Wavelet Based Automatic Phase Picking Algorithm for 3-Component Broadband Seismological Data

A. Ahmed[1], M.L.Sharma[2], and A. Sharma[1]

1. Department of Electrical Engineering, Indian Institute of Technology, Roorkee, India
2. Department of Earthquake Engineering, Indian Institute of Technology, Roorkee, India, email: mukutfeq@iitr.ernet.in

ABSTRACT: *A phase picking algorithm has been developed for P-wave arrival detection and picking on a 3-component broadband seismogram. The algorithm is based on Wavelet Transform, Akaike Information Criteria (AIC) picker and voting mechanism. The mathematical functions have been used to cut up data into different frequency components and each component with a resolution matched to its scale is studied. Continuous Wavelet Transform (CWT) has been applied to pre-filtered 3-component broadband seismograms to compute scales of interest. Wavelet coefficients have been calculated in a sliding time window and AIC picker is applied to these coefficients. Consistency of 3 components picks is the primary criteria of phase detection. Further, for phase picking, AIC picker is applied on the filtered seismogram in the same time window where phase is detected. The comparison of the present algorithm has been checked with the time domain automatic phase pickers and manual picks by seismologists. The algorithm has been tested on regional earthquake data acquired through deployment of broadband seismological station in Garhwal Kumaon Himalayan region and shows a good agreement between analyst picks and auto picks.*

Keywords: Phase picking algorithm; P-wave detection; 3-component broadband data

1. Introduction

Seismic data collection is a prerequisite for the disaster mitigation and by introducing digital seismic data acquisition, the long term continuous recording and archiving of seismic signals has became a demanding technical problem. A seismic network or even a single station operating continuously at high sampling frequency produces an enormous amount of data, which is often difficult to store (and analyze) locally or even at the recording center of a network. It is of paramount importance to devise automatic trigger algorithms, which automatically picks the first P-wave arrival so that only the data on events is collected and the information is available online. Automatic phase picking here means detecting and picking accurately the first P-wave arrival, as quickly as possible. It is immensely helpful for finding event location and its identification for source mechanism analysis. Phase picking is also helpful in systems, which releases alert messages quickly after an earthquake. Such warning systems are dependent on automatic, quasi real-time procedures for detecting and onset picking in seismogram recordings. Automatic methods are needed for such picking as manual picking is time consuming and subjective.

Automatic phase picking algorithm remains one of the current research topics and a number of different methods have been implemented in the past to detect and pick seismic phases and estimate their onset time, from single-component as well as three component (3-*C*) recordings. Most of the event detectors and phase pickers are based on the

STA (short term average) and *LTA* (long term average) algorithm [29]. Allen [3] used Characteristics Function which are the trace amplitude and the time derivative of the trace which is compared to some threshold value. Sharma et al [22] have used the modified Characteristics Function which includes prefiltering in time domain and *STA-LTA* algorithm for digital data acquired from Garhwal Kumaoun Himalaya. Baer and Kradolfer [5] gave amplitude envelope function, with dynamic threshold value. Morita and Hamaguchi [19] used a statistical adaptive algorithm to estimate the onset time using single component seismogram. A similar approach was taken by Pisarenko et al [20], Takanami and Kitagawa [24], and Tarvainen [25] using 3-component data. Rudd and Husebye [21] combined signal polarization and *STA/LTA* to devise a 3-Component phase identifier, mainly designed for P-waves. Detectors and pickers specially designed for S-waves and Rayleigh waves have been suggested by Cichowicz [7] and Chael [6] respectively. Other methods include energy analysis [9], polarization analysis [27] and autoregressive techniques [10, 14, 15, 16, 23]. Detection is the process for finding presence of seismic phases, whereas picking involves accurately marking the onset time of phases.

In this work, the use of wavelet *AIC* picker for detecting and picking first *P*-wave arrival on 3-*C* broadband seismogram is presented. Such wavelet *AIC* picker has already been implemented by Zhang et al [30] using single component. Wavelet transform has been used to detect and pick the onset time of several seismic phases. Anant and Dowla [4], Tibuleac and Herrin [26] and Gendron et al [11] have used wavelet transform for determination of different seismic phases in the incoming seismogram time series. Since the data obtained from Garhwal Kumaon Himalayan Region is very noisy, the broad-band seismograms are first treated with appropriate filters to enhance the signal to noise ratio (*SNR*). The signal is then decomposed into different scales, in order to enhance features of phase arrival at different resolutions. Important features (such as phase arrival) are retained over several resolution scales while irrelevant one (i.e. noise) decay quickly at larger scales [8]. The *AIC* picker [16] is applied to these wavelet coefficients over multiple scales. Consistency of these picks at different scales is our criterion for detection. Once there is detection in the current moving time window, again *AIC* picker is applied to the filtered seismogram, in the same time window, to obtain the onset time.

The algorithm has been tested on the regional earthquake data set obtained from Garhwal Kumaon Himalayan Region, which is recorded at the sampling frequency of 100 samples per second using broad band seismometers (*CMG*-3*T*). The broad band array has been deployed by Department of Earthquake Engineering, Indian Institute of Technology Roorkee in the Garhwal Kumaoun Himalaya under a research scheme funded by Department of Science and Technology, New Delhi. The comparison of the autopick with the analyst pick has been done and reported here in the present study.

