

A novel technique for automatic seismic data processing using both integral and local feature of seismograms

Ping Jin · Chengliu Zhang · Xufeng Shen ·
Hongchun Wang · Changzhou Pan ·
Na Lu · Xiong Xu

Received: 16 January 2014 / Accepted: 26 March 2014 / Published online: 24 May 2014
© The Seismological Society of China, Institute of Geophysics, China Earthquake Administration and Springer-Verlag Berlin Heidelberg 2014

Abstract A novel technique for automatic seismic data processing using both integral and local feature of seismograms was presented in this paper. Here, the term integral feature of seismograms refers to feature which may depict the shape of the whole seismograms. However, unlike some previous efforts which completely abandon the DIAL approach, i.e., signal detection, phase identification, association, and event localization, and seek to use envelope cross-correlation to detect seismic events directly, our technique keeps following the DIAL approach, but in addition to detect signals corresponding to individual seismic phases, it also detects continuous wave-trains and explores their feature for phase-type identification and signal association. More concrete ideas about how to define wave-trains and combine them with various detections, as well as how to measure and utilize their feature in the seismic data processing were expatiated in the paper. This approach has been applied to the routine data processing by us for years, and test results for a 16 days' period using data from the Xinjiang seismic station network were presented. The automatic processing results have fairly low false and missed event rate simultaneously, showing that the new technique has good application prospects for improvement of the automatic seismic data processing.

Keywords Seismic · Automatic data processing · Feature of seismograms

1 Introduction

Automatic of seismic event detection and location can significantly improve efficiency of routine seismic network processing. This is especially important for seismic monitoring of CTBT verification. As magnitude of a 1 kt fully decoupled explosion may be as small as m_b 2.0+ (Murphy 1996), events routinely need to be timely detected and located may amount to such a large number that are hard to be manually picked up. For example, it is estimated that annually there are over 1 million earthquakes with magnitude above m_b 2.0 occurring in the world. In addition to this, nowadays a station-network operator may find that it is relatively easy to install more seismic stations to improve the network's coverage. However, due to the deficiency of reliable automatic data processing system, often it is the case that the increasing amount of data is hard to be fully utilized for it is too large to be analyzed by a limited number of human analysts.

Although it may not be always necessary by some scholars (e.g., Ryzhikov et al. 1996; Withers et al. 1998), most of the practical automatic processing systems used by seismic station-network operators follow the procedure of signal detection, phase-type identification, signal association, and location, which is abbreviated as the DIAL in this paper. By this approach, possible seismic signals are automatically picked up by some kind of detectors [e.g., STA/LTA (Allen 1978; 1982; Evans and Allen 1983; Baer and Kradofer 1987; Earle and Shearer 1994; Withers et al. 1998); PMCC (Cansi 1995); artificial neural network (Dai and Macbeth 1995, 1997; Zhao and Takano 1999); F-detector (Arrowsmith et al. 2009; Selby 2008, 2011)]. As a detection is triggered, its onset time is further estimated using some more sophisticated procedures [e.g., AR-AIC (GSE/JAPAN/40 1992; Sleeman and van Eck 1999);

P. Jin (✉) · C. Zhang · X. Shen · H. Wang ·
C. Pan · N. Lu · X. Xu
Northwest Institute of Nuclear Technology, Xi'an,
Shaanxi 710024, China
e-mail: jinping668@sohu.com

CUMSUM (Inclán and Tiao 1994; Der and Shumway 1999); HOS (Saragiotis et al. 2002; Küperkoch et al. 2010b); Wavelet Transform (Tibuleac et al. 2003; Zhang et al. 2003); MPX (Stefano et al. 2006); Tpd (Hildyard et al. 2008)], as well as its characterization parameters, including amplitude, dominant frequency, backazimuth, slowness, polarization pattern, etc., are measured. Based on these characterization parameters, specific algorithm such as artificial neural network usually is used to identify the phase type of the triggered detections (e.g., Roberts et al. 1989; Suteau-Henson 1991; Tong 1995; Wang and Teng 1995, 1997; Tong and Kennett 1996; Anant and Dowla 1997; Bai and Kennett 2000, 2001; Wang 2002). Search algorithms are then used to associate detections possibly from same events together and hypocenter of these events are then located (e.g., Anonymous (IDC-5.2.1). 1999; Küperkoch 2010).

The seismic data processing system that is currently used at the Vienna International Data Center is a typical example of applying DIAL approach. This system was designed to automatically detect and locate seismic events recorded by the seismic component of the International Monitoring System before they can be presented to human analysts for review. According to the IDC Documentation IDC-5.2.1, the processing of seismic data at IDC is divided into two stages, i.e., the station processing stage and the network processing stage. During the station processing stage, a running STA/LTA detector is applied to detect potential seismic signals after raw waveform data from individual stations have passed through sequential measures for data quality checking and signal-to-noise ratio improvement. Then, characteristics include arrival time, amplitude, period, azimuth, slowness, and polarization feature for declared detections are extracted. Based on these characteristics, two different methods, artificial neural network for three-component seismic stations and simple rule for seismic arrays, are used to categorize the detected signals as one of the four wave type, i.e., noise, teleseismic P, regional P or regional S. Then, detections from the same events are then grouped into station association groups. For each group, the first arrival is called as the generator, and its signal type is a local P, regional P, or teleseismic phases, and subsequent phases in each group are identified based on their compatibility with the generator, as determined by the arrival times and compatibility tests, which details are expatiated in IDC-5.2.1. Finally the phases identified during the station processing is feed to the network processing stage to associate with events using a grid-based event search procedure and to locate and estimate the magnitude of the events using arrival time and amplitude of the associated signals. Detail explanation about the grid-based event search procedure can also be found in IDC-5.2.1.

Though the automatic processing of seismic data at IDC is very complex, its performance hardly can be said satisfying by producing too many false events, while missing a fair portion of real (e.g., Procopio et al. 2009; Pearce 2009; Le Bras et al. 2011; Draelos et al. 2012), which has already attracted extensive attention among the CTBT monitoring community and some new techniques have been proposed to deal with this problem (e.g., Le Bras et al. 2011; Draelos 2012; Arora et al. 2012; Moore et al. 2012). However, most of the new techniques (e.g., Draelos 2012; Arora 2012) are still based on the existed signal characterization parameters and performance improvement by these techniques remains to be validated.

In fact, the poor performance of conventional automatic processing method is largely because the existing parameters fail to sufficiently characterize the signals' feature. The characterization parameters exploited by conventional automatic processing systems (as that of IDC) are nearly all extracted from short waveform segments of seismograms. Though they may reflect signals' transient behaviors well, they can neither reflect the whole shape of seismograms, nor fully reflect relationships between different seismic detections, both of which are important for seismogram interpretation.

Here, in this paper, we present a novel technique for automatic seismic data processing which utilizes both integral and local feature of seismograms. Here, the term integral feature of seismograms largely refer to feature about the envelope shape of seismograms. However, unlike some previous efforts which completely abandon the DIAL approach and seek to use envelope cross-correlation to detect seismic events directly (e.g., Ryzhikov et al. 1996; Withers et al. 1998; Moore et al. 2012), we keep to follow the DIAL approach, but include integral feature of the whole seismograms for phase identification and association. Results of preliminary application of this technique show that the inclusion of integral feature in data processing can greatly reduce false and missed events simultaneously and have the potential to realize accurate automatic seismic data processing.

2 Definition and extraction of signals' integral feature

Integral feature about the shape of the whole seismograms play important roles in routine analyses by human analysts. Usually an experienced analyst can tell whether a seismogram is local, regional or teleseismic simply by glimpsing at it, picking up major phases such as P, S, and Lg by their position in the seismogram. However, because most feature utilized by human analysts are not clearly defined and parameterized, it is difficult to directly employ these feature for automatic seismic data processing.

