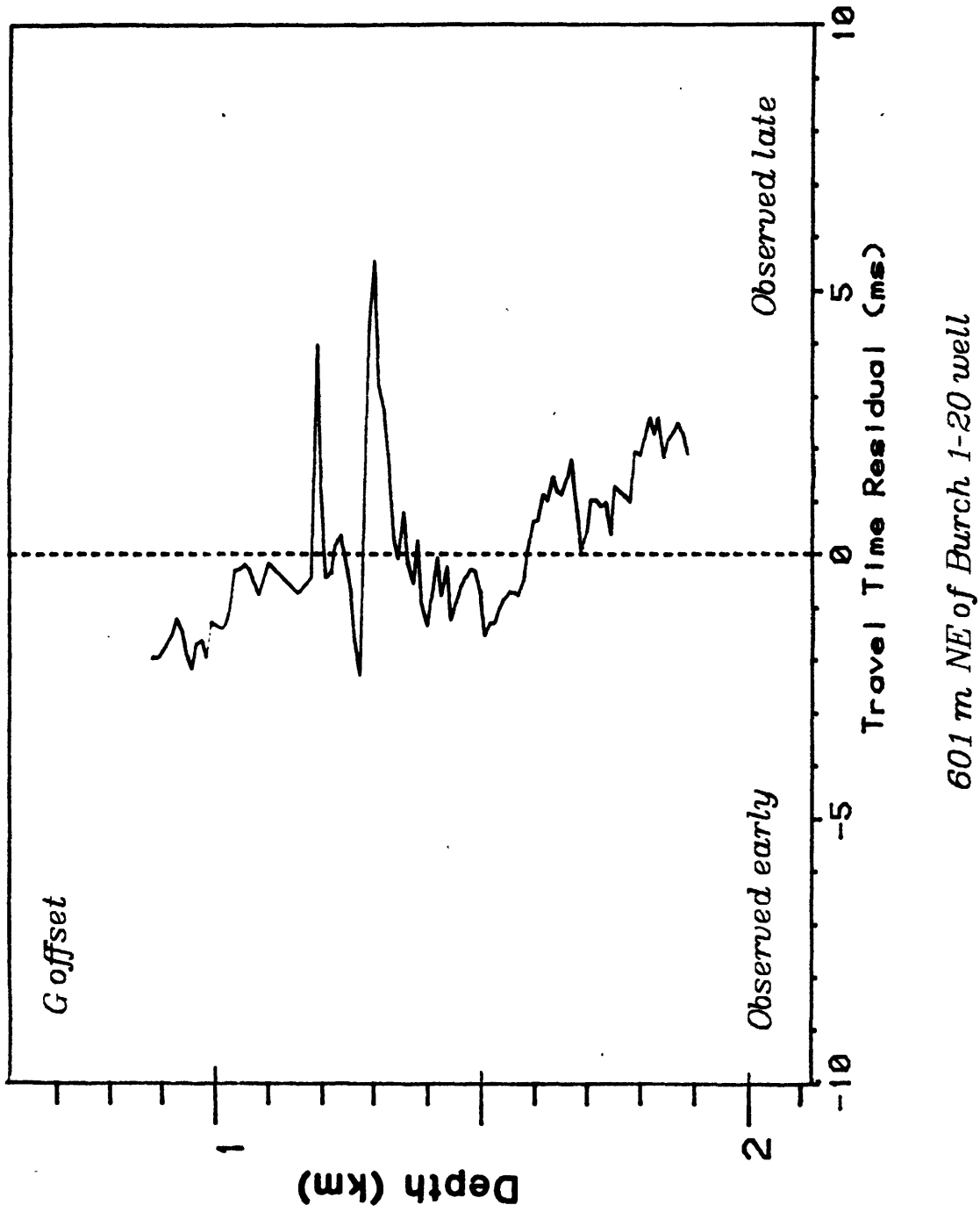

601 m NE of Burch 1-20 well
Travel Time Residual (ms)
10
5
0
-5
-10
Observed late
Observed early
G offset
Depth (km)
1
2

Figure 5.26

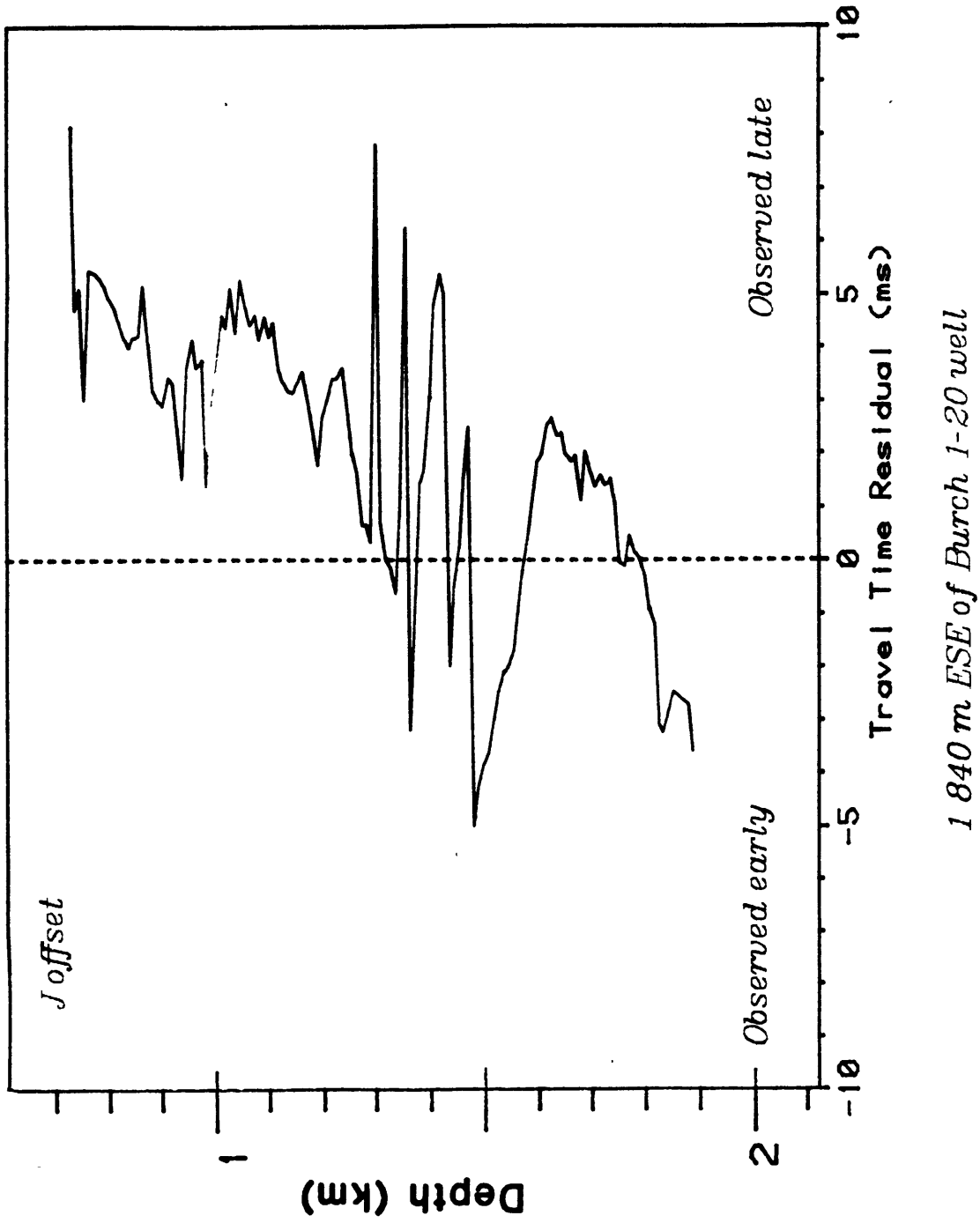

1 840 m ESE of Burch 1-20 well
Travel Time Residual (ms)
10
5
0
-5
-10
Observed late
Observed early
J offset
Depth (km)
1
2

