Quality Assessment of Seismic Data Sets and the Impact on Interval Velocity Estimates in DST

Erick Baziw
Baziw Consulting Engineers Ltd., Vancouver, Canada
Gerald Verbeek
Baziw Consulting Engineers Ltd., Vancouver, Canada

ABSTRACT: Downhole Seismic Testing (DST) is an important geotechnical testing technique for site characterization as it provides low strain ($<10^{-5}$) *in-situ* interval compression (*Vp*) and shear (*V*s) wave velocity estimates. These velocities are determined by obtaining relative arrival times of source waves as they travel through the stratigraphy and are recorded by one or more vertically offset seismic sensors. A challenging aspect of this process is to characterize the seismic data sets to determine the analysis method that will result in the most accurate interval velocity values. This paper introduces an assessment technique which utilizes linearity estimates (i.e., hodograms fitting straight lines) from the polarization analysis in conjunction with cross-correlation coefficient calculations of the full waveforms as well as the deviation of the source wave frequency spectrum from a desirable bell-shaped curve. The paper will provide an overview of the technique and include some examples of actual data sets where the technique was applied to analyze those data sets.

1. INTRODUCTION

In general terms, Downhole Seismic Testing (DST) such as Seismic Cone Penetration Testing (SCPT) is a geotechnical technique for measuring in-situ shear and compression wave velocities (V_S and V_P respectively). The main goal in DST is to obtain arrival times as the source wave travels through the soil profile of interest, and from these arrival times the velocities are then calculated. Figure 1 shows a schematic of the typical DST configuration: a seismic source is used to generate a seismic wave train at the ground surface. One or more downhole seismic receivers are used to record the seismic wave train at predefined depth increments. The downhole receiver(s) may be positioned at selected test depths in a borehole or advanced as part of an instrumentation package as in the case of SCPT. When triggered by the seismic source a data recording system records the response of the downhole receiver(s).

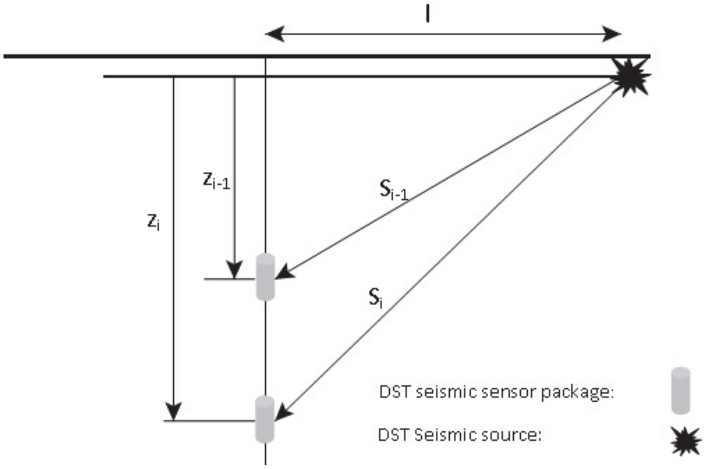

Figure 1. Schematic of the typical DST configuration (Baziw and Verbeek, 2014a)

The most common form of DST is the analysis of horizontally polarized shear waves (SH waves). These waves are commonly generated by applying a hammer blow laterally to the sides of special designed plate that is pushed into the soil at the surface. Typically shear waves are generated on both the right side (RS) and left (LS) of the seismic probe, which provides two independent sets of seismic traces at each depth increment and thus two independent interval velocity estimates. The field data examples outlined in this paper are obtained in this manner. Details of the seismic cone and comparisons with the crosshole results at several sites have been described by Campanella et al., and the test method is also described in ASTM D7400.

Once the test has been performed three important issues arise during the analysis of the test data. First and foremost there is the issue of the quality of the acquired seismic data sets (Seismic Trace Characterization (STC)). Secondly the analyst needs to determine the most appropriate signal processing techniques to obtain accurate interval velocity estimates, and finally what is the appropriate confidence level in the calculated interval velocities estimates (Interval Velocity Characterization (IVC))?

In recent years the authors have focused on the development of techniques and algorithms to address these three issues, with the ultimate goal to develop signal processing algorithms that can be applied in batch mode. A major component of this effort was to define independent STC parameters by analyzing in both the time and the frequency domain a large number of DST data sets (incl. those with poor signal-to-noise ratios, which are typically due to near field responses, source wave reflections, or “dirty” signals due to poor source-ground coupling). These parameters can then be fused together into a single classification, and also provide guidance concerning the most appropriate signal processing technique. Initial work in this area (Baziw and Verbeek, 2016a and 2016b) resulted in the development of three seismic trace parameters:

- the Cross Correlation Coefficient (CCC) of the full waveforms at the particular depth and the preceding depth.
- the linearity estimates (LIN) from polarization analysis.
- the Signal Shape Parameter (SSP).