Nevertheless, it is beneficial to think about how human analysts pick up and identify seismic signals. Unlike

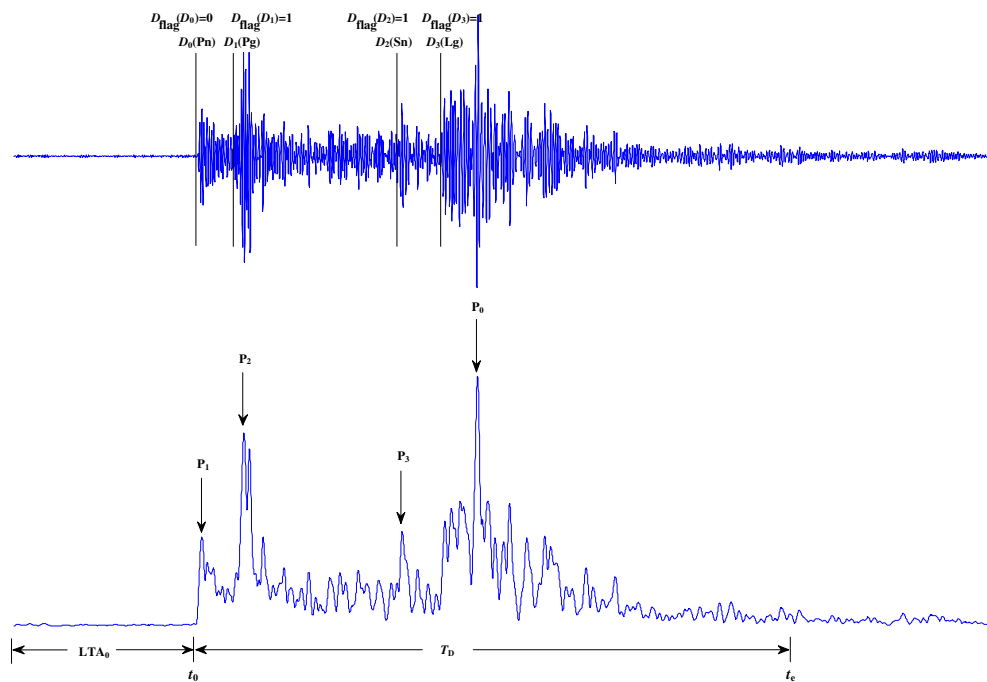


Fig. 1 Illustration of the concept of wave-train and its integral feature. The *top panel* is the waveform of a regional seismogram of epicentral distance about 3.4°. The practical time separation between D_0 and D_3 , which actually corresponding to Pn and Lg, respectively, in this case, is about 50 s. The waveform has passed a 1.5–2.5 Hz bandpass filter to make major regional phases, e.g., Pn, Pg, Sn, and Lg, clear. The bottom panel is its envelope represented by the STA curve

conventional automatic processing systems, human analysts usually pick the whole wave-train of a seismogram first, then determine its epicentral distance type by its envelope shape and frequency characteristics, and finally pick individual phases by relative amplitude variations and identify their phase types by their relative position in the seismogram along with their polarization and frequency feature.

Enlightened by this process, here, we propose a strategy to parameterize signals' integral feature. The basic idea of this strategy is illustrated in Fig. 1. This strategy uses a modified STA/LTA detector (see Appendix for detail) to detect individual signal detections and extracts local feature such as amplitudes, dominant frequency, and polarization characteristics of them as well. However, in the process of picking up individual signals, we simultaneously identify continuous wave-trains and associate them with various detections simultaneously. To do so, we first define that a wave-train is a segment of waveform that starts from its onset time to the time when STA/LTA_0 drops back to be less than a given threshold ϵ (it is set to be 1.1 in practice) where LTA_0 is the LTA value just before the first triggered detection of it, i.e.

$$LTA_0 = \frac{1}{L} \sum_{i=t_0-L}^{t_0} |x(i)|, \tag{1}$$

where L is the length of LTA window. For convenience, here we denote the onset time of the first detection as t_0 ,

and the time when STA/LTA_0 drops back to below ϵ as t_e , and the time difference

$$T_D = t_e - t_0 \tag{2}$$

as the wave-train's duration time. Second, we categorize all declared detections into the beginning type and the following type, i.e., detections at the beginning of a continuous wave-train (e.g., D_0 in Fig. 1) and detections during its duration (e.g., D_1 , D_2 , and D_3 in Fig. 1). In practice, we use a parameter D_{flag} , which may be called as triggered state, to flag these two kinds of detections, i.e., $D_{flag} = 0$ corresponds to beginning detections and $D_{flag} = 1$ for following ones. Third, we extract characterization parameters, including amplitude, signal-to-noise ratio, and time delay of envelope peaks of wave-trains and assign them to various detections. For example in Fig. 1, assume that a wave-

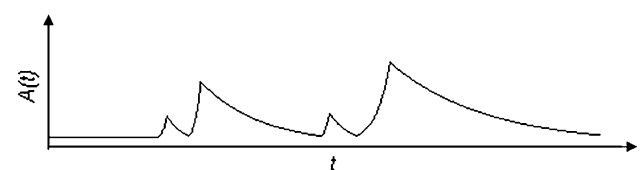


Fig. 2 Inferred envelope shape of the seismogram in Fig. 1 using arrival times and the integral characterization parameters defined in the paper

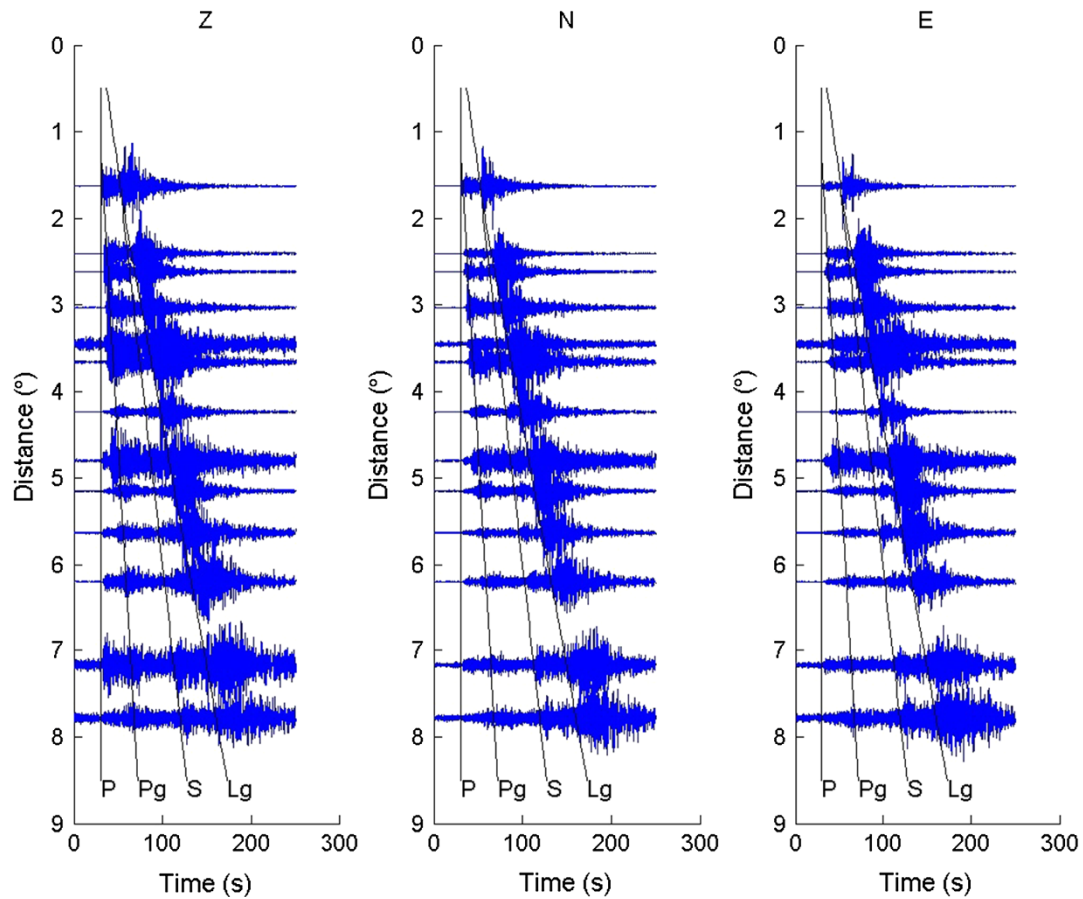


Fig. 3 Illustration of typical three-component seismograms recorded at regional distances. The waveforms are from an earthquake of M_L 4.2 occurring in Southwest Xinjiang, China on 2010/06/02. The waveform columns from left to right are 0.5–5 Hz filtered waveforms on vertical, south-to-north and west-to-east component, respectively



