Geophys. J. Int. (1993) **113**, 562–574

Master-event correlations of weak local earthquakes by dynamic waveform matching

Manfred Joswig and Hartwig Schulte-Theis

Institut für Geophysik, Ruhr-Universität Bochum, Postfach 102148, *D*-4630 *Bochum* 1, *Germany*

Accepted 1992 October 1. Received 1992 August 4; in original form 1991 November 8

SUMMARY

Dynamic waveform matching (DWM) performs a non-linear correlation between two seismograms that are similar in shape but may be squeezed or stretched relative to each other. It extends the application of master-event comparisons to seismograms of greater spatial distance and retains the high-timing resolution of correlation techniques that act on the original time series. The DWM approach is applied to data recorded by a small array being part of the BOCHUM UNIVERSITY GERMANY (BUG) network which monitors the mining-induced seismicity in the Ruhr basin of NW Germany.

The observed epicentres occur in clusters and therefore display only a limited number of seismogram waveform types. In one application an automatized cluster association with DWM obtains a resolution of about 100 metres at an epicentral distance of 20 to 40 km, using 10–20 defined master events for each region. These results are confirmed both by seismograms from a near-site station for mining-induced events from the *Hamm* region and by blast reports for a quarry region near *Wuppertal*.

In another application of DWM, array traces from the BUG array are correlated to yield azimuth and slowness for epicentre location. As for the master event application, this approach is tuned for high performance on weak local events using *a priori* information about the approximate epicentral region.

The implemented processes are shown to be capable of locating events with a rate of success equal to the performance of an experienced seismologist when processing all seismograms of four years BUG registration.

Key words: automated epicentre determination, cluster analysis, master-event technique, non-linear correlation, phase picker.

INTRODUCTION

The BOCHUM UNIVERSITY GERMANY (BUG) Network, shown in Fig. 1(a), consists of a tripartite small array of short period, vertical seismometers (stations KLB, SHA and TEZ) on the grounds of the Ruhr University. Furthermore, a three-component broad-band station (KLS) is deployed at the same site as station KLB; two short-period three-component stations are situated 40 km west (RPM) and 40 km east (HRH) of the small array within Ruhr Basin coal mines about 500 m and 1000 m below the surface, respectively. The main task of the network is to observe the mining-induced seismicity in the Ruhr Basin (Casten & Cete 1980; Hinzen 1982; Gibowicz, Harjes & Schäfer 1990), typically displaying clusters of very similar events [Fig. 1(b)]. Additionally, the network detects quarry blasts from limestone pits near Wuppertal which is some 25 km SW of KLB. Hence, the local seismicity monitored by the BUG network is restricted to an area of about 50 km around the small array.

For all epicentral regions covered by the BUG network, locating an event consists of determining array azimuth and distance; the hypocentral depth is within some 500 m and cannot be resolved. Azimuths are calculated from time differences between coherent P phases of three array stations. Distances are obtained by picking P and S onsets on KLS. Event locations near HRH and RPM can be further refined by azimuth determinations using particle motion and the distance constraints $t_S - t_P$ at these stations.

While for strong events phase picking does not impose any problems, the evaluation of weak events is often difficult because of a high noise level that causes ambiguous phase

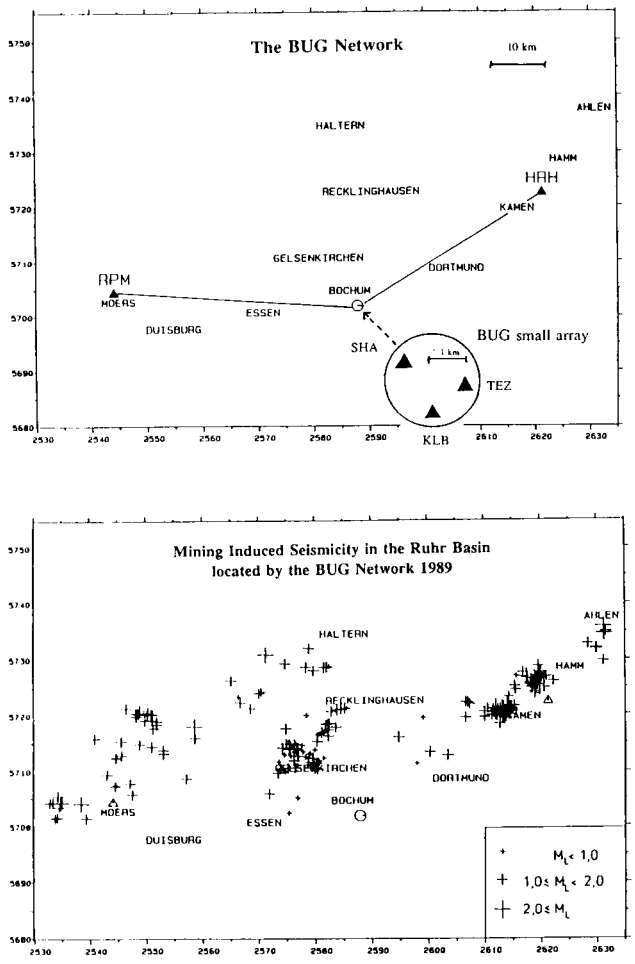

Figure 1. (a) Map of the BOCHUM UNIVERSITY GERMANY (BUG) Network. It consists of a small array and single stations. The small array is situated on the ground of the Ruhr University near Bochum. For further information about BUG see text. (b) Mining-induced seismicity located by the BUG Network in 1989.

identifications. Unfortunately, the bulk of the events are weak events and their processing represents a time-consuming task in the routine work of a seismologist Therefore, automated location of weak events would, if possible, be an attractive aiding tool. This is normally difficult to realize. However, the frequent occurrence of clusters in our region evoked the idea to evaluate and locate weak events by comparing them to a set of strong master events instead of processing them individually.

In this paper, we present an automatic method of evaluating BUG seismograms based on master-event correlation. Two different kinds of applications are presented: in the first method, each event cluster is represented by a strong master event. By correlating weak events with master events, a cluster association can be performed and strong constraints can be placed on the epicentre of the weak event relative to the master event. In the second method, different BUG small array traces are correlated to obtain relative phase onset times. Doing this for several pairs of traces, azimuth and slowness can be estimated. This second approach does not require master events. We demonstrate that conventional cross correlation is unsuitable for these applications. Instead, we present a method that uses a non-linear correlation scheme based on dynamic waveform matching (DWM) which yields good results for the automatic processing of BUG data.

THE PURPOSE OF EVENT CORRELATION

It is well known for all kinds of seismicity that seismograms of weak events originating from the same epicentral region and recorded at the same station are often very similar (Stauder & Ryall 1967; Ishida & Kanamori 1978). The spectral corner frequency of events with a local magnitude smaller than 2–3 generally remains constant (Archuleta *et al.* 1982; Hanks 1982; Frankel & Kanamori 1983). Finally Geller & Mueller (1980) presumed that earthquakes causing similar seismograms are based on the same focal mechanism and that their hypocentres lie within one quarter of the shortest wavelength to which similarity extends. This so-called $\lambda/4$ criterion was confirmed by Frankel (1982), Pechman & Kanamori (1982), Motoya & Abe (1985) and Thorbjanardottir & Pechman (1987). For the seismicity monitored by the BUG network, a cluster of similar mining induced events was investigated by Gibowicz *et al.* (1990).

By classical master-event techniques, an earthquake is relocated relative to a strong master event whose hypocentre is well known (Dewey 1972; Fitch 1975; Spence 1980). The relative coordinates are found by inverting a system of linear equations which contains P-wave traveltime differences for the two events at different stations. Prior to this step, the degree of similarity between a master event and an actual event should be estimated by cross correlation. However, different authors suggest different time windows for the correlation analysis. While Geller & Mueller (1980) take the whole seismograms, Pechman & Kanamori (1982), Frankel (1982) and Thorbjanardottir & Pechman (1987) define a comparison window based on the P and S phase. Even more restrictive is Harris (1991) who determines similarity for the P-phase wavetrain only.

When investigating local seismic data by master-event techniques, the number of events correctly attached to clusters strongly depends on the chosen master events. Correlating all the data with all the master events from the appropriate region, a cluster association is performed (Israelson 1990; Harris 1991). Usually the master events are taken directly from the data set, but they may also be optimized linear combinations of several events (Harris 1991) or synthetic events (Macbeth & Redmayne 1989). Instead of cross correlation in the time domain, cross-spectral analysis can be used for master-event relocation (Poupinet, Ellsworth & Frechet 1984; Ito 1985; Scherbaum & Wendler 1986). However, the latter approaches are useful in automatization only for a sufficiently high signal-to-noise ratio, while we focus on data of weak events. Cross-spectral techniques are preferred for interactive detailed analysis of some selected events with a theoretical subsample timing resolution. In our investigations, we cannot utilize this benefit because AD conversion at different sites is synchronized by our equipments within one sample—obviously, subsample resolution of cross correlation needs subsample accuracy in all involved AD conversions.

The general idea of matching similar seismic waveforms from different sources which are measured at one site in Fig.

