

Comparison of recent S-wave indicating methods

Katarzyna Hubicka^{1,*} and Jakub Sokolowski^{2,**}

¹ Faculty of Pure and Applied Mathematics, Janiszewskiego 14a, Wrocław University of Science and Technology, Poland

² Machinery Systems Division, Na Grobli 15, Wrocław University of Science and Technology, Poland

Abstract. Seismic event consists of surface waves and body waves. Due to the fact that the body waves are faster (P-waves) and more energetic (S-waves) in literature the problem of their analysis is taken more often. The most universal information that is received from the recorded wave is its moment of arrival. When this information is obtained from at least four seismometers in different locations, the epicentre of the particular event can be estimated [1]. Since the recorded body waves may overlap in signal, the problem of wave onset moment is considered more often for faster P-wave than S-wave. This however does not mean that the issue of S-wave arrival time is not taken at all. As the process of manual picking is time-consuming, methods of automatic detection are recommended (these however may be less accurate). In this paper four recently developed methods estimating S-wave arrival are compared: the method operating on empirical mode decomposition and Teager-Kaiser operator [2], the modification of STA/LTA algorithm [3], the method using a nearest neighbour-based approach [4] and the algorithm operating on characteristic of signals' second moments. The methods will be also compared to well-known algorithm based on the autoregressive model [5]. The algorithms will be tested in terms of their S-wave arrival identification accuracy on real data originating from International Research Institutions for Seismology (IRIS) database.

1 Introduction

The indication of seismic event moment of arrival is vital in the processing of seismic signals, e.g. when the problem of event's energy evaluation is considered (which is mostly done by computing the integral of the registered signal from it's moment of arrival till the end). Moreover, the result has to be rescaled in accordance to the distance between the sensor and the event's hypocenter (due to inverse-square law). Naturally when this approach is conducted one has to estimate the moment of wave arrival and the distance (which is commonly solved with the aid of onset moment indicated by at least four different single-axial sensors).

Single excitation indicate various waves which differ in terms of propagation patterns. They can be divided into [6]: body waves, which travel through the Earth interior, and surface waves which propagate along the interface between differing media. As their name suggest the main difference between the body and surface waves is the place of their occurrence. Another difference is their

*e-mail: 221667@student.pwr.edu.pl

**e-mail: jakub.sokolowski@pwr.edu.pl

carrier frequency which for body waves is much higher thus their identification on seismometer data is relatively easy.

The body waves can be qualified as Primary waves (P-waves, pressure waves) and Secondary waves (S-waves, shear waves). As the P-waves are the fastest one (especially faster than the S-wave), sensors register their moment of arrival as first. When the sensors are relatively close to the event epicentre the delay between event's onset moments are small thus the indication of arrivals other than P-wave is difficult. The indication of waves arrival may be considered as a case of finding a structural break detection [7–10]. Recently couple new methods to indicate the S-wave arrival were created. However, the results of each are obtained from different set of data. Thus this is quite hard to tell which of them is the most accurate. It was decided to compare these new methods altogether with one well-known method. All of the methods employs the same real data set.

2 Methodology

2.1 Kirbas-Peaker Algorithm

Kirbas-Peaker algorithm [2] was proposed in 2016. The method is based on Empirical Mode Decomposition (EMD) and Teager-Kaiser energy operator (TKEO).

First part of the algorithm, EMD, is useful for processing natural signals as it doesn't need assumption about its linearity and stationarity [11]. The EMD divides the time series into Intrinsic Mode Functions (IMF) which form complete and nearly orthogonal basis for the signal. Each mode function fulfils two constraints: having only one extreme between zero crossing, and the mean value of their upper and lower envelope must equal zero.

TKEO is a nonlinear operator which is used to estimate the energy of signal. For n -th sample of discrete signal $x[n]$ it is defined as [12]

$$\Psi(x[n]) = x[n-1]^2 - x[n]x[n-2]. \quad (1)$$

In this method the mean TKEO is used. It is described as:

$$\Psi(x[n]) = \log \left(\frac{1}{N} \sum_{m=n-N+3}^n x[m-1]^2 - x[m]x[m-2] \right), \quad (2)$$

where N is the window length. Kirbas-Peaker algorithm has the following steps:

1. The signal is filtered by 10-point moving average and IRR Butterworth band-pass filter [13].
2. The resulting signal is divided into windows of N samples.
3. The EMD algorithm is applied to each window. The IMF with the most visible S-wave is taken.
4. In each window the mean TKEO is calculated.
5. First window with mean TKEO exceeding predefined threshold th_s implies the period with S-wave onset.
6. The sample with maximum TKEO value (in the corresponding window) is chosen as actual S-wave arrival moment.