2. Wavelet Transform

Wavelets are necessarily mathematical functions that cut up the time series into different frequency components, and then study each component with a resolution matched to its scale. This type of processing has the advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes. Wavelets were developed independently in the fields of mathematics, quantum physics, electrical engineering, and seismic geology/geophysics. The multidisciplinary approaches during the last decades have led to many new wavelet applications such as image compression, turbulence, human vision, radar, and earthquake prediction. Because seismic waves traveling through complex media are composed of time-frequency-localized waveforms, it is a better choice to represent the seismogram locally both in time and frequency domains. Wavelet transform, which was evolved from the work on seismic signal [12-13], is a very powerful tool for the analysis of such non-stationary signals.

Wavelet transform can be accomplished in two ways: Continuous Wavelet Transform (*CWT*) and Discrete Wavelet Transform (*DWT*). The *CWT* is defined as follows.

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) g\left(\frac{t-b}{a}\right) dt \tag{1}$$

Where $g(t)$ is the wavelet function, and a and b are the scale and translation factors, respectively. The wavelet function $g(t)$ decays rapidly to zero with increasing t and has zero mean. The domain (range) of nonzero values of the wavelet is called the support. The scale factor controls the dilation or compression of the wavelet, whereas at lower scales, the wavelet is compressed and characterizes the rapidly changing details of the signal. At higher scales, the wavelet is stretched over a greater time span and the slowly changing and coarse features are better resolved [30].

In practical applications, the *DWT* is generally preferred because waveforms are recorded as discrete time samples. *DWT* can be implemented quickly via the Mallat algorithm [17-18]. A low-pass filter and high-pass filter can be used to calculate the wavelet coefficients of a discrete time series recursively. The following tree clearly shows how *DWT* is implemented using recursively half band filters, see Figure (1). Fourier transform gives us the global information of the frequency content available in the signal. Local frequency content can not be analyzed using Fourier transform. Wavelet transform has advantage over the Fourier transform because of its ability to characterize the signal features locally with a detail matched to its scale.

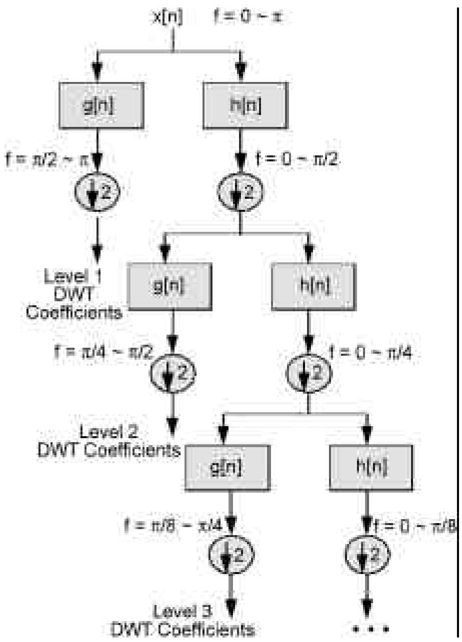

Figure 1. The Subband Coding Algorithm.

3. AIC Picker

AIC picker is a mathematical function whose value can be calculated for each sample of a time series. The global minimum value of *AIC* gives us the optimal separation of time series and this is the feature of *AIC* picker which is mostly exploited in many ways in various applications. There are two equations available for calculating *AIC* values viz using Autoregressive coefficients and without using them.

For *AR-AIC* (Autoregressive-*AIC*) approach, the seismogram, i.e., the time series is divided into locally stationary segments, each one is modeled as an *AR* process, and intervals before and after the onset time are two different stationary processes [23]. The *AIC* is then used to determine the order of *AR* process when fitting the portion of the seismogram, which indicates the badness and unreliability of the model fit [1]. When the order of *AR* process is fixed, the *AIC* function gives us the measure of the model fit. The point where *AIC* gives a minimum value (in least square sense) is the optimal separation of the two stationary time series [23]. This is known as *AR-AIC* picker approach [2, 14, 23]. The *AIC* of the two interval model for seismogram x of length N is represented as function of merging point k [23];

$$AIC(k) = (k - M)\, log(\sigma^2_{1,max}) + (N - M \_ K)\, log(\sigma^2_{2,max}) + C_2 \tag{2}$$

Where M is the order of an *AR* process fitting the data, C_2 is a constant, and $\sigma^2{}_{1,max}$ and $\sigma^2{}_{2,max}$ indicate the variance of the seismogram in the two intervals not explained by the autoregressive process.

The second approach of computing the *AIC* function is without using *AR* coefficients, which can be calculated directly from the time series, i.e., the seismogram. The onset point is again the minimum of the *AIC* function. For seismogram x of length N, the *AIC* value is defined as [16];

$$AIC(k) = K.log\{var(x[1,K])\} + (N - k - 1).log\{var(x[K+1,N])\} \tag{3}$$

where k ranges through all the seismogram samples.