In this paper these parameters are succinctly outlined, after which it is described how each parameter can provide guidance during the data analysis process to obtain more accurate values for the calculated interval velocities. This is illustrated through actual field data that were obtained with high precision and high bandwidth (1 Hz to 10 KHz) piezoelectric accelerometers with integrated operational amplifiers. These accelerometers have highly desirable rise and decay times of approximately 5 μs, which ensures that the acoustic waves and ambient noise are recorded with minimal or no sensor distortion.

2. THE STC PARAMETERS

2.1 Seismic Trace Characterization Parameters

STC Parameter 1: Cross- Correlation Coefficient

The cross-correlation between two time or distance offset seismograms is given as (Gelb 1974)

$$\varphi_{xy}(\tau) = \sum_{k} X_k \, Y_{k+\tau} \tag{1}$$

where $\varphi_{xy}(\tau)$ is the cross-correlation function, Y_k the sampled data at distance 1 and at sample time k, X_k the sampled data at distance 2 at sample time k, and τ the time shift between the two sets of recorded waves (note: distance 2 > distance 1). The value of the time shift at the maximum cross-correlation value is assumed to be the relative travel time difference, Δt, for the source wave to travel the distance increment. This technique has several advantages over selecting time markers within the seismogram (Baziw 1993, 2002), among others the human bias associated with visually selecting a reference point or time marker is minimized.

Normalizing the cross-correlation of the zero mean seismic signals by their standard deviations gives the cross-correlation coefficient:

$$\rho_{xy}(\tau) = \sum_k X_k Y_{k+\tau} \Big/ \sqrt{\sum_k X_k^2} \sqrt{\sum_k Y_K^2} \tag{2}$$

The CCC between the two DST waves is typically used to assess the quality of the interval velocity estimate as this parameter gives an indication of the similarity between the two waves being correlated. While on its own the CCC has proven to be an unreliable indicator of the overall quality of a seismic trace (since it is highly dependent on the digital filter applied to the raw seismic signals), it is still a useful component of seismic trace characterization. As an STC parameter the CCC value is calculated on the full waveforms after applying polarization analysis.

STC Parameter 2: Linearity Estimates from the Polarization Analysis

Polarization Analysis (PA) is applied when rotating the acquired X(t), Y(t) and Z(t) seismic recordings onto the full waveform axis. In PA the full seismic waveform's angle of incident is determined using hodograms, and rectilinearity estimates are obtained by calculating the covariance matrix of the orthogonal X(t), Y(t) and Z(t) seismic trace recordings (Kanasewich, 1981; Baziw *et. al.*, 2004b).

The PA begins with applying a time window to the seismic event of interest, after which a hodogram is created by plotting the X(t), Y(t), and Z(t) component seismic time series amplitudes against one another within this time window. Least squares straight line best fits are then applied to the hodograms and the slopes of these straight lines provide angle of incidence information, which allow the the X(t), Y(t) and Z(t) seismic responses to be rotated onto the full waveform. If SH(t) wave analysis is being carried out then only one Y(t) vs X(t) hodogram is required to obtain an angle of incidence estimate and subsequently the X(t) and Y(t) responses are rotated onto the SH(t) axis.

Next a covariance matrix is calculated for the X(t), Y(t) and Z(t) recordings over the hodogram time window specified (obviously in case of a SH wave analysis only one two dimensional covariance needs to be calculated for the X(t) and Y(t) axis recordings). This covariance matrix is defined as follows (using the notation of Kanasewich (1981):

$$\boldsymbol{V}_{xyz} = \begin{bmatrix} Var[X(t)] & Cov[X(t),Y(t)] & Cov[X(t),Z(t)] \\ Cov[Y(t),X(t)] & Var[Y(t)] & Cov[Y(t),Z(t)] \\ Cov[Z(t),X(t)] & Cov[Z(t),Y(t)] & Var[Z(t)] \end{bmatrix} \tag{3}$$

In (3) Var and Cov are abbreviations for variance and covariance, respectively. The Var of a variable (e.g., X(t)) is given as

$$Var[X(t)] = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_x)^2 \tag{4}$$

where μ_x denotes the mean of the variable X(t). The Cov of two variables (e.g., X(t) and Y(t)) in discrete form is given as

$$Cov[X(t), Y(t)] = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu_x)(Y_i - \mu_y) \tag{5}$$

An estimate of the rectilinearity of the particle motion over the specified hodogram time window is obtained by diagonalizing the covariance matrix ((3)) and subsequently calculating the ratio of the principal axis of the diagonalized matrix. A measure of the rectilinearity is referred to as linearity and it is calculated as follows:

$$F(\lambda_1, \lambda_2) = 1 - \left({}^{\lambda_2}/_{\lambda_1}\right) \tag{6}$$

where λ_1 and λ_2 denote the largest eigenvalue and next largest eigenvalue of the diagonalized covariance matrix, respectively. The linearity approaches unity when the rectilinearity is high ($\lambda_1 \gg \lambda_2$) and approaches zero when the rectilinearity is low ($\lambda_1 \approx \lambda_2$).