Fig. 4 DRSN data processing flow

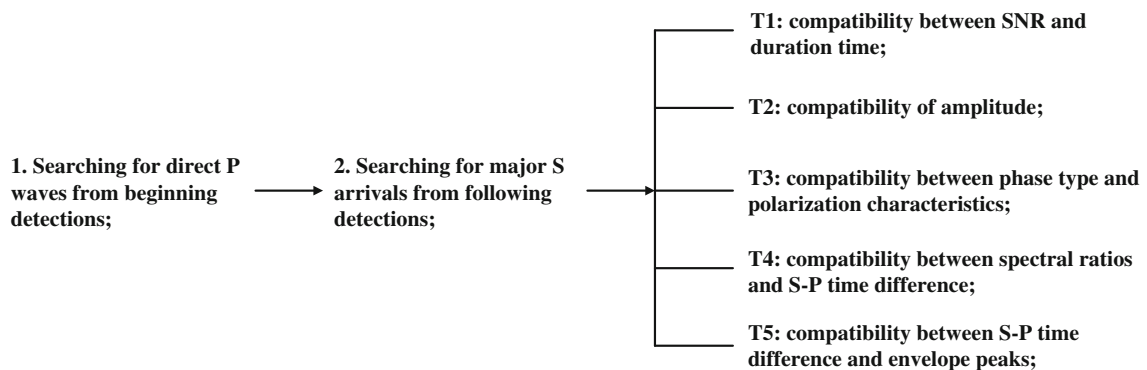


Fig. 5 Illustration of the method to search for regional P/S pairs by integral and local feature together

train starts at t_0 and ends at t_e , and that there are $n + 1$ triggered detections, i.e.,

D_0, D_1, \dots, D_n between this time period, where D_0 (with arrival time t_0) is the beginning detection and D_i (with arrival time t_i where $i = 1, \dots, n$) is the following detections. Denote the envelope peak of the whole seismogram as P_0 (with arrival time t_{P_0}) and the envelope peak of the waveform segment between detections D_{i-1} and D_i as P_i (with arrival time t_{P_i} , $i = 1, \dots, n$), then characterization parameters for P_0 are

$$A_{P_0} = \max_{t_0 < t < t_e} STA(t) = STA(t_{P_0}),$$

$$SNR_{P_0} = A_{P_0}/LTA_0, \quad T_{P_0} = t_{P_0} - t_0, \quad (3)$$

while characterization parameters for P_1 are

$$A_{P_1} = \max_{t_0 < t < t_1} STA(t) = STA(t_{P_1}),$$

$$SNR_{P_1} = A_{P_1}/LTA_0, \quad T_{P_1} = t_{P_1} - t_0 \quad (4)$$

and so on. With these parameters, a wave-train’s general envelope shape can be reflected. For example, for the seismogram in Fig. 1, a general shape of its envelope can be inferred to be similar to that in Fig. 2 by assuming that envelope amplitudes rise or decay exponentially. In practice, these parameters, as well as wave-trains’ duration time can be estimated within a series of frequency bands.

Integral feature defined as above can be easily extracted in the process of signal detection and can be easily embedded into conventional DIAL approach to improve its performance as well. First, each separate continuous wave-train can be represented by its beginning detection. And in cases of local or regional seismograms, beginning detections normally should be the first P arrivals (e.g., Pg or Pn). It is impossible that a following detection is the first P arrival of an event except for rare situations of overlapping signals of multiple events. For this reason, beginning detections naturally are good candidates of direct P phases. Second, following detections during the duration of a wave-train can be readily associated to respective beginning detections, considering their compatibility in amplitude, frequency, and azimuth. Further more, with characterization parameters of envelope peaks, we may determine a following detection’s position in its related wave-train, which is of importance for phase-type identification. One simple way to characterize a detection’s relative position within the seismogram is to use the ratio

$$r = \frac{t - t_0}{T_{P_0}}, \quad (5)$$

where t is the detection’s arrival time, t_0 is the onset time of respective beginning detection, and T_{P_0} the time delay of the envelope peak of the whole seismogram. In case of local or regional seismograms, the arrival time of regional S phases (e.g., Sn and Lg) normally should be a little

before the envelope peak of the whole seismogram (e.g., Fig. 3), especially within low-frequency band such as 0.5–1.5 Hz and on horizontal components. In contrast, detections of secondary P phases, e.g., Pg of regional seismograms, normally are far before the envelope peak of the whole seismograms.

3 Application of the seismograms’ integral feature in seismic data processing

Based on the strategy described above, we have developed a prototype software, which is called as DRSN (i.e., Detector for Regional Seismic Networks) by us, for automatic seismic data processing for regional seismic station networks. As illustrated in Fig. 4, the processing flow of this prototype software can be split into 5 stages. The task of the first stage is to detect possible signals and extract their feature, both local and integral, simultaneously. The task of the second stage is to categorize detections into P waves, S waves and noise types using neural networks similar to that of Wang (2002). The task of the third stage is to search for P/S pairs of local or regional events, and the

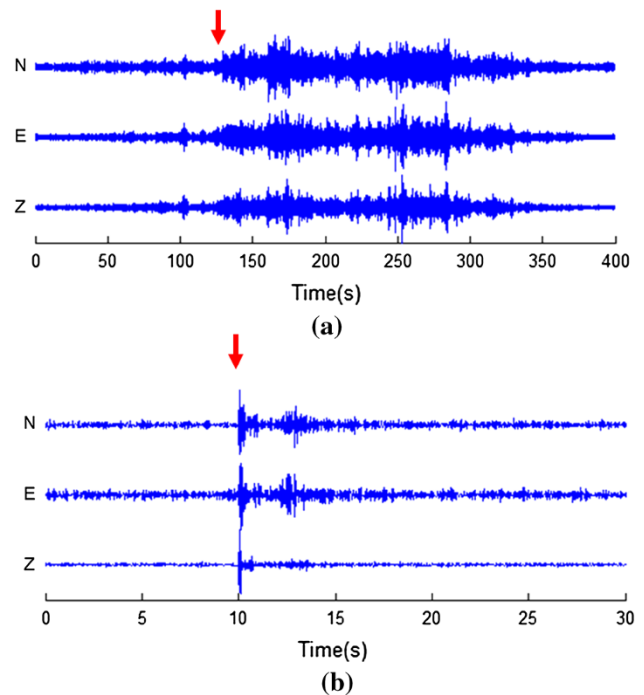


Fig. 6 Testing the compatibility between the beginning detections’ high-frequency SNR and wave-train’s duration to screen out noise bursts (a) and microseismic signals (b), marked by the red arrow in both cases. Detections due to both cases can be viewed as disturbances to meaningful processing results. Waveforms illustrated in both panels are 8–16 Hz bandpass filtered, and readers are reminded to note the difference of the timing scale between them

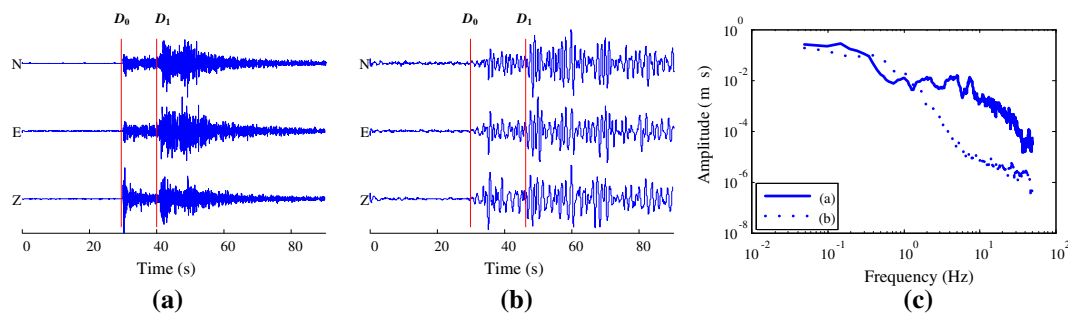


Fig. 7 Testing the compatibility between signals' spectral ratio and assumed S–P time difference to search for regional P/S pairs. The waveforms in **a** are the 3C seismogram of a local earthquake, while the waveforms in **b** are the seismogram of a teleseismic event. *Plot c* compares the displacement spectra of the beginning detections of the two seismograms, both spectra were computed using data within the 10 s long window beginning from the onset time of D_0 . It is clear that the signal in *a* has rich high-frequency content which is compatible with the assumption that D_0 and D_1 consist a P/S pair with S–P time difference a little more than 10 s. In contrast, the signal in *b* is deficient of high-frequency content which would contradict the assumption that D_0 and D_1 consist a P/S pair as the arrival time difference between them is about 15 s