2.2 Rawles-Thurber Algorithm

The method [4] was presented in 2015 and uses a nearest neighbours-based approach of Nikolov [14]. A method does not need to use parameters estimated from data, but uses data itself to build a model. The main issue is to aggregate the set of reference data with P-waves and S-waves picks which is compared to the analysed seismic signals. The method may somehow be considered as a clusterisation problem [15, 16]. The Rawles-Thurber algorithm uses 3 traces of a signal. To indicate the S-wave, firstly the P-wave onset is pointed. The method consists of following steps:

1. In the first step, two reference data sets are built: for P-waves from vertical trace and for S-waves from horizontal traces. Each reference set consists of positive \mathcal{R}_+ and negative examples \mathcal{R}_- , where positive examples are N -length windows surrounding the P-wave/S-wave onsets and negative are windows with length of N containing for example noise or P-wave/S-wave coda.
2. The next move is windowing and data preprocessing by applying the bandpass filter. All of signal's traces are divided into moving-windows of the same length as reference windows. Each window is scaled as follow: the absolute amplitude of window is divided by the mean of its first half.
3. For each window of N -length, the similarity metric $d(r, s)$ to reference examples is computed, where $d(r, s)$ is the sum of squares of the Euclidean distance between the sample of reference signal r_i and observed signal s_i :

$$d(r, s) = \sum_{i=1}^N (r_i - s_i)^2. \quad (3)$$

4. Subsequently the score function $R(s)$ is determined for each window as the ratio of the negative to positive distances of the particular trace:

$$R(s) = \frac{\sum_{r \in \mathcal{R}_-} d(r, s)}{\sum_{r \in \mathcal{R}_+} d(r, s)}. \quad (4)$$

As for S-wave identification two horizontal traces are used, to appoint the S-wave score function $R_S(s)$, two horizontal scores $R_{H1}(s)$ and $R_{H2}(s)$ are multiplied:

$$R_S(s) = R_{H1}(s) \cdot R_{H2}(s). \quad (5)$$

5. The centre of the window with the biggest value of score $R_P(s)$ for a vertical trace indicates the moment of P-wave onsets, and the biggest $R_S(s)$ – the moment of S-wave onset. In addition, the $R_P(s)$ and $R_S(s)$ scores must exceed the thresholds and the difference between P-wave and S-wave picks has to be located in user-supplied range.

2.3 Tasic-Runovc Algorithm

Tasic-Runovc approach [3] is a modification of STA/LTA algorithm and needs three-component seismic station monitors with velocities in east-west (E), north-south (N) and vertical (V) directions. The STA/LTA algorithm is based on the energy analysis by the ratio of the short-term average (STA) energy to long-term average (LTA) energy level from the same seismogram [17]. In T-R method the modification of STA/LTA is applied through the transformation of signal $f(x_i)$. The nominator still gives the information about the sensitivity for local changes in the amplitude as the STA segment, however the denominator informs about the amplitude variation from the sample i to the end of the signal. The algorithm is as following:

1. Each trace of signal is saved as vectors \mathbf{Y}_E , \mathbf{Y}_N , \mathbf{Y}_V :

$$\mathbf{Y}_E = (e_1, e_2, \dots, e_k) \quad (6)$$

$$\mathbf{Y}_N = (n_1, n_2, \dots, n_k) \quad (7)$$

$$\mathbf{Y}_V = (v_1, v_2, \dots, v_k). \quad (8)$$

The length of the signal is k and an i -th sample is in the range $1 < i_p < i_s < k$, where i_p indicates the moment of P-wave and i_s – moment of S-wave arrival.

2. To estimate the relative total energy, the vector \mathbf{O} consisting of the sum of squares of the components of 3 traces is used:

$$o_i = e_i^2 + n_i^2 + v_i^2. \quad (9)$$

3. To appoint the S-wave pick, the $f(x_i)$ function is made:

$$f(x_i) = \frac{\frac{1}{l} \sum_{j=i}^{i+l-1} |x_j|}{\frac{1}{k-i+1} \sum_{j=i}^{k-l} |x_j|} \quad \text{for } i = 1, \dots, k-l. \quad (10)$$

where l is the window's length.

