A novel algorithm for identifying arrival times of P and S Waves in seismic borehole surveys

P. Anbazhagan [*], Sauvik Halder

Department of Civil Engineering, Indian Institute of Science, Bangalore, 560012, Karnataka, India

ARTICLE INFO

ABSTRACT

The arrival times of P and S waves, originating from earthquakes, diverse seismic tests, and events, are crucial geotechnical parameters. Derived from the inversion of these travel times, V_P (P-wave velocity) and V_S (S-wave velocity) are pivotal in geotechnical engineering, correlating directly with dynamic soil properties and enabling calculations of Poisson's Ratio (ν), Young's modulus (E), Shear modulus (μ), and Bulk modulus (B). Both V_P and V_S are crucial for evaluating soil behaviour under various conditions, aiding in modelling soil for settlement, wave propagation, seismic wave interaction, liquefaction potential analysis, seismic response analysis, and many more. The selection of arrival times for seismic tests, including Crosshole, Downhole, and Uphole tests, is done manually, which is time-consuming and potentially erroneous. To address this issue, various algorithms have been developed to automate the picking process. Some of these algorithms use wavelet transforms and Bayesian information criteria, while others use machine learning techniques such as artificial neural networks. These methods vary in terms of their accuracy, yet each one possesses inherent limitations when it comes to processing data with different levels of signal-to-noise ratio. The advancement of automated algorithms for determining arrival times is an ongoing and dynamic field of research. Apart from the existing research focused on determining the arrival time of P waves, there is a dearth of studies investigating the detection of S wave arrival times. To fill this gap, this study proposes new approaches for detecting both P and S wave arrival time(s). One approach entails the utilization of an iterative optimization algorithm to accurately fit a curve to the leading edge of the P waveform. The arrival time is determined by calculating a fraction relative to the highest point obtained from the fitted peak. The second approach entails identifying the exact moment of the S wave's arrival by determining the points of intersection between the oppositely polarized S waveforms. These methods provide a promising approach for automatically detecting both P and S wave arrival time(s), which has the potential to improve the precision and efficiency in picking up arrival time(s).

1. Introduction

The measurement of compressional/primary wave velocities (V_P) and shear/secondary wave velocities (V_S) through seismic wave tests such as crosshole, downhole, and uphole tests are crucial in assessing the dynamic properties of soil. These measurements enable the calculation of key soil parameters such as Poisson's Ratio (ν), Young's modulus (E), Shear modulus (μ), and Bulk modulus (B). The comprehensive evaluation of VP & VS supports various geotechnical analyses, including stability for static and dynamic loads, liquefaction potential, seismic response, and many more (Kramer, 1996; Hussien and Karray, 2016). So, accurately detecting the arrival time of recorded P and S waves holds significant importance in seismic data processing, as it profoundly

influences the precision of the computed V_P and V_S profile. Manual picking of the arrival times is time-consuming as well as subjective. The accuracy also depends on the intensity of zoom and the limitations of the software being used. Thus, automating the entire procedure will increase the precision, save time and, more importantly, eliminate bias.

Numerous techniques have been introduced in academic research over the years, employing diverse mathematical models, including but not limited to surpassing threshold levels, utilizing deep learning techniques, and so on. Zhang et al. (2003) introduced an algorithm capable of automatically identifying the P-wave arrival by utilizing wavelet transform and the Akaike Information Criteria (AIC) picker. The AIC algorithm fails when the Signal to Noise Ratio (SNR) is extremely low. The AIC algorithm had a picking accuracy of 81% within 0.2 s and 93%

---

within 0.1 s of the corresponding analyst pick for the Dead Sea and Parkfield datasets, respectively, when applied to earthquake data. However, these results show a significant error compared to more advanced algorithms that have been developed since the publication of this research. It was mentioned in the study that the accuracy of the analyst picks themselves may also be subject to error. Rawles and Thurber (2015) proposed an innovative technique for detecting P- and S-wave arrivals automatically. Their approach, based on Nearest Neighbours, differs from traditional methods that rely on estimating parameters from the data. Instead, this method leverages the data directly to establish the model. Erol Kalkan (2016) proposed the P-PHASE PICKER algorithm, which transforms the seismic signals into the response domain of a single-degree-of-freedom (SDOF) oscillator with viscous damping. This transformation tracks dissipated damping energy, allowing for precisely identifying P-wave arrival times by analysing significant changes in the damping energy function. The P-PHASE PICKER has proven highly effective, reducing missed picks caused by noise from 15.1% (AIC picker) and 7.2% (STA/LTA picker) to just 0.9%. Erol Kalkan (2024) employed a similar technique for the S-PHASE PICKER, transforming seismic signals into the response domain of a single-degree-of-freedom (SDOF) oscillator with a high damping ratio (60% of critical damping). This method tracks the dissipated damping energy, which stays near zero before the P-wave and surges rapidly with the S-wave's arrival. This rapid build-up is then used to detect the S-wave onset. Saad et al. (2017) have proposed an approach to detect the arrival time of earthquakes based on the Fuzzy Possibilistic C-Means (FPCM) clustering algorithm. The FPCM algorithm is more accurate than other algorithms like Short-Term-Average (STA)/Long-Term-Average (LTA) Ratio, Akaike Information Criterion (AIC), and Fuzzy C-Means Clustering (FCM), but its accuracy is restricted to when the SNR is greater than −10dB. Saad et al. (2018a) introduced a method for earthquake onset time detection based on the Modified Laplacian of Gaussian (MLoG) filter. The MLoG filter incorporates a denoising-filter algorithm to smoothen the background noise and utilizes a Dual Threshold Comparator to accurately detect the onset time of the seismic event. The MLoG algorithm was compared with other algorithms (like STA/LTA and AIC), but its accuracy, though it increased, is still restricted to an SNR of −12dB. In contrast to conventional methods, more advanced techniques leverage deep learning and machine learning techniques. Saad et al. (2018b) proposed a Stacked Denoising Auto Encoder (SDAE), a deep learning technique, which acts as a denoising filter for the seismic data, thus smoothing the background noise. A threshold is used to detect the onset time of the event. It can detect arrival times for events when SNR is greater than −14dB. Hara et al. (2019) developed a machine-learning model that utilizes Convolutional Neural Networks (CNNs) to accurately determine the arrival time of P-waves in observed seismic waveforms. The inputs for the model were determined by human experts. Similarly, Wang et al. (2019) used a deep-learning method called PickNet to automatically determine the P and S wave arrival time(s) of local earthquakes. Both the CNN model and PickNet were trained by humans, which might introduce bias due to individual differences. Moreover, achieving a high level of accuracy necessitates the utilization of a substantial volume of training data. Zhang and Sheng (2020) proposed using the Residual Link Nested U-Net Network (RLU-Net) and an improved Wasserstein Generative Adversarial Network (WGAN) to pick the first arrival of microseismic signals. Saad et al. (2021) presented SCALODEEP, a deep learning framework for real-time earthquake detection that utilizes a sophisticated network architecture and a scalogram, a time-frequency representation of earthquake data, to extract high-order features from three-component seismograms. The model demonstrates superior performance, achieving high accuracy across diverse datasets: 88.5% (Japanese), 90.7% (Texas), 87.8% (Egyptian), and 79.6% (Arkansas). SCALODEEP outperforms traditional methods (STA/LTA, FAST, and template matching) and advanced frameworks like CRED and Earthquake transformers.

Picking of the arrival time is used in other disciplines of engineering and science as well. Zhang et al. (2020) proposed a method based on the wavelet transform (WT) and Bayesian information criteria (BIC) to determine the arrival time of the ultrasonic echo signal, which is used for measuring the ultrasonic wind speed. Sedlak et al. (2008) proposed a technique to automatically determine the first arrival of acoustic emission (AE) signals in thin metal plates, which is based on the Akaike Information Criterion (AIC), that uses the specific characteristic function.

The arrival time of seismic waves plays a crucial role in geotechnical engineering to characterize the properties of soil, as it is used to calculate wave velocities, which in turn provides the values of Poisson ratio and Modulus at specific levels of strain. This study aims to establish a mathematical model to accurately detect the arrival times of Primary (P) and Secondary (S) waves in borehole seismic surveys, which are currently predominantly determined through manual methods. The proposed model includes two methods: one for detecting P-wave arrival times using Iterative Curve Fitting to fit multiple Gaussian curves to complex peak shapes, with the start point determined by the best-fit model, and another for detecting S-wave arrival times by identifying the intersection point of positively and negatively polarized S waves which demarcates the 180° phase difference between the waves. The goal of this study is to create a more reliable and consistent method for detecting wave arrival times for seismic borehole tests, eliminating any potential human bias in the process.

2. The proposed algorithms

P waves, also known as primary waves or pressure waves, are characterized by a compressional motion. P wave arrival time is determined by considering the starting of the first peak or valley. S waves, also known as secondary or shear waves, are characterized by a shearing motion - a side-to-side movement. They are slower than P waves, and their arrival time is determined by several factors, including a significant increase in amplitude, a decrease in frequency, and most importantly, a 180° phase difference between oppositely polarized waves.