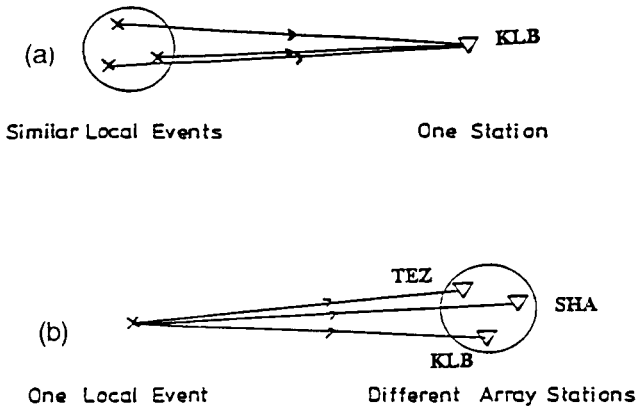

Figure 2. Application of DWM to Seismology, in (a) a weak event is evaluated by comparison with its stronger master event. Every cluster of similar events will be characterized by a representative master event; comparing on alternate choices performs automated cluster association. In (b) different array traces of the same event are correlated to determine traveltime differences of coherent *P* phases. This gives azimuth and slowness for epicentre beam steering.

2(a) can also be applied to the inverse problem. This is the situation of array seismology in Fig. 2(b) where different traces of one event are correlated to investigate local inhomogeneities at the receiver site (VanDecar & Crosson 1990). In many cases the inter-station distance of arrays is greater than the distance between adjacent epicentres for the master-event technique, so the correlation window must be restricted to the *P* onset in this case.

Figure 3 presents the registrations of two similar events from the same epicentre region [Fig. 2(a)]. They are located near Hamm in the eastern part of the Ruhr Basin and were recorded at the best site, KLB, of the BUG small array. Choosing window lengths of 2, 8 and 40 s will result in correlations that are performed on the *P* onset, the *P* coda and the whole seismogram, respectively. The two seismograms appear to be similar when compared over their whole length, but the 2s windows yield periodic maxima for all frequency bands. So the *P* onset association gets ambiguous, while the low-frequency surface waves remain stable and thus can be used to adjust an unequivocal peak and trough mapping. However, this correlation between low-frequency phases has insufficient timing resolution and often some shift relative to the high-frequency maximum. All together, the cross correlation over the whole seismogram length yields a clear, broad maximum in the lower-frequency bands. Unfortunately, the more unequivocal the maximum is, the greater the timing inaccuracy becomes. Also, the exact position of the main maximum varies slightly because of different residuals in the phase velocity. We found that for local events with low S/N ratio the similarity tends to be characteristic more for the whole seismogram than for any short time windows. High values in the cross correlation are necessary for all the time windows, but a single high correlation value for any window length is not sufficient for cluster association or event relocation.

Figure 4 once again shows a local *Hamm* event, this time measured at the two BUG array stations KLB and TEZ which are 1 km apart. This is the situation of Fig. 2(b). Our aim is now to correlate coherent phases to determine the time differences for the correct beam steering. As seen in this example, linear cross correlations over the longer time windows degrade in all frequency bands because the two signals are stretched and squeezed relative to each other. This effect is caused by the different residuals in phase velocity over the array aperture. The only correlations that yield unambiguous maxima are the 2 s and 8 s windows in the lowest passband (0.5–4 Hz), however, they suffer from poor timing resolution and are of no value for the determination of beam steering delays.

When the seismologist examines seismograms of Figs 3 and 4 interactively, he/she handles the problem of time warping—often unconsciously—by independent correlation

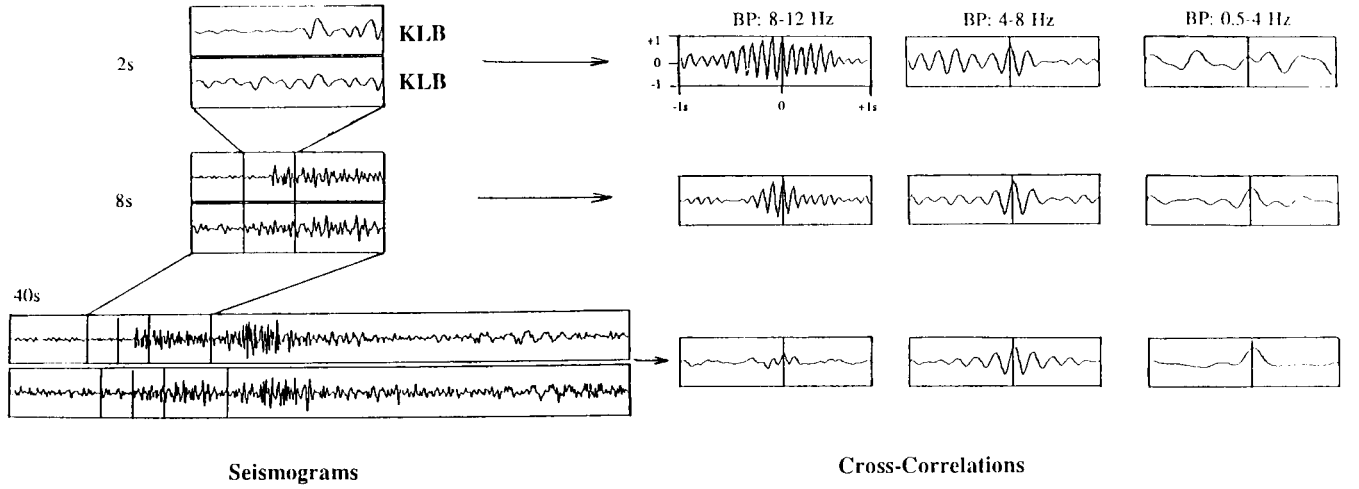

Figure 3. Cross correlation of two similar earthquakes: seismograms are displayed in 2/8/40 s windows for *P* onset, *P* coda and whole seismogram, measured at station KLB of the BUG small array. The event data are 3.1.1989 02:21:40, M_l = 1.7, Δ = 40 km for the upper trace and 18.1.1989 13:06:53, M_l = 1.5, Δ = 40 km for the lower one, the source region is *Hamm*. The cross correlation is performed in three pass-bands, it either lacks in timing resolution or shows significant ambiguity. Also the maximum peak shifts to later times for the longer time windows.

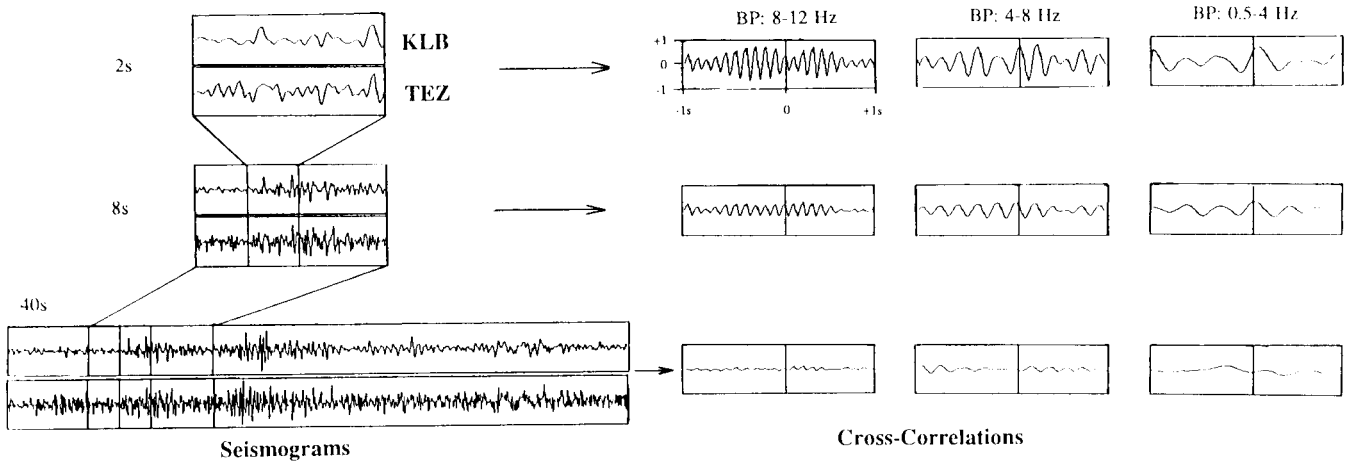

Figure 4. Cross correlation of two different array traces: seismograms are displayed like in Fig. 3 but this time measured at the two BUG small array station KLB and TEZ. The data of the event are 17.2.1989 13:14:02, $M_L = 1.2$, $\Delta = 40$ km, the source region is *Hamm*. This time the decay for longer time windows is even more obvious while the 2 s cross correlations are periodic.

over several wavetrains. A common procedure is to put paper seismograms one upon another, turn them both to the light and shift them relative to each other. By this procedure, possible phase correlations of limited duration are searched. Automatizing this process is rather complex and demands a non-linear correlation method.

NON-LINEAR CORRELATION METHODS

Non-linear correlation procedures are well known in speech processing as *dynamic time warping* (DTW). While common cross correlation matches two signals by linear time shift, DTW methods admit the correlation of signals which are compressed or stretched relative to each other, for example the same word pronounced by different speakers (Myers & Rabiner 1981). DTW performs a non-linear shifting and warping of the time axis until the dominant features of both waveforms are correctly matched, i.e. matching over the whole signal length can be achieved with more degrees of freedom as in conventional cross correlation.

In geophysics, non-linear correlation methods are used in the exploration industry to perform automatic stratigraphic correlation. Two completely different ways of non-linear matching can be distinguished in the literature. The first group of procedures carries out a feature extraction by converting the geophysical data into strings of primitives. The matching is exclusively performed on these waveform descriptions. The second group are methods working on the original data with possible prefiltering. They are extensions of linear cross correlation and will be designated here as dynamic waveform matching (DWM), a term first suggested by Anderson & Gaby (1983).

