

An Effective Algorithm to Correct the Orientation Error and Time Shift of Downhole Seismograms

Hung-Chie Chiu^{1,*} and Huey-Chu Huang²

(Manuscript received 16 October 2002, in final form 26 February 2003)

ABSTRACT

Estimating the orientation error of a downhole seismometer, in general, is a two-parameter optimization problem. It requires determining both the time shift and orientation error at same time. Although this estimation involves a tedious computation, the resolution of this optimization is still limited by the grid size selected for global searching. To improve the computing efficiency and resolution, we derive an analytic formulation for calculating the maximum cross-correlation function between synthetic and rotated seismograms and apply it to an orientation-error-correction method developed by Chiu and Huang (2003). This modified method is capable of giving an accurate and efficient estimation of the orientation error in one step for a given time shift. This algorithm consists of two stages: (1) computing the synthetic seismograms at the downhole station, and (2) searching a time shift that yields a maximum cross-correlation between the rotated seismograms (observations) and the synthetic seismograms at the downhole station. Results show that this modified algorithm allows for a more flexible selection of data for analysis, and gives a fast, stable and accurate estimation of the orientation error.

(Key words: Orientation errors, Downhole seismometer, Synthetic seismogram)

1. INTRODUCTION

For a downhole seismometer (velocity-type seismometer or accelerometer) without a built-in compass, orientation errors must be corrected from recorded data after its first installation and upon any later reinstallation. In general, this is a two-parameter optimization problem. It requires a global search for the time shift due to unsynchronized recording, and the orientation error of the downhole seismometers due to the uncontrolled installation. Numerical optimization can be done by grid search; the possible range of these two parameters (time shift and

¹Institute of Earth Sciences, Academia Sinica, Taipei, Taiwan, ROC

²Institute of Seismology, National Chung-Cheng University, Chia-Yi, Taiwan, ROC

* *Corresponding author address:* Dr. Hung-Chie Chiu, Institute of Earth Sciences, Academia Sinica, P. O. Box 1-55, Nankang, Taipei, Taiwan, ROC; E-mail: chiu@earth.sinica.edu.tw

orientation error) can be divided into discrete values with a small interval and the best solution can be found by selecting a pair of time shift and orientation error that gives the highest cross-correlation coefficient. For example, the orientation error could be any value between 0 and 360° and the possible time shift can be any value between -2 and 2 s, if we assign the increment of the orientation error at 1 degree and time accuracy at 0.01 s, then we will have 1.44×10^5 grid points for numerical optimization with a resolution of one degree in orientation error and 0.01s in time shift.

Most correction algorithms based on the similarity of waveforms between the surface and the downhole waveforms (e.g., Seale and Archuletta 1989; Aster and Shearer 1991; Chiu et al. 1994). However, this similarity is low and limited in the narrow window of a seismogram. Chiu and Huang (2003) suggested selecting the SH wave on the free surface as the reference direction and to search the maximum cross-correlation coefficient between the downhole synthetic and observed SH waves in various orientations. The corrected angle corresponding to the maximum cross-correlation coefficient then can be considered as the orientation error of a downhole seismometer. Their method includes more data points for analyses and thus reduces the estimation uncertainty; however, all these methods have common difficulties in increasing the computing efficiency and resolution. A rise in resolution requires reducing the grid size on both of the time shift and orientation error, thus causing a tremendous increase in the computation time. Although the CPU time is not a big issue, a reliable estimation in both the time shift and the orientation error is still a tedious process and its resolution is limit by the grid interval selected for both the time shift and the orientation error. Using an elegant optimization method such as the Genetic Algorithm (e.g., Goldberg 1989) can increase the searching efficiency but it cannot improve the resolution. Developing an effective and robust algorithm that can reduce the computation time and increase the resolution is necessary. In this study, we make use of the advantages of the algorithm developed by Chiu and Huang (2003) and improve its orientation error estimation to reduce the computing time and increase the resolution.

2. METHOD

The steps involved in the present algorithm are two-fold: first, computing the synthetic seismograms at a downhole station, and secondly, searching for orientation and time shift that give a maximum cross-correlation between the rotated seismograms (observations) and the synthetic SH waves at the downhole station.

The basic assumption behind this algorithm is that the synthetic (derived from the surface recording) and the corrected downhole seismograms (observation) have a maximum cross-correlation coefficient (hereafter CCC) when both of their SH waves are in the same direction and the recordings are chronologically synchronized. Ideally, these two waveforms should be the same; however, deviations occur in an imperfect velocity model and an imperfect separation of the P, SH and SV waves. Nevertheless, compared with a surface recording, this synthetic seismogram has much higher similarity with downhole recordings.

