ATTENUATION ESTIMATION FROM ACOUSTIC WELL LOGS

By

STEVEN WAYNE PATTON

Bachelor of Science in
Mineral Engineering - Physics
Colorado School of Mines
Golden, Colorado
May, 1980

Master of Science in
Electrical Engineering
Stanford University
Stanford, California
June, 1981

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
May, 1986

Thesis
[illegible]
[illegible]
[illegible]

ATTENUATION ESTIMATION FROM

ACOUSTIC WELL LOGS

Thesis Approved

Thesis Advisor

Dean of the Graduate College

AKNOWLEDGEMENT

I will always be grateful to my adviser, Dr. Rao Yarlagadda, for his advice and constant support while I studied under him at Oklahoma State University. Working with him convinced me that I will never finish learning. His patience with me while I ran past many deadlines for reports and this thesis was truely amazing. I also wish to thank the rest of my graduate committee, Dr Marvin Keener, Dr David Soldan and Dr. Randall Reininger for their many helpful suggestions.

This research was supported by the members of the Oklahoma State University Research Consortium for Enhancement of Well Log Data Via Signal Processing The members of this consortium include Amoco Production Company, Arco Oil and Gar Company, Cities Service Oil and Gas Corporation, Conoco, Exxon, IDM, Mobil Research and Development, Phillips Petroleum Corporation, Sohio Petroleum Company, and Texaco. In particular, I would like to thank Dr Robert Baumel of Conoco for a great deal of advice regarding acoustic logs, and for his help in supplying the model and real data used in this study.

I would like to thank Dr. Keith Teague and Sherry Mayes for allowing me to stay in their home many nights during the preparation of this thesis Kelly Whitfield went above and beyond the call of duty in typing the manuscript Most importantly, I would like to thank my wife, Laurie, without her love and support I would never have finished.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The tool used in full waveform acoustic or sonic logging is a long narrow cylindrical object, with cylindrically symmetric source (transducer) near the top of the tool, and with 2 or more evenly spaced receivers offset some distance from the source. The receivers are also cylindrically symmetric electro-acoustic transducers. As this tool is pulled up a borehole, the source periodically emits an acoustic pulse, and the borehole response at each of the receivers is sampled and recorded.

As the acoustic pulse propagates from the source to the receivers, it is attenuated. This means the wave decays as it travels through the subsurface, but in such a manner that the higher frequencies decay more rapidly than the lower frequencies. Attenuation of acoustic energy in the earth is known to occur over a broad range of frequencies, from earthquake frequencies (about 1 Hz) through the sonic logging range (10-25 kHz) on up through the ultrasonic (MHz) range. The traveling acoustic wave also decays due to other factors, including geometrical spreading losses, scattering and reflections. These other loss factors are considered to be separate from true intrinsic attenuation. However, it may not prove to be possible to measure these losses separately from intrinsic attenuation.

The most popular attenuation model uses exponential amplitude decay where the exponent is linear in frequency. The constant coefficient in the exponent can be written in terms of Q^{-1}, where Q is the "quality factor", commonly used to describe any oscillating system. For a general oscillating system

$$Q = \frac{2\pi E}{\Delta E} \tag{1}$$

where E is the peak energy stored in the oscillator and ΔE is the energy loss per oscillation. If the oscillator is "perfect" (no energy loss), then ΔE is zero and Q is infinite. Alternatively, the more lossy the medium the lower the quality factor is.

For a propagating acoustic wave in an elastic medium, there are actually two quality factors describing two losses. The directly measureable attenuation is the loss per wavelength of propagation (spatial attenuation). This is distinct from the loss per stationary oscillation (temporal attenuation) which can't be measured directly from a travelling wave. Indeed, Knopoff (1964) has established the following relationship

$$u\, Q_{temp} = c\, Q_{spatial} \tag{2}$$

where u = group velocity, c = phase velocity, Q_{temp} = temporal Q, and $Q_{spatial}$ = spatial Q. If the medium is dispersive, that is if the group velocity and phase velocities differ, then the two Q values should also differ. However, according to Willis (1983), Q_{temp}, sometimes referred to as intrinsic Q, is approximately equivalent to Q spatial (measured with wave propagation techniques) when attenuation losses are fairly small ($Q > 10$). Fortunately, estimates of Q for rock are nearly always

greater than 10. Typical values of Q are in the range 50 to 200. Hamilton (1972) concluded that dispersion for the P_{wave} is not significant in marine sediments, so Q_{temp} and $Q_{spatial}$ are essentially the same. Hence this thesis shall assume no dispersion.

Many different types of waves propagate down the borehole corresponding to different modes of propagation. The one of interest here is the P-wave (pressure or compressional wave). This wave moves from the source to the receiver through the fluid as a compressional wave, then is refracted to a P-wave in the rock and travels down through thr rock near the borehole. Some of this energy is refracted back into the fluid as another P-wave. Part of the compressional wave energy in the fluid is converted to a refracted shear wave at the borehole wall. This wave travels down the borehole at the shear velocity, and is also converted back into the fluid as a P-wave. These two modes, the compressional and shear (also known as "body-waves") are both considered non-dispersive.

There is another set of acoustic modes in the borehole. Waves of this type are known as "tube-waves" and include the Stonely and pseudo-Rayleigh waves. While these tube-waves are very dispersive, meaning that phase velocity varies with frequency, their velocities asymptotically approach the fluid velocity as frequency increases. These waves greatly complicate attempts to analyze shear wave behavior since the fluid compressional velocity and rock shear velocity are about the same. In fact, if the shear wave velocity is less than the fluid velocity, then the refracted shear wave can't exist. But if the shear velocity is slightly higher than the fluid velocity, the shear arrival will roughly coincide with the arrival of the larger amplitude tube-waves,

and the shear wave arrival is masked. The Q values for tube waves are a function of the Q values of the P-wave, shear wave, and fluid P-wave. So, calculating useful Q values (P-wave or shear wave) is difficult (see, for example Cheng, Toksoz, Willis, 1981).

Therefore, since the P-wave can usually be extracted from the total received waveform by windowing, only the P-wave will be considered. For the remainder of this thesis, it will be assumed that the "data" represents a windowed version of the P-wave only. In this research, a Hamming window is used and it gives good results. However, the attenuation estimates are not very sensitive to window choice.

Given a monochromatic plane-wave (frequency ω) traveling a distance Z in a constant Q medium, the amplitude coefficient of the wave function is

$$A(Z) = A_0 \exp(-\omega Z/2Qc) \tag{3}$$

where ω = angular frequency (radians/second), A_0 = initial amplitude (at Z = 0), c = phase velocity, and Z = distance travelled. Assuming Q to be frequency independent, then by Equation (3), for every wavelength travelled, the wave amplitude decreases by the same fraction. This means that if one observes a plane wave with a broad spectrum propagating in the Z direction, then at any one particular point Z_0, the attenuation undergone by the wave is an exponential function of frequency. The following section presents a discussion on attenuation by the earth and reasons for investigating this process.

1.1 Attenuation by the Earth

Many researchers agree that Q is not dependent on frequency for dry rocks, and there is a large body of laboratory data to support this. This includes Nur and Winkler (1980), and Toksoz et al (1979). However, most of these attenuation measurements have been conducted in the ultrasonic range (0.1 to 1.0 MHz), and it is not obvious that these results can be extrapolated down to sonic log frequencies (10 to 23 kHz) or further to seismic frequncies (10 to 100 Hz).

There has been a great deal of effort spent to study the effects of fluid and gas saturation, pressure, and wave amplitude on attenuation. All of the variables can have a significant effect on Q, and there is some evidence to indicate that Q is much more dependent upon frequency when the rock is partially or totally fluid saturated. The conclusions reached by Nur and Winkler (1980), as well by Tittman, Nadler, Clark, et al (1981) show a strong dependence of Q both on frequency and pressure for a water-saturated Wingate sandstone. The conclusions to be drawn from the available laboratory data seems to be that there is evidence both for and against Q being frequency dependent.

Field measurements of the variation of Q with frequency are primarily limited to seismic frequencies and there have been fewer field experiments than high frequency lab experiments. The classic experiments for seismic frequencies include those by McDonal et al (1958), Tullos and Reid (1969), and Hamilton (1972). These experiments were primarily vertical seismic profile (VSP) experiments consisting of recording the arrivals, due to seismic energy sources at the surface, at various geophones located at different depths in a borehole. They all concluded that Q is approximately independent of frequency. However,

differences do exist in attenuation coefficients for different subsurface layers, as subsurface layers consist of different rock types, at different pressures, and at different degrees of fluid saturation.

The mechanisms of attenuation are not well understood and the models used to explain attenuation tend to be physically complex. These mechanisms are summarized by Johnston and Toksoz (1981). These include frictional dissipation due to movement of grain boundaries (Walsh, 1966), fluid flow (Walsh, 1968), relative motion of frame due to fluid inclusions (Stoll and Bryan, (1970), "squirting" (Mavko and Nur, 1975), gas pockets (White, 1975), and geometrical effects (Kuster and Toksoz, 1974). From the point of view of acoustic wave propagation along a borehole, it is likely that a combination of many of the above mechanisms contribute to attenuation.

The earth can affect a seismic wave in an attenuative manner without true intrinsic attenuation. That is, the affect of intrabed multiples can appear to be that of a frequency filter with the same general shape one would expect from a constant Q attenuation model. The works of Schoenberger and Levin (1974), and O'Doherty and Ansty (1971) lead to the conclusion that the "tuning" effect of intrabed multiples may account for 1/3 to 1/2 of the attention in seismic data, and that the frequency filtering done by intrabed multiples can be modeled as an attenuative phenomenon. Scattering and geometrical spreading may also appear to be attenuative, and the borehole geometry may lead to frequency dependent tuning effects. Therefore, it may prove impossible to completely separate intrinsic attenuation from other "attenuative" phenomenon, and so attenuation estimates from sonic log data may represent an effective Q, rather than a true intrinsic Q.

However, many investigators including Willis (1983) and Anderson and Castagna (1984) concluded that geometrical spreading can be assumed to be frequency independent. In fact, Anderson and Castagna (1984) state that the theoretical geometrical spreading formula for a point source (on the borehole axis) is

$$\frac{1}{Z \log^2(Z)} \tag{4}$$

where Z is the ratio of source-receiver offset to borehole radius. They also, state that for small offsets, the spreading formula above approximately reduces to 1/Z and that their data confirm that this formula is valid for both the Schlumberger long and short tools. Willis (1983) also states that geometrical spreading loss formula for the P-wave is Z_1^{-1}, where Z_1 is the source receiver offset. Therefore, in this thesis, geometrical spreading losses are assumed to be frequency independent, and the amplitude decays inversely proportional to offset due to spreading losses.

Attenuation estimates from sonic logs should be useful because such estimates would provide another physical parameter describing the subsurface which is independent of other parameters, such as velocity and density. Since the variation of Q with frequency seems to be related to the amount of saturation and pressure on the sample, an accurate estimate of Q as a function of frequency could provide some knowledge of water saturation (or porosity) and pressure. Knowledge of Q obtained from sonic logs could be used to improve seismic data and also to improve the correlation between seismic data and synthetic seismograms calculated from well logs.

Synthetic seismograms from well logs frequently do not match seismic data very well. Many factors contribute to this problem, one of which is that the reflected wavefield represented by the seismic section has undergone a significant amount of attenuation. But the synthetic seismogram calculated from well log data is based on velocity, density, and source estimates with no attenuation taken into account. The correlation could improve significantly when an attenuation filter based on the estimated Q values is applied to the synthetic seismogram.

Hale (1982) and Bickel (1982) demonstrated that it is possible to design inverse-Q filters for seismic data. However, the design of such filters require some knowledge of the variation of Q with depth, and this information is not generally available. Hence, Q estimates from sonic logs could provide information useful for the processing of seismic data and increasing the resolution of seismic data. Finally, it is possible that a technique developed for estimating attenuation from sonic logs could form the basis for a method of estimating the amount of attenuation in seismic or VSP data directly from the data itself.

1.2 Attenuation Estimation

Attenuation measurements of acoustic waves in the earth have been of interest to scientists for some time. Most of the early efforts were aimed at estimating attenuation in the frequency range generally used in surface seismic esploration (10-100 Hz). When higher frequencies were used, no special effort was made to estimate attenuation in the acoustic well log frequency range (10-30 kHz). The classic experiment is that by McDonal et al (1958) while similar experiments have been done by Tullos and Reid (1969), and Spencer et al (1982). These experiments were of

the vertical seismic profile type. Experiments to estimate the earth's acoustic attenuation also include those by Taylor and Toksoz (1982) to estimate Q from earthquake data, and by Jacobson et al (1981) to estimate Q from offshore ocean bottom refraction data. Hale (1982) discussed estimating and removing attenuation effects from surface seismic data. Kuc and Schwartz (1979) used an interesting method to estimate the attenuation coefficient for the liver from ultrasonic data.

With the introduction of sonic logging tools capable of recording the entire received waveform, rather than just estimating P-wave velocity, attempts were made to make use of the extra data available. If properly interpreted, it is possible to acquire knowledge about the compressional and shear formation velocity, as well as the attenuation coefficient. The waveform shapes themselves may hold even more information about the condition of the hole and about the formations. Recent efforts to estimate formation parameters such as shear velocity and attenuation coefficient include Anderson and Castagna (1984), Cheng et al (1981 and 1982), Willis (1983), Aron et al (1978), Goldberg et al (1984), and Parks et al (1983).

Since the attenuation estimation problem may involve non-Gaussian noise, then the least-squares techniques may not work well. So, maximum likelihood and robust estimation techniques are introduced in Chapter III for use as attenuation estimators.

The spectral ratio method is the fundamental method used to estimate the attenuation coefficient, Q. It has served as the basis for more sophisticated Q estimation techniques, and the calculations are relatively simple and fast compared to some of the more complex techniques. Unfortunately, modelling has shown that in the presence of a

noisy signal the variance of the estimate may be large. Further, the algorithm tends to become unstable when the source spectrum has zeros or near zeros.

The spectral ratio method, as applied to acoustic logs, is based upon the model of a sonic tool in the borehole shown in Figure 1, and upon Equation (3). Compressional waves will originate at the source or transmitter "T" and will propagate out through the fluid. Some of the compressional wave will impinge upon the wall near "B". The energy which actually goes into the rock will be converted to P and shear waves in the rock. The P-waves refracted at or near critical angle will travel down the walls of the borehole, emitting compressional waves back into the fluid. When the re-emitted compressional waves reach the receivers "R1" and "R2", the wavetrain is recorded. Since the P-wave velocity in the rock is usually much greater than the shear velocity of the rock and the fluid velocity, then the P-wave which travelled through the rock arrives first. That is, it usually arrives before the "tube-waves" or the converted shear wave, and is easy to window out and separate from the rest of the data.

1.2.1 Development of the Spectral Ratio Method

The amplitude of a plane wave of frequency ω which has travelled a distance Z in a medium with attenuation coefficient Q, and phase velocity c is

$$A(Z,\omega) = A_0 e^{-\frac{\omega Z}{2Qc}}$$

where A_0 is the initial amplitude at Z = 0. For a plane wave with

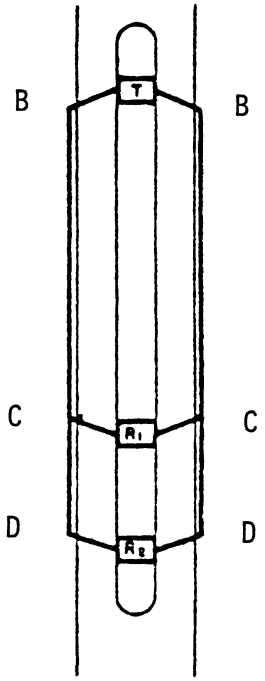

Figure 1 Model of Sonic Tool in a Borehole

source spectrum $S(\omega)$, the spectrum of the wave Z units away from the source can be written as

$$R(Z,\omega) = B_0 S(\omega)\, e^{-\frac{\omega Z}{2Qc}} \tag{5}$$

where B_0 is a constant. Rewriting Equation (5) and taking the logarithm of both sides results in

$$-\ln \frac{R(Z,\omega)}{S(\omega)} = \frac{\omega Z}{2Qc} - \ln (B_0) \tag{6}$$

From equation (6), it is obvious that the negative of the logarithm of the ratio of the spectra from the receiver and source is linear in frequency ω. The slope of that line is the same as the derivative of the right hand side of Equation (6), which is

$$\text{slope} = \frac{Z}{2Qc}$$

So, if the source-receiver offset and the velocity are known, then Q can be calculated from the slope of the line.

In full waveform acoustic logs, the spectrum of the received P-wave can be calculated via Fourier transform of the windowed time-domain P-wave arrival. The spectrum of the source can not be easily measured down hole. Also, the source spectra is affected by source-fluid-borehole acoustic coupling, which is usually different every time the source fires as the tool is pulled up the hole

Since the source spectrum must be regarded as unknown, the spectral ratios used must be those from adjacent receiver pairs.

Using Equation (5), the ratio of the spectrum at R_1 to the spectrum at R_2 is

$$\frac{R_2(\omega)}{R_1(\omega)} = \frac{1}{X} e^{-\frac{\omega X}{2Qc}} \tag{7}$$

where X = distance between R_1 and R_2, and c and Q are the phase velocity and attenuation coefficient of the subsurface between receivers R_1 and R_2. As previously mentioned the 1/X term approximately represents the geometric spreading losses for the P-wave. We shall define the log-spectral ratio, SR(ω) to be

$$SR(\omega) = - \ln \frac{R_2(\omega)}{R_1(\omega)} = \frac{\omega X}{2Qc} + \ln (X) \tag{8}$$

where $R_1(\omega)$ and $R_2(\omega)$ are the spectra of the P-wave arrivals at receivers "R_1" and "R_2". For real data, SR(ω) will not be a straight line. An estimate of the slope, given by,

$$\frac{1}{2Qc}$$

is made by finding the best-fit line through the data, and using the slope of that best-fit line Since X is the known receiver spacing and the phase velocity c can be calculated from P-wave first breaks, Q follows easily from the slope

The ln(X) term in (8) is due to the geometrical spreading loss, that loss was assumed to be of the form 1/X But, if the geometrical spreading loss is not exactly 1/X, then as long as it is frequency independent, the effect of the loss will appear only in the intercept of SR(ω) in Equation (8) Thus the loss will have no affect on the Q estimate

Since the spectral ratio deals only with ratios of received spectra, the offsets are typically much shorter than the total source-receiver spacing This has advantages as well as disadvantages The primary advantage is resolution That is, the estimated Q is that for

the subsurface layer between receivers "R_1" and "R_2" only. If the source-receiver spectral ratio were used, the resolution would be much poorer, since Q would be measured over a much greater distance In addition, P-wave velocities can be picked accurately by analyzing timing differences between adjacent receivers The disadvantage of small offset receiver to receiver spectral ratios is that relatively little attenuation has taken place. For example, assume the following

$c = 10000$ ft/sec , $1 = 2\pi f = 2\pi$ 10 kHz, $X = 2$ ft., $Q = 100$

Then $\frac{\omega X}{2Qc} = \frac{2\pi}{100} = .063$

So $e^{-\frac{\omega X}{2Qc}} = 939$, or 6.1% attenuation has taken place between receivers.

In contrast, if the source spectrum was known, and assuming a source-receiver offset of 12 ft., then

$$\frac{\omega X}{2Qc} = 0.38,$$

and

$$e^{-\frac{\omega X}{2Qc}} = 0\ 68$$

has taken place between source and receiver

1.2 2 Attenuation Estimation from Acoustic Logs-Previous Work

Cheng, Taksoz, and Willis (1981 and 1982) discuss estimating attenuation from full waveform acoustic logs. They argue that the effective bandwidth of the source spectrum is too narrow for the spectral ratio method to work well. See Figure 2 for P-wave amplitude spectra from

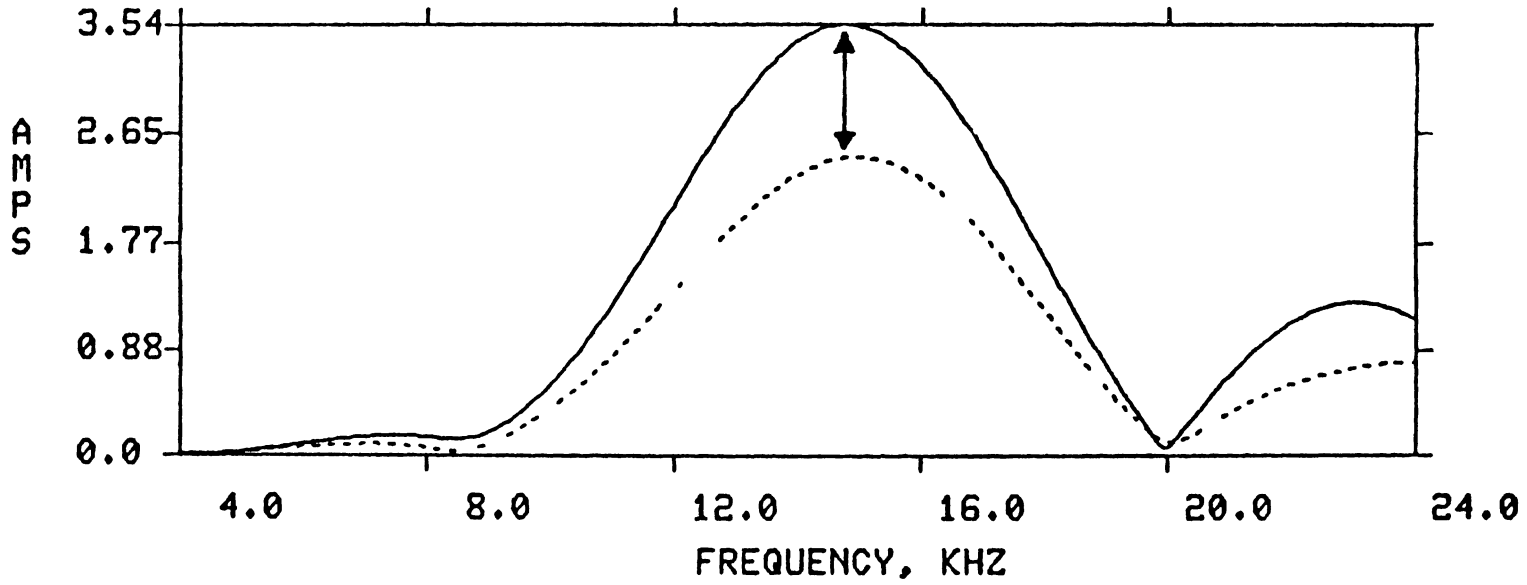

Figure 2 P-Wave Amplitude Spectrum from Cheng, et al (1981)

Cheng (1981). So, they estimate the attenuation from the decrease in amplitude of the second received spectrum at the peak. This is roughly the same as using the spectral ratio method, but only using the frequency point at the maximum of the spectrum. This method has the advantage of only using the frequency where the signal is strongest. Further, for reasonable values of receiver offset, Q, and velocity, the drop in peak amplitude is significant.

On the other hand, there are several difficulties involved in using this technique. For the spectral ratio method, any frequency independent geometrical spreading losses will appear in the intercept rather than the slope of the best-fit line. Thus, spreading effects need not be taken into account to estimate Q. However, for this method, spreading effects must be accounted for. Theoretically, the spreading loss is proportional to 1/Z, where Z is the source-receiver offset, and this is the correction factor used by Cheng. This is fine as long as there are no other frequency independent losses (otherwise the Q estimate is affected), so having to account for spreading losses is one of the disadvantages of this technique. Another disadvantage is that this technique only uses one frequency point. So, if the entire spectrum is corrupted by noise, then the estimate based upon one frequency point will be less reliable than an estimate based upon many points.

It is interesting to note that the shift in the spectral peak predicted by Kuc and Schwartz (1979) (discussed later) is only barely discernable in Figure 2. So, it appears that the spectral shifts present in acoustic log data is not be enough to allow the use of their method.

In his Ph.D. thesis, Willis (1983) uses the maximum likelihood method (Pisarenco, 1970) to estimate the transfer function (or earth response) from the input (first receiver data) to the output (second receiver data). The idea of estimating the transfer function is the same as that used by the Wiener filter technique (Taylor and Toksoz, 1982), except that the actual estimation is done using a different algorithm. The maximum likelihood method of Pisarenko (1970) attempts to achieve an optimum estimate of an unknown transfer function when n different examples of noise corrupted input and output functions are known. Pisarenko (1970) showed that his estimator is asymptotically unbiased, whereas the simple averaging of spectral ratios results in a significant bias. In addition, Pisarenko's method has a lower mean square error.

Goldberg, Kan, and Castagna (1984) estimated Q from synthetic waveforms and from real data. They modelled the received power spectrum, $A(Z,\omega)$, as

$$A(Z,\omega) = T(Z)\ S(\omega)\ e^{-\ aZ} \tag{9}$$

where $T(Z)$ is the frequency independent power loss, Z is the source-receiver offset, $S(\omega)$ is the total system response, and the exponential term, e^{-aZ} is the attenuation operator. The constant, a, in the exponent is assumed to be linearly dependent on frequency, thus following the constant-Q assumption.

The system response, $S(\omega)$ includes the source signature, receiver response, cable transmission, all frequency dependent coupling effects, and the electronics. $S(\omega)$ is also assumed to be offset-invariant and thus is calculated by averaging over the available offsets for a shot.

T(Z) represents the geometrical spreading losses (assumed frequency independent) and the coupling losses which are frequency independent, and is calculated from the average power contribution at each receiver. The "a" is calculated in the usual way via the spectral ratio method. The estimated parameters are then adjusted to minimize a power-weighted measure of the mean square error between the data and the fit from the estimated model.

Goldber, Kan, and Castagna (1984) concluded that the estimate of Q is adversely affected by "deep nulls" in the power spectra (what have been described as zeros in this thesis). In addition, they determined the estimate to be sensitive to window shape and size. They also modelled the effects of a tilted tool, and demonstrated that the effects were readily apparent in the received waveforms. This may explain the "spectral nulls" or zeros in the spectra.

Anderson and Castagna (1984) presented a paper on the analysis of amplitudes of compressional waves on sonic logs. They use the "borehole compensation" (BHC) technique to remove from the log the coupling and focusing affects which greatly affect amplitudes. The BHC method is based upon ray-path analysis, and adds (or subtracts) signals from various receivers to approximately cancel effects such as borehole focusing due to tool position and tilt, attenuation in the fluid, and fluid-borehole coupling.

The authors use the BHC correction technique to produce amplitude, attenuation, and coupling logs, as well as the usual transit time (slowness) log. They note that the amplitudes did approximately decay as theoretically predicted, but the corrected amplitude is still

sometimes less than predicted. This was assumed to be caused by imperfect compensation by their BHC method.

1.2.3 Attenuation Estimation from Data Other Than Acoustic Well Log Data

Taylor and Toksoz (1982) used a Wiener filtering approach to estimate attenuation of the earth from earthquake seismograms. The primary advantage of this technique is that it yields a more reliable estimate of Q from data with spectral zeros than does the spectral ratio method. The method is based upon estimating a time-domain filter to approximate the impulse response of the earth between adjacent receivers. The Fourier transform of this filter is an estimate of the spectral ratio from which Q can be calculated.

The spectral ratio relation from Equation (6) actually represents the frequency domain transfer function or Green's function between receivers R_{i-1} and R_i, $G(\omega)$, and could be rewritten as

$$SR(\omega) = - \ln [G(\omega)],$$

where

$$G(\omega) = \frac{R_i(\omega)}{R_{i-1}(\omega)} \tag{10}$$

This could also be written in the time-domain as a convolution

$$r_i(t) = g(t) * r_{i-1}(t) \tag{11}$$

where $g(t)$ is the impulse response, and $r_i(t)$ is the signal recorded by the i^{th} receiver.