Procedures of the first kind should extract data features that are best suited to the specific matching problem. Le & Nyland (1990) describe seismic signals as pattern consisting of simple subpatterns like half waves, to which the attributes *amplitude* and *wavelength* are attached. Liu & Fu (1983) use a pattern representation where seismic waveforms are transformed into strings of characters in three steps: the processing starts with signal segmentation and feature extraction for each segment. Then similar segments are grouped into clusters and named by letters. A primitive recognition can be performed on the resulting string representations by determining the similarity or *Levensthein* distance which is the minimum number of edit operations to transform one string into the other. Edit operations are, for example, deletions, insertions or changes of characters. Effectively, the problem is solved by pathfinding through a matrix formed by the two character strings of pattern and actual data. It is typical for this kind of approach to minimize a cost function of distance instead of maximizing the correlation product.

An alternative to character strings for feature representation is the description by relational trees to describe the relative size and placement of peaks and troughs in a waveform (Shaw & DeFiguereido 1990). Like any other kind of pattern representation, they should emphasize the underlying, more global structure of the waveforms to reduce the dimensions of problem space (i.e. the matrix size). If applied correctly, they also succeed in making the procedure more robust. The disadvantage of relational trees like all other parameter approaches is the irrevocable loss in timing accuracy which could in turn be obtained by cross correlation and which is often necessary for processing the seismograms.

The second group of DWM methods also determines some matching path but overcomes the loss in resolution by processing the original time series. Minimizing the similarity measures is substituted by searching for the cross-correlation maximum. The predictable weakness of these DWM methods is the restriction to some simple waveforms, more than 10 peaks and troughs render the matrix inversion very complex and often unstable. Martinson, Menke & Stoffa (1982) introduce a matching function which is parametrized by a sum of simple harmonic functions with unknown coefficients. Those coefficients that maximize the coherence between the two waveforms are determined by iterations based on a suitable start model (Shure & Chave 1984; Martinson, Menke & Stoffa 1984). The high frequency parts in the signal would cause a rough topography in the

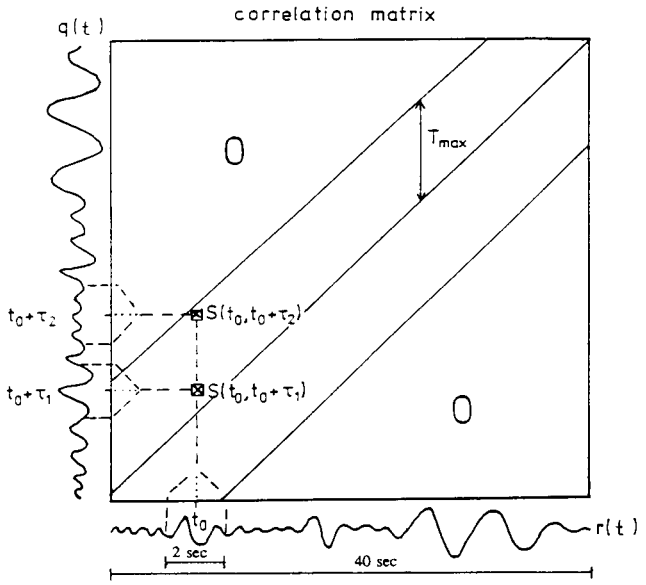

Figure 5. Layout of the similarity matrix **S**: two signals $r(t)$ and $q(t)$ are matched by correlating local similarities between short windows for different relative time shifts τ. The final path will be searched within the maximum time lag T_{max} which corresponds to the greatest expected delay.

coherence matrix, the iteration might converge to a local maximum. To avoid this effect the data are initially lowpass filtered to start iteration near the global maximum.

In our investigations we follow a slightly different approach that is based on the *multiple dynamic matching* by Leany & Ulrych (1987). Instead of one path, a whole network of possible matching solutions is determined. This makes sense when we assume that the globally correct function might locally appear suboptimal. Allowing several solutions in the first instance, one can take supplementary criteria as the boundary condition in the final path finding for any special matching problem. Their approach starts by a piecewise linear approximation of the matching function based on sliding cross correlations over the initial waveforms. Fig. 5 describes the layout of the similarity matrix. Then the local maxima or *ridge points* are extracted, adjacent ridge points are connected to *ridge trends*. Connecting all the ridge trends yields the network of global paths through the similarity matrix. Each one represents a possible correlation over the whole signal length. The matching function is linear over the short time windows, but non-linear over the whole seismogram length. Finally, some supplementary criteria are introduced to rank the most probable solutions. Obviously, these path constraints are strongly dependent on the content of the data.

The method described so far yields high timing resolution and is applicable to large sets of data due to the simple calculation of short cross-correlation segments. So the resulting similarity matrix and the ridge trend extraction are suited to match seismograms of local earthquakes, while the subsequent path evaluation must significantly differ from the lithostratigraphic matching problem treated by Leany & Ulrych (1987).

DYNAMIC WAVEFORM MATCHING OF WEAK LOCAL EARTHQUAKES

When matching sonic logs of two adjacent boreholes by DWM in exploration geophysics, the correct mapping of every peak and trough is significant because they can be associated with some lithologic boundaries. The complexity of the correlation path will depend on the variation in geology, the boundary conditions for path finding must cope with the possibility of intersections, dipping or thinning beds, pinch outs and slip faults. So very large shifts between the signals which bridge missing layers are possible. In this path finding, even 'detours' can make sense. In general, path finding has to rely more on local criteria than on global constraints; a necessary prerequisite is that there is no problem with S/N ratio.

On the contrary, when matching local seismic events in our application we already know the approximate epicentre region by sonogram detection (Joswig 1990). The stretchings and squeezings of the signals are smaller and not as complicated as in the lithostratigraphic correlation. We have rather precise and more global boundary conditions for the path-finding algorithm. Opposite to the borehole example, no great significance can be attributed to the single peaks of a noisy signal. Besides the non-linearity of the path, our main problem is the low S/N ratio which yields ambiguous peak and trough mappings in the correlation path when the program tries to match coherent phases. These uncertainties are not caused by program deficiencies, instead they reflect the principle limitation of our knowledge of signal waveforms in noise and can't be resolved in principle.

In our application, maximum timing resolution is necessary for precise event location. So for DWM we will work on the original seismograms without further transformation. Filtering is done only to suppress pure noise portions in the different seismic phases. The appropriate filter settings for P and S waves (0.5–8 Hz) versus surface waves (0.5–3 Hz) are applied to the seismograms prior to the correlation.

(a) Correlation of similar local events

By DWM two seismograms are matched by correlating local similarities between short windows but tolerating different time shift residuals on the course of the whole path. In our application we work on 100 Hz data. The epicentres are less than 50 km apart and the resulting seismograms have less than 40 s duration. The local correlations are performed on 2 s windows with centre times t_0. The shift increment for the 2-s windows in the local cross correlations is 0.01 s. The maximum time lag T_{max} is determined by the greatest shift that must be expected between coherent phases. Analysis is performed by 50 per cent overlap, that means in 1-s steps along the whole seismogram. If $r(t)$ is the master event and $q(t)$ is the actual signal to match, the normalized cross correlation is calculated by

$$s(t_0,\tau)=\frac{\sum_{t=t_0-1s}^{t_0+1s} r(t)\cdot q(t+\tau)}{\sqrt{\sum r^2(t)\cdot\sum q^2(t+\tau)}};\qquad |\tau|<T_{max}, \tag{1}$$

