Errors of Estimating Near-Surface $Q$ - A Statistical Approach

Chih-Sung Chen, Yih Jeng*
Department of Earth Sciences, National Taiwan Normal University, Taipei
E-mail: geofv001@ntnu.edu.tw
Received December 18, 2010; revised February 7, 2011; accepted March 16, 2011

Abstract

Although various estimating methods have been developed for measuring $Q$ from near-surface seismic data, less thought has been given to the accuracy of $Q$ obtained. The errors of $Q$ depend on the ways of measuring $Q$ and the computation techniques used in estimating. The main purpose of this paper is to give a comprehensive evaluation for the accuracy of measuring near-surface $Q$. We discuss the possible origins from which errors may develop, and provide a statistical guide to the accuracy that can be expected. A set of real data based on the improved spectral ratio method for near-surface $Q$ was used as an example of validation and sensitivity analysis. The Bonferroni procedure was adopted for deriving the joint confidence intervals for $k$ and $n$ of the power law model. The same approach with modest modification may be applied to analyze the accuracy of $Q$ estimated by other methods.

Keywords: $Q$, Near Surface, Spectral Ratio, Seismic, Statistical Errors

1. Introduction

Wave energy attenuation is an important physics issue for many applications, and has been defined in geosciences as the quality factor $Q$ indicating the inverse of intrinsic attenuation. In other words, a large $Q$ implies small energy dissipation when wave propagates through the rock. By contrast, rocks with small $Q$ values will be a poor transmitter for wave propagating. In geophysical investigation, $Q$ is useful for describing rock properties, implementing inverse filters [1], and indicating fluid saturation better than the velocity ratio. In particular, anisotropy of $Q$ could give unique information on subsurface properties, such as the lithology and orientation of fractures [2,3]. Although the importance of $Q$ is fully recognized by geophysicists, the estimation of $Q$ is still controversial perhaps due to the practical difficulties of in-situ measurement using geophysical method and insufficient information of physical mechanisms to establish appropriate models. In view of this, some geophysicists suggest that directly measuring variations of wave velocity, amplitude, and frequency content due to absorption effects is more practical than estimating $Q$ [4]. However, in near surface seismic survey, an accurate $Q$ model is still exclusively needed for performing inverse $Q$ filtering to recover the deep reflection energy and to improve the resolution of the subsurface image. Two basic hypotheses are considered in the estimate of $Q$ for shallow earth. The conventional one is that the $Q$ model is independent of frequency; but some investigators suggest that $Q$ may be frequency dependent [5,6]. A theoretical derivation proposed by Jeng et al. [7] indicates that the $Q$ of shallow seismic waves is frequency dependent. The frequency dependent $Q$ has gained more attentions in recent inversion studies. Li et al. [8] have developed $Q$ estimates using both $t^*$ and spectral amplitudes for frequencies between 1 and 40 Hz, and yielded consistent results that $Q$ increases with depth and frequency up to 20 Hz within the Seattle Basin. Although the frequency dependent model is more practical in use, the standard error in the estimation of $Q$ for each given frequency may be enormous. Most seismologists attribute the limitations of the $Q$ estimation to the source factor, the data quality, the effect of time window length, and the velocity model. However, some peculiar problems may not be so easy to answer; for instance, the negative $Q$ is noted in some cases. Therefore, we suspect the assumption of energy dissipation of traveling plane waves may not be true for the shallow earth model. In addition to this, the analysis problems should also be a crucial factor. In order to give a comprehensive evaluation for the accuracy of $Q$, this paper examines the
commonly used definition of $Q$ from the near-surface point of view and the basic theory for $Q$ estimation along with a statistical analysis. We present a framework of determining the errors in near-surface $Q$, and provide a statistical guide to the accuracy that can be expected. A field example of data acquisition and processing methodology is also briefly described. The physics behind the data scatter may give us a better understanding of many systematic errors of near-surface $Q$ measurements.