In general, g(t) will be an infinitely long sequence. However, it could be approximated by a finite length sequence f(t) using a criterion such as minimizing the least squares error

$$E = \sum_t [f(t) * r_{1-1}(t) - r_1(t)]^2 \tag{12}$$

It is clear that this is the classical discrete-time Wiener filter problem. To cast this as a least-squares linear algebra problem, let $b_J = r_{1-1}(t_J)$, $d_J = r_1(t_J)$, and $f_J = f(t_J)$. Then minimizing E in (12) is equivalent to solving (see for example Willis, 1983)

$$B\underline{f} = \underline{d} + \underline{e},$$

where

$$B = \begin{bmatrix} b_0 & & & 0 \\ & \ddots & & \\ b_n & & \ddots & \\ & & & b_0 \\ 0 & & & b_n \end{bmatrix}, \qquad \underline{f} = \begin{bmatrix} f_0 \\ \vdots \\ f_m \end{bmatrix}$$

$$\underline{d} = \begin{bmatrix} d_0 \\ \vdots \\ d_{m+n} \end{bmatrix}, \qquad \underline{e} = \text{error vector} = \begin{bmatrix} e_0 \\ \vdots \\ e_{m+n} \end{bmatrix}$$

The solution

$$\underline{f} = (B^T B)^{-1} B^T \underline{d} \tag{13}$$

minimizes the error energy, $E = \underline{e}^T \underline{e}$. The matrix $(B^T B)$ is the autocorrelation matrix and has the Toeplitz form

$$B^T B = \begin{bmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_1 & a_0 & & a_{n-2} \\ \vdots & & \ddots & \\ a_{n-1} & a_{n-2} & & a_0 \end{bmatrix}$$

where $a_1 = \Sigma_j\ b_j\ b_{j-1}$. The vector

$$B^T\underline{d} = \begin{bmatrix} h_o \\ \vdots \\ h_{n-1} \end{bmatrix}$$

is the cross correlation between the input and output, where $Z_1 = \Sigma_j\ b_{j-1}\ d_j$.

A zero in $R_{1-1}\ (\omega)$ corresponds to a singularity in the matrix (B^TB). Also, if zeros were present in the spectrum $R_{1-1}(\omega)$, then the autocorrelation matrix will again be singular and it must be diagonally loaded to have a stable estimate of Q. This diagonal loading of the matrix (B^TB) is roughly equivalent to adding a constant to the spectrum $R_{1-1}(\omega)$.

Once a solution for $\underline{f}$ is calculated to approximate the original impulse response g(t), the Fourier transform of $\underline{f}$ can be calculated. Thus, the Fourier transform of $\underline{f}$, $F(\omega)$, is an estimate of $G(\omega)$. So, equation (10) can be restated as

$$-\ln [F(\omega)] = RR_1\ (\omega) = \frac{\Delta z}{2Q_1 c_1}\ \omega. \qquad (14)$$

Now, Q can be estimated by doing a least-squares linear regression on $-\ln [F(\omega)]$ to find the best-fit slope.

Unfortunately, diagonal loading of the autocorrelation matrix results in errors in the estimation of Q. Taylor and Toksoz, (1982), suggested a technique for dealing with the inaccuracies. Diagonal loading B^TB is equivalent to replacing B^TB by $(B^TB + I\lambda)$ where λ is some small positive number and I is an identity matrix. The frequency domain analog of Equation (13) is

$$F(\omega) = \frac{C(\omega)}{A(\omega)} \quad , \tag{15}$$

where $C(\omega)$ and $A(\omega)$ are Fourier transforms of $z(t_1) = z_1$ and $a(t_1) = a_1$ respectively.

In the frequency domain, diagonal loading of B^TB corresponds to replacing $A(\omega)$ by $A(\omega) + \lambda$ in (15).

Now,

$$F'(\omega) \; Z \; \frac{C(\omega)}{A(\omega)+\lambda} \tag{16}$$

where $F'(\omega)$ = the Fourier transform of $f'(t)$,

From this, we can write

$$F(\omega) = F'(\omega) \quad \frac{A(\omega)+\lambda}{A(\omega)} \tag{17}$$

Hence, multiplying $F'(\omega)$ by $[(A(\omega)+\lambda)/A(\omega)]$ corrects the errors introduced by diagonal loading. However, we still have not solved the problem since $A(\omega)$ may be zero at certain frequencies. To alleviate this problem Taylor and Toksoz (1978) suggested smoothing $A(\omega)$ by windowing the autocorrelation and applying the correction only if $A(\omega)$ is above some minimum threshold.

To correct the diagonal loading induced errors, apply the correction factor in equation (17) whenever $A(\omega)$ is large enough for the correction to be stable. Thus, a threshold must be set and Equation (14) only applied when $A(\omega)$ is above that threshold. This threshold is defined as a fraction of the maximum $A(\omega)$. Similarly, the amount of diagonal loading applied to the autocorrelation matrix is defined as a fraction of the diagonal value. These two fractions must be input parameters for the Q-estimation algorithm, and a poor choice for either

may lead to very poor results. Another approach might be to interpolate across frequencies in the vicinity of the spectral zero.

The classic experiment to estimate in-situ attenuation of the earth is that by McDonal, Angona, Mills, et al (1958). They used geophones clamped to the sidewalls of a borehole through a fairly homogeneous 500 ft. thick layer of shale. The geophones measured the response of the earth at various depths from the same shot. They used the spectral ratio method to estimate the attenuation, and made the following conclusions (1) The shale attenuation followed the constant Q model very well (2) No velocity dispersion was measured, meaning the propagation speeds of various frequencies were the same (over the range 20-450 Hz) (3) The earth does not behave as a classical visco-elastic medium because the velocity dispersion which should accompany attenuation was not present.

McDonal, Angona, Mills, et al (1958) found the lack of dispersion very surprising, and admitted that it may not be valid to extrapolate the lack of dispersion to higher frequencies (such as those used in sonic logging). For the measured frequency range, evidence continues to support the conclusions about the constant Q model and that the amount dispersion is negligible. Although some authors maintain that the dispersion is measurable, even from the data of McDonal, Angona, Mills, et al (1958).

Tullos and Reid (1969) used an experimental layout very similar to that used by McDonal, Angona, Mills, et al (1958) to estimate attenuation of Gulf Coast sediments. They, too used the spectral ratio method, but averaged as many as 156 spectral samples to reduce the effects of noise. They also concluded that, within a specific layer, Q

is frequency independent. Further, Q may vary greatly from layer to layer. It is typically very low (high attenuation) near the surface, and the value of Q usually increases with depth.

Spencer, Sonnad, and Butler (1982) modeled the VSP layout and used the spectral ratio method to analyze the data and estimate Q. They averaged their spectral ratios in a manner similar to Tullos and Reid (1969). Spencer, Sonnad, and Bulter (1982) did an error analysis to go along with the attenuation estimates. They observed that the variance of the Q estimate was very large for small receiver separation, but the variance decreased quickly to a much smaller amount as the receiver spacing increased. They also conclude that local interference due to multiples accounts for much of the measured attenuation, and that the Q-estimation problem is ill-posed (i.e. small data errors lead to large estimation errors).

From their data, it appears as though the variance is very high for receiver separation below 200 ft., for a frequency band of 0-125 Hz. The center of the band, 62.5 Hz, is 200 times lower than a typical frequency in sonic logs (12.5 kHz). Thus the minimum receiver spacing scales down to about one foot for sonic log frequencies, hopefully this means that receiver spacings over one foot will give adequate accuracy in estimating Q.

Two recent papers discuss attenuation affects present in surface seismic data. Bickel (1982) shows that for a band-limited source, a filter can be both a reasonably accurate inverse Q filter and stable as well. Unfortunately, Bickel (1982) ignores the very difficult question of Q-estimation by showing that if Q = 50 for a model, inverse filters use Q = 40 or Q = 60 work fairly well to remove the attenuation

effects. He then concludes that all that is needed is a rough estimate of Q, and he makes no effort to handle a depth-variable attenuation coefficient.

The work of Hale (1982) is significantly different in that he describes an algorithm to estimate Q. Assumptions made by Hale include (1) Q does not vary with depth (2) the source waveform is minimum phase. To preserve strict causality in order to maintain the minimum phase assumption, Hale (1982) changes the attenuation model by replacing ω in equation (3) with

$$\omega' = |\omega| + J\, H(|\omega|) \tag{18}$$

where $|\omega|$ is the absolute value of angular frequency ω, $J = [-1$, $H(\bullet)$ is the Hilbert transform operator and ω' is the "new" frequency. Hale and Bickel both use as their basic inverse attenuation operation the "inverse" of that given by equation (3)

$$F(\omega,z) = e^{+\frac{\omega Z}{2Qc}} \tag{19}$$

where ω, Z, Q, c are as defined before. Hale then forces causality with the substitutions given by equation (18).

Hale then does the following. Let x(t) be the input trace to inverse attenuation operator, $\hat{r}(t)$ be the output trace, and $U(\omega) = |\omega| + J\, H(|\omega|)$. Then, by the principle of superposition

$$\hat{r}(t) = \sum_{\omega} e^{J\omega t}\left[e^{\frac{U(\omega)}{2Q}}\right]^{t} \tag{20}$$

where $X(\omega)$ = Fourier Transform (F.T.) of x(t). This can be rewritten as

$$\hat{r}(t) = \sum_{\omega} e^{(J\omega t)} X(\omega)\left[1+(t/2Q)U(\omega)+\frac{1}{2}\left(\frac{t}{2Q}\right)^2 U^2(\omega)+\ldots\right]$$

or,

$$\hat{r}(t) = \sum_{i=0}^{B} \frac{1}{i\pi} \left(\frac{t}{2Q}\right)^{i} [u^{*i}(t) * X(t)] \qquad (21)$$

where * denotes convolution, $U(\omega)$ is the F.T. $(u(t))$, and u^{*i} means $u(t)$ convolved with itself i times. According to Hale, the number of terms, necessary to estimate $\hat{r}(t)$ in Equation (21) is a function of t and Q, but the computational effort needed to implement (21) is comparable to other time-variable deconvolution schemes. Most of the effort to evaluate (21) is involved in the convolution. Thus the Q-independent convolution only need be calculated once. Then $\hat{r}(t)$ can be evaluated for various guesses of Q, the optimum being the Q which minimizes

$$E = \sum_{t} [\hat{r}(t)\ e^{\frac{\pi t}{4Q}}]^2 \qquad (22)$$

Unfortunately, the convolutional model used by Hale and Bickel does not fit the models used for sonic logs, and Hale's assumption of depth-invariant Q is not a realistic one.

The seismic refraction experiment used by Jacobson, Shor, and Dorman (1981) consisted of depth charges set to explode in deep water, and seafloor hydrophones as receivers. The P-wave energy from the source reaches the seafloor, and some of it is refracted along the floor and rock layers below it. As in the acoustic log case the travelling refracted wave emits energy which travels back up to the receivers on the seafloor. Their method of Q-estimation is based upon the J^{th} observation of the spectral ratio as previously defined. According to Jacobson et al

$$-\frac{SR_J(\omega,Z)}{\omega/2} = \sum_{i=1}^{m} \frac{t_{iJ}}{Q(Z)} \qquad m \le n \qquad (23)$$

where $SR_J(\omega,z)$ is the J^{th} observation of the spectral ratio at depth Z

and frequency ω, $Q(Z)$ is the attenuation coefficient at depth Z, n is the total number of layers, and t_{ij} is traveltime in the i^{th} layer for the j^{th} observation. Rewriting the right hand side of (23) using k_i for the attenuation for the i^{th} layer measured in dB/kHz/m and PL_{ij} for the path length for layer i and the j^{th} apparent velocity yields

$$- \frac{SR_j(\omega,Z)}{\omega/2} = \sum_{i=1}^{m} k_i \, PL_{ij} \qquad (24)$$

Now, note that (24) now describes a linear system of equations of the form $A\underline{b} = \underline{c}$. This underdetermined system of equations would normally be solved by the minimum norm least-squares technique. But, Jacobson, Shor, and Dorman (1981) chose to use an inverse technique from Wiggins (1972). This technique involves solving the system by decomposing the matrix A into orthonormal eigenvectors and ranking these eigenvectors. The total number of eigenvectors used is based upon a chi-square analysis of the result using different numbers of eigenvectors. Jacobson, Shor, and Dorman state that this method is similar to least-squares analysis, and indeed it appears to be very similar to using the so called QR decomposition with column pivoting to solving a least squares system of equations (Golub and VanLoan, 1983). In fact, the only thing noticeably different about Wiggins technique as used by Jacobson is that a chi-square test is used to determine when enough eigenvectors have been used. This is in contrast to some cut-off for minimum eigenvalues suggested by Golub and VanLoan.

Jacobson, Shor, Dorman (1981) concluded that their data analysis method produced less variance than other methods. Their data agreed with the usual conclusion that, over a frequency range of 0 to 100 Hz, Q is frequency independent and that Q generally increases with depth.

They did note a possible sharp increase in attenuation (decrease in Q) around a depth of 600 meters and an average Q around one hundred. However, it must be pointed out that their data showed very wide confidence intervals, with possible ranges of Q at some depths from about 30 to 100 or from 100 to infinity.

To estimate the acoustic attenuation coefficient of the human liver from ultrasound data, Kuc and Schwartz (1979) used an interesting technique which is not based upon the spectral ratio method (the attenuation coefficient yields data about liver cirrhosis). Kuc amd Schwartz (1979) showed that if the source power spectrum is Gaussian, and if the constant-Q model holds, then the received signal will also have a Gaussian power spectra. Furthermore, the received spectrum will be shifted down in frequency and the size of the shift will be proportional to Q^{-1}.

Suppose the input spectrum is Gaussian, centered at frequency ω_0 with variance s^2. Then,

$$R_1(\omega) = e^{-\frac{(\omega-\omega_0)^2}{2s^2}} \tag{25}$$

If a pulse with the above spectrum passes through an attenuative, constant-Q medium of thickness Z and velocity c, then the output spectrum would be

$$R_2(\omega) = e^{-\frac{\omega Z}{2Qc}} e^{-\frac{(\omega-\omega_0)^2}{2s^2}} = e^b \tag{26}$$

Then

$$-2b = \frac{\omega Z}{Qc} + \frac{(\omega-\omega_0)^2}{s^2} = \frac{\omega^2}{s^2} - a + \frac{\omega_0^2}{s^2} \tag{27}$$

where $a = \frac{2\omega_o Q_c - Zs^2}{Qcs^2}$.

Completing the square on (27) requires that the new center frequency, ω_o', be given by

$$\omega_o' = \frac{s^2 a}{2} = \omega_o - \frac{Zs^2}{2Qc} \tag{28}$$

where the second term ($Zs^2/2Qc$) defines the shift in the peak and it is proportional to Q^{-1}.

The shift in the peak is determined by the cross correlation between $R_1(\omega)$ and $R_2(\omega)$. The cross-correlation shift detector is optimum in the least-squares sense. If the noise in the system is Gaussian, then the cross-correlation detector will also be a maximum likelihood (ML) detector. This method has the potential to be applicable to acoustic well log data because the source spectrum may be roughly Gaussian in shape. For example, Figure 3 is a plot of the source spectrum of a typical acoustic logging tool (Aron, Murray, and Seeman, 1978). The main lobe of this spectrum is approximately Gaussian in shape. However, the tuning effects could easily distort this shape and make this method inaccurate.

1.3 Summary of This Chapter and Description of This Thesis

Applying plane-wave attenuation models to a borehole is discussed in Section 1.2, as is the basic or fundamental method of estimating Q, known as the spectral ratio method. This method is based upon the observation that from Equation (3), the logarithm of the ratio of power spectra should be linear in frequency. The slope of this line should be inversely proportional to Q. On real data, Q is estimated from the

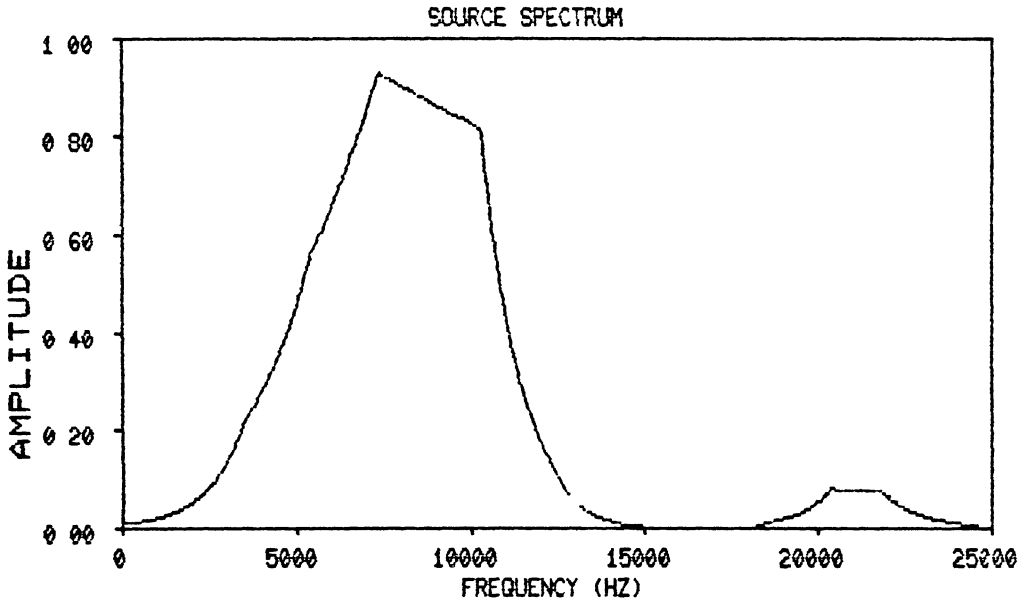

Figure 3 Digitized Source Spectrum (Amplitude Rather than Log-Amplitude)

slope of the least squares best fit line to the log-spectral-ratio of the data.

Many different methods have been used to estimate the acoustic attenuation coefficient, Q, from many different types of acoustic data. These data types include surface seismic, vertical seismic profile (VSP), sonic logs, offshore refraction, earth quakes, and ultrasonic data. Most methods are based upon the spectral ratio method, and Section 1.2 also contains a discussion of these methods.

Two previously used methods for estimating attenuation, the spectral ratio method and the Wiever filter method, are studied in more depth and tested on simple model data in Chapter II. The performance of these estimators in the presence of additive noise is evaluated, as is the performance on data arrising from source spectra with "zeroes" or "nulls" in them. Finally, the performance of these methods is compared with model data from a source spectrum with "zeroes" which is also contaminated with noise.

Chapter II also introduces an analysis of the attenuation estimation problem from a different point of view. A value of Q can be calculated from the received P-wave arrivals from two adjacent receivers, at each frequency point. These Q estimates can be cast as a matrix, with the column number representing the frequency value and the row number representing the adjacent receiver pair number (or depth). This matrix representation of Q estimates leads to several new ways of estimating Q. These new methods involve eigenvector-eigenvalue decomposition as well as robust and maximum likelihood estimation.

Since the attenuation coefficient, Q, is assumed to be frequency independent, then all of the columns of the data matrix will be the

same. Thus this matrix will be a rank one matrix. For real data, of course, the estimates will be noisy and the borehole will not conform to the ideal model. Therefore, the matrix will not be a rank one matrix.

However, if the ideal model is fairly accurate, and if the noise level is low, then the matrix will be "almost" rank one. In other words, the matrix will have one dominant eigenvalue. For non-square matrices, the dominant eigenvector (of the column space) corresponding to the dominant eigenvalue is found using singular value decomposition. It can be shown that this dominant eigenvector is an optimum least-squares estimate for the values of Q. Thus, the most significant eigenvector of the column space of the Q matrix is an estimate of Q as a function of depth or receiver pair number. This method can be compared to an optimum (least squares) weighted sum of the columns of the Q matrix.

Estimation of Q from power spectra is an ill-posed problem because small errors in the spectral estimate may lead to large errors in the final estimate of Q. Even for simple additive input noise models, the noise ditribution which contaminates the Q estimate cannot be written in closed form. Rather, the errors in Q estimation can only be analyzed using computer simulations. These simulations show that even if the input noise is Gaussian, the error in attenuation estimation can be highly non-Gaussian. So, estimators which are more robust than least-squares methods, and thus less sensitive to very bad data points, are needed for more reliable estimates of Q. Several classes of robust estimators and the underlying principle of maximum likelihood estimation are introduced in Chapter III.

Robust estimators of attenuation are also "justified" in Chapter III. A fairly simple model for noise in attenuation estimation is also introduced in that chapter, and several types of robust estimators are tested and compared for various noise models using computer simulations. These estimators are also compared with some of the previously discussed least-squares methods on the new noise models introduced in Chapter III.

More realistic noise models applicable to acoustic log data are introduced in Chapter IV. The effect of the noise in these models on the attenuation estimate from various robust estimators is analyzed. Computer simulations are used to compare these robust estimators with other types of estimators, including median, mean, -trimmed mean, and least-squares estimators. The comparisons are based upon accuracy of the estimation, ability to handle non-assumed noise distributions, and computational considerations.

The models used to simulated borehole propagation to this point are fairly simple, one dimensional, idealized models. These models do not include geometrical spreading losses and other complications due to the borehole itself. The tuning effect of the borehole acting as a waveguide may result in geometrical losses which may not conform to the losses predicted by theory. In addition, these losses may be very dependent on frequency. Therefore the robust estimation methods as well as simpler least-squares methods (spectral ratios, eigenvector decomposition) should be tested on more realistic data. Chapter IV also contains a discussion of the testing of the above estimators on "realistic" borehole model data from Conoco.

As a final test of the Q-estimation algorithms, their performance on real full waveform sonic data is evaluated. Unlike the models, the actual value of Q is unknown. Therefore the preformance of the estimators is more difficult to determine since the "answer" remains unknown. The real data used for this study comes from sonic logs from Conoco's borehole test facility, and some results from this data are also given in Chapter IV.

CHAPTER II

LEAST SQUARES ATTENUATION ESTIMATION METHODS

The spectral ratio method is the fundamental method of attenuation estimation. In order to develope an understanding of attenuation estimation in general and the spectral ratio method in particular, the performance of the spectral ratio method on simple model data is analyzed. In addition, the effects of noise and zeros in the source spectra on the spectral ratio method are studied. Since adding a constant to the diagonal of the autocorrelation matrix in the Wiener filter method roughly corresponds to whitening the spectra of the received signals, then the Wiener filter method should be able to handle spectral zeros better than the spectral ratio method. So, the Wiener filter method is also analyzed with regard to the effects noise and spectral shape have on the attenuation estimate. This method is also compared to the spectral ratio method.

There are other attenuation estimation methods which are different from the spectral ratio from just two adjacent receivers. First, remember that an estimate of attenuation can be calculated from every frequency value used in the spectral ratio method, rather than just one value of Q from the best-fit line through many frequency points. This is assuming that geometrical spreading has already been accounted for. These Q estimates can be cast as a matrix where the row number corresponds to the reciever pair from which the Q-estimate is made, and the

column number corresponds to frequency. Then for a reasonable model, this leads to a new technique where the dominant eigenvector of the column space of the data matrix forms an optimum estimate of Q for each receiver pair. These receiver pairs correspond to depth. This forms the basis of the eigenvector decomposition technique for the estimation of attenuation. The eigenvector decomposition method should perform better than the spectral ratio method for data with spectral zeros. To evaluate this method, it is compared to both the Wiener filter and spectral ratio methods on several data sets. These data sets contain various levels of noise and some of the data sets are from sources with spectral zeros.

2.1 Numerical Modeling of the Spectral Ratio and Wiener Filter Methods

Since many researchers have used the spectral ratio method or a variant of it, it is worthwhile to show a few simple examples to illustrate the use of the method. Consider a single layer, one-dimensional model with the following parameters

receiver spacing 2 ft.

velocity 10,00 ft./sec.

Q 50

noise 1%

Also, consider the broad-banded, cosine shaped spectrum is shown in Figure 4a for the input. The frequency range for this example is 10 kHz to 30 kHz. The spectrum of the signal after attenuation by the single-layer model with the above parameters is shown in Figure 4. Figure 5 shows the log spectral ratio, $SR(\omega)$ (see Equation 6) for this data set.

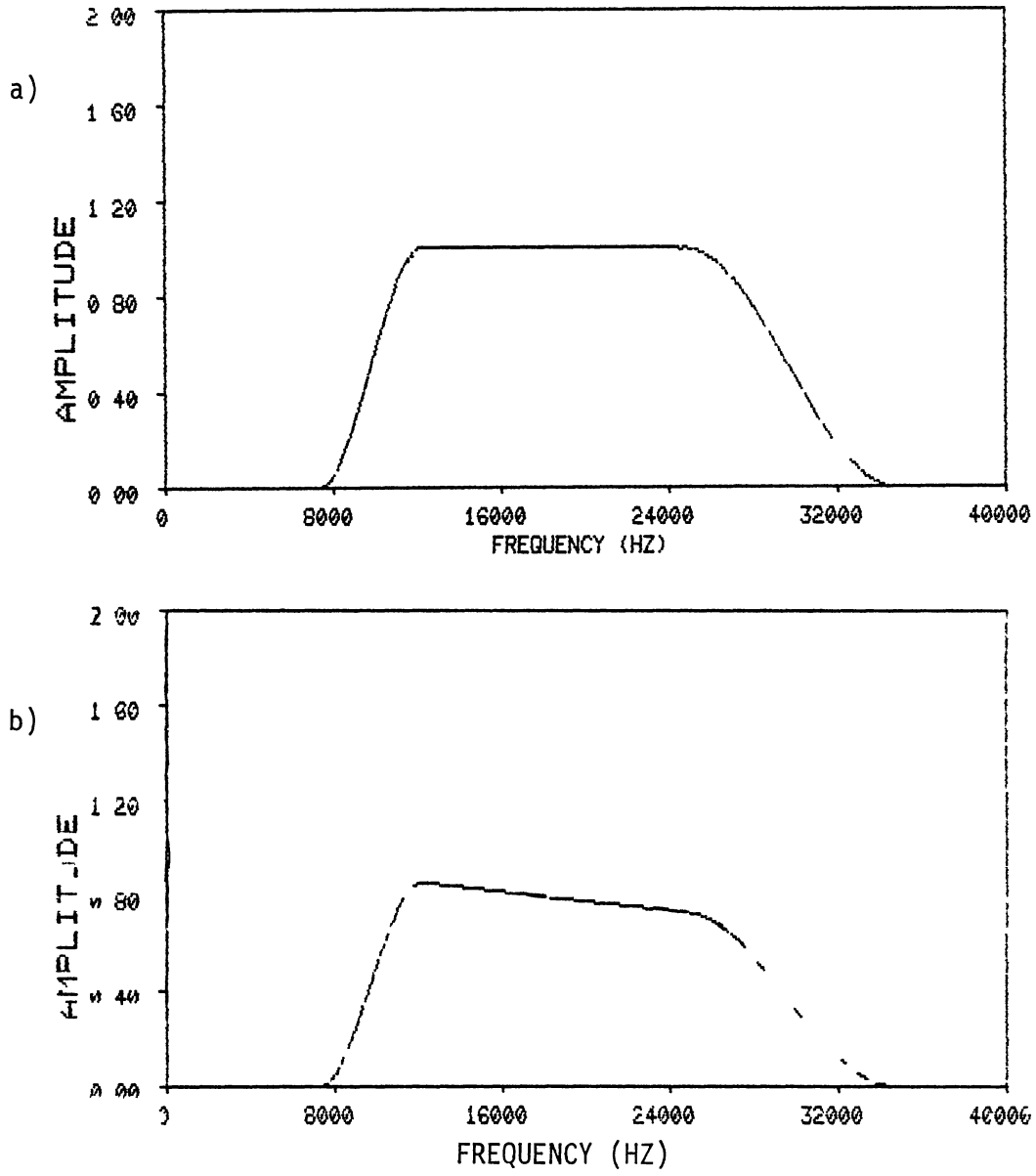

Figure 4 Input and Output Spectrum a) Input Spectrum-Cosine Shaped, b) Output Spectrum after Attenuation

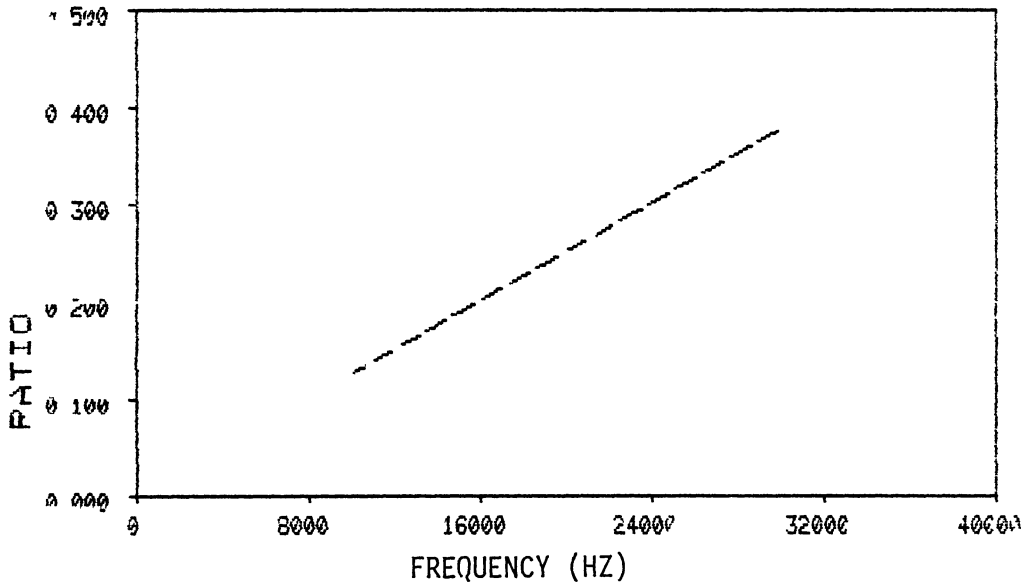

Figure 5 The Log-Spectral-Ratio of the Data in Figure 4 (a and b)