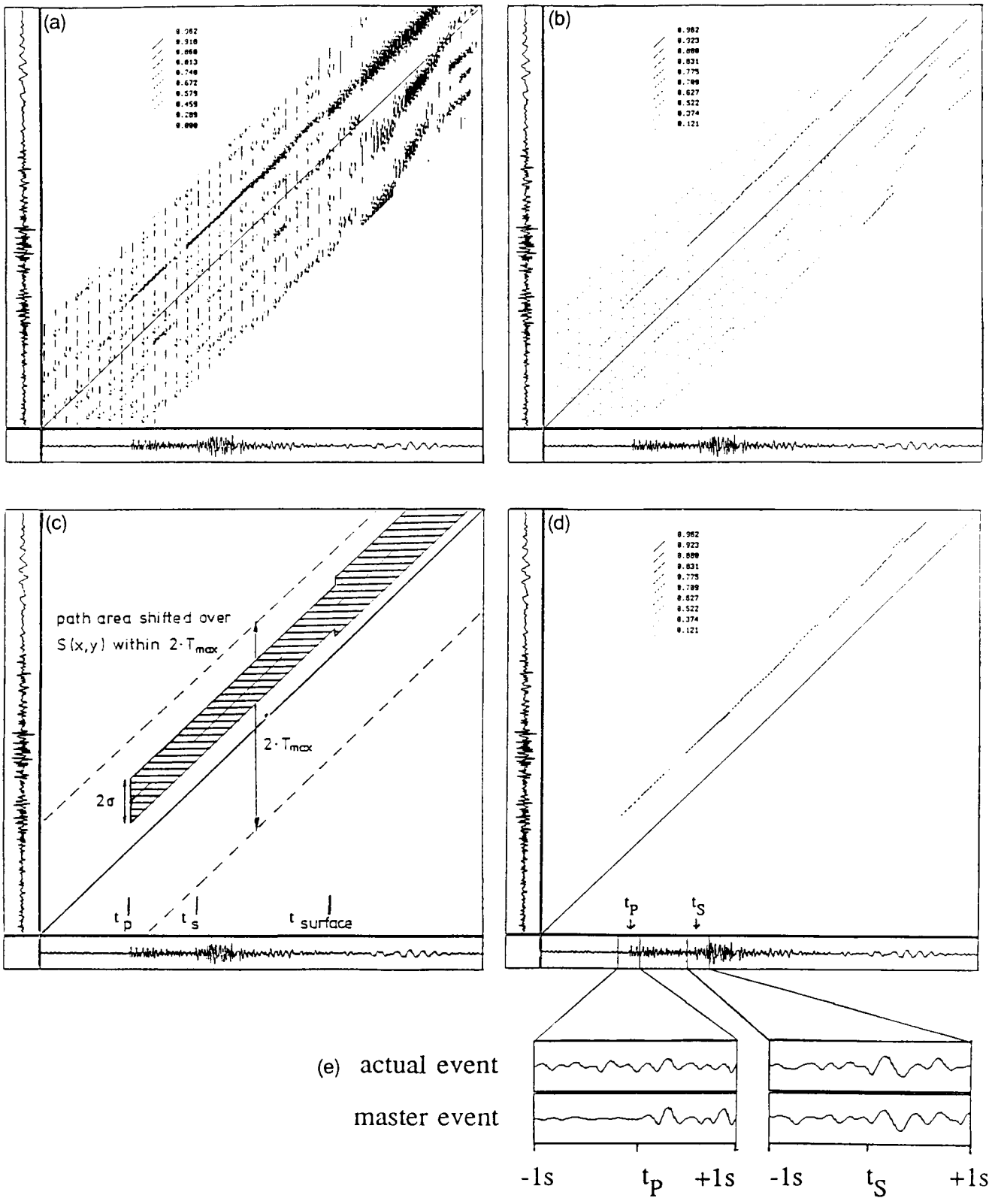

Figure 6. DWM for master-event correlation: (a) similarity matrix **S** for the two Hamm events of Fig. 3, the horizontal trace shows the master event. The 40 * 4000 matrix is expanded to quadratic shape by displaying each pixel of **S** as diagonal. (b) Matrix of ridge trends derived by eq. (3)–(5). (c) The theoretically expected path guides the search for the final matching path that contains the greatest sum of single correlation values. (d) Extracted find correlation path. In (e) the corresponding time windows are shown, the matching is exact within one sample of the initial time series. The windows show the *P*- and *S*-phase correlation.

which is arranged into the similarity matrix **S** of Fig. 5 via

$$S(x, y) = \begin{cases} s(x, y - x) & \text{if } |y - x| < T_{\max} \text{ and } s(x, y - x) > 0 \\ 0 & \text{else} \end{cases} . \quad (2)$$

With the x-axis sampled every second (half the window length) and τ being incremented by 0.01 s, we get the matrix dimensions $40 * 4000$ with a sparse distribution around the main diagonal. For the data display on computer screen the matrix is expanded to quadratic scale with different values of the cross correlation represented by nine colours. Fig. 6(a) shows the screen-dump reproduction by nine grey levels for the two similar *Hamm* events of Fig. 3. The horizontal trace is the reference seismogram of the master event, the vertical trace is the seismogram of the weak event. $T_{\max} = 2$ s, reflecting the assumed uncertainty in P-onset timing of the weak event in a fully automated system.

The next step in DWM is to determine the ridge points which make up the lines of ridge trends. The ridge points $R(x, y)$ are matrix values $S(x, y)$ that are above a given threshold and that are greater than their three immediate neighbour values

$$R(x, y) = \begin{cases} S(x, y) & \text{if } (I) \text{ and } (II) \text{ hold} \\ 0 & \text{else} \end{cases} \quad (3)$$

$$\text{Condition } (I): \begin{cases} S(x, y) > S(x, y + 0.01s) \text{ and} \\ S(x, y) > S(x, y - 0.01s) \text{ and} \\ [S(x, y) > S(x + 1s, y + 1s) \text{ or} \\ S(x, y) > S(x - 1s, y - 1s)] \end{cases} \quad (4)$$

$$\text{Condition } (II): \begin{cases} S(x, y) > 0.12 & \text{if body waves} \\ S(x, y) > 0.3 & \text{if surface waves} \end{cases} . \quad (5)$$

The resulting matrix **R** is shown in Fig. 6(b). It contains the relevant information when searching the best non-linear correlation between two seismograms. Since we expect a correlation path parallel to the main diagonal, left and right neighbours in eq. (4) are chosen in respect to the diagonal. The threshold in eq. (5) differs for body-versus-surface waves, the latter have higher correlation values because of the lower frequencies.

For path finding, we need some kind of *a priori* information about predictable path trends which is displayed in Fig. 6(c). We know P- and S-onset times of the master event exactly while assuming an uncertainty of ± 2 s for the detected P-onset time of the weak event. Since we cannot predict the location of the actual epicentre relative to the master event, the path variance must be assumed to be symmetric around a diagonal. All paths that start within the P-onset column of the master event and that lie within the appropriately shifted variance area must be considered. The path which yields the highest sum of correlation values is regarded as the best path. Once this final path is found, waveform similarity is normalized with respect to the 'ideal' path by

$$\text{sim} = \frac{\sum_{\text{path}} R_{r(t), q(t)}}{N} , \quad (6)$$

where N is the total number of correlations within the path. Fig. 6(d) shows the final correlation path for our example. It maps the peaks and troughs of both waveforms with an accuracy of one sample, so Fig. 6(e) can display the aligned 2 s windows of P and S onsets. Because of the extremely low S/N ratio, independent phase picking on the upper trace would not have been possible.

Finally for the cluster association we define an empirical similarity threshold. We found a value of sim = 0.35 with minor corrections for the different epicentre regions. Thus an actual event is compared to all the master events of its particular region, it is associated to the most similar one if this similarity measure exceeds the given threshold.

(b) Correlation of array traces

Figure 7(a) shows the similarity matrix **S** for the two BUG array traces KLB and TEZ when analysing the one *Hamm* event of Fig. 4. Once again the horizontal trace is the 'master trace' of KLB, the vertical one is TEZ. In case of array trace correlation, $T_{\max}$ must cover the difference of the lowest waves over the array aperture; for the 2 km maximum extension of BUG a limit of 1 s was chosen. Fig. 7(b) shows the matrix of ridge trends.

For correlating different array traces, the *a priori* information for pathfinding consists of the rough epicentral region derived by pattern recognition and the theoretical traveltime delays for dominant phases. Thus the path deviation from the main diagonal can be predicted here for any station combination, resulting in unsymmetric path tolerances [Fig. 6(c)]. Especially the difference in P-onset time is known very well while the absolute onset time is determined only within the error limits of the previous detection process. Therefore, when searching for the final path we must shift the onset of the given variance area parallel to the main diagonal now.

As for the master events, all possible paths are considered and the best one is chosen [Fig. 7(d)]. The distance between array stations KLB and TEZ is 1 km, which is an order of magnitude larger than the distance between two epicentres in Fig. 6, so the correlation path disintegrates clearly into P, S and surface waves. Fig. 7(e) shows the resulting array trace correlations for P and S phases.

RESULTS OF THE AUTOMATED MASTER-EVENT CORRELATION

Master-event correlation by DWM was developed as a software module within a fully automated process of earthquake analysis. In order to test the capabilities of DWM alone, we chose to process observatory data that was initially analysed by seismologists. We tried to relocate all the mining-induced seismicity for four year period (1987–1990) of the epicentre region *Hamm* which is 40 km NE of the BUG array. A second data set includes all explosions of the quarries near Wuppertal 25 km SSW of BUG in 1990. For both areas we can verify the master-event association either by the 3C-station HRH which is at 4 km distance of the *Hamm* cluster or by checking the limestone pit locations and their blast reports for *Wuppertal*. The DWM master-event correlation amounts to 470 events times 12 master traces for *Hamm* and 175 explosions times 22 master traces for the quarries, i.e. some 10 000 comparisons