The reasons for selecting the SH waves as the reference were given by Chiu and Huang (2003), and one numerical method that can calculate the downhole synthetic seismogram in a

layered medium using the surface recording as an input was given by (Huang and Chiu 1996). Here we will emphasize deriving a one-step algorithm for estimating the orientation error. To give a detailed description of this method, consider that a set of downhole seismograms in the SH-SV plan has two perpendicular components x_i and y_i . If x_i and y_i are rotated counter-clockwise by an angle θ to correct the orientation error, then we have two new components x'_i and y'_i . i.e.

$$\begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix}. \tag{1}$$

Let s_i be the synthetic seismogram at downhole station, then the CCC between s_i and x'_i is defined as

$$R_{x'_i s} = \frac{\sum_i s_i (x_i \cos \theta + y_i \sin \theta)}{\sqrt{\sum_i s_i^2} \sqrt{\sum_i (x_i \cos \theta + y_i \sin \theta)^2}}, \tag{2}$$

where summation is taken over the discrete data within a window.

Let $C_{xx} = \sum_i x_i x_i$, $C_{yy} = \sum_i y_i y_i$, $C_{xy} = \sum_i x_i y_i$, $C_{sx} = \sum_i s_i x_i$ and $C_{sy} = \sum_i s_i y_i$, and solve

the equation $\frac{\partial R_{x'_i s}}{\partial \theta} = 0$, we have

$$\tan \theta_o = \frac{C_{sx} C_{xy} - C_{sy} C_{xx}}{C_{sy} C_{xy} - C_{sx} C_{yy}} \quad \text{or} \quad \theta_o = \tan^{-1} \left(\frac{C_{sx} C_{xy} - C_{sy} C_{xx}}{C_{sy} C_{xy} - C_{sx} C_{yy}} \right). \tag{3}$$

The $R_{x'_i s}$ in Equation (2) can be either a maximum or a minimum at θ_o . To determine it, we need to compute the second derivative of $R_{x'_i s}$ at θ_o . Let $A = \sqrt{\sum_i s_i^2}$, $B = \sqrt{\sum_i (x_i \cos \theta + y_i \sin \theta)^2}$, and $r = \left[(C_{sx} C_{xy} - C_{sy} C_{xx})^2 + (C_{sy} C_{xy} - C_{sx} C_{yy})^2 \right]^{1/2}$ and by their definition, all the A, B and r are non-negative values. After substituting the θ_o into the second derivative of $R_{x'_i s}$, we have

$$\frac{\partial^2 R_{x'_i s}}{\partial \theta^2} \Big|_{\theta_o} = \frac{r}{AB^3} \geq 0. \tag{4}$$

Equations (4) implies that $R_{x'_i s}$ is only a minimum at the θ_o (we name it as θ_{\min} hereafter) as given in Equation (3). Fortunately, we can determine the θ_{\max} from θ_{\min} because the difference between these two θ s is always 180°. This relation can be obtained from equation (2) by substituting θ by $\theta+180$. After some manipulations we have

$$R_{x'_i s}(\theta + 180) = -R_{x'_i s}(\theta). \tag{5}$$

This equation further implies that the maximum and the minimum of R_x 's always have an 180° difference in its location in θ space, and we can determine θ_{\max} from a given θ_{\min} . The relationship between θ_{\max} and θ_{\min} depends on the definition of θ . If we define $0 \leq \theta < 360^\circ$ then we have

$$\theta_{\max} = \theta_{\min} + 180^\circ \quad \text{when} \quad 0 \leq \theta_{\min} < 180^\circ. \quad (6a)$$

$$\theta_{\max} = \theta_{\min} - 180^\circ \quad \text{when} \quad 180 \leq \theta_{\min} < 360^\circ. \quad (6b)$$

In the following section, we will outline three tests to validate this method step by step.

3. NUMERICAL TESTS

In the first test, we consider a noise-free experiment that assumes the synthetic and the observed waveforms at downhole are identical when orientation error has been corrected. The waveform selected for testing and the results of this test are shown in Fig. 1. Two orthogonal components of displacement waveforms x and y are considered as the downhole seismograms as shown in (a) and (b) in Fig. 1. If the downhole seismometer has an orientation error of 37.3° , the synthetic seismogram can be obtained by making a counter-clockwise rotation on waveforms x and y by a $\theta=37.3^\circ$ using Equation (1). The resulting waveform is shown in (c). Based on the waveforms (a), (b) and (c), we can estimate the θ_{\min} by applying Equation(3) which gives 217.3° . This θ_{\min} corresponding to a $\theta_{\max}=37.3^\circ$ and the corrected waveform as shown in (d). The CCC for various orientations is given at the bottom of Fig. 1. This result gives an exact estimation in the orientation error ($\theta_{\max}=37.3^\circ$ and $\text{CCC}=1$ at θ_{\max}). The CCC plot also indicates that θ_{\max} and θ_{\min} have a difference of 180° .