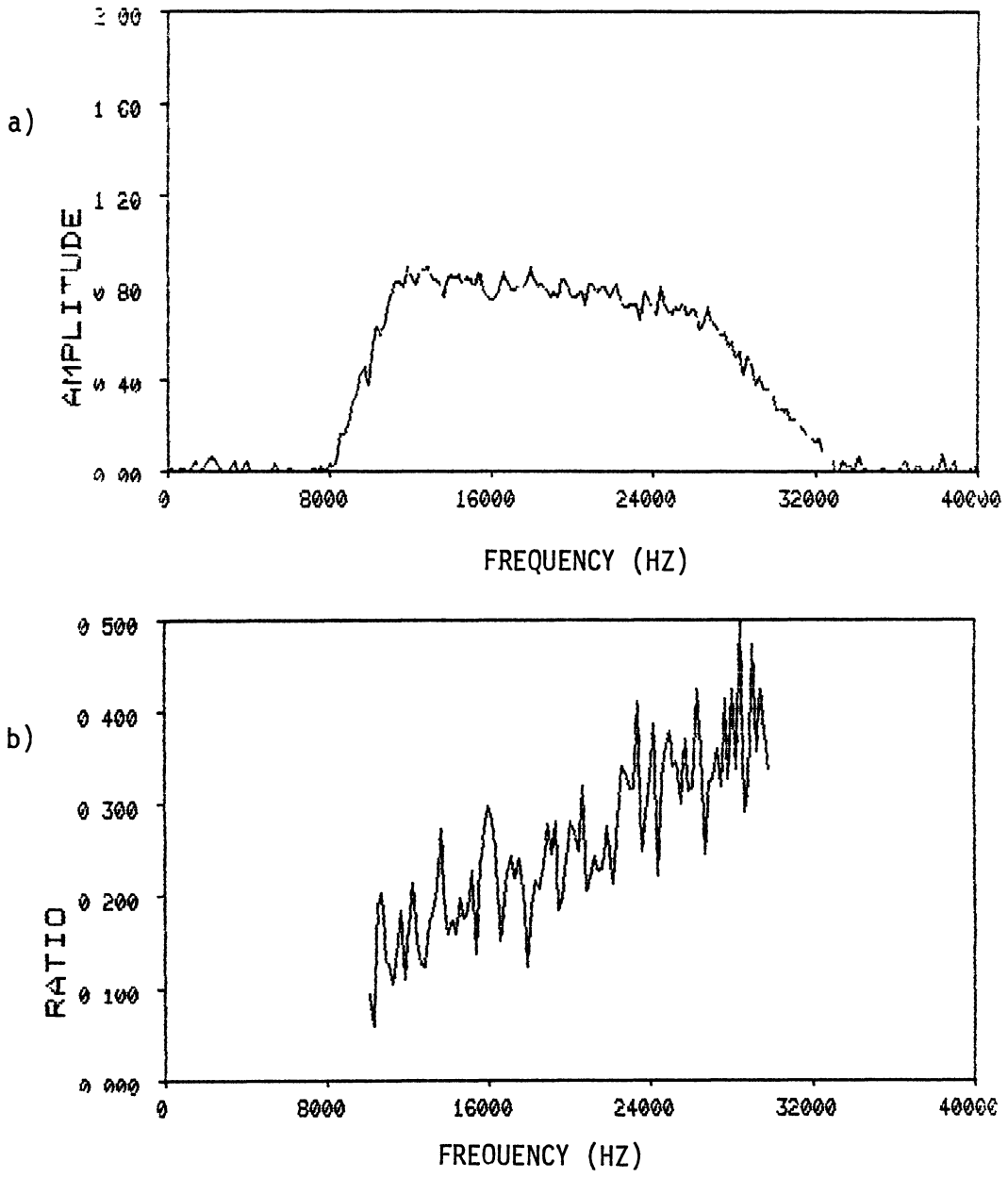

Figure 6 Output Spectrum with 1% Noise and its Log-Spectral-Ratio a) Output Spectrum with 1% Noise, after Attenuation, b) The Log-Spectral-Ratio of the Data in 6a

TABLE I

AVERAGE Q ESTIMATE AND STANDARD DEVIATION OF THE ESTIMATE
FOR SOURCE FROM FIGURE 4A

SNR	Average Q Estimate	Standard Deviation of Estimate
10000.	50.1	1.3
1000.	50.1	0.2
100.	49.6	4.0
10.	49.8	14.4
1.	59.5	539.

TABLE II

AVERAGE Q ESTIMATE AND STANDARD DEVIATION OF THE ESTIMATE
FOR SOURCE FROM FIGURE 7

SNR	Average Q Estimate	Standard Deviation of Estimate
10000.	50.9	3.3
1000.	50.3	9.3
100.	295	1900
100.	56.3	28.7*

*Using the best 90 of 100 trials from the line above.

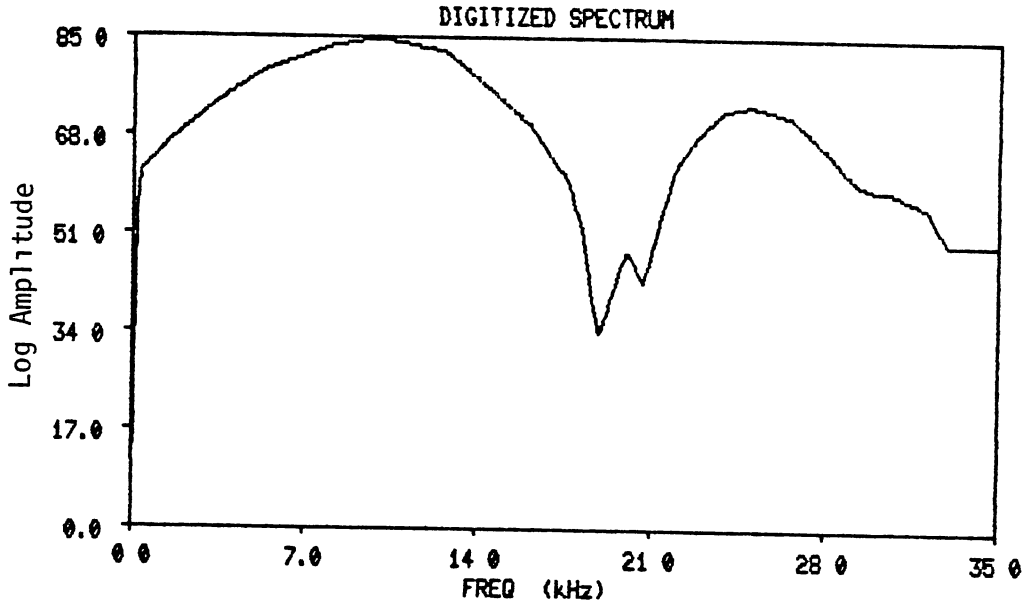

Figure 7 Frequency Spectrum of Source (from Aron, et al, 1978)

To more accurately simulate the borehole environment, Gaussian noise has been added to the received spectra with the noise power level of one percent of the signal power. Figure 6a shows the spectrum of the signal after attenuation and with the noise added on a typical test run. The log spectral ratio is shown in Figure 6b, and the attenuation estimated by the least-squares fit is Q = 47 8, and the root-mean-square error of the fit is 0.048.

In order to gain a better understanding of the accuracy of the attenuation estimate, the experiment described above was repeated 100 times with the same parameters. The average value of Q was 49.6, and the calculated standard deviation was 4.0. So, Q could be calculated fairly reliably in this case. Table 1 shows the average and standard deviation of Q for various signal-to-noise ratios.

Next, let us consider spectra that have both uncorrelated noise and zeros. Figure 7 shows a source spectral shape from Aron et al (1978), and the source obviously has a zero in its spectrum The effect of this zero on the numerical calculations is very significant. Assuming the borehole behaves as a linear system, then very little energy will propagate with frequencies in the vicinity of the zero. Thus the spectral ratio will be the ratio of a small number to (perhaps) an even smaller number. This obviously leads to numerical instabilities. Furthermore, since random noise may be present at all frequencies, overwhelming the signal near spectral zero, then the spectral ratio may involve ratios of noise to noise. This leads to completely unreliable spectral ratio estimates, with potentially very high errors

This spectral ratio method is unable to handle the more realistic case of a source with spectral zeros (see Figure 7) and additive random

noise. To demonstrate this, the same model as before was used The spectral ratio method was used on this data set, with and without noise. When the full frequency range (7-25 kHz) was used to calculate Q, negative (non-physical) values of Q resulted. When the frequency range was reduced to include just the main lobe (7-12 kHz), the results are much better. Table 2 lists average Q estimates, and the standard deviation of the estimate for several different noise levels. While the spectral ratio method works for a limited bandwidth, it is unable to handle the spectral zero in the presence of any significant amount of noise.

Table 3 lists the estimated Q and the standard deviation in the estimate of Q at varying noise levels for the sorce spectrum shown in Figure 7 (over the full 7-25 kHz frequency range). Compared to the spectral ratio method, the Wiener filter method does a much better job of Q estimation when the source spectrum has zeros. However, the Wiener technique is also very sensitive to parameter choice. In order to stabilize the estimate of Q when the source spectrum contained zeros, the autocorrelation matrix (see Equation (10)) must be diagonally loaded As previously mentioned, this is approximately equivalent to adding white noise to the spectrum of the input signal. Unfortunately, the Q-estimate is very sensitive to the amount of diagonal loading If there is too little diagonal loading, then the attenuation estimate will be poor due to the previously mentioned spectral zeros. If too much diagonal loading occurs, then the errors induced by the loading itself cannot be corrected.

For the single examples, it has proven possible to set the loading and threshold parameters to get reasonably accurate answers (of course,

TABLE III

AVERAGE Q ESTIMATE AND STANDARD DEVIATION OF THE WIENER METHOD ESTIMATE FOR SOURCE FROM FIGURE 7

SNR	Average Q Estimate	Standard Deviation of Estimate
10000.	50.6	3.0
1000.	50.8	10.1
100.	51.7	27.3
10.	52.6	47.9

the Q is already known). It appears that by noticing the change in Q estimates versus the two parameters, it is possible to pick values corresponding to good estimates, by choosing parameters such that small changes in the parameters lead to very small change in Q. Unfortuately, for multi-layered models (corresponding to data from a multi-receiver tool), choosing the correct parameters becomes much more difficult. Because the Q values differ, and because the spectral shape of the signal changes as it propagates through the model, the "best" choice of the parameters is different for the different layers (corresponding to different receiver pairs from the same tool). For a "noisy" source with spectral zeros, it has not proved possible to pick the correct parameters well enough to make reliable estiamte of Q vs depth. Therefore, the Wiener filter method may not prove to be practical for "real world" well log data, and it is dropped from further consideration.

2.2 Matrix Representation of Data

Most of the techniques discussed so far use or are based upon the method of spectral ratios to actually estimate the attenuation coefficient, Q Unfortunately, as examples have shown and several authors have pointed out, the spectra of received signals typically have zeros or "nulls" in them. These zeros result in extreme instability in the calculation of Q from the slope of the spectral ratio plot. Since the signal values in the frequency range surrounding a zero is low, the noise may predominate. Furthermore, since a spectral zero at one receiver frequency means a spectral zero at an adjacent receiver, then the spectral ratio at the frequency corresponding to the zero represents the

ratio of noise to noise. Therefore, in the vicinity of a spectral zero the calculated values of the spectral ratio may be very inaccurate

Several of the papers discussed earlier contain various methods of dealing with the problem of relatively meaningless data in some spectral regions of the spectral ratio. One approach is to only use data from the spectral peak, where the signal-to-noise ratio is presumably the best. Cheng, Toksoz, and Willis (1981 and 1982) measured the amplitude decay of the spectral peak, and from this Q is estimated. This does solve the problem of spectral zeros However, the spectrum may be contaminated by noise even at the peak, and only using the peak value ignores meaningful data at other frequencies. Note the difference between this and the spectral ratio method, where all frequencies are treated equally

An effective method of treating the above problem is to weight the frequencies used in the Q-estimate according to the reliability of the data. Consider the approach by Goldberg, Kan, and Castagna (1984), where the parameters are estimated by minimizing the error between the predicted model and the data, with the error weighted by signal power The following is an alternative formulation of the problem which leads to some useful results and it allows weighting of different parts of the spectrum by different amounts.

2.3 Eigenvector Decomposition

The definition of the spectral ratio, SR(1) (Equation 8) can be generalized from 2 receivers to m receivers, and can be written as

$$Q_1(1) = \frac{X1}{2c_1} \frac{1}{-\ln\left[\frac{R_1(1)}{R_{1-1}(1)}\right]} \tag{29}$$

where, X = receiver offset, ω = frequency, c_1 = phase velocity for the strata between the depths of receivers R_1 and R_{1-1}, $R_1(\omega)$ is the received spectrum of the ith receiver, and Q_1 is the corresponding attenuation estimate. This equation only holds true if the geometrical spreading losses are previously accounted for. For the time being, all frequency independent losses (such as spreading loss and possibly fluid borehole coupling) are assumed to have been corrected for.

Equation (29) represents an estimate of Q based upon one frequency point and if ω is chosen to be the frequency of the spectral peak, then this is essentially the estimator used by Cheng, Toksoz, and Willis (1981 and 1982) However, this single data point may be contaminated by noise, and if ω corresponds to a spectral zero, then the calculated spectral ratio may contain no useful information In terms of discrete frequency, ω can be written as $\omega = j(\Delta\omega) = \omega_j$, where ω_j is the jth frequency value. Then the attenuation measurements from one shot with a multi-receiver tool can be written as a matrix A, where the (i,j) element of A is

$$a_{ij} = Q_i(\omega_j) = \frac{z\omega}{2c} \frac{1}{-\ln\left[\frac{R_i(\omega)}{R_{i-1}(\omega)}\right]} . \qquad (30)$$

In this form, each element a_j) of the matrix A represents an estimate of Q from the ith receiver pair at and the jth frequency The rows of A correspond to Q estimates for a given receiver across all frequencies, and the columns of A correspond to Q estimates for a given frequency for all the receivers If the data were perfect (no noise) and assuming the constant-Q model holds, then all of the columns would be the same Therefore A will be rank one, and the non-zero eigenvector

of the column space of A (an eigenvector of AA^T) is proportional to the identical columns of A.

Of course, any real data will have noise and the rank of A will never be one. But, since the columns should be nearly the same, then one eigenvalue of A will be much larger than the rest, and the corresponding eigenvector of the column space will be a good estimate of the columns of A. In fact, the following will show that the eigenvector of the column space of A corresponding to the largest eigenvalue is the same as the vector most nearly parallel to the columns of A Furthermore, this vector is also a minimum mean square error estimate for the columns of A.

Let us consider the singular value decomposition of an M by N real matrix A, which can be expressed as

$$A = UDV^T \tag{31}$$

where U and V are orthonormal matrices determined by the eigenvalue - eigenvector decomposition. This decomposition leads to

$$(AA^T) = U(DD^T)U^T \tag{32}$$

$$(A^TA) = V(D^TD)V^T$$

where (T) represents transpose The martix D has the general form

$$\begin{bmatrix} D_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{33}$$

where D_1 is a diagonal matrix with

$$D_1 = \text{dia}\ (d_1, d_2, \quad , d_k)$$

and 0's are null matrices of appropriate dimensions The diagonal entries in D_1 are positive square roots of the nonzero eigenvalues of

AA^T or A^TA For future use, we will assume that the diagonal values in D_1, $d_1, d_2, \ldots d_k$, are ordered. That is,

$$d_1 \geq d_2 \geq \ldots \geq d_k$$

Now, we want to define a column vector $\underline{c}$ that is most nearly parallel to the columns of A. In other words, we want to find a column vector $\underline{c}$ such that

$$E = \sum_{i=1}^{N} S\ (\underline{A}_i - \underline{c})^T\ (A_i - \underline{c}) \tag{34}$$

is minimum, where $\underline{A}_i$ is the ith column in A and $\underline{c}$ is constrained such the $\underline{c}^T\ \underline{c}$ is a constant, say one The error E can be expressed as

$$E = \left(\sum_{i=1}^{N} \underline{A}_i^T\ \underline{A}_i \right) + (N(\underline{c}^T\underline{c})) - 2 \sum_{i=1}^{N} \underline{A}_i^T\ \underline{c}$$

where the first two terms are positive. It is clear that E is minimized when the last tern is maximized. The vector $\underline{c}$ can be determined by maximizing $(\Sigma \underline{A}_i^T\ \underline{c})^2$, which can be written as

$$C = \underline{c}^T\ A\ A^T\ \underline{c} \tag{35}$$

Expand $\underline{c}$ in term of columns of U, where U is defined in Equation (31) This is,

$$\underline{c} = U\ \underline{b} \tag{36}$$

Substituting Equation (36) and (32) into Equation (35) yields

$$C = \underline{b}^T U^T U (DD^T) U^T U \underline{b}.$$

Let b_i be the ith component of $\underline{b}$. Since $\underline{c}^T\underline{c} = \underline{b}^T\underline{b}$, then choosing a $\underline{c}$ to maximize C is equivalent to choosing $\underline{b}^T = [1\ 0\ \ \ 0]$ Therefore we can state the following theorem

Theorem 1 The column vector $\underline{c}$ is the eigenvector corresponding to the largest eigenvalue d_1 of the symmetric matrix AA^T.

In the above analysis we assumed that the vectors $\underline{A}_i$ are deterministic. We can generalize this by writing the matrix A in the form

$$A = [\underline{x}+\underline{n}_1 \ \underline{x}+\underline{n}_2 \ . \ . \ \underline{x}+\underline{n}_N] \tag{37}$$

where $\underline{x}$ is an M dimensional column vector and $\underline{n}_i$, i=1,2, ..,N are M dimensional noise vectors, where the entries are from $N(0,s_i^2)$ That is, n_{ij}, the (i,j) entry in the M by N matrix given by

$$Y = [\underline{n}_1 \ \underline{n}_2 \ . \ \underline{n}_N],$$

satisfies the following n_{ij} is a white Gaussian random variable, with nij

$$E[\, n_{ij}n_{kl}] = \begin{cases} s_j^2 \text{ if } i = k \text{ and } j = l \\ 0, \text{ otherwise} \end{cases} \tag{38}$$

Note that the entries in A are really estimates of Q and also the purpose of the model in (36) for the matrix A is to model the case when the received spectra have zeros or nulls. This means that for some frequencies, the corresponding columns in A will contain data contaminated by noise, while other columns will contain more reliable data.

If A is as given in Equation (36), then finding a vector $\underline{c}$ to minimize the error $(\underline{c} - \underline{x})^T (\underline{c} - \underline{x})$ from A corresponds to finding a $\underline{c}$ to maximize $\underline{c}^T\underline{x}$ subject to the previously mentioned constraint on $\underline{c}$. This is the same $\underline{c}$ which maximizes $E(\underline{c}^TAA^T\underline{c})$, where E is the expected value operator. Therefore, the eigenvector corresponding to the largest eigenvalue of AA^T is an optimum (least squares) estimate of the deterministic vector $\underline{x}$ Since the elements of A are actually Q estimates described by Equation (32), then $\underline{x}$ is an optimum estimate of frequency independent Q values as a function of depth

This interesting approach of estimating the values of Q as a function of depth using eigenvectors is not free from problems when the data matrix is as modelled in equation (36). The problem results from the fact that the matrix, $E(AA^T)$, is not rank 1 due to the noise Let the matrix X consisting of n identical column vectors be written in the form

$$X = [\underline{x}\ \underline{x}\ .\ .\ \underline{x}] \tag{39}$$

Then from Equations (38) and (39), we have

$$AA^T = XX^T + XY^T + YX^T + YY^T$$

From this it follows that

$$E[AA^T] = XX^T + (\sum_{i=1}^{N} s_i^2)I \tag{40}$$

where I is an identity matrix of dimension M and E is the expected value operator. From (40), it follows that the rank of $E(AA^T)$ is M, and not equal to 1 The effect of the variance of the noise is to add a positive constant to the eigenvalues, but the eigenvectors remain unchanged Therefore, while the noise does change the structure of the AA^T matrix, the eigenvector corresponding to the largest eigenvalue should still be a reasonable estimate of the vector $\underline{x}$.

The eigenvector decomposition method using the eigenvector corresponding to the most significant eigenvalue can be easily and efficiently implemented by the power method (Golub and Van Loan, 1983). This simple iterative method converges to the eigenvector corresponding to the dominant eigenvalue provided the eigenvalue is larger in magnitude than the second largest eigenvalue in magnitude In fact, the ratio of the largest to second largest eigenvalue controls the rate of convergence. So, as long as the constant-Q model holds and the noise is

not too large, the algorithm converges fairly quickly to the dominant eigenvector.

This method of signal estimation was tested with a model based upon Equation (37). The results were initially worse for the eigenvector algorithm than for simply averaging the columns. The reason appeared to be because the realizations of AA^T were frequently quite different from its expected value. This resulted in the YY^T matrix not being diagonal. Furthermore, the diagonal values, while usually larger than off diagonal values, were not at all the same. To simulate data from a typical tool from only one shot, with eight receiver pairs, an eight by eight matrix was used to simulate (AA^T). Obviously, this is not a large enough sample.

The results were drastically improved when the diagonal entries in (AA^T) were modified to make the rank of the matrix as close as possible to rank one. This construction actually involves adding a diagonal matrix, say dia $(a_1,a_2,\ldots a_M)$, such that the matrix $S = (AA^T) +$ dia$(a_1,a_2,\ldots,a_M)$ is close to a rank 1 matrix. Note that this handles a more general case than the case where a_i's are equal. The diagonal entries are computed successively by using the following method First, note that the determinants of all 2x2 submatrices of a rank 1 matrix are all zero. Second, assuming that $a_1,a_2,\ldots a_{i-1}$ are computed earlier, and assuming further $a_k = 0$, $k>i+1$, compute all possible 2x2 subdeterminants of S involving s_{ii}. Clearly, these determinants will have a_i as a variable. Now compute a set of a_i to make these determinants be zero Then, use the median of that set for a_i Once all the values of a_i are computed, then the process can be repeated. This resulted in a significant improvement over simply averaging the columns.

The following derivation from Lanczos (1961) shows that $\underline{c}$, the most significant eigenvector of the column space of the M by N matrix A, is a weighted sum of the columns of A. Let

$$S = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

be a (M+N) by (M+N) matrix. Since S is a normal matrix, it has (M + N) orthogonal eigenvectors. Let

$$\underline{w} = \begin{matrix} \underline{c} \\ \underline{g} \end{matrix}$$

be the eigenvector of S corresponding to the largest eigenvalue, denoted by d. Then the eigenvalue equation is

$$S\underline{w} = d\underline{w}$$

This implies $A\underline{g} = d\underline{c}$, and $A^T \underline{c} = d\underline{g}$.

Therefore, $\underline{c}$ is a weighted sum of the columns of the matrix A, where the weights are proportional to the elements of $\underline{g}$.

Recall that $\underline{c}$ is constrained such that $\underline{c}^T \underline{c} = 1$. This is convenient since $\underline{c}$, the solution to an eigenvector is fixed in terms of its direction but not its length. In orther words, $\underline{c}$ is only determined to within an overall scale factor.

To be used as an estimate of the vector $\underline{x}$, $\underline{c}$ must be rescaled in amplitude. Idealy,

$$\underline{c}^T\underline{c}\ q = \underline{x}^T\underline{x}$$

where q is scale factor. But since $\underline{x}$ is the unknown, $\underline{x}^T\underline{x}$ must be approximated, the approximation used here is the following. Denote the average of the columns of the matrix A by the vector $\underline{a}$. Then approximate the value of $\underline{x}^T\underline{x}$ by $\underline{a}^T\underline{a}$

An alternate way to formulate the problem is to find an optimum, unbiased solution. This is accomplished by directly solving

$$A^T A\underline{g} = d^2\underline{g}$$

for $\underline{g}$. Then, rescale $\underline{c}$ to a new vector $\underline{c}'$ where

$$\underline{c}' = \frac{1}{d} A\,\underline{g}$$

If g is normalized so that

$$d = \sum_{J=1}^{N} g_J$$

where the g_J's are the elements of g, then $\underline{c}'$ would be an unbiased estimate of $\underline{x}$.

The weights for an optimal weighted sum of the columns of A can also be calculated using straight forward least squares minimization. Let

$$\underline{c} = \sum_{J=1}^{N} \underline{A}_J w_J \tag{41}$$

Then choose $\underline{w}^T = [w_1, w_2, \ldots, w_N]$ such that the square error between $\underline{c}$ and the $\underline{A}_J$ vector

$$E_1 = E[\sum_{I=1}^{M} (c_I - x_I)^2] \tag{43}$$

is minimized, where c_I and x_I are the Ith elements of the vectors $\underline{c}$ and $\underline{x}$, respectively. In order to match the assumptions used in the eigenvector decomposition estimation, an appropriate constraint on $\underline{c}^T\underline{c}$ must be made. Such a constraint could take the form $E(\underline{c}^T\underline{c}) = \underline{x}^T\underline{x}$. Adding this constraint to (38) and using the Lagrange multiplier approach would result in a new error function

$$E_2 = E[\sum_{i=1}^{M} (c_i - x_i)^2] + p[E(\underline{c}^T\underline{c}) - \underline{x}^T\underline{x}] \qquad (43)$$

where p is the Lagrange multiplier. Finding the optimum weights, $\underline{w}$, would involve setting

$$\frac{\delta E_2}{\delta w_k} = 0 \text{ for } k = 1, 2, \ldots, N \text{ and } \frac{\delta E_2}{\delta p} = 0$$

Unfortunately this leads to a system of non-linear equations for which no closed form solution exists. Therefore, the equations can only be solved numerically.

The above constraint optimization problem, due to its lack of simple solution, does not lead to an understanding of the properties of the weigted sum formulation. However, minimizing the errer E in Equation (42) with an unbaised constraint does lead to a simple solution. Further, the properties of this solution should approximate those of the constrained least-squares problem in Equation (43). The unbiased constraint is that

$$\sum_{j=1}^{N} w_j = 1 \qquad (44)$$

which guarantees $E(\underline{c}) = \underline{x}$. For this constraint, define the error function E_3 by

$$E_3 = E[\sum_{i=1}^{M} (c_i - x_i)^2] + p(\sum_{j=1}^{N} w_j - 1)$$

Setting

$$\frac{\delta E_3}{\delta w_k} = 0, \text{ for } k = 1, 2, \ldots, N$$

leads to

$$\sum_{i=1}^{M} E\left[\sum_{j=1}^{N} (a_{ij} w_j - x_i) a_{ik}\right] + p = 0$$

Since $a_{ij} = x_i + n_{ij}$, then

$E(a_{ij} \cdot a_{ik}) = x_i^2 + \sigma_k^2 \delta_{jk}$

and

$E(x_i \, a_{ik}) = x_i^2$.

This reduces to

$$\sum_{i=1}^{M} \left[\sum_{j=1}^{N} w_j (x_i^2 + \sigma_k^2 \delta_{jk}) - x_i^2\right] + p = 0$$

thus,

$$\sum_{i=1}^{M} x_i^2 \left(\sum_{j=1}^{N} W_j - 1\right) + M \quad w_k \; \sigma_k^2 + p = 0$$

But the constraint, Equation (44), reduces this to

$$w_k = \frac{-P}{M \quad \sigma_k^2} \tag{45}$$

Applying the constraint, Equation (44), to (45) gives

$$\left(\frac{-P}{M}\right) = \frac{1}{\sum_{j=1}^{N} \frac{1}{j^2}}, \quad k = 1, 2, \ldots, N \tag{46}$$

Since these weights, w_k, are a function of the column variances, σ_k^2, they cannot be used on real data because the column variance are not known. However, it is possible to estimate these variances from the data. Unfortunately, there will be errors in the estimation of the variances which will in turn effect the results in the estimation of the weights These errors will lead to less than optimal performance and may, in some cases, result in worse performance than a simple average of the columns. Equation (46) does provide useful information, however

The ideal weights can be used to calculate the expected value of the minimum square error, E_{min}, which would result from an optimum weighted sum of the columns. The error should serve as an approximate lower bound for the error from the eigenvector estimator as well as for the error from any weighted sum of the column of A. The maximum expected error of a weighted sum of columns would actually be $M\ \sigma_{max}^2$, where σ_{max} is the variance of the "worst" column. However, the error using the simple column average should serve as a reasonable upper bound for most data corresponding to the Gaussian model used here (see Equation (36)).