Two methods have been suggested to determine the arrival times of Primary and Secondary waves, and their details are presented in the subsequent sections. The algorithms for these techniques have been implemented in MATLAB, which is a widely adopted commercial programming language extensively employed in the fields of engineering and science.

2.1. Primary (P) Wave

Selecting the arrival time, specifically the onset of the initial peak, for P-Waves can be subjective since the shapes of the peaks tend to gradually approach the baseline without intersecting it. This can be solved by using a simple threshold detection to determine the first arrival of the generated waveforms, i.e., the arrival time will be a value in the time domain when the amplitude reaches a small fraction of the peak amplitude. One drawback of this approach is that the random noise present in the baseline can often be a significant portion of the amplitude at that specific point. As a result, this compromises the accuracy of the measurement and makes it useless. Various strategies can be adopted to overcome this, such as Smoothing or Fitting each peak to a model shape, but in our case, these are not very effective. Smoothing can reduce the noise, but it will distort and broaden the peaks (especially with overlapping peaks), effectively changing their start and endpoints. Fitting each peak to a specific mathematical model, such as a Gaussian or Lorentzian shape, can effectively reduce noise in the data. However, this approach requires having an analytical peak model that accurately represents the data, which is often not feasible. An alternative and more robust method is to iteratively fit a model to the peak data. In the case of complex peak shapes, the model can be constructed by combining multiple simpler shapes, such as Gaussians. The start point can be determined with surprising precision because it is calculated based on
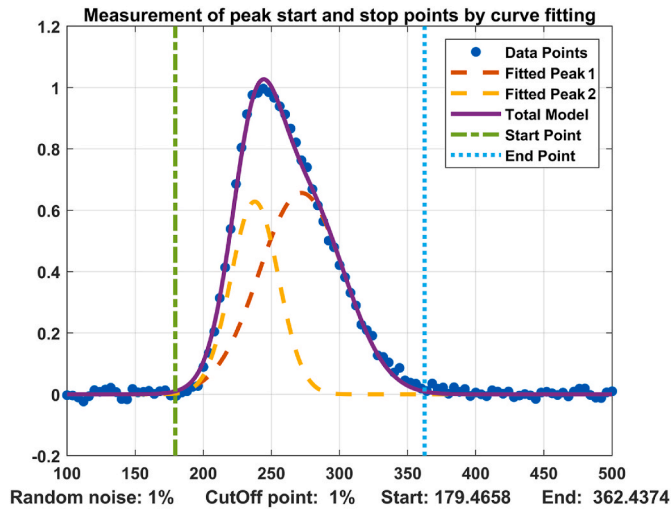
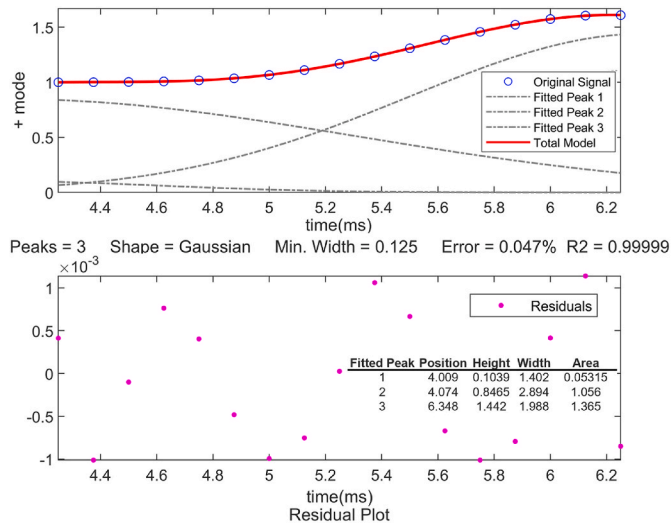**Fig. 1.** Measurement of peak start and end points by curve fitting (Modified from O'Haver).

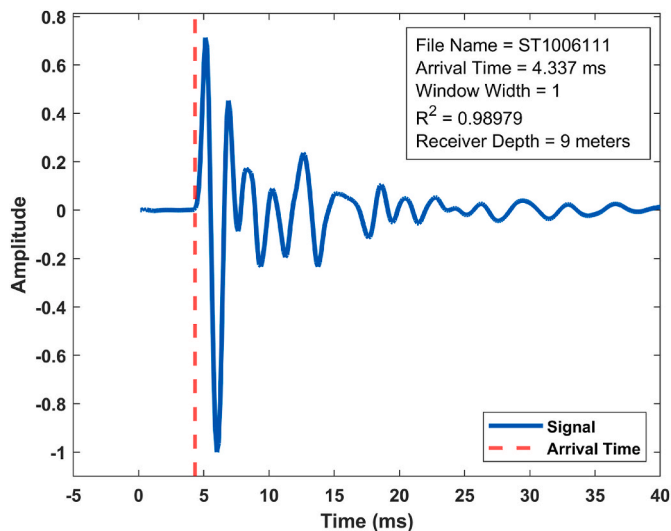**Fig. 2.** Fitting of the first peak of the waveform.

Fig. 3. Calculated P-Wave arrival time.

the best-fit model, which averages out the noise over the entire signal. This iterative curve fitting approach provides a more reliable way to analyze complex peaks and minimize the impact of noise. Here, the cut-off amplitude (or the threshold) is considered as 1% of the peak height. In a nutshell, we propose the use of the Iterative Curve Fitting technique to fit two or more Gaussian curves (where the complex peak shape is the sum of those curves).

2.1.1. Iterative curve fitting of Gaussian curves

Gaussian functions can describe many scientific and mathematical processes. The fitting of Gaussian functions has been used widely in signal processing. A Gaussian function (or simply Gaussian) can be expressed as.

$$f(x) = Ae^{\frac{-(x-\mu)^2}{2\sigma^2}}$$ where A, μ, and σ ($\sigma \neq 0$) are arbitrary real constants.

The graph of a Gaussian function turns out to be 'bell shaped' where the parameter 'A' represents the height of the peak, the parameter 'μ' represents the center of the peak, and the parameter 'σ' controls the width of the curve.

A straightforward and efficient approach to fitting a curve involves utilizing the method of linear least squares, wherein the dependent variable can be represented by a polynomial with linear coefficients. Caruana et al. (1986) proposed a simpler method (known as Caruana's Algorithm) of coordinate transformation for the Gaussian function by taking the natural logarithm on both sides.

$$\ln(y) = \ln(A) + \frac{-(x-\mu)^2}{2\sigma^2} \tag{1}$$

$$= \ln(A) - \frac{\mu^2}{2\sigma^2} + \frac{2\mu x}{2\sigma^2} - \frac{x^2}{2\sigma^2} \tag{2}$$

$$= a + bx + cx^2 \tag{3}$$

The non-linear equation involving unknowns A, μ and σ is converted into a linear equation involving unknowns a, b, and c. Caruana's algorithm is computationally efficient since it does not require iteration. However, its accuracy significantly decreases when noise is present. Additionally, the transformation of the non-linear function into log-log space alters the distribution of errors in the data.

The Iterative Curve Fitting Method employs a trial-and-error approach, wherein the model parameters are systematically adjusted until the equation closely matches the given data. This iterative process continues until the desired level of fit is achieved.

Firstly, a model is considered for the data (Gaussian for our case), and then a first guess is made for the non-linear parameters (position and width of overlapping peaks). The parameters are continuously adjusted until the desired fitting accuracy is attained or the maximum number of iterations is reached.

In our MATLAB implementation, users load signal data and select a region near the first peak using an interactive rectangular resizable region selector. The P-wave arrival time determination involves a 1% threshold of the maximum height of the total model. This threshold remains consistent irrespective of the selected region's size. Subsequently, the chosen peak undergoes precise iterative fitting with three Gaussian curves, enabling an accurate determination of the arrival time. The iterative fitting process employs a non-linear optimization algorithm to decompose complex, overlapping-peak signals into their component parts. Gaussian model parameters are iteratively adjusted to minimize disparities between the model's predictions and observed experimental data, ensuring convergence. This iterative approach ensures adaptability to data intricacies, such as noise or complex peak shapes, accurately representing the underlying trend.