Figure 5.27

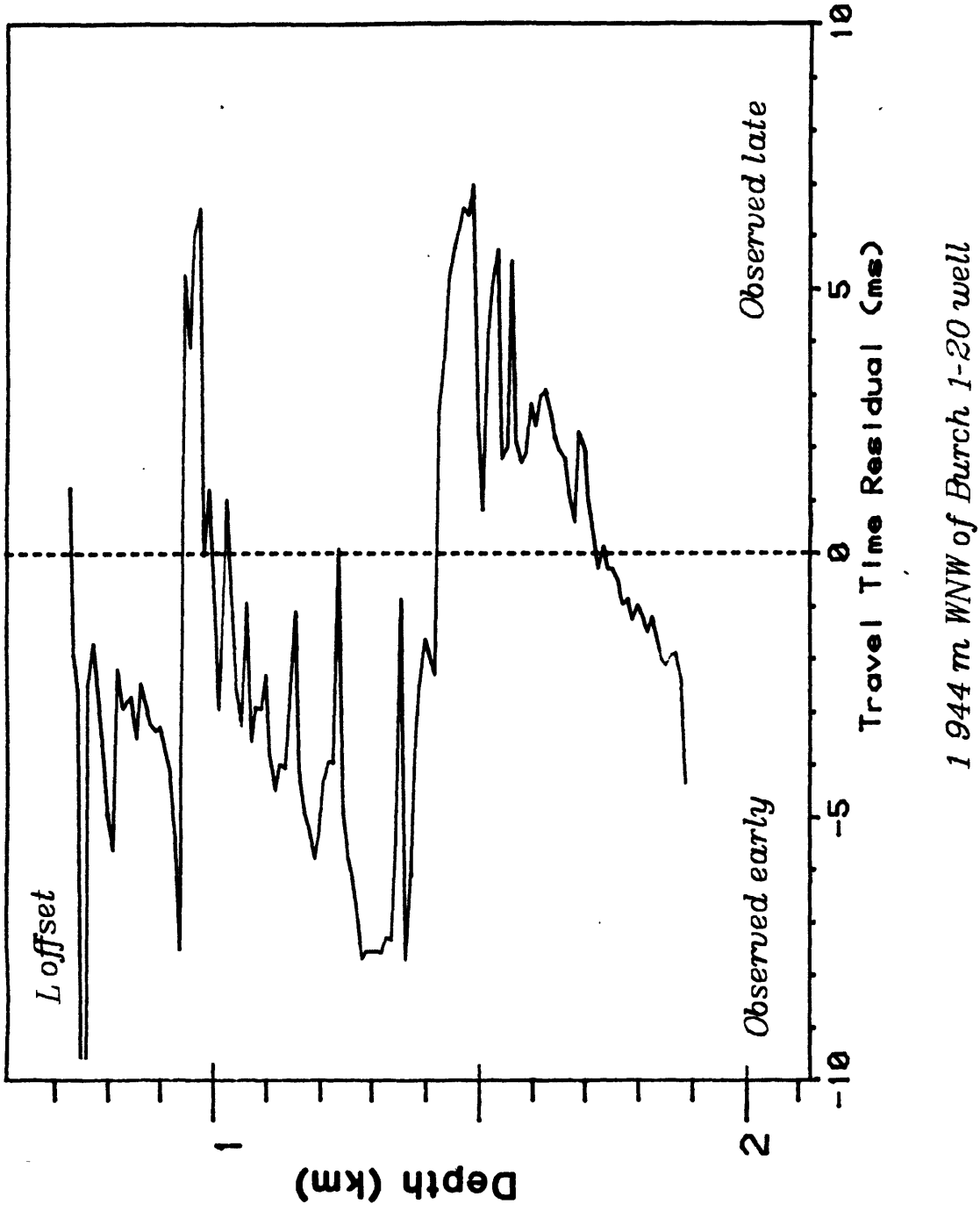

1 944 m WNW of Burch 1-20 well
Travel Time Residual (ms)
10
5
0
-5
-10
Observed late
Observed early
L offset
Depth (km)
1
2

Figure 5.28

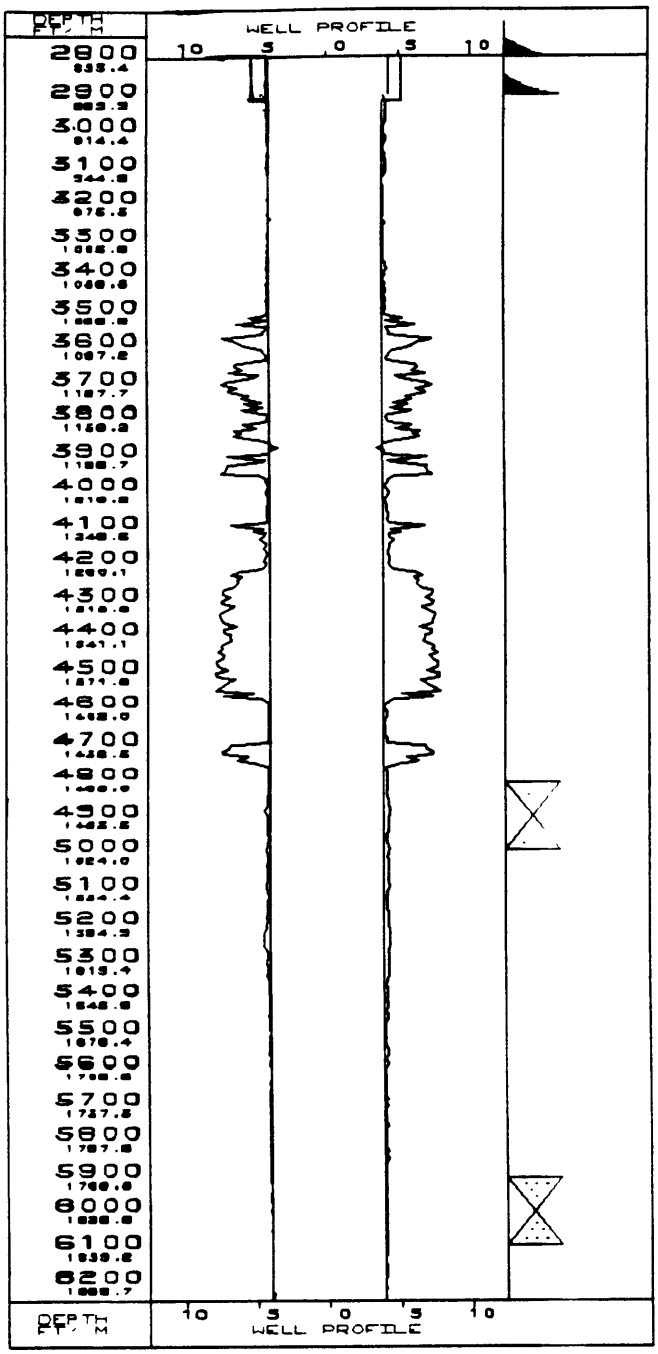

DEPTH
FT. M
WELL PROFILE
10 5 0 5 10
2800
2900
3000
3100
3200
3300
3400
3500
3600
3700
3800
3900
4000
4100
4200
4300
4400
4500
4600
4700
4800
4900
5000
5100
5200
5300
5400
5500
5600
5700
5800
5900
6000
6100
6200
DEPTH
FT. M
10 5 0 5 10
WELL PROFILE
BURCH #1 - 20
MANISTEE CO. MICH.

