

## AUTOMATED REEVALUATION OF LOCAL EARTHQUAKE DATA BY APPLICATION OF GENERIC POLARIZATION PATTERNS FOR P- AND S-ONSETS

ERIC KLUMPEN and MANFRED JOSWIG\*

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

**Abstract**—The particle motion of local earthquake seismograms is affected strongly by the fine structure details of upper crust. The angle of incidence gets frequency dependent, shear wave splitting occurs and strong P-SV conversions contaminate the P-coda. Contrary to teleseism, it is not possible any more to detect S-onsets by conformance tests between data and simple models. Instead, we must derive polarization images in the time–frequency plane that display particle motion without any assumptions. By suitable scaling, these images neutralize all high frequency effects and allow for onset recognition by simple patterns. The method was applied to the 800 events of 1989 evaluated by the Bochum University Germany (BUG) observatory. We determined a 67% success rate with 13% wrong and 20% rejected because of unstable phase energy. For two source regions, the automated results are shown to be more reliable than interactive routine evaluation by man.

*Key Words:* S-phase picker, Automated epicenter determination.

### INTRODUCTION

In most schemes of automated seismogram evaluation, the exact determination of P- and S-onset times is a prerequisite for the event location. To call these timing procedures 'phase picker' instead of the general term 'detector' we follow a suggestion of Allen (1982). There he outlines as differences: detectors are the initial processing steps and must be optimized for low false alarm rate whereas phase pickers act on preselected time windows and should yield maximum timing resolution and high sensitivity. For the P-phase, we can achieve this alternate tuning by a lower detection threshold. This approach will fail obviously for the S-onset which is disguised somewhere in the P-coda. So it can not be identified by signal amplitude but—at least in theory—is marked by the beginning of transversal particle motion. Its unequivocal identification demands three-component data. These data, however, are difficult to capture for humans because the additional time dependence introduces the fourth dimension. In the observatory routine of interactive phase determination, the necessary visualization is supported by hodograms or by transforming the initial Z–NS–EW seismograms to P–SH–SV axes (Plesinger, Hellwig, and Seidl, 1986).

### MODEL-BASED AUTOMATIZATION

For the automated S-onset determination, another way was the preferred one: the records of particle

motion are reduced in dimension to be tested against some threshold level. The new functions represent the conformance of the actual data to some given model of shear-wave property. Models start with the assumption of rectilinearity for the polarization filter (Montalbetti and Kanasewich, 1970), utilize 'predicted coherence' (Roberts, Christoffersson, and Cassidy, 1989; Ruud and Husebye, 1992) and extend to the complex interaction of signal and noise which is fitted to the data by maximum likelihood (Christoffersson, Husebye, and Ingate 1988; Ruud and others, 1988; Roberts and Christoffersson, 1990). However, the reported performances can be summarized in the statement that model-driven phase picking performs well for teleseismic data and P-onsets, but tends to fail for the S-phase of local earthquake data as well as for the weak signals. We noted these restrictions confirmed by our own applications of polarization filtering to the local seismicity of the Ruhr Basin, NW Germany, which is monitored by the Bochum University Bochum (BUG) network.

The reason is because of the complexity of local seismograms that are different significantly even within small epicenter areas such as the 'Gelsenkirchen' region at 18 km distance (see Fig. 1). As shown in Figure 2, this complexity remains even after rotating into P–SH–SV axes. This observation is widely known (MacBeth and Redmayne, 1989; Buchbinder and Haddon, 1990; Ebel, 1989) where Menke and others (1990) determined the expected patterns of linear polarization to exist only in the beginning of the P-phase. So our first intention for a new S-phase picker was to get an unbiased image of the full complexity in local

