

## CLUSTERING AND LOCATION OF MINING INDUCED SEISMICITY IN THE RUHR BASIN BY AUTOMATED MASTER EVENT COMPARISON BASED ON DYNAMIC WAVEFORM MATCHING (DWM)

HARTWIG SCHULTE-THEIS and MANFRED JOSWIG\*

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

**Abstract**—Most of the local seismicity in the Ruhr Basin can be separated into characteristic clusters of similar, mining induced earthquakes. Each cluster can be represented by a strong master event. Therefore, it is possible to associate weak events to the corresponding clusters by master event comparison. The seismic signal matching is performed by a nonlinear correlation termed DWM for the entire seismogram length. DWM permits stretchings and shortenings between the two signals and overcomes the ambiguities in phase correlation by a consistent matching path.

The automatic cluster association searches for the best DWM-correlation between the actual event and all master events of the appropriate epicenter region. Knowing the P- and S-onsets of the master event, they can be transposed to the actual event by the correlation path with one sample accuracy. The method has been applied to all BUG small array recordings 1987–1990 of local events from the Hamm-region to investigate spatial and temporal clustering. Within the clusters, a high percentage of weak events could be located relative to its master event. The temporal clustering resolved seismic activities that typically last a few months per cluster, but single aftershocks occur in the following years.

*Key Words:* Master event technique, Cluster analysis, Nonlinear correlation, Automated epicenter determination.

### INTRODUCTION

By the BUG small array (Joswig, 1993), a location consists of determining azimuth and distance; the hypocentral depths of the mining induced seismicity are within some 500 m and cannot be resolved. Azimuths are calculated from time differences between coherent P phases in the tripartite array, whereas distances are obtained by  $t_s - t_p$ . Additionally, the location can be refined around the stations HRH and RPM by the azimuth of particle motion and a second distance constraint by  $t_s - t_p$ .

Although for strong events the phase picking is simple, the location of weak events is difficult because higher level of noise causes some ambiguities in the phase identification. However, weak events make up the significant part of the event distribution, hence an automated location of weak earthquakes would be attractive.

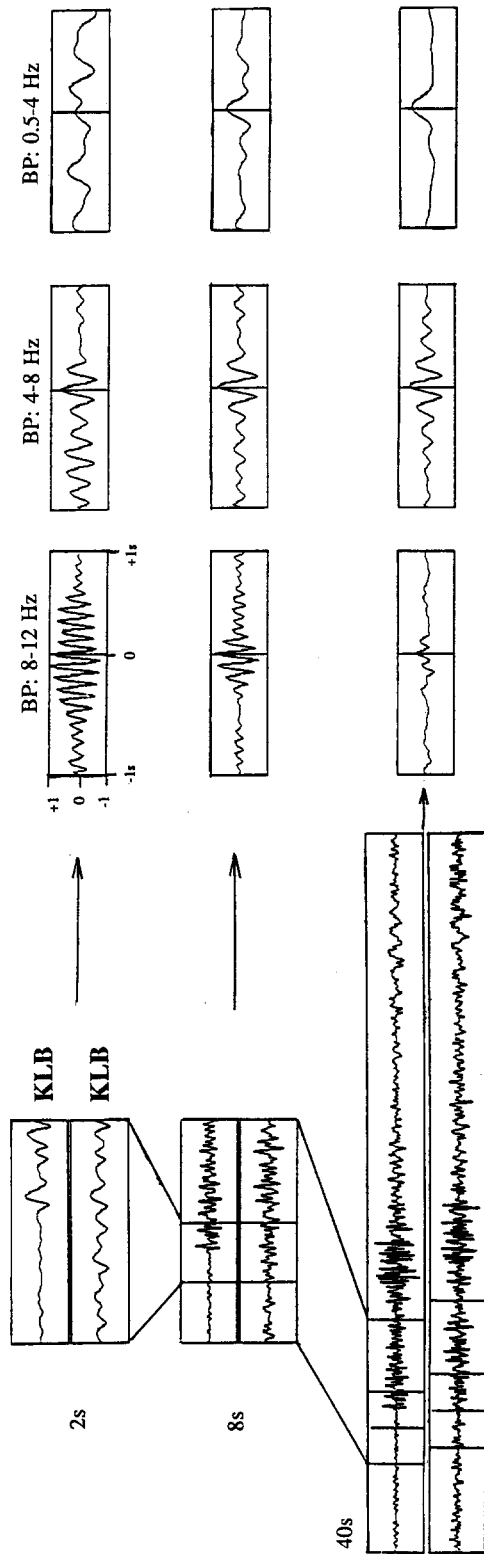
One way is to evaluate weak local earthquakes by matching them to a similar master event. In this paper, we present an automatic location procedure based on a nonlinear signal matching scheme termed Dynamic Waveform Matching (DWM). Each event cluster is represented by a strong master event. By correlating a weak event with all

master events of a specific region, a cluster association can be performed which yields strong epicenter constraints.