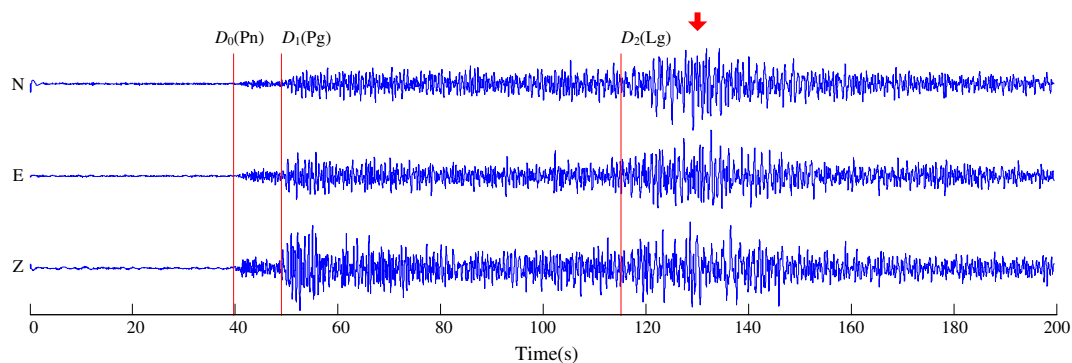


Fig. 8 Testing the compatibility between the S–P time difference and envelope peaks for phase-type identification for detections on regional seismograms. Waveforms displayed have been 0.5–1.5 Hz filtered, and the *arrow indicates* the general location of *envelope peaks* of the seismograms

tasks of the fourth and fifth stage are to complete network association and localization for regional and global events, respectively. Currently integral feature are mainly applied in the third stage, though their application in the second stage has also been investigated by us. We will be back to discuss this point later.

Finding major P and S arrivals and grouping them together is of key importance for detecting and locating regional events. Figure 5 is a diagram to explain how integral feature defined in Sect. 2 can be used for this objective. The searching process for P/S pairs is divided into two steps. The first step is to search out candidate direct P arrivals from beginning detections; and the second step is to search out major S arrivals from respective following detections. During the searching process, candidate direct P arrivals and respective P/S combinations have to go through a series of tests, including: (1) compatibility between the beginning

detection's signal-to-noise ratio (SNR) and the duration time of respective wave-train (T1); (2) compatibility between amplitude of assumed P and S phases from same events (T2); (3) compatibility between assumed phase type and measured polarization characteristics (T3); (4) compatibility between amplitude ratios among different frequency bands and the assumed S–P time difference (T4); and (5) compatibility between assumed S–P time difference and position of respective seismogram's envelope peaks (T5).

Among these tests, T2 and T3 are conventional and reasons of them are easy to understand. For test T1, the reason behind it is that if a signal is of very low SNR, but with very long-duration time, especially when it occurs within high-frequency bands such as 8–16 Hz, then corresponding signal usually is a noise burst, such as that in Fig. 6a. On the other hand that if a signal is of great SNR, but very small duration time also within high-frequency

band, then it may be signals emitted by microseismic events very close to the recording station and normally cannot be associated to other regional stations (e.g., Fig. 6b). The reason for test T4 is that, if an assumed P/S pair has a time difference corresponding to local distances, then respective signals, especially the P waves, should have rich high-frequency compositions. In practice this can be used to verify a local P/S pair by checking comprising signals' spectral ratios against the assumed S–P time difference (Fig. 7). The reason for test T5 has already been mentioned in Sect. 2, i.e., in case of local or regional seismograms, the arrival time of major S arrivals normally should be a little before the envelope peak of the whole seismogram, at least within some low-frequency bands such as 0.5–1.5 Hz and at horizontal components. This is true even for underground explosions which can generate P waves more efficiently than earthquakes. Based on this phenomenon, we apply the following condition, i.e.

$$-\varepsilon_1 < T_{P_0} - (t_S - t_P) < \varepsilon_2 \quad (6)$$

to test whether the assumed P/S pair is compatible with the envelope shape of respective seismogram, where T_{P_0} is the time delay of the peak of the whole wave-train, ε_1 and ε_2 are empirically specified parameters, and in practice ε_1 can be set to be 1 s; and ε_2 may take different values between 5 and 30 s, depending upon different $t_S - t_P$ range for the assumed P/S pair. The reason to set ε_1 to be 1 s instead of zero is for the consideration of that possible error may exist for S arrival time estimation in practice and in some cases the time interval $T_{P_0} - (t_S - t_P)$ may be very small (e.g., S waves with very sharp onset). An example of applying this condition is illustrated in Fig. 8. The envelope peaks on the seismograms in the figure is about 90 s after its beginning detection D_0 . By in Eq. (6) we can avoid mistaking the following detection D_1 as a S phase, which often occurs when only local feature are applied for phase association, especially in the case of 3-component stations.

4 Test results using data from Xinjiang seismic station network

The approach presented here has been applied to routine data processing by us for regional seismic station networks of various scales for years. Promising results have been achieved with both false and missed events rate normally less than 10 %, in comparison prior to its application the false events rate had accounted about 50 %; and missed events rate for regional events having signals recorded at least at three stations accounted more than 15 %. As an example, here we present the test processing results of using two weeks' continuous data of the full-scale Xinjiang Seismic Station Network (XJSSN) which is illustrated in

Fig. 9. The time period of data that we processed is from 06/01/2010 to 06/16/2010. Automatically detected events were manually checked, and the results of regional events were compared with XJEB, the event bulletin produced by the Seismological Bureau of Xinjiang Province, China, while results of teleseismic or far regional seismic events outside the network are compared with that of the Reviewed Event Bulletin of the Vienna International Data Centre.

For the time period above mentioned, there are 3,654 events detected by our prototype software, only 284 of them are false or split events, approximately accounting for 7.8 %. To evaluate the detection rate for regional and local events, events listed in XJEB were compared to events automatically detected. Within the time period from 06/01/2010 to 06/16/2010, there are 1,025 events detected in XJEB, among which 136 events are of magnitude above M_L 2.0, 325 events between M_L 1.0–2.0, 48 events below M_L 1.0, and 516 events are underdetermined, i.e., events defined by P and S waves at only one station. For an underdetermined event, its epicenter estimation is not given in the XJEB if S–P time difference at the defining station is greater than 3 s and it is assigned to be the station location otherwise. Examples of automatically detected regional events are given in Fig. 10, and detection rate for these four event groups is summarized in Fig. 11. The detection rate is 100 % for events of $M_L > 2.0$, 98.5 % for M_L between 1.0–2.0, 95.8 % for $M_L < 1.0$, and 52.1 % for events underdetermined. The major cause why detection rate for underdetermined events is much lower than that for other three event groups is that most of these events are micro-earthquakes very close to the recording station with S–P time difference only a few seconds, and our software especially its signal detection program is not designed to

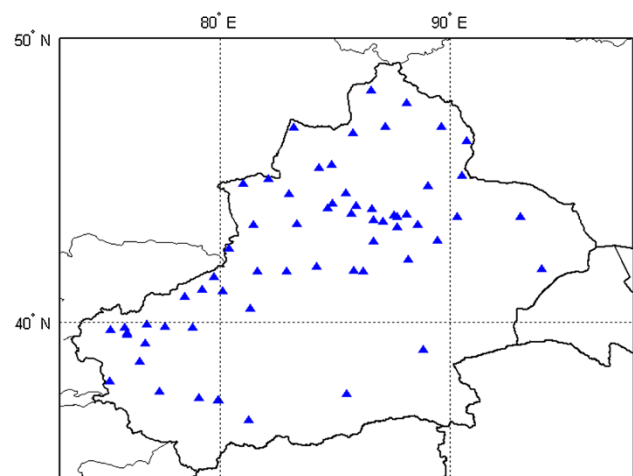


Fig. 9 Locations of seismic stations of the Xinjiang seismic station network (XJSSN)

