A Robust and Sufficient Algorithm For Automatic First Arrival Picking Using Higher-order Statistics

A. Guirguis, A. El-Dahshan and A. Yahia

***Abstract*—**A demand for a reliable automated first break times picking algorithm is incredibly growing among service petroleum companies. This is because its significance to build an accurate velocity profile through further processing stages. Several algorithms suffer from poor conditions of recorded seismic data and low Signal-noise-ratio (S/N). In this paper, a comparative study reports concerns with both second and third order statistics including the cross-correlation and bispectral correlation techniques with the aim of examining their performances of time delay estimation. Based on results, a successful picking algorithm is proposed with few picking parameters and then test on real seismic data sets. Here, we illustrate the capability of third order statistics (bispectral correlation) to cope with Gaussian noises and nonlinearity issues of real seismic data sets

***Keywords*—** Seismology; Digital seismic Processing; correlation techniques; algorithms.

I. Introduction

Over the last few decades, Time delay estimation has been solving several problematic issues in different disciplines in industry including radar [1-3], sonar [4,5], weather forecasting, and environment measurements [6]. Time delay estimation always targets to extract vital hidden information inside received signals based on obtained delays. This target is achieved by evolving mathematical theories, statistics and numerical analysis.

One of the most crucial processes in seismic exploration is to substitute time delays in recorded seismic data due to the spatial distribution of sensors, this process is known as First Break Picking process. the quality of obtained first arrival times is strongly related to the accuracy of estimated delays. The first break picking is extremely important to build an accurate velocity model through the static correlation stage that is vital in further processing stages [7]. Although the picking process is usually carried out manually by the visual inspection of the amplitudes of refractions waves, this method suffers from various drawbacks, for instance, time costs and human errors. Therefore, this may lead to building an imprecise picking model.

Several techniques have been developed to carry out the picking process automatically such as energy ratio functions [8], neural networks [9], and wavelet techniques [10]. second order statistics are also used. For example, the cross-correlation technique [11,12].

The authors are with the Physics group, Basic Science Department, Faculty of Engineering, The British University in Egypt, Cairo 11837, Egypt

Unfortunately, most of these techniques do not work well with real seismic data because of the low signal-to-noise ratio (S/N) and nonlinearity of ambiguous signals. Other nonconventional methods are involved like higher-order statistics, for example, third order statistics including bispectral correlation technique [13,14] as they have an extraordinary ability to suppress Gaussian noises and this is proven theoretically [15].

In this context, we compare between the Cross-correlation and Bispectral techniques based on their Time Delay Estimation' performances with the aim of emphasizing the Third order statistics (bispectral correlation) ability to suppress Gaussian noises in real seismic data conditions. An automatic First break picking algorithm also proposes based on time delay estimators' results. In addition, some picking parameters suggest involving time window, threshold time delay and hypothetical velocity model in order to guide the estimation process and neglect mispicks.

II. Theory

Time delay estimation technique is used to expect the time difference between two versions of received signals from separated spatially sensors. In most common applications, discrete-time signals can be mathematically described by eq. (1).

$$x(n) = s(n) + q_1(n) \;\&\; y(n) = s(n \pm \tau) + q_1(n) \qquad (1)$$

Where $s(n)$ is the recorded signal, $q_1(n)$ & $q_2(n)$ are two independent noises terms, $s(n \pm \tau)$ is a shifted version of signal with certain leg or shift (τ) due to the spatial distribution of sensors and (N) is the number of samples in each seismic trace.

There are two well-known classes of correlation theories. The first class is Second-order statistics mainly known by the cross-correlation method. This method can be defined as a shifted form of the convolution technique [16]. Other versions of the technique are developed to tackle drawbacks of the conventional cross-correlation technique. For example, Coherence ratio technique that is a normalized version of the cross-correlation technique by the reference power "autocorrelation" [17]. The Phase delay approach also belongs to the second-order statistics family, which depends on the determining phase shift in the cross-correlation pattern with respect of the autocorrelation's phase [18, 19]. Unfortunately, these earlier techniques are very sensitive to noise and mainly depend on the amplitude of arrived seismic events. The cross-correlation and autocorrelation techniques can be calculated in frequency domains by eq. $(2,3)$ [20] as following:

$$\mathrm{CC}_{\mathrm{xy}}(\lambda) = \mathrm{X}(\lambda) \times \mathrm{Y}(\lambda)^* \quad (2)$$

$$\mathrm{AC}_{\mathrm{xx}}(\lambda) = \mathrm{X}(\lambda) \times \mathrm{X}(\lambda)^* \quad (3)$$

Where CC_{xy} & AC_{xx} are the cross correlation and autocorrelation patterns in the frequency domain respectively, $X(\lambda), Y(\lambda)$ are two seismic received signals in frequency domain, and (*) is a mathematical operator "complex conjugate operator".