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

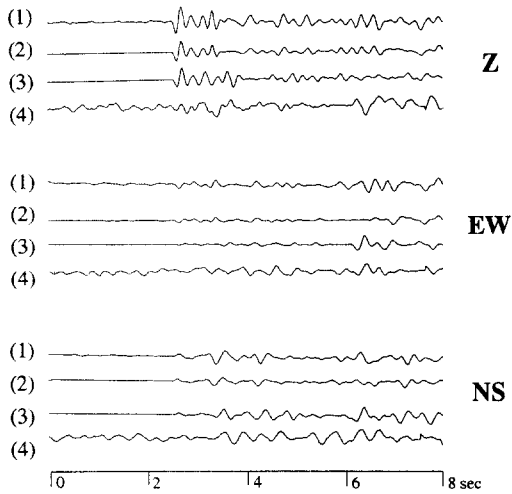


Figure 1. Four local earthquakes from 'Gelsenkirchen' region (distance = 18 km, azimuth = 320°). Although from small source area, seismograms show significant differences in S to P amplitude ratios.

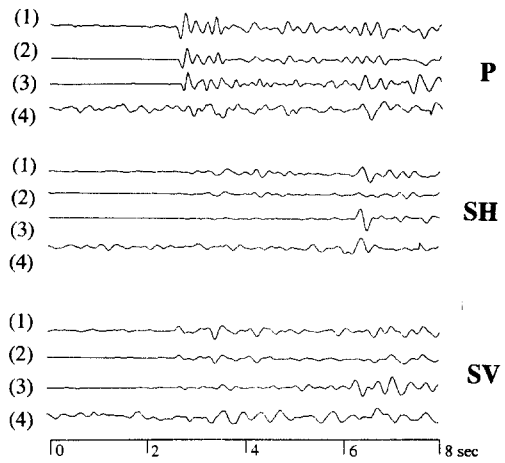


Figure 2. Same events of Figure 1 in P-SH-SV axes: even after rotation deviations remain, most notably for second event.