In DST interval velocity estimation it is desired to have data sets with linearity values near unity. This will be the case for seismic traces recorded in a transverse isotropic medium with minimal measurement noise, clean source waves, and no signal distortions (e.g., reflections).

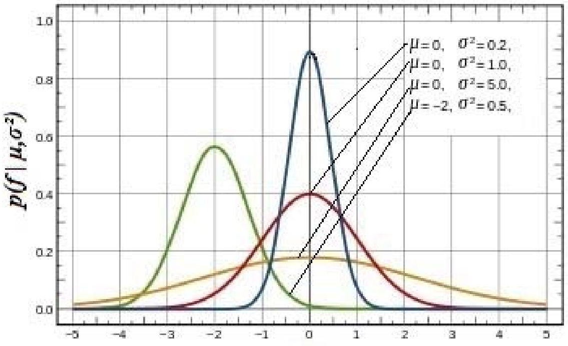

Figure 2: Example of normal pdfs for varying μ and σ^2 values. (Baziw and Verbeek (2016a)

STC Parameter 3: Signal Shape Parameter from Frequency Spectrum "Bell Curve" Fitting

Based upon frequency spectrum analysis of large sets of DST data it was determined that the shapes of high SNR DST data sets had frequency spectrums closely resembling Gaussian bell-shape pdf curves (Baziw and Verbeek, 2016b), which can be described as follows:

$$p(f\,|\,\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(f-\mu)^2}{2\sigma^2}} \tag{7}$$

where μ denotes the mean or expectation of the distribution and σ denotes the standard deviation with variance σ^2. Based on this observation a STC parameter was developed which quantified the deviation of the shape of the frequency spectrum of the seismic trace under analysis from a bell-shaped pdf curve.

2.2 Seismic Trace Classification Methodology

For the classification of microseismic source location estimation a methodology was developed by Ge (Ge and Mottahed 1993 and 1994; Ge 2003). As part of this method various parameters of the acquired microseismic data set are established, and then these parameters are fused together based upon analytic, derived and evolving empirical relationships to generate a microseismic source location estimate "Rank" varying from A (very good), B (good), C (acceptable), and D (not acceptable). In similar fashion the three STC parameters are fused together to provide for a quality assessment of the seismic traces using the following equation:

$$STC = 0.4 \times CCC + 0.18 \times (LIN_1 + LIN_2) + 0.12 \times (SSP_1 + SSP_2) \tag{8}$$

with CCC the cross correlation coefficient between the full waveforms obtained at depth1 and depth2 (where depth2 > depth1).
LIN_1 - the linearity value for seismic traces acquired at depth 1
LIN_2 - the linearity value for seismic traces acquired at depth 2
SSP_1 - the signal shape parameter for seismic traces acquired at depth 1
SSP_2 - the signal shape parameter for seismic traces acquired at depth 2

This value is then converted into a grade ranging from A to F as shown in Table 1, where A is highly desirable and F is unusable. However, the STC rank is automatically set to D (if not already set to D or F) when LIN < 0.78, SSP < 0.6 or CCC < 0.7. It should be noted that the proposed classification is based on a re-evaluation of many data sets previously processed (from over 40 different sites around the world, covering over 4000 seismic traces), but as additional data sets are analyzed the classification may well be refined either through adjustments of the constants in equation (8) and/or by adding additional parameters.

Table 1. Seismic Trace Classification and Description

STC Numeric Value [0-1]	STC Rank [A-F]	STC Description
0.9 to 1.0	A	very good to good
0.8 to 0.9	B	good to acceptable
0.7 to 0.8	C	acceptable to questionable
0.65 to 0.7	D	questionable to unacceptable
< 0.65	F	Unacceptable