The shape of the CCC depends upon the seismograms x , y and s . In the second test, we select another set of waveforms [(a) and (b) in Fig. 2] for testing and let the orientation error be 37.621° . In addition, we let the synthetic seismogram have a time shift of 1.54 s. The final seismogram is shown in Fig. 2(c). This waveform set has a richer signal than that of the displacement waveform in the previous test. Furthermore, we can increase the decimal shift of the orientation error to 3 digits. In order to search for the optimized orientation error and time shift, we set the possible time shift in a range between -2.5 and 2.5 s with an increment of 0.005s. We should emphasize here that this grid point increment is the same with the sampling interval of these of waveforms. For each time shift, we can calculate its maximum CCC using equations (3) and (5). Comparing maximum CCC among all the possible time shifts, we found that the global maximum CCC is 1, which occurred at $\theta_{\max}=37.621^\circ$ and $t=-1.54$ s. Since the positive time shift is defined as an advance in the time shift, the negative time shift represents the time delay of the observed seismogram relative to the synthetic seismograms or the surface recording. The corrected waveform is shown in (d), which is identical to the waveform in (c). Again, in this noise-free case, we can obtain an exact solution for both the time shift and orientation error.

In the previous two tests, we calculated θ_{\max} indirectly from θ_{\min} using Equation (5). To show the detail of the CCC and the relationship between θ_{\max} and θ_{\min} , we now calculate the

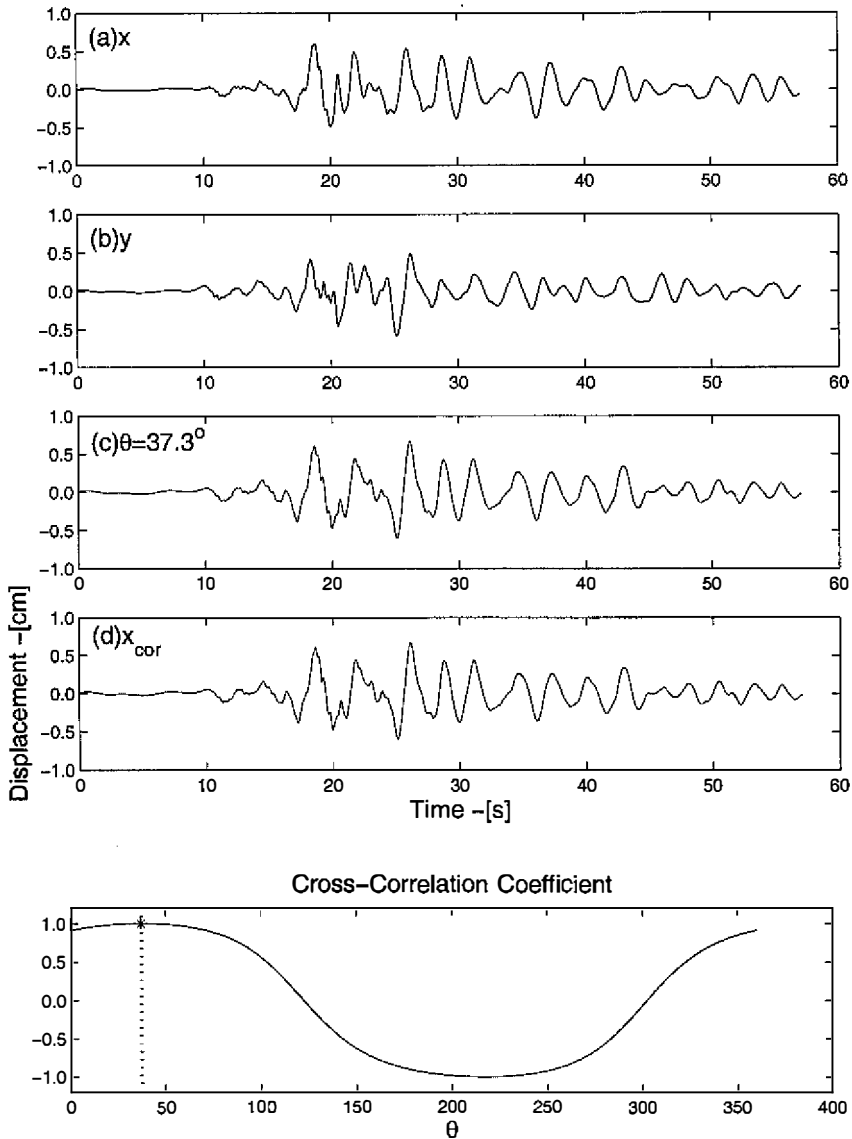


Fig. 1. Displacement waveforms in (a) and (b) are x and y components in the Cartesian coordinate system. (c) is the projection of (a) and (b) in a direction of new x axis (x') after a 37.3° counter-clockwise rotation. In the Test Model 1, waveforms in (a) and (b) are considered as observation, while (c) is considered as the noise-free synthetic seismogram at the downhole station. The corrected waveform is given in (d). The cross-correlation coefficient between various rotated components of the observation (x in equation (2)) and the synthetic seismogram s is shown in the bottom of the Figure. A dotted line marks the orientation error of 37.3° and a star marks the maximum cross-correlation coefficient.