Substituting Equation (42) into (43) and calculating the error, E_3, results in

$$E_3 = E \sum_{i=1}^{M} [\sum_{j=1}^{N} (x_i + n_{ij})w_j - x_i]^2$$

which becomes

$$E_3 = E \sum_{i=1}^{M} [x_i(\sum_{j=1}^{N} w_j - 1) + \sum_{j=1}^{N} w_j \cdot n_{ij}]^2$$

This simplifies to

$$E_3 = M \sum_{j=1}^{N} w_j^2 \sigma_j^2 \tag{47}$$

If a simple average is computed, then w=1/N and

$$E_{avg} = \frac{M}{N^2} \sum_{j=1}^{N} \sigma_j^2 \tag{48}$$

If the optimum weights given by Equation (45) are used, then

$$E_{min} = \frac{M}{\sum_{j=1}^{N} \frac{1}{\sigma_j^2}} \tag{49}$$

According to Gimlin, Keener, and Lawrence (1982), it can be shown, using

Cauchy's inequality and the theorems of arithmetic and geometric means, that

$$\sum_{J=1}^{N} \sigma_J^2 \sum_{J=1}^{N} \frac{1}{\sigma_J^2} \geq N^2$$

This is sufficient to prove that

$$E_{min} = \frac{M}{\sum_{J=1}^{N} \frac{1}{\sigma_J^2}} \leq E_{avg} = \frac{M}{N^2} \sum_{J=1}^{N} \sigma_J^2$$

with equality occuring when $\sigma_J = \sigma$.

2.4 Results of Eigenvector Decomposition

To test the eigenvector estimator, and to compare it with simply averaging the columns, a matrix A is generated to match the model given by Equation (37). Remember from Equation (37), that the (I,J) element of the matrix A is given by

$$a_{IJ} = x_I \, n_{IJ}, \; I=1, 2, \ldots, M, \; J=1, 2, \ldots, N$$

The "signal" vector $\underline{x}$ is composed of elements which are computer generated, pseudo-random numbers from a probability density function which is unifrom on (0,1). The noise vectors, $\underline{n}_I$, are composed of random numbers which are $N(0, \sigma_I^2)$. The column variances, σ_I^2, are uniform random variables which are uniform on the interval $(0, V_{max})$, where V_{max} is the user determined maximum noise variance.

The eigenvector estimator was used to estimate the signal vector, $\underline{x}$, from the model data For each test, the total square error between $\underline{x}$ and its estimate was calculated In addition, a median estimator which calculates the median of $\{a_{I1}, a_{I2}, \ldots, a_{IN}\}$ for $I = 1, 2, \ldots, M$, and a column average estimator which calculates the average (mean) of $\{a_{I1},$

a_{12}, .. , a_{1N}} for i=1, 2, . , M are also used on the model. The total square error for these methods was calculated. Based on the values of the randomly generated variances, the weighted-sum minimum error from Equation (41) is computed for comparison with actual errors. The results are shown in Table IV For every row listed in Table IV, a signal vector $\underline{x}$ and a set of column variances are generated. Column one lists the maximum column variances for the row, which is actually the upper limit of the uniform distribution from which the column variances are chosen Once the signal vector and column variances are determined, ten sets of noise vectors are generated. For each set of noise vectors, the square errors of the previously mentioned estimators are calculated and the average square errors from the 10 sets are listed in appropriate columns. For the data shown in Table IV, the size of matrix A chosen to be 10 by 10 (M = N = 10).

Note that 3 to 4 runs were made at each listed level of maximum column variance. Each run used a different signal vector and noise (column) variances and 10 sets of noise vectors to calculate average errors. Surprisingly there is a large variation error levels. Unfortunately, this makes conclusions regarding the relative performance of the estimators difficult. There are, however, several important observations which can be made from this data. The eigenvector estimator outperformed the average for all cases except when the noise variance (maximum = 1 0) is greater than the average signal power, which is $E(s^2)$ = 1/3. Clearly, poorer estimates of optimum weights should result from large noise variances Since the eigenvector estimator implicitly uses a set of optimum weights (the vector $\underline{g}$ in equation (39)), then its performance should depart from the optimum attainable from Equation

TABLE IV

TOTAL SQUARE ERROR IN SIGNAL ESTIMATION

Maximum Noise Variance	Eigenvector Error	Average Error	Median Error	Minimum Error
0.0001	1.7×10^{-9}	1.9×10^{-9}	7.4×10^{-9}	1.3×10^{-11}
0.0001	4.5×10^{-9}	4.8×10^{-9}	4.2×10^{-9}	3.0×10^{-11}
0.0001	3.2×10^{-9}	3.7×10^{-9}	5.0×10^{-9}	2.8×10^{-11}
0.001	3.5×10^{-7}	3.8×10^{-9}	4.7×10^{-7}	1.3×10^{-7}
0.001	3.3×10^{-7}	3.8×10^{-9}	1.6×10^{-7}	2.2×10^{-8}
0.001	3.0×10^{-7}	3.4×10^{-9}	1.8×10^{-7}	5.0×10^{-8}
0.01	2.5×10^{-5}	2.8×10^{-5}	3.2×10^{-5}	1.4×10^{-5}
0.01	2.9×10^{-5}	3.3×10^{-5}	1.3×10^{-5}	2.1×10^{-6}
0.01	1.2×10^{-5}	1.3×10^{-5}	3.7×10^{-6}	1.9×10^{-7}
0.01	4.4×10^{-5}	5.0×10^{-5}	3.9×10^{-5}	7.5×10^{-6}
0.1	2.9×10^{-3}	3.2×10^{-3}	2.2×10^{-3}	6.0×10^{-6}
0.1	6.0×10^{-3}	6.6×10^{-3}	7.1×10^{-3}	1.7×10^{-3}
0.1	3.8×10^{-3}	4.3×10^{-3}	2.0×10^{-3}	9.4×10^{-4}
0.1	3.5×10^{-3}	3.6×10^{-3}	2.7×10^{-3}	1.2×10^{-4}
1.0	.60	.44	.45	.12
1.0	.66	.49	.42	3.0×10^{-2}
1.0	.29	.32	.32	.16
1.0	.34	.30	.89	1.8×10^{-3}

(49) In fact, the eigenvector methods performance generally falls well short of the optimum weighted performance However, as long as the noise variances are sufficiently smaller than the signal variance, the eigenvector technique does represent a considerable improvement over a simple column average.

The most important thing to note about the results in Table IV is that the median estimator performed about as well as the eigenvector estimator and the column average. It is, however, difficult to make definite statements regarding the performance of the median estimator. The ratio of eigenvector error to average error is fairly constant for the trials listed in Table IV, with the eigenvector method having typically about ten percent lower error But, the ratio of the median error to the eigenvector and average errors varies greatly So, a new method of analyzing the relative performance is needed. This is discussed next.

A table similar to Table IV is generated in a computer But instead of 10 sets of noise vectors generated for every choice of signal vector, $\underline{x}$, and 10 column variances, 30 test sets of noise are generated as input to the estimators. Thus the average errors that are listed in this table in the computer result from more tests and should be more reliable. Instead of only a few signal vectors and sets of column variances, 30 are generated This means that for a particular maximum noise variance, 30 different signal vectors and column variances are generated, and for each signal vector and set of column variances, 30 sets of noise are generated Unfortunately, simply averaging all the rows with a particular maximum column variance does not yield anything meaningful. This is because some sets of column variances may have

significantly higher noise levels than others even with the same maximum column variances. So, taking an average of these tests would mean that the tests which resulted in higher errors would overwhelm those with lower errors. Thus, the average performance of the median compared to the other estimators would tend to mimic that for the tests with the highest noise levels, obviously this is not good. The first step towards combining the data is to normalize all of the errors by dividing each error by the error for the column average from the same row. Then the median of the normalized errors for each estimator is used to represent that estimator's relative performance for the particular maximum column variance. The results are tabulated in Table V. In addition, a count is also kept of the number of trials for which each of the three estimators had the first or second lowest average noise levels The results of this ranking are shown in Table VI.

Since each row of Table V represents the median of 30 trial runs with 30 different sets of noise per run, and since the results are normalized so that each trial is weighted about equally, then meaningful conclusions can be made from the data Except for noise variance greater than or equal to signal variance, the eigenvector technique always yielded lower average errors than the column average. Furthermore, the eigenvector error is typically about 10 percent lower than that for the column average error However, the median estimator performed even better. A typical error from the median method is 30 percent lower than for the column average method, and it is about 20 percent lower than for the eigenvector method From the above results, it appears that the median is the superior estimator However, the

TABLE V

MEDIAN NORMALIZED ERROR FOR 30 TRIALS
WITH 30 SETS OF NOISE PER TRIALS

Maximum Noise Variance	Eigenvector Error	Average Error	Median Error	Minimum Error
0.0001	0.898	1.0	0.740	0.116
0.001	0.9000	1.0	0.674	0.057
0.01	0.913	1.0	0.696	0.114
0.1	0.898	1.0	0.693	0.112
1.0	1.135	1.0	0.683	0.110

TABLE VI

NUMBER OF TIMES (OUT OF 30 TRIALS) EACH ESTIMATE
RESULTS IN FIRST OR SECOND
LOWEST ERROR

Maximum	Eigenvector		Column Average		Median	
Noise Variance	1st	2nd	1st	2nd	1st	2nd
0.0001	12	18	0	7	18	5
0.001	5	25	1	4	25	1
0.01	11	19	0	7	19	4
0.1	7	23	0	3	23	4
1.0	1	4	0	24	24	2

results listed in Table VI show that the median is not always the best estimator.

Each entry in Table VI is the number of times, out of 30 trials, that the particular estimator had the lowest or second lowest average error. It is important to keep in mind that the average error for each trial is actually the average error for one specific signal vector and one set of column variances, averaged over 30 trials (30 sets of noise) for the given column variances. For example, suppose one reads from Table VI that for a particular one maximum noise variance, the eigenvector method had the lower average error 11 times and the median had the lower average error 19 times This should be interpreted to mean that for eleven different choices of a signal vector and column variances, the eigenvector error (averaged over 30 different sets of noise for those variances) was lower than the other 11 times and the average median error was lower 19 times.

Since each one of these trials is really the average of 30 trials on 30 noise sets, then one might say that for 19 trials, the "expected value" of the error of the median estimator is lower than that for the eigenvector method. The term "expected value" in quotation marks is really an approximation of the expected value by an average of thirty sets of noise.

For the most part, the results from Table VI are as expected given the results in Table V. However, there are a few points about Table VI which need to be stated. Even when the median of the average errors of the median estimator is lower than for all other estimators, the following is true For some noise levels, the eigenvector technique resulted in the lower average error nearly as often as the median

technique. Also, as often as 7 out of 30 times the average error of the median estimator is higher than the average errors for the sample average or eigenvector method. So, the median estimator is not as consistent as the other estimator, but it is usually better

The median outperformed the eigenvector and column average methods even though the latter two methods are supposed to be optimum for Gaussian noise (used in the tests) while the median is not This is because the answer for the average estimator is an optimum estimator of a parameter corrupted by Gaussian noise as long as the noise variance is the same for every sample point. This obviously does not apply to this case as the variances vary a great deal. The following example from Huber (1981) illustrates this point well. Let

$$F = (1-e)X + eY,$$

where $X \sim N(0, \sigma^2)$ and $Y \sim N(0, 9\sigma^2)$

Here,

$$d_N = \frac{1}{N} \sum_{i=1}^{N} |x_i|, \; s_N = \left[\frac{1}{N} \sum_{i=1}^{N} (x_i)^2 \right]^{1/2}$$

are being compared as estimates of the scatter (variance) of the random variable F For a single Gaussian random variable, s_N is the optimum estimator. Define the asymptotic relative efficiency, ARE, as

$$ARE = \lim_{N \to \infty} \frac{\dfrac{var(s_N)}{[E(s_N)]^2}}{\dfrac{var\ (d_N)}{[E(d_N)]^2}}$$

where var is the variance operator

Huber shows that for $e \leq .001$, $ARE(e) < 1$, which means that s_N is the more efficient estimator. However, for $.002 \leq E \leq .998$, d_N is the more efficient estimator. Therefore, if the data departs at all from the nominal Gaussian model, then the performance of the least-squares based estimators deteriorate rapidly compared to other types of estimators such as the median estimators. The eigenvector method does not reach the potential lower limit of error because of problems previously discussed in the Section (2.3). This leads to the search for better estimators using maximum likelihood methodology, and to the concept of robust estimators which will be discussed in the next chapter.

2.5 Chapter Summary

Many least-squares based Q estimators have been evaluated. The Wiener filter method has been dropped from consideration because it is very sensitive to the choice of parameters. Further, it is not generally possible to set the parameters so that the estimator works well on multi-layered models. The spectral ratio method works well on broad band data, but does not do well on data with a realistic spectrum. In fact, the spectral ratio method only yields reasonable results when the bandwidth of the estimation is reduced to include only the main lobe in the received spectrum, and not the spectral zero. Limiting the bandwidth to the main lobe or peak is to be avoided if possible because it reduces the number of data points over which the attenuation estimation is made.

Attenuation estimates at each frequency and for each receiver pair can be formulated in terms of matrices. This matrix formulation leads to a new model for attenuation, given by Equations (36) and (37), and

subsequently leads to new methods of Q estimation. They include the eigenvector decomposition method and the sample median and sample mean of the columns.

The median estimator is generally the best of these methods, although the eigenvector decomposition method does sometimes outperform the median method Both the eigenvector decomposition and median methods are better estimators than the average of the columns. Of these two, the median estimator is preferred because it is simpler to compute and it never fails to converge. The excellent performance of the median estimator leads to the consideration of a group of estimators known as robust estimators, since the median is a common example of a robust estimator.

CHAPTER III

ROBUST ESTIMATION

The principle attenuation algorithms discussed so far, the spectral ratio method, the Wiener filte method, and the eigenvector method, are all designed using a least-squares criterion. That is, the parameters are chosen to minimize the square error between the data and the model prediction based on the chosen parameters. The least-squares criterion is frequenctly chosen for many theoretical reasons. These reasons include (1) least-squares optimization problems generally lead to more tractable mathematics (2) filters and estimators based upon least-squares criterion frequently have nice linear properties (3) least squares estimators are maximum likelihood estimators for Gaussian noise.

There are also many reasons not to use least-squares estimators. While they are maximum-likelihood estimators for Gaussian noise, the estimators are not necessarily optimum for any type of real data Gaussian noise models are very popular for the same reasons as least-squares techniques, they lead to pleasing theoretical results For example, the maximum likelihood estimate of the data corrupted by Gaussian noise is the sample mean. In fact, according to Watt (1983), the reason Gauss chose the Gaussian noise distribution is because it led to the arithmetic mean as an optimum estimator, not because it is a good descriptor for any sets of real data However, there is one theoretically sound reason for assuming a Gaussian noise distribution,

and that is the Central Limit theorem. This theorem states that as n increases to infinity, the distribution of any linear sum of n random variables (of any continuous distribution) will approach a Gaussian distribution

The greatest weakness of least-squares estimators is they lack robustness. Estimators are generally considered to be robust if they are relatively immune to extreme outlying data points or if their performance does not deteriorate significantly when the actual noise distribution encountered differs from the assumed model. Unfortunately, least-squares estimators are extremely sensitive to outliers because they give equal weight to all of the data points. This means that one very bad data point may pull the estimate away from many good points due to the very large size of the error. In addition, as demonstrated by the example from Huber (1981) discussed in this thesis, least-squares estimators are also sensitive to deviations from the assumed noise distribution. From this point, a reasonable next step is to analyze the errors and see if the estimates are robust.

3.1 Data Modelling of the Q Estimation Problem

The model for the data matrix, A, given by equation (37) has been used as a working model for the matrix whose elements are Q estimates In this model, the Q estimates themselves are considered to be contaminated by Gaussian noise. On the other hand, it would seem more physically reasonable if the Gaussian noise was modelled as additive noise to the original signal. That is

$$r(t) = s(t) + n(t)$$

where r(t) is the received P-wave signal, n(t) is Gaussian white noise, and s(t) is the "pure" signal. Since the Fourier transform is a linear operator, it follows that

$$R(\omega) = S(\omega) + N(\omega)$$

where $R(\omega)$, $S(\omega)$, and $N(\omega)$ are the Fourier transforms of r(t), s(t), and n(t) respectively. The spectral ratio defined by Equation (8) corresponding to the received signal $r_1(t)$ and $r_2(t)$ is

$$SR(\omega) = - \ln\left[\frac{R_2(\omega)}{R_1(\omega)}\right] = - \ln\left[\frac{S_2(\omega) + N_2(\omega)}{S_1(\omega) + N_1(\omega)}\right] \qquad (50)$$

where $R_1(\omega)$ = Fourier transfer of $r_1(t)$.

It would be convenient if Equation (50) could be reduced to the form

$$SR(\omega) = SR_0(\omega) + N \qquad (51)$$

where

$$SR_0(\omega) = - \ln\left[\frac{S_2(\omega)}{S_1(\omega)}\right]$$

and N is a noise term independent of $S_1(\omega)$ and $S_2(\omega)$. If Equation (50) could be reduced to the form of (51), then a fairly simple maximum likelihood estimator could be derived from $SR(\omega)$ based upon the functional form of N (as a function of $N_1(\omega)$ and $N_2(\omega)$). Unfortunately, separation of the terms in Equation (51) is not possible and therefore the distribution of the errors in $SR(\omega)$ must be analyzed numerically using computer simulations

Because of the symmetry of the model in Equation (50), analyzing $1/SR(\omega)$, which is proportional to Q, should give the same result as

analyzing SR(ω) To implement the simulation for attenuation estimation, or equivalently analysis for the estimation of SR( ), a simple one-dimensional model is constructed. This model is similar to the ones used to test the spectral ratio and Wiener filter Q estimators earlier in this work. The model is as follows

$$r_1(t) = s_1(t) + n_1(t)$$
$$r_2(t) = s_2(t) + n_2(t)$$

where $r_1(t)$ and $r_2(t)$ are the received P-wave signals at two adjacent receivers, and n_1 and n_2 are white Gaussian noise processes such that the signal to noise ratio (SNR) is 100. Here, $SR_0(\omega)$, the ideal log-spectral ratio is given by

$$SR_0(\omega) = \frac{Z}{2Qc}. \tag{52}$$

As before, Z is the model layer thickness, c is the P-wave phase velocity, and the attenuation, Q, is assumed to be frequency independent. Then Q can be estimated from $SR(\omega)$ by

$$Q = \frac{Z}{2c\ SR(\omega)} \tag{53}$$

where $SR(\omega)$ is as defined in Equation (50). Note that a value for Q can be calculated for every frequency point. With Q = 50, and the source spectrum as shown in Figure 4a. Q is calculated for the 100 frequency points which lie within the fequency band of the source. This process is repeated for 50 different sets of noise, resulting in a total of 5000 Q estimates.

These 5000 Q estimates are put into bins and a histogram of the values is made. A plot of this histogram appears as the solid line in

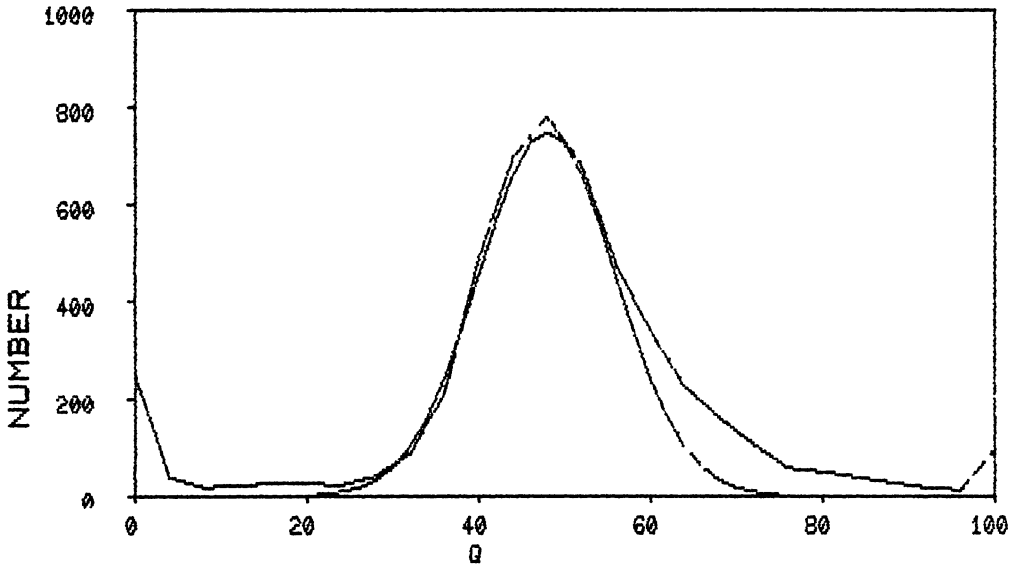

Figure 8 Histogram of Q Estimates

Figure 8. The data points at each end of the graph which appear to rise sharply simply represent the total number of values which exceed the limits of the histogram on either side. The main lobe of the graph, lying approximately between 35 and 65 on the x-axis, appears to have roughly a Gaussian shape. But it is important to determine how well the data really fits a Gaussian curve. A Gaussian curve is generated to approximately match the histogram in plot amplitude, mean, and width of the main lobe using "eyeball" fit. The variance of the histogram data is very high due to the large number of outliers (the variance = 2300). So the variance of interest is that of the Gaussian curve which is fitted to the histogram (the Gaussian is the dotted curve in Figure 7). Note that the Gaussian curve fits the main lobe of the histogram data very well except for that part of the main lobe which lies to the right of Q = 55.

The variance of the best-fit Gaussian curve is 64, so the standard deviation is σ= 8. The histogram plot deviates widely from the Gaussian in the tails, because of the large number of data points with very large deviations from the mean. For example, approximately 350 of 5000 data points lie outside the limits of the histogram That is, 350 of 5000 data points lie outside plus or minus six standard deviations from the mean To demonstrate how badly the model deviates from the Gaussian model, consider the following probabilities. The probablity of one data point from a Gaussian distribution being outside plus or minus six standard deviations from the mean is less than 10^{-8}. By using a Gaussian approximation for a binomial distribution, it can be shown that the probability of finding 350 or more of 5000 data points outside of plus or minus six standard deviations is about 10^{-4x10^8} Obviously, the

tails of this distribution are not Gaussian, though perhaps the main lobe could still be modelled as a Gaussian distribution. Clearly, an estimator which is less sensitive than least-squares estimators to these frequently occuring outliers than is needed. A more rigorous approach to deriving estimators which are optimum for a particular noise distribution is known as maximum likelihood estimation theory. The rest of this chapter deals with the development of maximum likelihood estimators and their application to some data models.

3.2 Maximum Likelihood Estimation

The description here of the fundamental ideas of maximum likelihood theory are taken mainly from Van Trees (1968), while the sections which relate to robust estimation come mainly from Huber (1981), Watt (1983), and Kassam and Poor (1985)

Bayesian estimation theory can be thought of as an extension of Bayesian detection theory. In Bayesian detection theory, the goal is to make a "good" guess about which of two possible hypothesis, H_0 or H_1, is true. Since the guess is to be as "good" as possible, a quantitative measure of the "goodness" of the guess must be used The risk R is the mathematical measure of the decision quality and the optimal detector minimizes the risk, R. Denote $Pr(H_i|H_j)$ to be the probability of the event $H_i|H_j$. That is, $P_r(H_i|H_j)$ is the probability of guessing H_i given that H_j is true (i and j are 0 or 1) Then the risk, R_1, is defined by

$$R = C_{00}P_0P_r(H_0|H_0) + C_{10}P_0P_r(H_1|H_0) + C_{11}P_1P_r(H_1|H_1) + C_{01}P_1P_r(H_0|H_1)$$

where the C_{ij} are the cost functions associated with the events $H_i|H_j$, and P_0 and P_1 are the a priori probabilities of the events H_0 and $H1$.

For Bayesian estimation the cost function is, in general, a function of the parameter to be estimated, a, its estimate, a, and the observed random variable, $\underline{x}$. Note that $\underline{x}$ may be vector valued. Most cases of interest are limited to cost functions C which are functions only of the estimation error, $a - \hat{a}(\underline{x})$ Then, the risk R can be written

$$R = E\{C[a - \hat{a}(\underline{x})]\}$$

or

$$R = \int_{-\infty}^{+\infty} da \int_{-\infty}^{+\infty} C[a - \hat{a}(\underline{x})]\, P_{\underline{a},\underline{x}}(a, \underline{x}) d\underline{x} \tag{54}$$

where $P_{a,\underline{x}}(a,\underline{x})$ is the joint probability density of the estimation parameter a and the observation, $\underline{x}$. The estimation parameter, a, is considered at this point to be a random variable However, it will later be restricted to be a non-random parameter. Common cost functions include

$$C(e) = e^2 \tag{55}$$

$$C(e) = |e| \tag{56}$$

$$C(e) = \begin{cases} 1, & |e| > \Delta/2 \\ 0, & |e| \leq \Delta/2 \end{cases} \tag{57}$$

where the error $e = a - \hat{a}(\underline{x})$, and Δ is an undefined parameter It is interesting to note that if C is chosen as in Equation (55), the square error cost function, then setting

$$\frac{\delta R}{\delta \hat{a}} = 0$$

to minimize R results in

$$\hat{a}(\underline{x}) = \int_{-\infty}^{+\infty} a\, P_{a|\underline{x}}(a|\underline{x}) da$$

Now, $a(\underline{x})$ is the conditional mean, and $P_{a|\underline{x}}(a|\underline{x})$ is the conditional probability density of a given $\underline{x}$. If C is chosen as in Equation (56), the absolute error, then

$$\int_{-\infty}^{\hat{a}(\underline{x})} da\, P_{a|\underline{x}}(a|\underline{x}) = \int_{a(\underline{x})}^{+\infty} da\, P_{a|\underline{x}}(a|\underline{x}).$$

where $\hat{a}$ is the conditional median, or the median of the a posteriori density.

The cost function in (57), known as the uniform cost function, is the most important of the three in that it leads to the maximum likelihood estimate The risk expression for this cost function leads to

$$R = \int_{-\infty}^{+\infty} d\underline{x}\, P_{\underline{x}}(\underline{x}) \left[1 - \int_{a - \Delta/2}^{a + \Delta/2} P_{a|\underline{x}}(a|\underline{x}) da \right], \tag{58}$$

where $P_{\underline{x}}(\underline{x})$ is the probability density function of the observation, $\underline{x}$ Minimizing R in (58) requires maximizing the inner integral. As Δ becomes arbitrarily small, the value of $\hat{a}$ which minimized R is the maximum of the a posteriori density, $P_{a|\underline{x}}(a|\underline{x})$. Since the natural logarithm is a monotone increasing function, and since all probabilities are non-negative, then maximizing the conditional probability is equivalent to maximizing the natural logarithm of the conditional probability density Remember that the conditional probability density is also the a posteriori density Working with the logarithm of probability densities is convenient because the natural log of $P_{a|\underline{x}}(a|\underline{x})$ often has a fairly simple form The following, known as the MAP (Maximum a Posteriori) equation is a necessary, but not sufficient, condition for locating the maximum of R.

$$\frac{\delta \ln P_{a|x}(a|\underline{x})}{\delta a}\Big|_{a = \hat{a}(\underline{x})} = 0. \tag{59}$$

For many problems of interest, the estimation parameter, a, is not actually a random variable, but is actually an unknown constant. It is possible to rewrite the maximum a posteriori equation to take this into account. From Bayes' Theorem

$$P_{a|\underline{x}}(a|\underline{x}) = \frac{P_{\underline{x}|a}(\underline{x}|a) \cdot P_a(a)}{P_{\underline{x}}(\underline{x})} .$$

Finding an $\hat{a}$ to maximize $\ln P_{a|\underline{x}}(a|\underline{x})$ is equivalent to maximizing

$$L(a) = \ln P_{\underline{x}|a}(\underline{x}|a) + \ln P_a(a)$$

since $P_{\underline{x}}(\underline{x})$ is not dependent on a Now, assuming a is no longer a random variable, then the second term can be dropped. Thus,

$$L(a) = \ln P_{\underline{x}|a}(\underline{x}|a) \tag{60}$$

According to Van Trees (1965), this corresponds to the limiting case of a maximum a posteriori estimate in which the a priori knowledge approaches zero. The log-likelihood function, L(a), or the likelihood function, $P_{\underline{x}|a}(\underline{x}|a)$, are now the functions to be maximized by the choice of a. As before, a necessary condition for maximization is

$$\frac{\delta L(a)}{\delta a}\Big|_{a=\hat{a}} = \frac{\delta}{\delta a}\left[\ln P_{\underline{x}|a}(\underline{x}|a)\right]_{a=\hat{a}} = 0 \tag{61}$$

which is known as the likelihood equation. The estimate derived from Equation (61) is known as the maximum likelihood estimate.