Figure 5.29

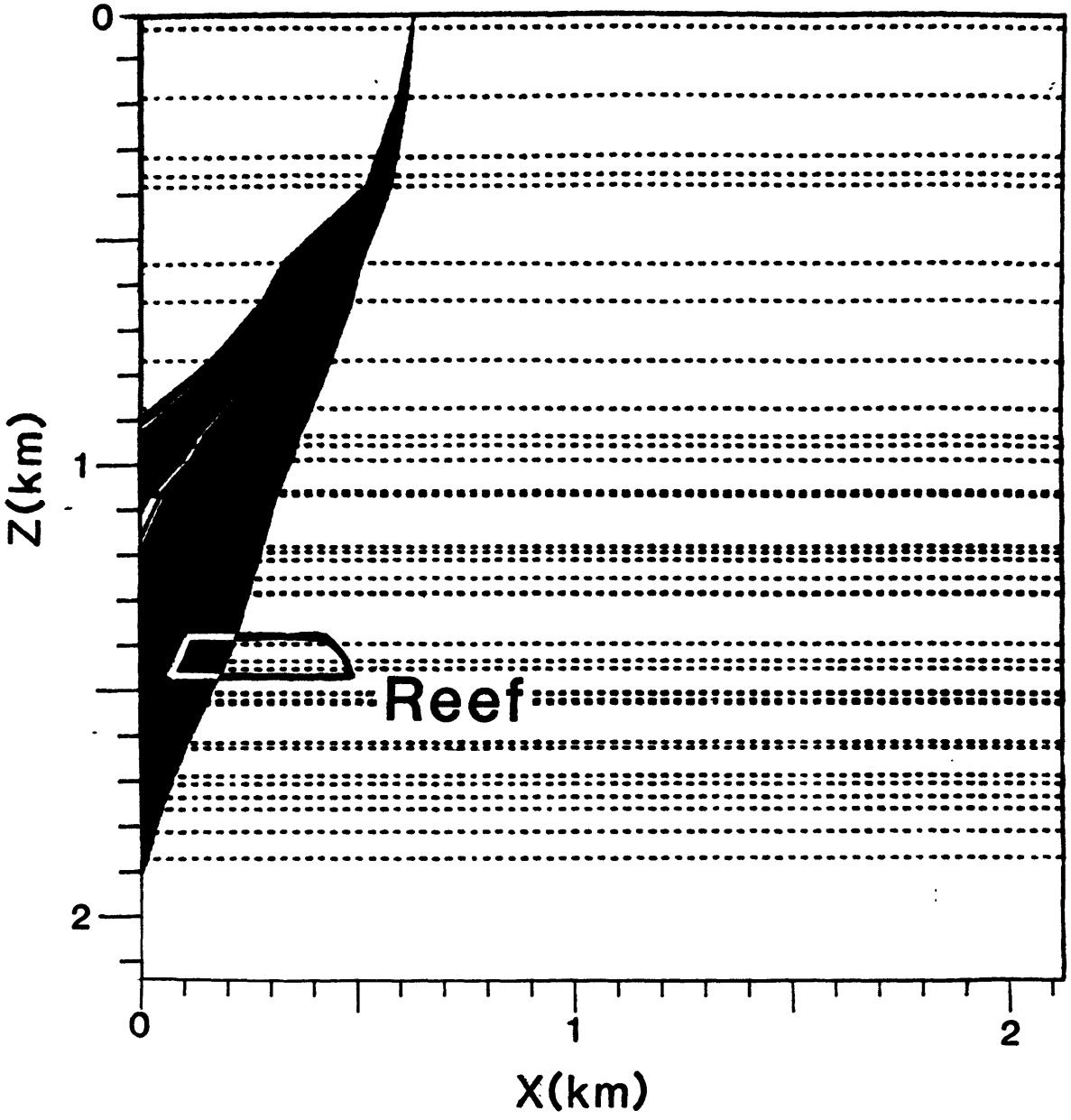


Figure 5.30

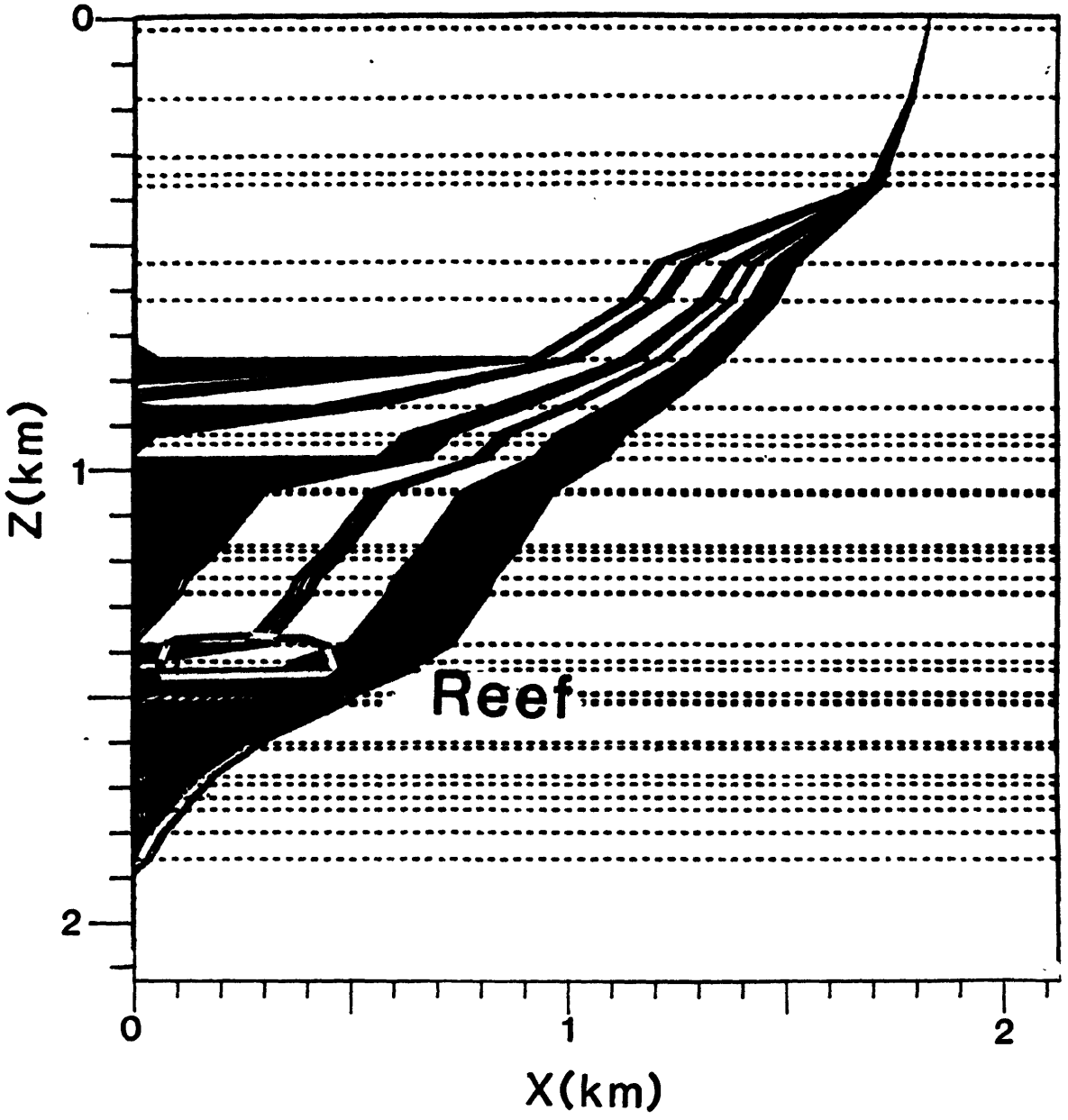


Figure 5.31

CHAPTER 6

Conclusions

The topic of travel time inversion applied to the vertical seismic profile has been explored before. In this work, an attempt was made to thoroughly understand many of the subtleties of the method in one dimension, especially in the extension to wide offset and multi-offset inversion. A Gauss' method style algorithm was developed to simultaneously invert multi-offset travel time data for a model with any sort of horizontal layering. Headwave refractions are considered and the elements of the partial derivative (Jacobian) matrix are found analytically rather than by finite differencing which results in a large time savings. A very general VSP geometry is assumed so that a deviated well or buried source, for example, do not present a problem. Further, the existing algorithm could be used to invert first arrivals from a transposed VSP.

Efforts have been made to explain and circumvent potential problems with, for example, one-dimensional ray tracing and the application of damping to Gauss' method. The nonlinearity of offset VSP travel time calculation has been investigated and shown to play a part in slowing iterative convergence. A simple layer stripping velocity inversion has been described which helps solve the sometimes formidable problem of choosing a starting point for Gauss' method.

Synthetic data sets were used to test the velocity inversion algorithm. The effects of source offset, over-parameterization, wrong parameterization, and random noise were studied and found to degrade the solution to the inverse problem only slightly except in the case of very high levels of noise. In that case, the resulting velocity model was found to oscillate about the true model. The effect of incorrect parameterization on the inversion is an averaging of the true model which, when used to calcu-

late travel times, produces a very good fit. This indicates that if the layering is not known *a priori*, the solution to the inverse problem is not unique. Well logs have been used to provide this *a priori* information.

Since accurate picking travel times from VSP records is an issue of key importance to a successful VSP inversion, methods for improving the pick were studied. Cubic spline interpolation was found to be a quite accurate and time-efficient method of increasing resolution and a good alternative to sampling theorem interpolation.

Three sets of field examples were investigated. While the first two tests were comparatively simple in the sense that the experiments had nearly zero source offset and the wells studied were shallow and situated in fairly simple structure, the results were good and demonstrated the utility of the method. The third field example involved multiple source offsets about a reasonably deep well. Besides yielding good results, it showed the potential for using observed departures from average structure to make at least qualitative statements about the three-dimensional structure.

During this work, a robust, versatile and possibly well-documented set of computer programs were written which may be of use in future VSP analysis. They will be used for routine VSP analysis and for such research endeavors as the design of optimal experiments.

References

Acton, F. S., *Numerical Methods That Work*, Harper and Row, New York, 1970.

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

Allen, R., Automatic phase pickers: their present use and future prospects, *Bull. Seis. Soc. Amer.*, *72*, S225-S242, 1982.

Balch, A. H., M. W. Lee, J. J. Miller, and R. T. Ryder, The use of vertical seismic profiles in seismic investigations of the Earth, *Geophysics*, *47*, 906-918, 1982.

Beck, J. V. and K. J. Arnold, *Parameter Estimation in Engineering and Science*, John Wiley and Sons, New York, 1977.

Beydoun, W. B., Asymptotic wave methods in heterogeneous media, Ph. D. Thesis, Mass. Inst. of Tech., Cambridge, MA, 1985.

Brent, R. P., *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

Brown, K. M. and J. E. Dennis, Derivative free analogues of the Levenberg-Marquardt and Gauss algorithm for nonlinear least squares approximations, *Numerische Mathematik*, *18*, 289-297, 1972.

Caughlin, W. G., F. J. Lucia, and N. L. McIver, The detection and development of Silurian reefs in northern Michigan, *Geophysics*, *41*, 646-658, 1976.

Compagnie Général de Géophysique, Experimental Group Shoot for the Study of Offset Vertical Seismic Profiling Applied to Reservoir Delineation, Final Report, C.G.G., Denver, CO, 1985.

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

deBoor, C., *A Practical Guide to Splines*, Applied Mathematical Sciences, 27, Springer-Verlag, New York, 1978.