4. The arguments of the function above are components of vectors \mathbf{Y}_E , \mathbf{Y}_N and \mathbf{O} . The next step is to make a characteristic vector \mathbf{C} with components:

$$c_i = f(e_i)f(n_i)f(o_i) \quad \text{for } i = 1, \dots, n-k. \quad (11)$$

5. If the sample of the characteristic function exceeds the user-supplied threshold as the first one, sample i_s is taken as the moment of S-wave arrival.

2.4 Sokolowski-Wylomanska-Zimroz algorithm

The algorithm is expansion of P-wave indicating algorithm [18]. The procedure is based on the characteristic of signals' empirical second moment. It is known that when X_i is consisted of i.i.d. random variables with mean μ and variance σ^2 they fulfill [19]:

$$E \sum_{i=1}^n X_i^2 = \sum_{i=1}^n EX_i^2 = \sum_{i=1}^n (E(X_i - EX_i)^2 + (EX_i)^2) = n(\sigma^2 + \mu^2). \quad (12)$$

This clearly induces that when the cumulative sum of squared signal $\sum_{i=1}^n x_i^2$ is considered, one has to be aware that if x_i is i.i.d. the regression line can be fitted. It is common to assume that the seismic noise fulfils the assumption of being i.i.d. However, when signal's part after the P- or S-wave onset moment is considered the signal is clearly not i.i.d.

It was noticed that the statistics after the P-wave onset moment (and before the S-wave moment) becomes concave. It was also noted that when applying the logarithm to the statistics the resulting function is in general concave from the beginning till the onset moment and from the onset moment till the end (if on whole its course it is concave it means that no impulse was registered).

In the paper it was also observed that the logarithmic statistic when considered for the part of signal which includes the seismic wave is well approximated with exponential function. This led the authors to the following algorithm:

1. Determine the moment of P-wave arrival (p) for input signal s_t of length n . Take $x_t = s_{t+p-1}$ for $t \in \{1, 2, \dots, n - p + 1\}$.
2. Compute $L_t = \log(\sum_{i=1}^t x_i^2)$ for $t \in \{1, 2, \dots, n - p + 1\}$.
3. For $i = \{3, 4, \dots, n - p + 2\}$ let $L_t^1 = L_t$ for $t = 1, 2, \dots, i$, $L_t^2 = L_{t+i}$ for $t = 1, 2, \dots, n + 1 - p - i$. Fit $e^{a_1 x + b_1} + c_1$ to L_T^1 and $e^{a_2 x + b_2} + c_2$ to L_T^2 sum their squared errors (SSE). Find i which minimises SSE.
4. S-wave arrival is indicated for $i + p$.

2.5 Takanami-Kitagawa Algorithm

Algorithm of Takanami-Kitagawa [5] was proposed in 1988 and uses the AR model to analyse seismic signal. Despite the fact of seismic signals being non-stationary, they can be divided into two subseries and each of them can be approximated by the AR model.

The method to obtain the moment of S-wave arrival is as follows:

1. The signal is divided into two subseries before and after the unknown moment of S-wave arrival:

$$x_n = \sum_{m=1}^{M(1)} a_m^1 x_{n-m} + \varepsilon_n^1, \quad (1 \leq n < s_1) \quad (13)$$

$$x_n = \sum_{m=1}^{M(2)} a_m^2 x_{n-m} + \varepsilon_n^2, \quad (s_1 \leq n \leq N) \quad (14)$$

where ε_n^i is a Gaussian noise with mean zero and variance σ_i^2 , a_m^i is the autoregressive coefficient, $M(i)$ is the order of the i -th model, s_1 is the unknown arrival time and N is the length of series. The [13] is considered as the "noise" part of signal and [14] as the "earthquake" part of signal, where S-wave appears. The length of "noise" window increases together with the s_1 , while the "earthquake" window's length decreases.

2. To measure the quality of estimated model, the Akaike Information Criterion (AIC) is used:

$$AIC = -2(\log L_{max}) + 2(M_P), \quad (15)$$

where L_{max} is the maximised value of the likelihood function for the model and M_P is the number of independently estimated parameters.