Fig. 1 illustrates the measurement of peak start and stop points using the iterative curve fitting process with two overlapping Gaussian functions, which are shown as dashed lines in the plot. The cut-off point is
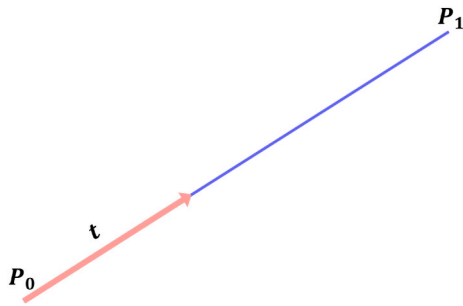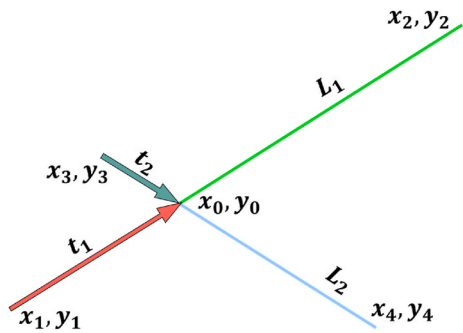
Fig. 4. Linear Bézier curve, t ∈ [0,1].

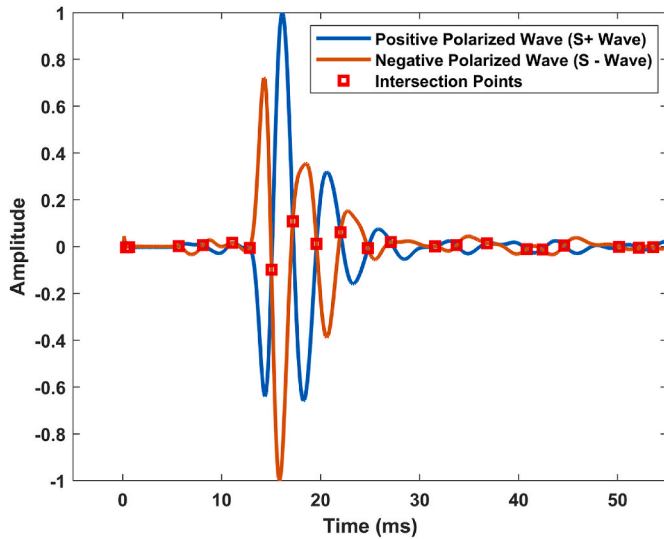**Fig. 5.** Representation of L$_1$ and L$_2$ with all the points marked.

Fig. 6. Marked intersection points over the entire waveform.

defined as 1% of the peak height, and random white noise has been added to the data set. The dash-dotted line and the dotted line represent the start and end points, respectively. The plot in Fig. 2 shows the fitted peak of a sample wave, and the corresponding residual plot demonstrates that the model is a good fit to the data. This is further reinforced by the strong R^2 value of 0.99999, suggesting that the model accounts for a significant amount of the variability observed in the data. The randomness of the residuals suggests that the model is capturing the underlying trend in the data accurately. Fig. 3 illustrates the marked arrival time calculated from the algorithm for a sample wave.

2.2. Secondary (S) Wave

Conventionally, shear wave arrivals are estimated by overlapping

two seismic traces of alternating strikes and identifying the beginning of a 'bow', which is a result of the 180° phase difference between two shots. As mentioned by Mok et al. (2016), the onset of the butterfly (or the bow) pattern marks the arrival time of the shear wave signals generated by reversing impact. The points where intersections occur distinctly depict the butterfly or bow pattern. We are considering the intersection point (may or may not be the first one) of the positively and negatively polarised S waves as the arrival time. As discussed earlier, manually selecting those points might give varying results based on the user and the intensity of zoom.

2.2.1. Intersection points

A straight line with two distinct points P_0 and P_1 and can be represented as $B(t) = P_0 + t(P_0 - P_1) = (1 - t)P_0 + tP_1$, where t is the linear Bezier parameter, which represents the distance from the starting point with $0 \leq t \leq 1$ and this form is known as the Linear Bezier Curve. This form is equivalent to Linear Interpolation. Fig. 4 depicts a linear Bezier curve with the indicated parameters.

If two-line segments L_1 and L_2 are given with endpoints: x_1, y_1 & x_2, y_2 for L_1, and x_3, y_3 & x_4, y_4 for L_2, as shown in Fig. 5.

Let's consider the point of intersection between two-line segments, L_1 and L_2, as (x_0, y_0). The parameter t_1 represents the distance from the starting point of L_1 to the intersection point, relative to the length of L_1. Similarly, the parameter t_2 represents the distance from the starting point of L_2 to the intersection point, relative to the length of L_2.

Thus, we can write four equations as

$$x_0 = x_1 + t_1(x_2 - x_1) \tag{4}$$

$$x_0 = x_3 + t_2(x_4 - x_3) \tag{5}$$

$$y_0 = y_1 + t_1(y_2 - y_1) \tag{6}$$

$$y_0 = y_3 + t_2(y_4 - y_3) \tag{7}$$

When rearranged and expressed in matrix form, the given expression can be rewritten as follows:

$$\begin{bmatrix} x_2 - x_1 & 0 & -1 & 0 \\ 0 & x_4 - x_3 & -1 & 0 \\ y_2 - y_1 & 0 & 0 & -1 \\ 0 & y_4 - y_3 & 0 & -1 \end{bmatrix} \times \begin{bmatrix} t_1 \\ t_2 \\ x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} -x_1 \\ -x_3 \\ -y_1 \\ -y_3 \end{bmatrix} \tag{8}$$

The given system of linear equations can be represented as $A \times T = B$. Therefore, to solve for the variable T, we can express it as $T = A^{-1}B$,

where $T = \begin{bmatrix} t_1 \\ t_2 \\ x_0 \\ y_0 \end{bmatrix}$, $A = \begin{bmatrix} x_2 - x_1 & 0 & -1 & 0 \\ 0 & x_4 - x_3 & -1 & 0 \\ y_2 - y_1 & 0 & 0 & -1 \\ 0 & y_4 - y_3 & 0 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} -x_1 \\ -x_3 \\ -y_1 \\ -y_3 \end{bmatrix}$.

The matrix T represents the parameters needed to describe the intersection point between two-line segments L_1 and L_2, denoted as (x_0, y_0), along with the parameters t_1 and t_2. Upon restricting $t_1, t_2 \in [0, 1]$, the intersection point lies within the line segments L_1 and L_2. By constructing matrices A and B from the endpoint coordinates of the two lines L_1 and L_2, and then solving the equation $T = A^{-1}B$, we can determine the parameters describing the intersection point between the two-line segments, i.e., x_0, y_0.

The MATLAB implementation for S-wave arrival time detection involves loading recorded signal data, including one positively and one negatively polarized wave. Given the discrete nature of the recorded wave(s), the algorithm interprets the relationship between consecutive data points as a linear Bézier curve. It systematically solves the system of linear equations (Equations (4)–(8)) for each pair of adjacent data points, deriving Bézier parameters that define the intersection point.

In constructing the matrices for solving the system of linear equations, matrix A comprises coefficients derived from the coordinates of the two adjacent points. Matrix T is a column vector containing the unknown Bézier parameters and the intersection coordinates, while
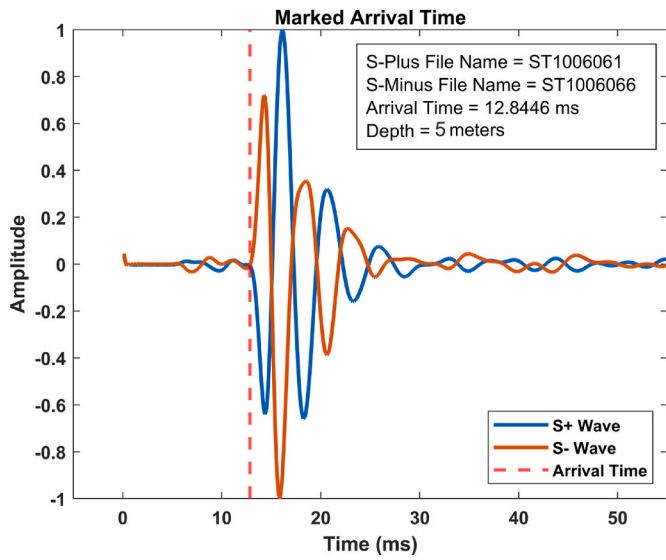
Fig. 7. Calculated S-Wave arrival time.

matrix B holds the constant terms on the right-hand side of the equations. The equation $A \times T = B$ concisely summarizes this relationship, and this approach is systematically applied to cover the entire signal.

A critical validity check is employed to ensure that these parameters fall within the range of 0–1, signifying a valid intersection. These intersection points represent instances of a 180-degree polarity difference. With the calculated intersection points between oppositely polarized waveforms, a predefined threshold is applied to identify the arrival time at the first bow, marked by a substantial increase in waveform amplitude. Alternatively, the user can also interactively pinpoint this juncture for arrival time selection, providing a more consistent and mathematically calculated alternative to manual

selection by zooming into the wave. This meticulous mathematical procedure robustly identifies intersection points, significantly enhancing the precision of S-wave arrival time determination.

Figs. 6 and 7 show the entire waveform with the marked intersections and the marked arrival time, respectively. The intersection points are highlighted along the entire waveform in Fig. 6.