The *AIC* picker defines the onset point as the global minimum. For this reason, it is necessary to choose a time window that includes only the seismogram segment of interest [30]. In our work, Eq. (3) is used for calculating *AIC* function. Figure (2) shows the seismic signal in the upper traces and the corresponding *AIC* values computed from it in the lower traces. For a seismogram with very clear onset, *AIC* values have a very clear global minimum, Figure (2a), which corresponds to P-wave arrival. For a seismogram with relatively low *SNR*, there may be several local minima but the global minimum still gives accurately the P-wave onset, see Figure (2b). If *SNR* is very low, global minimum is not very clear and can not be identified as a phase onset, see Figure (2c). So here lies the shortcoming that our algorithm may not detect phase onset if *SNR* is low (which is also the problem with other phase picker algorithms). Also for multiple phases available in a single time window, the *AIC* picker will show its global minimum at the strongest phase, see Figure (2d). On the other hand, since there is always a global minimum in a time window, our picker will always pick an onset for any part of data irrespective of its being a phase or noise, see Figure (2e). Therefore there is a need to choose an appropriate time window and instead of looking into the *AIC* values of data only, we go for calculation of *AIC* values of seismogram segment over multiple scales.

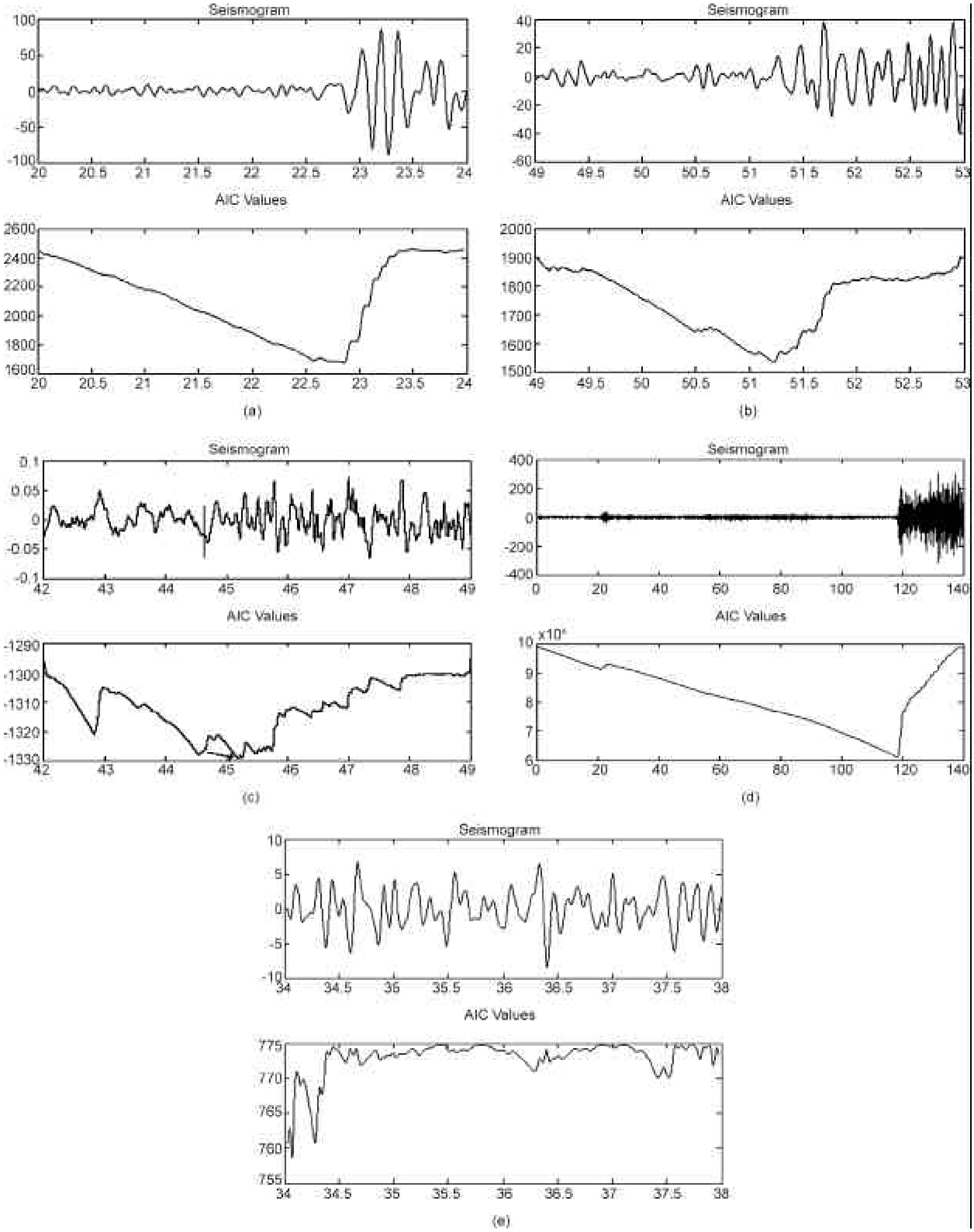

Figure 2. Seismograms and their corresponding AIC values. (a) Clear P-wave arrival. (b) Clear P-wave arrival with relatively lower S/N ratio (c) Very low SNR seismogram (d) Multiple phases available in a single window. (e) Random noise, where global minimum is not a phase arrival.