### MASTER EVENT CORRELATION OF WEAK LOCAL EARTHQUAKES

Seismograms of small events originating from the same epicenter region being recorded at the same station may have the same focal mechanisms and look similar. Their hypocenters lie within one-quarter of the shortest wavelength to which similarity extends (Geller and Mueller, 1980). This so-called  $\lambda/4$ -criterion was confirmed by Frankel (1982). Pechman and Kanamori (1982), and Thorbjarnardottir and Pechman (1987). For the seismicity monitored by the BUG network, a cluster of similar mining induced events was investigated by Gibowicz, Harjes, and Schäfer (1990). The occurrence of earthquake clusters evokes the idea of locating weak events by matching them to similar master events from the same region. By classical master event techniques, an earthquake is relocated relative to a strong master event, whose hypocenter is well-known (Spence, 1980). The degree of similarity between master event and actual event must be estimated by cross-correlation. When investigating local seismic data, the number of events correctly attached to clusters strongly depends on an appropriate set of master events. Correlating all the data with all available master events yields a

\*Present address: Lehrstuhl für Meßtechnik, RWTH, Templer Graben 55, D-5100 Aachen, Germany.



**Cross-Correlations**

**Seismograms**

Figure 1. Cross-correlation of two similar earthquakes: seismograms are displayed in 2/8/40 sec windows for P-onset, P-phase, and entire seismogram, measured at station KLB of BUG small array. Event data are 3.1.1989 02:21:40,  $M_L = 1.7$ ,  $\Delta = 40$  km for upper trace and 18.1.1989 13:06:53,  $M_L = 1.5$ ,  $\Delta = 40$  km for lower one, source region is *Hamm*. Cross-correlation is performed in three pathbands, it either lacks in timing resolution or shows significant ambiguity. Maximum peak shifts to later times for longer time windows.

cluster association (Israelson, 1990; Harris, 1991). Instead of cross-correlation in the time domain, cross-spectral analysis can be used for master event relocation when the data have sufficient S/N ratio (Poupinet, Ellsworth, and Frachet, 1984; Scherbaum and Wendler, 1986).

For the correlation analysis in the time domain, different window lengths are used. Only P-phases may be used for correlation (Harris, 1991) of P- and S-phases (Pechman and Kanamori, 1982; Frankel, 1982; Thorbjarnardottir and Pechman, 1987), whereas Geller and Mueller (1980) use the whole seismogram length. Figure 1 is intended to discuss the effect of different correlation lengths. The events are located near Hamm in the eastern part of the Ruhr Basin 40 km apart. The upper trace represents the master event, whereas the lower one represents a weak event with a low S/N ratio. Selecting window lengths of 2, 8, and 40 sec, respectively, will result in correlations that are performed over the P onset, the P phase wavetrain or the entire seismogram. We regard any two seismograms as similar, if there is a distinct, unique maximum in the cross-correlation function. This maximum exists in the 40 sec correlations, especially in the low-frequency passband (0.5–4.0 Hz), but not for shorter cross-correlation windows. For local events with low S/N ratio, the similarity tends to be characteristic more for the entire seismogram than for any shorter phase. Unfortunately, the more unequivocal the maximum is, the greater is its timing inaccuracy. This is obvious in Figure 1 for the different time windows and pathbands. Also the exact position of the main maximum differs slightly because of the different phase velocities. If we use the short-time windows for cross-correlation of weak events, they yield high correlation values and a high-timing accuracy but ambiguous phase associations because of the periodic maxima. In our application, we look for a correlation with great timing accuracy but without ambiguity in phase association. Obviously, the common master event correlation for any window length does not satisfy our requirements. Instead, we have to consider a nonlinear correlation scheme that can handle stretchings and shortenings of the signals relative to each other. So we can correlate both seismograms over 40 sec with high timing accuracy. We then can use the stable low-frequency surface waves to adjust an unequivocal peak and trough mapping. When the seismologist examines the seismograms interactively, he/she handles this problem of 'time warping' by independent correlation for several seismogram parts. A usual procedure is to put paper seismograms one upon another, turn them to light to see them both and shift the sheets relative to each other. By this procedure, the existence of phase correlations is searched. It was this process being rather complex which inspired our development and subsequent implementation of an automatic nonlinear correlation method based on DWM.

## MASTER EVENT CORRELATION BY DYNAMIC WAVEFORM MATCHING

Nonlinear correlation methods were introduced firstly in exploration geophysics to perform automatic stratigraphic correlation. In literature, two completely different ways of nonlinear matching can be distinguished. The first group of procedures performs a feature extraction by converting the geophysical data into strings of primitives. The matching is based exclusively on these waveform descriptions (Anderson and Gaby, 1983; Le and Nyland, 1990; Liu and Fu, 1983; Shaw and DeFiguereido, 1990). The second group works on the original data with possible prefiltering (Leany and Ulrych, 1987; Martinson, Menke, and Stoffa, 1982). These methods are extensions of the linear cross-correlation and will be designated here as *Dynamic Waveform Matching* (DWM), a term suggested first by Anderson and Gaby (1983). In DWM, two seismic signals are correlated by measuring local similarities between short-time windows for different relative time shifts.