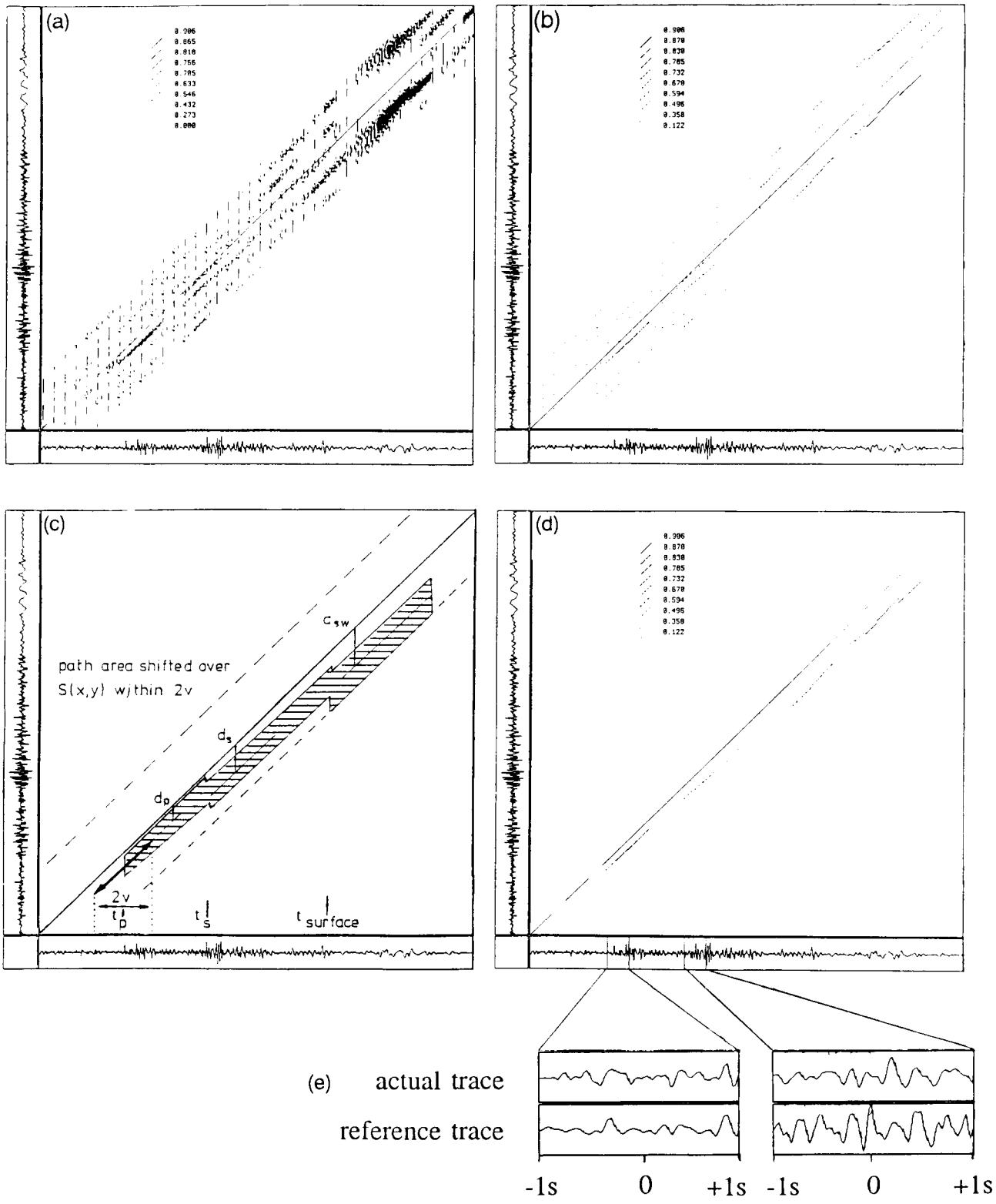

Figure 7. DWM for array trace correlation: (a) similarity matrix **S** as in Fig. 6, this time for the seismograms of Fig. 4. The horizontal registration of KLB acts as the reference trace. (b) Matrix of ridge trends. (c) Theoretically expected path. The significant shifts within the path depend on the difference phase velocities over the array aperture. For different epicentre regions, different path shifts occur. (d) Extracted path for the actual data. (e) Correlation windows for *P* and *S* phases.

Table 1. Table of master-event correlation for the *Hamm* clusters: for the shown time period of 1989, only four clusters were active and some events could be associated to the corresponding master events. For each one selected, the determined relative shift of $S-P$ onsets is given also (with one sample resolution). Note the large capture domain which goes far beyond the limits of the $\lambda/4$ criterion.

master events:	Hamm1		Hamm4		Hamm7		Hamm9	
	11.03.1987		27.05.1988		06.01.1989		15.12.1989	
	00:44:58:72		20:02:27:84		19:11:27:85		09:31:13:19	
test (s) events	sim_{cq}	Δt (10 ms)	sim_{cq}	Δt (10 ms)	sim_{cq}	Δt (10 ms)	sim_{cq}	Δt (10 ms)
[illegible]	0.66	-4	0.26		0.31		0.30	
[illegible]	0.37		0.26		0.76	0	0.24	
[illegible]	0.26		0.39		0.16		0.18	
[illegible]	0.18		0.19		0.23		0.22	
[illegible]	0.63	-1	0.26		0.33		0.27	
[illegible]	0.29		0.21		0.22		0.24	
[illegible]	0.60	+3	0.29		0.37		0.26	
[illegible]	0.65	+3	0.24		0.40		0.29	
[illegible]	0.60	0	0.23		0.32		0.29	
[illegible]	0.43	+4	0.24		0.27		0.23	
[illegible]	0.48		0.32		0.77	-1	0.28	
[illegible]	0.22		0.19		0.22		0.19	
[illegible]	0.27		0.25		0.22		0.22	
[illegible]	0.17		0.20		0.19		0.23	
[illegible]	0.26		0.36	-2	0.25		0.26	
[illegible]	0.31		0.44	-1	0.25		0.24	
[illegible]	0.44	+3	0.26		0.33		0.30	
[illegible]	0.13		0.11		0.08		0.14	
[illegible]	0.24		0.50	+2	0.22		0.20	
[illegible]	0.26		0.45	-2	0.20		0.35	
[illegible]	0.21		0.17		0.15		0.13	
1989/06/26/08:29	0.24		0.24		0.28		0.21	
1989/08/10/03:27	0.31		0.25		0.29		0.22	
1989/08/23/12:29	0.44		0.27		0.78	+3	0.26	
1989/09/03/12:11	0.22		0.31		0.22		0.21	
1989/09/15/21:49	0.20		0.64	+3	0.21		0.20	
1989/09/20/19:10	0.26		0.74	+1	0.26		0.28	
1989/09/20/19:31	0.34	-2	0.30		0.28		0.19	
1989/09/26/01:59	0.19		0.42	+3	0.25		0.26	
1989/09/29/21:50	0.30		0.68	+4	0.18		0.25	
1989/10/02/08:59	0.29		0.27		0.24		0.47	+1
1989/10/05/01:44	0.21		0.65	-2	0.27		0.26	
1989/10/06/02:01	0.28		0.70	0	0.22		0.27	
1989/10/06/19:22	0.23		0.63	+3	0.28		0.25	
1989/10/07/03:13	0.23		0.65	+3	0.27		0.29	
1989/10/07/10:38	0.16		0.67	+5	0.16		0.27	
1989/10/09/05:23	0.29		0.23		0.20		0.29	
1989/10/09/21:15	0.24		0.75	0	0.25		0.24	
1989/10/10/13:57	0.28		0.60	-4	0.24		0.23	
1989/10/11/01:41	0.26		0.65	+3	0.26		0.24	
1989/10/11/11:24	0.23		0.53	+4	0.20		0.19	
1989/10/11/21:56	0.17		0.70	+3	0.33		0.22	
1989/10/13/01:31	0.25		0.22		0.29		0.23	
1989/10/13/06:31	0.25		0.64	-2	0.23		0.25	
1989/10/13/23:28	0.19		0.25		0.20		0.18	
1989/10/14/10:31	0.29		0.72	+5	0.29		0.21	
1989/10/17/07:03	0.19		0.62	+2	0.21		0.20	
1989/10/17/14:22	0.20		0.62	+4	0.26		0.29	
1989/10/17/23:51	0.23		0.75	-3	0.24		0.27	
1989/10/19/01:33	0.25		0.68	0	0.26		0.24	
1989/10/19/17:41	0.27		0.68	-8	0.21		0.21	
1989/10/20/02:14	0.21		0.67	0	0.22		0.31	
1989/10/20/16:24	0.21		0.71	0	0.18		0.26	
1989/10/20/17:31	0.26		0.32		0.25		0.41	+8
1989/10/20/19:38	0.25		0.67	+4	0.26		0.18	
1989/10/21/10:24	0.19		0.72	-2	0.23		0.19	

that can only be handled by an automated approach for an event evaluation.

Table 1 shows the results of event comparison for *Hamm* in 1989. Only the four clusters significantly active at that time are listed. For each correlation, the table contains *sim* from eq. (6). A similarity threshold of 0.33 was chosen empirically. Above this value, a distinct path can be seen within the similarity matrix, so the deviation of t_S-t_P to the master event (Δt, in 10 ms) is reported, too. In Fig. 8, 31 P codas (6s) from cluster *Hamm4* are shown, with aligned P-onset times and the master event on top. This example is intended to demonstrate the power of DWM which can correlate even weak events with extremely low S/N ratios. Prior to the cluster association, some of these plots were also used by the seismologist to display events from one region when searching for the master events. Until now, this step has to be performed interactively by visual inspection, i.e. master events are determined empirically from a restricted set of stronger events, while the much greater number of weak events have to be excluded.