2. Methodology and Background

The work discussed in this paper is part of a series of experiments related to the near-surface seismic exploration project that has been carried out at the geophysical laboratory of National Taiwan Normal University. Several test sites located mainly in northern and central Taiwan provide the real data for analysis. The test sites are very similar in geology, and consist basically of Recent alluvium and unconsolidated Quaternary sediments [9,10]. In the previous experiments, Jeng et al. [7,11] proposed an improved field measurement of near-surface $Q$ using a frequency dependent spectral ratio method. The field lay-out is shown in Figure 1. A multichannel seismograph with 4.5 Hz three-component geophones and 28 Hz, 40 Hz, and 100 Hz vertical geophones were used to acquire data of different dominant frequencies. A mechanical striking seismic source (Figure 2) and a 7-kg sledgehammer were used to generate P and S waves of different impacting energy. Figure 3 demonstrates the way of generating shear wave using the sledgehammer. A typical shot gather acquired in the field for picking the first arrivals is shown in Figure 4. This technique firstly estimates the $Q$ of any given frequency by fitting a straight line to the slope of the spectral ratio over a finite range of arrival-time difference of two receivers, and then a frequency dependent equation of $Q$ is determined by an optimum power law model. The final $Q$ of any particular frequency is then evaluated from the equation. In practice, similar to the problem of the conventional spectral ratio method, the ratio values rarely fall on a straight line that has been described by Patton [12]. Therefore the estimated $Q$ is far from a definite value no matter what method is applied. It is more like an approximate value with an estimating error.

The origin of the estimation errors of near-surface $Q$ values from spectral ratios can be considered from two aspects, i.e., the inherent and processing errors. The inherent error can be interpreted as a trade-off error that stems from the measurement of spectral ratios itself, and places fundamental limits on the accuracy of estimates of $Q$ [13]. The processing error depends predominantly on

Figure 1. Standard field layout of the experiment. Geophone intervals vary from 0.5 m to 2.0 m based on the site environments and sources applied. An engineering seismograph was employed to record data.

Figure 2. VAKIMPAK, a mechanical striking seismic source used in the field for generating P- and S-waves. The striking energy is much stronger than the sledgehammer, about 2500 Nm for P-wave and 2000 Nm for SH-wave.

Figure 3. Shear wave generating by using a sledgehammer striking against the end of a block of a wooden plank.
Figure 4. Typical raw shot gather data acquired in the experimental sites.

The model of measurement. Conventional estimating method is based upon the assumption of frequency independence that may be problematic in shallow $Q$ values [7,14] due to heterogeneity, variations in saturation and porosity of the near surface layers, thin bed scattering, leaky modes and so on. Given this, we investigated the accuracy of near-surface $Q$ in a broad view by way of the improved method, developed by Jeng et al. [7], which assumes frequency dependence.

3. Theoretical Consideration

3.1. Errors of Small Dissipation Assumption

The concepts of the seismic attenuation and the estimation from the amplitude spectral ratios in near-surface seismic data will be reviewed briefly before we discuss the accuracy of near-surface $Q$.

The fundamental definition of $Q$ is based upon the intrinsic property of rock, and can be described by the harmonic oscillation theory

$$Q = -2\pi \frac{E}{\Delta E},$$

where $E$ represents the elastic energy stored at maximum stress and strain, and $\Delta E$ is the energy loss per harmonic cycle. $Q$ is also defined as

$$Q = \frac{\pi f}{\alpha V}$$

where $\alpha$ is the exponential decay constant, $V$ is the wave velocity, and $f$ is the frequency [6,15]. The above definitions are valid if the attenuation is small, i.e., large $Q$ is assumed. A more appropriate equation is proposed by Hamilton [16] for large-dissipation (small $Q$):

$$\frac{1}{Q} = \frac{\alpha V}{\pi f} - \frac{\alpha^2 V^2}{4\pi f}$$

Let $Q_s$ denote the estimated $Q$ on the small-dissipation assumption ($Q > 10$) [15], and $Q_l$ the estimated $Q$ on the large-dissipation assumption, then Equation (3) becomes

$$Q_l = \frac{\pi f}{\alpha V} - \frac{\alpha V}{4\pi f} = Q_s - \frac{1}{4Q_s}$$

Most $Q$ values estimated in near surface are lower than $2\pi$, thus the error on small dissipation assumption can be determined by Equation (4).