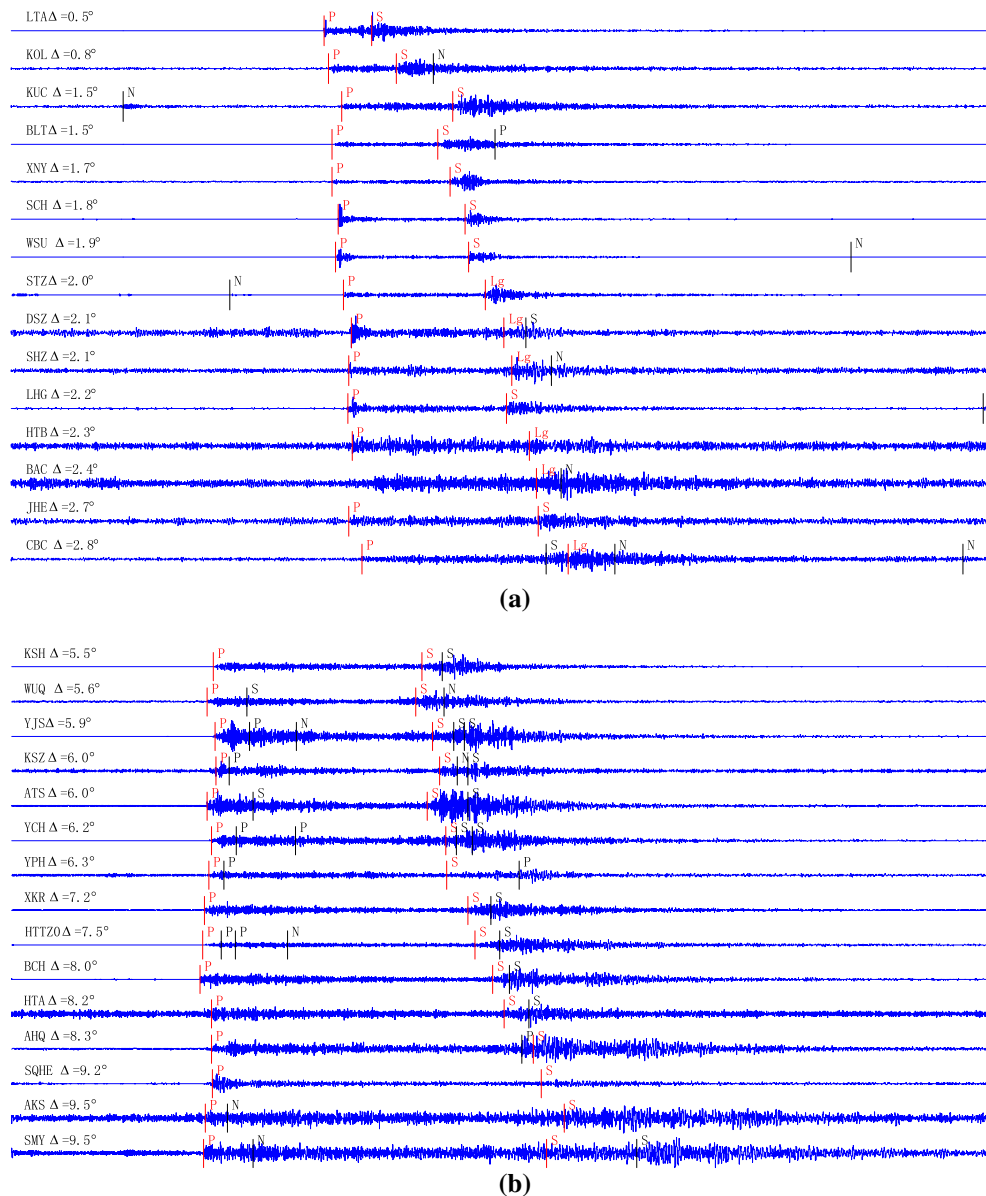


Fig. 10 Examples of automatic processing results for two regional events. **a** Seismograms of an M_L 2.6 Southern Xinjiang earthquake occurred at 2010/06/09 10:58:59. The epicenter of this earthquake is (42.33°N , 84.82°E) in XJEB, our automatically determined epicenter is 11 km away from it. **b** Seismograms of an m_b 3.4 earthquake occurred at 2010/06/09 16:01:08 in the Hindu Kush Region, Afghanistan. For the epicenter of this earthquake is outside of XJSSN, no information is given in the XJEB. However, the IDC REB determined its epicenter and depth is (35.89°N , 69.57°E) and 82.5 km, respectively, and our automatically determined result is epicenter (36.1143°N , 69.771°E), depth 107.8 km. For each seismogram in this figure, respective recording station and epicentral distance are given at its *left side*, and all detections triggered within the displayed waveform time window are added, where *red ones* are automatically associated defining phases for corresponding event, *black ones* are unassociated, and N stands for noise. Readers should be aware that phase label for only associated detections is determined by integral feature, and that for unassociated phases are determined by a neural network with only local feature as the input

detect such events. For example, the signal detection program of our software requires that a detection to be triggered should be at least 3 s later than the last detection before it. However, among the 516 underdetermined events, 174 of them are of S–P time difference less than that. Besides nearly half the underdetermined events, there

are another 7 events listed in XJEB were not detected by DRSN. These events typically did not have good signal-to-noise ratio and had both clear P and S waves at stations less than three. However, for the undetected events, P and S waves at these key stations had not been both detected or though detected, but not grouped together due to causes

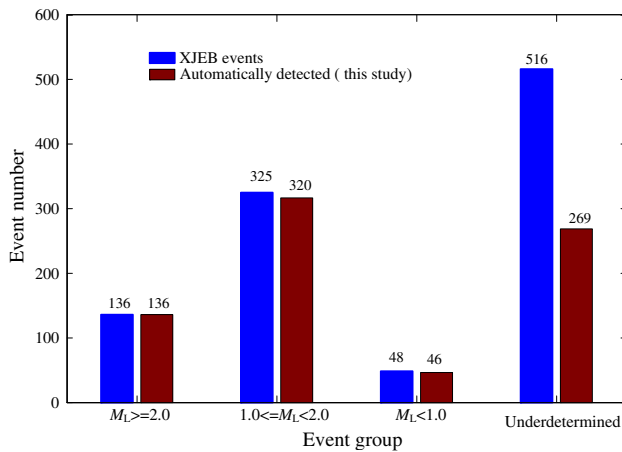


Fig. 11 Comparison of automatically detected events with events listed in the XJEB

include: (1) emergent P wave; (2) very short S–P time interval (<5 s); (3) data dropout occurring just ahead of signal’s arrival time; (4) error in estimating arrival time or polarization characteristics; (5) the signal were affected by coda wave of other big events and the detection corresponding to the first P wave was identified as the following type; and (6) the only stations with both P and S waves detected just had a small amplitude S. The first three causes may lead to P or S wave fail to be detected, and the later three may lead to P and S waves fail to be grouped together. And for the 7 undetected events, generally there were more than one causes happened together and so they failed to be detected.

Comparison between automatically estimated epicenters and epicenters in XJEB for events of $M_L > 2.0$ are given in Fig. 12. The average deviation between them is 15.6 km. Considering that the Xinjiang seismic station network is quite sparse and average minimum epicentral distance for the earthquakes is close to 100 km; this is a fairly good result in practice and suggests that the application of integral feature helps improve the performance of not only events detection but also localization.

It is worth to point out that the number of local or regional events detected by our prototype software is much larger than that listed in XJEB (almost triple of it). Local or regional events detected by us, but not listed in XJEB include events outside Xinjiang, events corresponding to quarry blasts and events that are too weak to be efficiently picked up by human analysts or cannot meet the quality requirement of XJEB. Typical examples of the unmatchable events are given in Fig. 13.

Besides the good performance to detect local and regional events, the prototype software also performs well to detect teleseismic or far regional events outside of the network’s coverage. The method about how to utilize a regional seismic station network to effectively and reliably

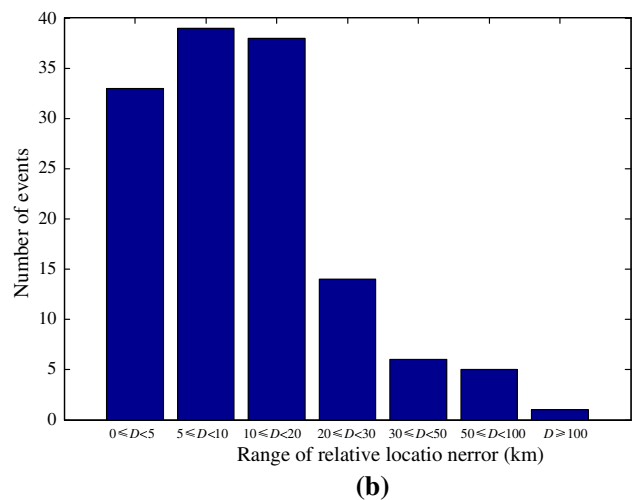
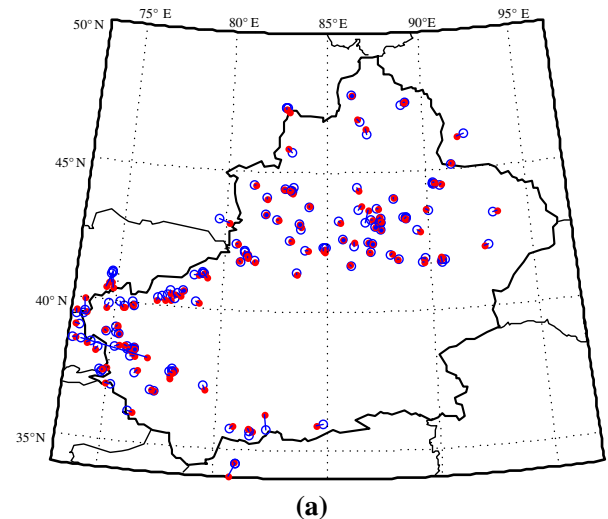


Fig. 12 a Comparison of automatically determined epicenters (red dot) with that in XJEB (open blue circle) for events of $M_L > 2.0$ and b the distribution of relative location error between them

detect teleseismic events is beyond the scope of this paper and will be discussed in another one. Here, we just mention that the integral feature introduced in this paper are also important for the identification and association of teleseismic signals for it can help to greatly reduce the number of detections need to be considered in the signal association process by screening out the following detections, as well as obvious noises and local signals.