A set of experimental data is analyzed using these methods, which includes estimated arrival time and determination of V_P and V_S values. These results are then compared with the conventional manual pick-up time and the estimated V_P and V_S values in the subsequent section.

3. Field survey

Crosshole, Downhole, and Uphole tests are geophysical techniques

Fig. 9. Field setup for crosshole seismic testing at IISc, Bangalore.
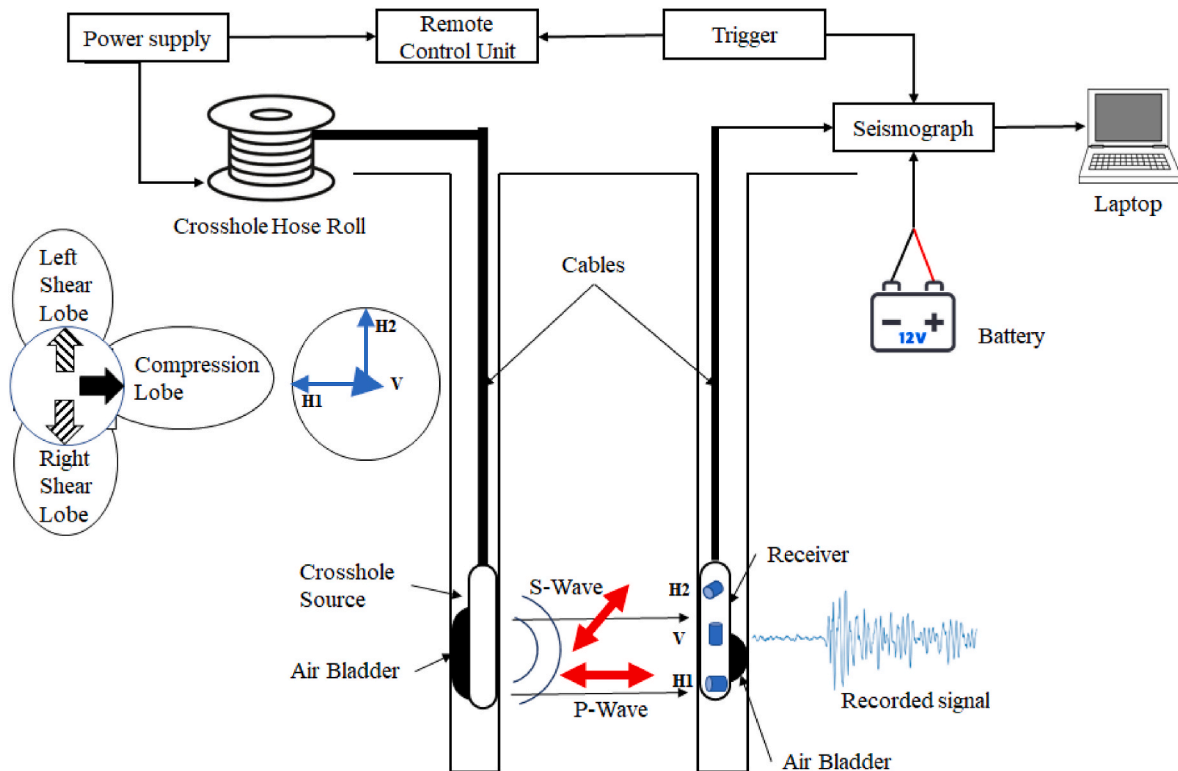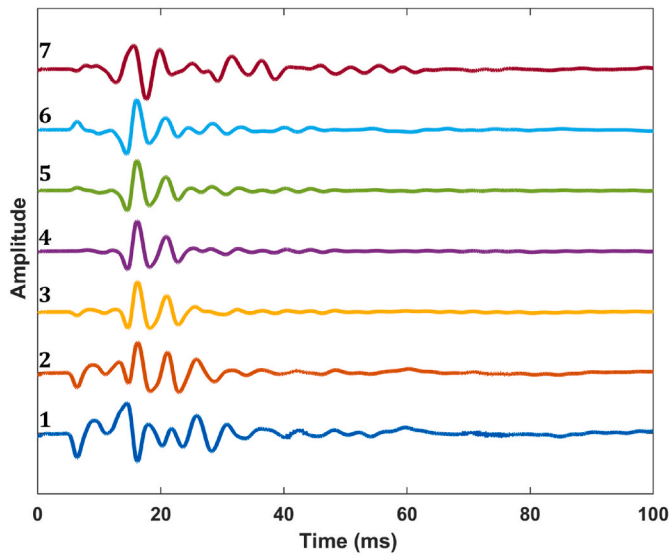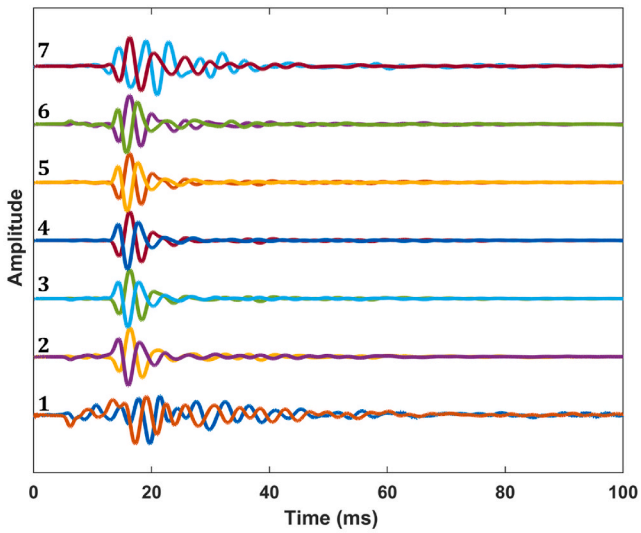

**Fig. 8.** Schematic illustration of crosshole seismic testing.

(a)

(b)

Fig. 10. Multi-Channel Representation of Crosshole Seismic Waves recorded by BGK-7 (a) P – Waves (b) S – Waves (both positively and negatively recorded waves).

used to measure the properties of subsurface materials. Crosshole tests involve sensors in two boreholes, downhole tests use sensors in a single borehole, and uphole tests use sensors on the surface. Vibrations are induced in the subsurface material for crosshole and uphole tests using impact inside the borehole, and for downhole tests using impact at the surface.

A crosshole test was conducted at the Department of Civil Engineering, Indian Institute of Science (IISc) in Bangalore, using a BIS-SH Crosshole source and a BGK-7 geophone, both manufactured by Geotomographie GmbH. The test reached a depth of 15 m with an interval of 1 m. Fig. 8 presents a visual schematic illustration of the experimental setup used in the testing process. The illustration showcases the components of the testing process, including the boreholes, geophones, and energy source, as well as the direction of wave propagation. Fig. 9 depicts the typical field arrangement and setup used for crosshole seismic testing conducted at IISc.
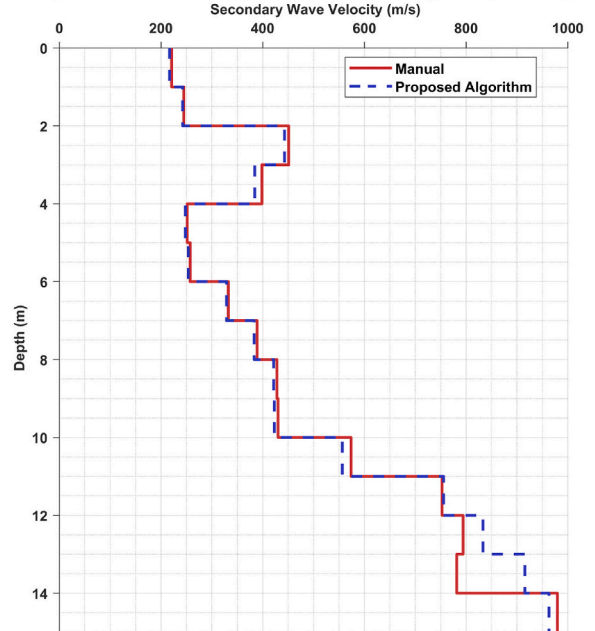
The BGK-7 geophone has seven active channels - six in the horizontal

(a)
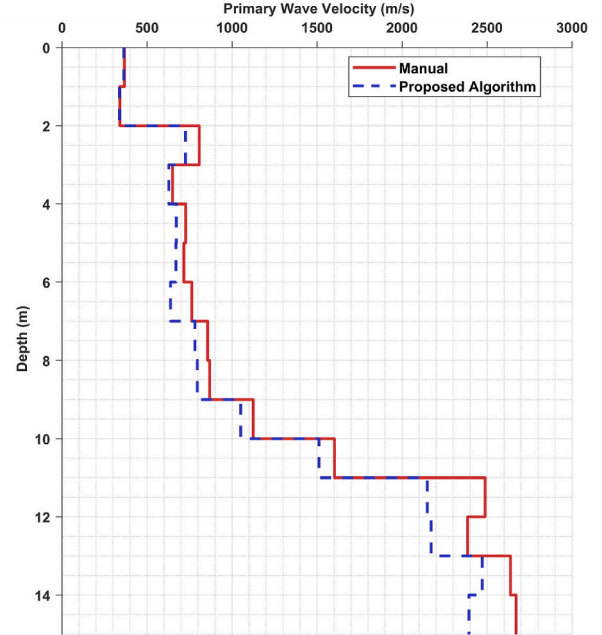
(b)

Fig. 11. Comparison of (a) V_P & (b) V_S profile.

direction and one in the vertical direction - which can record data simultaneously. The first channel of the geophone was aligned in the direction of the source borehole, and P-waves were recorded when hit in the forward direction. Two oppositely polarized hits were made in the transverse directions, and the resulting S-waves were recorded in the fourth channel.