3.2. Errors of the Over-Parameterized Model

White [14] has proposed that it is essential to parameterize the estimation model based on the principle of parsimony. According to this principle, fundamental limits on the accuracy of estimating $Q$ can be set up. The error of the over-parameterized model mainly originates from the calculation of spectral ratios. If the estimation of spectral ratios comes from multiple coherence analysis of two time intervals in surface seismic data, then the equation of the relative standard error is

$$s.e.(\bar{Q}) = \left( \frac{6Q_s^2}{\pi^2 (\Delta t)^2 F^3 T} \right)$$

where $Q_s$ denotes the sample mean of $Q$, $F$ is the bandwidth over which useful measurements can be made, and $T$ is the duration of the data segment [14].

3.3. Noise Effect

The primary assumption of the spectral ratios is the linearity of the data to be fitted. This assumption is sometimes invalid due to the peaks occurring in the spectral ratio. The spikes in the spectral ratio obtained from the conventional method are the contributions from the P-wave leaky modes and other noises. The leaky modes are the energy of surface waves (normal modes) below certain cutoff frequencies that leak through the half-space as body waves, and the phase velocity may exceed...
the body wave's velocity. Because the energy of leaky modes attenuates exponentially with distance, the spikes occurring at the spectral ratio can be reduced by increasing the source-receiver offset, but it is still a probable cause for unstable spectral ratios. The improved spectral ratio method reduces the effect of leaky modes by taking an optimum linear model from a variety of geophone pairs of different source-receiver offsets, then assuming a simple nonlinear power law regression model to estimate the Q as a function of frequency. The data uncertainties due to other noises are poorly known; therefore, the noises usually are assumed to be a zero-mean Gaussian process, and ignored in the inversion algorithm [17]. This is an oversimplified assumption about the noise, and adds errors to the Q estimation. Therefore, the standard error of the slope of the optimum linear regression model may reflect the residual error originating from the leaky modes and other noises.

4. Processing Errors

4.1. High and Low Estimates of the Regression Coefficient from the Least-Squares Fit

Equation (1) can be generalized for an arbitrarily small Q by taking the differential form:

$$-\frac{2\pi}{Q} = \frac{dE}{E} \cdot \frac{\tau}{d\tau} = \frac{\tau}{E} \cdot \frac{dE}{d\tau}$$

(6)

where \( \tau \) is the period of the wave or oscillation [18]. Integrating Equation (6) yields

$$E = E_0 e^{\frac{2\pi \tau}{Q}}.$$  

(7)

Since energy is proportional to the square of amplitude A, Equation (7) can be expressed as

$$S = S_0 e^{\frac{4\pi f}{Q}} = S_0 e^{\frac{2\pi f}{Q}}.$$  

(8)

A conventional equation of spectral ratios is derived from Equation (8) as

$$\ln \frac{S_1(Z_1, f)}{S_2(Z_2, f)} = C + \frac{\pi(Z_2 - Z_1)f}{Q}$$  

(9)

where \( S_i(Z_i, f) \) is the amplitude at frequency \( f \) of the signal recorded at offset \( Z_i \) and \( C \) is a constant that takes into account the source function, receiver function, and geometric function.

Jeng et al. [7] have improved the conventional spectral ratio method by measuring the two-way travel time difference rather than the difference of the offset. Accordingly, the equation of spectral ratios is given by

$$\ln \frac{S_1(Z_1, f)}{S_2(Z_2, f)} = C + \frac{\pi(t_2 - t_1)f}{Q} = C + \frac{\pi\Delta t f}{Q}$$  

(10)

In order to obtain a stable \( Q \), the improved spectral ratio method assumes that \( Q \) is frequency dependent. For each given frequency \( f \), the natural logarithm of the amplitude ratio \( S_1(Z_1, f)/S_2(Z_2, f) \) is taken; then plot it against the arrival time difference \( \Delta t \) of the two receivers at offsets \( Z_1 \) and \( Z_2 \). A large number of different \( \Delta t \)s and corresponding spectral ratios are estimated for the same frequency (normally 30 to 60 different \( \Delta t \)s were taken for one frequency); then the \( Q \) of the given frequency is determined by the slope of an optimum linear regression model.