Analysis of the variance of maximum likelihood estimators is often difficult, but there are various bounds on the variance of the estimates. One very useful bound is known as the Cramer-Rao bound The

Cramer-Rao bound is a lower bound on the variances of any unbiased estimate, $\hat{a}(\underline{x})$, of a. It can be stated in two forms, they are

$$\mathrm{Var}[\hat{a}(\underline{x}) - a] \geq \frac{1}{E\left\{\left[\frac{\delta L(a)}{\delta a}\right]^2\right\}} \tag{62}$$

or

$$\mathrm{Var}[\hat{a}(\underline{x}) - a] \geq \frac{1}{-E\left[\frac{\delta^2 L(a)}{\delta a^2}\right]} \tag{63}$$

provided

$$\frac{\delta P_{\underline{x}|a}(\underline{x}|a)}{\delta a}$$

and

$$\frac{\delta^2 P_{\underline{x}|a}(\underline{x}|a)}{\delta a^2}$$

exist and are absolutely integrable. Here, Var is the variance operator. It can be shown that the equalities in Equation (62) and (63) hold if and only if

$$\frac{L(a)}{a} = [a(\underline{x}) - a] \cdot k(a) \tag{64}$$

where k(a) is some function of a. An estimate is said to be efficient if the equality in Equation (62) or (63) holds (e g. the estimate meets the Cramer-Rao bound). It can be shown that if an efficient estimate $\hat{a}(\underline{x})$ of a exists, then the estimate is the maximum likelihood estimate of a. However, if an efficient estimate does not exist, meaning that Equation (64) does not hold, then the only thing known about the variance of the maximum likelihood estimate is that it must exceed the Cramer-Rao bound It is important to remember that the Cramer-Rao bound

only applies to unbiased estimates, although there exists similar bounds for biased estimates.

To investigate the use of maximum likelihood estimation and Cramer-Rao bounds, consider the following simple but useful model

$$x_i = a + n_i, \quad i = 1, 2, \ldots, N \tag{65}$$

where $\underline{x} = [x_1, x_2, \ldots, x_N]^T$ is the observation vector, a is the parameter to be estimated, and n_i is a white noise process which is $N(0, \sigma_i^2)$. For this model,

$$P_{\underline{x}|a}(\underline{x}|a) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma_i} e^{-\left[\frac{(x_i-a)^2}{2\sigma_i^2}\right]}. \tag{66}$$

Then,

$$L(a) = \ln P_{\underline{x}|a}(\underline{x}|a) = \ln\left[\frac{1}{(2\pi)^{N/2}}\right] + \sum_{i=1}^{N}\left[\frac{-(x_i-a)^2}{2\sigma_i^2} - \ln(\sigma_i)\right]$$

Applying Equation (59) yields

$$\frac{\delta L(a)}{\delta a}\Big|_{a=\hat{a}} = \sum_{i=1}^{N} \frac{(x_i-a)}{\sigma_i^2} = 0 \tag{67}$$

Therefore,

$$\hat{a} = \frac{\sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \tag{68}$$

Note that if $\sigma_i = \sigma$ for $i = 1, 2, \ldots, N$, then

$$\hat{a} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

which is the familiar sample mean. Obviously, $E(\hat{a}) = a$. So, the estimate is unbiased and the Cramer-Rao bound applies. Because $\frac{\delta L(a)}{\delta a}$ from Equation (67) has the form of Equation (64), then the estimate must be efficient. Since

$$\frac{\delta^2 L(a)}{\delta a^2} = - \sum_{i=1}^{N} \frac{1}{\sigma_i^2}$$

then by Equation (63)

$$\mathrm{Var}[\hat{a}(\underline{x}) - a] = \frac{1}{\sum_{i=1}^{N} \frac{1}{\sigma_i^2}} \tag{69}$$

If $\sigma_i = \sigma$ for $i = 1, 2, \ldots, N$, then the variance simplifies to

$$\mathrm{Var}[\hat{a}(\underline{x}) - a] = \frac{\sigma^2}{N}.$$

Note that Equation (69) is the same as that given by Eqution (49) as the optimum minimum error from a weighted average Therefore, Equation (49) describes a Cramer-Rao bound.

According the Van Trees (1965), under "reasonable general" conditions the maximum likelihood estimate converges "in probability" to the correct value, a, as N approaches infinity Also, the maximum likelihood estimate is asymptotically Gaussian with mean a. Finally, even if the maximum likelihood estimate is not efficient, the estimate is asymptotically efficient. In other words,

$$\lim_{N \to \infty} \frac{\mathrm{Var}[\hat{a}(\underline{x}) - a]}{\left\{ -E\left[\frac{\delta^2 \ln p_{x|a}(\underline{x}|a)}{\delta a^2} \right] \right\}^{-1}} = 1.$$

If the maximum likelihood estimate is efficient, then no unbiased estimate with a lower variance exists. On the other hand, if the maximum likelihood estimate is not efficient, then there may exist an unbiased estimator with a lower variance Unfortunately, there is no simple rule for finding these estimators It is also possible that biased estimates exist which may have lower variances than do maximum likelihood unbiased estimates But, since they are relatively easy to

find optimum estimators, maximum likelihood estimators are very popular. They are optimum in the sense that the maximum likelihood estimate yields the estimate which is most likely to have produced the given observed parameter set.

3 3 Robust Estimation

There are three basic types of robust estimates. Two of these types will be dealt with in this thesis. The first type is called the M-estimate which has an maximum likelihood form, and may actually be a maximum likelihood estimate The second type of estimate is known as the L-estimate, which uses a linear combination of statistics That is, the L-estimate uses a linear combination of the data or a linear combination of some function of the data. Means and medians are both examples of L-estimates, and they are also M-estimates for particular noise models.

The M-estimate is the solution of an equation which can be of two forms. The first is the equation

$$e = \sum_{1} f(x_1, a) \tag{70}$$

where $\hat{a}$ is the estimate, and f is some function of the estimate and the observed data, $\underline{x}$ The estimate is found by choosing the a to minimize e in Equation (70). The alternative form is the equation

$$\sum_{1} F(x_1, a) = 0 \tag{71}$$

where

$$F(x_1,y) = \frac{\delta}{\delta y} f(x_1,y)$$

If f is of the form $f(x,y) = -\ln p(x,y)$ where p is a probability density function, then the M-estimate is a maximum likelihood estimate.

In many problems, including the attenuation estimation problem discussed here, the estimate of interest is a location estimate Thus, Equation (71) becomes

$$\sum_i F(x_i - \hat{a}) = 0 \tag{72}$$

which can be written as

$$\sum_i w_i(x_i - \hat{a}) = 0 \tag{73}$$

where

$$w_i = \frac{F(x_i - a)}{x_i - \hat{a}} . \tag{74}$$

Therefore, a can be written as

$$\hat{a} = \frac{\Sigma_i \; w_i \cdot x_i}{\Sigma_i \; w_i} \tag{75}$$

This means that the location estimate, $\hat{a}$, can be written as a weighted average of the observed data values $\{x_1, x_2, \ldots, x_N\}$ If the model is as given by Equations (65) and (66), then the estimate given by Equation (68) is of the same form as Equation (75). Note that Equation (75) has arisen without assuming a maximum likelihood estimation form.

Important properties of any estimate are the bias and the variance of that estimate A good estimator is hopefully unbiased, but the most important quality of an estimator is the variance of the estimate Obviously, the variance of the estimate should be as small as possible. As has been discussed, it is also important that the estimate be robust A good indication of the robustness of an estimator is the influence function of the estimator The influence function is a

description of the effect a particular observed data value has on the estimate. For M-estimates of location, the influence function as defined by Huber (1981) is proportional to the function F in Equation (72). If the function is bounded, then the estimate which is derived from Equation (72) is generally considered to be robust. In other words, if

$$\lim |F(t)| < \infty$$

then observed data values which are extreme outliers have only a limited role in determining the optimum estimate, a. In fact, some statisticians prefer a "redscending" influence function where

$$\lim |F(t)| = 0$$

so that extreme outliers, which may be "bad" data points, have very little influcence on the estimate

As previously stated, an estimator is robust if the estimates are reasonably efficient estimates on data which deviates from the nominal model, or if the estimate is relatively immune to outlying data points However, estimators which have bounded influence functions generally meet the other two criteria, and hence are robust. If the influence function, F(t), is linear in t, then it is obviously unbounded Since this corresponds to the sample average as an estimate, then it is obviously not robust. An estimator which has an influence function which is not bounded but which increases less rapidly than a linear function is considered to be somewhat robust. A common example of a robust estimator is the median, for which the influence function F can be written

$$F(t) = \begin{cases} -1, & t < 0 \\ 0, & t = 0 \\ +1, & t = 0 \end{cases}$$

Obviously, this influence function is bounded

The L-estimate is a linear combination of a function h of the observations. The estimate is of the form

$$\hat{a} = \sum_i w_i h(x_i)$$

Estimation using L-estimates requires finding the optimum set of weights $\{w_1, w_2, \ldots\}$ based upon a particular criterion The weights should be chosen so that the estimate is unbiased or nearly so. There is then a tradeoff between minimum variance and robustness, where the best estimator has a bounded influence function but is also fairly efficient. Both the sample mean and the median are L-estimates as well as M-estimates It is possible to find L-estimates which are as efficient as M-estimates. In fact, it can be shown that for most distributions there is an optimum L-estimate which has the same asymptotic efficiency as the optimum M-estimate. However, there is no general method for finding optimum L-estimates. Because of the difficulty of finding optimum L-estimates in general, and because the alpha-trimmed mean, defined later, has good properties, the only L-estimate under consideration will be the alpha-trimmed mean

3.4 M-estimates

Consider the data model given by Equation (37)

$$x_{ij} = a_i + n_{ij}, \quad i = 1, 2, \ldots, M, \quad j = 1, 2, \ldots, N \tag{76}$$

where the x_{ij} is an observation, a_i is the location parameter to be estimated, and n_{ij} represents independent, uncorrelated random noise The noise process, n_{ij} is such that

$$E[n_{ij} \cdot n_{kl}] = \begin{cases} \sigma_j^2, & \text{if } i=k \text{ and } j=1 \\ 0, & \text{otherwise} \end{cases} \tag{77}$$

Note that this model is essentially the same as given by Equations (37) and (38) except for notation and that no distribution has been chosen for the noise As in Equation (37), if the data have been cast as a matrix where x_{ij} is the (i,j) element of the matrix, then the noise in each column has the same variance. In this section, maximum likelihood estimators based upon the model in Equation (76) will be derived.

First, consider the case of n_{ij} being Gaussian, that is, n_{ij} is $N(0, \sigma_j^2)$. Since the standard deviations, σ_j^2, are unknown, then a maximum likelihood estimate of $\underline{a} = [a_1, a_2, \ldots, a_M]^T$ must also estimate the standard deviations $\underline{\sigma} = [\sigma_1, \sigma_2, \ldots, \sigma_m]^T$. This is the problem considered by Gimlin, Keener, and Lawrence (1982) The log-likelihood function for this model with Gaussian noise is

$$L(\underline{a}) = -\frac{MN}{2} \ln(2\pi) - \sum_{j=1}^{N} M \ln(\sigma_j) - \frac{1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{1}{\sigma_j^2}(x_{ij} - a_i)^2 \tag{78}$$

Setting

$$\frac{\delta L(\underline{a})}{\delta a_k}\Big|_{a_k=\hat{a}_k} = 0, \quad k = 1, 2, \ldots, M$$

and

$$\frac{\delta L(\underline{a})}{\delta \sigma_i}\Big|_{\sigma_i = \hat{\sigma}_i} = 0, \quad i = 1, 2, \ldots, N$$

to solve for $\hat{a}_k$ and $\hat{\sigma}_i$, the estimates of a_k and σ_i, results in the following system

$$\hat{a}_k = \frac{\sum_{j=1}^{N} \frac{x_{kj}}{\hat{\sigma}_j^2}}{\sum_{j=1}^{N} \frac{1}{\hat{\sigma}_j^2}}, \quad k = 1, 2, \ldots, M \tag{79}$$

and

$$\hat{\sigma}_i^2 = \frac{1}{M} \sum_{j=1}^{M} (x_{ij} - a_i)^2, \quad i = 1, 2, \ldots, N \tag{80}$$

There is no closed form solution to Equations (79) and (80), so the equation must be solved numerically In their paper, Gimlin, Keener, and Lawrence (1982) use a steepest ascent algorithm to maximize Equation (78) to find the optimum $\underline{\hat{a}}$ and $\underline{\hat{\sigma}}$

They did not prove that the steepest ascent algorithm will converge. They did succeed, however, in proving the existence of at least one solution to Equation (79) and (80). Gimlin, Keener, and Lawrence (1982) also reported that it was necessary to constrain the steepest ascent algorithm to keep the values of the estimate vector $\underline{\hat{a}}$ from approaching too closely to any one column of the data matrix X (with elements x_{ij}) They also discussed the uniqueness of the solution to the problem. While they were unable to give a proof of the uniqueness of a solution either to Equations (79) and (80) Gimlin, Keener, and Lawrence (1982) do state that there is strong numerical evidence that the steepest ascent method converges to a unique maximum of Equation (78). However, the form of Equations (79) and (80) suggest that an iterative technique may be used to solve the system given by Equations (79) and (80). The iterative method is as follows

1. Find the estimate $\hat{a}_k$ by using $\hat{a}_k$ = median $\{x_{k1}, x_{k2}, \quad , x_{kN}\}$, k=1,2,. ,M.
2 Using the current $\underline{\hat{a}}$, calculate $\underline{\hat{\sigma}}$ using Equation (80)
3. Calculate a new $\underline{\hat{a}}$ from Equation (79) using the current $\underline{\hat{\sigma}}$
4 Go to step (2) and repeat as necessary

While it has not been proven that this iterative algorithm

converges, it seems reasonable that it should for many cases. Assume first that the data is noisy, but good enough so that the median in step 1 gives a reasonable estimate of $\underline{\hat{a}}$. If we assume that the initial estimate of $\underline{\hat{a}}$, is close enough to $\underline{a}$ so that the variance estimates given by Equation (80) yield higher variance for "bad" data columns relative to "good" columns, then the next guess for $\underline{\hat{a}}$ should be closer. The estimate $\underline{\hat{a}}$ of $\underline{a}$ in Equation (79) is a weighted sum of the columns of the data matrix X. This can be shown by rewriting Equation (79) as

$$\hat{a}_k = \sum_{J=1}^{N} X_{kJ} W_J \quad , k=1,2, \ . \ ,M \tag{81}$$

where

$$w_J = \frac{\frac{1}{\sigma_J^2}}{\sum_{p=1}^{N} \frac{1}{\sigma_J^2}} \quad , J = 1, 2, \ .., N. \tag{82}$$

If the jth column is noisier than the kth column, then most of the time, $\sigma_J 2 > \sigma_k 2$. Therefore in the updated estimated $\underline{\hat{a}}$, the noisier jth column contributes less than the "cleaner" kth column. Thus the new estimate $\underline{\hat{a}}$ should be a better approximation for $\underline{a}$ than the previous estimate, leading in turn to even better estimates of $\underline{\hat{\sigma}}$. So, while convergence has not been proven, it seems plausible that this iterative algorithm should converge

Maximizing the log-likelihood function given by Equation (78) is equivalent to minimizing

$$H(S) = \sum_{J=1}^{N} M \ln(S_J + M) + \frac{1}{2} \sum_{I=1}^{M} \sum_{J=1}^{N} \frac{(x_{IJ} - S_I)^2}{S_J^2 + m} \tag{83}$$

where

$$\underline{S} = [a_1, a_2, \ . \ , a_M, \sigma_1, \sigma_2, \quad , \sigma_N]^T$$

is the unknown column vector of length (M + N) The initial guess necessary for the steepest descent method is found by following the first two steps for the iterative solution to this problem That is, estimate $\underline{\hat{a}}$ from the median or mean of the observed data points and use this estimate for $\underline{a}$ to estimate the standard deviations, $\underline{\hat{\sigma}}$. Locating the minimum of $H(\underline{S})$ in Equation (83) using the steepest descent algorithm involves evaluating the gradient of $H(\underline{S})$ for the previous guess Then a search is made, in the direction of the negative gradient, for the minimum $H(\underline{S})$ This means that a parameter t is chosen to minimize $H(\underline{S}^{(i)} - t\underline{u})$ where $\underline{S}^{(i)}$ is the ith guess at the solution and $\underline{u}$ is the gradient of H at $\underline{S}^{(i)}$. Then the guess is updated according to

$$\underline{S}^{(i+1)} = \underline{S}^{(i)} - t^* \quad \underline{u}$$

where t^* is the scale factor which produces the desired minimum in H

It is interesting to note from the form of Equation (83) that the function H to be minimized is not just the weighted sum of the square errors. There is a second term,

$$\sum_{J=1}^{N} M \ln(S_J + M)$$

and the purpose of this term is to guarantee that the estimate which minimizes H in (83) is an unbiased estimate. Unfortunately, this minimization problem has an instability which can easily occur Remember that finding the zero of the gradient, or the minimum of $H(\underline{S})$, corresponds to simultaneously solving Equation (79) and (80). If the $\underline{\hat{a}}$ portion of some guess $\underline{S}$ happens to lie too close to an observed data vector

$$\underline{X}_I^T = [X_{1I}, X_{2I}, \quad , X_{MI}]$$

then the sum in Equation (80) will be very small and so will be $\hat{\sigma}_l^2$. This means that the weight, w_l, (Equation (81)) on the lth column for the next estimate of $\underline{a}$ will be much higher on the next iteration, pushing the estimate even farther away The result is convegence to the data vector which is not generally the optimum solution. This does happen for the iterative method. The gradient of H, $\underline{f}$, in Equation (83) is given by

$$f_k = \begin{cases} \sum_{J=1}^{N} \dfrac{(S_k - x_{kJ})}{S_{J+M}^2}, & k=1,2,\ldots,M \qquad (84) \\ \dfrac{M}{S_k} - \dfrac{1}{S_k^3} \sum_{i=1}^{M} (X_{i,k-m} - S_i)^2, & k = m+1,\ldots,M+N \qquad (85) \end{cases}$$

While it is not clear from Equations (84) and (85) that the steepest descent may converge to a data vector, in practice it occassionally will unless the algorithm is constrained so as not to approach to closely to the data vector.

The Cramer-Rao bound for this model must be calculated using Fisher's information matrix, but is has the same form as Equation (69) Since this problem cannot be expressed in the form given by the vector equivalent of Equation (64), then the estimate is not efficient and will not reach the Cramer-Rao bound This vector case maximum likelihood estimator is not actually a robust estimator by most definitions of the term. However, it does have some robust-like properties. If one data vector is particularly "bad", then it will still have a significant effect on an initial estimate (if the initial estimate is done using an average). As the steepest descent converges, the variance estimate of the "bad" vector increases, and the

corresponding contribution to the gradient decreases. This can be seen from Equations (84) and (85). Thus, the "bad" data vector contributes less and less to the estimates as the method converges. Since "bad" data points contribute less than "good" data points, this method can be considered to be somewhat robust.

A wider-tailed noise distribution, which is more likely to produce "bad" data points than the Gaussian, and, which results in medians as maximum likelihood estimates is the Laplacian or double-exponential distribution. The probability density function for Laplacian noise has the form

$$P(x) = \frac{b}{2} e^{-b|x|} .$$

Consider the previously studied model

$$X_i = a + n_i, \quad i=1, 2, \ .., N$$

where x_i is the observed data, n_i is Laplacian noise, and a is the parameter to be estimated. The maximum likelihood estimate of a, $\hat{a}$, is

$$a = \text{median } \{x_1, x_2, \ . \ , x_N\}$$

Given a vector model for the data given by Equations (76) and (77), with a Laplacian noise process, the conditional probability density is given by

$$P_{x|\underline{a}}(x|\underline{a}) = \prod_{i=1}^{M} \prod_{j=1}^{N} (\frac{b_j}{2}) e^{-b_j|x_{ij} - a_i|} \tag{86}$$

This leads to a log-likelihood function given by

$$L(\underline{a}) = \frac{-MN}{2} \ln(2) + M \sum_{j=1}^{N} \ln(b_j) - \sum_{i=1}^{M} \sum_{j=1}^{N} b_j|x_{ij} - a_i| \tag{87}$$

Setting

$$\frac{\delta L(a)}{\delta a_k} \Big|_{a_k = \hat{a}_k} = 0, \quad k = 1, 2, \quad , M$$

and

$$\frac{\delta L(a)}{\delta b_l} \Big|_{b_l = \hat{b}_l} = 0, \quad l = 1, 2, \ldots, N$$

to maximize $L(\underline{a})$ results in the following system of equations, which is analygous to Equation (79) and (80) for Gaussian noise,

$$\sum_{j=1}^{N} b_j \cdot \operatorname{sgn}(x_{kj} - a_k) = 0, \quad k = 1, 2, \ldots, M \tag{88}$$

$$b_l = \frac{M}{\sum_{i=1}^{M} |x_{il} - a_l|}, \quad l = 1, 2, \quad , N \tag{89}$$

Here, sgn(x) is the sign operator defined as

$$\operatorname{sgn}(x) = \begin{cases} 0, & x=0 \\ -1, & x<0 \\ +1, & x>0 \end{cases}$$

Note that from Equation (88), $\hat{a}_k$ is a weighted median of data points from the kth Equation row of the matrix X, and b_l is inversely proportional to an L_1 estimate of the scatter of the data values.

This system can be solved in the same manner as that for Gaussian noise, using either a steepest ascent method on the function $L(\underline{a})$ or by an interative method using Equations (88) and (89). When an estimate approaches too closely to a data vector, both methods demonstrate the same type of instability as did the Gaussian noise based methods Again, the steepest ascent must be constrained to stay away from the data vectors. Due to the non-linear nature of this median-type estimator, a lower bound on the variance of the estimate has not been calculated

3 5 L-Estimates

An L-estimate is defined as being an estimate which is a sum of order statistics. This means that, an L-estimate is a sum of the observed data values or a sum of a function of the data values The sample mean and median are both examples of L-estimates A useful L-estimate which is not also a maximum likelihood estimate is the alpha-trimmed mean. The alpha-trimmed mean is an optimum L-estimate for the "least-informative" distribution given by the following density function (Huber, 1981)

$$f(x) = \begin{cases} C_1 \, e^{-X^2/2} \, , & |x| \leq c \\ C_1 \, e^{-c|x| + c^2/2}, & |x| > c. \end{cases}$$

The alpha-trimmed mean of a set of data values $Y = \{Y_1, \ . \ , Y_N\}$ can be found in the following manner. Let a set $Z = \{Z_1, \ . \ , Z_N\}$ be an ordering of the elements of a set Y such that $Z_1 \leq Z_2 \leq .. \leq Z_N$ Then the alpha-trimmed mean of the set Y is the average of the elements of the set Z less k data points on each end. In other words, $\hat{Y}$, the alpha-trimmed mean of the set Y is given by

$$\hat{Y} = \frac{\sum_{i=k+1}^{N-k} Z_i}{N - 2k} .$$

The number of points trimmed off each end of the set Z is related to the trimming parameter α by $k = \text{int}(\alpha N)$, where int(x) denotes the largest integer less than or equal to x

The alpha-trimmed mean has many useful properties. This estimator retains the robust features of the median because extreme outliers or "bad" data points only influence the estimate in that they help to determine which other data points contribute to the estimate. Yet the

alpha-trimmed mean can be more efficient than the median for distributions close to Gaussian. In fact, the alpha-trimmed mean may be a very good estimator in a situation where the noise distribution is primarily Gaussian but is contaminated by a longer-tailed distribution such as the Laplcaian distribution. In this case, the few outliers from the longer-trailed distribution will be trimmed away and the rest will be averaged. Thus for a Gaussian distribution contaminated by a Laplacian distribution, the alpha-trimmed mean may be expected to outperform maximum likelihood estimates based upon either pure Laplacian or Gaussian noise models.

3.6 Results

Consider a model of the form given by Equation (76). In this case, though, the noise n_{IJ} is a random process given by

$$n_{IJ} = (1 - e)\, ng_{IJ} + (e)\, ne_{IJ} \tag{90}$$

where ng_{IJ} is a white Gaussian noise process, ne_{IJ} is a white Laplacian noise process, and e is a contamination parameter. Thus n_{IJ} is a Gaussian noise process which is contaminated by Laplacian noise This contaminated distribution is used because it is a good model for a noise distribution shown in Figure 8

An experiment is conducted to test estimators on the model given by Equation (76), where the noise model is given by Equation (90) This experiment is essentially the same as that one described in the previous chapter which produced the results shown in Tables V and VI The primary differences between this experiment and the one previously described are that this experiment uses contaminated noise with varying

TABLE VII

MEDIANS OF NORMALIZED AVERAGE NOISE VARIANCES - LAPLACIAN CONTAMINATION

Contamination	Max Noise Variance	Alpha-Trimmed Mean	Average	ML Laplacian	ML Gaussian	Median	Normalization
0 0	0 001	73	1 0	39	28	74	$4\ 0\times10^{-7}$
0 0	0 01	78	1 0	37	27	66	$3\ 6\times10^{-5}$
0 0	0 1	78	1 0	40	31	67	$3\ 2\times10^{-3}$
0 0	0 5	73	1 0	40	27	69	$7\ 9\times10^{-2}$
0 01	0 001	73	1 0	46	32	74	$3\ 9\times10^{-7}$
0 01	0 01	69	1 0	34	29	62	$4\ 2\times10^{-5}$
0 01	0 1	73	1 0	33	21	63	$3\ 2\times10^{-3}$
0 01	0 5	74	1 0	40	29	72	$9\ 9\times10^{-2}$
0 1	0 001	57	1 0	51	52	57	$5\ 1\times10^{-6}$
0 1	0 01	70	1 0	50	39	67	$7\ 8\times10^{-5}$
0 1	0 1	74	1 0	41	34	73	$3\ 3\times10^{-3}$
0 1	0 5	52	1 0	.24	17	45	$8\ 3\times10^{-2}$
0 5	0 001	55	1 0	51	59	55	$1\ 2\times10^{-4}$
0 5	0 01	52	1 0	44	45	52	$1\ 3\times10^{-3}$
0 5	0 1	60	1 0	57	58	60	$1\ 4\times10^{-2}$
0 5	0 5	67	1 0	54	51	67	$8\ 5\times10^{-2}$
0 9	0 001	52	1 0	45	46	52	$4\ 5\times10^{-4}$
0 9	0 01	50	1 0	48	51	48	$3\ 6\times10^{-3}$
0 9	0 1	51	1 0	47	50	51	$3\ 8\times10^{-2}$
0 9	0 5	51	1 0	44	46	50	$1\ 8\times10^{-1}$
1 0	0 001	56	1 0	51	54	56	$4\ 1\times10^{-4}$
1 0	0 01	53	1 0	49	58	54	$4\ 9\times10^{-3}$
1 0	0 1	55	1 0	51	55	56	$5\ 2\times10^{-2}$
1 0	0 5	56	1 0	49	52	53	$2\ 6\times10^{-1}$

amounts of contamination and that different estimators are being evaluated. In this case, the estimators are the following sample average, sample median, alpha-trimmed mean, a maximum likelihood estimator based upon a Laplacian noise model, and a maximum likelihood estimate based upon a Gaussian noise model. Also, to reduce computation time, only 20 sets of signal vectors and noise variances were generated and for each set 20 different sets of noise were generated. As was the case for the previous experiment, both M and N were chosen to be ten to approximate what one might expect from real data.

As previously mentioned in Section 3.1, the most important measure of an estimator is the variance or error produced by the estimtor. So, the quality of the estimtor is measured by the variance of the final esimate. Since the actual noise distribution and signal-to-noise ratios in real data are unknown, it is necessary to test the proposed estimators on a wide range of contamination levels and over a wide range of signal-to-noise ratios.