For each potential time of S-wave arrival s_1 AIC is calculated for the first part of signal AIC_i^1 and second part AIC_i^2 .

3. Then the sum of two AIC is taken:

$$AIC_i = AIC_i^1 + AIC_i^2. \quad (16)$$

4. The sample i_{min} with minimum value of AIC_i is searched.
5. The sample $i_{min} + 1$ responds the S-wave pick.

3 Applications

The above-mentioned algorithms are tested on two signal sets from International Research Institutions for Seismology (IRIS) database with reference picks. Each of them contains the events from West Coast of South America.

The first set consists of 40 signals registered in Brazilian seismic station SAML from different earthquakes in 2003-2017, while the second one contains 100 series from LVC station events registered in 2002-2017. Both data sets are sampled at 20 Hz, the magnitudes of events are above 5.5 and the distances between stations and earthquakes are between 7° and 20°.

If the method does not provide otherwise, the input seismic signals are not denoised as such step can make onset moment indication easier [20, 21]

In K-P, R-T and T-R algorithms, the setting of few parameters is essential and most of them is determined by the minimum mean squared or absolute errors between reference and auto S-wave picks on the mesh with different values of parameters.

In Kirbas-Peaker approach the optimal length of window N , the number of IMF and threshold th_s are chosen from the mesh of mean absolute errors between references and auto picks for particular parameters.

The least error is reached for north-south trace and parameters which are showed in table 1.

Table 1. Parameters used in Kirbas-Peaker algorithm

	1st data set	2nd data set
length of window N (in samples)	72	50
the number of IMF	1	5
S-wave threshold th_s	0.7207	0.05

In Rawles-Thurber method the thresholds th_P and th_S are appointed from the mesh of mean absolute errors between picks and the length of window was taken as in Rawles-Thurber paper [4]. The range of distance between P-wave and S-wave picks is determined on the base of reference picks in data sets. A different reference set from other observed signals (39 for 1st set and 99 for 2nd set) is created for each signal. The parameters used in this method are presented in table 2.

Table 2. Parameters used in Rawles-Thurber algorithm

	1st data set	2nd data set
length of window N (in samples)	67	67
P-wave threshold th_P	0.0001	0.03
S-wave threshold th_S	0.0001	0.0001
minimum distance between P and S-wave SP_{min} (in samples)	200	900
maximum distance between P and S-wave SP_{max} (in samples)	3100	3100

To chose the best parameters for Tasic-Runovc algorithm the mesh with mean squared errors is applied. Optimal threshold th_s and the length of window l are presented in Table 3.

Table 3. Parameters used in Tasic-Runovc algorithm

	1st data set	2nd data set
length of window l (in samples)	50	100
S-wave threshold th_s	2	2

In Takamami-Kitagawa algorithm, the AR models are matched to data starting from the sample $i_p + 200$, where i_p is the moment of P-wave onset and ending on the end of the signal. In the first step of fitting autoregressive models the "noise" part consists of the samples i_n :

$$i_n \in [i_1, \dots, i_{199}] ,$$

where $i_1 = i_p + 200$ and the "earthquake" part from the samples i_e :

$$i_e \in [i_{200}, \dots, i_N] ,$$

where N – the length of signal. The first and second part was moving on the right in the each consecutive step (the last element in "noise" part and the first one in "earthquake" part are changing) and the best models of AR are fitted to each part. Maximum order of AR model to first part of signal is 10 and to second part – 15 in both data sets.

3.1 Application to the particular signal

In the figures below the application of all five methods to the one signal from the first analysed set of data is shown. The upper panels of all figures present the signal with indicated reference S-wave onset.

Figure 1 presents the implementation of Kirbas-Pekaer to the exemplary signal. The lower panel illustrates the mean TKEO value computed for each window of signal. The horizontal solid line represents the S-wave threshold th_S and the distinguished box indicates the window exceeding the value of threshold as the first one. The sample with maximum value of mean TKEO in that window responds the S-wave pick.