Originally it was assumed that the rank of the seismic traces also indicated the accuracy of the calculated interval velocities. However, as the classifications were applied on more data sets it became apparent that instead it points towards the most appropriate signal processing technique for the data sets under review, resulting in a more accurate assessment of the interval velocities. Based on that finding the recommended analysis method for seismic traces now begins with the application of a minimal digital frequency filter to acquired time series (to eliminate obvious measurement noise components from the seismic trace), after which the STC parameters and the associated STC rank is established. For those traces with a low STC rank the values of the individual STC parameter are used to define the appropriate signal processing technique. Using these techniques the interval velocities are then estimated for the RS and the LS for each depth interval. In addition the spread is calculated using the following calculation:

$$\text{Spread} = \tfrac{1}{2} \times (\text{LS Interval Velocity} - \text{RS Interval Velocity})/\text{Average Interval Velocity} \qquad (9)$$

The spread can then be used as the basis for the IVC. If it is less than 10 % then there is good correlation between the left and the right side and therefore the calculated average interval velocity results can be used for engineering. In case the spread is more than 10 % greater weight should be given to the results from the side with the higher STC rank and the calculated interval velocity results should be considered indicative.

In case only one set of seismic traces is available (i.e. only LS or RS) the STC rank can give some indication of the reliability of the derived interval velocities (e.g. those from a set with an A rank at each depth interval are more than likely more reliable than those for a set with only D ranks), but this assessment should be used with extreme care.

3. GUIDANCE FOR SIGNAL PROCESSING BASED ON STC PARAMETERS

3.1 Low cross-correlation coefficients

The CCC values are derived from the calculated full waveforms and they quantify the similarity in shape and form between the traces at each depth increment. The CCC value can be very important for relative large depth increments where significant absorption is taking place and higher frequency components of the source wave are attenuated. For standard DST investigations, where the depth increment is relatively small, the CCC value generally reflects issues that also affect the LIN and/or SSP values, but in certain cases the CCC offer another "look" at the data independent of those two parameters. This can be especially true for near surface investigations where relatively small source-sensor radial offsets can result in significant near-field recordings, in which case the LIN and SSP values are still relatively high, while the CCC is relatively low.

Figure 3(A) illustrates low SNR X and Y axis seismic recordings at depths 1m and 2m with the LIN, SSP and CCC values as shown in Table 2:

Table 2. STC parameter values

Depth [m]	LIN [0-1]	SSP [0-1]	CCC [0-1]
1.000	0.8011	0.744	N/A
2.000	0.8818	0.916	0.75

The relatively low CCC value is more than likely due to source wave skewing as illustrated in Fig. 3(B). A signal decay was applied on the 1m full waveform recording to address the trough skewing and the corresponding low CCC value as illustrated in Fig. 3(C)