Fig. 10 displays the visual representation of the recorded waves from the seven channels of the geophone during the crosshole seismic test, at a depth of 6 m. Based on the spatial distribution of the sensors, the
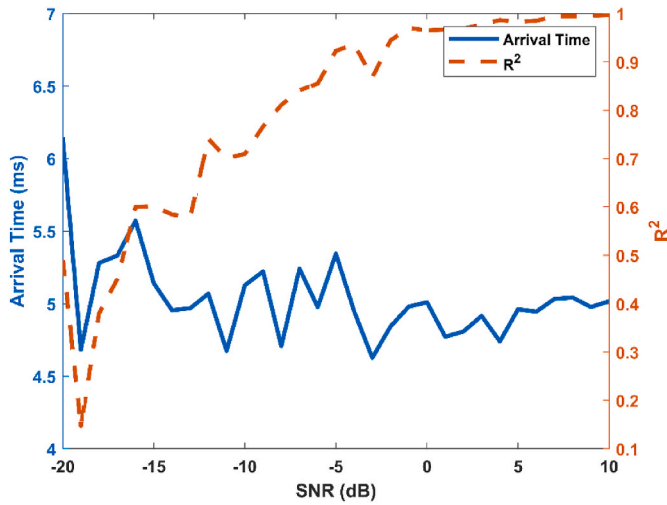
Fig. 12. Variation of arrival Time and R^2 with SNR

arrival times of P-waves were determined by utilizing data from the first channel, whereas the arrival times of S-waves were determined based on the waveform obtained from the fourth channel. This is due to the fact that these channels are exactly in the same direction as the direction of the hit made by the source, and thus, these channels are most likely to capture the wavefront of the wave generated by the energy source, with the least amount of interference from other waves or noise. The use of these channels, which were directly aligned with the direction of the wavefront, improved the accuracy of arrival time estimation. It should be noted that while arrival times varied between all channels, the channel in the same direction as the wavefront provided the most accurate results.

The proposed algorithms were employed to determine the arrival times of the waves. The horizontal distance between the boreholes on the surface is 3.05 m. To refine the accuracy of measurements, the corrected distance at each depth was calculated following the procedures outlined in ASTM D4428 (Standard Test Methods for Crosshole Seismic Testing), incorporating data obtained from the Borehole Deviation Survey. This correction ensures that the recorded distances align with the true subsurface distances, contributing to the precision of subsequent velocity calculations and the construction of reliable P-wave and S-wave velocity profiles. The velocities (V_P and V_S) are computed using the corrected distances and the corresponding arrival times at each depth using the relation *Velocity* (V) = *Corrected Distance* (L) / *Arrival Time* (t), where V represents either P-wave (V_P) or S-wave (V_S) velocity. These calculated velocities can help identify lithological variations, distinguish between different rock types, estimate the dynamic properties, and much more.

Subsequently, a comparison was made between the resulting P-wave and S-wave velocity profiles and those obtained through manual picking of the arrival time(s).

Fig. 11 illustrates the velocities calculated from both the conventional manual technique and the proposed algorithm. It can be observed that the velocities (P and S wave) estimated from the algorithm are very
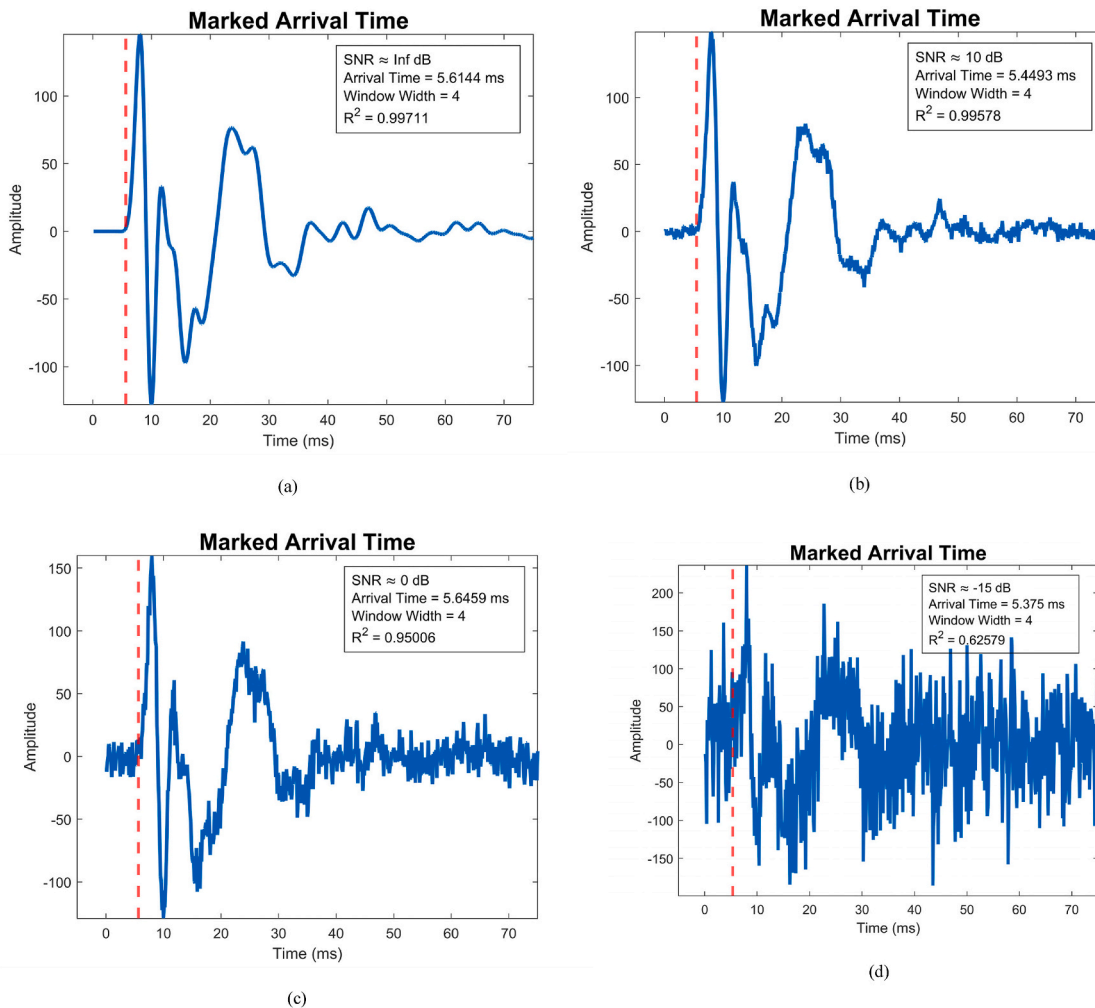

(a)

(b)

(c)

(d)