DWM admits the correlation of signals which are compressed or stretched relative to each other by a nonlinear shifting and warping of the time axis until the corresponding features of both waveforms are correctly matched. In our investigations, we follow a slightly different approach that is based on the *Multiple Dynamic Matching* by Leany and Ulrych (1987). It starts by a piecewise linear approximation of the matching function based on sliding time windows over the initial waveforms. Figure 2 describes the layout of the similarity matrix where local maxima or *ridge points* are extracted and adjacent ridge points form *ridge trends*. Connecting all the ridge trends yields a network of global paths through the similarity matrix. Each path represents a possible

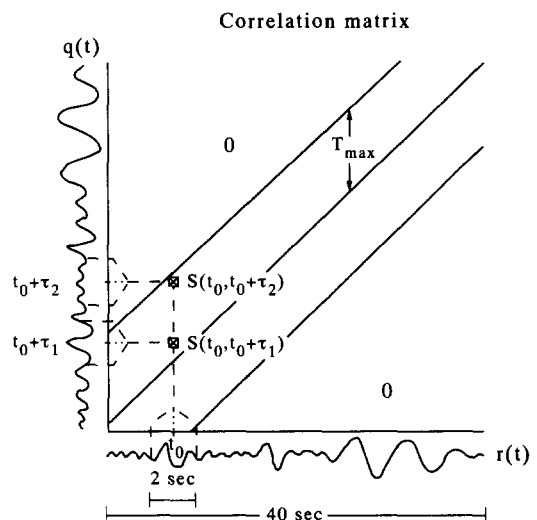


Figure 2. Layout of similarity matrix  $S$ : two signals  $r(t)$  and  $q(t)$  are matched by correlating local similarities between short windows for different relative time shifts  $\tau$ . Final path will be searched within maximum time lag  $T_{max}$  which corresponds to greatest expected delay.

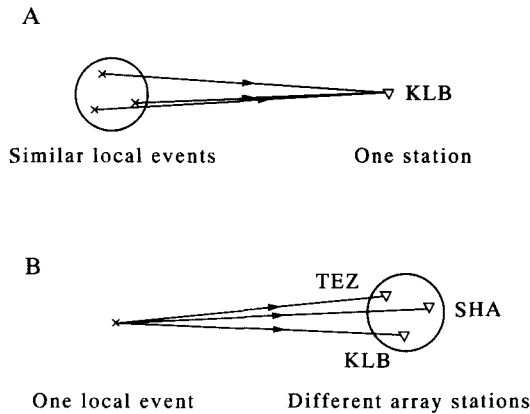


Figure 3. Application of DWM to BUG events: in A weak event is located by matching it to strong master event. Every cluster of similar events will be characterized by representative master event; comparing weak events with all master events performs automated cluster association. In B different array traces of same event are correlated to determine travel time differences of coherent P-phases. This gives azimuth and slowness for epicenter beam steering.

correlation for the entire signal length. In a last step, the 'best' path has to be selected.

Figure 3 introduces the two possible applications of DWM to local seismic data in the BUG network. In (A) a weak event is evaluated by comparison to its stronger master event. Because the P and S onset time of the master event are known within one-sample, the P and S onset of the weak event can be derived from the correlation path with the same timing accuracy. Application (3B) correlates different array traces of the same event (Schulte and Joswig, 1990). This approach is valuable especially for sparse arrays where one station has the best S/N ratio and thus can act as the master trace.

In the Appendix, we explain the principle of DWM in a simple example before applying it to real data. On the left side we point out the essential computing steps in C-code. We calculate local cross-correlations of length  $L$  or the whole signal length  $I$  with a 50% overlap. The windows are shifted relative to each other by  $\pm T$ . Applied to the time series  $r(i)$  and  $q(i)$ , such a correlation yields a two-dimensional similarity matrix  $S(x,y)$  where one column represents the similarity of a fixed window in  $q(i)$  to a sliding window in  $r(i)$ . The next step in DWM is to determine the ridge trends where 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 neighbor values. Because we expect a correlation path in parallel to the main diagonal, the left and right neighbors are taken in respect to the diagonal.

Then we must extract one final correlation path from the matrix of ridge trends. We expect a path passing approximately parallel to the main diagonal. For the path, we allow a variation within an area of  $\pm V$ . This area being parallel to the main diagonal, the path is calculated by shifting it for the matrix

of ridge trends in the  $x$  direction.  $V$  is selected, appropriately, so we do not have more than one matrix value  $>0$  in each column of the area. The best path is defined as the one which leads to the greatest sum of ridge points within its area.

The extension to real seismological data differs only slightly from our example in the Appendix. We work on 100 Hz data and on seismograms with a length of 40 sec. The traces are bandpass filtered with corner frequencies of 0.5–8.0 Hz for P- and S-waves and 0.5–3 Hz for surface waves, respectively. The local correlations are performed on 2 sec windows with center time  $t_0$  (see Fig. 2). These windows are shifted over the whole seismogram length in steps of 1 sec. The maximum time lag  $T_{\max}$  is determined by the greatest shift that must be expected between coherent phases. Because the  $x$ -axis is sampled every second only (one-half of the window length), we get matrix dimensions of  $40 \times 4000$  samples with a sparse distribution around the main diagonal. To display on computer screen, the matrix is expanded to quadratic scale by padding along the diagonals. The values of cross correlation are represented by nine colors. Figure 4A shows the reproduction by gray-scale levels for the Hamm events already introduced in Figure 1. The horizontal trace is the master event, the vertical one is the weak event.