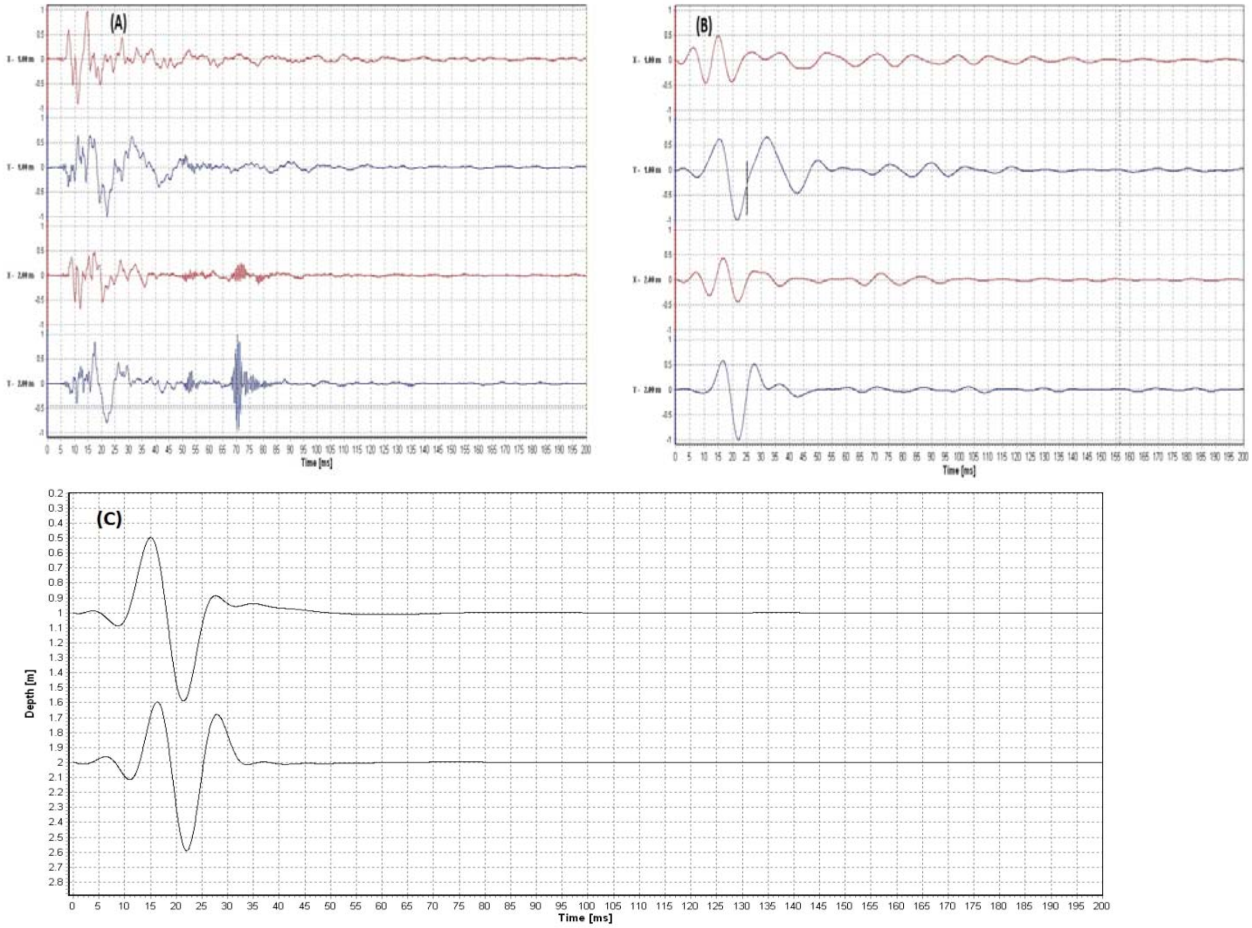

Figure 3: (A) Low SNR near field unfiltered recordings. (B) Filtered traces of (20Hz to 130Hz) outlining source trough skewing on the 1m Y axis recording. (C) Full waveforms with trough skewing on 1m recording minimized by applying signal decay.

3.1 Low linearity values

If the LIN parameter is low there are two possible ways to address it. First, by applying a narrower time window the dominant source wave response (either peak or trough) is more clearly defined and as a result the incorporation of measurement noise is minimized. The outcome is then most likely an improved rotation of the individual response onto the full waveform axis.

However, there are also cases when the low linearity is the result of very low SNR recordings on a specific seismic sensor axis, in which case time windowing will not resolve the issue. In those situations the appropriate action is to drop the trace with the low SNR recording and only using the higher SNR recording. This phenomenon is outlined in Fig. 4 where the X axis response has significant interference on the source wave peak compared to the corresponding source wave trough recording on the Y axis, resulting in a linearity value of 0.49. This interference results in both a broadening and skewing of the peak, and this adverse effect can result in significant errors in arrival time estimation. Since the Y-axis response has clearly a higher SNR the appropriate action is to use only the Y axis response for the analysis.

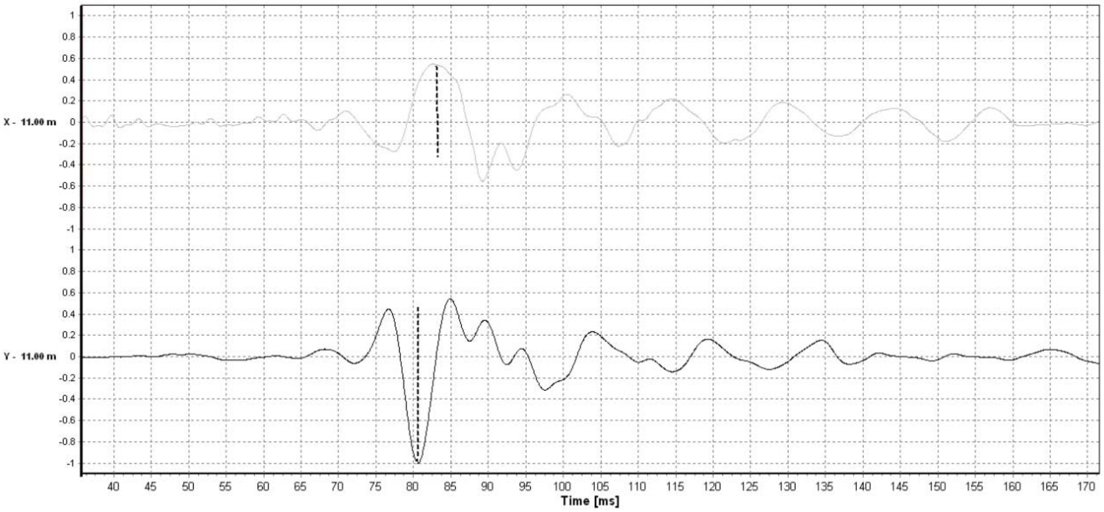

Figure 4: Significant source wave interference on the X axis resulting in peak skewing and broadening.