Figure 9 shows the 94 epicentres of *Hamm* in 1989. In Fig. 9(a) the bulletin results from the BUG small array, obtained by interactive analysis, are displayed, the uncertainty of ±2°

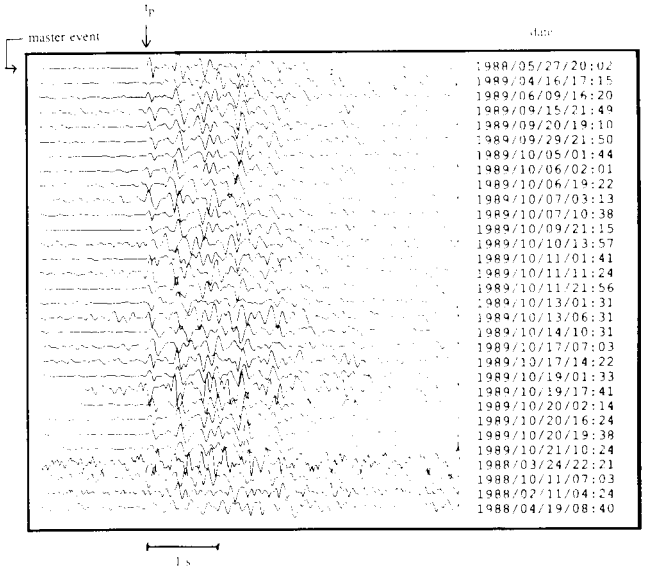

Figure 8. 24 Seismograms of cluster *Hamm4*: the master event is shown above, the other traces are associated and aligned automatically showing the good performance of DWM for poor S/N ratio. Such plots also guided the interactive determination of master events by the seismologist.

in azimuth (i.e. ±2.8 km) and ±1 km in distance is too large to resolve any detailed structure. In Fig. 9(b) the 76 events that could be associated to one of the four master events by DWM were relocated according to the t_S-t_P deviation. This step is performed on the base of seismograms from the array station KLB only, which is situated in 40 km distance of the Hamm clusters. These results can be compared to the epicentre locations in Fig. 9(c) determined interactively from the data of the station HRH in 4 km distance. Taking the uncertainty of ±200 m into account, the clusters can also be resolved and the association by DWM in (b) can be confirmed.

In Fig. 10 the cumulative history of cluster activities in 1987–1990 is given. As for the previous figure, some events remain unassociated either because they were indeed single events or because their S/N ratio was too low or because the similarity was just below the threshold. While the absolute number of events has increased, the ratio remained stable with 80 per cent or 379 from 470 events that could be associated to one of the 12 master events within the four years. Apart from the significant spatial clustering, a temporal clustering can be resolved as well. In general, the clusters extend over two months with single pre- and aftershocks. The distribution of seismic activity is not uniform over the year, we can recognize event accumulations and very calm months. But coal mining near Hamm is performed continuously over the years; obviously, there is only indirect interdependency between coal mining and the mining-induced seismicity.

Figure 11 shows a map of the limestone-pit region at 25 km distance from BUG where Fig. 11(a) displays the bulletin results obtained interactively by the small array. Data of 175 explosions were available for 1990, the assumed location uncertainty is ±1.7 km in azimuth and ±1 km in distance. The quarry blasts with yield of 2–4 tons were

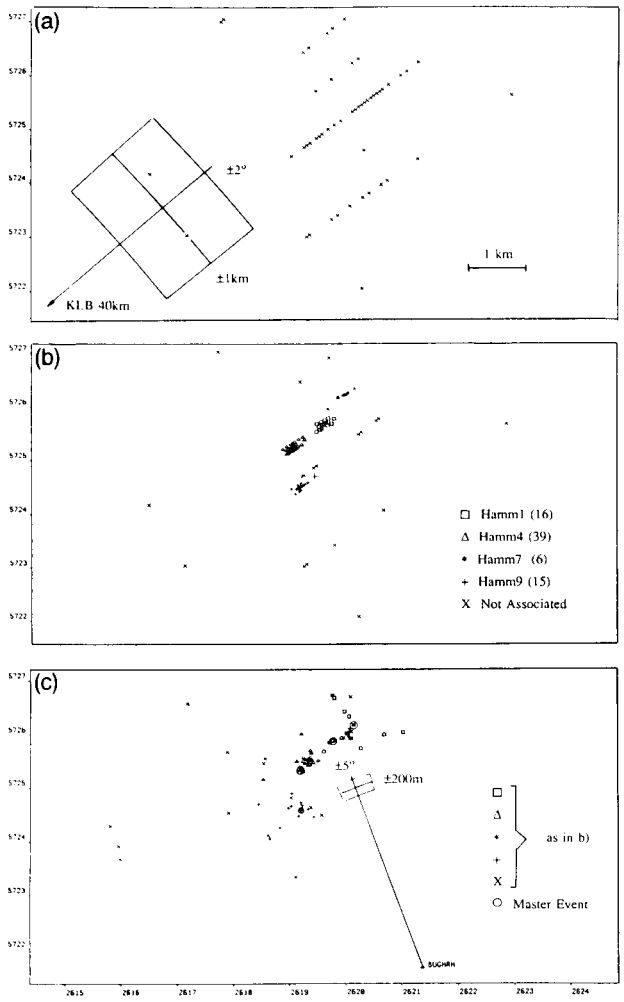

Figure 9. Epicentre map of the region *Hamm:* (a) the interactive results of BUG-array evaluation do not show any significant clustering. The aligned structures are due to the principal uncertainties in location which are ±2° for the azimuth and ±1 km for the distance. (b) Improvement of location by DWM for KLB master-event association (in 40 km distance). Some events remain unassociated, their epicentre estimate remained unchanged. In (c)

carried out in four different pits; for each pit 4–7 master events were chosen in the DWM run. The path finding was complicated by the fact that for quarry blasts the correlation of *S* phases was significantly worse than the *S*-phase correlation of Hamm events. Nevertheless, 138 events or 79 per cent could be associated using a threshold of 0.4. The success rate of correct associations was 85 per cent or 117 events. Characterizing each pit by more than one master event agrees with the results reported by Thorbjarnardottir & Pechman (1987) and Harris (1991). We found no hint for clustering in time. When the epicentre locations in (a) are compared to actual limestone pit locations in Fig. 11(c), we have to admit some systematic deviation due to inhomogeneities under the array aperture and due to the propagation path. Event association by DWM in Fig. 11(b) allows a determination of correct relative distances if the master events are adjusted according to the blast reports. Once again this improvement was achieved on the base of KLB seismograms only.

In Fig. 9 and Fig. 11, we determined the location of actual events relative to their similar master event. The basic assumption was vicinity, its obvious limit would be the $\lambda/4$ radius of ±100 m for 8 Hz as the upper cut off in frequency prefiltering and $v_P = 3.4\,\mathrm{km\,s^{-1}}$. On the other hand, we observe the path deviation between *P* and *S* phases of master and actual event which yields a radial residual

$$\Delta r = \frac{v_S v_P}{v_P - v_S} v_P[(t_S - t_P)_{\mathrm{actual}} - (t_S - t_P)_{\mathrm{master}}]$$

$$= \frac{v_P}{\sqrt{3}-1} v_P[(t_S - t_P)_{\mathrm{actual}} - (t_S - t_P)_{\mathrm{master}}] \qquad (7)$$

with $v_P/v_S = \sqrt{3}$. The given 100 m limit taken in radial direction would correspond to a maximum shift of 20 ms while Table 1 reports correlations of up to 80 ms difference which indicates a much larger capture area of master events by DWM. When Thorbjarnardottir & Pechman (1987) derived their constraints on relative earthquake location, they performed the correlation over the whole *P* and *S* phase and just neglected eq. (7). If we assume that critical degradation in their correlation results is caused by the $\lambda/4$ shift of both phases, we get as the upper frequency limit of two events with distance *d*

$$\frac{1}{4f_u} = \frac{d}{v_S} - \frac{d}{v_P}. \qquad (8)$$

For our data, we get $d = 150$ m by 8 Hz cut off, $v_P = 3.4\,\mathrm{km\,s^{-1}}$ and $v_S = 2.0\,\mathrm{km\,s^{-1}}$. This distance is of the same magnitude as the above $\lambda/4$ limit. Thorbjarnardottir & Pechman attribute the decay of cross correlation to waveform changes by source area inhomogeneities, however, it could also be caused by the delay between *P* and *S* phase (Spudich & Bostwick 1987). Therefore, a correct waveform correlation should either be restricted to one type of seismic wave or performed by DWM to yield larger similarities. However, it was not the scope of this paper to revisit the $\lambda/4$ criterion since investigations for this purpose will demand seismograms with good *S*/*N* ratio. We focus on weak events where the noise is significant and will introduce a distance threshold that cannot be lowered because of the limited similarity in the signals. When relocating our associated events, we thus follow the more exact eq. (7) for the radial distance but rely on the $\lambda/4$ of 8 Hz in DWM for the uncertainty in azimuth.