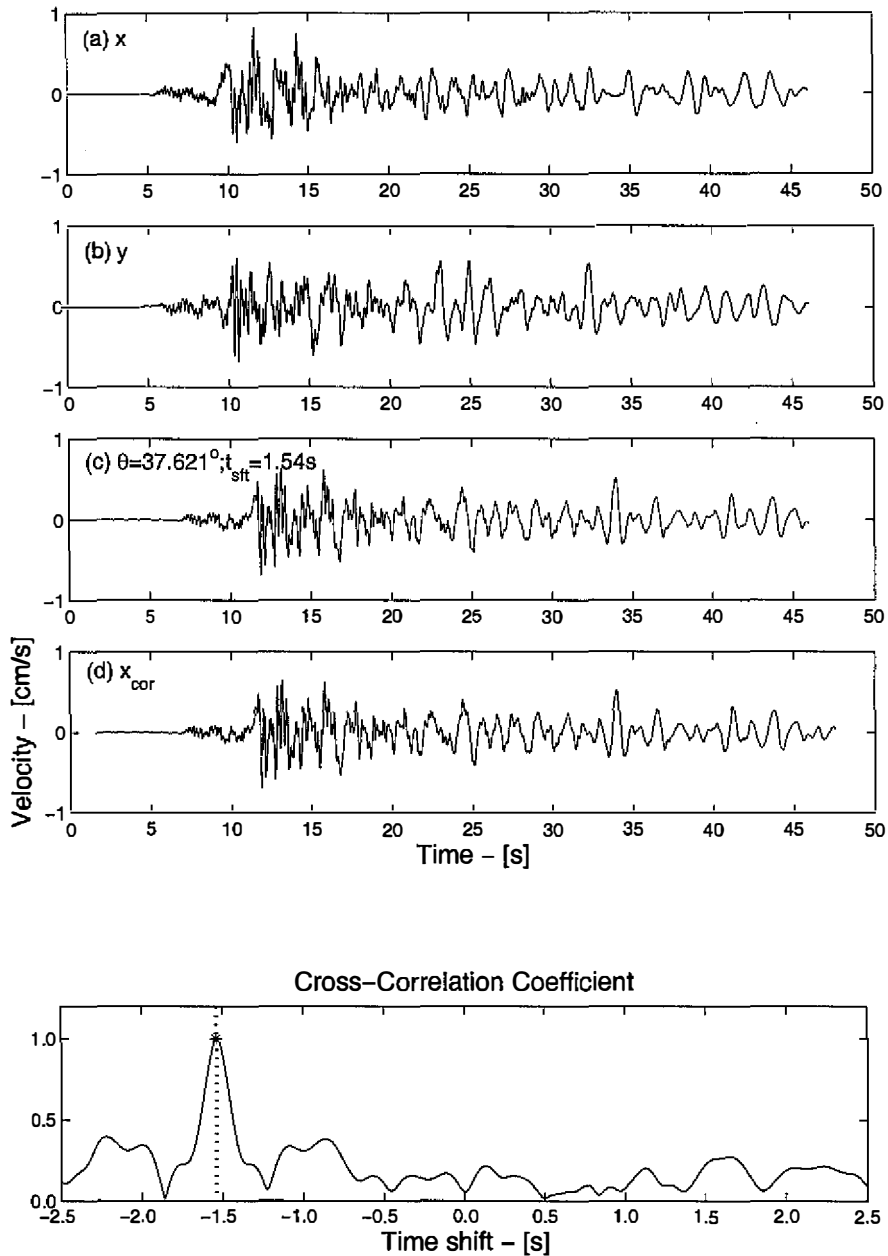


Fig. 2. The selected waveforms and results for the Test Model 2. The format of this figure is similar to that of the Fig. 1 except that the waveform in (c) has a time shift 1.535 s and the cross-correlation coefficients are plotted for various time shifts.

CCC over a range of the time shift between -2.5 and 2.5 s and orientation error in a range between 0° and 360° and present the result in a 3D plot as shown in Fig. 3. As expected, this two-parameter optimization has several local maxima and minima. The global maximum and minimum of these CCC are marked by solid dots in the figure and their corresponding time shift, orientation error and CCC are labeled beside the dots. For this noise-free case, the maximum CCC is 1 while the value for the minimum CCC is -1. These two extremes have the same t_{st} while the θ has a difference of 180° . This numerical example also shows that θ_{max} and θ_{min} have a difference of 180° .