It has been reported that the high-order statistics can suppress all Gaussian noises, this is because all Polyspectra of sinusoidal signals are vanished theoretically in this domain [21,22]. To define the Third-order statistics technique, it correlates any received signal with the reference signal at a synthetic frequency that equals to the sum of frequency components of both signals. Third-order statistics can also estimate and determine the phase shift between two different frequency components of the same signal [23]. Bispectral correlation is considered as the most dominant techniques in the third-order statistics which can compute as the ratio between the cross-bispectrum $BC_{xyx}(\lambda_1, \lambda_2)$ and auto-bispectrum $AC_{xxx}(\lambda_1, \lambda_2)$ that usually computed in the frequency domain by eq. $(4-6)$ [24].

$$\mathrm{BC}_{\mathrm{xyx}}(\lambda_1, \lambda_2) = \mathrm{Y}(\lambda_2) \times \mathrm{X}(\lambda_1) \times \mathrm{X}(\lambda_1 + \lambda_2)^* \quad (4)$$

$$\mathrm{AC}_{\mathrm{xxx}}(\lambda_1, \lambda_2) = \mathrm{X}(\lambda_2) \times \mathrm{X}(\lambda_1) \times \mathrm{X}(\lambda_1 + \lambda_2)^* \quad (5)$$

$$\beta_{\mathrm{xxx}}(\lambda_1, \lambda_2) = \frac{\mathrm{BC}_{\mathrm{xyx}}(\lambda_1, \lambda_2)}{\mathrm{AC}_{\mathrm{xyx}}(\lambda_1, \lambda_2)} \quad (6)$$

Where λ_1, λ_2 are the indexes of frequency components and can be written as $\lambda_1 = \lambda_2 = 0,1, N-1$.

In the last few decades, third-order statistics have been used for several purposes including providing consistent time picking in seismogram analysis [25, 26] and determining phase shifts in regional tomography models [27]. Most of the previous publications had worked on synthetic seismic data with different noises ratio, or on a couple of real seismic traces [28]. In this work, we propose a robust algorithm to pick/track times of any seismic event automatically. This algorithm also can be usedwith different data sets in various applications.

III. First Break Arrival Times Algorithm

Our algorithm mainly depends on the estimation of delays between received signals using correlation techniques with the help of some picking parameters. The picking process can be done through three separate stages as follows:

A. Stage I: Initial Inspection & Picking Parameters.

Each seismic data set has certain conditions that need to be taken into consideration during the picking process. These conditions can be considered in the algorithm through the picking parameters. These parameters can be:

- The reference signal & its reference time (t_o) to start with it.
- A picking time window (W), which is very crucial for minimizing the computational time of the picking process.
- A hypothetical velocity model to preserve the move-out of certain events in the seismic data.
- A Threshold time delay to avoid mispicks. Here, the first arrival event "refractions" is chosen to pick their times up.

B. Stage II: Time Delay Estimation

The time delay estimation can be computed via correlating the reference signal with it is neighbor which finds out the time lag between them; this crossponds to the peak value of the computed correlation pattern. Then, it can be used to express the time delay via knowing the sampling rate (Δt) of recorded seismic data. In this algorithm, the time delays can be determined by two different correlation techniques: the cross-correlation and bispectral techniques through calling subroutine function that is responsible for one of them at the beginning of algorithm.