Fig. 13. Estimated arrival time for signals with varying SNRs (a) Original data (b) 10 dB (c) 0 dB (d) −15 dB.
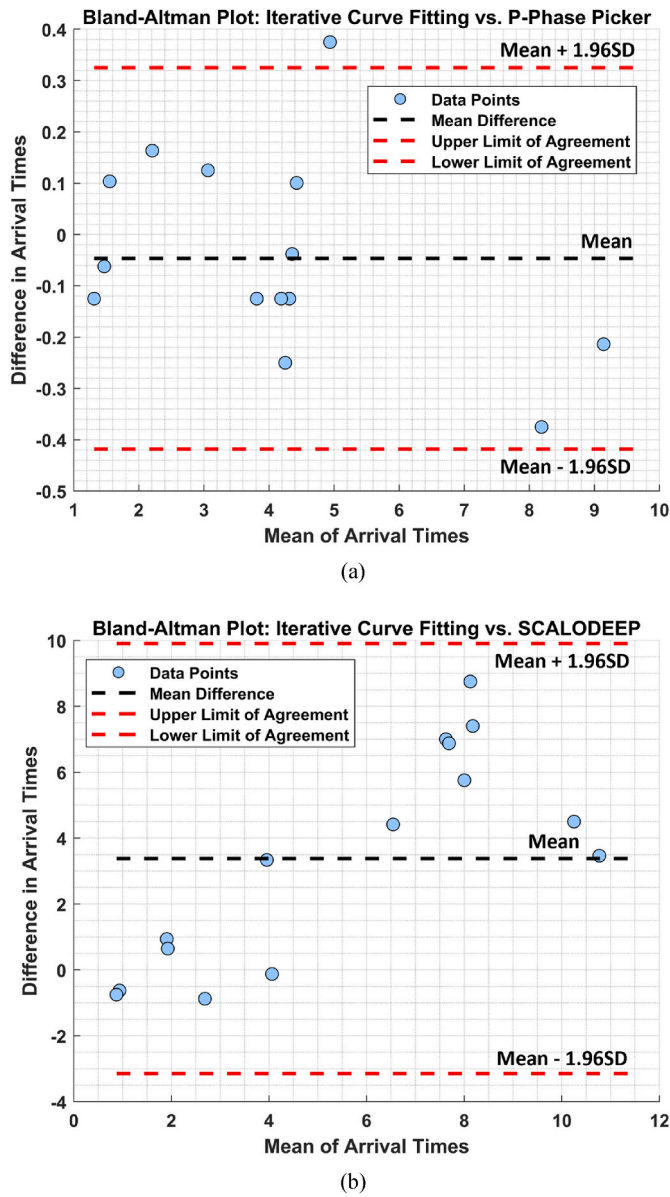
(a)

(b)

Fig. 14. Bland-Altman Plot Comparing P-Wave Arrival Time Algorithms (a) Iterative Curve Fitting vs P-Phase Picker (b) Iterative Curve Fitting vs SCALODEEP.

close to the ones made manually. But at a few depths, there are significant differences. This variability is due to the inherent inconsistency of manual picks, as well as the influence of human bias. It can be noted here that manual wave pick-up was done by experienced research students using these test data for more than a year. The effectiveness of the proposed approach in handling noisy data and accommodating users with varying levels of expertise in seismic borehole testing will be discussed in the following sections.

4. Evaluating the accuracy of the proposed algorithms

To assess the accuracy of the proposed algorithms for detecting the arrival time of P and S waves, the human bias in arrival time picks was assessed by comparing the predicted arrival time to a set of manual picks, and its reliability and performance were evaluated with decreasing signal-to-noise ratios.
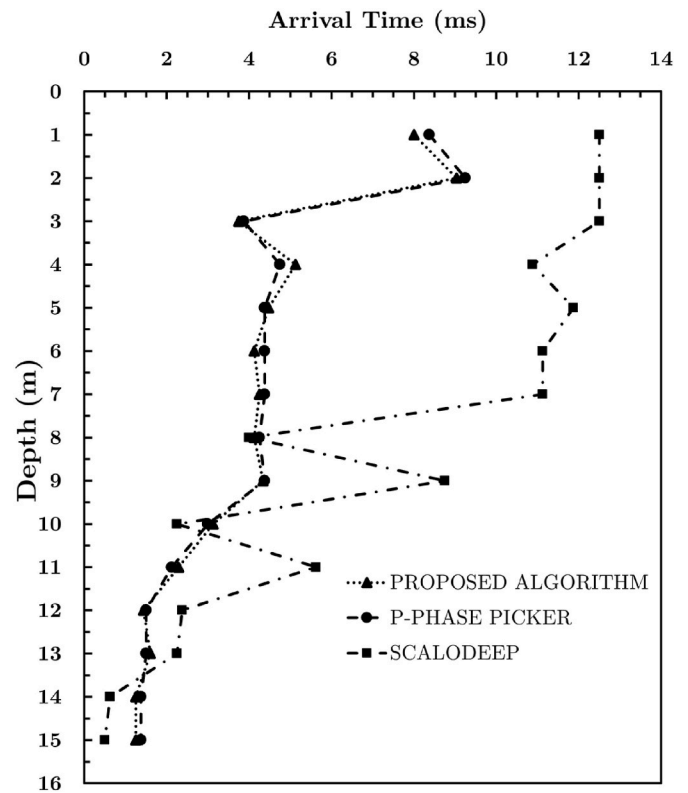

Fig. 15. Comparison of arrival times by the proposed algorithm, P-Phase Picker, and SCALODEEP.

4.1. Robustness of the proposed algorithms

The performance of the algorithms was evaluated by subjecting them to varying signal-to-noise ratios by applying white noise to recorded waveforms. White noise is a type of random noise that has a flat frequency spectrum. In other words, white noise contains equal amounts of energy at all frequencies within a given range.

4.1.1. Primary waves

A downhole waveform with a signal-to-noise ratio of -1.6214 dBc (decibels relative to the carrier) was selected and subjected to randomly generated white noise such that the SNRs range from 10 dB to -20 dB. The results show that the proposed algorithm was able to evaluate the arrival time with a fair level of precision, despite the decreasing R^2 value. The maximum difference in arrival time when compared to the original waveform was just 1.14 ms. It is possible to estimate the arrival time of a wave with high precision, even if the signal-to-noise ratio is low, as long as the shape of the first peak is clearly evident. Fig. 12 demonstrates the relationship between the SNR and the Arrival Time and R^2 values. The arrival time values remain relatively constant as the SNR decreases, while the R^2 values decrease as the SNR decreases.

Fig. 13(a)–(d) illustrates the marked arrival time for the original waveform and waveforms with added white noise at SNR levels of 10 dB, 0 dB, and -15dB. It can be observed that, even as the waveform becomes increasingly disrupted with decreasing SNR, the prediction of the arrival time remains accurate.

To further validate our proposed P-wave arrival time detection algorithm, we conducted a comprehensive comparative analysis with established methods, including the P-PHASE PICKER by Kalkan (2016) and SCALODEEP by Saad et al. (2021). Fig. 14 illustrates Bland-Altman plots used for this assessment, where the central line (mean) signifies
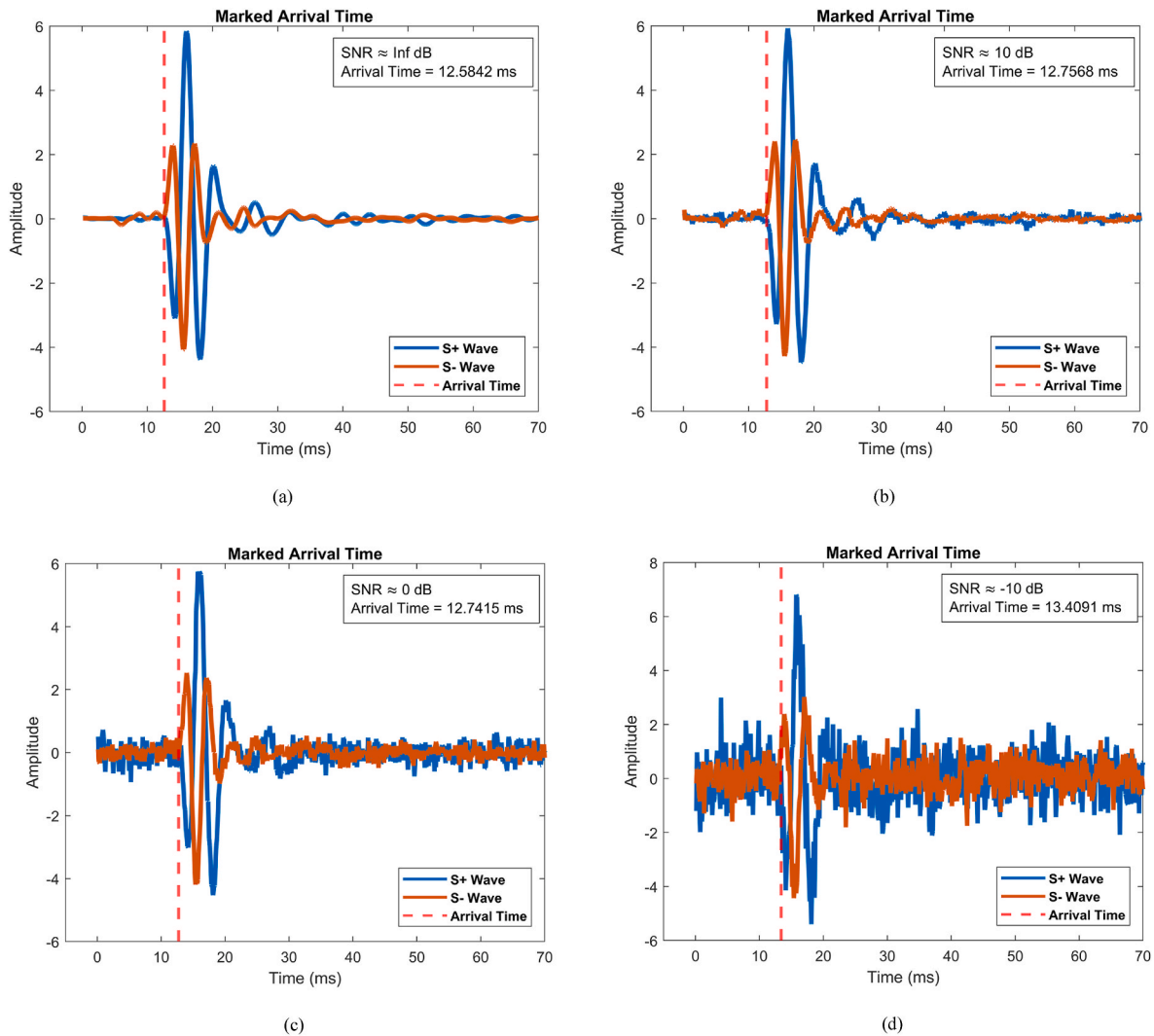
Fig. 16. Estimated arrival time for S wave signals with varying SNR (a)Original data (b)10 dB (c)0 dB (d) −10 dB.
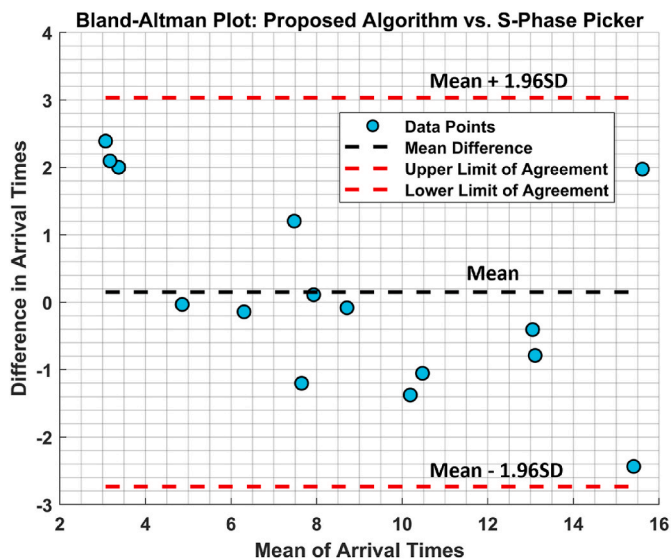