Equation (3) is an exact solution in numerical form. It doesn't have any limitation on resolution. However, the numerical calculation may introduce an estimation error. The possible errors would be caused by the numerical error due to the rounding calculations of a computer, data noise, and the error caused by the grid point selected for finding the time shift. The rounding error of a computer can be ignored because it is relatively small. The amount of the data noise is case by case and will not be discussed here. In the next test, we will discuss the effect of the grid size of the time shift on the estimation of the orientation error and time shift. In this test, we still use the same waveforms as those in the Fig. 2. If the grid point coincides with the data sampling points, we always can obtain an exact time shift and orientation error and the corresponding CCC is 1 (as shown in Fig.2). In the following test, we will examine the estimation error when the grid points do not fall on the sampling points. To design

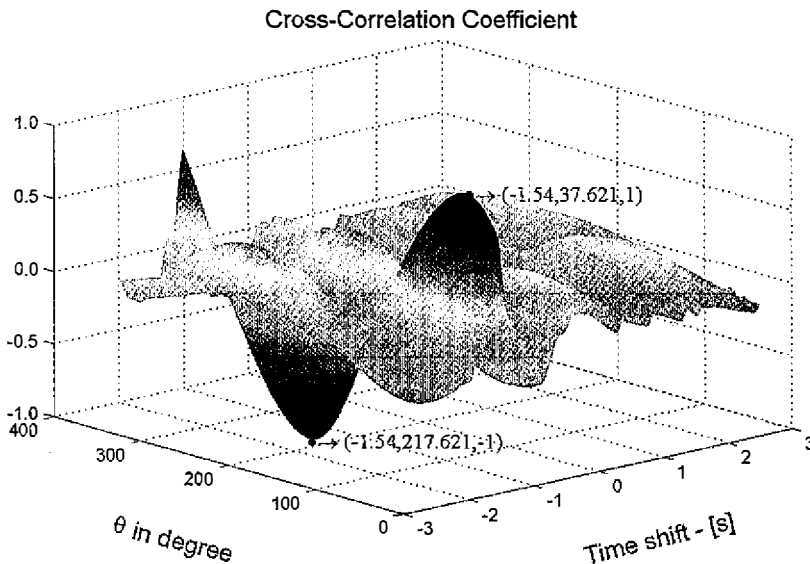


Fig. 3. The 3D plot of the cross-correlation coefficients based on the waveforms in Test Model 2. The range of the time shift is between -2.5 and 2.5 s and the orientation range is between 0° and 360° . Two dots in this figure mark the global maximum and minimum. Three numbers in the labels next to the dots are the corresponding time shift, orientation error and cross correlation coefficient respectively.

this experiment, we repeat the same calculation as that in the Fig. 2 for 20 time shifts. These time shifts distribute uniformly on three data sampling points (from 1.535 to 1.545) with an interval of 0.0005 s (1/10 of the sampling interval). Results are shown in Fig. 4. The top plot shows its effects on the coefficient of cross correlation. The solid line marks the exact solution, which has a value of 1. Among these 20 points, only three points which fall on the data sampling point (marked by a circle) have an exact solution. The worst correlation (the minimum CCC between 1.535 to 1.545 s) is found at the center between two sampling points and has a value of 0.9994. Nevertheless, this CCC is still very close to the noise-free case, which has a value of 1. The value of CCC increases as the grid points approach the sampling points. The effects of the grid point on the estimation of the time shift are shown in (b). The symbols in this plot are the same as that in (a). The worst estimation is still at the center between two sampling points. The maximum error of the estimation is one half of the sampling interval and the estimated value is rounded to be the value at the sampling point. The effects of the grid point on the estimation of the orientation error are shown in (c). The same symbols are for this plot.