RESULTS OF THE AUTOMATED ARRAY TRACE CORRELATION

In the second application, we determine azimuth and slowness for single events by a correlation of their array traces. No additional master events are needed, in each case the trace with the best *S*/*N* ratio acts as the reference seismogram. The input to DWM is the approximate epicentral region characterized by Fig. 12(b). It was derived for the off-line test by the exact bulletin data of BUG that are displayed in Fig. 12(a); the *P*-phase onset was expected to be known within ±2 s. Both assumptions, derived from previous test run results (Joswig 1992b), reflect a system performance that can be expected from Sonogram detection with subsequent automatic coincidence evaluation. The *a*

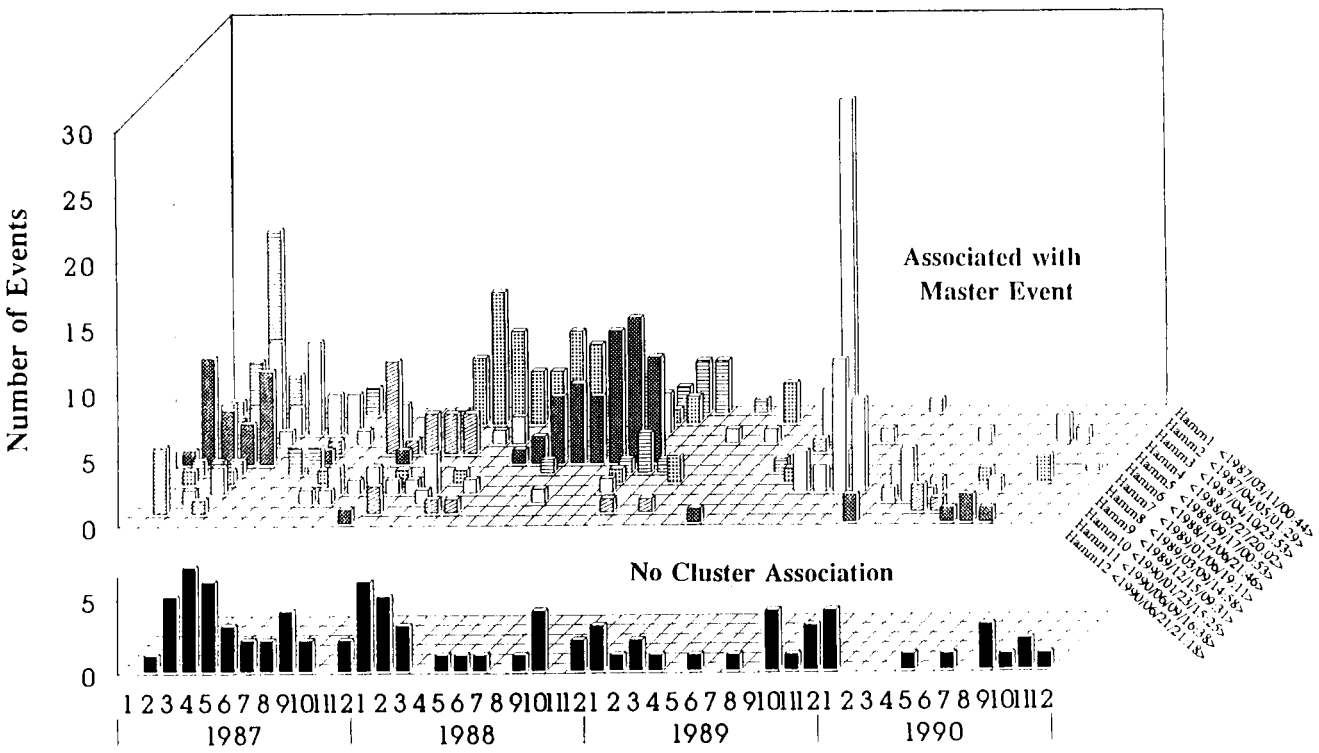

Figure 10. Clustering in the *Hamm* region over time: for the mining-induced seismicity, a clear temporal clustering can be resolved besides the spatial clustering of Fig. 9. However, this effect can not be attributed to mining activities which were constant over the years.

priori information for DWM consists of expected path shifts like the one shown in Fig. 7(c), derived for each region and each pair of station combinations KLB-SHA and KLB-TEZ.

The test run was performed on all earthquakes in 1989 originating from the epicentre regions *Ahlen*, *Hamm*, *Kamen* and *Moers* [see Fig. 1 and Fig. 12(a)] which produce half of the mining-induced seismicitiy in the Ruhr Basin. From these 246 events, azimuth and slowness could be calculated by DWM for 187 events. All results were correct within $\pm 2^{\circ}$ and ± 0.05 s km^{-1}. From the rest, 22 events were not evaluated because of corrupted data structure, while in 37 cases DWM did not succeed. This unsuccessful performance was mainly caused by errors for KLB-SHA (30 cases), the correlation KLB-TEZ failed once and both correlations missed six times.

In a fully automated system, comparing array traces by DWM will improve our *a priori* knowledge of azimuthal resolution but will not refine the *S*-phase picking. The shift from *P* to *S* phase is of course found in the correlation path but the absolute onset time is only exact for ± 1 s, i.e. half the correlation window. This result is similar to the resolution which has already been achieved by the Sonogram detector, so other approaches like a three-component *S*-phase detector must be used. Therefore we display the 187 epicentres of our off-line test in Fig. 12(c) based on the automated azimuths by DWM, but with distances determined by interactive analysis. The association of events to distinct epicentre regions is once again obvious, but as opposed to the master-event correlation, no cluster association can be performed within the epicentre regions.

DISCUSSION AND CONCLUSIONS

Automatic evaluation of local events by DWM and path finding appears to be most suited for weak events which cannot be evaluated individually because of ambiguous phase identifications. Both methods, array-trace correlation and master-event correlation, have been developed for fine structure analysis of local seismic data. They are knowledge-based method being integrated in an automated earthquake-processing system and dependent on *a priori* information to achieve the necessary robustness.

By master-event comparison of seismograms from one station, a fine structure of clusters can be resolved within different epicentre regions. Even event correlation of very weak events yields significant correlation paths and an unequivocal phase correlation for the whole seismogram. Traditional interactive methods could only resolve a similar cluster fine structure by using additional stations close to the epicentre region. In this respect, automatic evaluation by DWM is superior to the interactive evaluation of isolated events as it is routinely used in the BUG array. However, good results are crucially dependent on a suitable and complete set of master events which requires much experience in its compilation.

For path finding, the master-event comparison requires other constraints than the array-trace correlation, while both methods must be tuned to handle data with low *S*/*N* ratio. This is only possible on the basis of *a priori* information by predictable path trends—see Fig. 6(c) and Fig. 7(c). The *a priori* information can be taken from bulletin listings in case of post processing or it must be derived on-line from the

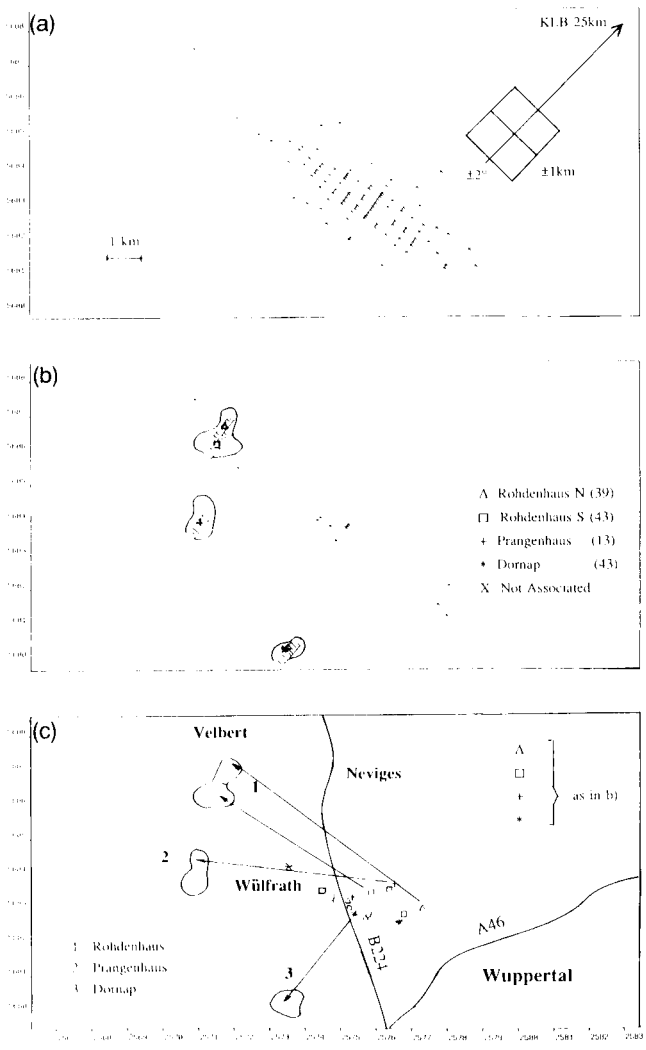

Figure 11. Epicentre map of the quarry blast region near *Wuppertal:* (a) BUG-array locations show an error that is significantly larger than the theoretical uncertainty. It must be attributed to local inhomogeneities on the path and within the array aperture. (b) By master-event correlation of KLB traces most of the events can be associated once the master events are adjusted correctly. (c) Each quarry blast will demand several master events, their initial location in (a) was random and did not show any structure.