Table VII shows the results of this experiment run for maximum noise variances of 0.001, 0.01, 0.1, and 0.5, and for contamination levels (e) of 0., .01, 0.1, 0.5, 0.9, and 1 0. Remember that the signal vector is generated from a uniform distribution on (0,1). So, a maximum column noise variance of 0.5, resulting in an "expected" column variance of 0.25 is more than the signal vector variance of 0.083 But, the "expected" column variance of 0 05 from a maximum variance of 0.1 is less than the expected signal variance.

Since a table corresponding to Table IV is not shown, and since all of the errors given in Table VII are medians of normalized average errors, there needs to be another column to give an indication of the

absolute level of the errors, rather than just relative levels of error. For this purpose, a separate column, labelled Normalization, is added. The normalization is the median of the error values used to normalize the data in Table VII. Multiplying the normalization values by the error values on the same row of the table would give an example of typical error values for the various method.

From Table VII, many observations can be made regarding the relative efficiency of the estimators. Since all of the normalized average errors are less than one, and since the errors were normalized to that for the sample average method, then all of the other methods consistently outperformed the sample average. Thus the sample average should be dropped from further consideration as an estimator on these models.

As one might expect, the maximum likelihood estimators based upon pure Gaussian or pure Laplacian noise outperformed all others when the noise is pure Gaussian (e=0) or pure Laplacian (e=1.0). The surprising thing about the comparison between the two maximum likelihood estimates is that even for noise which is mostly Laplacian, the performance of the maximum likelihood Gaussian estimator is still close to that for the maximum likelihood Laplacian estimator. It is interesting to note from a comparison of the normalization values for different levels of contamination that the performance of all of the estimators became worse as the amount of contamination became larger This is in spite of the fact that the overall noise variance remained unchanged.

The trimming parameter for the alpha-trimmed mean was adjusted to provide optimum performance for each case. In terms of average error, there is not a significant difference between the performances of the

alpha-trimmed mean and the median estimators on the data This is somewhat surprising for the data has been contaminated with primarily Gaussian noise One would think that the alpha-trimmed mean would be a somewhat better estimator than the median but it is not. In fact, the opposite is true. The alpha-trimmed mean seems to do slightly worse than the median for Gaussian data, but it does about as well as the median for primarily Laplacian noise. Overall, the performance of the alpha-trimmed mean and median estimators is definitely worse than that of the two maximum likelihood estimators when the noise is primarily Gaussian. But, the simpler to compute median and the alpha-trimmed mean do nearly as well as maximum likelihood estimators when the noise contains 10% or more Laplacian noise

Table VIII offers a different viewpoint of the results from the previously described experiment, and is analagous to Table VI in the previous chapter. That is, Table VIII lists the total number of times (out of 20 tests) the various estimators had the 1st, 2nd, or 3rd lowest average errors.

There is nothing particularly surprising about the results in Table VIII considering the results in Table VII. The relative performances of the two ML estimators as shown in Table VIII are as expected from the data in Table VII. While there were a few cases when for nearly pure Gaussian noise the sample average did well, the overall performance of the sample average was poor. In fact, the sample average nearly always had the highest average error. Probably the only observation to be made regarding the relative performance of the estimators for Table VIII which could not have been made from Table VII is the following When the noise is half or more Laplacian, $e \geq 0.5$, the alpha-trimmed mean

TABLE VIII

NUMBER OF TIMES THE AVERAGE ERRORS ARE FIRST, SECOND, OR THIRD LOWEST-LAPLACIAN CONTAMINATION

Contamination e	Max Noise Variance	Alpha-Trimmed Mean			Average			Median			ML Laplacian			ML Gaussian		
		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0	0 001	0	0	5	0	1	3	0	0	12	0	19	0	20	0	0
0	0 01	0	0	6	0	0	1	0	0	13	0	20	0	20	0	0
0	0 1	0	0	4	1	0	2	0	0	14	0	19	0	19	1	0
0	0 5	0	0	5	0	0	3	0	0	12	0	20	0	20	0	0
0 01	0 001	0	0	10	1	0	1	0	0	9	0	19	0	19	1	0
0 01	0 01	0	0	3	1	0	3	0	0	14	0	19	0	19	1	0
0 01	0 1	0	0	4	0	0	1	0	0	15	0	20	0	20	0	0
0 01	0 5	0	1	7	0	0	1	0	0	11	0	19	1	20	0	0
0 1	0 001	2	6	6	0	0	0	3	3	7	6	8	6	9	3	1
0 1	0 01	1	2	8	0	0	1	0	0	9	3	14	2	16	4	0
0 1	0 1	0	0	7	0	1	0	0	0	13	1	18	0	19	1	0
0 1	0 5	0	0	4	0	0	0	0	0	16	0	20	0	20	0	0
0 5	0 001	4	5	5	0	0	0	2	8	8	14	0	5	0	7	2
0 5	0 01	4	2	5	0	0	0	1	6	10	10	5	5	5	7	0
0 5	0 1	2	3	6	0	0	0	1	4	10	12	5	3	5	8	1
0 5	0 5	2	0	11	0	0	0	0	0	9	3	15	0	15	5	0
0 9	0 001	5	0	6	0	0	0	1	7	10	10	6	4	4	7	0
0 9	0 01	3	3	6	0	0	0	2	4	10	10	7	3	5	6	1
0 9	0 1	2	4	6	0	0	0	2	5	8	10	7	3	6	4	3
0 9	0 5	7	2	1	0	0	0	2	5	12	9	4	4	2	9	3
1 0	0 001	3	5	8	0	0	0	4	5	6	7	7	6	6	3	0
1 0	0 01	5	4	4	0	0	0	0	6	12	14	3	3	1	7	1
1 0	0 1	6	3	6	0	0	0	2	8	8	11	4	5	1	5	1
1 0	0 5	2	3	5	0	0	0	1	5	11	10	7	2	7	5	2

estimator had the lowest average error much more often than did the median and about as often as did the maximum likelihood (Gaussian) method. This is in spite of the fact that from Table VII, the median estimator average error for the alpha-trimmed mean is about the same as for the median and higher than that for the maximum likelihood (Gaussian) estimator. Therefore the alpha-trimmed mean is a better estimator than indicated in Table VII.

3.7 Chapter Summary

Overall, there seems to be no clear winner among the estimation methods discussed here. The sample average is obviously the clear loser. The maximum likelihood estimators outperform the others, but at a significant computational cost. Further, while the steepest descent maximum likelihood estimators generally converge in about 5 iterations, they do occasionally fail to converge and the estimates appear to bounce around between data column vectors. Hence, the steepest descent maximum likelihood estimators may not prove to be reliable for real data.

There are many advantages to the sample median and the alpha-trimmed mean. They can be calculated easily and they have no convergence problems. Furthermore, for levels of contamination of Gausssian noise by Laplacian noise which are greater or equal to 10%, the median and the alpha-trimmed mean performed nearly as well as the maximum likelihood methods. Therefore, the median and the alpha-trimmed mean may be the best estimators for use on real data, although the maximum likelihood method will also remain under consideration.

CHAPTER IV

ATTENUATION ESTIMATION FROM REALISTIC MODELS AND REAL DATA

It is important to model the effects of additive noise in the original input signal on the attenuation estimate. This results in a more accurate model of the types of noise (or errors) which might be present in the final attenuation estimates Also, the robust estimators described in the previous chpater should be tested on the same models used to test the spectral ratio and Wiener filter methods.

To truly evaluate the accuracy of the various attenuation estimation methods, they should be tested on realistic three dimensional model data. This step is very important because borehole geometry may strongly effect the estimates, and borehole geometry has not previously been taken into account Of course, the best attenuation estimation methods must be tested on real data. Unfortunately, the results are very difficult to interpret This is because the actual values of Q are completely unknown, so no conclusion can be made regarding the accuracy of the techniques.

4 1 Noise Models

The results in Figure 7 showed that the contaminating noise (or errors) in the attenuation problem has a long-tailed distribution This justifies the use of robust estimators. For the model given by

Equations (66) and (67), with Laplacian contaminated Gaussian noise, it has been shown that the sample average is a very poor estimator compared to some other estimators Since the spectral ratio method is a least-squares method which does not attempt to weight the data values, the performance of the spectral ratio method is similar to that of the sample average. Therefore if the model given by Equation (66) is very realistic, then these robust methods should yield more reliable estimates of attenuation than the spectral ratio method.

It seems physically more reasonable to assume that additive noise is present in the original input signal than at some later point in the analysis. Furthermore, if additive Gaussian noise is present in the time domain, then the noise in the frequency domain will also be additive Gaussian noise. In addition, it appears that much of the noise present in any seismic signal is not simply additive, independent, white noise, but is a noise process that is very much correlated with the signal. In fact, the noise may take the form of a convolution between the signal and a noise that can be characterized. Therefore the following noise model is proposed

$$R_1(\omega) = S_1(\omega) + N_1(\omega) \tag{91}$$

$$R_2(\omega) = E_2(\omega) \cdot S_2(\omega) + N_2(\omega) \tag{92}$$

Here $R_1(\omega)$ and $R_2(\omega)$ are the amplitude coefficients of the Fourier Transforms of the time domain signals $r_1(t)$ and $r_2(t)$ The signals $r_1(t)$ and $r_2(t)$ are the compressional wave forms recorded at adjacent receivers for the same shot.

In order to model the noise in the attenuation estimation problem, it is necessary to analyze the effect that noise terms in Equations (91)

and (92) have on the actual attenuation estimate The following leads to an approximate model for the noise or error in the Q estimate which results from the model given by Equations (91) and (92)

Substituting $R_1(\omega)$ and $R_2(\omega)$ from Equation (91) and (92) into the Equation (66) to find $SR(\omega)$ yields

$$SR(\omega)=-\ln\left[\frac{S_2(\omega)}{S_1(\omega)}\right] - \ln\left[E_2(\omega)\right] - \ln\left[1 + \frac{N_2(\omega)}{E_2(\omega)S_2(\omega)}\right] + \ln\left[1 + \frac{N_1(\omega)}{S_1(\omega)}\right] \quad (93)$$

Let $SR_0(\omega)$ be the ideal log-spectral ratio without noise and

$$SR_0(\omega) = -\ln\left[\frac{S_2(\omega)}{S_1(\omega)}\right] .$$

Then, expand the logarithms and assume that the additive noise terms are very much smaller than the signal terms. In other words, assume that

$$N_1(\omega) << S_1(\omega)$$

and

$$N_2(\omega) << S_2(\omega).$$

Then,

$$SR(\omega) = SR_0(\omega) - \ln\left[E_2(\omega)\right] - \frac{N_2(\omega)}{E_2(\omega)S_2(\omega)} + \frac{N_1(\omega)}{S_1(\omega)} \quad (94)$$

Now, define the total noise term $N_t(\omega)$ by

$$N_t(\omega) = SR(\omega) - SR_0(\omega).$$

An experiment is conducted to test estimators on the model given by Equation (66), where the noise model is given by Equations (91) and (92). This experiment is essentially the same as the described in the previous chapter which produced the results in Tables V and VI. The elements of the signal vector $\underline{a}$ are computer generated random numbers which have a uniform distribution on the interval (0,1) Each element

of the variance vector, $\underline{s}$, is the variance of the noise in the corresponding column of the data matrix X The variance are also computer generated random numbers which are uniform on the interval (0,MNV), where MNV is the maximum noise variance parameter. For a particular given maximum noise variance, MNV, and contamination parameter, e, 30 different signal and variance vectors are generated. Then,

$$N_t(\omega) = -\ln\left[E_2(\omega)\right] - \frac{N_2(\omega)}{E_2(\omega)S_2(\omega)} + \frac{N_1(\omega)}{S_1(\omega)} \tag{95}$$

Since Q is inversely proportional to $SR(\omega)$, then Q is proportional to

$$\frac{1}{SR(\omega)} = \frac{1}{SR_0(\omega)\left[1 + \frac{N_t(\omega)}{SR_0(\omega)}\right]} = \frac{1}{SR_0(\omega)} - \frac{N_t(\omega)}{[SR_0(\omega)]^2}$$

Hence, Q is proportional to the $N_t(\omega)$ given by Equation (95). Unfortunately, the signal terms do not separate from the noise in this problem. However, if the signal terms, $S_1(\omega)$ and $S_2(\omega)$, are treated as constants for the analysis, then the effect of the noise terms (independent of the signal) can be investigated. Then the estimate of Q is contamined by three noise terms which are proportional to $\ln\left[E_2(\omega)\right]$, $N_2(\omega)/E_2(\omega)$, and $N_1(\omega)$

In order to gain some insight into the noise model, consider the form of the three contaminating terms if $N_1(\omega)$ and $N_2(\omega)$ are $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively, and if $E_2(\omega)$ is $N(0, \sigma_e^2)$ The distribution of the first term is given by

$$f_1(z) = \frac{e^z}{\sqrt{2\pi}\,\sigma e}\left[e^{-\left(\frac{e^z-1}{2\sigma^2}\right)^2} - e^{-\left(\frac{e^z+1}{2\sigma^2}\right)^2}\right]$$

This distribution is approximately zero mean, asymetric, and has second and fourth moments which are comparable to or less than those of the process $E_2(\omega)$. This distribution was studied primarily using computer modelling. The second term has a distribution given by

$$f_2(z) = \left[\frac{2\sigma_2^2\,\sigma_e^2}{z^2\,\sigma_e^2 + \sigma_2^2}\right] \exp -\left[\frac{1}{\sigma_e^2}\right] \exp \left[\frac{\sigma_2^2}{2(z^2\sigma_e^2 + \sigma_2^2)}\right] \qquad (96)$$

Note that as $z \to \infty$, $f(z)$ becomes proportional to the Cauchy distribution given by

$$g(z) = \frac{\sigma_x/\sigma_y}{z^2 + \sigma_x^2/\sigma_y^2} \qquad (97)$$

The term N_1, is of course Gaussian as assumed.

The Cauchy distribution given by Equation (97) is the distribution of the random variable z where $z=x/y$, x is $N(0, \sigma_x^2)$, and y is $N(0,\sigma_y^2)$ The Cauchy distribution is a symmetric distribution with a maximum of

$$g(z) = \frac{\sigma_y}{\pi\sigma_x}$$

at z=0. However, the tails die off so slowly that the integral defining the second moment, $E(z^2)$, does not converge. This could be interpreted as meaning the Cauchy distribution has infinite variance. In addition, the integral defining the mean, E(z), also fails to converge. Since the mean can be thought of "physically" as the center of mass of the distribution, then the failure of E(z) to converge could be interpreted to as meaning the tails are so long that the distribution could be "balanced" anywhere (or perhaps, nowhere)

Since the second term, with distribution given by Equation (96), is asymptotically Cauchy, the distribution also has "infinite" variance and

the distribution is very wide-tailed Now, the noise terms appear to be two fairly well behaved terms plus one that has very wide tails The sum of these can be approximated as a Gaussian distribution contaminated by a Cauchy distribution.

To test the effectiveness of the previously discussed robust estimators, consider a repeat of the previous experiment (results in Table VI, VII and VIII) with the following changes. First, the contaminating noise is Cauchy rather than Laplacian, and the standard deviation of the Cauchy noise is approximated by the analagous Cauchy parameter given by (σ_x/σ_y) from Equation (97). Secondly, since the sample mean has proved to be a very poor robust estimator and since the Cauchy noise should make things "worse", the sample mean was not used as an estimator. Rather, it is replaced by a maximum likelihood estimator based upon Cauchy noise and the model from Equation (67). This maximum likelihood estimator is implemented using a steepest descent algorithm similar to that used for the other maximum likelihood estimates The average errors are normalized to the average error for the median estimator, since the sample average is not used as an estimator in this experiment. The results from this experiment are presented in Tables IX and X which are analogous to Tables VI and VII from the previous experiment

As was the case for Laplacian contamination, when the noise is primarily Gaussian ($e \leq 0.1$) the best estimator is clearly the maximum likelihood Gaussian estimator The second best estimator for mostly Gaussian noise ($e \leq 0\ 1$) is the maximum likelihood Laplacian estimator These conclusions are supported by data from both Tables VIII and IX. However, when the contamination is large ($e \geq 0\ 5$), the

TABLE IX

MEDIAN OF NORMALIZED AVERAGE ERRORS FOR
CAUCHY CONTAMINATION

Contamination e	Max. Noise Variance	Alpha-Trimmed Mean	ML Cauchy	ML Laplacian	ML Gaussian	Median	Normalization
0.0	0.001	1.07	1.82	1.82	1.82	1.0	2.3×10^{-7}
0.	0.01	1.12	5.88	.521	.413	1.0	3.8×10^{-5}
0.	0.1	1.06	6.30	.617	.396	1.0	2.6×10^{-3}
0.	0.5	1.03	1.62	.798	.378	1.0	5.7×10^{-2}
0.01	0.001	1.03	2.93	.764	.496	1.0	1.7×10^{-7}
0.01	0.01	1.05	6.30	.555	.446	1.0	3.6×10^{-5}
0.01	0.1	1.06	5.19	--	.433	1.0	2.9×10^{-3}
0.01	0.5	1.07	4.67	--	.423	1.0	8.1×10^{-2}
0.1	0.001	1.04	9.64	.967	.892	1.0	2.1×10^{-7}
0.1	0.01	1.05	13.9	.731	.609	1.0	1.2×10^{-5}
0.1	0.1	1.04	21.0	.800	.685	1.0	1.8×10^{-3}
0.1	0 5	1.14	5.26	--	.517	1.0	6.4×10^{-2}
0.5	0 001	1.10	6.59	3.57	2.48	1.0	1.7×10^{-7}
0.5	0.01	1.11	9.07	1.94	2.48	1.0	2.2×10^{-5}
0.5	0.1	1.05	24.1	1.80	2.31	1.0	2.9×10^{-3}
0 5	0.1	1.14	7.93	--	1.88	1.0	2.9×10^{-2}
0.9	0.001	1.30	--	4.81	3.80	1.0	3.2×10^{-7}
0.9	0 01	1.17	--	4.07	3.51	1.0	4.3×10^{-5}
0.9	0.1	1.23	--	--	--	1.0	3.0×10^{-3}
0 9	0.5	1.26	--	--	--	1.0	9.7×10^{-2}
1.0	0.001	1.25	--	7.02	4.30	1.0	4.9×10^{-7}
1.0	0.01	1.23	--	2.82	3.92	1.0	7.6×10^{-5}
1.0	0.1	1.20	--	--	--	1.0	6.2×10^{-3}
1.0	0.5	1.18	--	--	--	1.0	8.9×10^{-2}

TABLE X

NUMBER OF TIMES THE AVERAGE ERRORS ARE FIRST, SECOND, OR THIRD LOWEST-CAUCHY CONTAMINATION

Contamination e	Max. Noise Variance	Alpha-Trimmed Mean			ML Cauchy			ML Median			ML Laplacian			ML Gaussian		
		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
0.	0.001	4	13	0	0	0	2	13	4	1	3	0	17	0	3	0
0.	0.01	0	0	3	0	0	0	0	1	16	20	0	0	0	19	1
0.	0.1	0	0	5	0	0	0	1	0	14	19	1	0	0	19	1
0.	0.5	0	0	5	0	0	3	0	0	12	20	0	0	0	20	0
0 01	0.001	1	5	7	0	0	0	4	2	6	14	1	5	1	12	2
0.01	0.01	0	0	6	0	0	0	0	0	14	19	1	0	1	19	0
0.01	0.1	0	1	3	0	0	0	0	0	16	17	3	0	3	16	1
0.01	0.5	0	6	14	0	0	0	0	14	6	20	0	0	0	0	0
0.1	0.001	4	5	2	0	0	0	2	4	11	13	3	3	1	8	4
0.1	0.01	1	2	4	0	0	0	0	1	15	16	2	1	3	15	0
0.1	0.1	0	1	7	0	0	0	0	0	13	15	4	0	5	15	0
0.1	0.5	0	6	14	0	0	0	0	14	6	20	0	0	0	0	0
0.5	0.001	1	16	2	0	0	0	16	3	1	3	0	13	0	1	4
0.5	0.01	2	18	0	0	0	0	18	2	0	0	0	5	0	0	15
0.5	0.1	2	13	6	0	0	0	18	1	0	0	6	1	0	0	13
0.5	0.5	0	20	0	0	0	0	19	0	1	0	1	0	0	0	0
0.9	0.001	2	18	0	0	0	0	18	2	0	0	0	8	0	0	12
0.9	0.01	4	16	0	0	0	0	16	4	0	0	0	9	0	0	11
0.9	0.1	3	17	0	0	0	0	17	3	0	0	0	0	0	0	0
0.9	0.5	1	19	0	0	0	0	19	1	0	0	0	0	0	0	0
1 0	0.001	3	17	0	0	0	0	17	3	0	0	0	2	0	0	18
1.0	0.01	4	16	0	0	0	0	16	4	0	0	0	8	0	0	12
1.0	0.1	7	13	0	0	0	0	13	7	0	0	0	0	0	0	0
1.0	0.5	9	11	0	0	0	0	11	9	0	0	0	0	0	0	0

maximum likelihood estimators are outperformed by the alpha-trimmed mean and median estimators by a large margin. The median estimator appears to be only slightly better than the alpha-trimmed mean from when the data from Table IX is used. But, from Table X, it is quite clear that the median is almost always better than the alpha-trimmed mean as an estimator for contaminated distributions.

The dashes in Table IX indicate a failure of the algorithm to converge. When the Cauchy contamination of the Gaussian becomes significant, the steepest descent maximum likelihood method fails to converge Furthermore, even if the methods do converge, they produce large errors. This is perhaps an indication that the estimate is converging to a local maximum or to a data vector. Notice that the maximum likelihood method based on Cauchy noise generally failed to converge to anything with a reasonably low error. This could be due to any number of reasons, and no clear answer has been found However, it is quite possible that the method does converge, but to a "bad" solution, such as a local maximum This is definitely a possiblity for many reasons One reason is that the standard deviaion is used as an initial guess for the Cauchy parameter, $\alpha = \frac{\sigma_x}{\sigma_y}$, and this may not produce a reliable enough initial guess Unfortunately, the maximum likelihood equations for the Cauchy distribution do not lead to a system of equations from which an estimate of the Cauchy parameter can be made A significant error induced by using the standard deviation for α may result in such a poor initial guess that the maximum likelihood Cauchy method fails to converge The maximum likelihood Cauchy method did converge on small data sets, but not to the same solution as did the other methods It appears that without a proper initial guess for the

parameters, the Cauchy method (when it converges) may be converging to a local maximum rather than minimum.

The main conclusion to be reached from the data in Tables IX and X is that for e $\leq$ 0.1, and for Cauchy contamination less than or equal to 10 percent, the maximum likelihood estimator for a Gaussian noise model is the best. Also, when e $\leq$ 0.5, the sample median is the best estimator and the alpha-trimmed mean does nearly as well For large amounts of Cauchy contamination, the maximum likelihood methods generally fail due to its inability to adaquately estimate the data values and the "error" parameters (estimates of spread of the distribution) The maximum likelihood methods have a potential disadvantage for use on real data This is because even the input noise models and distributions are unknown, so the distributions could be non-Gaussian. As can be seen from Cauchy contaminated Gaussian noise, these maximum likelihood methods are much more sensitive to the noise distribution than is the median Because of its reliability for all types of noise, the median appears to be the best choice for an estimator at this point.

4.2 Robust Attenuation Estimation from 1-D Model Data

To simulate data from Conoco's borehole model and to test the robust estimates on a more physical model, a simple earth model is constructed. There are 10 identical 2 foot thick layers with a velocity of 15000 ft./sec and Q=100 The source used is the digitized source plotted in Figure 4 The bandwidth used is 7-12 KHz (the peak area of the spectrum), and 10 runs with 10 noise sets per run are used in the

experiment The results, in terms of standard deviation of the estimates, are given in Table XI for various maximum likelihood and robust estimates, as well as for several attenuation methods previously discussed. White Gaussian noise is added to the time-domain input signals so that the signal-to-noise ratio's (SNR) are 100 and 1000. That is,

$$r_i(t) = s_i(t) + n_i(t)$$

where $r_i(t)$ is the P-wave signal at the ith receiver, $s_i(t)$ is the ideal signal, and $n_i(t)$ is a Gaussian noise process.

When the signal-to-noise ratio is 1000, then all of the method shown in Table XI worked fairly well. By far the best methods are the median and the alpha-trimmed mean. The maximum likelihood estimator for a Gaussian model did not do well. The results for maximum likelihood Gaussian method were not included because they frequently failed to converge even for the relatively high signal-to-noise ratio of 1000.