The estimated standard error of the slope of the linear regression model is the square root of our estimator of variance, thus

$$s.e.(\beta) = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$$

(11)

where \( b \) is the sample mean of the slope and \( x \) represents the value of \( \Delta t \). The numerator \( s = \sqrt{ \frac{rs}{n-2} } \) is the square root of the residual variance in which \( rs \) is the residual sum of squares. The standard error in the estimate of the slope can be related to the standard error in estimated \( Q \) through Equation (10). The estimated limits for the slope \( b \) of the regression model with a 95% confidence interval are

$$b \pm \frac{t(0.025,n-2)s}{\sqrt{\sum(x_i - \bar{x})^2}}$$  

(12)

where \( t \) is the test for significance from the \( t \) table, 0.025 is one-half of the significance level of \( t \) test (usually denoted by \( a \)), and \( n - 2 \) is the degrees of freedom of the \( t \)-distribution [19-21]. These are also the estimated limits for the \( Q \) of a given frequency with a 95% confidence interval.

4.2. Standard Errors and Confidence of the Optimum Power Law Model

Since near-surface seismic data show strong evidence of frequency dependence [7], the true value of \( Q \) can only be approximated by estimating an optimum regression model related to the frequency. We calculated the error by estimating the deviation of our calculated \( Q \) from the power law model \( Q = kf^n \) where \( f \) is the frequency, \( k \) and \( n \) are constants to be determined. For a very small \( Q \), Equation (4) is suggested to correct the small-dissipation (large \( Q \)) assumption error before calculating the uncer-
tainty at this stage. Because the nonlinear regression model consists of two parameters, \(k\) and \(n\), thus we use confidence regions that involve separate intervals in each parameter. The joint confidence intervals for multi-regression parameters in nonlinear regression can be derived by the Bonferroni procedure [21]. A \([100(1 - \alpha)]\%\) joint confidence interval for \(k\) and \(n\) is given by

\[
k \pm p \left(1 - \frac{\alpha}{2}\right) \left(k_s\right)
\]

where \(\alpha\) is the significance level, \(p \left(1 - \frac{\alpha}{2}\right)\) is the \(100\) percentile of the standard normal distribution, and \(k_s\) and \(n_s\) are the standard error of parameters \(k\) and \(n\), respectively.

5. Example of Real Data

We use some typical field data acquired at the Keelung experimental site in northern Taiwan to demonstrate the statistical approach of the previous section. Figure 5 shows a representative numerical plot of the spectral ratios of \(P\)-wave versus arrival time difference at one particular frequency (60 Hz) for any two geophones of different offsets in one geophone spread (Table 1). The linear regression model of the data is \(Y = 0.05442 + 0.09025X\). The estimated standard error of the slope of the regression model was \(\text{s.e.}(b) = 0.012149\), and a 95\% confidence interval for the slope of the regression model was \(0.09025 \pm 0.0245\) since \(t(0.025, n - 2) = t(0.025, 37) = 2.02\) in Equation (12). The standard error in the estimate of \(Q\) can be related to the standard error of the slope of the regression model.

Table 2 is the frequency dependent near-surface \(Q\) values estimated for the \(P\)-wave data recorded at the same experimental site [22]. Inspection of the examples and Table 5 in Jeng et al. [7], these values are reasonable for the near surface \(Q\) estimation and more stable than those obtained by using the conventional spectral ratio method. Let us take the data set of \(\text{SHOT NO. P0430-2}\) for example. Given that \(Q = 5\) for the representative \(Q\) value in our case, data of \(T = 30\) ms with \(\Delta t = 5\) ms were analyzed. The seismic bandwidth was from 60 to 300 Hz normally. Then the error of the over-parameterized model for our near surface \(Q\) in this case is about 1.2. A desirable way of reducing the relatively large inherent standard error is to expand the duration of the analyzed data segment and their time separation because the bandwidth is wide enough for seismic frequency ranges.
Correction for the error of small dissipation assumption is another way to improve the accuracy of near surface \( Q \). However, it produces only an insignificantly small difference (about 1% of difference for \( Q \) with a value of 5) when compared with the standard errors from other origins.