Figure 4B shows the matrix of ridge trends. The threshold is 0.12 for body versus 0.3 for surface waves; the latter have higher correlation values because of the lower spectral noise. Figure 4D shows the final correlation path. We know P and S onset time of the master event exactly but must assume an uncertainty of  $\pm 2$  sec for the detected P onset time of the weak event. Because we cannot predict the location of the actual epicenter relative to the master event, the path variance must be taken symmetrically around any straight diagonal in Figure 4C. All paths that start within the P onset column of the master event and that lay within the appropriately shifted variance area must be considered. For body and surface waves we select different values for  $V$ . For each path, the sum of correlation coefficients is normalized with respect to the 'ideal' path, the auto-correlation of  $r$ . The path with the greatest value is taken as the final correlation and shown in Figure 4D. It maps the peaks and troughs of both waveforms with one sample accuracy, so Figure 4E can display the aligned 2 sec windows of P and S onsets. For cluster association, a similarity threshold must be defined. Empirically, we determined a value of 0.35 to work well: any actual event is compared to all the master events of its particular region but is associated to the most similar one only if its correlation exceeds the threshold as given.

#### RESULTS OF AUTOMATED MASTER EVENT ASSOCIATION

As an example for the claimed robustness in noise, we tried to relocate all the mining induced

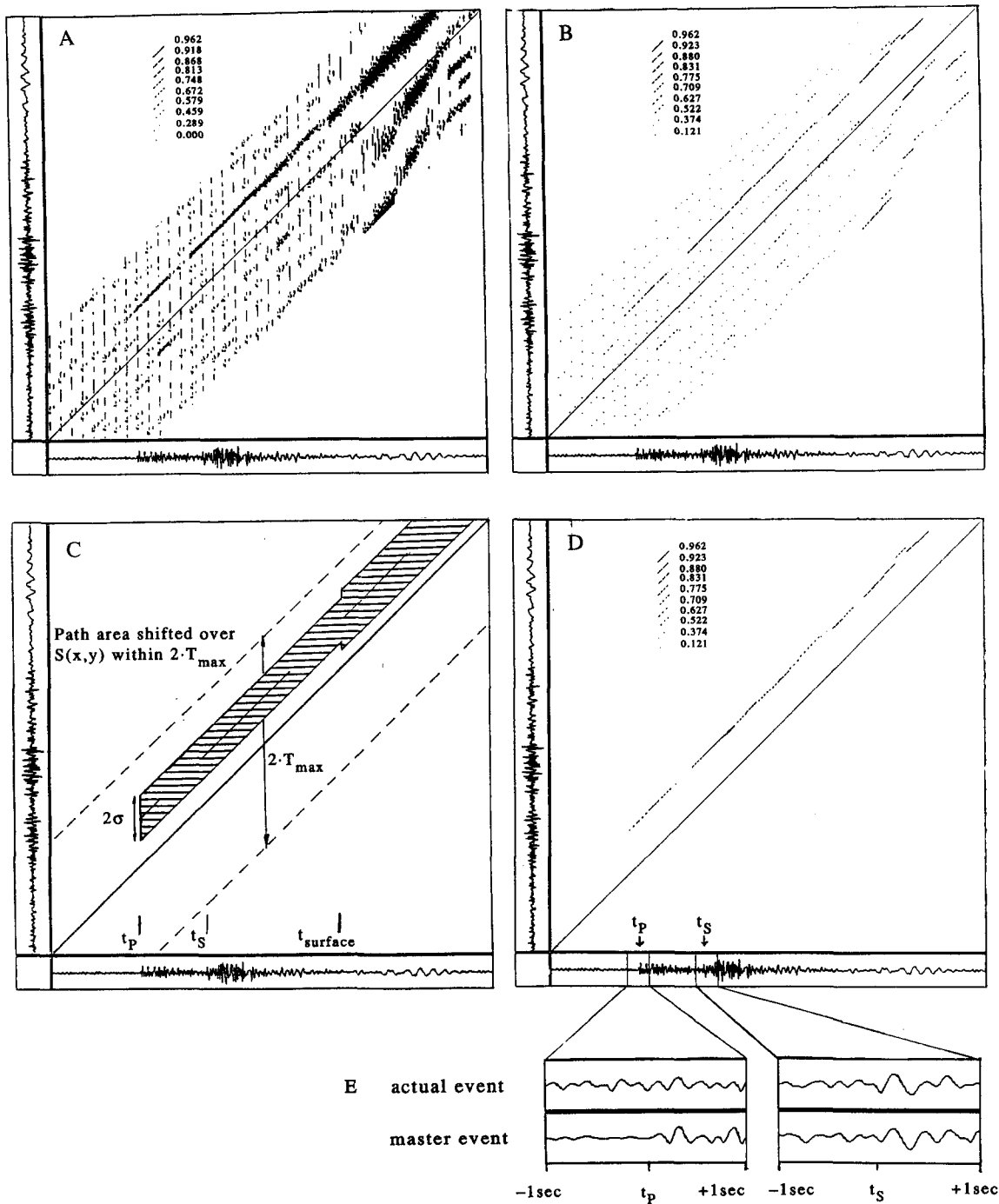


Figure 4. DWM for two Hamm events of Figure 1: A—similarity matrix  $S$ . Horizontal trace shows master event; B—matrix of ridge trends; C—theoretically expected path guides search for final matching path that contains greatest sum of single correlation values as in D. In E corresponding time windows are shown, matching is exact within one sample of initial time series.