Fig. 17. Bland-Altman plot comparing S-wave arrival time: Proposed Algorithm vs. S-PHASE PICKER.

bias, and the outer lines (±1.96 SD, indicating a 95% confidence interval) represent the limits of agreement. This statistical approach provides a comprehensive perspective on the consistency and reliability of each algorithm compared to the others.

The P-PHASE PICKER transforms seismic signals into the response domain of a single-degree-of-freedom (SDOF) oscillator with a short natural period and high damping ratio, specifically designed to track dissipated damping energy. Notably, the damping energy function exhibits significant changes at the onset of the P-wave, enabling precise identification of the arrival time by analyzing this change. On the other hand, SCALODEEP is a deep learning framework that processes three-component seismograms using a scalogram and a deep architecture with skip connections to extract high-order features from the input data. It then classifies each time sample as a seismic signal or noise based on its probability value, allowing it to identify and distinguish earthquake signals from background noise.

The comparative assessment was applied to crosshole data introduced in Section 3, where arrival times calculated by the P-PHASE PICKER were compared with those obtained from our algorithm. The results, depicted in Fig. 14a, indicate a mean difference of −0.0464 when compared to the P-PHASE PICKER, a value remarkably close to zero. The Limits of Agreement (0.3253 and −0.4182) are relatively narrow, suggesting minimal systematic bias between our algorithm and the established P-PHASE PICKER. This emphasizes the consistency and reliability of our algorithm, showcasing its performance in agreement
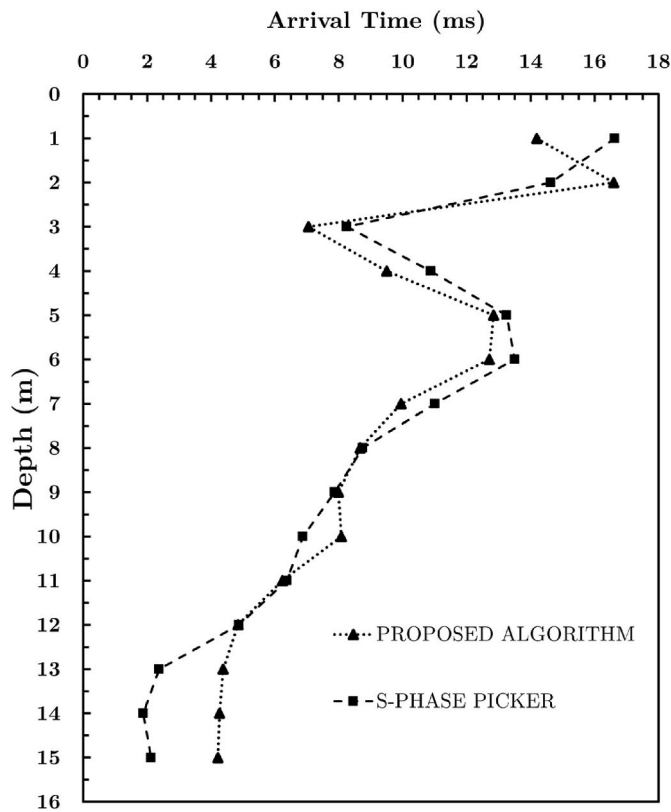
Fig. 18. Comparison of arrival times by the proposed algorithm and S-PHASE PICKER.

with the existing method.

The P-PHASE PICKER algorithm was then tested for noise robustness, repeating the same SNR test as mentioned at the start of this section. While it performed well from 10 dBc to −4 dBc, it failed from −5 dBc to −20 dBc. In contrast, our proposed algorithm exhibited superior noise robustness, successfully detecting arrival times even at −20 dBc. This underscores the effectiveness of our algorithm, surpassing the P-PHASE PICKER in challenging, low SNR conditions.

In the comparison with SCALODEEP, as illustrated in Fig. 14b, our algorithm exhibited a mean difference of 3.3801 and relatively wide limits of agreement, ranging from 9.9061 to −3.3801. Despite the higher

mean difference, the consistent differences suggest comparability, though not in strong agreement. This discrepancy may be influenced by the difference in training data characteristics, as SCALODEEP was trained with earthquake data at 100 Hz, while our algorithm used seismic borehole test data at 8000 Hz. Typically, data from seismic borehole tests are sampled at higher frequencies, ranging from 1 kHz to 20 kHz, to capture sharper peaks and finer details of these waves for accurate analysis. This underscores the need for further research tailored to signals generated during seismic borehole tests. Additionally, Fig. 15 visually compares arrival times calculated by the proposed algorithm, P-PHASE PICKER, and SCALODEEP across different depths, corroborating the findings. Notably, the proposed algorithm and P-PHASE PICKER are closely aligned, while SCALODEEP shows a consistent but distinct difference from the other two.

4.1.2. Secondary waves

A positively and negatively polarized waveform with signal-to-noise ratios of 5.1123 and 3.9609 dBc (decibels relative to the carrier), respectively, were selected and subjected to randomly generated white noise. The resulting SNR ranged from 10 dB to −15 dB. The proposed algorithm was able to accurately evaluate the arrival time, with a maximum difference of 0.8249 ms when compared to the original waveform. It is possible to estimate the arrival time of a wave with high precision, even in challenging situations where the signal-to-noise ratio is low, as long as the high amplitude peaks of the wave are distinctive and can be easily distinguished from the surrounding noise. As shown in Fig. 16(a)–(d), the marked arrival time is depicted for the original waveform and waveforms with added white noise at SNR levels of 10 dB, 0 dB, and −10dB. The prediction of the arrival time remains reliable, even with the increasing degradation of the waveform at lower SNR values.

It is important to highlight that interpreting P-wave arrivals shares similarities between seismic records generated by earthquakes and those obtained from borehole surveys. However, when it comes to S-waves, borehole surveys present a unique challenge: two oppositely polarized waves must be interpreted together, a departure from the approach used in earthquake records. A thorough literature review reveals a lack of automated algorithms tailored for seismic borehole records. While numerous automated algorithms exist for detecting S-wave onsets in earthquake records, such as the S-PHASE PICKER by Erol Kalkan (2024), none are specifically designed for borehole surveys. The S-PHASE PICKER, like the P-PHASE PICKER, converts seismic signals into the response domain of a Single Degree of Freedom (SDOF) oscillator with high damping and detects the S-wave arrival when the damping energy

Table 1
Arrival time picks made by the participants.