C. Stage III: Computing First Break & Quality Control

The First break time can be computed via adding the estimated time delay (t_d) between received signals to the reference time (t_o). It is important to check the validity of estimated delay relative to the threshold time delay for preserving the event move-out and ignoring wrong picks' times. This can be done via comparing obtained time delay with a threshold time delay according to a controlled margin of error factor (ζ) that will be defined inside the algorithm. Following this, the reference signal will be updated to be the second signal in the data matrix and the neighboring signal will be the third one. Next, the reference time is added to the estimated time delay. By this routine, all seismic traces will be correlated to each other. We also recommend dividing the data matrix into two parts to avoid the accumulative error as much as possible. The flowchart of the proposed algorithm with all specific steps is shown
Figure 1.

IV. Results and Discussion

2D real seismic data were gathered from a survey that was mainly concerned with the shallow reflection survey, Kansas City, USA [29]. The Kansas data collected by 96 geophones, with 0.5 offset. The recorded data was 500 (ms) long and sampling rate equals 0.25(ms). The geometry of the Kansas survey and the first examination of data with some identified events are presented in **Figure 2**.

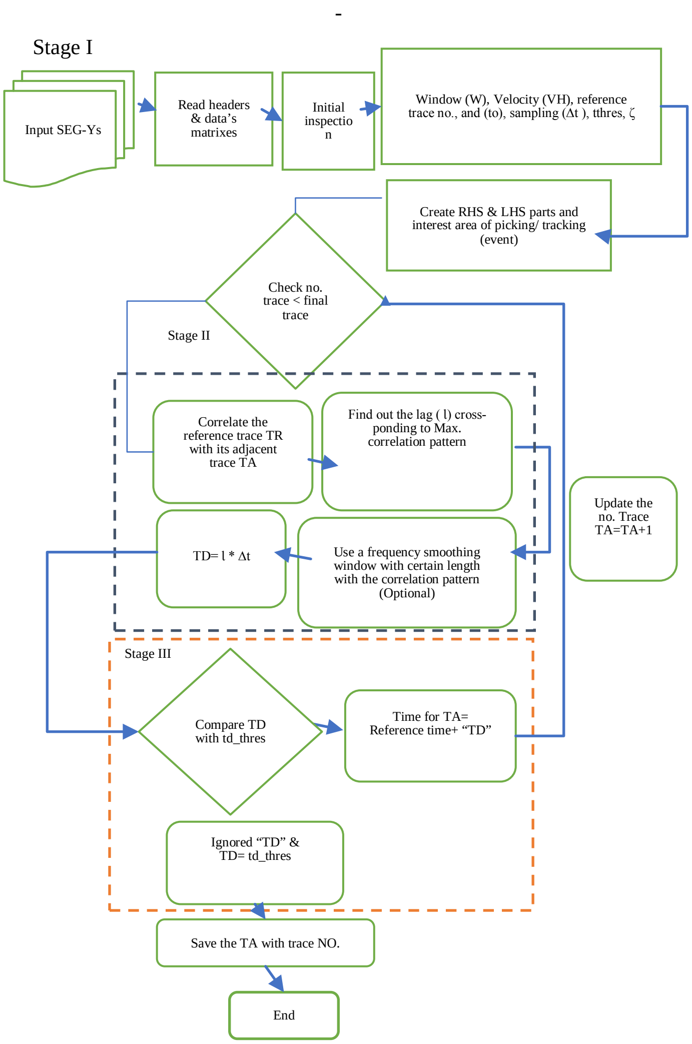


Figure 1. The flowchart of automated first arrival times picker algorithm

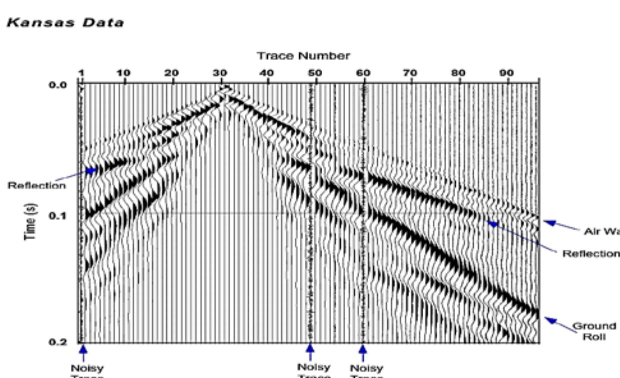


Figure 2. A diagram of the geometry of the Kansas survey and Initial inspection of the Kansas seismic shot @Copyright to Ref. [30].