When the data contains more noise (signal-to-noise ratio is 100), all the estimators are higher. The eigenvector method failed to converge and the maximum likelihood (Gaussian) method occasionally failed to converge. The Wiener filter method worked fairly well, but the results shown in Table XI is the best result for the method and it was reached only after a great deal of experimentation to find the optimum parameters. The spectral ratio method did not do well. For this model of attenuation estimation, the simplest and most robust scalar methods, namely the median and alpha-trimmed mean, performed the best

TABLE XI

ATTENUATION ESTIMATION - STANDARD DEVIATION
OF ERROR FOR A SIMPLE, ONE
DIMENSIONAL MODEL

Signal-to Noise Ratio	Spectral Ratio	Wiener Filter	Eigenvector Decomposition	Average	Max. Likelihood Gaussian	Median	Alpha-Trimmed Mean
1000.	9.3	10.1	8.9	7.7	14.0	6.3	6.0
100.	1900	27.3[2]	--	381.	72.8[1]	25.4	25 4

1 = Did Not always Converge
2 = Required Careful Adjustment of Parameters

4 3 Attenuation Estimation from Borehole Model Data

Synthetic data which accurately models the propagation of acoustic energy in a borehole has been supplied by both Amoco and Conoco The data set from Conoco was chosen for study because attenuation is readily apparent in the Conoco data This data set contains groups of 31 traces with source-receiver offsets ranging from 3 0 meters to 6 0 meters at 0 1 meter increments. A complete set of 31 traces is modelled for many different combinations of borehole parameters (e g. shear and P-wave velocity, shear wave attenuation, and P-wave attenuation). Two sets were chosen for detailed study

The first data set, shown in Figure 9, contains model data with the following set of parameters Borehole radius = 10. cm, fluid velocity = 1600. m/sec., fluid Q = 50 , P-wave velocity = 4000 m/sec , P-wave Q = 100 , shear wave velocity = 2300 m/sec., shear wave Q = 100. Figure 10 is a plot of just the near offset (3.0 m) trace from Figure 9. The feature of interest in this figure is the P-wave which begins at about 0.8 milliseconds and ends at the onset of the shear wave arrival at about 1.3 milliseconds Note that the P-wave is much lower in amplitude than the other waves, but it is separated in time from the other waves and can easily be isolated using a window. Figure 11 is a plot of just the P-wave from the trace in Figure 10

A Hamming window is used to isolate the P-wave arrival from the rest of the trace The total length of the window is 0 75 milliseconds (60 samples) and the shoulder width is 0 1 milliseconds (8 samples) Figure 12 shows the camplitude spectrum of the P-wave in the near offset (3 0 m) trace. Figure 13 shows the amplitude spectrum of the far offset

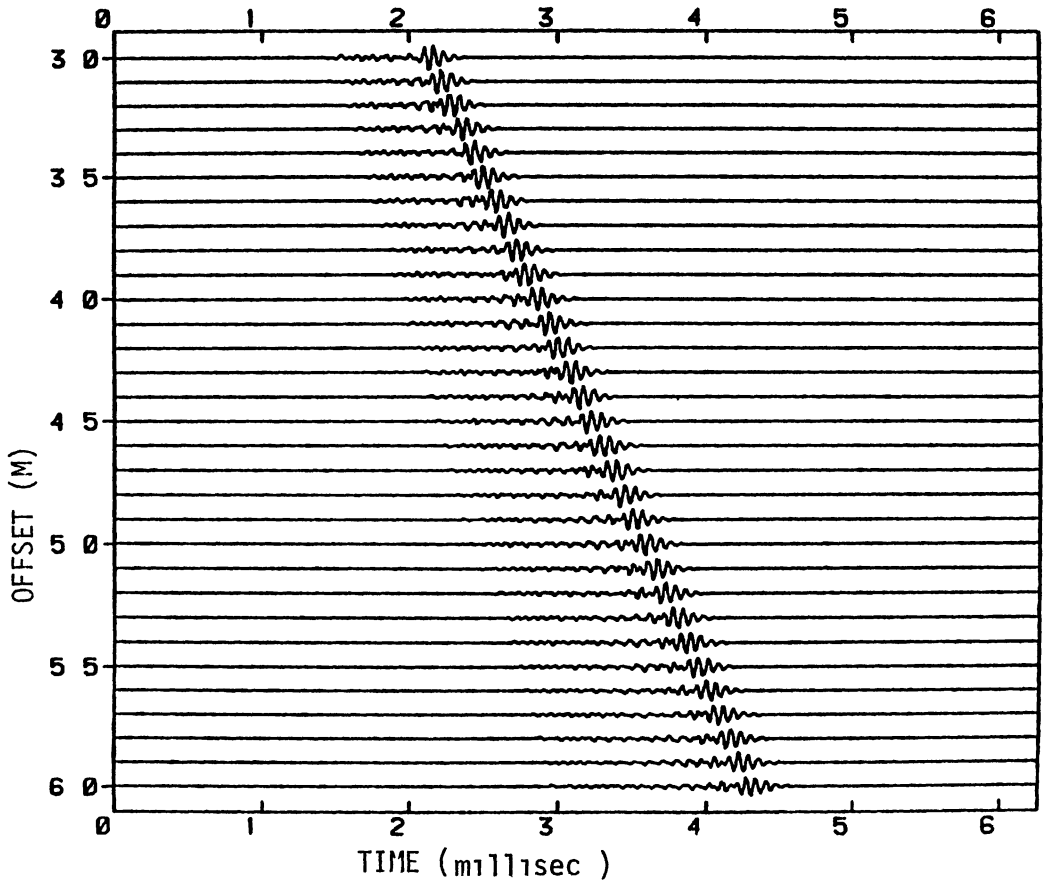

Figure 9 Conoco Borehole Model Data

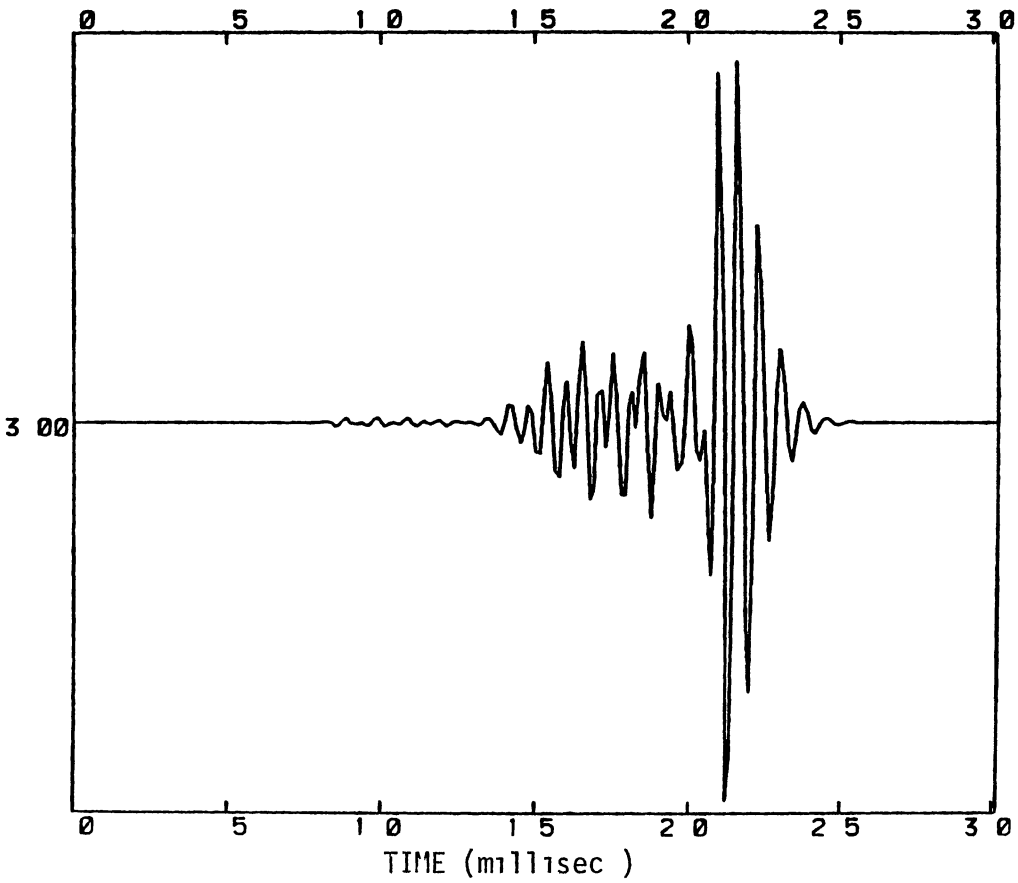

Figure 10 Near Offset Trace-Model Data

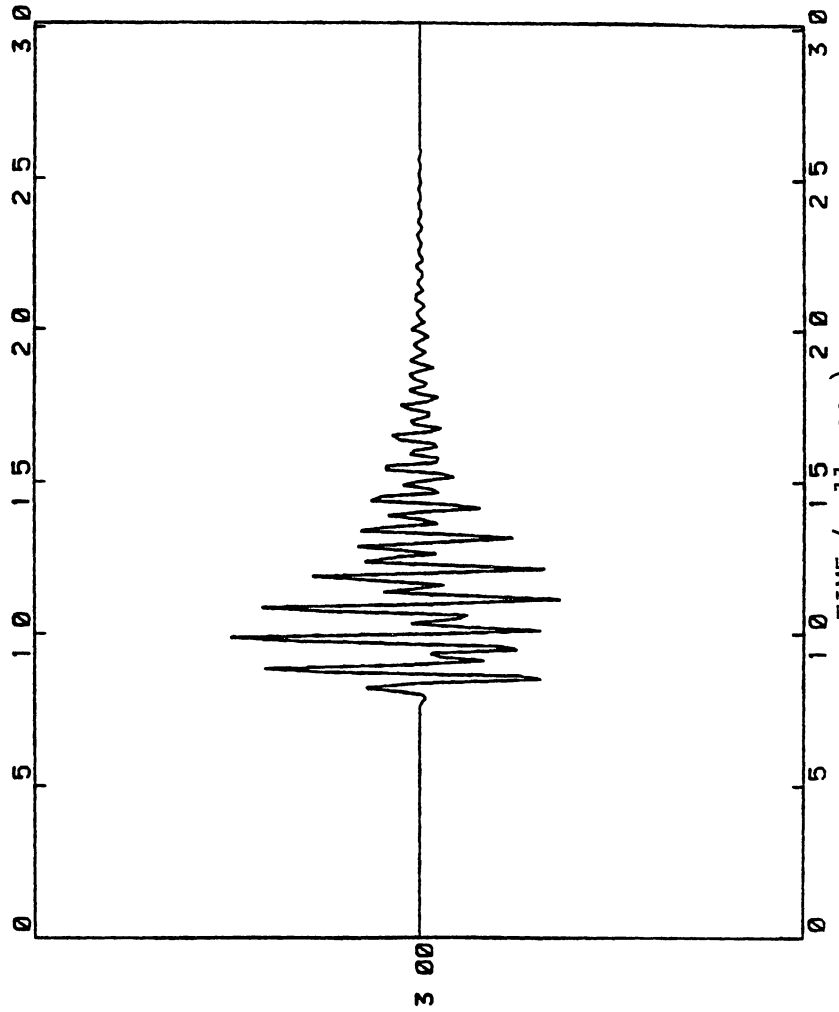

Figure 11 P-Wave Component Only

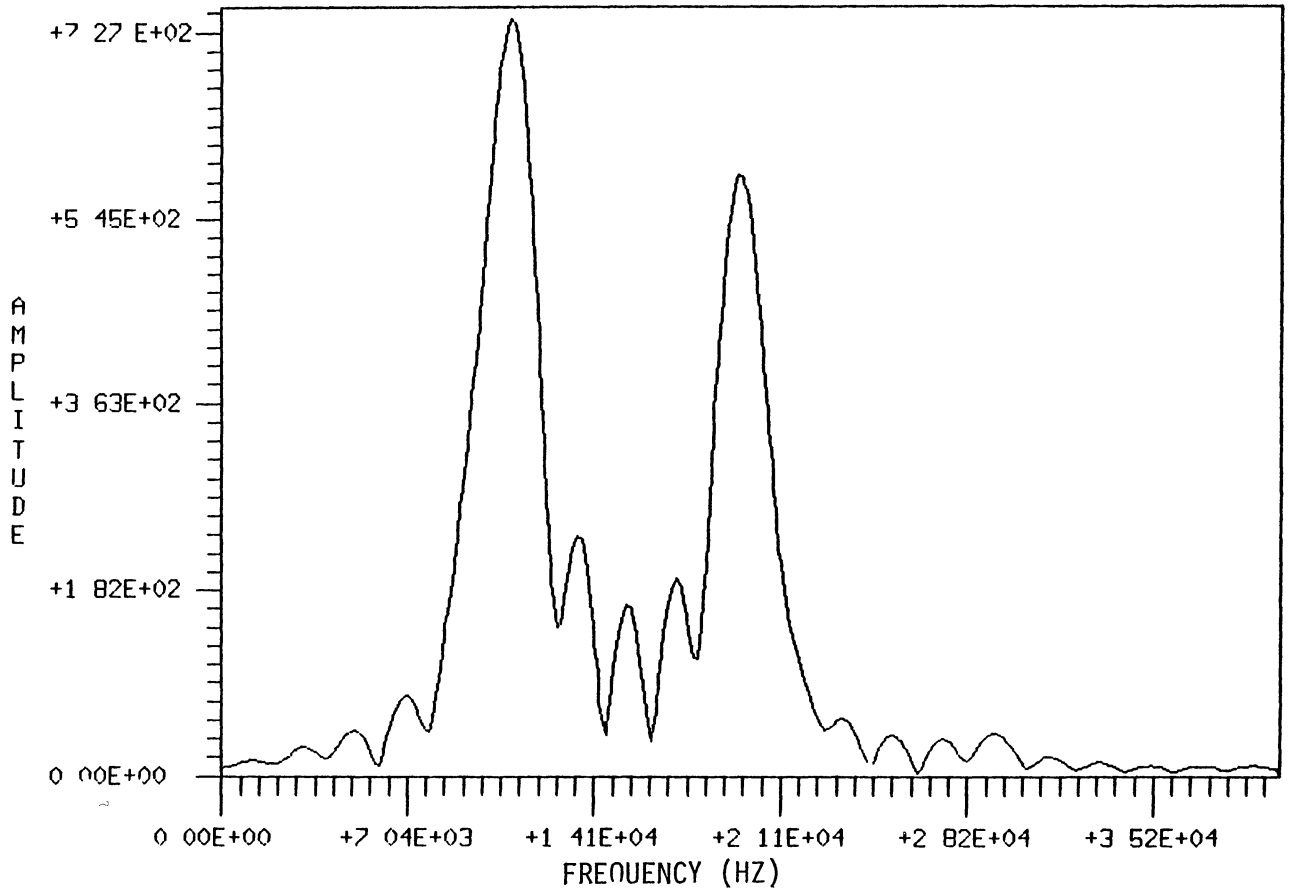

Figure 12 Amplitude Spectrum of Near Offset Trace

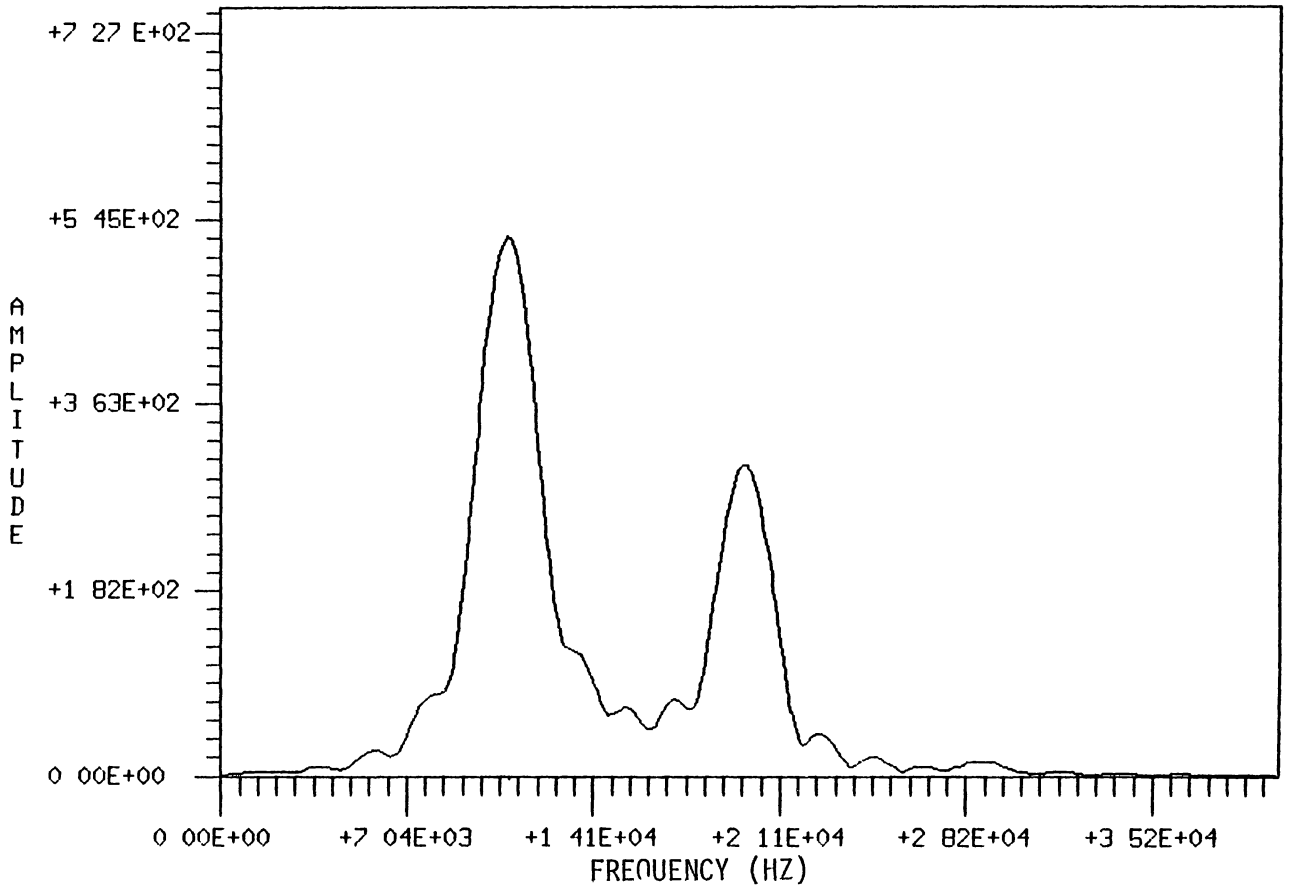

Figure 13 Amplitude Spectrum of Far Offset Trace

trace (6 0 m) Both traces have been multiplied by the offset to approximately correct for geometrical spreading losses. For a different set of traces without attenuation, multiplying by the offset did a good job of correcting for geometric spreading loss (to within a few percent) The effect of attenuation is readily apparent from a comparison of Figures 12 and 13. It is important to note here that the P-wave spectral shapes and the attenuation estimates are fairly insensitive to changes in the method of windowing. As long as the window contained most of the P-wave arrival and has some taper at the shoulders, the resulting spectra are not significantly affected. Other windows tried include trapezoidal, raised cosine and rectangular windows Only the rectangular window gave poor results.

In order to produce reliable attenuation estimates it is necessary to carefully choose the frequencies over which the attenuation estimate is made. For this model data the best range is from 9.5 kHz to 11 5 kHz. The final attenuation estimates are much more sensitive to the frequencies used than to the windowing function It is important that the range be restricted to some neighborhood in the vicinity of a spectral peak. Using less than the optimum number of frequency points causes a slight deterioration in the accuracy of the estimate. However using too many frequency points results in very inaccurate estimates It is interesting to note that the 2.0 kHz frequency band used for the attenuation estimates results in 13 frequency points Since acoustic logging tools typically have 10-12 receivers, the size of the data matrix using for many estimation method is close to the ten by ten size previously used for simulations

The results of applying some of the previously discussed attenuation estimators to the Conoco borehole model data are shown in Table XII The estimators listed in Table XII are the sample average, the median, the alpha-trimmed mean, the maximum likelihood Gaussian estimator and the maximum likelihood Laplacian estimator The following attenuation estimators were tested on the model data but the results are not listed becasue they failed to produce physically reasonable estimates spectral ratio, eigenvalue decomposition, Wiener filter, and maximum likelihood Cauchy estimators. Even though Q = 100 is the correct value, these estimators usually produced Q estimates in the range $-10 \leq Q \leq 10$

The first row of Table XII lists the results when Q = 100, no noise is added, and the frequency range (9.5 - 11 5 kHz) is optimum. For this case, all of the estimators listed did fairly well, except for the sample average. The median had the lowest average error, and the maximum likelihood Gaussian estimator performed nearly as well as the median When the attenuation estimation is based upon frequencies from 19.5 kHz to 21.5 kHz (the second lobe of the spectrum) the results of the estimates are much worse for all of the estimators.

The Conoco model contained one set of traces which were generated with Q = 30. The results are much worse than for Q = 100, because the low Q (high attenuation) results in very little energy at the far offsets. The attenuation estimate for the near offset is more accurate than for the far offsets, but the improvement is slight Because the Q is low, and because dispersion can only be considered negligible if when $Q >> 10$, it is possible that dispersion effects are reducing the accuracies of the estimate

TABLE XII

ROOT MEAN SQUARE ERROR OF ATTENUATION
ESTIMATES FROM CONOCO MODEL DATA

Percen Noise	True Q	Frequency (kHz)	Average	Median	Alpha-Trimmed Mean	ML Gaussian	ML Laplacian
0	100.	9 5-11.5	62.0	12.3	17.2	13.4	17 1
0.	100	19.5-21.5	42 7	22 0	21.2	27.0	34 6
0	30	9.5-11.5	26.2	30 0	27.6	27.0	--
0 1	100	9 5-11.5	63.8	43.0	41.5	--	--
1.0	100.	9.5-11.5	66.8	44.7	43.4	--	--

When noise is added to the model, all of the estimators gave inaccurate estimates. The maximum likelihood Gaussian and maximum likelihood Laplacian estimators failed to converge The three estimators which did give estimates have average root mean square errors about half as large as the parameter being estimated. Even 0 1 percent noise added to the signal produced estimates which have very large variances Actual Q estimates range from near zero to around 150. Any decrease in the amount of additive noise by an order of magnitude has little effect on the accuracy of the estimates.

The median attenuation estimator had the best overall performance on the Conoco model data, and the alpha-trimmed mean also did well. The column average did not do well at all. The maximum likelihood Gaussian estimator did very well when no random noise was present, but it didn't converge for even very small levels of random noise. The maximum likelihood Laplacian estimator did moderately well on the best no-noise data set, but this method frequently failed to converge Overall, the median and the alpha-trimmed mean should perform the best on real data based upon the performance given in Table XII. The maximum likelihood Gaussian estimator may also do well on real data provided the noise levels are very low

4 4 Attenuation Estimation from Borehole Data

The real acoustic log data used in this study comes from Conoco and was recorded using a 12 receiver Schlumberger tool The source-receiver offsets range from 13 ft to 18.5 ft , with a receiver to receiver offsets of 0 5 ft. To thoroughly investigate the use of several methods of attenuation estimation on the data set, one set of 11 traces is

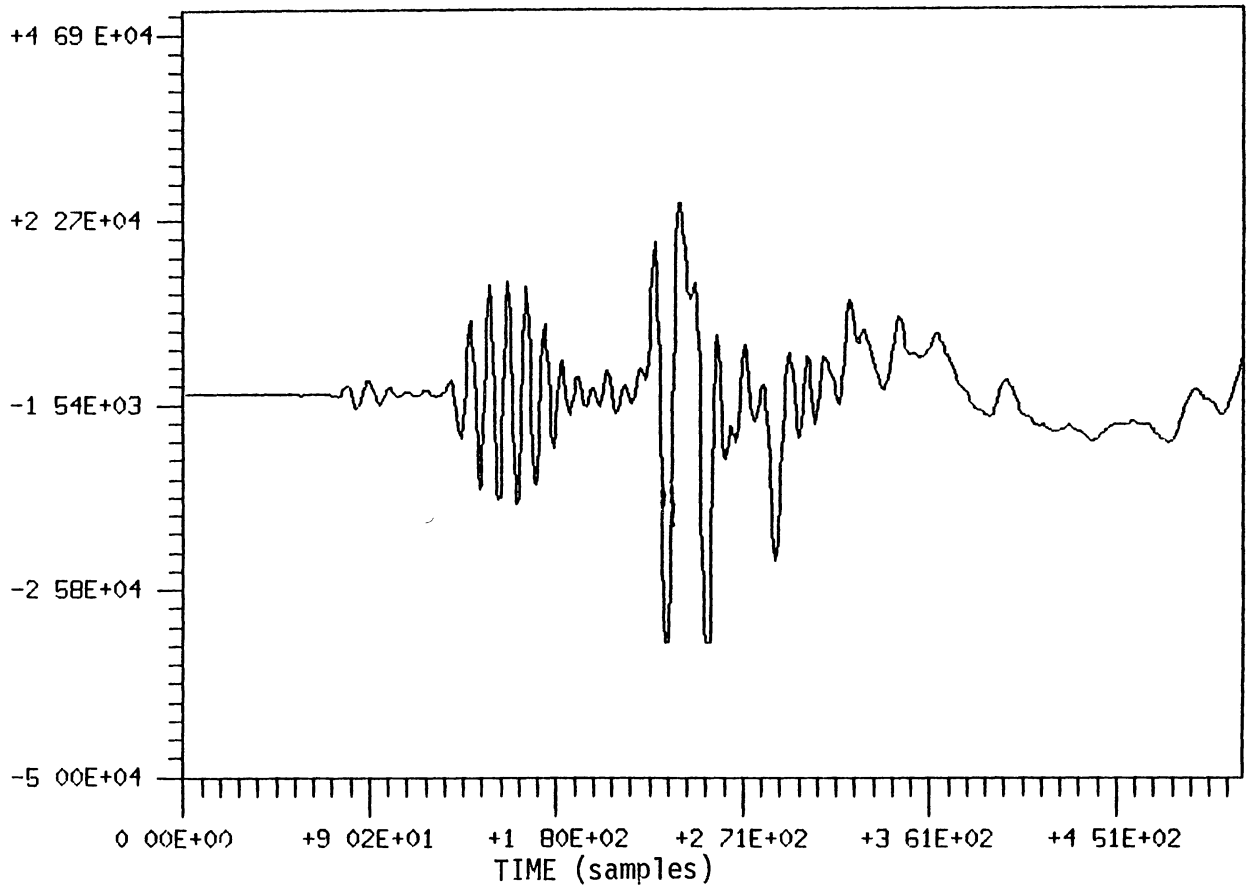

Figure 14 Trace 1 (Near Offset) Real Data

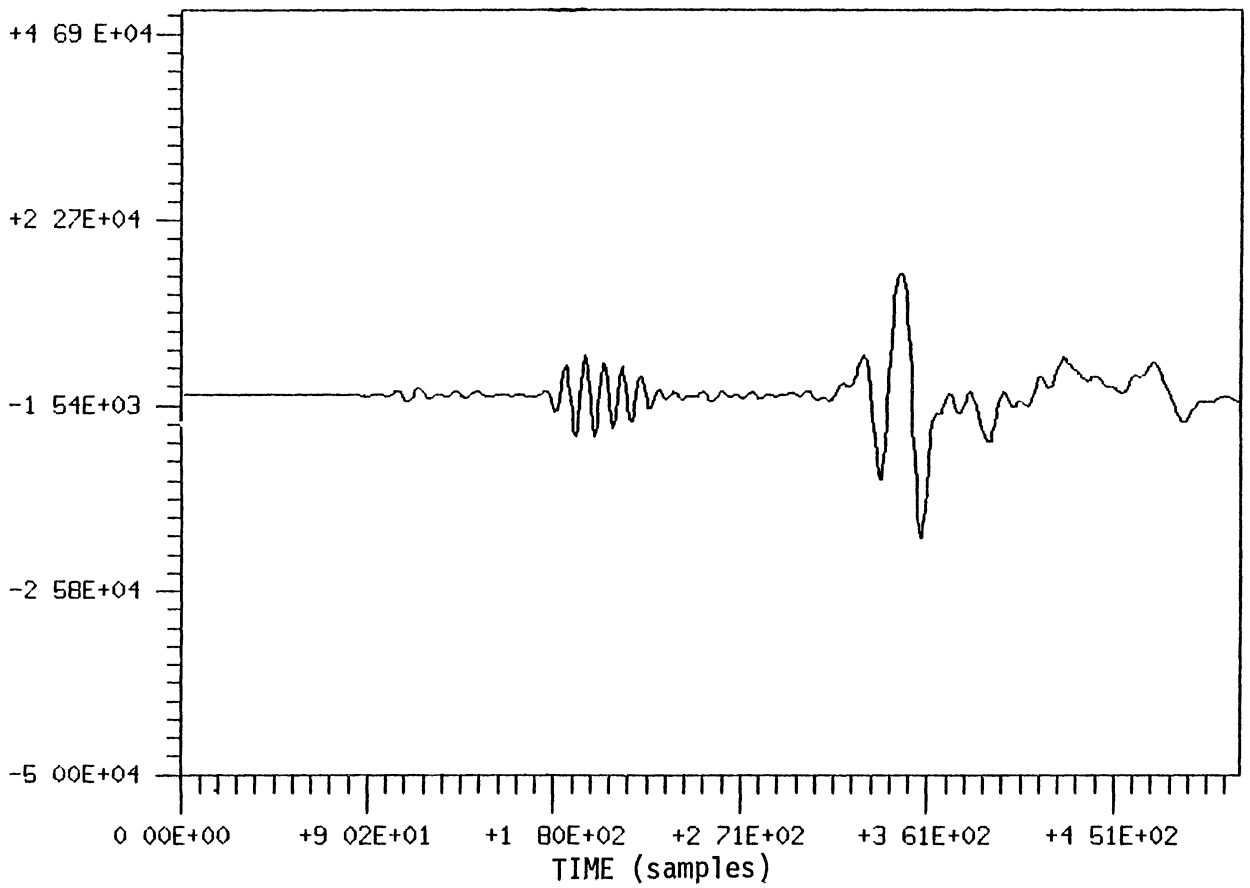

Figure 15 Trace 11 (Far Offset) Real Data)

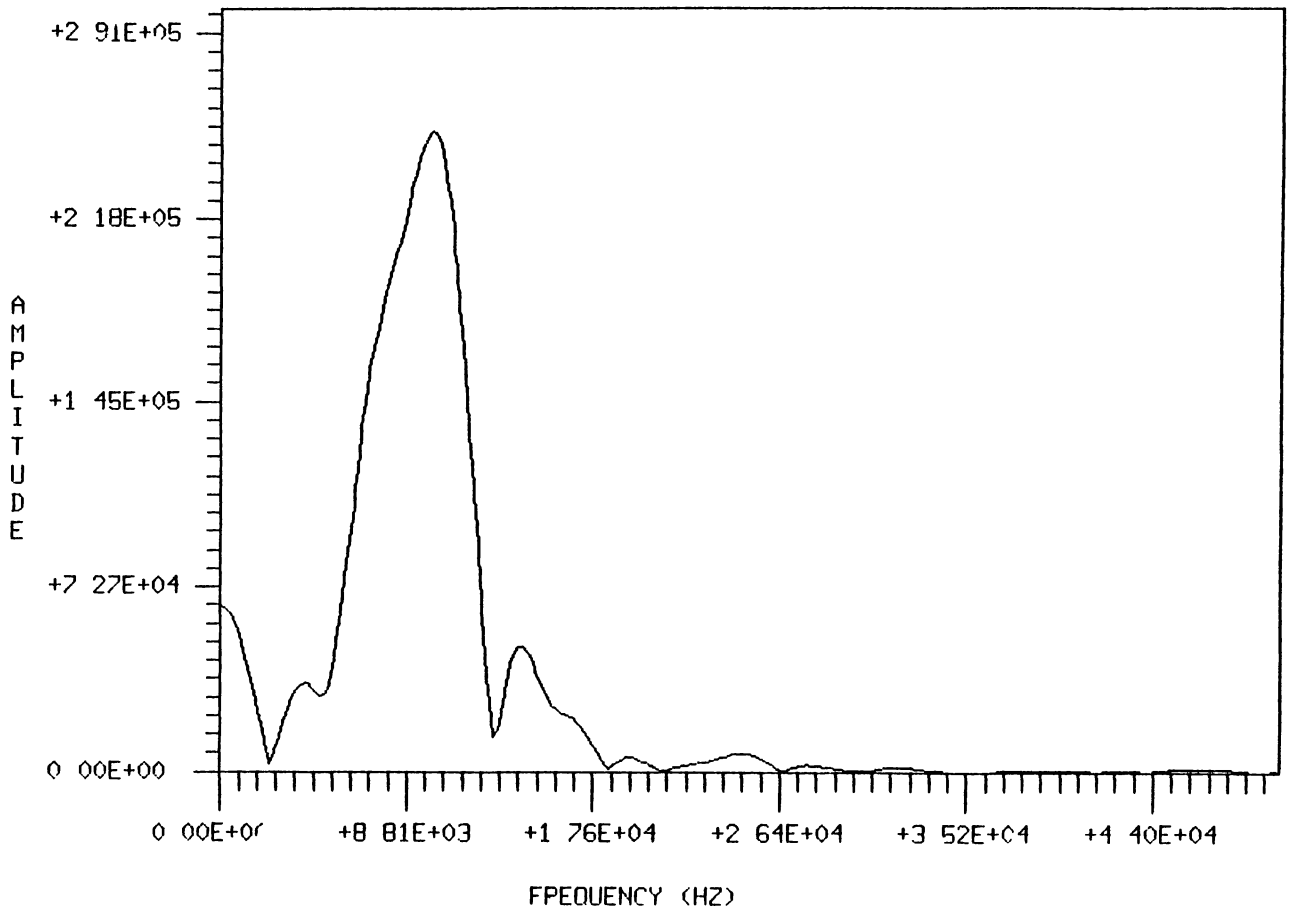

Figure 16 Spectrum of P-Wave, Trace 1 (Real Data)