3.2 Low signal shape parameter values

"Ringing" due to measurement noise can be pronounced within a DST seismic trace in the presence of a very strong source wave maximum peak or trough. This "ringing" will decrease the SSP value and can lead to errors in the automated arrival time calculation from a batch analysis utilizing the cross-correlation function. Figure 5 illustrates this phenomenon. It shows a very strong source wave peak on the X axis (light grey trace) and a corresponding strong trough on the Y axis (black trace). Superimposed on the high SNR source responses is measurement noise which has "ringing" characteristics. The frequency spectrum for the Y axis response illustrated in Fig. 5(a) is illustrated in Fig. 5(b) (light grey trace). The best fit "bell-curve" is the black bold trace in Fig. 5(b), which corresponds with a SSP value of 0.41. Figure 5(c) illustrates the isolation of the dominant peak and trough by applying a time windowing algorithm

Another cause of low SSP values is excessive measurement noise, which can be addressed by applying more aggressive digital filtering to increase the SNR of the recorded traces. For example, traces recorded in high electrical magnetic noise environments (e.g, near power lines) may produce significant electronic noise at 60 Hz (or harmonics of 60 Hz) decreasing the SNR of the recorded DST traces. In this case the investigator can apply notch filters to remove the electronic noise. In other cases simply narrowing the bandpass filter may results in higher SNR traces. For example, the near surface (2m depth) trace illustrated in Fig. 6(A) has a minimally applied low pass filter of 300Hz and corresponding SSP value of 0.52. Figure 6(B) shows the trace illustrated in Fig. 6(A) with a bandpass filter of 20Hz to 130Hz. In this case the SSP value has been increased to 0.7.

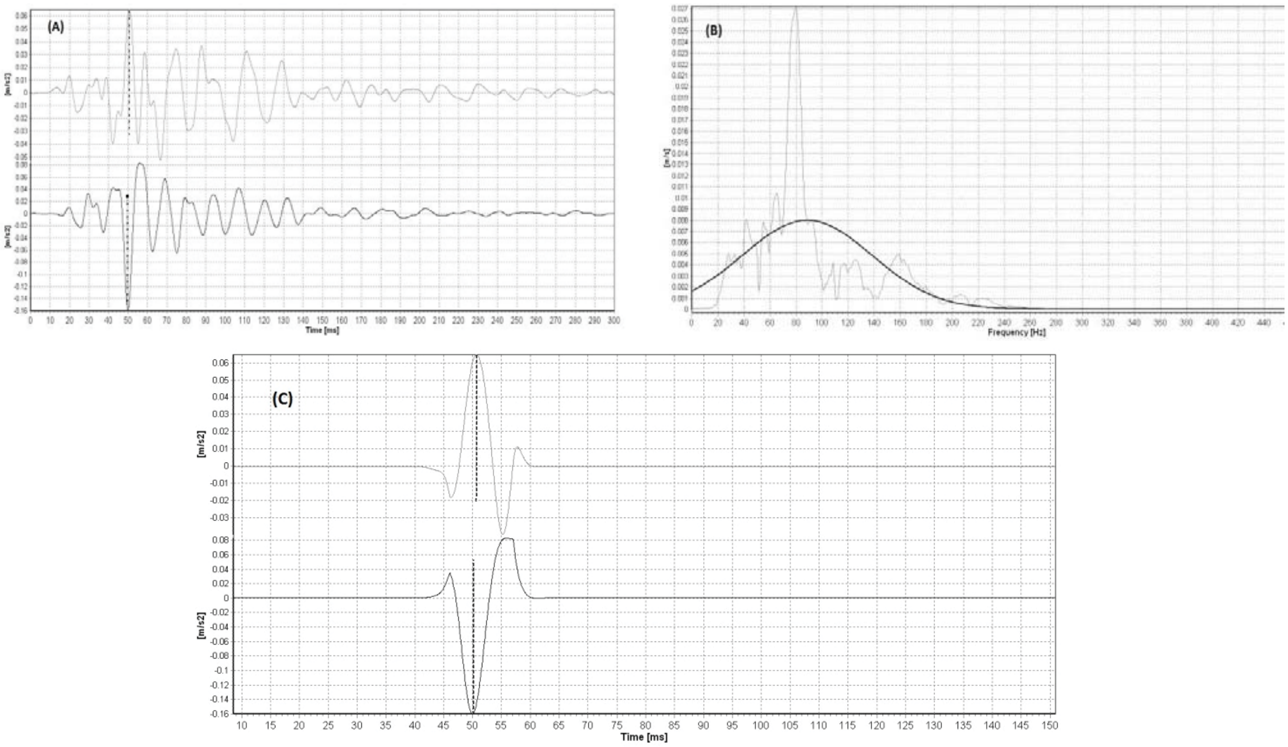

Figure 5: (A) DST seismic trace recorded with evidence of strong measurement noise "ringing". (B) After application of signal decay where first peak and trough are isolated. (B) Frequency spectrum for the Y axis response illustrated in (A) (light grey trace) and the best fit "bell-curve" (black bold trace). (C) Isolation of the dominant peak and trough by applying time windowing.