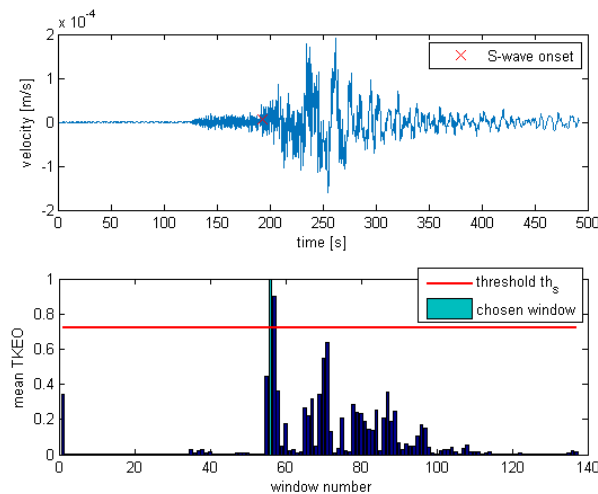


Figure 1. Application of Kaiser-Pekaer method to the particular signal.

The implementation of Thurber-Rawles algorithm was presented in the figure 2. The lower panel illustrates the score function R_S , which was computed for each window of the signal. The dot represents the minimum value of R_S and thereby the moment of S-wave onset.

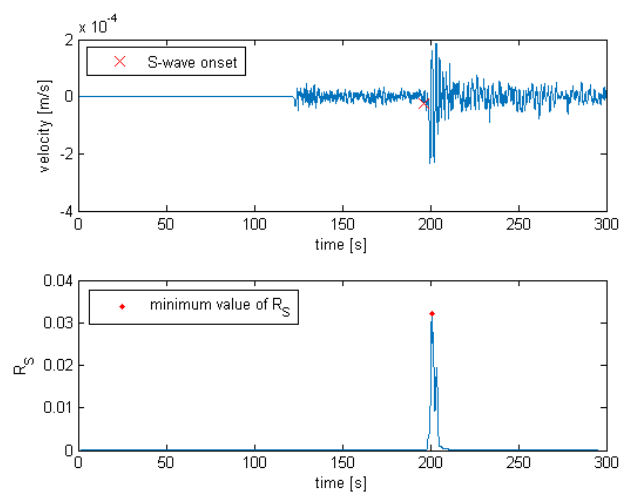


Figure 2. Application of Rawles-Thurber method to the particular signal.

In Figure 3 the application of Tasic-Runovc method can be seen altogether with the statistic which determines the results. It can be seen that with the S-wave onset moment the statistics considerably rises which confirms the algorithm idea.

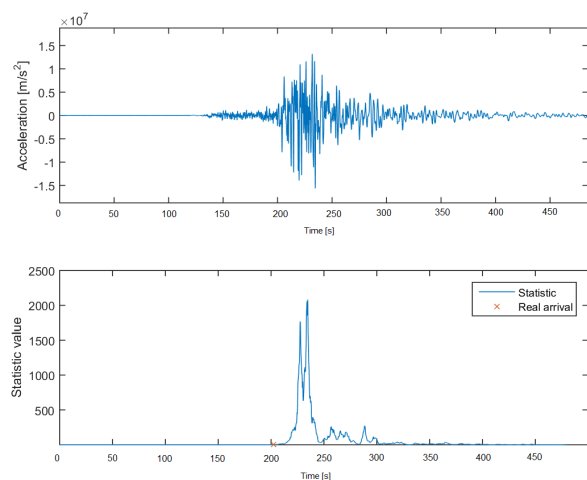


Figure 3. Application of Tasic-Runovc method to the particular signal.

In Figure 4 the exemplary signal is presented in upper figure and its L statistic. It can be seen that this statistic consists of 3 consequent concave functions which have their limits in moments of P- and S-wave arrivals. Thus when the algorithm finds these limits it finds the moment of wave arrival.

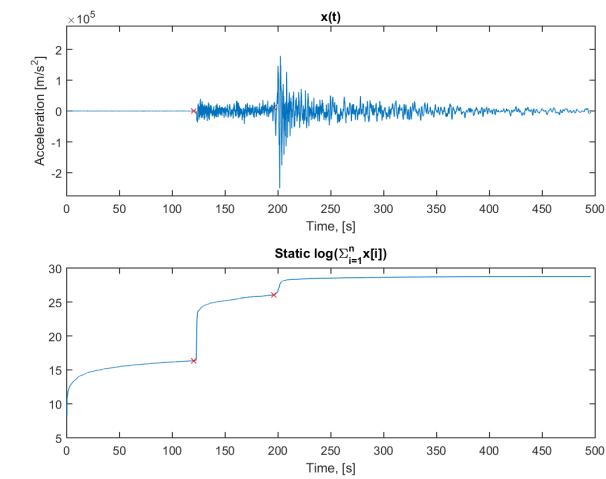


Figure 4. Application of Sokolowski-Wylomanska-Zimroz method to the particular signal.