Foulds, L. R., *Optimization Techniques*, Springer-Verlag, New York, 1981.

Gal'perin, E. I., *Vertical Seismic Profiling*, Society of Exploration Geophysicists, Tulsa, OK, 1974. (Originally published in Russian by Nedra, Moscow, 1971).

Geyer, R. L., *Catalog of Klauder Wavelets*, Seismograph Service Corporation, Tulsa, Oklahoma, 1972.

Grant, F. S. and G. F. West, *Interpretation Theory in Applied Geophysics*, McGraw-Hill, New York, 1965.

Hardage, B. A., *Vertical Seismic Profiling. Part A: Principles*, Geophysical Press, London, 1983.

Hoover, G. M., J. G. Gallagher, and H. K. Rigdon, Vibrator Signals, *Proc. I.E.E.E.*, *72*, 1290-1301, 1984.

Keho, T. H., M. N. Toksöz, C. H. Cheng, and R. M. Turpening, Wave Dynamics in a Gulf Coast VSP, in *Vertical Seismic Profiling, Part B: Advanced Concepts*, edited by M. N. Toksöz and R. R. Stewart, Geophysical Press, London, 1984.

Lee, M. W. and A. H. Balch, Computer processing of vertical seismic profile data, *Geophysics, 48*, 272-287, 1983.

Levenberg, K., A method for the solution of certain non-linear problems in least squares, *Quart. Appl. Math., 2*, 164-168, 1944.

Lines, L. R., A. Bourgeois, and J. D. Covey, Traveltime inversion of offset vertical seismic profiles – A feasibility study, *Geophysics, 49*, 250-264, 1984.

Luenberger, D. G., *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1973.

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

Oppenheim, A. V., A. S. Willsky, and I. T. Young, *Signals and Systems*, Signal Processing Series, Prentice Hall, Englewood Cliffs, NJ, 1983.

Pujol, J., R. Burridge, and S. B. Smithson, Velocity determination from offset vertical seismic profiling data, *J. Geophys. Res., 90*, 1871-1880, 1985.

Stewart, R. R., Vertical seismic profiling: The one-dimensional forward and inverse problems, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA., 1983.

Tarantola, A. and B. Valette, Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. of Geophys. and Space Physics, 20*, 219-232, 1982.

Telford, W. M., L. P. Geldart, R. E. Sheriff, and D. A. Keys, *Applied Geophysics*, Cambridge University Press, New York, 1976.

Toksöz, M. N. and R. R. Stewart, *Vertical Seismic Profiling, Part B: Advanced Concepts*, Geophysical Press, London, 1984.

Waters, K. H., *Reflection Seismology*, John Wiley and Sons, New York, 1978.

Wingo, J. R., Velocity and attenuation determinations from a vertical seismic profile, M. S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1981.

APPENDIX A

Formulation of Linearized Inversion

As stated in Chapter 3, the linearized least squares method seeks to minimize the sum of the squares of data residuals. This appendix gives an alternative derivation (which is more common though perhaps less rigorous) of the linearized inversion formulation for travel time inversion. The partial derivatives with respect to layer velocities are given for the direct and refracted rays.

In travel time inversion the travel time residual is defined by

$$r = t_{obs} - t_{calc} = \Delta t. \tag{A.1}$$

The travel time residual is expanded to first order in a Taylor expansion in the model parameters, in this case n layer velocities:

$$\Delta t = \sum_{j=1}^{n} \frac{\partial t}{\partial V_j} \Delta V_j \tag{A.2}$$

where the partial derivatives $\frac{\partial t}{\partial V_j}$ depend on the path taken.

The travel time equation for a direct wave from the surface to a receiver in the mth layer, equation (2.9), may be differentiated giving

$$\frac{\partial t}{\partial V_j} = \frac{-d}{V_j^2 \sqrt{1 - p^2 V_j^2}} \tag{A.3}$$

where

$$d = \begin{cases} h_j & \text{if } j < m \\ z - \sum_{k=1}^{m-1} h_k & \text{if } j = m \\ 0 & \text{if } j > m. \end{cases}$$

The partial derivative in equation (A.3) differs from that presented by Pujol et al. (1985). Their expression (in the present notation),

$$\frac{\partial t}{\partial V_j} = \frac{-d}{V_j^2 \sqrt{1 - p^2 V_j^2}} + \frac{p^2 d}{(1 - p^2 V_j^2)^{3/2}}, \tag{A.4}$$

contains a second term which results from differentiating the square root term $\sqrt{1 - p^2 V_j^2}$ as well. This is a higher order partial derivative, however, which is really a ray path perturbation since $\sqrt{1 - p^2 V_j^2} = \cos\vartheta_j$.

To see this, suppose that the velocity $V(z)$ is perturbed slightly to $V(z)+\delta V(z)$. The travel time between two points A and B on a ray, given by

$$t = \int_A^B \frac{ds}{V(z)}, \tag{A.5}$$

is affected in two ways. First, the velocity in the integrand of equation (A.5) becomes $V(z)+\delta V(z)$. The second effect is a ray path perturbation. Fermat's Principle, however, states that the travel time is stationary with respect to the ray path so that the contribution of the ray path perturbation to δt is not a first order effect (Aki and Richards, 1980). Only the velocity perturbation δV contributes. The perturbed travel time becomes

$$t+\delta t = \int_A^B \frac{ds}{V(z)+\delta V(z)}. \tag{A.6}$$

Subtracting equation (A.5) from this expression gives, to first order,

$$\delta t = -\int_A^B \frac{ds\,\delta V}{V^2}. \tag{A.7}$$

which corresponds to equation (A.3).

There are several cases to consider for refracted waves. For a ray from the surface which bottoms on the top of the rth layer and reaches a receiver in the mth layer, there are four different partial derivatives depending on the layer's position in the stack.

$$\frac{\partial t}{\partial V_j} = \frac{-h_j}{V_r^2\sqrt{1-\frac{V_j^2}{V_r^2}}} - \frac{h_j\sqrt{1-\frac{V_j^2}{V_r^2}}}{V_j^2}, \quad j < m \tag{A.8}$$

$$\frac{\partial t}{\partial V_j} = \frac{d-2h_j}{V_r^2\sqrt{1-\frac{V_j^2}{V_r^2}}} - \frac{(2h_j-d)\sqrt{1-\frac{V_j^2}{V_r^2}}}{V_j^2}, \quad j = m \tag{A.9}$$

$$\frac{\partial t}{\partial V_j} = \frac{-2h_j}{V_r^2\sqrt{1-\frac{V_j^2}{V_r^2}}} - \frac{2h_j\sqrt{1-\frac{V_j^2}{V_r^2}}}{V_j^2}, \quad m < j < r \tag{A.10}$$

$$\frac{\partial t}{\partial V_j} = \sum_{i=1}^{j-1}\frac{h_i V_i}{V_j^3\sqrt{(1-\frac{V_i^2}{V_j^2})}} + \sum_{i=m+1}^{j-1}\frac{h_i V_i}{V_j^3\sqrt{(1-\frac{V_i^2}{V_j^2})}}$$

$$+ \frac{(h_m-d)V_m}{V_j^3}\sqrt{1-\frac{V_m^2}{V_j^2}} - \frac{x}{V_r^2}, \quad j = r \tag{A.11}$$

If there are l observations and an n layer velocity model, a linearized system of l equations in n unknowns can be formulated. It follows that

$$\begin{bmatrix} \frac{\partial t_1}{\partial V_1} & \frac{\partial t_1}{\partial V_2} & \cdots & \frac{\partial t_1}{\partial V_n} \\ \frac{\partial t_2}{\partial V_1} & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial t_l}{\partial V_1} & \frac{\partial t_l}{\partial V_2} & \cdots & \frac{\partial t_l}{\partial V_n} \end{bmatrix} \begin{bmatrix} \Delta V_1 \\ \Delta V_2 \\ \vdots \\ \vdots \\ \Delta V_n \end{bmatrix} = \begin{bmatrix} \Delta t_1 \\ \Delta t_2 \\ \vdots \\ \vdots \\ \Delta t_l \end{bmatrix} \tag{A.12}$$

which, following the notation of Chapter 3, may be written as

$$\mathbf{G}(\mathbf{m}_o)(\hat{\mathbf{m}} - \mathbf{m}_o) = \mathbf{d} - \mathbf{g}(\mathbf{m}_o) \tag{A.13}$$

where $\hat{\mathbf{m}}$ is the true velocity model,

$\mathbf{m}_o$ is an estimate of $\hat{\mathbf{m}}$,

$\mathbf{d}$ is the observed travel time vector,

and $\mathbf{g}(\mathbf{m}_o)$ is the calculated travel time vector.