4. Automatic Phase Picker Development and Testing

The first step in the processing is the filtering of the seismogram. Seismogram is filtered for signal enhancement prior to applying the wavelet *AIC* picker. For this algorithm, 5 pole Butterworth band pass filter is used to pass 2-10*Hz* band. This band is chosen after looking into Short Time Fourier Transform (*STFT*) of several P-wave arrivals, which lie in this band. Wavelet *AIC* picker combines the *AIC* picker with multiscale wavelet analysis, in which the *AIC* picker is applied to absolute wavelet coefficients at different scales. For P-wave arrival detection, the *AIC* picker must pick the phase at different scales, within a proximity to each other, see Figure (3a) and Figure (3b). All three components of seismogram and decompose each component are used into 3 scales. After choosing a suitable time window, the *AIC* values are calculated for all 3 decomposed components and the consistency of picks were considered. This is achieved by voting mechanism such as checking the consistency of each pick with the other as well as counting the number of times the picks was checked. Maximum voting for such consistency check is 36($^9C_2$). If the count of votes crosses a pre-decided threshold, a phase is detected in the current time window. Each component is decomposed into 3 scales which is most appropriate [30]. If more scales are used, the *COIs* of those singularities that are not isolated will have more overlap and cause more ambiguities. If only two scales are used, however, the singularity due to the noise will still be significant at scales 1 and 2 in some cases, resulting in too much false detection [30].

After detection of event, the picking is achieved

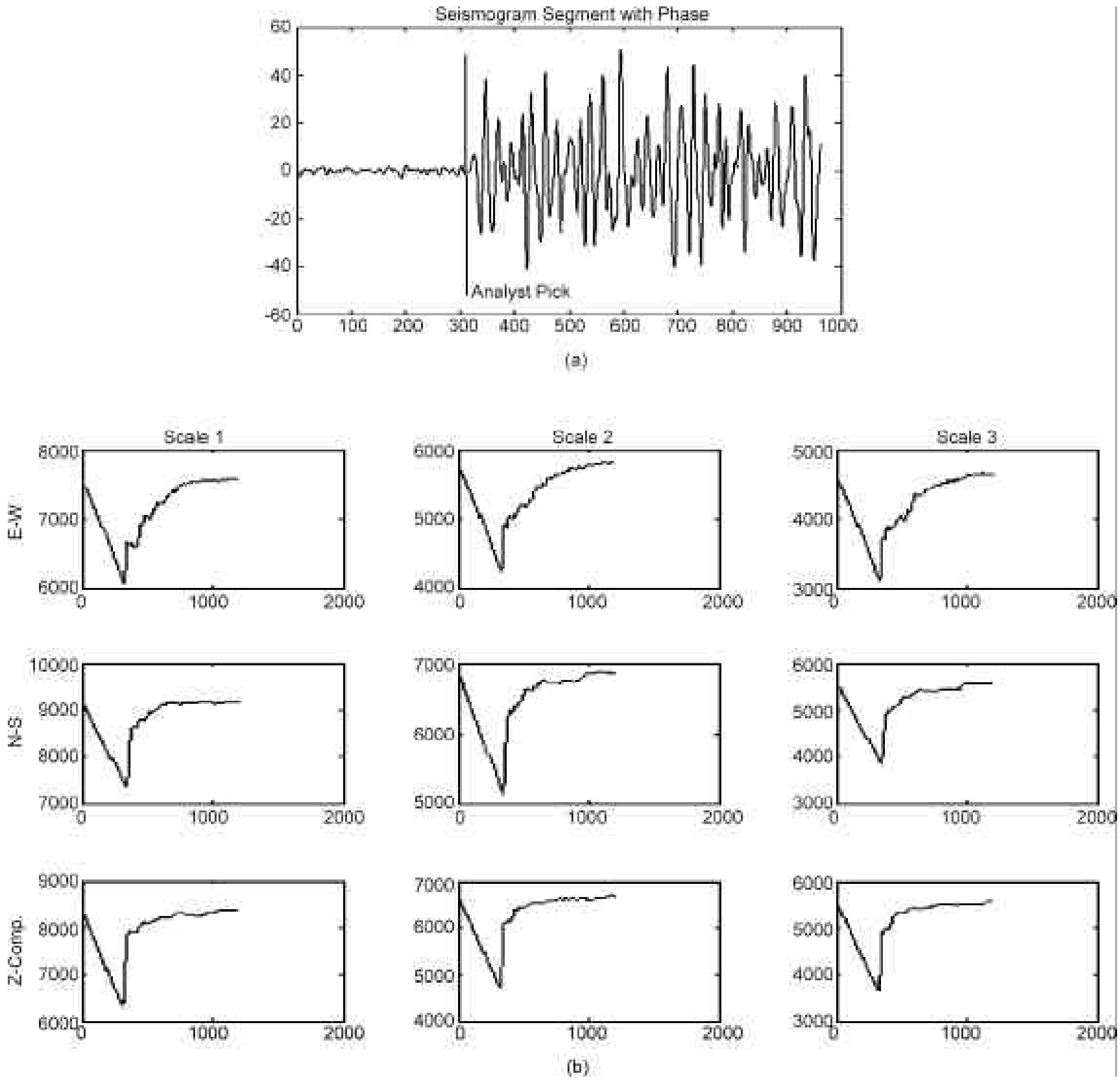

Figure 3. (a) Seismogram segment with phase arrival (vertical component). (b) Corresponding AIC values for 3-C seismogram at 3 different scales, consistency of these picks detects the availability of phase.

by again applying *AIC* picker to the vertical component of filtered seismogram, in the current time window. This pick is our final P-wave onset. Vertical component is chosen for picking as it has the maximum sensitivity to P-wave arrival.