5 Discussion and conclusions

Integral feature such as overall envelope shape are important for seismogram interpretation. It is surprising that to-date they are not utilized in main-stream automatic seismic data processing possibly due to the fact that they

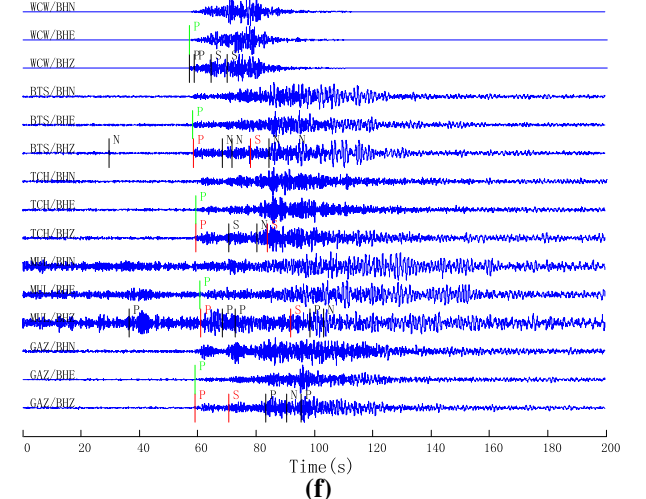
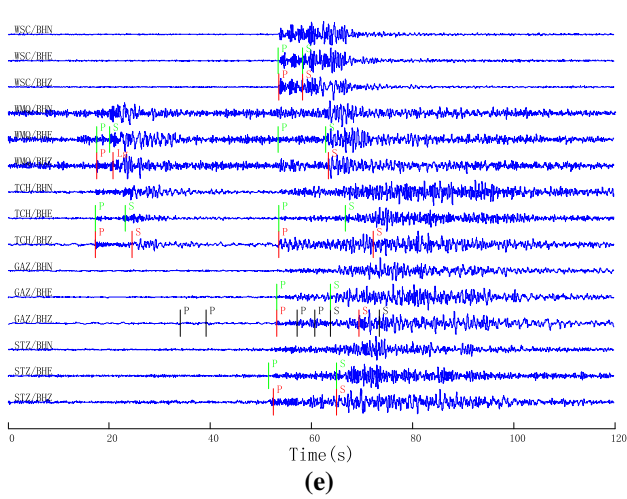
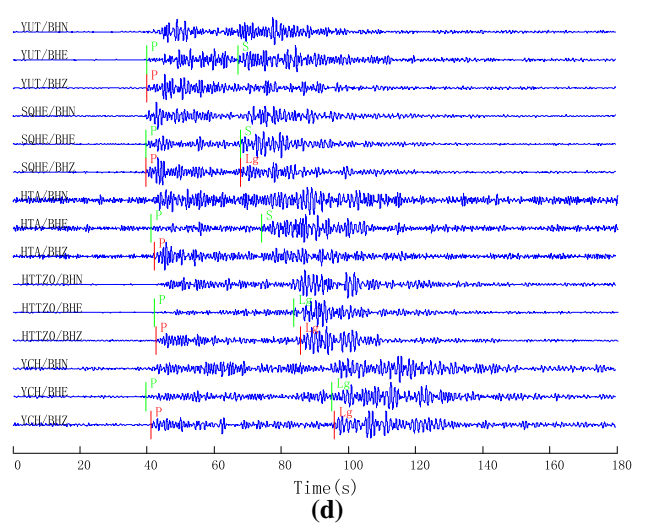
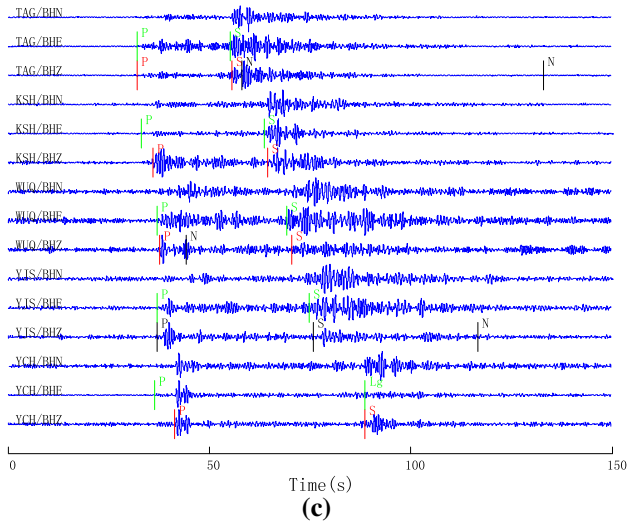
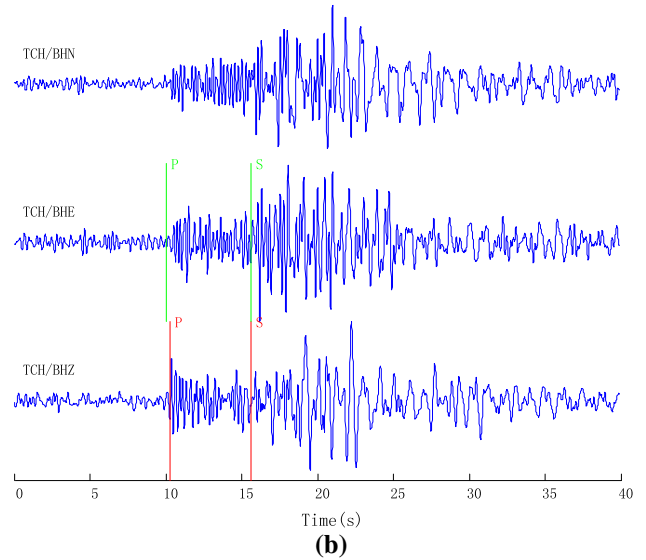
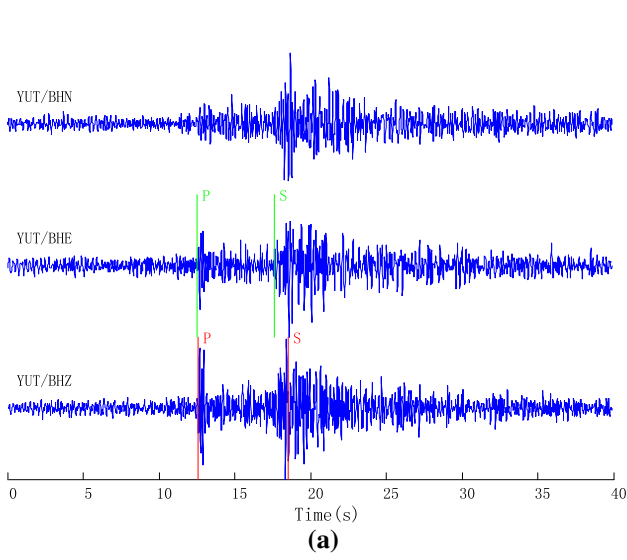