The system is similar to the usual least squares problem

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

which is solved by the steps

$$\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$$

$$\mathbf{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

This result is akin to equation (3.12):

$$\mathbf{m}_{k+1} = \mathbf{m}_k + [\mathbf{G}^T(\mathbf{m}_k)\mathbf{G}(\mathbf{m}_k)]^{-1}\mathbf{G}^T(\mathbf{m}_k)[\mathbf{d} - \mathbf{g}(\mathbf{m}_k)] \tag{A.14}$$

Note that in most cases equation (A.13) represents an overdetermined system so that the step analogous to going from $\mathbf{A}\mathbf{x} = \mathbf{b}$ to $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ represents great space savings in computer memory as the full sensitivity matrix $\mathbf{G}$ is never actually used. Since $\mathbf{G}^T\mathbf{G}$ is symmetric, the storage requirements are reduced from $l \cdot n$ to $n \cdot (n + 1)/2$. All of the algorithmic variations presented in Appendix B also have this property.

APPENDIX B

Comparison of Inversion Methods

B.1. Introduction

In order to choose the best inversion scheme for this work, a number of different methods were pitted against each other in trial inversions of synthetic travel time data. This appendix introduces the contestants which are variations on Gauss' method involving three aspects of the inversion: damping, model perturbation, and partial derivative calculation.

B.2. Scaled Levenberg Damping

An example of the first kind of variation is one of the methods suggested by Beck and Arnold (1977). It is a modification of Levenberg's (1944) method which scales the amount of damping applied to $\mathbf{G}^T\mathbf{G}$ in equation (3.13) by choosing the value of Ω to be not necessarily equal to $\mathbf{I}$. Instead Ω is set equal to the diagonal elements of $\mathbf{G}^T\mathbf{G}$. They choose the damping parameter λ to be

$$\lambda = \frac{\mathbf{r}^T\mathbf{G}\Omega\mathbf{G}^T\mathbf{r}^T}{S} \tag{B.1}$$

where $\mathbf{r}$ is $\mathbf{d} - \mathbf{g}(\mathbf{m}_o)$
and S is the sum of squares of the data error or $\mathbf{r}^T\mathbf{r}$

This formulation has the additional benefit of making the results invariant under scale changes in the parameters, something that cannot be said about the $\Omega = \mathbf{I}$ approach. This turns out not to be of great consequence in this particular problem since all of the parameters are velocities.

B.3. Modified Brown and Dennis Damping

Another updating scheme for the damping parameter is a modification of a method due to Brown and Dennis (1972). The method, which is unscaled, sets $\Omega = \mathbf{I}$ and decreases λ in a very simple manner as the solution is neared. Far from the solution, the correction follows the method of steepest descent by choosing values of λ so large as to swamp the Gauss' method contribution. As the solution is approached, damping is no longer required to maintain stability and the value of λ is reduced accordingly. A small improvement on their method was implemented to achieve speedier final convergence in this study. The modified Brown and Dennis method calls for

$$\lambda = c \, \| \mathbf{r}(\mathbf{m}_0) \|_\infty \tag{B.2}$$

with

$$c = \begin{cases} 10^{-7} & \| \mathbf{r}(\mathbf{m}_0) \|_\infty < 5\times10^{-5} \\ 10^{-4} & 5\times10^{-5} \le \| \mathbf{r}(\mathbf{m}_0) \|_\infty < 5\times10^{-3} \\ 10^{-2} & 5\times10^{-3} \le \| \mathbf{r}(\mathbf{m}_0) \|_\infty < 1 \\ 1 & 1 \le \| \mathbf{r}(\mathbf{m}_0) \|_\infty < 10 \\ 10 & 10 \le \| \mathbf{r}(\mathbf{m}_0) \|_\infty \end{cases}$$

where $\| \mathbf{r}(\mathbf{m}_0) \|_\infty$ (called the infinite order norm) is simply the largest travel time residual. This method was the most successful when a large number of parameters were involved.

B.4. The Modified Box-Kanemasu Method

The second type of variation on Gauss' scheme involves adjusting the model parameters. Gauss' method advocates simply adding the perturbations to the previous model as in equation (3.12). A different way of adjusting the model perturbations is presented by Beck and Arnold (1977) as a modification of the Box-Kanemasu interpolation method. The algorithm seeks to minimize the sum of squares of the travel time

residuals ($S \equiv \mathbf{r}^T\mathbf{r}$) by approximating S at each iteration by a parabola in error space and trying to adjust the model just far enough to minimize S. If equation (3.12) is rewritten as two equations,

$$\Delta\mathbf{m}_k = (\mathbf{G}(\mathbf{m}_k)^T\mathbf{G}(\mathbf{m}_k))^{-1}\mathbf{G}(\mathbf{m}_k)^T[\mathbf{d} - \mathbf{g}(\mathbf{m}_k)] \tag{B.3}$$

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta\mathbf{m}_k \tag{B.4}$$

then the interpolation is achieved by introducing a scalar factor h_{k+1} into equation (B.4):

$$\mathbf{m}_{k+1} = \mathbf{m}_k + h_{k+1}\Delta\mathbf{m}_k. \tag{B.5}$$

Finding the optimum h_{k+1} involves fitting the parabola to S, minimizing S in that context, and then using the appropriate value of h_{k+1} to scale $\Delta\mathbf{m}_k$ in equation (B.5). Beck and Arnold add steps to insure that the improvement $h_{k+1}\Delta\mathbf{m}_k$ actually will reduce S instead of possibly allowing it to grow. This method is remarkably stable but also very slow since the value of S must be calculated twice instead of once on every iteration as part of solving for h_{k+1}. This requires more of the time-consuming ray tracing and associated root-finding discussed in Chapter 2. The modified Box-Kanemasu method is also slow to converge.

B.5. Finite Difference Levenberg-Marquardt Algorithm

The last departure from the straight Gauss' method formulation involves using finite difference approximations to the elements of the partial derivative matrix $\mathbf{G}(\mathbf{m}_k)$ (Brown and Dennis, 1972; Stewart, 1983). This method is easy to implement since there exists a finite difference Levenberg-Marquardt subroutine in the IMSL package. It is not, however, recommended if it is convenient to calculate the values of the partial derivative matrix analytically. This is especially the case in travel time inversion where calculating travel times is usually a great deal more time consuming than calculating partial derivatives. The partial derivative calculation typically involves only

one or two terms and no root-finding. In the test runs, the finite difference algorithm typically took more than ten times as much CPU time as other methods. It was, however, very stable since the damping parameter in this particular package is modified by the Brown and Dennis (1972) method which was discussed above.

APPENDIX C

The Data Variance Surface

Contoured plots of data variance are full of interesting information. Their major drawback is that they are limited to only two dimensions. These two dimensions will be sufficient for the following discussion which can be generalized to higher dimensions. The units used in this appendix are *cubits* in flagrant violation of the SEG guidelines.

Figure C.1 shows data variance plots for four geometries which differ only in their source offsets. The true velocities are V_1=5, V_2=9. Several observations can be made based on these plots.

As the source offset is increased, the error surface begins to look less and less like the quadratic error surface characteristic of the linear least squares problem. This effect is due to the larger and larger deviations from straight rays that are observed as source offset increases. If the ray paths were straight, velocity inversion would be a linear inverse problem. An additional source of nonlinearity at larger offsets is the presence of headwave refractions as first arrivals.

Another effect of increasing source offset is that the interdependence between the two model parameters increases and the trend of the error contours turn from vertical toward diagonal. The error analysis taken up in section 3.5, because it assumes model parameter independence, is only strictly valid when the trend of the contours (or, equivalently, the eigenvectors of the linearized system) are either vertical or horizontal. The reason these contours drift away from vertical is that the fastest rays are more often headwave refractions as the source offset becomes larger. With the possibility of headwave refractions comes a trade-off between V_1 and V_2. Since there can be no critical refractions at zero offset (Fig. C.1c) the best velocity for

Since there can be no critical refractions at zero offset (Fig. C.1c) the best velocity for V_1 is 5 no matter what velocity is used for V_2. The velocities are nearly independent. At the largest source offset (Fig. C.1f), the velocities are completely dependent on each other.

A third observation can be made from the curves shown in Figure C.1. The innermost contour is $\sigma_d^2 = 5\times10^{-4}$ for all four plots yet the contours become tighter as the source offset increases indicating that a given poor fit generally has less data variance (or a more gently dipping error surface) for smaller source offsets. This observation is not as important as the previous two since it is just a scaling effect. If larger offsets are used, the overall path lengths and travel times increase so that magnitudes of the travel time errors increase while the relative size of the error remains the same.

The rotation of the error surfaces seen in Figure C.1 suggests that the simultaneous use of multi-offset VSP data might provide more stability to the problem. Stability problems in least squares inversion occur as $\mathbf{G}^T\mathbf{G}$ (see Chapter 3) becomes singular which happens when one or more of its eigenvalues nears zero. This is illustrated by the data variance plots. The eigenvectors corresponding to the smaller eigenvalue is directed along the long axis of the data error contour. As the smaller eigenvalue decreases, the ellipse lengthens. An interpretation of this is that the Jacobian matrix exerts less control over variations of the model parameters in the direction corresponding to this eigenvalue. The large eigenvalue eigenvector, on the other hand, is better behaved. Variations in its direction are minimal. The rotation of the contours (or eigenvectors) in Figure C.1 shows that at each the various source offsets, the travel time data is best capable of resolving a particular direction in (V_1, V_2) space. Multi-offset data, therefore, would be capable of pinning down velocity variation from many sides. Figure C.2 shows this to be true. While $\lambda_{max}/\lambda_{min}$ varies from approximately 6 to 11 in the four experiments taken separately, $\lambda_{max}/\lambda_{min}$ is just 3 when they

are treated together.