Figure 5 presents the application of Takanami-Kitagawa algorithm. The lower box pictures the sums of AIC values from "noise" and "earthquake" auto-regressive models. The cross indicates the sample i_{min} with minimum sum of AIC and the sample $i_{min} + 1$ represents the S-wave onset.

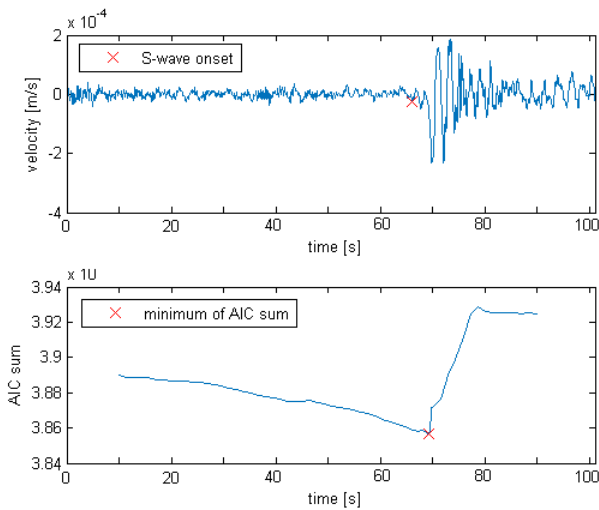


Figure 5. Application of Takanami-Kitagawa method to the particular signal.

3.2 Algorithms’ results

In this section all methods are compared. Their results are summarised by such statistics as number of events whose evaluated S-wave arrival are closer than 10 seconds, mean of absolute differences between the estimation and real moment, its standard deviation, and standard deviation of differences. In Table 4 the results of algorithms tested on the first set (which consists of 40 events) are presented. The bold numbers symbolise the algorithm with best performance in specific statistics.

Table 4. Algorithms’ results for the first set

	K-P	R-T	T-R	S-W-Z	T-K
$\#i : s_i - \hat{s}_i \leq 10 \text{ s}$	7	7	15	14	15
$\frac{1}{n} \sum_{i=1}^n s_i - \hat{s}_i \text{ [s]}$	46.31	42.59	36.30	30.37	34.00
$\text{std}(s_i - \hat{s}_i) \text{ [s]}$	68.90	44.37	51.20	32.90	33.23
$\text{std}(s_i - \hat{s}_i) \text{ [s]}$	50.65	32.97	36.25	29.71	29.76

In Table 4 it can be seen that the Takanami-Kitagawa and Tasic-Runovc algorithms has the most number of indications closer than 10 seconds (15 out of 40 which is 37.5%). However other statistics (mean value of errors, and their standard deviations) clearly show the superiority of Sokolowski-Wylomanska-Zimroz method. The Kirbas-Peaker algorithm seems to be the least accurate.

Table 5. Algorithms’ results for the second set

	K-P	R-T	T-R	S-W-Z	T-K
$\#i : s_i - \hat{s}_i \leq 10 \text{ s}$	14	10	1	26	31
$\frac{1}{n} \sum_{i=1}^n s_i - \hat{s}_i \text{ [s]}$	41.97	58.51	31.18	17.71	27.57
$\text{std}(s_i - \hat{s}_i) \text{ [s]}$	36.57	41.58	20.70	16.07	28.31
$\text{std}(s_i - \hat{s}_i) \text{ [s]}$	55.05	67.49	31.04	22.93	36.46

In Table 5 the results of methods applied to the second set (100 signals) are presented. This time the results are almost the same as in Table 4. The biggest change is the small number of accurate picks of Tasic-Runovc algorithm (Previously 37.5% of indications and now 1%). Also the results of Sokolowski-Wylomanska-Zimroz algorithm seem to be more accurate than for the previous set.

4 Conclusions

In this paper the problem of S-wave arrival has been presented. Within this four relatively new algorithms indicating moment of S-wave arrival were compared altogether with one earlier and more commonly used method. The algorithms were tested on real data which was gained from the International Research Institutions for Seismology database.