Main criteria for choosing a wavelet function from a vast family are its support, symmetry, regularity and number of vanishing moments [28]. In our study, Daubechies wavelet of order 2 is used, which is appropriate for detecting signal singularity and has tighter support of 4.

The main objective of using the wavelet transform is to guide the work of the *AIC* picker by choosing an appropriate time window for it, in which a P-wave arrival exists [30]. The time window is chosen via a trial and error process and we found 12*sec* window quite suitable. Smaller time windows may lead to sluggish detection process as there is a need to overlap consecutive windows in order to reduce the border effect. On the other hand, larger windows may include multiple phases and our picker will pick the stronger phase which may not be first P-wave arrival. For decomposing the signal into different scales, *CWT* was used and for checking the consistency among picks, the proximity of picks within 1 sec (i.e. 100 samples) was checked. If picks in a time window are consistent more than 20 times (i.e. vote count is ≥ 20), out of 36, we consider it as P-wave detection. The *AIC* picker picks the onset time after applying it on the vertical component of seismogram in current time window.

Our algorithm allows the user to choose any filter combination, wavelet function, percentage overlap of neighboring windows, time duration for consistency check and the length of time window (with a default setting of 2-10*Hz*, "*db*2" wavelet, 20% overlapping, 1*sec* and 12*sec* respectively). Figures (4a), (4b), (5a) and (5b) gives the examples

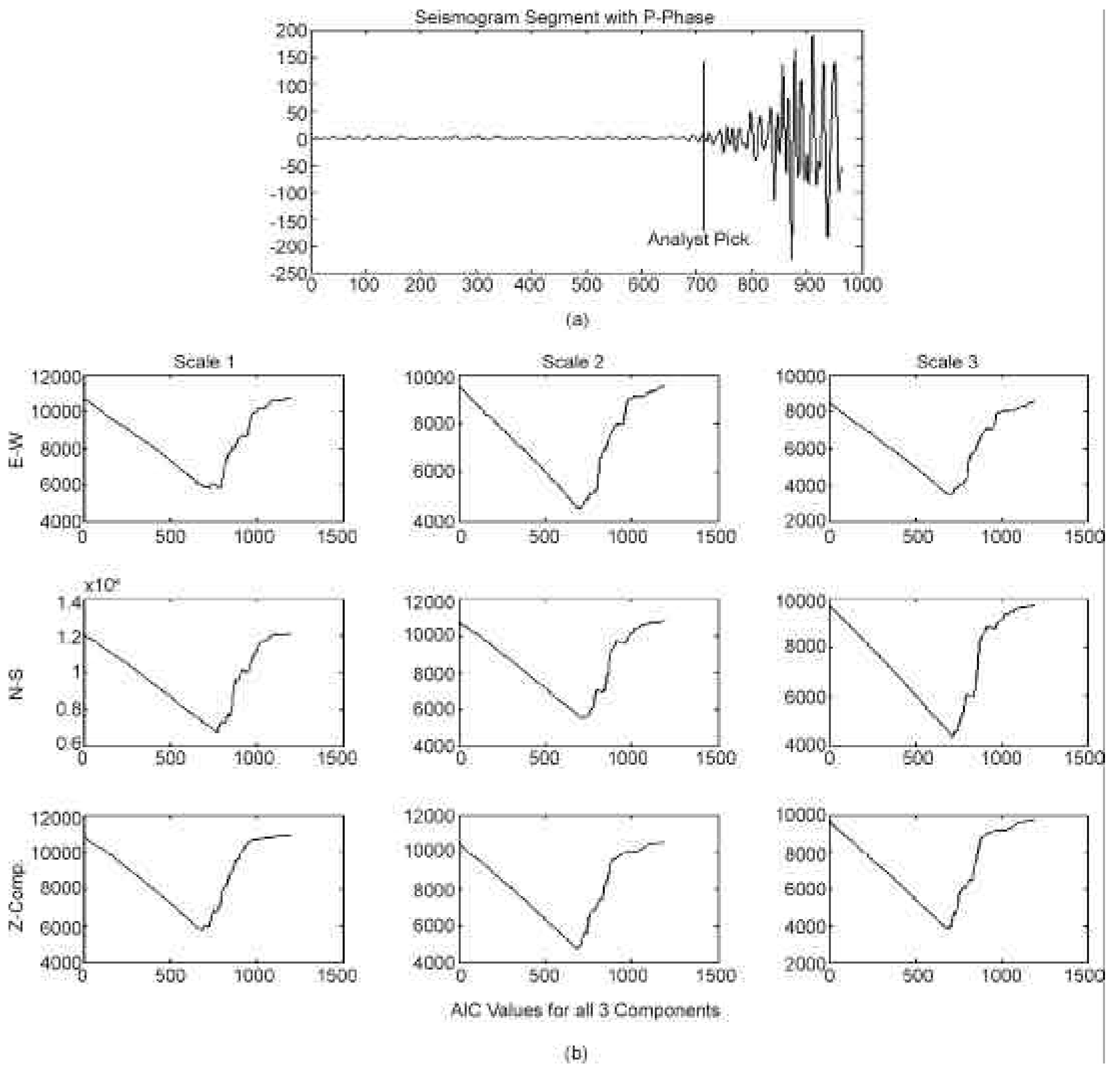

Figure 4. (a) Seismogram segment marked with Analyst pick (vertical component). (b) AIC values corresponding to the seismogram in Figure (4a), minima of AIC values clearly shows the consistency.