Figure Captions

Figure C.1. Contoured data variance for a simple two-layer model at four offsets. a) Velocity model. b) Travel time versus depth curves for the four source offsets. c) Data variance for zero offset. d) Short offset (2000). e) Wide offset (4000). f) Very wide offset (6000).

Figure C.2. Contoured data variance plot for four-source VSP experiment. The source offsets used vary from zero to 6000.

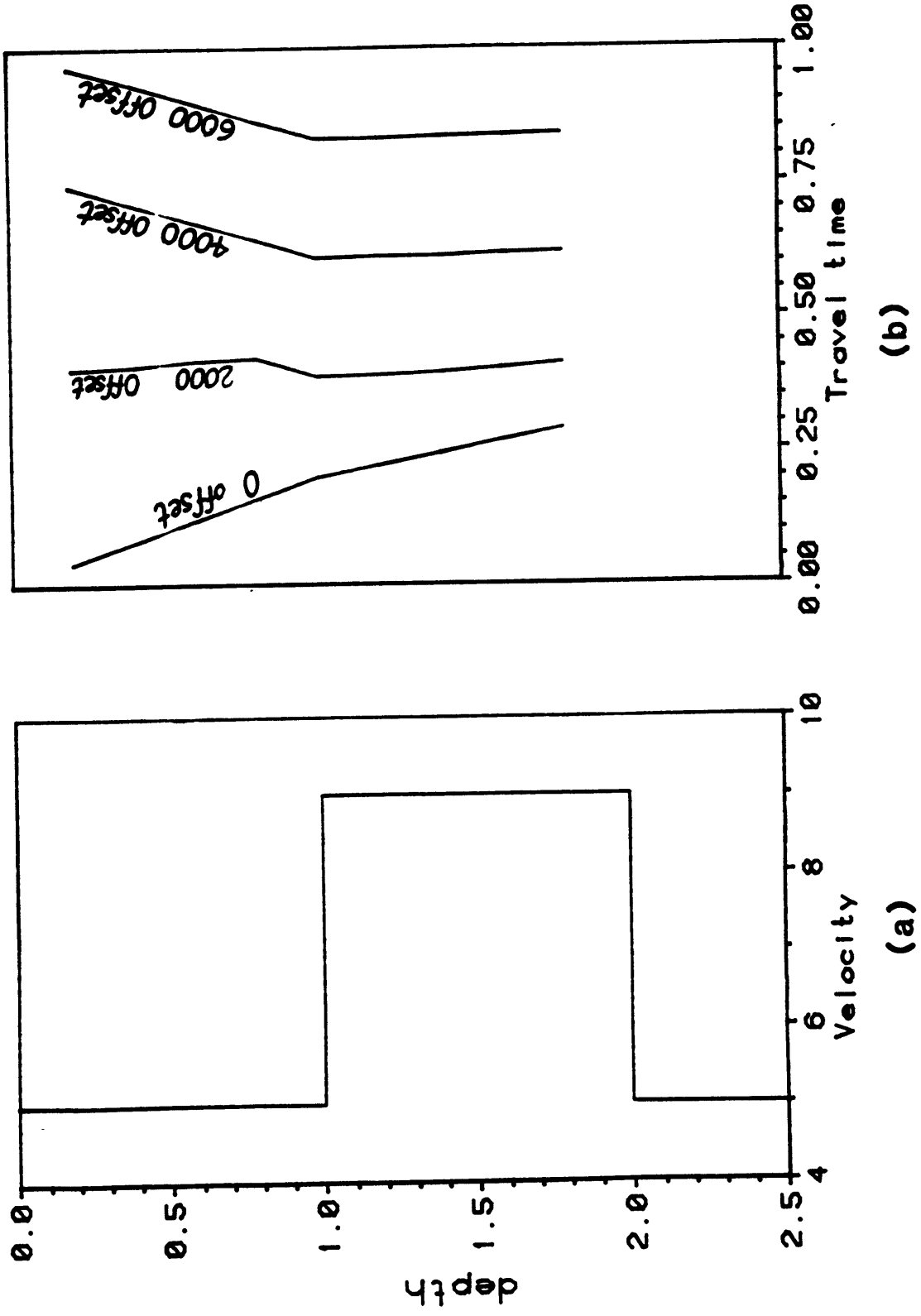

6000 offset
4000 offset
2000 offset
0 offset
0.00
0.25
0.50
0.75
1.00
Travel time
(b)
0.0
0.5
1.0
1.5
2.0
2.5
depth
4
6
8
10
Velocity
(a)

Figure C.1

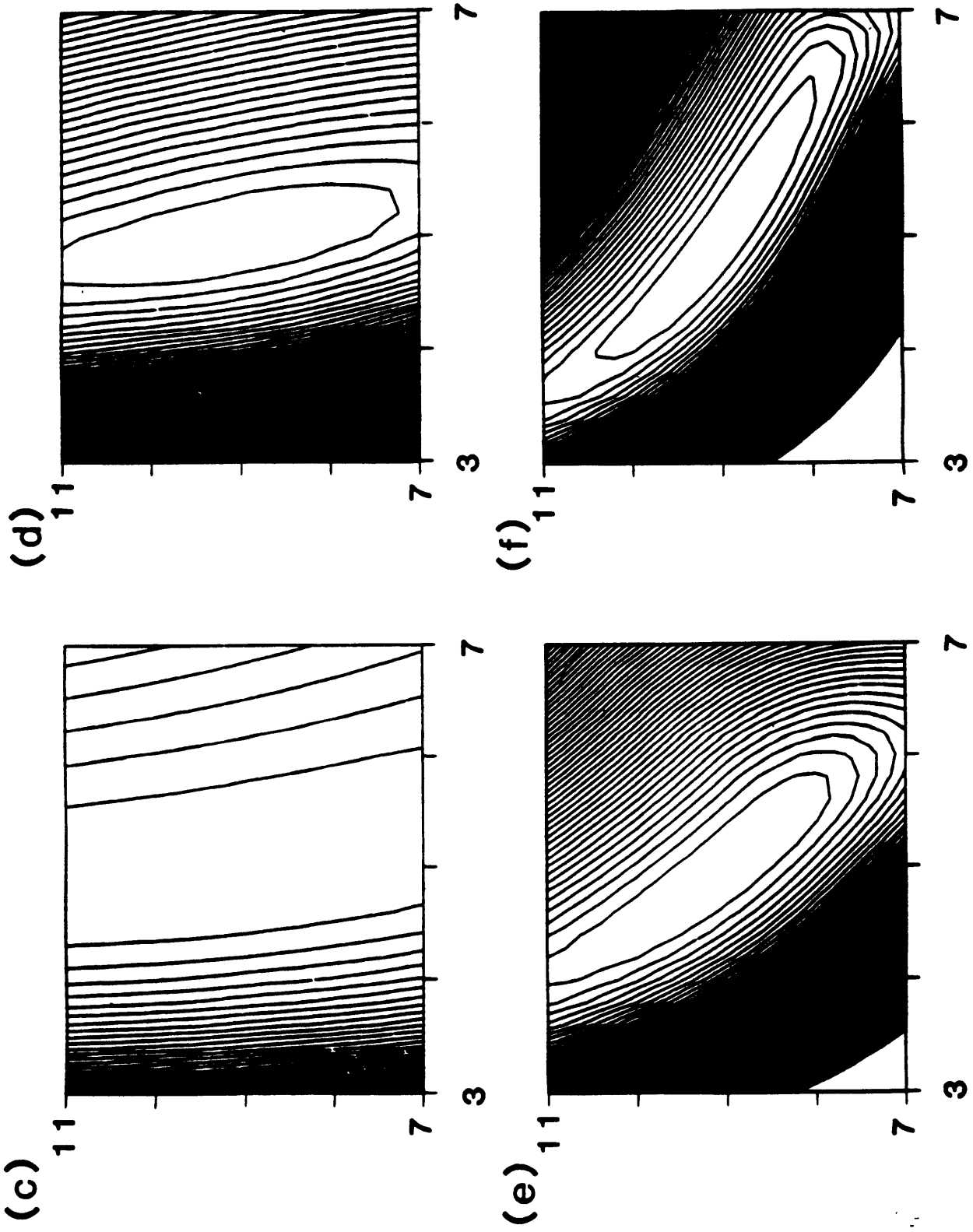


Figure C.1 (continued)