We tested the proposed algorithm on an arbitrary seismic shot from the Kansas data based on different two-time delay estimators': cross-correlation and Bispectral correlation measures to pick times of first arrivals. There are some essential picking parameters that were adjusted and unified for both estimators, these are summarized in **TABLE 1** .

PARAMETER	RECOMMENDED VALUES IN (MS)
Reference time, to	60
Time window, W	15
Sampling interval, Δt	0.25
Threshold Time delay, $t_{thres.}$	0.03
Error factor, ζ	1e-4

TABLE 1. Some recommended picking parameters for the reported picking results

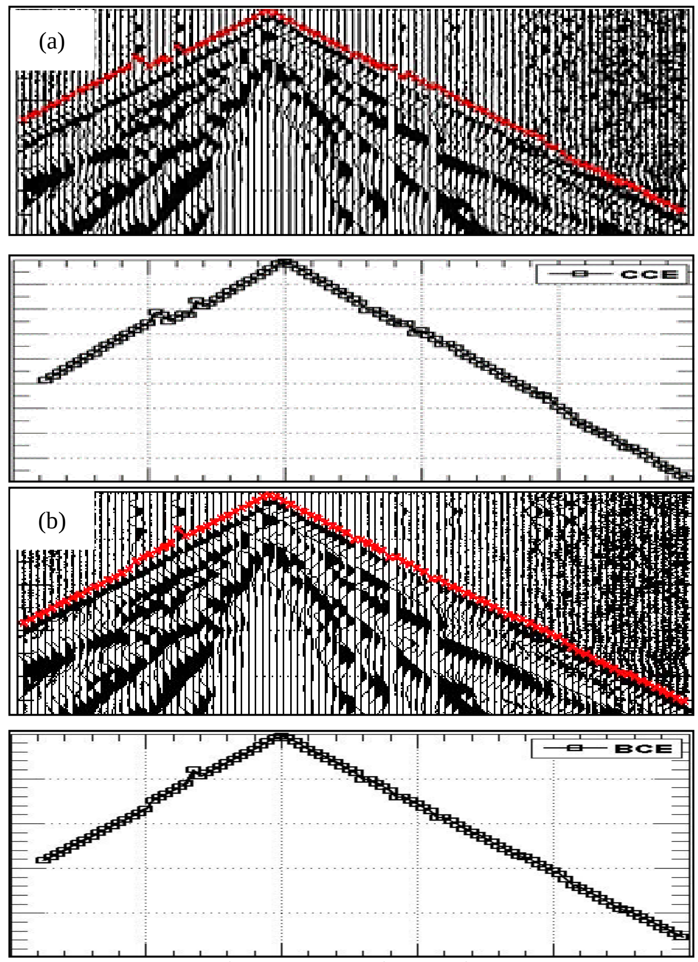


Figure 3. The automatic first break times of arrivals for seismic shot overlaid on the seismic view and the picked time model separately below it determined by a) the cross-correlation and b) bispectral-correlation estimators.

It is clear that picked times based on the cross-correlation technique are entirely affected by embedded Gaussian noises, in contrast, to a consistent time model built up via the bispectral correlation technique. All seismic traces have been used without any kind of pre-filtration techniques. This means that bispectral correlation is able to suppress Gaussian noises in real seismic data and preserve all details for velocity calculation stage.

This algorithm also can be used to track any seismic event in the seismic shot by adjusting the picking parameters. **Figure 4** shows a time model obtained from tracking certain seismic event's "reflection", these are represented by a blue line on the seismic view of the shot. Again, the cross-correlation technique is corrupted with noises and cannot tackle the nonlinearity problem, whereas the bispectral correction tracked the whole event well and saved the moveout without any mispicks with weak, or even missing amplitudes in the original data. This certainly shows the capability of the bispectral correlation technique and the algorithm to pick and track times accurately in real seismic data conditions.