Figure 5 (overleaf). Epicenter map of region Hamm: A—interactive results of BUG array evaluation do not show significant clustering. Aligned structures are result of principal uncertainties in location which are  $\pm 2^\circ$  for azimuth and  $\pm 1$  km for distance; B—improvement of location by DWM for KLB master event association (at 40 km distance). Some events remain unassociated, their epicenter estimate is unchanged. In C, results of HRH local station (4 km distance, three-component registration) are shown. For better comparison, the epicenters are marked by same symbols as in B although they are not associated to master events now. Location of master events in B was determined on base of HRH results.

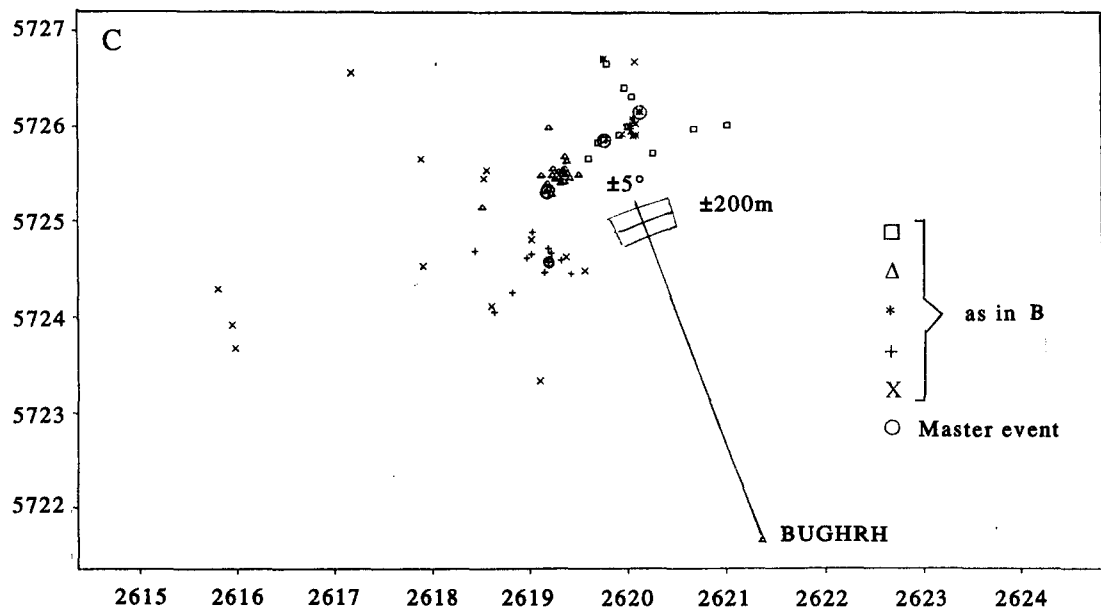
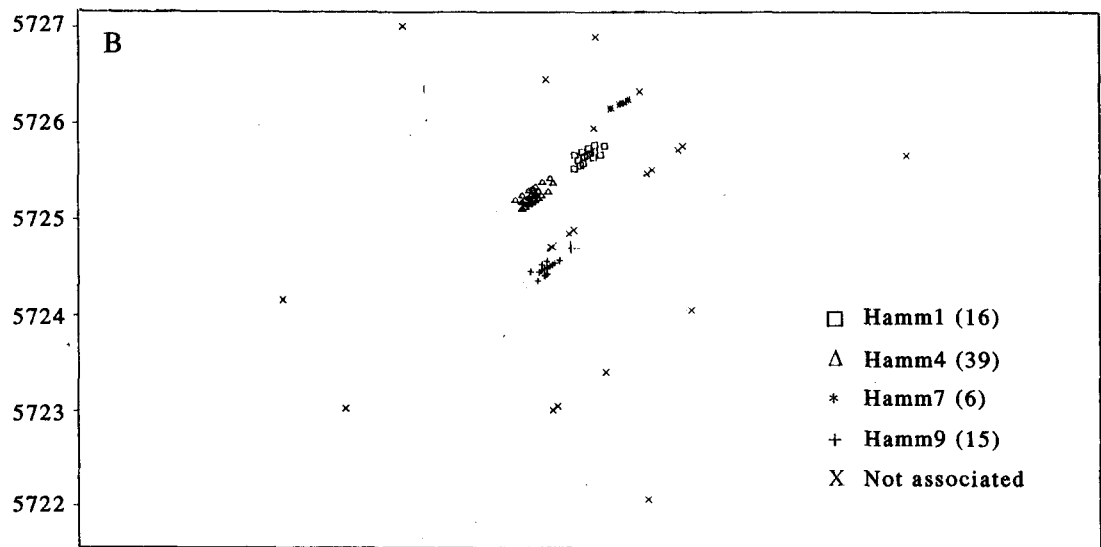
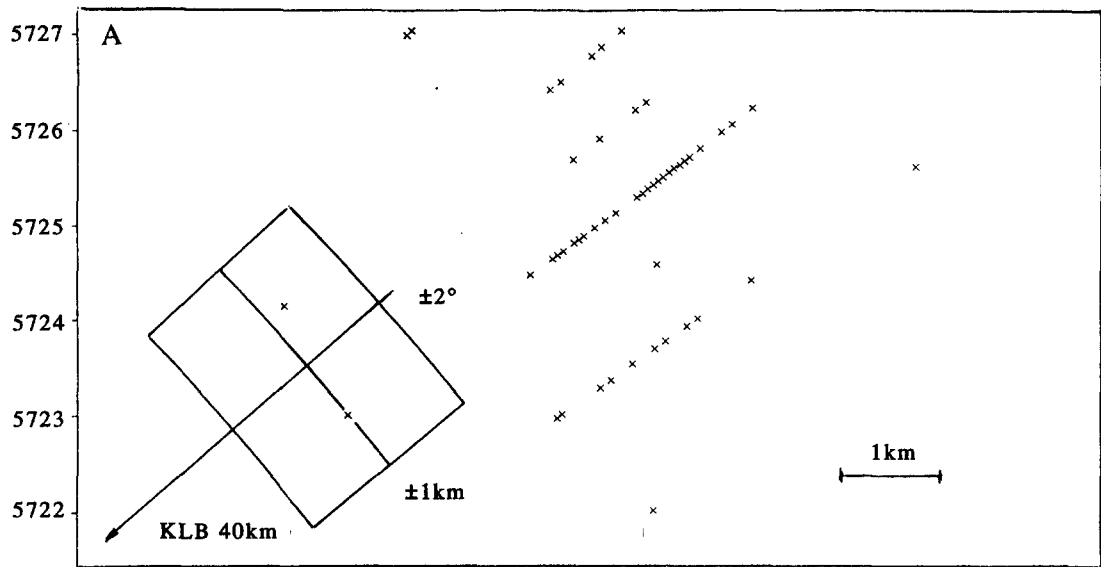