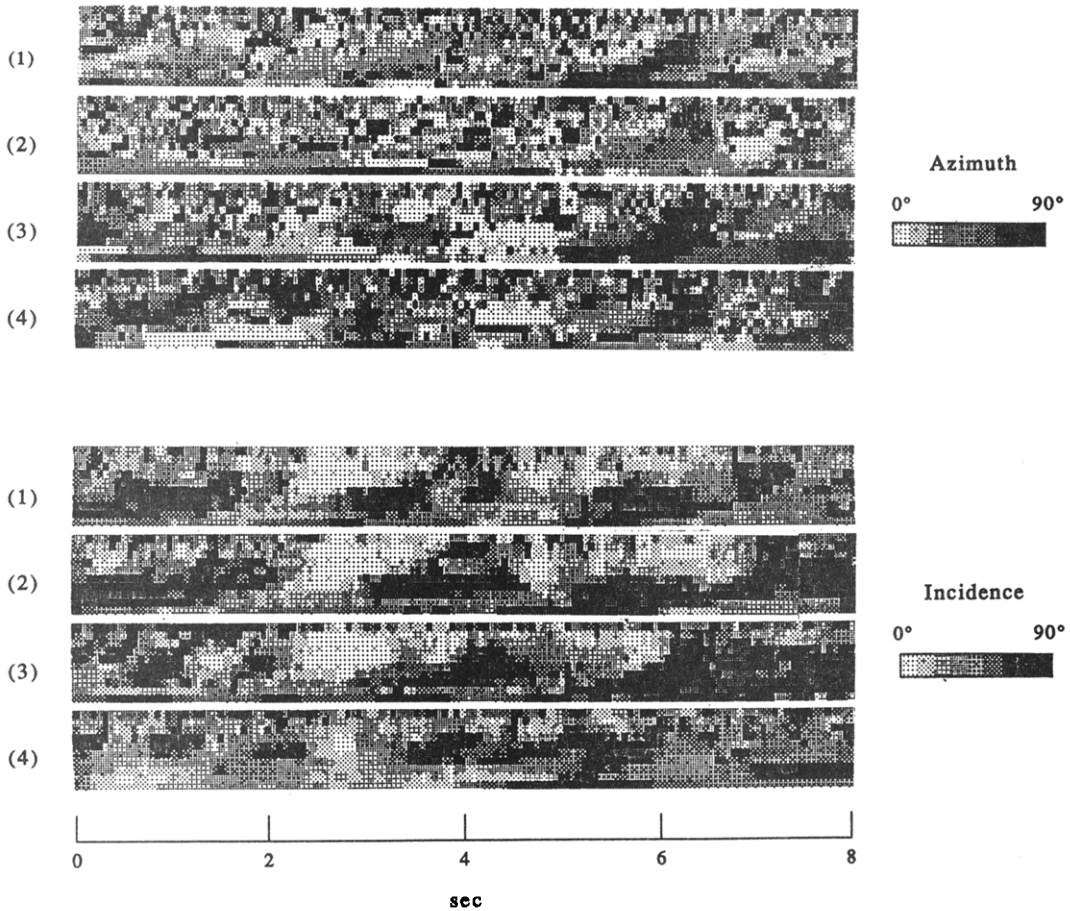


Figure 3. Instantaneous angles of azimuth and incidence in  $f-t$  plane for events of Figure 1. Whereas for azimuth distribution seems arbitrary (white = 0, black = 180), there is clear pattern for incidence (white = 0, black = 90). In P-onset, incidence gets more perpendicular with higher frequencies caused by more detailed crustal layers for shorter wavelengths.

earthquake data. We therefore developed an interactive analysis tool that displays the particle motion in different representations but does not rely on any model assumptions.

#### CALCULATION OF POLARIZATION IMAGES

In general, we will follow the recognition of time-frequency images already introduced for single-

trace analysis as 'sonogram-detection' by Joswig (1990a, 1993a). For three-component data, the dependencies on frequency were outlined by Jurcevics (1988). All the calculations start by principal value decomposition of the correlation matrix (Flinn, 1965). For our time-frequency images, we can assume the particle motion to be sufficiently monochromatic by bandpass filtering; then the smallest eigenvalue gets zero.