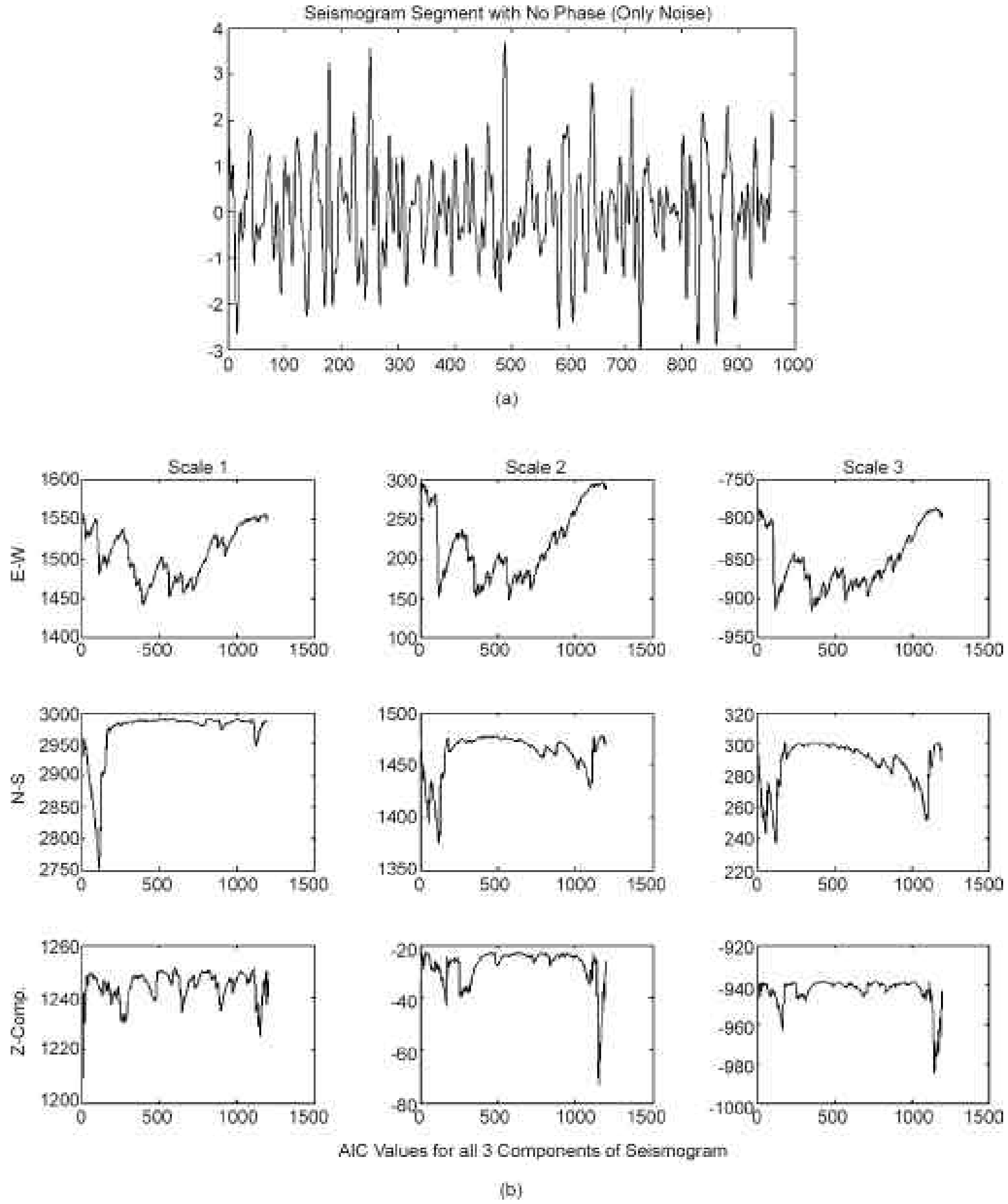

Figure 5. (a) Seismogram segment with noise only. (b). AIC values corresponding to seismogram in Figure (5a), which shows clearly that there is no consistency among the picks.

for seismogram containing strong p-phase and with no phase, respectively.

This algorithm has been tested on 450 seismograms from different earthquake events recorded on a single station. The parameters i.e. filter band, wavelet function, percentage overlap, length of time window and number of samples for proximity or consistency check, are all optimized to some extent by trial and error method, applying over the different subsets of the main data set again and again with different parameter settings. Figure (6) shows percentage of events picked accurately by this algorithm within 0.5 sec of analysts' picks, wrong detection (or false alarms), undetected events and later phases detected (i.e. S-phase).