Figure 5 (Caption on p. 237).

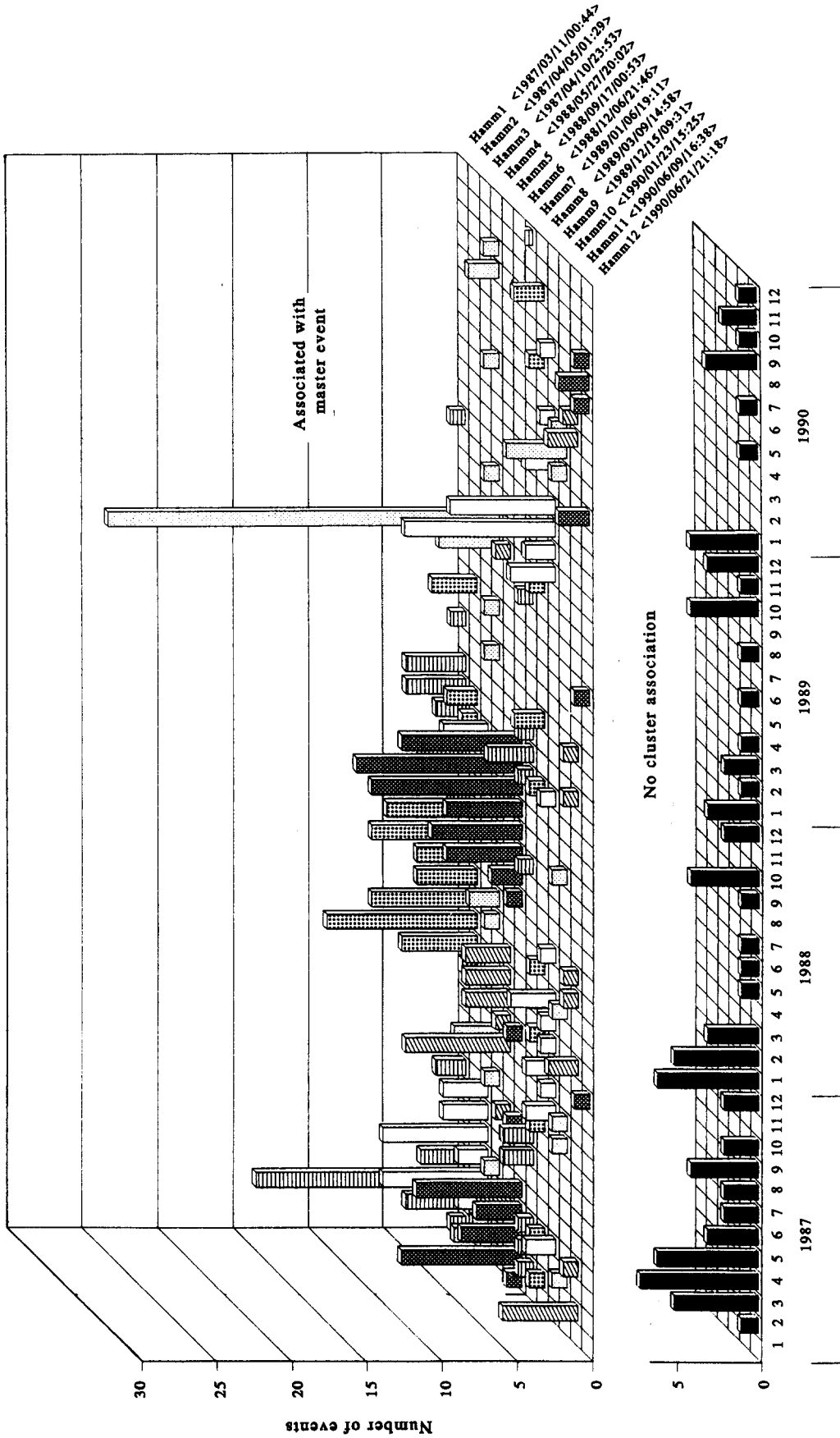


Figure 6. Clustering in Hamm region through time: for mining induced seismicity, clear temporal clustering can be resolved besides spatial clustering of Figure 5. However, this effect cannot be attributed directly to mining activities which were constant through years.