Fig. 13 Examples of local/regional events detected by DRSN but not listed in XJEB. Waveforms displayed have already been filtered using appropriate filters. Phases displayed on the trace of BHE components in green are manual picks, and meaning of picks displayed on BHZ component is the same as that in Fig. 10. **a** small earthquake recorded by only one station (2010/06/01 07:12:42, 36.4°, 81.5°, $M_L \sim 0.5$); **b** small quarry blasts recorded also by only one station with P, S and Rg discernible (2010/06/01 06:57:39, 44.2°, 88.5°, $M_L \sim 0.5$); **c** small earthquake near Xinjiang-Tajik border region (2010/06/01 06:50:47, 37.7°, 72.8°, $M_L \sim 2.5$); **d** small earthquake near Xinjiang-Xizang border region (2010/06/01 06:58:03, 34.5°, 80.9°, $M_L \sim 2.3$); **e** quarry blast recorded at multiple stations (2010/06/01 07:10:23, 43.3, 87.7, $M_L \sim 1.5$); note that signals of another detected event (2010/06/01 07:09:54, 44.0, 87.7, $M_L \sim 0.6$) can also be seen at two of the stations; **f** another quarry blast with signals at multiple stations (2010/06/01 07:19:43, 44.8, 89.3, $M_L \sim 2.0$), S phases for the event cannot be reliably determined. Note that all these events occurred in half an hour (2010/06/01 06:50–2010/06/01 07:20). Actually DRSN detected 11 events during this time period, among them only two are listed in XJEB

have not been clearly defined and parameterized before. Here in this paper we have shown that the overall envelope shape of the whole seismograms can be depicted by only a few parameters. On the base of this progress, we have developed a novel technique to incorporate seismograms' integral feature into the conventional DIAL approach for seismic data processing. In addition to picking up individual signal detections, the new technique picks up separate wave-trains and extracts their characterizing parameters at the same time. In comparison with conventional local feature for individual signals, the newly introduced characterizing parameters for the whole wave-trains can better reflect the relationship among different detections, as well as their relative position within respective seismograms, and its combination with signals' local feature including amplitude, frequency, azimuth, and polarization can lend help to resolve problems about signal association and phase-type identification, especially in the case of 3-component seismic stations, and consequently reduce the rate of both false and missing event rate a lot.

Nevertheless, some may argue that in case of very weak signals or signals of overlapping wave-trains of multiple events, the effectiveness of the integral feature would be discounted. Indeed, these are two situations that should be cautious to deal with and why local feature are still needed. For signals of overlapping wave-trains of multiple events, the direct P wave of the second events may have D_{flag} characterization of 1 other than 0. Fortunately, the frequency of this kind of situation is low, and often it can be identified by testing the compatibility between its features (e.g., amplitude, frequency, azimuth, etc.) against that of the signals of the first one. For example, if we have a detection D_n that is really a direct P wave, however, its D_{flag} is mislabeled as 1 for the interference of another wave-train arriving at the station just a little while before it. For convenience, here denote detections of the interference

wave-train with arrival times previous to D_n to be D_0, D_1, \dots, D_{n-1} , where D_0 is the beginning detection of $D_{\text{flag}} = 0$. In practice, D_n can be accepted as a possible direct P wave to test if one of the following situations emerges. (1) The amplitude of D_n is significantly larger than that of from D_0 to D_{n-1} ; (2) D_n has richer high-frequency content than D_0 does; (3) The azimuth and slowness indicates that D_n cannot be coda P or S waves of D_0 in case of array stations, and in case of 3-component stations this may be surrogated by polarization characteristics, however, it is much less reliable for the later situation and one should be cautious about applying it; (4) An S wave has already been associated to D_0 and D_n has characteristics inconsistent with a S coda wave. For very small events, the problem that should be dealt with is different. In this case, it is possible that P and S waves even at regional stations beyond some distance cannot constitute a continuous wave-train due to amplitude of P coda waves have dropped to below that of background noises before the arrival of S waves. Under this situation, we abandon the search for regional P and S pairs and will deal with them later in the network processing stage in case of three component stations, or they may be still grouped together taking consideration of their compatibility in azimuth, slowness, amplitude, and dominant frequency in case of array stations. Anyway, the application of integral feature would have no negative effect upon processing result in comparison with conventional approaches even for the two kinds of special situations.

As we have mentioned earlier that currently we have not apply the integral feature in the artificial neural network for phase-type identification, though we believe that these feature may improve its performance. The neural network that we currently used in DRSN for phase-type identification is similar to that described by Wang (2002). The later used three cascade neural network to categorize detections into four wave types, i.e., noise, teleseismic, regional P or regional S, where the first neural network distinguishes between noise and signal, the second distinguish between teleseismic or regional, and the third distinguish between regional P and regional S. There are 15 attributes used together for the input of each of the neural networks, including dominant period, polarization attributes, contextual attributes, and a spectral representation of the horizontal-to-vertical power ratio whose exact definition was explained in Wang (2002). The neural networks that we practically used differ from that in Wang (2002) in three aspects: (1) two instead of three cascade neural networks are used by us and signals are categorize into only three wave types, i.e., noise, P and S; (2) there are only 10 attributes all together for the input of each of the two neural networks, include signal-to-noise ratio (SNR) within 5 different frequency bands in the range of 0.5–16 Hz, rectilinearity, planarity determined by polarization analyses

(e.g., Jurkevics, 1988) within the frequency band of the best SNR, horizontal-to-vertical amplitude ratio also measured within the frequency band of the best SNR, and pseudo-duration time within 0.75–1.5 Hz and 8–16 Hz, respectively. Unlike the duration time defined in Eq. (2) which is measured for the whole wave-train, the pseudo-duration is measured for individual detections and it is cut off if it is longer than 20 s. The performance of the cascade neural networks is generally ok, however, still with a large portion of detections was mis-identified. For this reason, in the subsequent association processing stages, the output of the cascade neural networks actually was completely ignored. For example, in the process of searching P/S pairs, the major requirement for a detection to be accepted as the generator for searching P/S pairs is that its D_{flag} should be 0 or it has been identified as the direct P arrival buried in a previous wave-train as we just discussed in the last paragraph, and the requirement for the S phase is to pass tests which has been described in Sect. 3. So output of the neural networks actually has no impact on the final associated events. However, we still keep this processing stage for at least it can give a label to detections that are not associated to any event.

Our experience suggests that this novel technique presented in this paper have promising application prospect for reliable automatic seismic data processing. Actually we have applied this technique to regional seismic station networks of various scales with results generally as good as we have shown for the XJSSN. Nevertheless, it is worth to point out that so far the application of seismograms' integral features by us is far to be optimized for many decision conditions are set by experience. It is necessary to optimize its application through statistical studies about the joint distribution behavior of both integral and local feature for different types of seismograms, as well as different kinds of seismic phases in the future to achieve more satisfying result.

Appendix: the multiple index STA/LTA detector

To detect individual signals, we use a modified STA/LTA detector which exploits two other indices, i.e., MTA/MTA_{old} and STA/STA_{old} as supplement to STA/LTA itself. For a time series $x(n)$, the running average LTA, STA, STA_{old} , MTA and MTA_{old} is obtained by

$$LTA(n) = \frac{1}{L} \sum_{i=n-L}^{n-1} |x(i)|, \quad (\text{A1})$$

$$STA(n) = \frac{1}{S} \sum_{i=n}^{n+S-1} |x(i)|, \quad (\text{A2})$$

$$STA_{\text{old}}(n) = \frac{1}{S} \sum_{i=n-S}^{n-1} |x(i)|, \quad (\text{A3})$$

$$MTA(n) = \frac{1}{M} \sum_{i=n}^{n+M-1} |x(i)|, \quad (\text{A4})$$

$$MTA_{\text{old}}(n) = \frac{1}{M} \sum_{i=n-M}^{n-1} |x(i)|, \quad (\text{A5})$$

where S , M , L are window length for STA, MTA and LTA and in practice they are set to be 1 s, 6 s, and 30 s respectively. A signal is declared when either the following two conditions, i.e.

$$\begin{aligned} \text{cond1} &= STA/LTA > Th_{11} \text{ and } MTA/MTA_{\text{old}} \\ &> Th_{12} \text{ and } STA/STA_{\text{old}} > Th_0 \end{aligned} \quad (\text{A6})$$

$$\begin{aligned} \text{cond2} &= STA/LTA > Th_{21} \text{ and } MTA/MTA_{\text{old}} \\ &> Th_{22} \text{ and } STA/STA_{\text{old}} > Th_0 \end{aligned} \quad (\text{A7})$$

is true, where the symbol “&” stands for the logical operator “and”, Th_0 , Th_{11} , Th_{12} , Th_{21} , Th_{22} represent respective threshold values. In practice, $Th_0 = 1.1$, $Th_{11} = 4.0$, $Th_{12} = 1.5$, $Th_{21} = 3.5$, $Th_{22} = 2.2$ are used for all stations and for all frequency bands.