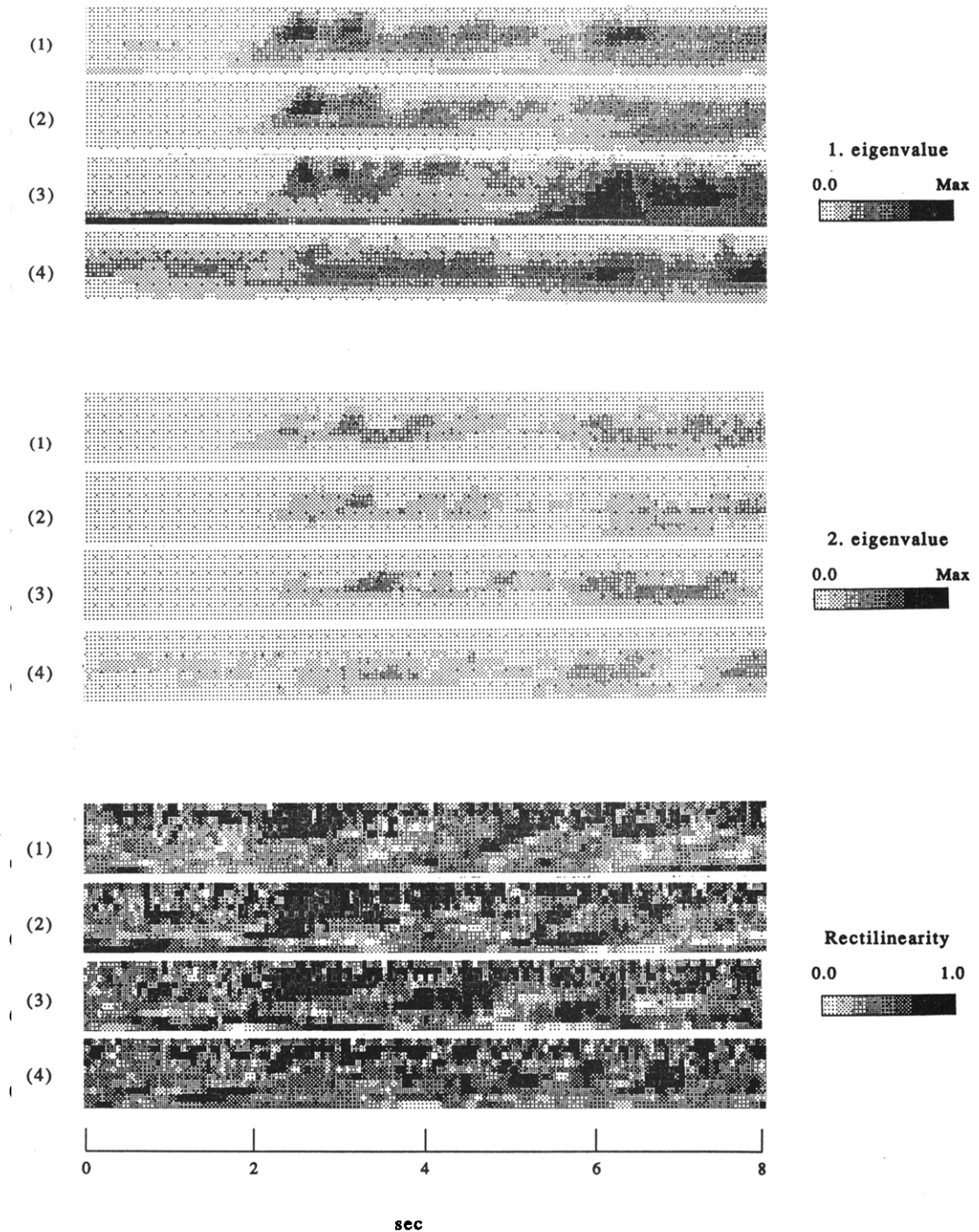


Figure 4. Parameters of polarization filtering in  $f-t$  plane for events of Figure 1. Whereas energy increase is visible for both P- and S-onset, ratio does not show significant change in rectilinearity for shear waves.