events of the epicenter region *Hamm* between 1987 and 1990. The master event comparisons of the 470 seismograms are based on the single station KLB. The epicenter region Hamm is 40 km NE of KLB. Our master event association can be verified by locating the events individually with the three-component station HRH which is only 4 km distant from the *Hamm* cluster. The 12 master events were selected interactively by some visual inspection. Matching them by DWM did not yield a correlation path. When 470 events are correlated with the 12 master events, 5640 comparisons have to be performed. Such an amount of calculation can only be handled by automated approaches.

Figure 5 maps the 1989 subset of 94 epicenters geographically. In Figure 5A the bulletin results from the BUG small array obtained by interactive analysis are displayed, the uncertainty of  $\pm 2^\circ$  in azimuth (i.e.  $\pm 2.8$  km) and  $\pm 1$  km in distance is too large to resolve any detailed structure. In Figure 5B the 76 events that could be associated by DWM to one of the active master events were located according to the  $t_s - t_p$  deviation. These results of the station KLB being at 40 km distance can be compared to the epicenter locations in Figure 5C having been determined interactively from the data of the station HRH at 4 km distance. Taking the uncertainty of only  $\pm 200$  m for the HRH locations into account, we can indeed resolve the clusters and confirm the cluster association of DWM in Figure 5B.

In Figure 6 the cumulative history of cluster activities in 1987–1990 is given. 379 from 470 events (80%) could be associated to one of the 12 master events. Some events remain unassociated either because they were single events or their S/N ratio was too low. Apart from the significant spatial clustering, a temporal clustering could be resolved as well. In general, the clusters extend over 2 months but some single pre- and aftershocks occur as well. The distribution of seismic activity is not uniform over the year, event accumulations and calm months can be recognized. However, the coal mining near Hamm is performed continuously through the years, so there is only an indirect interdependency between coal mining and the mining induced seismicity.

### CONCLUSIONS

Clustering and location of local seismicity by master event comparison based on Dynamic Waveform Matching especially is suited for weak events which cannot be evaluated individually because of ambiguous phase correlations. DWM and pathfinding through a similarity matrix leads to a significant correlation path and an unequivocal phase correlation for the entire seismogram. The method was designed for an automated earthquake processing system. By DWM, the fine structure of clusters can be resolved within different

epicenter regions that otherwise is obtained only by stations close to the epicenter regions. In this respect, automatic evaluation by DWM is superior to the interactive evaluation methods of isolated events, but the results are dependent on a set of master events. A useful improvement in future will be the automated clustering and selection of master events.

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

### REFERENCES

- Anderson, K. R., and Gaby, J. E., 1983, Dynamic waveform matching: Proc. Third Intern. Symp. on Computer-aided Seismic Analysis and Discrimination, Washington, D.C., p. 98–108.
- 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. America*, v. 72, no. 4, p. 1277–1294.
- Geller, R. J., and Mueller, C. S., 1980 Four similar earthquakes in Central California: *Geophys. Res. Letters*, v. 7, no. 10, p. 821–824.
- Gibowicz, S. J., Harjes, H.-P., and 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. America*, v. 80, no. 1, p. 88–109.
- Harris, D. B., 1991, A waveform correlation method for identifying quarry blasts: *Bull. Seism. Soc. America*, v. 81, no. 6, p. 2395–2418.
- Israelsson, H., 1990, Correlation of waveforms from closely spaced regional events: *Bull. Seism. Soc. America*, v. 80, no. 6, pt. B, p. 2177–2193.
- Joswig, M., 1993, Automated seismogram analysis for the tripartite BUG array: an introduction: *Computers & Geosciences*, v. 19, no. 2, p. 203–206.
- Le, L., and Nyland E., 1990, Pattern analysis of seismic records: *Geophysics*, v. 55, no. 1, p. 20–88.
- Leany, W. S. P., and Ulrych, T. J., 1987, Multiple dynamic matching: a new approach to well log correlation: *Geoexploration*, v. 24, no. 6, p. 503–515.
- Liu, Hsi-Ho, and Fu, King-Sun, 1983, An application of syntactic pattern recognition to seismic discrimination, *IEEE Trans. Geoscience Remote Sensing*, v. GE-21, no. 2, p. 125–132.
- Martinson, D. G., Menke, W., and Stoffa, P., 1982, An inverse approach to signal correlation: *Jour. Geophys. Res.*, v. 87, no. B6, p. 4807–4818.
- Pechman, J. C., and Kanamori, H., 1982, Waveforms and spectra of preshocks and aftershocks of the 1979 Imperial Valley, California, Earthquake: evidence for fault heterogeneity?: *Jour. Geophys. Res.*, v. 87, no. B13, p. 10579–10597.
- Poupinet, G., Ellsworth, W. L., and Frechet, J., 1984, Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California: *Jour. Geophys. Res.*, v. 89, no. B9, p. 5719–5731.
- Scherbaum, F., and Wendler, J., 1986, Cross spectral analysis of Swabian Jura (SW Germany) three-component microearthquake recordings: *Jour. Geophys.*, v. 60, no. 3, p. 157–166.
- Schulte, H., and Joswig, M., 1991, Automated phase picking for weak local earthquakes by Dynamic Waveform Matching, in Proc. and Activity Rept. XXII General Assembly Roca, A., and Mayer-Rosa, D., eds., of ESC 1990 in Barcelona, v. 1, p. 211–218.