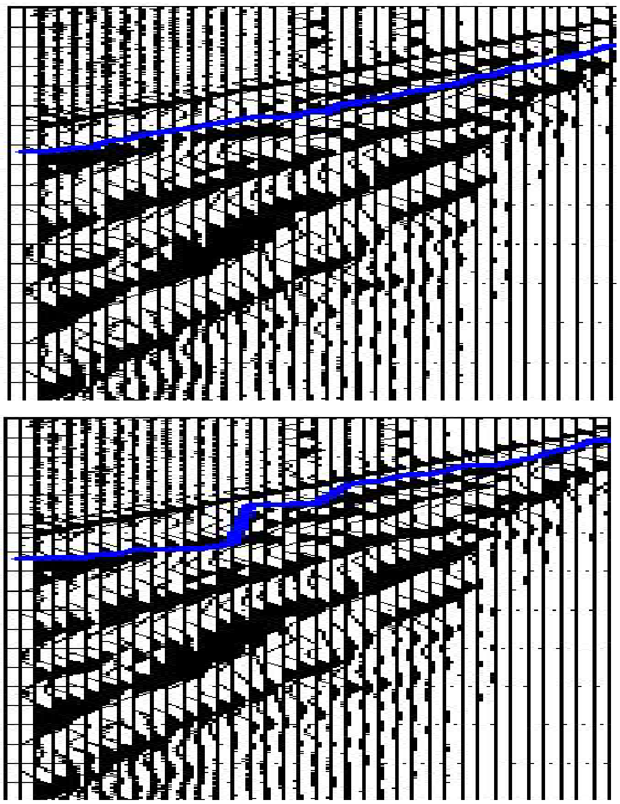
(a)
(b)


Figure 4. The automatic time tracking of seismic event in certain seismic shot, which is presented with a blue line overlaid on the seismic view that are determined by a) the cross-correlation and b) bispectral-correlation estimators.

V. Conclusion

In this paper, we reported on a comparative study of the performance of the time delay estimation based on the cross-correlation and bispectral correlation techniques. These time delay approaches are involved to compute the first arrival times of seismic events automatically. We also proposed an algorithm that works with some picking parameters, which are picking window, reference time, hypothetical velocity, threshold time delay and error factor. We realized that most of the time delay estimators based on the second-order domain statistics have been suffering from low signal-to-noise ratio and the nonlinear nature of seismic data even when using filtration techniques. In contrast, with others time-delay estimators based on the third order domain statistics that have extraordinary ability to suppress the embedded Gaussian noises with seismic data without using any pre-filtration processes that may remove some valuable features of processed seismic section interpretation. Although the bispectral correlation might take long computational time, it provides a consistent time model in real conditions. At the end, we recommend selecting picking parameters wisely because they play crucial factors in picking results.

VI. Acknowledgments

The authors gratefully acknowledge the support of the Schlumberger Cambridge Research for student internship grant, especially Dr. Larry Velasco for his technical comments. Authors are appreciated the time and effort of Prof. Vaneeta D'Andrea to revise manuscript’s language before publishing.

VII. References

[1]. G. Dimitris, Manolakis and J. G. Proakis, Digital Signal Processing: Principle, Algorithms and Application. New Jersey: Prince-Hall, 2013
[2]. P. Protiva, J. Mrkvica and J. Macha, “"Time delay estimation of UWB radar signals backscattered from a wall", Microwave and Optical Technology Letter, vol. 53**,** pp.1444, 2011.
[3]. K. V Mishra and Y. C. Eldar, “"Performance of time delay estimation in a cognitive radar", preprint arXiv:1701.03898, 2017.
[4]. G. Carter, “"Time delay estimation for passive sonar signal processing", IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.29**,** pp. 463, 1981.
[5]. C. Changlin and J. Wei, “"The Estimation Method on the Phase Delay Between Two Real Signals", International Conference on Communications, Circuits and Systems. IEEEXplore, pp. 394, 2006.
[6]. L. L. Lai, H. Braun, Q. P. Zhang, Y. N. Ma, W. C. Sun and L. Yang, “Intelligent Weather Forecast”, Proceedings of the 3rd International Conference on Machine Learning and Cybernetics (IEEE, 2005), Vol.8, pp. 4216-4221.
[7]. O. Yilmaz, Seismic data processing (SEG, 1987).
[8]. F. Coppens, "First arrivals picking on common-offset trace collections for automatic estimation of static corrections” Geophysical Prospecting, vol. 33, pp.1212-1231, 1985.
[9]. A. J. Rudman and M. E. Murt, "Automated first arrival picking: A neural Automated first arrival picking.", Geophysical Prospecting, vol. 40**,** pp. 587-604, 1992.
[10]. I. M. Tibuleac, E. T. Herrin, J. M. Britton, R. Shumway and A. C. Rosca, “Automatic secondary seismic phase picking using wavelet transform”, Proceeding of 25th Seismic Research Review (SRR-03), Nuclear Explosion Monitoring: Building the Knowledge Base, pp.352-359, 2003.
[11]. A Clement and R Peraldi, “Digital processing of refraction data Study of first arrivals", Geophysical Prospecting, vol. 20**,** pp. 529-548, 1972.
[12]. V. Shtivelman and B. Gelchinsky, “"Automatic picking of first arrivals and parameterization of traveltime curves", Geophysical Prospecting, vol. 31, pp. 915-928, 1983.
[13]. R. Pan and C. L. Nikias, “Time delay estimation in unknown Gaussian spatially correlated noise", IEEE