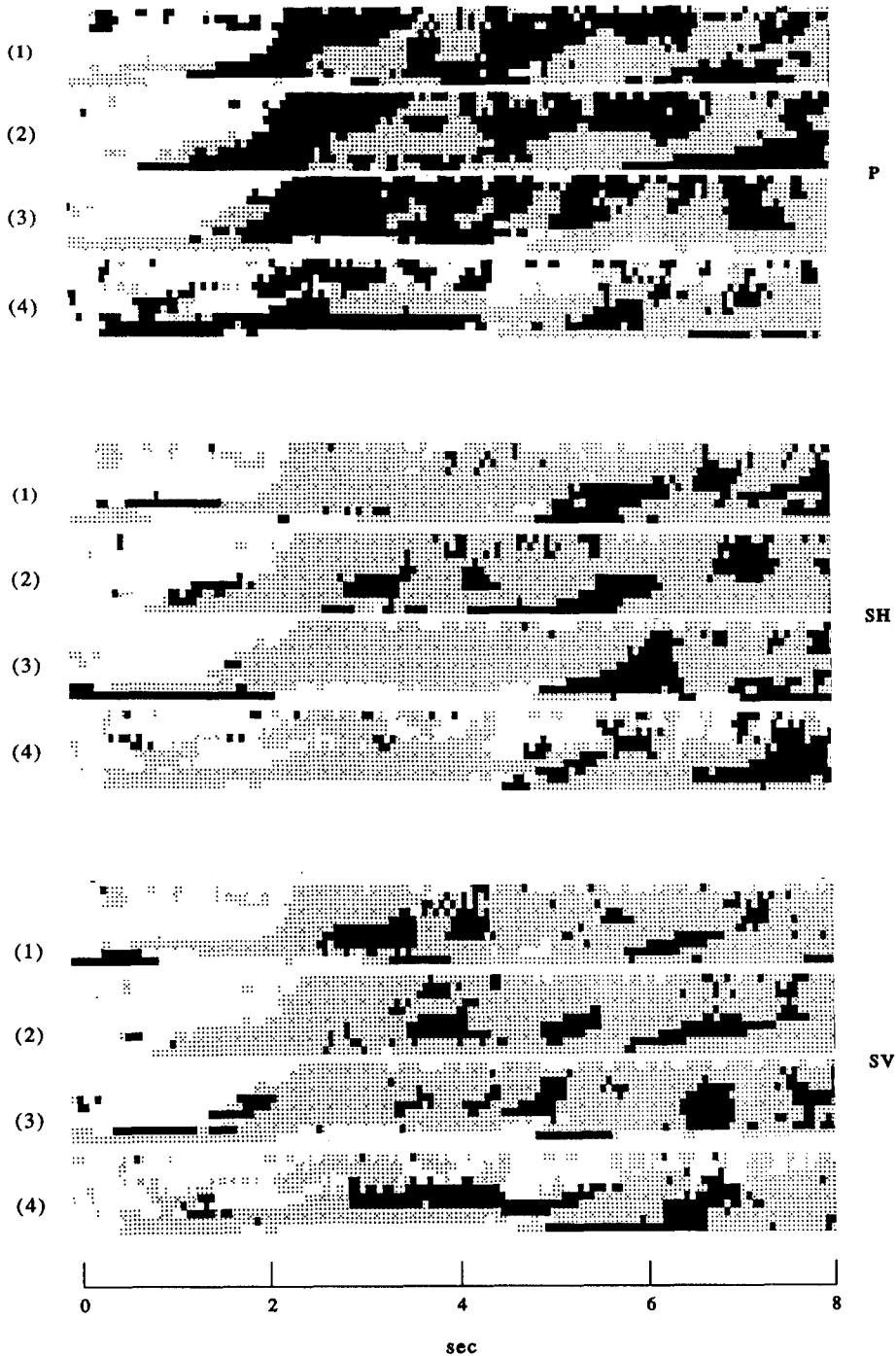


Figure 5. Polarization images for events of Figure 1. These images are composed by S-onset combination rule (see Appendix) to detect on SH energy onsets. Black areas stand for +1, dotted region symbols -1, and white is zero.

The sonogram-detector was tuned—just in the sense of Allen (1982)—to maximum robustness by small and stable image dimensions realized by constant window length for all frequencies. Now we are concerned with phase picking and must optimize for maximum timing resolution. As in the calculation of Gabor-matrices (Dziewonski, Bloch, and Landisman, 1969) or the wavelet-transformation (Rioul and Vetterli, 1991), this is best performed by

time windows that have constant ratio with respect to the analyzed wavelengths, that is they get shorter for higher frequencies. In the situation of the 100 Hz BUG data, the frequency axis is made up by 10 zero-phase, half-octave wide passbands centered from 0.8 to 20 Hz. The time windows have the minimum possible length of one wave train for a maximum in timing resolution. The analysis is performed with 50% window overlap up to

10 Hz. To avoid unequal sampling in the final image, we interpolate at larger wavelengths to get a common time increment of five data samples for all frequencies.