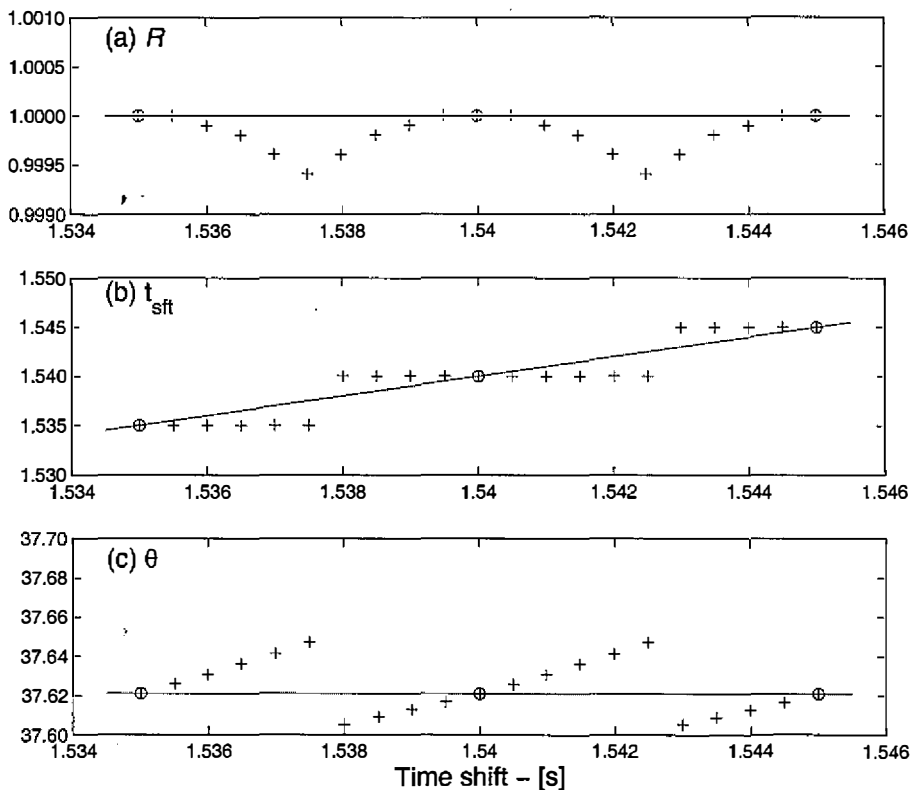


Fig. 4. The effects of grid points on the cross-correlation coefficient (a), time-shift estimation (b) and the orientation error estimation (c). ' + ' marks the test points, ' o ' shows the test point on the data sampling point and the solid line gives the exact solution.

Again, the worst case is still at the center of two sampling points. For this case, the maximum error in orientation estimation is 0.03° . Of course, the estimation errors shown in this test will change as the data or θ changes. Overall, the estimation error due to the selecting the grid point is proportional to the time difference between the sampling point and grid point. This error is small and a decrease in grid interval will cause a decrease in the estimation error.

The data selected for the last test is from the downhole data set recorded at the Dahan Downhole Array, Hualien, Taiwan. This earthquake occurred on 21 September 1999 in the southwest of the Array. The local magnitude of this earthquake was 5.4 and its Epicenter was about 30 km from the array. Before proceeding farther, a window must be selected for the data for analysis. To select a proper window, both the nonlinear effect and the effect of the imperfect separation of SH waves from a three-component recording need to be considered. The non-linear effects only occur when ground shaking is strong while the imperfect separation of SH waves is always found in the later arrivals (the later portion of a waveform). Since the non-linear behavior was not found in this data set, the selection of a 12-s window is mainly for avoiding in imperfect separation of the SH waves. The original seismogram of this data was recorded on acceleration; however, we selected the velocity waveforms for estimating the orientation errors because it is better than an acceleration or displacement waveform for such study (Chiu and Huang 2003). This data set was evaluated by Chiu and Huang (2003) and proved to be one of the best recordings suitable for estimating the orientation errors. The orientation errors of these downhole accelerometers at depths of 50, 100 and 200 m estimated by an independent approach (Chiu and Huang 2003) were 40° , 114° and 285° , respectively. In this case, all downhole recordings were recorded in the same time system and no time shift correction was necessary. The θ_{\min} estimated using equation (3) were 219.451° , 289.7310° and 104.8816° which corresponding to θ_{\max} were 39.4591° , 113.7310° and 284.8816° . The comparisons of the synthetic and the corrected waveforms are shown in Fig. 5. Three pairs of waveforms from the top to the bottom are synthetic-observed SH waveform pairs at depths of 50, 100 and 200 m. It is clear these waveform pairs are very similar.

Further comparisons of these velocity waveforms before and after orientation correction are shown in Fig. 6. The amplitudes of these waveforms are plotted on the same scale and the vertical distance between two traces is proportional to the depth of the recording accelerometers. Before the correction (top), no consistent phases (arrivals) could be found among these waveforms. After the correction, all the phases (bottom) displayed a very good alignment after the orientation correction. From the corrected waveforms, very clear up-going and down-going primary S waves were obvious. Both of these results suggest that the orientation errors of these downhole accelerometers have been corrected accurately.

4. DISCUSSION AND CONCLUSIONS

This algorithm has its disadvantages and advantages when compared with other methods. The disadvantage is that we must know the velocity model in advance. In general, this doesn't become a serious problem because all the downhole array sites must have a velocity well logging before installation. On the other hand, the main advantage of this algorithm is in its