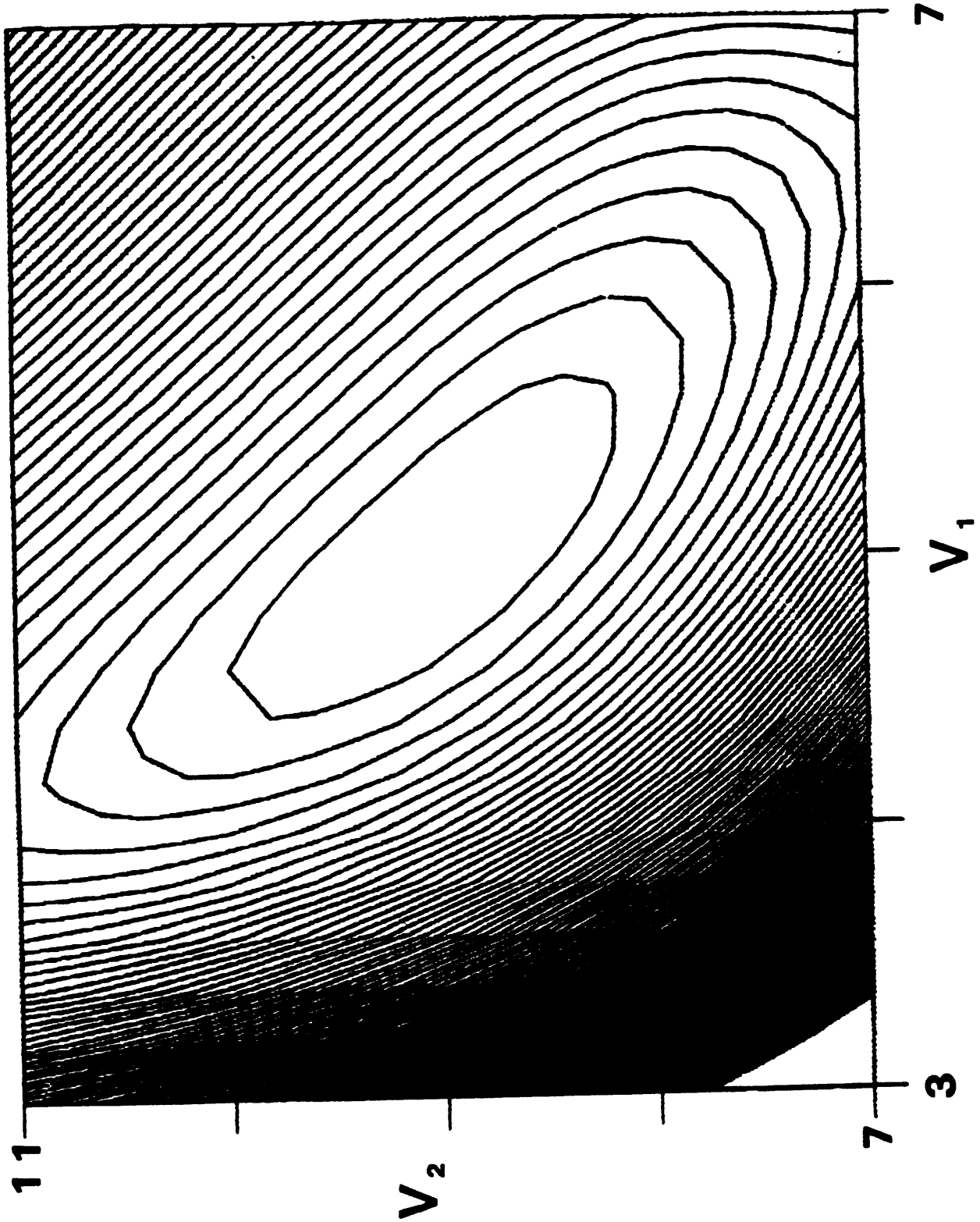


Figure C.2

APPENDIX D

Cubic Spline Interpolation of Sampled Data

This appendix examines the validity of using cubic spline interpolation instead of sampling theorem (also known as "sinc function") interpolation to augment sampled continuous time data.

A seismogram is composed of discrete samples of a continuous time velocity function. The analog signal is low-pass filtered before the A/D conversion so that it is bandlimited to (less than) half of the sampling frequency. The results of the sampling theorem (see equation 4.4) can be used to reconstruct the lowpass filtered continuous time signal as outlined in Chapter 4. Practical implementation of sampling theorem interpolation, however, is computationally slow and has additional problems which are also discussed in Chapter 4. Cubic spline interpolation, on the other hand, is very fast and, as will be shown, quite accurate.

Four kinds of interpolating functions will be discussed: the zero-order hold, the first-order hold (linear interpolation), the cubic spline, and sampling theorem interpolation. In addition, two ways of thinking about interpolation will be presented.

The time domain version involves convolving a discrete time sequence representing the sampled data which has been converted to an impulse train with various impulse responses. Figure D.1 shows the impulse responses for the four algorithms listed above. As can be seen from the waveforms, three of the four impulse responses are symmetric about the time origin which means that they are zero-phase filters. This is a desirable characteristic for interpolation filters to be used in travel time picking since phase distortion will seriously alter time picks. The zero-order hold (Fig. D.1a) is therefore a bad choice. Linear interpolation (Fig. D.1b) might be thought of as

the default scheme since it is what an analyst sees if the sample points are merely connected with straight lines by a plotting program.

Another way of visualizing interpolation is to imagine inserting one or more zeros between discrete time sequence values and then lowpass filtering (Oppenheim et al., 1983). The magnitudes of the frequency responses are shown in Figure D.2 for the four interpolation filters. The ideal case, sampling theorem interpolation, would be implemented with an ideal lowpass filter (Fig. D.2d). Cubic spline interpolation (Fig. D.2c) is clearly the second best choice as far as the lower frequencies are concerened. The relatively high stop-band gain is worse than linear interpolation but may not be a problem for data which are sampled at a higher than necessary rate or lowpass filtered prior to digitization. The predominant frequency content of a VSP seismogram is near 50 Hz (Fig. D.3) which is well below the 250 Hz Nyquist rate for the data set used in this study.

Although cubic spline interpolation is not bandlimited (see Fig. D.2c), it provides a very reasonable substitute for perfect bandlimited interpolation since the resulting waveform has no phase distortion, appears more continuous than linear interpolation, and has good amplitude response.

Figure Captions

Figure D.1. Impulse responses of several interpolation functions. (a) zero-order hold (b) first-order hold (c) cubic spline (d) sampling theorem.

Figure D.2. Magnitudes of frequency responses for the same interpolating functions as in Figure D.2. Functions (b), (c), and (d) are all zero-phase; (a) is linear phase.

Figure D.3. Magnitude of the Fourier transform of a typical VSP seismogram used in this study. The peak is near 50 Hz.