The results of our analysis tool are given for the four *Gelsenkirchen* events. In Figure 3, the momentary azimuths change rapidly in time and frequency. This effect seems arbitrary, no general pattern is visible at the phase onset times. In contrast, the angles of incidence get frequency dependent in a systematic way. For the P-onset, they are lower (light color) for high frequencies indicating the better resolution of crustal fine structure by shorter wavelengths. The image smears for S but dependencies on frequency are visible. When looking at the signal energy in Figure 4 all onsets cause an increase in eigenvalues. However, the rectilinearity  $R = 1 - l_2/l_1$  does only increase for higher frequencies of the initial P-phase whereas no linear polarization is seen in S. Additional effects were determined in the data too which are shear-wave splitting, large variations for the P/S amplitude ratio even within small epicenter regions and a systematic deviation between azimuths of P- and S-phase for individual events. For the recognition of shear-wave onsets, the worst perturbations are caused by some early but strong P-SV conversions which contaminate the whole P-coda.

#### PATTERN RECOGNITION ON POLARIZATION IMAGES

Some fragmentary explanations exist for the encountered effects but a complete description is still beyond our knowledge. So one may wonder how model-based data fitting should cope with the complexity of local earthquake data. On the other hand, humans obviously are able to determine S-phase onsets with sufficient reliability, and thus it should be possible to mirror this cognition by computer. Our principal idea is similar to the approach of sonogram-detection where we model the recognition of known seismogram patterns which also can not be confirmed by theoretical computations. We determined that it seems to be sufficient to transform the given situation into some stable representation of internal images that guide the human recognition (Joswig, 1990b, 1991). In the situation of S-phase determination, the search for these 'mental images' took us a year partly because of the great difficulties to visualize the particle motion. As for our version of specialized sonogram images, the calculation of the new polarization images should be as simple, robust, and restricted to the principle effects as possible. Once we succeed as shown by Figure 5, the complexity of local seismic data reduces to recurrent situations that can simply be recognized by the constant patterns of Figure 6.

Our calculation of polarization images is restricted to the maximum eigenvalue and its associated eigenvector per time-frequency pixel (see Appendix). It

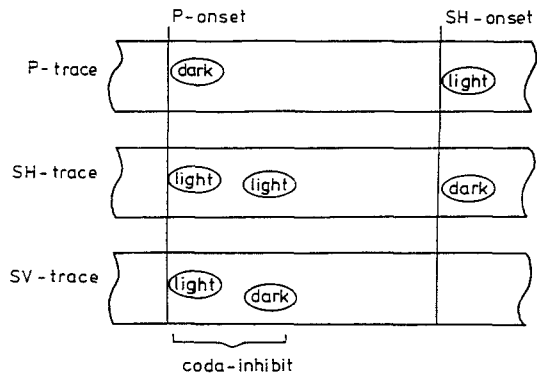


Figure 6. Principal shape of polarization onset patterns: dependent on type of onset, pattern must reflect different energy distributions. Coda inhibit will mask early P-SV conversion whereas S-onset is defined more precisely here as first occurrence of SH-energy. Numerical values can be seen in Appendix.

utilizes two principal characteristics of the given data. In the *phase criterion*, we look for the dominant direction relative to coordinate axes that are fixed in space. We select the P-SH-SV system where its angles to Z-NS-EW could be derived from the initial particle motion. To get better performance in noise, we alternatively can supply this information a priori, for example by the initial recognition of source regions in a complete evaluation system (Joswig, 1993b). In the *amplitude criterion*, we then determine those parts of our initial data that are significantly above noise. This step performs similar to the noise blanking in sonogram-detection; we now assume a log-normal distribution for eigenvalue  $l_{\max}$  and a threshold of mean plus once the variance.

Both criteria will reduce the data to binary representations [0,1]. To yield the final phase recognition, both matrices must be combined to one trivalent image of [-1,0,1]. As described by C-code and example in the Appendix, this combination obeys different rules for P- or S-onsets. Because a value of -1 will decrease the pattern fit, it is associated with  $l_{\max}$  below threshold for the P-onset but with *not aligned eigenvector* for the S-phase. Contrary, the zero will not effect the calculation and represents *significant energy out of phase* for P but *correct alignment of nonsignificant energy* in the situation of S-picking. These differences in the polarization image trace back to correspondent definitions in the patterns' matrices. So the determination of P-onset is performed by P pattern with the P combination rule whereas for S we accordingly take S pattern and S combination rule.