Shaw, S. W., and DeFigueiredo, R. J. P., 1990, Structural processing of waveforms as trees: IEEE-ASSP, v. 38, no. 2.  
 Spence, W., 1980, Relative epicenter determination using P-wave arrival-time differences: Bull. Seism. Soc.

America, v. 70, no. 1, p. 171-183.  
 Thorbjarnardottir, B. S., and Pechman J. C., 1987, Constraints on relative earthquake locations from crosscorrelation of waveforms: Bull. Seism. Soc. America, v. 77, no. 5, p. 1626-1634.

**APPENDIX**

We present a simple example to explain the essential computing steps of DWM. On the right side we calculate the similarity matrix, the ridge trends and the path for two short time series. On the left side we sketch the essential parts of the program in C-code. For further explanations see text.

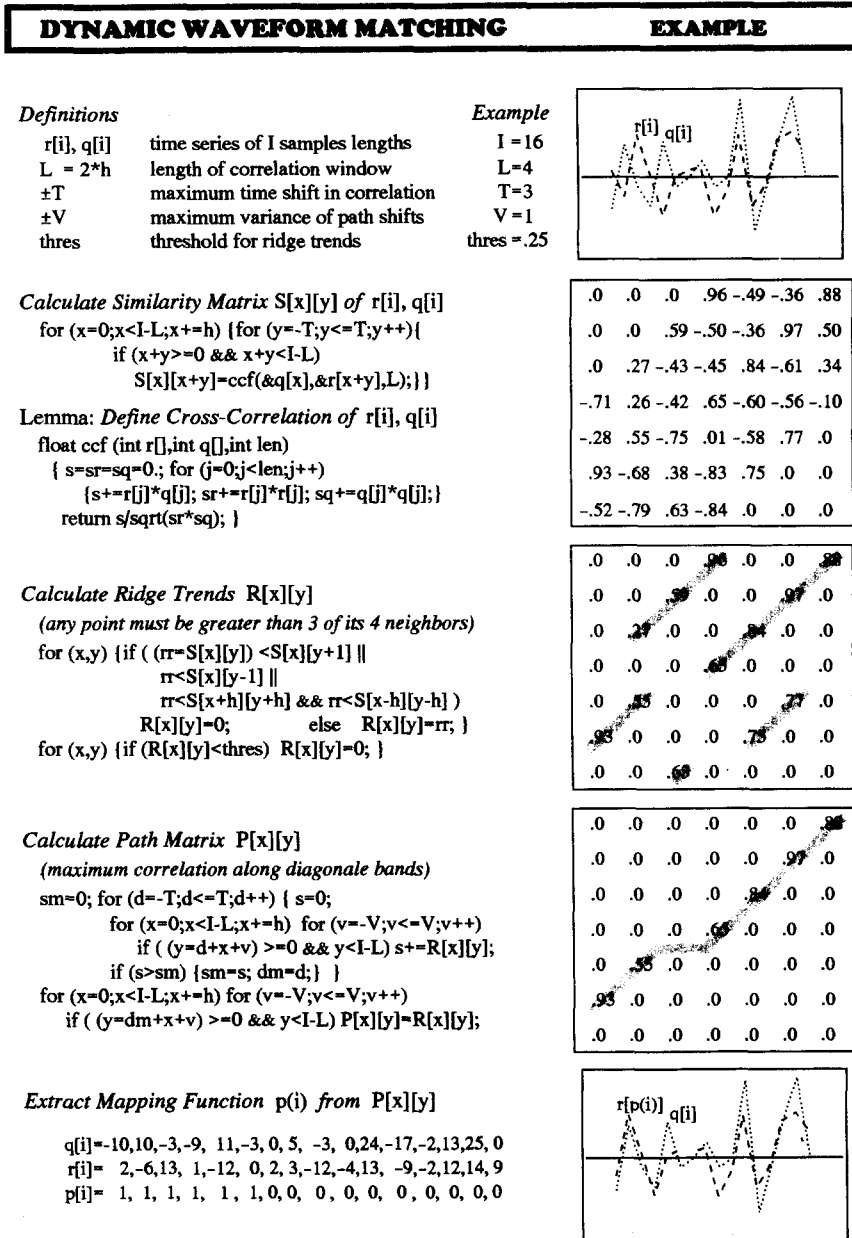


Figure A1