Trans. Acoust. Speech Signal Process, vol. 36, pp. 1706-1714, 1988.
[14]. A. P. Petropulu and C. L. Nikias, Higher-order spectra: A nonlinear signal processing frame work, Englewood Cliffs: Prentice-Hall, 1993.
[15]. L.T. Ikelle and K. Srinivasan, “"Applications of third-order statistics for the automatic time picking of seismic events", Geophysical Prospecting vol. 49, pp.58-155, 2001.
[16]. J. A. Jiang and L. A. Tan. Digital Signal Processing: Fundamentals and Applications (Elsevier Science, 2013).
[17]. A. H. Whale. Effects of Ionospheric Scattering on Very-Long-Distance Radio Communication, Boston, MA: Springer, 1969.
[18]. C. Changlin and J. Wei, “The Estimation Method on the Phase Delay Between Two Real Signals", International Conference on Communications, Circuits and Systems. IEEE, pp.394-397, 2006.
[19]. S. L. Marple, “Estimating group delay and phase delay via discrete-time “analytic” cross-correlation", IEEE Transactions on Signal Processing, vol. 47, pp. 2604-2607, 1999.
[20]. R. Bracewell, Pentagram Notation for Cross Correlation." The Fourier Transform and Its Applications: New York: McGraw-Hill, 1965.
[21]. A. P. Petropulu and C. L. Nikias, "Higher-Order Spectra Analysis." In A Nonlinear Signal Processing Framework”, Englewood Cliffs, N.J: Prentice-Hall, 1993.
[22]. J. M. Mendel, “Tutorial on Higher-Order Statistics (Spectra) in Signal Processing and Systems Theory: Theoretical Results", Proceedings of the IEEE, pp. 278-305, 1991.
[23]. M. R. Raghuveer and C. L. Nikias, "Bispectrum Estimation : A Digital Signal Processing Framework", Proceedings of the IEEE, pp. 869-891, 1987.
[24]. M. R. Raghuveer and C. L. Nikias, "Bispectrum Estimation : A Digital Signal Processing Framework", Proceedings of the IEEE, pp.869-891, 1987.
[25]. B.L. Kennett and C. Bai., "Automatic Phase-Detection and Identification by Full Use of a Single Three-Component Broadband Seismogram", Bulletin of the Seismological Society of America, vol. 90, pp. 187-198, 2000.
[26]. L. Küperkoch, T. Meier, J Lee and W Friederich, "Automated determination of P-phase arrival times at regional and local distances using higher order statistics", Geophysical Journal International, vol. 181, pp. 1159-1170, 2010.
[27]. T. Diehl, E. Kissling, S. Husen and F. Aldersons, "Consistent phase picking for regional tomography models: application to the Greater Alpine region", Geophys. J. Int. vol. 176, pp. 542-554, 2009.
[28]. L. T. Ikelle and S. K. Yung, "An example of seismic time picking by third-order bicoherence, GEOPHYSICS, vol. 6, pp.1947-1951, 1997.
[29]. S. G. Baker, Processing Near-Surface Seismic-Reflection Data: A Primer (SEG, 1999).
[30]. A. Guirguis, Time Delay Estimation Technique & Seismic Processing Applications, M.Sc. thesis, (Lambert Academic Publishing, 2017).