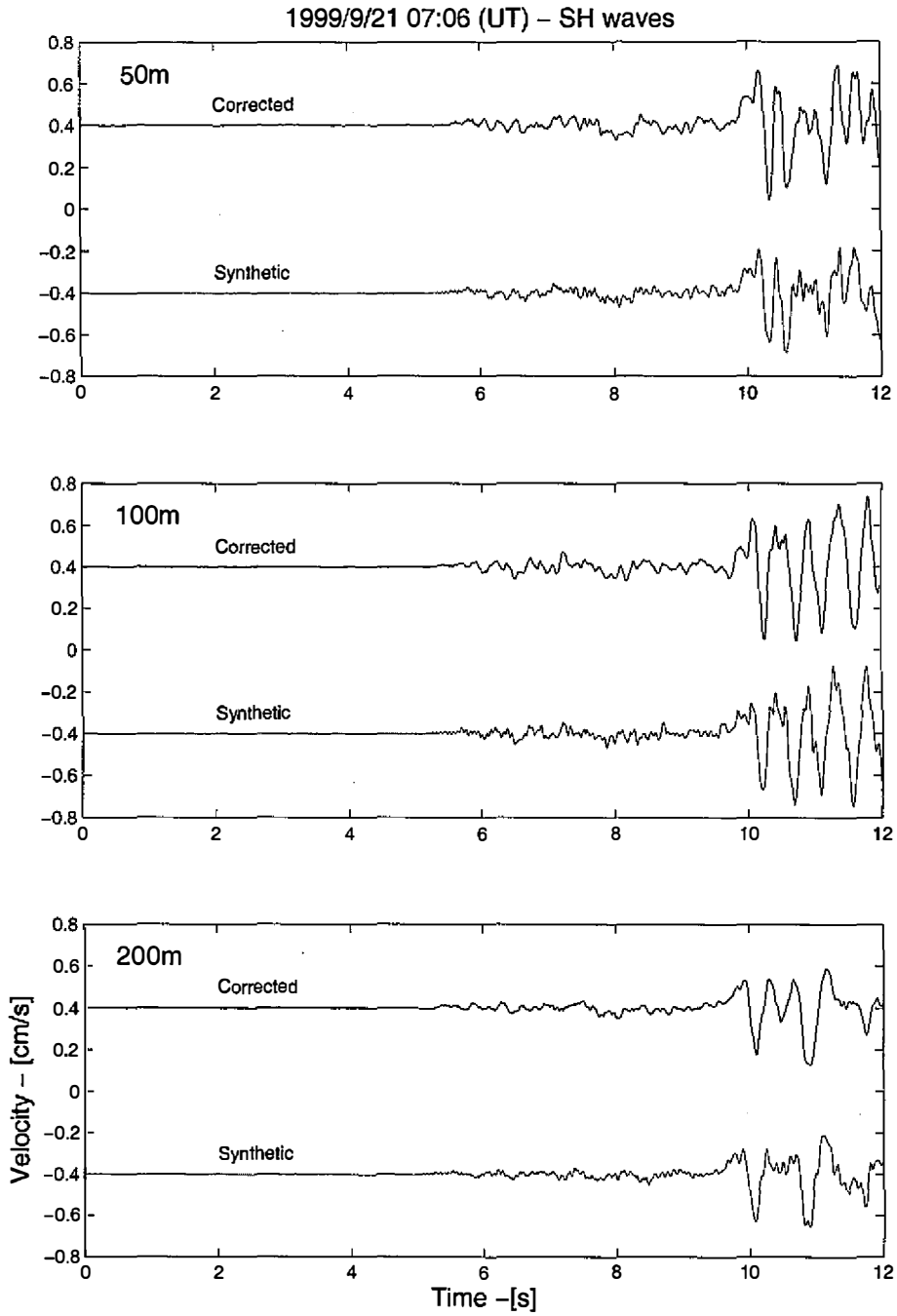


Fig. 5. Comparisons between the corrected and synthetic waveforms at the depths of 50 (top), 100 (middle) and 200 m (bottom). The synthetic SH waves are calculated base on the 1D layered model.