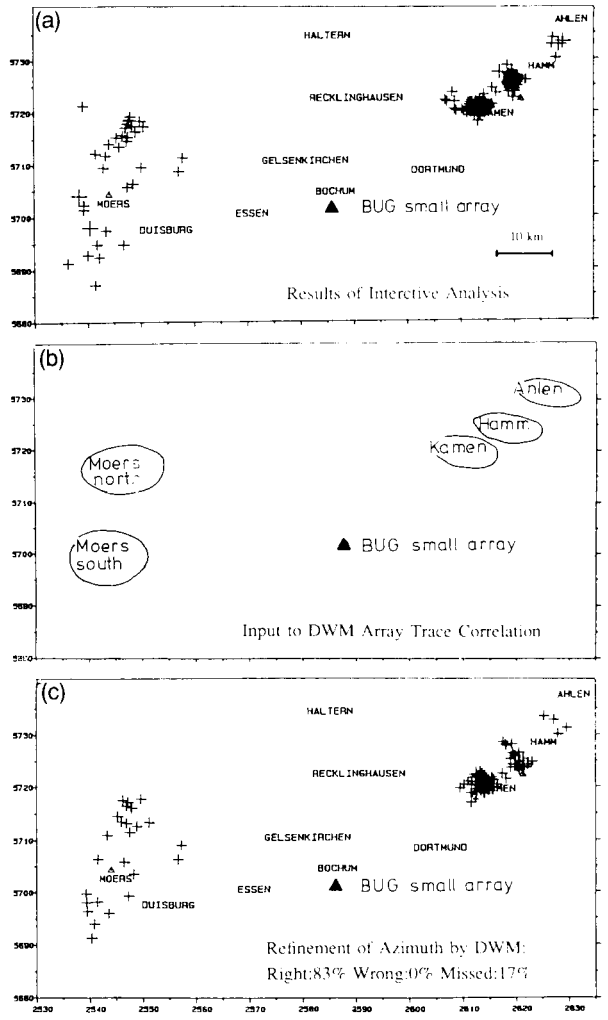

Figure 12. Epicentre map of the Ruhr Basin: (a) location of all events in 1989 based on BUG-array aximuths and distances determined interactively by the observer. (b) DWM on array traces needs the *a priori* information of rough epicentre regions to succeed for low *S/N* ratio. (c) The refinement of azimuths by DWM does not show any single mistake although some events remain uncorrelated. Because DWM does not improve the distance estimate here, the distances in (c) were taken from (a) to get a display by map once again.

incoming data in a completely automated system (Joswig 1992a). DWM, in this environment, is designed and optimized for a subsequent, fine-grain analysis with maximum possible timing resolution and good discrimination on cluster association. DWM can distinguish between 10 to 20 master events in an epicentral region where the initial Sonogram detector will demand and resolve just one pattern.

Of course, DWM can also be designed to work without *a priori* knowledge. Like in the applications for exploration geophysics, local path constraints must then be emphasized more than global constraints, thus excluding all the events with low *S/N* ratio from automated processing. This is exactly what we did not want to do, because small events are not the neglectable side effect in automatization, instead they make up the major part of earthquake activity monitored by any seismic network and dominate all routine works in the observatory.

Determining the beam azimuth by trace matching in a small array, DWM gives high success rates as well. Even for weak events, a correct automatic azimuth determination for the different epicentre regions can be done. One advantage is that no additional master events are needed. However, no cluster fine analysis is possible using the method separately. DWM separates the joint trace matching into distinct correlation pairs. This introduces possible weaknesses as by the (KLB/SHA) pair for BUG data, while the seismologist handles this situation by optimizing on all traces simultaneously. The connection realized in this paper by onset-time adjustment is fairly primitive. More advanced procedures should be developed to utilize the interdependency of correlation pairs better and to improve the evaluation rate for weak events even more.

ACKNOWLEDGMENT

This work was supported by Deutsche Forschungsgemeinschaft under grant Ha 842/8-1.

REFERENCES

Anderson, K. R. & Gaby, J. E., 1983. Dynamic waveform matching, *Inform. Sci.*, **31,** 221–242.

Archuleta, R. J., Cranswick, E., Mueller, C. & Spudich, P., 1982. Source parameters of the 1980 Mammoth Lakes, California earthquake sequence, *J. Geophys. Res.*, **87,** 4595–4607.

Casten, U. & Cete, A., 1980. Induzierte Seimizität im Beriech des Steinkohlebergbaus des Ruhrreviers, *Glückauf-Forschungshefte*, **41,** 1, 12–16, Bochum.

Dewey, J. W., 1972. Seismicity and tectonics of Western Venezuela, *Bull. seism. Soc. Am.*, **62,** 1711–1751.

Fitch, T. J., 1975. Compressional velocity in source regions of deep earthquakes: an application of the master event technique, *Earth planet. Sci. Lett.*, **26,** 156–166.

Frankel, A., 1982. Precursors to a magnitude 4.8 earthquake in the Virgin Islands: spatial clustering of small earthquakes, anomalous focal mechanisms, and earthquake doublets, *Bull. seism. Soc. Am.*, **72,** 1277–1294.

Frankel, A. & Kanamori, H., 1983. Determination of rupture duration and stress drop for earthquakes in Southern California, *Bull. seism. Soc. Am.*, **73,** 1527–1551.

Geller, R. J. & Mueller, C. S., 1980. Four similar earthquakes in central California, *Geophys. Res. Lett.*, **7,** 821–824.

Gibowicz, S. J., Harjes, H.-P., & Schäfer, M., 1990. Source parameters of seismic events at Heinrich Robert mine, Ruhr Basin, Federal Republik of Germany: evidence for non-double-couple events, *Bull. seism. Soc. Am.*, **80,** 88–109.

Hanks, T. C., 1982. f_{max}, *Bull. seism. Soc. Am.*, **72,** 1867–1879.

Harris, D. B., 1991. A waveform correlation method for identifying quarry explosions, *Bull. seism. Soc. Am.*, **81,** 2395–2418.

Hinzen, K.-G., 1982. Source paramers of mine tremors in the eastern part of the Ruhr-district (West Germany), *J. Geophys.*, **51,** 105–112.

Ishida, M. & Kanamori, H., 1978. The foreshock activity of the 1971 San Fernando earthquake, California, *Bull. seism. Soc. Am.*, **68,** 1265–1279.

Israelsson, H., 1990. Correlation of waveforms from closely spaced regional events, *Bull. seism. Soc. Am.*, **80,** B6, 2177–2193.

Ito, A., 1985. High resolution relative hypocenters of similar earthquakes by cross-spectral analysis method, *J. Phys. Earth*, **33,** 279–294.

Joswig, M., 1990. Pattern recognition for earthquake detection, *Bull. seism. Soc. Am.*, **80,** 170–186.

Joswig, M., 1992a. Knowledge-based earthquake processing in sparse arrays, in *Application of Artificial Intelligence in Seismology*, Cahiers du Centre Européen de Géodynamique et de Séismology, **6,** Luxembourg.

Joswig, M., 1992b. Single-trace detection and array-wide coincidence association of local earthquakes and explosions, *Comput. Geosci.*, in press.

Le, L. & Nyland, E., 1990. Pattern analysis of seismic records, *Geophysics*, **1,** 20–88.

Leany, W. S. P. & Ulrych, T. J., 1987. Multiple dynamic matching: a new approach to well log correlation, *Geoexploration*, **24,** 503–515.

Liu, Hsi-Ho & Fu, K.-S., 1983. An application of syntactic pattern recognition to seismic discrimination, *IEEE Trans. Geosci. Rem. Sens.*, **GE-21,** 2, 125–132.

MacBeth, C. D. & Redmayne, D. W., 1989. Source study of local coalfield events using the modal synthesis of shear and surface waves, *Geophys. J. Int.* **99,** 155–172.

Martinson, D. G., Menke, W. & Stoffa, P., 1982. An inverse approach to signal correlation, *J. Geophys. Res.*, **87,** 4807–4818.

Martinson, D. G., Menke, W. & Stoffa, P., 1984. Reply to L. Shure and A. D. Chave: "Comments on 'An inverse approach to signal correlation' by D. G. Martinson, W. Menke and P. Stoffa", *J. Geophys. Res.*, **89,** 2501–2504.

Motoya, Y. & Abe, K., 1985. Waveform similarity among foreshocks and aftershocks of the October 18, 1981, Eniwa, Hokkaido, earthquake, *Earth. Predict. Res.*, **3,** 627–636.

Myers, C. S. & Rabiner, L. R., 1981. A comparative study of several dynamic time-warping algorithms for connected-word recognition, *Bell System Tech. J.*, **60,** 1389–1409.

Pechman, J. C. & Kanamori, H., 1982. Waveforms and spectra of preshocks and aftershocks of the 1979 Imperial Valley, California, earthquake: evidence for fault heterogeneity? *J. Geophys. Res.*, **87,** B13, 10 579–10 597.

Poupinet, G., Ellsworth, W. L. & Frechet, J., 1984. Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California, *J. Geophys. Res.*, **89,** B9, 5719–5731.

Scherbaum, F. & Wendler, J., 1986. Cross spectral analysis of Swabian Jura (SW Germany) three-component microearthquake recordings, *J. Geophys.*, **60,** 157–166.

Shaw, S. W. & DeFigueiredo, R. J. P., 1990. Structural processing of waveforms as trees, *IEEE ASSP*, **38,** 328–338.

Shure, L. & Chave, A. D., 1984. Comments on 'An inverse approach to signal correlation' by D. G. Martinson, W. Menke and P. Stoffa, *J. Geophys. Res.*, **89,** B4, 2497–2499.

Spence, W., 1980. Relative epicentre determination using *P*-wave arrival-time differences, *Bull. seism. Soc. Am.*, **70,** 171–183.

Spudich, P. & Bostwick, T., 1987. Studies of the seismic coda using an earthquake cluster as a deeply buried seismograph array, *J. Geophys. Res.*, **92,** B10, 10 526–10 546.

Stauder, W. & Ryall, A., 1967. Spatial distribution and source mechanism of microearthquakes in Central Nevada, *Bull. seism. Soc. Am.*, **57,** 1317–1345.

Thorbjarnardottir, B. S. & Pechman, J. C., 1987. Constraints on relative earthquake locations from cross correlation of waveforms, *Bull. seism. Soc. Am.*, **77,** 1626–1634.

VanDecar, J. C. & Crosson, R. S., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross correlation and least squares, *Bull. seism. Soc. Am.*, **80,** 150–169.