Data File	Arrival time from the proposed algorithm (in ms)					Arrival Time by manual picking (in ms)				
	P1	P2	P3	P4	P5	P1	P2	P3	P4	P5
P – Wave Data Record										
1006201	5.61	5.61	5.61	5.61	5.61	5.44	4.58	4.23	5.27	5.46
1006206	6.07	6.07	6.07	6.07	6.07	6.06	5.64	5.57	5.80	6.03
1006212	7.45	7.45	7.45	7.45	7.45	7.23	7.30	6.72	7.40	7.30
1006218	9.28	9.28	9.28	9.28	9.28	8.98	8.81	8.94	8.90	9.43
1006224	11.40	11.40	11.40	11.40	11.40	11.20	10.02	11.09	10.91	11.55
S – Wave Data Record										
1006202 & 04	11.02	11.02	11.02	11.02	11.02	11.00	10.98	11.17	11.29	10.98
1006208 & 10	13.06	13.06	13.06	13.06	13.06	13.01	12.91	13.19	12.88	12.96
1006214 & 16	17.15	17.15	17.15	17.15	17.15	15.74	17.02	16.88	15.78	17.21
1006220 & 22	19.07	19.07	19.07	19.07	19.07	18.96	18.83	18.91	10.22	19.06
1006226 & 28	22.64	22.64	22.64	22.64	22.64	22.39	22.57	22.83	22.65	22.45

Note: All the participants used the same computer for the survey, an MSI GF63 Thin Core i5 9th Gen with 8 GB RAM.
P1: An individual with no background in geophysics, but with a good understanding of waves and arrival times.
P2: A newcomer to the field of geophysics, with a good understanding of waves.
P3: An individual with no background in geophysics, but with a decent understanding of waves.
P4: An individual with no background in geophysics and no understanding of waves.
P5: An individual with 5 years of experience in geophysics and expert-level knowledge.

function changes significantly. A comparison between the proposed algorithm and the S-PHASE PICKER, similar to the P-wave comparison in section 4.1, shows general agreement but slight bias, due to the latter's non-alignment with seismic signals from borehole surveys. The Bland-Altman Plot in Fig. 17 illustrates this with a mean difference of 0.1502, suggesting a small bias, and moderately wide limits of agreement (-2.7328 to 3.0333), indicating variability between the two algorithms. The absence of extreme outliers suggests reliability and consistency within the anticipated range. This is further supported by the arrival times plotted across different depths in Fig. 18 for both the proposed algorithm and the S-PHASE PICKER. While both methods largely agree with a small average difference, the moderate variability highlights the S-PHASE PICKER's design limitations for borehole surveys, leading to some discrepancies. Our approach introduces a novel dimension to this field, addressing an aspect that has not been extensively explored in the current body of research.

4.2. The human bias

A survey was conducted to address the issue of human bias in manual picking. A group of 5 individuals with varying expertise in the subject matter was chosen and were given an instruction manual on how to use the software CrossOverPlot (used in manual picking) and the MATLAB code for the proposed algorithms. The instruction manual contained all the necessary details for processing the provided data, including the parameters that are taken into account when selecting the arrival times.

The survey included a set of 5 files containing P and S waves with varying signal-to-noise ratios (SNRs). The participants were asked to estimate the arrival times using both the manual technique and the proposed algorithms for all of the files. Table 1 displays the arrival times manually selected and proposed by the algorithm for each participant. The results showed that the manual picks made by the individuals varied depending on their expertise, while the algorithm consistently produced the same results. The standard deviation of the manual picks varied from 0.2 to 3.9, which may not seem significant but can create a substantial error when calculating velocities. In contrast, the standard deviation of the picks made by the algorithm by the same person is zero. This indicates that the use of the proposed algorithms can reduce human bias in manual picking, as the results are not influenced by the individual making the picks. By providing a consistent and objective method for selecting P and S waves, the algorithm can improve the accuracy and reliability of data analysis in this field.

5. Conclusion

In geotechnical analysis, the estimated time of arrival of P and S waves is a very important parameter because it allows for the calculation of seismic wave velocities, which is essential for seismic site characterization. In this study, two new approaches were proposed for automatically detecting the arrival times of P and S waves. These techniques employ iterative curve fitting and waveform intersection to identify the timing of P and S waves' arrivals precisely, respectively. These algorithms are promising because they are based on well-established mathematical concepts and have been shown to be accurate even with very low signal-to-noise ratios. Overall, they have the potential to improve the precision and efficiency of geotechnical seismic borehole testing and interpretations.

CRediT authorship contribution statement

P. Anbazhagan: Writing – review & editing, Supervision, Resources, Conceptualization. **Sauvik Halder:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

ASTM D4428/D4428M-14, Standard Test Methods for Crosshole Seismic Testing. https://doi.org/10.1520/D4428_D4428M-14.

Caruana, R., Searle, R., Heller, T., Shupack, S., 1986. Fast algorithm for the resolution of spectra. Anal. Chem. 58 (6), 1162–1167. https://doi.org/10.1021/356ac00297a041.

Hara, S., Fukahata, Y., Iio, Y., 2019. P-wave first-motion polarity determination of waveform data in western Japan using deep learning. Earth Planets Space 71, 127. https://doi.org/10.1186/s40623-019-1111-x.

Hussien, M.N., Karray, M., 2016. Shear wave velocity as a geotechnical parameter: an overview. Can. Geotech. J. 53 (2), 252–272. https://doi.org/10.1139/cgj-2014-0524.

Kalkan, E., 2016. An automatic P-phase arrival-time picker. Bull. Seismol. Soc. Am. 106 (3), 971–986. https://doi.org/10.1785/0120150111.

Kalkan, E., 2024. An Automated S-phase Arrival Time Picker with SNR Output. MATLAB Central File Exchange. Retrieved May 18, 2024, from. https://www.mathworks.com/matlabcentral/fileexchange/70343-an-automated-s-phase-arrival-time-picker-with-snr-output.

Kramer, S.L., 1996. Geotechnical Earthquake Engineering. Prentice Hall, Upper Saddle River.

Mok, Y.J., Park, C.S., Nam, B.H., 2016. A borehole seismic source and its application to measure in-situ seismic wave velocities of geo-materials. Soil Dynam. Earthq. Eng. 80, 127–137. https://doi.org/10.1016/j.soildyn.2015.10.011.

O'Haver, T. C. "Peak finding and Measurement". Retrieved from http://terpconnect.umd.edu/~toh/spectrum/PeakFindingandMeasurement.htm.

Rawles, C., Thurber, C., 2015. A non-parametric method for automatic determination of P-wave and S-wave arrival times: application to local micro earthquakes. Geophys. J. Int. 202 (2), 1164–1179. https://doi.org/10.1093/gji/ggv218.

Saad, O.M., Shalaby, A., Sayed, M.S., 2017. Automatic arrival time detection for earthquakes based on Fuzzy possibilistic C-means clustering algorithm. 8th International Conference on Recent Advances in Space Technologies (RAST). Istanbul, Turkey, pp. 143–148. https://doi.org/10.1109/RAST.2017.8003001.

Saad, O.M., Shalaby, A., Samy, L., Sayed, M.S., 2018a. Automatic arrival time detection for earthquakes based on Modified Laplacian of Gaussian filter. Comput. Geosci. 113, 43–53. https://doi.org/10.1016/j.cageo.2018.01.013. ISSN 0098-3004.

Saad, O.M., Inoue, K., Shalaby, A., Samy, L., Sayed, M.S., 2018b. Automatic arrival time detection for earthquakes based on stacked denoising autoencoder. Geosci. Rem. Sens. Lett. IEEE 15 (11), 1687–1691. https://doi.org/10.1109/LGRS.2018.2861218.

Saad, O.M., Huang, G., Chen, Y., Savvaidis, A., Fomel, S., Pham, N., Chen, Y., 2021. SCALODEEP: a highly generalized deep learning framework for real-time earthquake detection. J. Geophys. Res. Solid Earth 126 (4). https://doi.org/10.1029/2020JB021473 e2020JB021473.

Sedlak, P., Hirose, Y., Enoki, M., Sikula, J., 2008. Arrival time detection in thin multilayer plates on the basis of Akaike information criterion. J. Acoust. Emiss. 26. Retrieved from. https://www.ndt.net/?id=10893.

Wang, J., Xiao, Z., Liu, C., Zhao, D., Yao, Z., 2019. Deep learning for picking seismic arrival times. J. Geophys. Res. Solid Earth 124, 6612–6624. https://doi.org/10.1029/2019JB017536.

Zhang, J., Sheng, G., 2020. First arrival picking of microseismic signals based on nested U-Net and Wasserstein Generative Adversarial Network. J. Petrol. Sci. Eng. 195, 107527. https://doi.org/10.1016/j.petrol.2020.107527. ISSN 0920-4105.

Zhang, H., Thurber, C., Rowe, C., 2003. Automatic P-wave arrival detection and picking with multiscale wavelet analysis for single-component recordings. Bull. Seismol. Soc. Am. 93 (5), 1904–1912. https://doi.org/10.1785/0120020241.

Zhang, W., Li, Z., Gao, X., Li, Y., Shi, Y., 2020. Arrival-time detection in wind speed measurement: wavelet transform and bayesian information criteria. Sensors 20 (1), 269. https://doi.org/10.3390/s20010269.