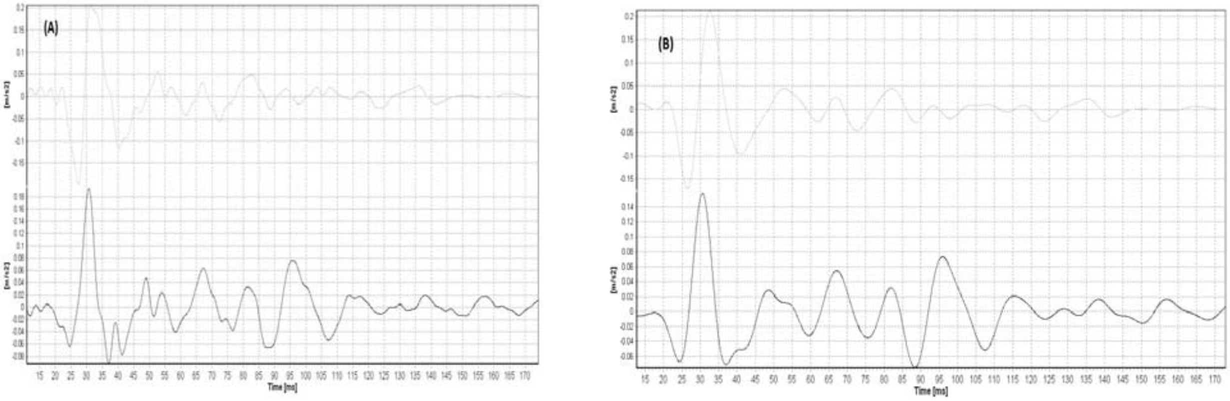

Figure 6: (A) Near surface DST trace with a 300 Hz low pass filter applied.
(B) Trace illustrated in (A) with a bandpass filter 20Hz to 130Hz applied

3.3 Low linearity and low signal shape parameter values

There are DST testing conditions which can result in strong source wave reflections (Baziw and Verbeek, 2014c), which can lead to source wave speak skewing and subsequently lowering of the both the linearity and the SSP values. To address this case, time windowing can be applied to the effected traces so that unaffected first peaks and troughs can be isolated. This process is outlined in Fig. 7.

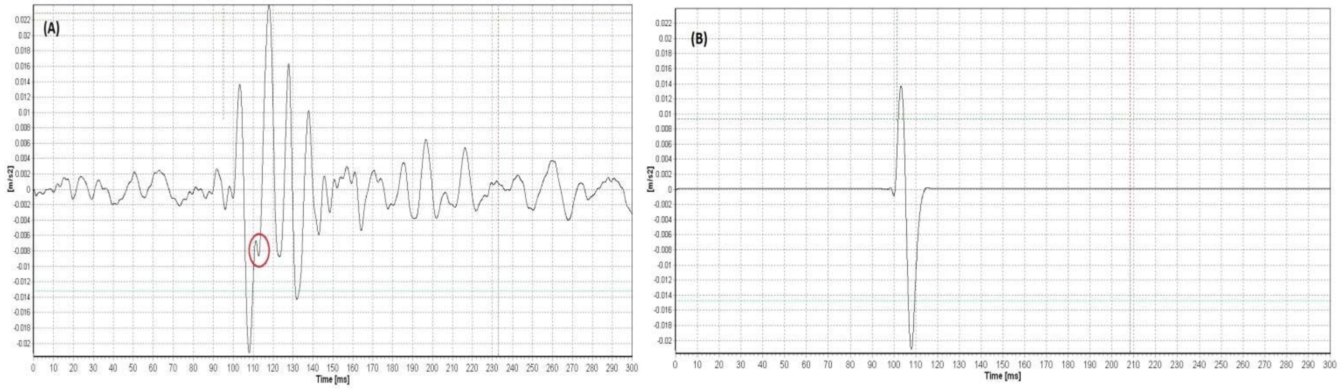

Figure 7: (A) DST seismic trace recorded with evidence of strong source wave reflections at 112ms. (B) After application of signal decay where first peak and trough are isolated.