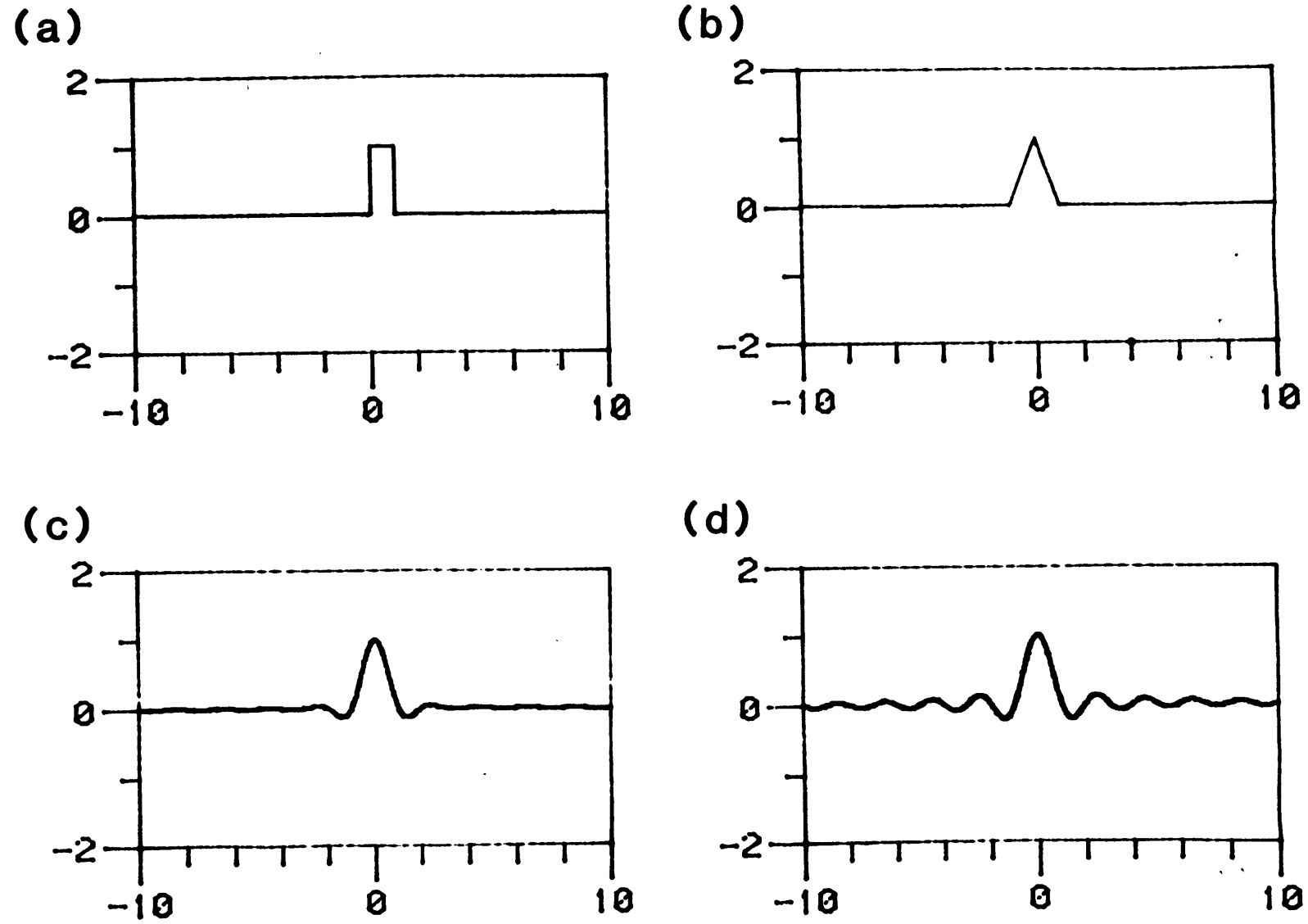


Figure D.1

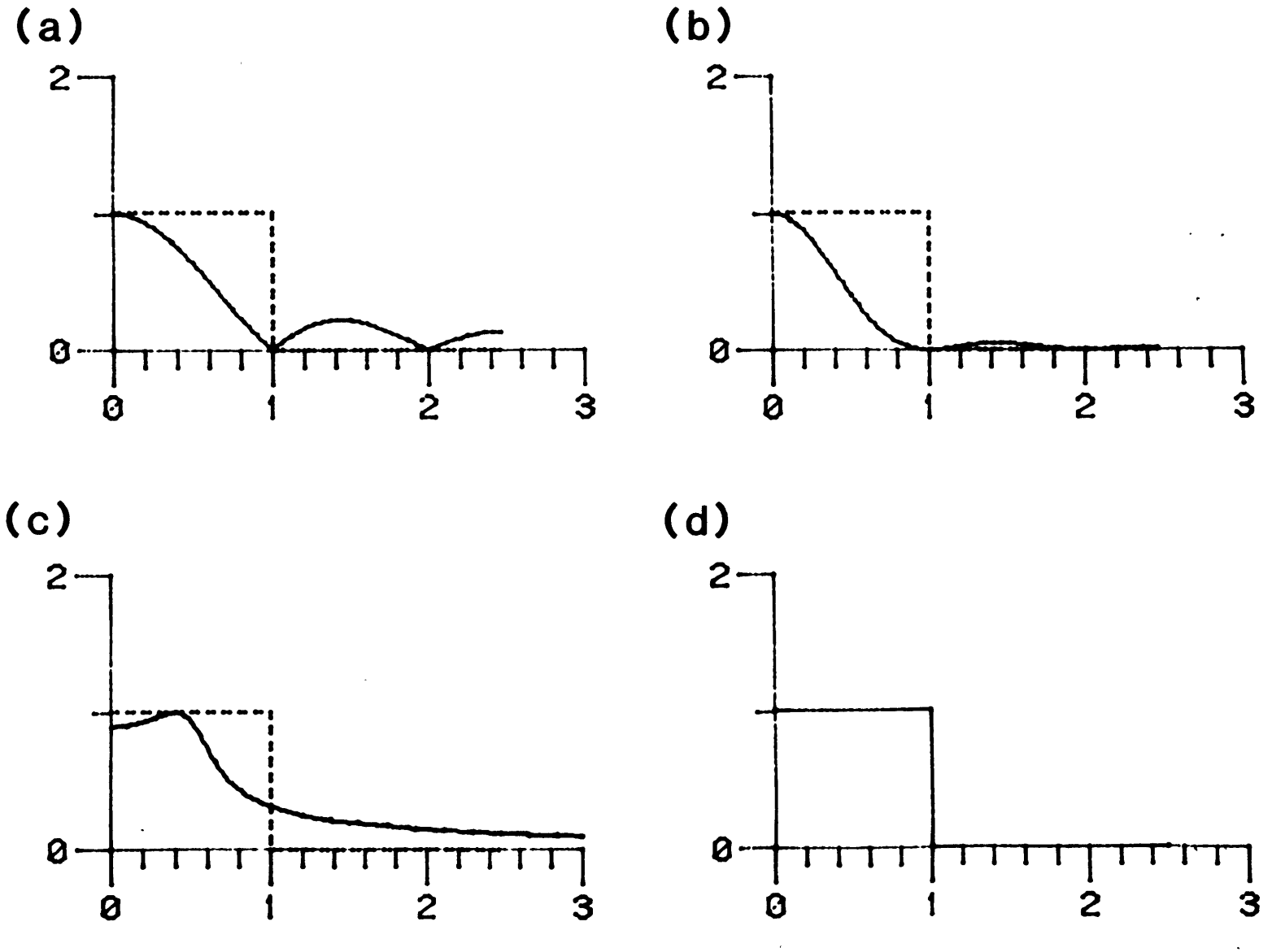


Figure D.2

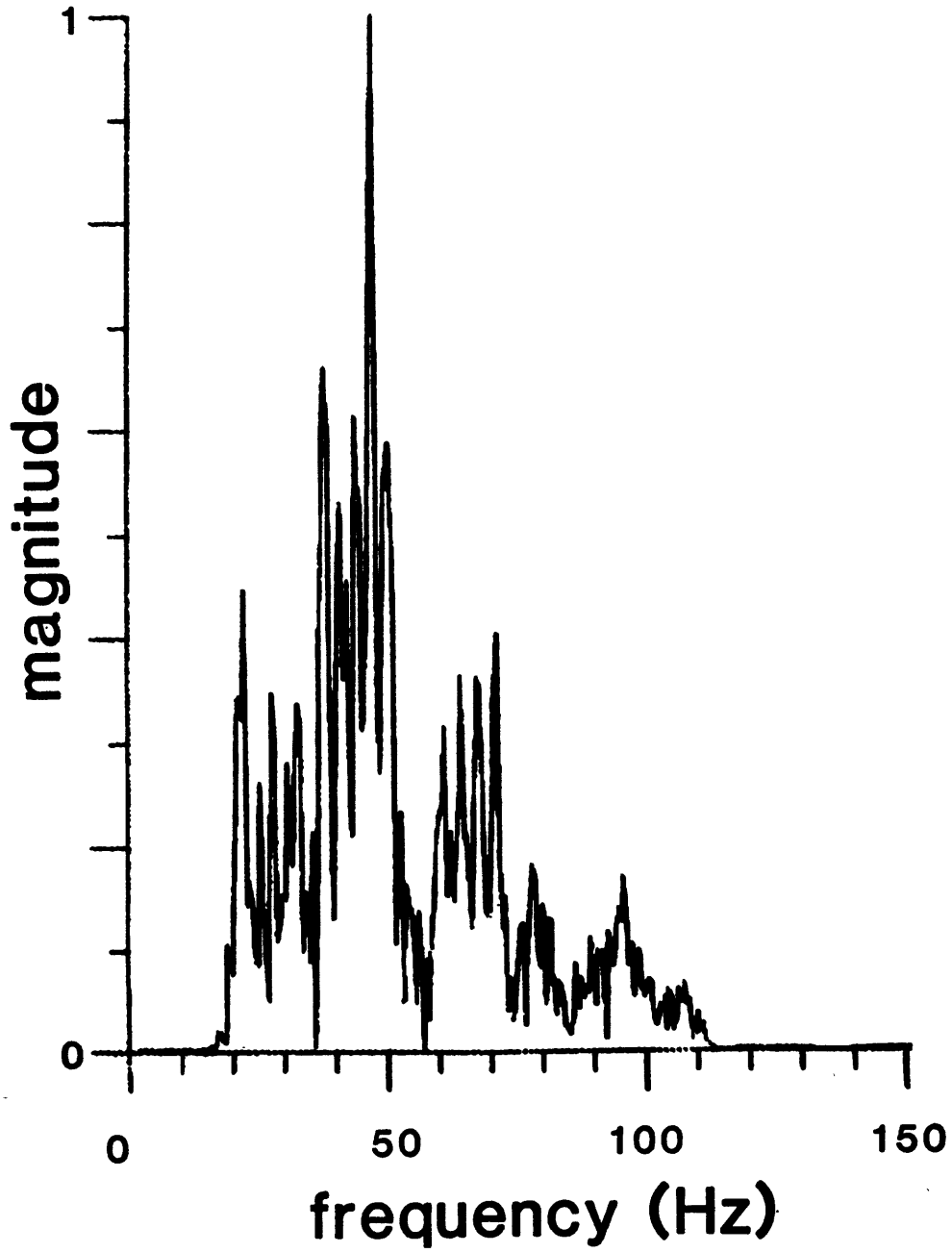


Figure D.3