The *AIC* based picker has also been checked with other available automatic phase pickers. The phase picking was carried out using the *STA/ LTA* base simple phase picker program available with the data processing software, Sharma et al [22] *APP* in which the filters were incorporated in the time domain while computing the *STA* and *LTA*, and the *AIC* phase picker using the single component i.e., without using voting criteria for three component. The results are summarized in Table (1). In all 450 seismograms which were checked for the four types of algorithms, the *AIC* picker using 3 component has detected 279 picks within 0.5 seconds, while using

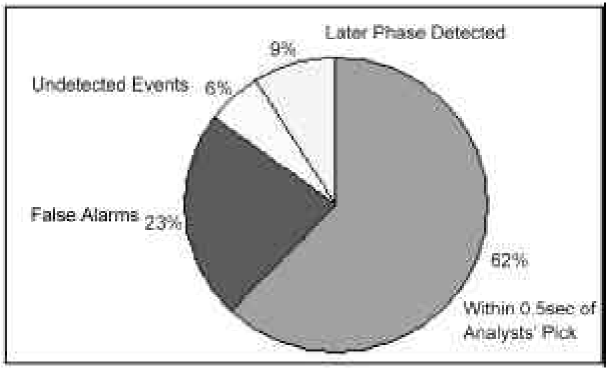

Figure 6. Pie chart showing percentage of events picked accurately by the algorithm within 0.5 sec of analysts' picks, wrong detection (or false alarms), undetected events and later phases detected (i.e. S-phase).

Table 1. Comparison of the Phase picking algorithm on the data.

Sl. No.	Name of the Picker Programe	Picks with in 0.5sec	False Alarms	Undetected Events	Later Picks Detected
1	AIC Picker 3 Comp.	279	103	27	41
2	AIC Picker Single Comp.	264	115	36	35
3	Sharma et al [22]	258	121	32	39
4	STA/LTA	246	136	35	33

single component detecting only 264 picks. The single component program was run on the vertical component only. The horizontal components showed almost similar results. The inclusion of filtering in the computing using Sharma et al [22] criteria of *STA* and *LTA* has increased the number of picks by 12. The comparison shows the advantage of the developed 3 component *AIC* picker.

The 3 component and the single component pickers using *AIC* have been compared for the picks within 0.5 seconds of the analyst picks in Table (2). The 3 component picker has picked 279 picks while the single component without voting criteria has picked only 264. The comparison shows the increase in accuracy while using the voting criteria.

Table 2. Percentage distribution of all picks picked by the AIC algorithm within 0.5 sec.

		Percentage of Picks as Compared to Manual Picking			
Sl. No.		With in 0.1sec	0.1 – 0.2sec	0.2 – 0.3sec	00.3 – 0.5sec
1	AIC Picker 3 Comp.	52	21	10	17
2	AIC Picker Single Comp.	45	20	15	20

5. Conclusions

An algorithm has been developed for automatic P-phase picking using *AIC* picker combined with multiscale wavelet analysis in the present study. The data acquired through *CMG*3*T* broadband seismometer has been used for checking the developed phase picking algorithm. As P-wave arrival in a seismogram is a significant feature therefore it will be retained over several scales, whereas noise or other incoherent features will disappear quickly over larger scales. One of the main advantage of this algorithm over the other algorithm is its checking of the signal in various frequency bands. On the other side, the available phase picking algorithms use the *STA/LTA's* in one frequency band only. *AIC* picker is applied directly to the absolute wavelet coefficients calculated from different components of seismogram. Detection criterion is to check consistency quantified as a minimum number, among the picks in different windows (3 components ~9 windows). Picking is achieved by again applying *AIC* picker to the vertical component of seismogram segment in the current time window. This algorithm has been tested on a dataset of seismic events occurring in the Garhwal Kumaon Himalayan Region. Around 450 events of varying length are checked using these algorithms. The comparison of other such algorithms has shown the advantage of using *AIC* over the classical phase pickers. The comparison with the manual picks by the analyst shows that using the present phase picker about 62% of auto picks are within 0.5 sec of analyst picks.

Acknowledgement

The data used in the study has been acquired under research scheme funded by Department of Science and Technology, New Delhi. The help provided by Ms Shipra Malik, Research Scholar is gratefully acknowledged. The suggestions made by the anonymous reviewers are thankfully acknowledged which has helped the authors to enhance the technical quality of the paper.

References

1. Akaike, H. (1973). "Information Theory and an Extension of the Maximum Likelihood Principle", 2nd *International Symposium on Information Theory*, B. Petrov and F. Csaki (Editors), Budapest Akademiai Kiado, 267-281.

2. Akaike, H. (1974). "Markovian Representation of Stochastic Processes and Its Application to the Analysis of Autoregressive Moving Average Processes", Ann. Inst. Stat. Math., **26**, 363-387.

3. Allen, R.V. (1978). "Automatic Earthquake Recognition and Timing from Single Traces", Bull. Seism. Soc. Am., **72**, S225-S242.

4. Anant, S.K. and Dowla, F.U. (1997). "Wavelet Transform Methods for Phase Identification in Three-Component Seismogram", Bull. Seism. Soc. Am., **87**, 1598-1612.