Some minor comments are necessary to explain the rest of the algorithm. Because pattern recognition can be understood as a cross-correlation of pattern and data, we look for the maximum of pattern fit and not for any initial threshold surpass to determine the correct onset timing. The polarization images do not have uniform timing resolution in all frequency bands, so the 'acausal' prerun in the patterns reflects

the greater analysis windows at lower frequencies. The achieved resolution for phase picking is one-half the shortest window length, that is  $\pm 5$  samples or

$\pm 0.05$  sec. The edge conversion by minus prerun in the patterns will enhance the timing, it is adjusted to  $-1/3$  by the same idea of *balance condition* as in

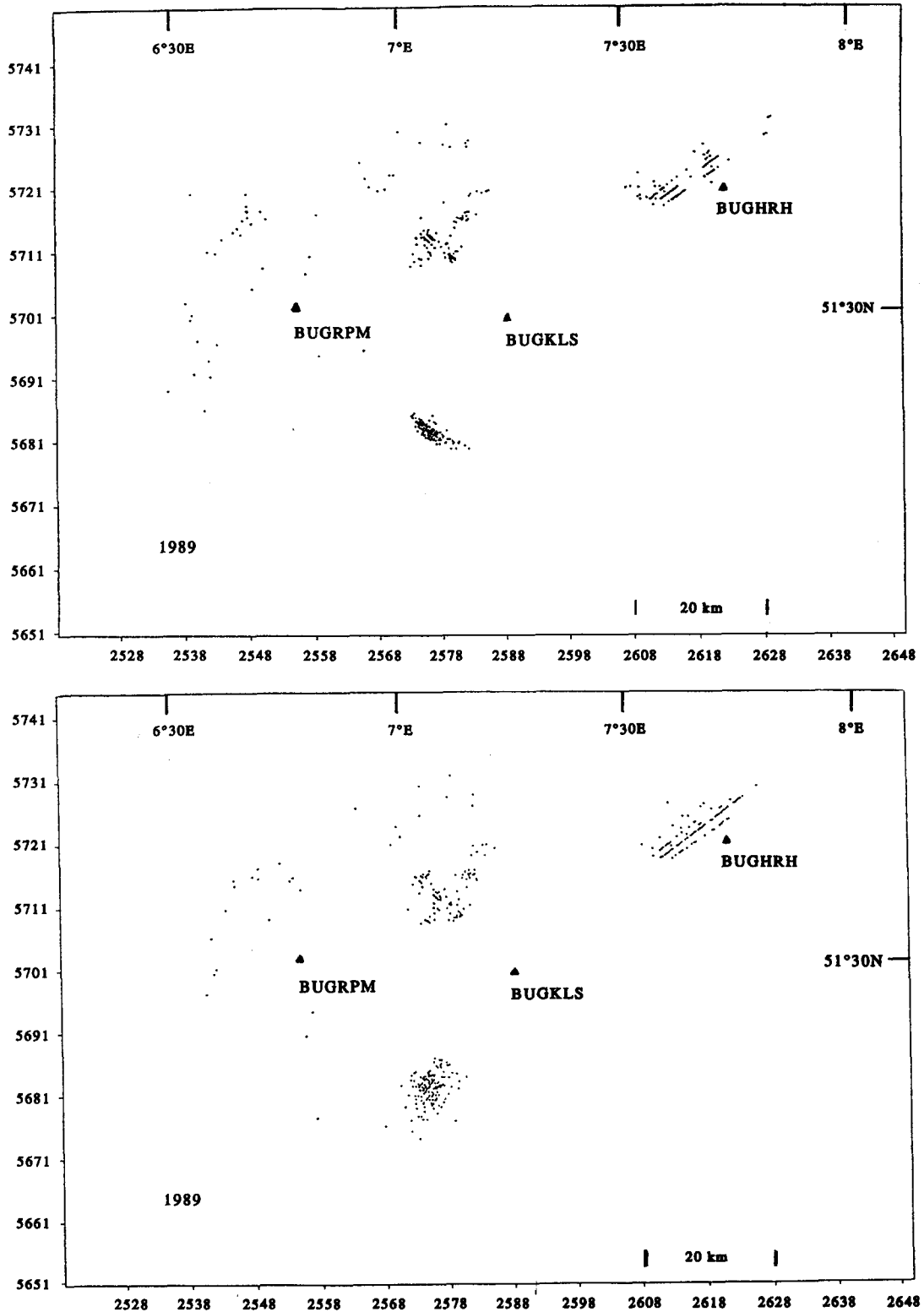


Figure 7. Epicenter maps for seismicity in 1989 monitored by BUG network. In upper map, 800 events are located manually whereas below distance is recalculated on base of P- and S-onset times determined automatically for 473 events.

sonogram-detection to yield a pattern fit of zero for noisy data. And finally, the two given patterns completely describe the task of phase picking. Thus we term them 'generic' to stress the difference