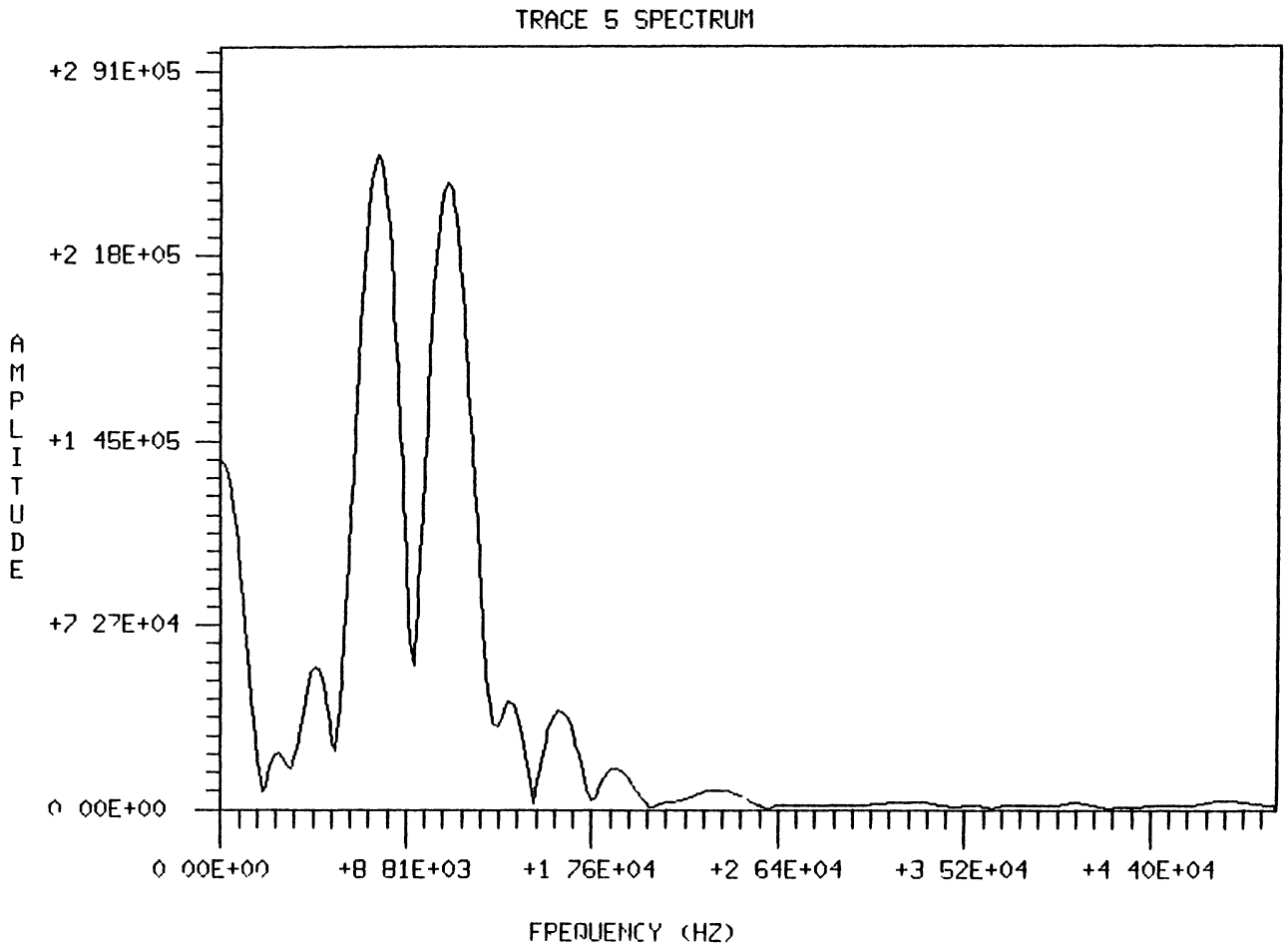

Figure 17 Spectrum of P-Wave, Trace 5 (Real Data)

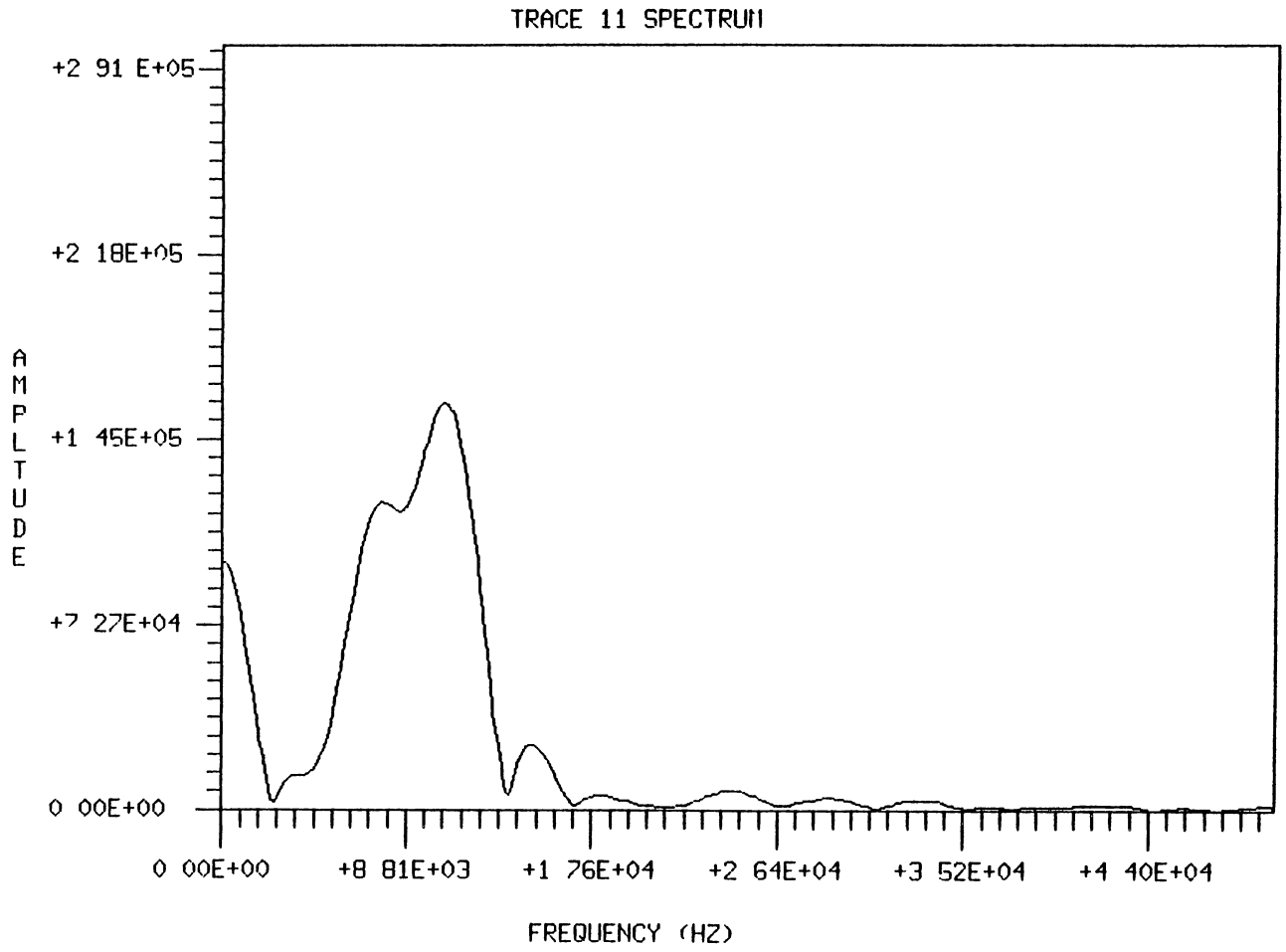

Figure 18 Spectrum of P-Wave, Trace 11 (Real Data)

studied. The first trace, known as the near offset trace (offset = 13.0 ft.) is shown in Figure 14. The far-offset trace (offset = 18.5 ft) is shown in Figure 15 The loss of amplitude with offset is apparent even though these traces have been approximately corrected for geometric spreading losses by multiplying each trace by its offset.

The borehole wave under investigation is the P-wave. The P-wave is the very small amplitude event which arrives first. In Figure 14, the P-wave arrives at 77 samples on the time axis and dies away at 125 samples on the time axis (the units of time in the plots are "samples"). The P-wave lasts about 50 samples, and so 50 samples is chosen as the length of the Hamming window applied to separate the P-wave data from the rest of the data. The shoulder width of the Hamming window is chosen to be eight samples. Figures 16, 17 and 18 the amplitude spectra of the windowed P-wave arrival at offsets of 13.0 ft., 15 0 ft., and 18.5 ft., respectively. Notice that the spectrum in Figure 16 is smooth, and the spectrum in Figure 18 is fairly smooth. However, the spectrum of Figure 17 has a null or zero which occurs set the same frequency as the peak of the main lobe in Figure 16. Unfortunately this spectral shape makes attenuation estimation very difficult.

Attempting to estimate Q for each receiver pair resulted in negative or nearly zero Q estimates for all methods. Only when the attenuation estimate is based upon the ratio of the spectra from the near and far offset receivers do physically reasonable answers result The results of a few attenuation estimators for various frequency ranges are shown in Table XIII. There is a strong correlation between Q values and P-wave velocity. When the P-wave velocity is high, there is usually

TABLE XIII

ATTENUATION ESTIMATES FROM CONOCO DATA

Layer Number	Receiver Spacing (ft)	Frequency (kHz)	Average Q	Median Q	Alpha-Trimmed Mean Q
1	5.5	10.0-11.5	38.3	36 0	37.1
1	5.5	7.5-11 5	20.8	14.9	17.2
1	2 0	10 0-11.5	700.8	19 1	20.9
2	2.0	10.0-11.5	22.7	21.6	24.5
1	2 0	7 5-11 5	35.34	8.4	13.3
2	2 0	7.5-11.5	20.7	18 4	18.7
1	--	---	. 08	3 8	3.8
2	--	---	-5.8	-6.0	-6.0
3	--	---	4.8	4.8	4.8
4	--	---	8.4	8.4	8.4
5	5	10 -11.5	8 0	8.0	8.0
6	--	---	4.4	-16.1	-16 1
7	--	---	.82	.81	81
8	--	---	-.82	-.79	-.79
9	--	---	2.5	2.5	2.5
10	--	---	-8.9	-8 0	-8.0

little attenuation (high Q) Since the P-wave velocity estimate for this data set measured from the onset of the P-waves is $2x10^4$ ft./sec (which is fairly high), then Q estimates should also be high (perhaps Q = 100).

As can be seen from Table XIII, most of the attenuation values are lower than should be expected based upon laboratory results. If the criterion for judging the performance of attenuation estimators is the laboratory values, then the best results are from using a narrow frequency band (10 kHz to 11.5 kHz) which corresponds to the neighborhood of the largest spectral peak in Figure 18 As previously mentioned, this result is based upon the spectral ratio involving the near and far offset traces only The only techniques which produce reasonable results are the sample average, the median, and the alpha-trimmed mean. The maximum likelihood methods and the eigenvector method do not work on just one spectral ratio. Unfortunately, attenuation estimates from real data can be judged for accuracy only by comparing the Q estimates with values measured in laboratories. These laboratory measurements indicate that Q values may range from 50 to 150 though values outside that range are possible. Estimators such as the spectral ratio and Wiener filter methods fail on this data set (they yield negative values). The results with the highest Q estimates from Table XIII range from 36.0 for the median to 38.3 for the sample average with the alpha-trimmed mean in the middle. The difference between these three estimates is probably not significant, and it is not possible to decide which is better based upon this one result.

4 5 Chapter Summary

Attenuation estimation methods have been tested on fairly realistic models which may have noise added. The only estimators able to perform adaquately are the median and the alpha-trimmed mean The sample average and the maximum likelihood Gaussian method work well occassionally but not consistently. Therefore, based upon performance on model data, and taking into account computational considerations, the median and the alpha-trimmed mean are superior to the others. This conclusion is not changed by the results on real data, because the estimators cannot be judged when the values to be estimated remain unknown, and because the median and alpha-trimmed mean do give physically reasonable estimates.

CHAPTER V

SUMMARY AND CONCLUSIONS

The estimation of the attenuation coefficient, Q, from acoustic well log data is a difficult problem. The problem is ill-posed, so that small errors in the estimation of the spectra of the received signals may result in large errors in the estimation of Q A physically reasonable noise model leads to an error distribution in the final attenuation estimate which is very long-tailed Adding small amounts of random noise to borehole model data caused the estimators to yield inaccurate estimates, the behavior of the estimators on model data with small amounts of noise matched what one would expect from the long-tailed noise distributions previously modelled. Some of the estimators worked well enough on real data to yield physically reasonable values of Q. Unfortunately, the accuracy of the Q estimates from real data remains unknown, because the actual Q values for the rock surrounding the borehole are unknown.

The classic attenuation estimator, the spectral ratio method, is the most commonly used method for attenuatione estimation on all types of acoustic data This method is computationally simple, and it allows for an unknown amount of frequency-indenpendent geometric spreading losses. However, this method also has many disadvantages. The ratio of noise contaminated spectra results in very unstable estimates of the pure spectral ratio Since the spectral ratio method is a least-squares

method and is thus sensitive to bad data points, the method does not work very well when the spectra contain zeros and noise Because the spectral ratio method can handle unknown geometric loss factors if it must estimate another parameter. This is actually a disadvantage because the estimating an unnecessary parameter (the intercept of line in the spectral ratio method) results in higher estimation errors. The Wiener filter method is a variation of the spectral ratio method and can handle spectral zeros better. Unfortunately, it is very sensitive to parameter selection, and since it is also a least squares technique, the Wiener fitler method is too sensitive to large spectral errors.

The Q estimates for each frequency and depth can be cast as a matrix, and the eigenvector corresponding to the largest eigenvalue is an optimum least-squares estimate for the values of Q versus depth. Unfortunately, the eigenvector method performs only slightly better than simply averaging the columns of the matrix to find Q. However, the matrix formulation itself leads to maximum likelihood vector estimators to find Q versus depth Maximum likelihood estimators based upon Gaussian, Laplacian, and Cauchy noise were implemented using the steepest descent method and tested on various data sets The Gaussian based estimator performed the best overall, and had fewer convergence problems However it did not work very well on model data with random noise.

Two very simple robust estimators, the sample median and the alpha-trimmed mean, were tested on many models, with varying levels of noise with different distributions These two methods proved to be superior to all of the methods for most test cases. This is because these estimators are very robust, and thus not sensitive to bad data values

Furthermore, their computational simplicity and lack of convergence problems makes them the optimum choices for use on real data.

However, the maximum likelihood estimator based upon a Gaussian noise model did work about as well as the median and alpha-trimmed mean on several data sets In particular, the maximum likelihood (Gaussian) method worked well on the synthetic borehole data for low noise levels. So, it should work on real data when the spectrum is smooth and the background noise level is low.

All of the estimators which have been tested on model data were also tested on the real data set from Conoco Unfortunately, only three estimators gave physically reasonable results on real data They are the sample average, the median, and the alpha-trimmed mean. The estimates of Q from real data are strongly affected by the choice of the frequency band over which the estimate is made. For models, the optimum results were achieved from a narrow band around the peak of the spectrum. The spectra from the real data is sometimes more complicated than the spectra in the models, but choosing a window in frequency which contains the spectral peak for most offsets did yield what is probably the best results This complicated spectrum is probably the cause of the failure of the maximum likelihood (Gaussian) method.

Since the actual Q values for the real data are unknown, the accuracy of the Q estimates is also unknown. However, based upon laboratory data and given the apparent P-wave velocity, estimates of Q should be near Q = 100. The "best" Q estimates from the real data are in the range of 36 to 38 This may be too low, but it is impossible at this time to make any significant conclusions regarding the accuracy of these estimates In fact, even if laboratory Q measurements are made

from a core of the well which the data is from, one would still not be able to reach a conclusion regarding the accuracy of the attenuation estimates. This is because attenuation is a strong function of the environment of the rock (pressure, fuild saturation, etc...) So, in order to compare laboratory values with well log values of Q, the measurements made in the lab must be under essentially the same environmental conditions as those in the borehole. This may not be possible. But, if it is possible to make such measurements, it should be done so that the accuracy of Q estimates from real data can be determined Only when this is completed will it be possible to improve attenuation estimation algorithms to the point that they can be used routinely on real data.

Futher research in the area of attenuation estimation from acoustic logs should emphasize two areas One area in which more research is necessary is understanding the effects of borehole and tool geometrics on the spectra of the received waveforms. Irregularities in the borehole and tilting of the tool produce features in the spectra of the received signal which makes attenuation estimation difficult. It may prove possible to model these irregularities with a simple model and remove them from the data in the same manner as multiple reflections are removed from surface seismic data.

Even if the problems with spectral irregularities were solved, much work would still need to be done to evaluate the accuracy of the attenuation estimates. Experiments to estimate the attenuation from cores should be conducted where the core is kept in conditions that are as close to those in the borehole as possible The frequences used should correspond to those of the source in the acoustic logging tool

In addition, attempts should be made to generate attenuation estimates from vertical seismic profile data and surface seismic data for comparison with estimates from acoustic logs. Once the accuracy of the attenuation estimates is established, then work can be done to develop rules for the interpretation of attenuation logs.

BIBLIOGRAPHY

Aki, K., and Richards, P G., 1980, Quautitative Seismology Theory and Methods W.H. Freeman and Co

Anderson and Castagna, 1984, Analysis of Sonic Log Compressional Wave Amplitudes using Borehole Compensation Techniques Presented at SPWLA 25th Annual Logging Symposium, June 10-13

Andrews, D.F., 1981, Robust Statistics John Wiley & Sons, Inc

Aron, J., Murray, J. and Seeman, B., 1978, Formation Compressional and Shear Interval-Transit-Time Logging by Means of Long Spacings and Digital Techniques Presented at 53rd Ann. Fall Conf. and Exhib of Soc. of Petroleum Eng. of AIME (SPE 7746).

Bednar, J.B , and Watt, T L., 1984, Alpha-Trimmed Means and Their Relationship to Median Filters Acoustics, Speech, and Signal Processing, Vol 32, No 1, pp 145-153

Bickel, S.H., 1982, Constant Q Inverse Filters Society of Exploration Geophysicists Annual Meeting, Dallas, Tx.

Cheng, C.H , Toksoz, M.N., Willis, and M.E., 1981, Velocity and Attenuation from Full Waveform Acoustic Logs Society of Professional Well Log Analysts Annual Symposium.

Cheng, C.H , Toksoz, M.N., Willis, and M E., 1982, Determination of in Situ Attenuation from Full Waveform Acoustic Logs Jour of Geophysical Res , Vol 87, No. B7, pp 5477-5484.

Engelhard, L., Gross, T , and Newport, F , 1984, Comment on "Determination of in Situ Attenuation From Full Waveform Acoustic Logs" Jour. of Geophysical Res., Vol 89, No. B5, p. 3400.

Gilbert, F., 1971, Ranking and Winnowing Gross Earth Data for Inversion and Resolution Geophys. J Roy Astr Soc , Vol. 23, pp. 125-128.

Gimlin, D R., Keener, M.S , and Lawrence, J F., 1982, Maximum Likelihood Stocking in White Gaussian Noise With Unknown Variances IEEE Transactions on Geoscience and Remote Sensing, Vol GE-20, No 1

Goldberg, D S., Kan, T.K , and Castagna, J.P , 1984, Attenuation Measurements From Sonic Log Waveforms Presented at SPWLA 25th Annual Logging Symposium, June 10-13

Hale, D , 1982, Q-Adaptive Deconvolution Society of Exploration Geophysicists Annual Meeting, Dallas, Tx.

Hamilton, E., 1972, Compressional Wave Attenuation in Marine Sediments Geophysics, Vol. 37, pp. 620-646.

Huber, P.J , 1981, Robust Statistics John Wiley & Sons, Inc.

Jacobson, R.S., Shor, G.G , and Dorman, L.M., 1981, Linear Inversion of Bodywave Data-Part II Attenuation Versus Depth Using Spectral Ratios Geophysics, Vol. 46, pp. 152-162

Johnston, D.H., and Toksoz, M.N., 1981, Seismic Wave Attenuation Society of Exploration Geophysicists (pub.), pp. 459.

Kassan, S A , Poor, and H.V , 1985, Robust Techniques for Signal Processing A Survey Proceedings of the IEEE, Vol. 73, No 3.

Knopoff, L., Aki, K., Archambeau, C.B., Ben-Menahem, A , and Hudson, J.A., 1964, Attenuation of Dispersed Waves Jour. of Geophys. Res., Vol. 69, No 8, pp 1655-1657.

Kuc, R., and Schwartz M , 1979, Estimating the Acoustic Attenuation Coefficient Estimation Technique for Liver Pathology Characterization IEEE Trans. Sonics Biomed. Eng BME-27, pp 312-319.

Kuc, R , 1981, Digital Filter Models for Media Having Linear With Frequency Loss Characteristics J Acoust. Soc. Amer., Vol 69, pp 35-40.

Kuc, R., 1984, Estimating Attenuation From Reflected Ultrasound Signals Comparison of Spectral-Shift and Spectral-Difference Approaches IEEE Transactions on Acous. Speech Sig. Proc., Vol. ASSP-32, No. 1, pp 1-6.

Kuster, G T , and Toksoz, M.N., 1974, Velocity and Attenuation of Seismic Waves in Two-Phase Media Part I Theoretical Formulations Geophysics, Vol 39, pp 587-606.

Lanczos, C , 1961, Linear Differential Operators D. Van Nostrand Co., Inc.

Mavko, G.M., and Nur, A , 1975, Melt Squirt in the Aesthenosphers J Geophysical Research, Vol 80, pp. 1444-1448

McCarley, L A , 1985, An Autoregressive Filter Model for Constant Q Attenuation Geophysics, Vol. 50, pp. 749-758.

McDonal, F J., Angona, F A., Mills, R.L , Sengbush, R.L , Von Nostrand, R.G , and White, J E , 1958, Attenuation of Shear and Compressional Waves in Pierre Shale Geophysics, Vol 23, pp 421-439

Nur, A., and Winkler, K., 1980, The Role of Friction and Fluid Flow in Wave Attenuation in Rocks (abst) Geophysics, Vol 45, pp. 591-592

O'Doherty, R F., and Austey, N.A , 1971, Reflections on Amplitudes Geophys. Prosp., Vol 19, pp. 430-458.

Parks, T W., McClellan, J.H., and Morris, C.F., 1983, Algorithm for Full-Waveform Sonic Logging Proceedings of the International Conference on Acoustics, Speech, and Signal Processing.

Pisarenko, V F., 1970, Statistical Estimates of Amplitude and Phase Corrections· Geophys. J.R., Astr Soc., Vol. 20, 89-98.

Schoenberger, M., and Levin, F.K , 1974, Apparent Attenuation Due to Introbed Multiples Geophysics, Vol 39, pp. 278-291.

Spencer, T.W., Sonnard, J.R., and Butler, T.M., 1982, Seismic Q-Stratigraphy or Dissipation Geophysics, Vol. 47, No. 1, pp. 16-24

Stoll, R.D., and Bryan, G.M., 1970, Wave Attenuation in Saturated Sediments J. Acous. Soc Am., Vol 47, pp. 1440-1447

Strang, G., 1980, Linear Algebra and Its Applications Academic Press, Inc.

Taylor, S.R., and Toksoz, M.N., 1982, Measurements of Interstation Phase and Group Velocities and Q Using Wiener Filtering Bull. Seis. Soc. Amer , Vol 72, No 1, pp. 73-91

Tittman, B.P , Nadler, H., Clark, V.A , Ahlberg, L A., and Spencer, T.W., 1981, Frequency Dependence of Seismic Dissipation in Saturated Rocks Geophys. Res. Lett., Vol. 8, pp. 36-38.

Toksoz, M.N., Johnston, D.H., and Timur, A., 1979, Attenuation of Seismic Waves in Dry and Saturated Rocks I. Laboratory Measurements Geophysics, Vol. 44, pp 681-690

Tullos, F.N , and Reid, A.C., 1969, Seismic Attenuation of Gulf Coast Sediments Geophysics, Vol 34, pp 516-528

Van Trees, H.L., 1968, Detection, Estimation, and Modulation Theory John Wiley and Sons, Inc.

Walsh, J B , 1966, Seismic Wave Attenuation in Rock Due to Friction J. Geophys Res , Vol 71, pp 2591-2599

Walsh, J B , 1968, Attenuation in Partially Melted Material J Geophys Res., Vol 73, pp 2209-2216

Watt, T.L , 1984, Some Applications of Robust Statistical Procedures to Problems in Seismic Signal Processing Ph D Thesis, University of Tulsa

White, J E , 1975, Computed Seismic Speeds and Attenuation in Rocks With Partial Gas Saturation Geophysics, Vol 40, pp. 224-232

Wiggins, R.A., 1972, The General Linear Inverse Problem, Implications of Surface Waves and Free Oscillations for Earth Structure Reviews of Geophys and Space Physics, Vol. 10, No. 1, pp 251-285.

Williams, D M., Zemanek, J., and Angona, F.A., et. al., 1984, The Long Spaced Acoustic Logging Tool Presented at SPWLA 25th Annual Logging Symposium, June 10-13.

Willis, M.E., 1983, Seismic Velocity and Attenuation From Full Waveform Acoustic Logs Ph.D. Thesis at Massachusetts Institute of Technology.

VITA

Steven Wayne Patton

Candidate for the Degree of

Doctor of Philosophy

Thesis ATTENUATION ESTIMATION FROM ACOUSTIC WELL LOGS

Major Field Electrical Engineering

Biographical

Personal Data Born in San Diego, California, February 16, 1958, the son of William K and Shirley J. Patton. Married to Laura M. Ondrake on May 31, 1980.

Education Graduated from Heritage High School, Littleton, Colorado, in May, 1976, received the Bachelor of Science degree in Mineral Engineering-Physics from Colorado School of Mines in May, 1980, received the Master of Science degree in Electrical Engineering from Stanford University in June, 1981, completed requirements for the Doctor of Philosophy degree at Oklahoma State University in May, 1986.

Professional Experience Research Geophysicist, Conoco Incorporated, Ponca City, Oklahoma, July, 1981, to December, 1984, Graduate Research Assistant, School of Electrical and Computer Engineering, Oklahoma State University, August, 1983, to December, 1985.

Awards/Affiliations Stanford School of Engineering Graduate Fellowship, Boettcher Foundation Scholarship, member Tau Beta Pi, Sigma Pi Sigma, Eta Kappa Nu, Institute of Electrical and Electronics Engineers, and Society of Exploration Geophysicists.