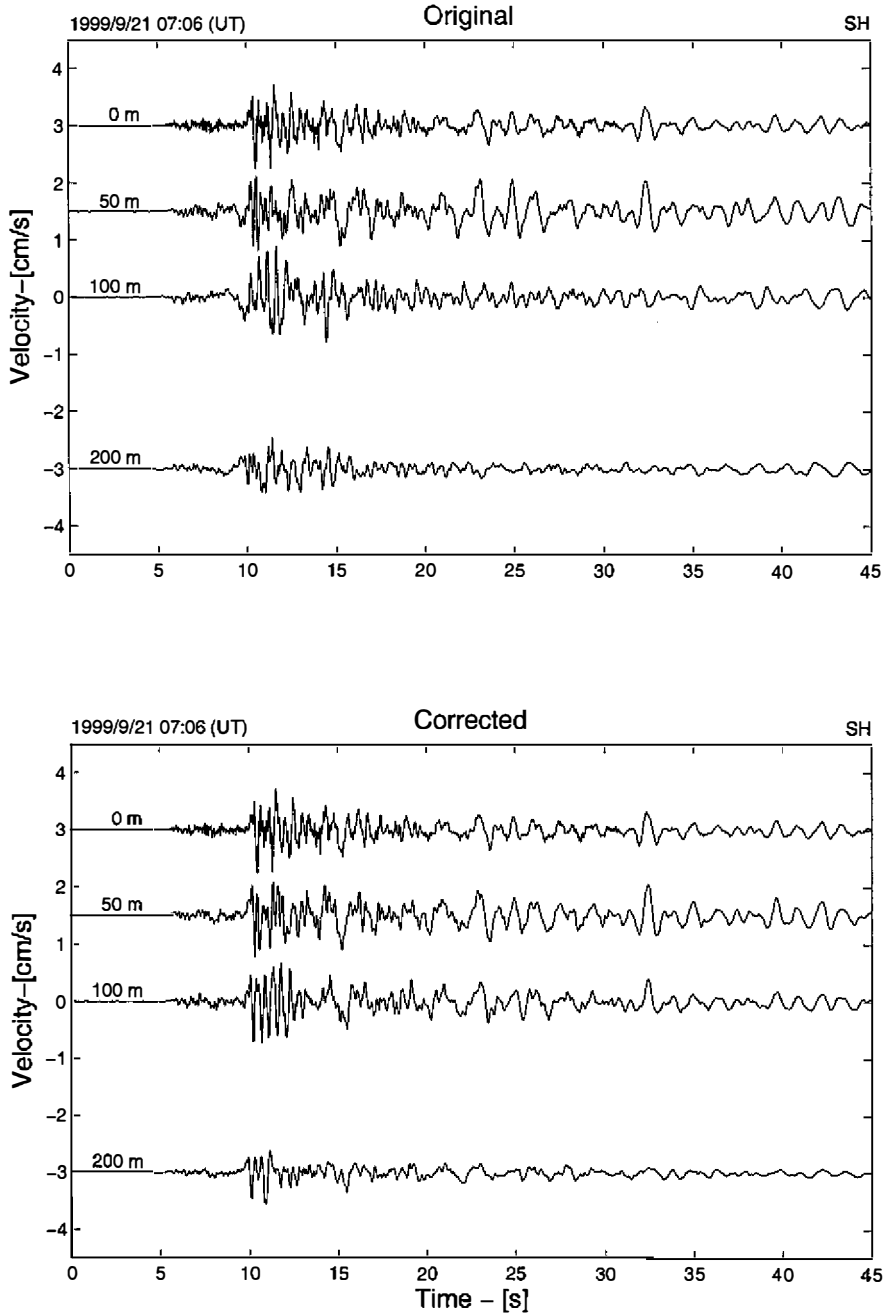


Fig. 6. The uncorrected (top) and the corrected (bottom) velocity waveforms for orientation error. The amplitude of these SH waves are plotted on the same scale and the distance between two traces is proportional to the depth of the accelerometers.

capacity to include more data points for analysis. The second advantage is that we can estimate the orientation error in one-step calculation. Compared with the grid-search method (Chiu and Huang 2003), this method can significantly reduce the computation time and increase the estimation resolution.

A good match in phase and the underestimation of the amplitude as shown in Figs. 5 and 6 imply that the average velocity model is sufficient. But, a further improvement on waveform matching requires a refined velocity model.

In most cases, the shallow structure can be a good approximation for a horizontal layered medium. For such cases, we can apply the Haskell-Thomson matrix to calculate the downhole seismograms. This study did not discuss the case of a non-horizontal layered medium. However, if the synthetic seismograms for such a velocity model can be obtained, we can extend this method for use in a non-horizontal layered medium.

Acknowledgments This research was supported by Academia Sinica and the National Science Council (under the Grant NSC 91-2119-M-001-017), Taiwan, Republic of China. The support from these organizations is gratefully acknowledged.

REFERENCES

- Aster, R. C., and P. M. Shearer, 1991: High-frequency borehole seismograms recorded in the San Jacinto fault zone, southern California: Part I. Polarization. *Bull. Seism. Soc. Am.*, **81**, 1057-1080.
- Chiu, H. C., 1997: Stable baseline correction of digital strong-motion data, *Bull. Seism. Soc. Am.*, **87**, 932-944.
- Chiu, H. C., 2001: Data Files from the SMART-2 Strong-Motion Array for the Chi-Chi Earthquake. *Bull. Seism. Soc. Am.*, **91**, 1391-1392.
- Chiu, H. C., and H. C. Huang, 2003: Estimating the orientation error using the maximum cross-correlation coefficient between the observed and synthetic waves. *Journal of Seismology* (in press).
- Chiu, H. C., H. C. Huang, C. L. Leu, and S. D. Ni, 1994: Application of polarization analysis in correcting the orientation error of downhole seismometers. *Earthquake Eng. and Struct. Dyn.*, **23**, 1069-1078.
- Goldberg, D. E., 1989: Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley.
- Haskell, N. A., 1953: The dispersion of surface waves on multilayered media. *Bull. Seism. Soc. Am.*, **43**, 17-35.
- Haskell, N. A., 1960: Crustal reflection of plane SH waves. *J. Geophys. Res.*, **65**, 4147-4150.
- Haskell, N. A., 1962: Crustal reflection of plane P and SV waves. *J. Geophys. Res.*, **67**, 4751-4767.
- Huang, H. C., and H. C. Chiu, 1996: Estimation of site amplification from Dahan downhole recordings. *Earthquake Eng. Struct. Dyn.*, **25**, 319-332.
- Seale, S. and R. Archuleta 1989: Site amplification and attenuation of strong ground motion. *Bull. Seism. Soc. Am.*, **79**, 1673-1695.