5. Baer, M. and Kradolfer, U. (1987). "An Automatic Phase Picker for Local and Teleseismic Events", Bull. Seism. Soc. Am., **77**, 1437-1445.

6. Chael, E.P. (1997). "An automated Rayleigh-Wave Detection Algorithm", Bull. Seism. Soc. Am., **87**, 157-163.

7. Cichowicz, A. (1993). "An Automatic S-Phase Picker", Bull. Seism. Soc. Am., **83**(1), 180-189.

8. Daubechies, I. (1992). "Ten Lectures on Wavelets", *In CBMS-NSF Regional Conference Series in Applied Mathematics*, **61**, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 357p.

9. Earle, P. and Shearer, P.M. (1994). "Characterization of Global Seismograms Using an Automatic-Picking Algorithm", Bull. Seism. Soc. Am., **84**, 366-376.

10. Earle, P. and Shearer, P.M. (1994). "Characterization of Global Seismograms Using an Automatic-Picking Algorithm", Bull. Seism. Soc. Am., **84**, 366-376.

11. Gendron, P., Ebel, J., and Manolakis, D. (2000). "Rapid Joint Detection and Classification with Wavelet Bases Via Bayes Theorem", Bull. Seism. Soc. Am., **90**, 764-774.

12. Goupillaud, P., Grossmann, A., and Morlet, J. (1984). "Cycle-Octave and Related Transforms in Seismic Signal Analysis", *Geoexploration*, **23**, 85-102.

13. Grossmann, A. and Morlet, J. (1984). "Decomposition of Hardy Functions into Square Integrable Wavelets of Constant Shape", *SIAM J. Math. Anal.*, **15**, 723-736.

14. Leonard, M. (2000). "Comparison of Manual and Automatic Onset Time Picking", Bull. Seism. Soc. Am., **90**, 1384-1390.

15. Leonard, M. and Kennett, B.L.N. (1999). "Multi-Component Autoregressive Techniques for the Analysis of Seismograms", Phys. Earth Planet. Interiors, **113**, 247-264.

16. Maeda, N. (1985). "A Method for Reading and Checking Phase Times in Autoprocessing System of Seismic Wave Data", Zisin-Jishin, **38**, 365-379.

17. Mallat, S. (1989). "A Theory for Multi-Resolution Signal Decomposition: The Wavelet Representation, IEEE Trans", Pattern Anal. Machine Intelligence, **11**, 674-693.

18. Mallat, S. and Hwang, W.L. (1992). "Singularity Detection and Processing with Wavelets", IEEE Trans. Inform. Theory, **38**, 617-643.

19. Morita, Y. and Hamaguchi, H. (1984). "Automatic Detection of Onset Time of Seismic Waves and Its Confidence Interval Using the Autoregressive Model Fitting", Zisin, **37**, 281-293.

20. Pisarenko, V.F., Kushnir, A.F., and Savin, I.V. (1987). "Statistical Adaptive Algorithms for Estimation of Onset Moments of Seismic Phases", *Phys. Earth Planet. Int.*, **47**, 4-10.

21. Rudd, B.O. and Husebye, E.S. (1991) "A New Three-Component Detection and Automatic Single Station Bulletin Production", Draft Paper for AFOSR Grant AFOSR-89-0259.

22. Sharma, M. L., Gupta, S.C., and Wason, H.R. (1990). "An Automatic Phase Picker for Earthquakes", *Proc. International Conference on Automatic Robotics and Computer Vision*, Singapore, 247-250.

23. Sleeman, R. and Van Eck, T. (1999). "Robust Automatic P-Phase Picking: An On-Line Implementation in the Analysis of Broadband Seismogram Recordings", Phys. Earth Planet. Interiors, **113**, 265-275.

24. Takanami, T. and Kitagawa, G. (1988). “A New Efficient Procedure for the Estimation of Onset Times of Seismic Waves”, *Journal of Physics of the Earth*, **36**, 267-290.

25. Tarvainen, M. (1992) “Automatic Seismogram Analysis: Statistical Phase Picking and Locating Methods Using One-Station Three-Component Data”, Bull. Seism. Soc. Am., **82**(2), 860-869.

26. Tibuleac, I.M. and Herrin, E.T. (1999). “An Automatic Method for Determination of Lg Arrival Times Using Wavelet Transforms”, Seism. Res.Lett., **70**, 577-595.

27. Vidale, J.E. (1986). “Complex Polarization Analysis of Particle Motion”, Bull. Seism. Soc. Am., **76**, 1393-1405.

28. Wickerhauser, M.V. (1994). “Adapted Wavelet Analysis from Theory to Software Algorithms”, A.K. Peters Ltd., Wellesley, Massachusetts, 504.

29. Withers, M., Aster, R, Young, C., Beiriger, J., Harris, M., Moore, S., and Trujillo, J. (1998). “A Comparison of Selected Trigger Algorithms for Automated Global Seismic Phase and Event Detection”, Bull. Seism. Soc. Am., **88**, 95-106.

30. Zhang, H., Thurber, C. and Rowe, C. (2003). “Automatic P-wave Arrival Detection and Picking with Multiscale Wavelet Analysis for Single-Component Recordings”, Bull. Seism. Soc. Am., **93**, 1904-1912.