CONCLUSIONS

Downhole Seismic Testing (DST) is an important geotechnical testing technique for site characterization and therefore it is essential that the quality of the acquired seismic traces can be assessed. The Seismic Trace Characterization (STC) provides the user with a good indication, but the individual parameters used to derive the STC (the cross-correlation coefficient of the full waveforms between successive depths of data acquisition, the linearity estimates from the polarization analysis, and the Signal Shape Parameter that quantifies the deviation of the source wave frequency spectrum from a desirable bell-shaped curve overview) also provide guidance on the most appropriate data analysis technique to ensure that the analysis results are as accurate as possible. More extensive use of the STC will provide further validation of the equation used to derive the STC and also an indication whether additional parameters need to be taken into account.

REFERENCES

ASTM (American Standards and Testing Methods). (2014). “D7400: Standard Test Methods for Downhole Seismic Testing.” ASTM Vol. 4.09 Soil and Rock (II): D5877-latest.

Baziw, E. and Verbeek, G., (2016a), "Interval Velocity Classification Technique in Downhole Seismic Testing", http://www.bcengineers.com/images/BCE_IVC_2016.pdf

Baziw, E., and Verbeek, G. (2016b), “Frequency spectrum “bell-curve” fitting as a component of SCPT interval velocity accuracy assessment.” In Proceedings of the 5th International Conference on Geotechnical Site Characterization (ISC-5), Queensland – Australia, 5-9 Sept. Australian Geomechanics Society, 1431-1436.

Baziw, E., and Verbeek, G. (2012). “Deriving Interval Velocities from Downhole Seismic Data.” Geotechnical and Geophysical Site Characterization 4 – Mayne (eds), CRC Press, 1019–1024.

Baziw, E. (1993). “Digital Filtering Techniques for Interpreting Seismic Cone Data.” Journal of Geotechnical Engineering, ASCE, Vol. 119, No. 6, pp. 98-1018.

Baziw, E., and Verbeek, G. (2014a). “Signal Processing Challenges when Processing DST and CST Seismic Data containing TIRs.” ASTM International - Geotechnical Testing Journal (GTJ), vol. 37, no. 3, 1-21.

Baziw, E. and Verbeek, G. (2014b). “Identifying Critical Layers using SCPT and Seismic Source Moveout.” In the Proceedings of the 3rd International Symposium on Cone Penetration Testing, CPT'14, May 12-14, 2014 - Las Vegas, Nevada, 357-364.

Baziw, E. and Verbeek, G (2014c), “Methodology for Processing Seismograms Containing Total Internal Reflections”, IEEE Transactions on Geosci. Remote Sensing (TGRS), vol. 52, Issue 11, 7073-7085.

Baziw, E. (2011). “Incorporation of Iterative Forward Modeling into the Principle Phase Decomposition Algorithm for Accurate Source Wave and Reflection Series Estimation”. IEEE Transactions on Geosci. Remote Sensing (TGRS), vol. 49, No. 2, 650-660.

Baziw, E., Nedilko, B., and Weir Jones, I. (2004). “Microseismic Event Detection Kalman Filter: Derivation of the Noise Covariance Matrix and Automated First Break Determination for Accurate Source Location Estimation.”, Pure appl. geophys. vol. 161, no. 2, 303-329.

Baziw, E. (2002). “Derivation of Seismic Cone Interval Velocities Utilizing Forward Modeling and the Downhill Simplex Method.” Can. Geotech. J., Vol. 39, 1181-1192.

Baziw, E., Tichy, J, and de Caprona, G (2000),"Data Acquisition in Seismic Cone Penetration Testing". In Proceedings of the 3rd International Symposium on Integrated Technical Approaches to Site Characterization (ITASCE), Argonne, IL, 11 14 Sept. 2000. Argonne National Laboratory. 69-72.

Campanella, R.G., Robertson, F.T.C., & Gillespie, D. (1986). “Seismic Cone Penetration Test.” In Proceedings of INSITU86. American Society of Civil Engineers (ASCE) Geotechnical Special Publication. No. 6, 116–130.

Gelb, A. (1974). Applied Optimal Estimation (4th ed.). MIT Press, Cambridge, Mass, USA.

Gibowicz, S.J. and Kijko, A. (1994). An Introduction to Mining Seismology. Academic Press,
San Diego, USA.

Ishihara, K. (1982). “Evaluation of Soil Properties for use in Earthquake Response Analysis.” International Symposium on Numerical Models in Geomechanics, Zurich, 237-259.

Kanasewich, E.R. 1981. Time Sequence Analysis in Geophysics (3rd ed.). The University of Alberta Press, Edmonton, AB. Canada.

Shearer, P.M. (1999). Introduction to Seismology. Cambridge University Press, Cambridge, U.K.