The results for the first data-set provides that the Takanami-Kitagawa and Tasic-Runovc algorithms have the most number of accurate picks for, however they are not the most accurate algorithms, as the Sokolowski-Wylomanska-Zimroz indicated picks that were on average most accurate with the results having the lowest variance (and in this terms the Takanami-Kitagawa is the second best). The results also show that the Kirbas-Peaker is the least one and the Rawles-Thurber is last but one in terms of results accuracy.

For the second data-set only Takanami-Kitagawa algorithm has the biggest number of accurate picks (Sokolowski-Wylomanska-Zimroz has the second best number and this time Tasic-Runovc method has only one accurate pick). In terms of accuracy once again S-W-Z algorithm is the most

accurate one, Takanami-Kitagawa has the second best average picks, however greater standard deviation than Tasic-Runovc. This time the Kirbas-Peaker algorithm has the fourth best results and Rawles-Thurber is the least accurate one.

The aggregated results provide that the Sokolowski-Wylomanska-Zimroz algorithm despite being second the best in terms of number of accurate picks has the most accurate results on average and the least average variance of results thus it is recommended to use.

In the paper no attention was paid on the computational complexity of algorithms which may be a proposition for future work.

This work was financed by the Grant no. 0401/0128/17.

References

- [1] S. J. Gibowicz, A. Kijko, *An Introduction to Mining Seismology* (ACADEMIC PRESS, INC., San Diego, 1994)
- [2] I. Kirbas, M. Peker, *Neural Computing and Applications*, 1-11 (2016)
- [3] I. Tasic, F. Runovc, *Acta Geotechnica Slovenica*, **6**, 46-55 (2009)
- [4] C. Rawles, C. Thurber, *Geophysical Journal International*, **202.2**, 1164-1179 (2015)
- [5] T. Takanami, G. Kitagawa, *Journal of Physics of the Earth*, **36.6**, 267-290 (1988)
- [6] P.M. Shearer, *Introduction to seismology* (Cambridge University Press, Cambridge, 2009)
- [7] D. Kucharczyk, A. Wylomanska, R. Zimroz, *Physica A: Statistical Mechanics and its Applications*, **471**, 499-511 (2017)
- [8] R. Hossa, R. Makowski, R. Zimroz, *International Journal of Rock Mechanics and Mining Sciences*, **91**, 29-39 (2017)
- [9] D. Kucharczyk, A. Wylomanska, J. Obuchowski, R. Zimroz, M. Madziarz, *Shock and Vibration*, **2016**, (2016)
- [10] R. Zimroz, M. Madziarz, G. Zak, A. Wylomanska, J. Obuchowski, *Journal of Vibroengineering*, **17.6**, 3111-3121 (2015)
- [11] A. Erturk, M.K. Gullu, S. Erturk, *IEEE Trans Geosci Remote Sens*, **51**, 2787-2798 (2013)
- [12] J.F. Kaiser, *ICASSP-90*, **1**, 381-384 (1990)
- [13] S. Butterworth, *Wireless Engineer*, **7.6**, 536-541 (1930)
- [14] S. Nikolov, M.Sc. thesis (2012)
- [15] J. Sokolowski, J. Obuchowski, M. Madziarz, A. Wylomanska, R. Zimroz, *Journal of Vibroengineering*, **18.3**, 1654-1667 (2016)
- [16] M. Polak, J. Obuchowski, M. Madziarz, A. Wylomańska, R. Zimroz, *Journal of Vibroengineering*, **18.1**, 267-275 (2016)
- [17] C.Y. Bai, B.L.N. Kennett, *Journal of Seismology*, **5**, 217-231 (2001)
- [18] J. Sokolowski, J. Obuchowski, R. Zimroz, A. Wylomanska, E. Koziarz, *Shock and Vibration*, **2016**, (2016)
- [19] J. Gajda, G. Sikora, A. Wylomanska, *Acta Physica Polonica B*, **44.5**, 1015-1035 (2013)
- [20] F. Chaari, J. Leskow, A. Napolitane, R. Zimroz, A. Wylomanska, *Cyclostationarity: Theory and Methods III*, (Springer International Publishing, Cham, 2017), 51-68
- [21] A Wylomańska, R Zimroz, J Janczura, J Obuchowski, *IEEE Transactions on Industrial Electronics*, **63.9**, 5612-5621 (2016)