The meaning of using MTA/MTA_{old} and STA/STA_{old} can be explained as the following. First, the requirement $STA/STA_{\text{old}} > Th_0$ with $Th_0 > 1$ means that a detection may be declared only when the signals' amplitude is rising. This can reduce redundant and meaningless detections caused by coda waves. Second, with a window length of 6 s, MTA in fact acts as a middle-term average and MTA/MTA_{old} may be more sensitive to emergent signals than STA/LTA and for signals of higher MTA/MTA_{old} the requirement for STA/LTA can be relaxed.

References

- Allen RV (1978) Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am* 68:1521–1532
- Allen R (1982) Automatic phase pickers: their present use and future prospects. *Bull Seismol Soc Am* 72:S225–S242
- Anant KS, Dowla FU (1997) Waveform transform methods for phase identification in three-component seismograms. *Bull Seismol Soc Am* 87:1598–1612
- Anonymous (1999) IDC Processing of Seismic, Hydroacoustic, and Infrasonic Data. IDC-5.2.1, http://www.rdss.info/librarybox/idc_docs/downloads/521.pdf
- Arora N, Given J, Tomuta E, Russell S, Spiliopoulos S (2012) Analyst evaluation of model-based bayesian seismic monitoring at the CTBTO. In: Proceedings of the 2012 monitoring research review: ground-based nuclear explosion monitoring technologies, LA-UR-12-24325, 191–199
- Arrowsmith SJ, Whitaker R, Katz C, Hayward C (2009) The F-detector revisited: an improved strategy for signal detection at seismic and infrasound arrays. *Bull Seismol Soc Am* 99:449–453
- Baer M, Kradolfer U (1987) An automatic phase picker for local and teleseismic events. *Bull Seismol Soc Am* 77:1437–1455

- Bai CY, Kennett BLN (2000) Automatic phase-detection and identification by full use of a single three-component broadband seismogram. *Bull Seismol Soc Am* 90:187–198
- Bai CY, Kennett BLN (2001) Phase identification and attribute analysis of broadband seismograms at far-regional distances. *J Seismol* 5:217–231
- Cansi Y (1995) An automatic seismic event processing for detection and location: the PMCC method. *Geophys Res Lett* 22:1021–1024
- Dai H, MacBeth C (1997) The application of back-propagation neural network to automatic picking seismic arrivals from single-component recordings. *J Geophys Res* 102:15105–15113
- Dai H, Macbeth C (1995) Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophys J Int* 120:758–774
- Der ZA, Shumway RH (1999) Phase onset time estimation at regional distances using the CUSUM-SA algorithm. *Phys Earth Planet Int* 113:227–246
- Draelos TJ, Ballard S, Young CJ, Brogan RA (2012) Refinement and testing of the probabilistic event detection, association, and location algorithm. In: Proceedings of the 2012 monitoring research review: ground-based nuclear explosion monitoring technologies, LA-UR-12-24325, 221–231
- Earle P, Shearer P (1994) Characterization of global seismograms using an automatic-picking algorithm. *Bull Seismol Soc Am* 84:366–376
- Evans J, Allen S (1983) A teleseismic-specific detection algorithm for single short period traces. *Bull Seismol Soc Am* 73:1173–1186
- GSE/JAPAN/40 (1992) A fully automated method for determining the arrival times of seismic waves and its application to an on-line processing system. Paper tabled in the 34th GSE session in Geneva GSE/RF/62
- Hildyard MW, Nippres SEJ, Rietbrock A (2008) Event detection and phase picking using a time-domain estimate of predominate period T^{pd} . *Bull Seismol Soc Am* 98:3025–3032
- Inclán C, Tiao GC (1994) Use of cumulative sums of squares for retrospective detection in the changes of variance. *J Am Stat Assoc* 89:913–923
- Jurkevics A (1988) Polarization analysis of three-component array data. *Bull Seismol Soc Am* 78:1725–1743
- Küperkoch L (2010) Automated recognition, phase arrival time estimation, and location of local and regional earthquakes, Dissertation zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften an der Fakultät für Geowissenschaften der Ruhr-Universität Bochum, Bochum
- Küperkoch L, Meier T, Lee J, Friederich W, EGELADOS Working Group (2010) Automated determination of P-phase arrival times at regional and local distances using higher order statistics. *Geophys J Int* 181:1159–1170
- Le Bras R, Russell S, Arora N, Miljanovic V (2011) Machine learning at the CTBTO. Testing and evaluation of the false events identification (FEI) and Vertically Integrated Seismic Association (VISA) Projects. In: Proceedings of the 2011 monitoring research review: ground-based nuclear explosion monitoring technologies, LA-UR-11-04823, 313–321
- Moore DA, Mayeda KM, Myers SM, Seo MJ, Russell SJ (2012) Progress in signal-based bayesian monitoring. In: Proceedings of the 2012 monitoring research review: ground-based nuclear explosion monitoring technologies, LA-UR-12-24325, 263–273
- Murphy JR (1996) Types of seismic events and their source descriptions. In: Husebye ES, Dainty AM (eds) *Monitoring a comprehensive test ban treaty*. Kluwer Academic Publishers, Dordrecht, pp 225–245
- Pearce RG, staff, Monitoring and Data Analysis Section, IDC (2009) Exploiting the skills of waveform data and analysts in the quest for improved automatic processing, poster presented at the ISS09 Conference, Vienna, 10–12 June 2009, [http://www.ctbto.org/fileadmin/user_upload/ISS_2009/Poster/DM-12A%20\(PTS\)%20-%20Robert_Pearce%20etal.pdf](http://www.ctbto.org/fileadmin/user_upload/ISS_2009/Poster/DM-12A%20(PTS)%20-%20Robert_Pearce%20etal.pdf)
- Procopio MJ, Young CJ, Lewis JE (2009) Using machine learning to improve the efficiency and effectiveness of automatic nuclear explosion monitoring system. In: Proceedings of the 2009 monitoring research review: ground-based nuclear explosion monitoring technologies, LA-UR-09-05276, 788–797
- Roberts RG, Christoffersson A, Cassidy F (1989) Real-time event detection, phase identification and source location estimation using single-station three-component seismic data. *Geophys J* 97:471–480
- Ryzhikov GA, Biryulina MS, Husebye ES (1996) A novel approach to automatic monitoring of regional seismic events. *IRIS Newsletter* 15:12–14
- Saragiotis CD, Hadjileontiadis LJ, Panas SM (2002) PAI-S/K: a robust automatic seismic P phase arrival identification scheme. *IEEE Trans Geosci Remote Sens* 40:1395–1404
- Selby ND (2008) Application of a generalized F detector at a seismometer array. *Bull Seismol Soc Am* 98:2469–2481
- Selby ND (2011) Improved teleseismic signal detection at small-aperture arrays. *Bull Seismol Soc Am* 101:1563–1575
- Sleeman R, van Eck T (1999) Robust automatic P-phase picking: an online implementation in the analysis of broadband seismogram recordings. *Phys Earth Planet Int* 113:265–275
- Stefano RD, Aldersons F, Kissling E, Baccheschi P, Chiarabba C, Giardini D (2006) Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity. *Geophys J Int* 165:121–134
- Suteau-Henson A (1991) Three-component analysis of regional phases at NORESS and ARCESS: polarization and phase identification. *Bull Seismol Soc Am* 81:2419–2440
- Tibuleac I, Herrin E, Britton J, Shumway R, Rosca C (2003) Automatic determination of secondary seismic phase arrival times using wavelet transforms. *Seismol Res Lett* 74:884–892
- Tong C (1995) Characterization of seismic phases-an automatic analyzer for seismograms. *Geophys J Int* 123:937–947
- Tong C, Kennett BLN (1996) Automatic seismic event recognition and later phases identification for broadband seismograms. *Bull Seismol Soc Am* 86:1896–1909
- Wang J (2002) Adaptive training of neural networks for automatic seismic phase identification. *Pure Appl Geophys* 159:1021–10
- Wang J, Teng T-L (1995) Artificial neural network-based seismic detector. *Bull Seismol Soc Am* 85:308–319
- Wang J, Teng TL (1997) Identification and picking of S phase using an artificial neural network. *Bull Seismol Soc Am* 87:1140–1149
- Withers M, Aster R, Young C, Beirger J, Harris M, Moore S, Trujillo J (1998) A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bull Seismol Soc Am* 88:95–106
- 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* 93:1904–1912
- Zhao Y, Takano K (1999) An artificial neural network approach for broadband seismic phase picking. *Bull Seismol Soc Am* 89:670–680