For estimating the optimum power law model we used the Simplex estimation algorithm to do the nonlinear regression (Figure 6). From Equation (13) and Equation (14), the estimated parameters \( k \) and \( n \) of the power law model are
\[
\begin{align*}
    k &= 0.1340208 \\
    n &= 0.6975394
\end{align*}
\]
and the standard errors of parameters \( k \) and \( n \) are
\[
\begin{align*}
    k_s &= 0.04637 \\
    n_s &= 0.06452
\end{align*}
\]
Figure 6 shows the power law model for the \( Q \) values obtained at the Keelung experimental site. Because the \( 1 - \frac{\alpha}{2} \) 100 percentile of the standard normal distribution for a 95% confidence interval is 1.960 in Equation (13) and Equation (14), the approximated 95% confidence interval for \( k \) is
\[
0.04314 \leq k \leq 0.22491
\]
and for \( n \) is
\[
0.57108 \leq n \leq 0.824
\]
The joint confidence interval for the family coefficient approximates 90% for
\[
\begin{align*}
    0.04314 \leq k &\leq 0.22491 \\
    0.57108 \leq n &\leq 0.824
\end{align*}
\]
Figure 6. Power law model for the \( Q \) values obtained at the Keelung experimental site. The data used is \( P0430-2 \) shown in Table 2. The nonlinear regression model of the data is \( Q = 0.1340208 \times f^{0.6975394} \) with a correlation coefficient of 0.97445.

Figure 7. Plot of the residuals against the model predicted values showing the significant relationship. Dashed lines define the confidence bands. The data used is \( P0430-2 \) listed in Table 1. The linear regression model for residual values \( R \) to predicted values \( P \) of the data is \( R = 0.08052 - 0.01390 P \) with a correlation coefficient of \(-0.0610\).
\[
\frac{1}{Q} = \frac{1}{Q_1} + \frac{1}{Q_2}
\]

Neep et al. [25] propose that this equation is valid if at least one full wavelength propagates through the medium, and the relationship appears to be valid in the spectral ratio method despite the scattering \( Q \) having both positive and negative values. The scattering \( Q \) in the near-surface is more significant due to thin bed scattering, and therefore, it may contribute more errors to the total \( Q \) because the geometrical spreading factor is insignificant in the near-surface \( Q \) estimation [7].

The real data example in this paper demonstrates that the standard error of the estimated \( Q \) is strongly related to the numerical values of \( k \) and \( n_f \) of the power law model; thus the accuracy of the estimated \( Q \) depends on the frequency bandwidth of data. This is also a coincident result proposed by White [14].

The total error associated with the near-surface \( Q \) estimation can be obtained by summing up all the relative errors if they are independent and that the change of \( Q \) is linear. Because these conditions may not be true in reality, the total error only describes the maximum or the worst-case scenario.

7. Conclusions

The attributes of uncertainties in estimating \( Q \) depend on the methods used. The approach discussed in this study suggests that the accuracy of near surface \( Q \) is affected by a variety of factors resulting from errors of small dissipation assumption, over-parameterized model, noise effect, slope of the optimum linear regression model, and optimum power law model.

Except under exceptionally favorable conditions, the value of \( Q \) is always unstable and far from a definite value no matter what method is applied. Our statistical analysis can help understand the stability of the \( Q \) estimated. However, the noise effect is the most uncertain factor that may affect the accuracy of near surface \( Q \). Since no any filtering technique may remove noises completely, it is suggested that every endeavor should be made to acquire accurate estimates of error in the field. The errors in estimating \( Q \) can be handled well by the proposed method if the data acquisition errors have zero mean, are approximately Gaussian, and have been well estimated.

8. Acknowledgements

The authors thank Jeffrey C. Green for his assistance in manuscript preparation. Special thanks go to the editor and one anonymous reviewer for their interest in this work and constructive comments. This research was partially funded by the National Science Council of Taiwan, ROC, Grant NSC 98-2116-M-003-007.

9. References