to sonogram-detection where the patterns must individually be adapted to every new situation of seismicity. Particularly, the S pattern is tuned to evaluate the onset of SH energy, so we more precisely will term it an SH phase picker.

#### REEVALUATION OF LOCAL EARTHQUAKE DATA

In a test installation, we run the phase picker offline on the 1989 data set of 800 events located interactively by the BUG observatory and displayed in the upper map of Figure 7. The data set spans a range of  $M_L = 0.5-2.7$  in magnitude and 10–50 km in distance. The epicenters in the lower map of Figure 7 are based on the automated phase picking for distance calculations; their azimuths are given a priori by the BUG small array located near KLS. In a fully automated system, we would get this information by DWM (Joswig and Schulte-Theis, 1993). The success rate was 67%, that is for 473 events both P- and S-onsets are determined correctly within  $\pm 50$  msec. 13% or 88 events were mislocated mostly because of

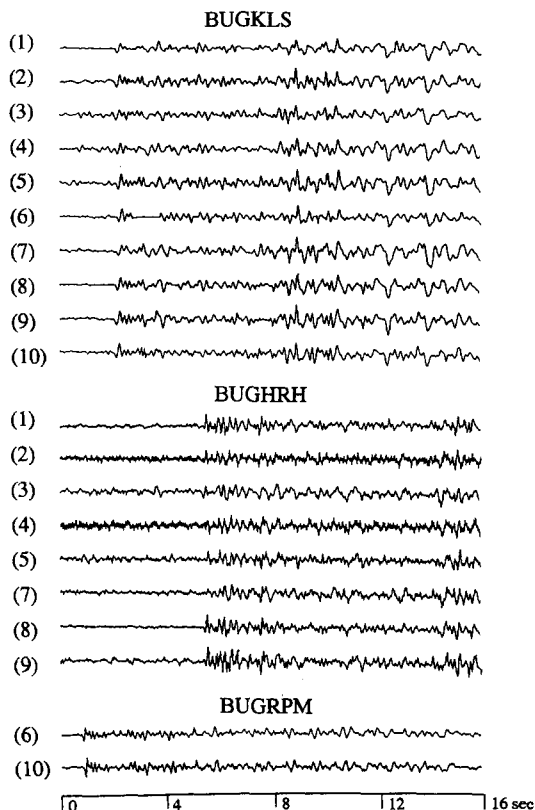


Figure 8. Ten events of 'Dorsten' region that are recorded by at least two of three 3-component stations KLS, HRH, and RPM of BUG network. Only rotated P-traces are shown.

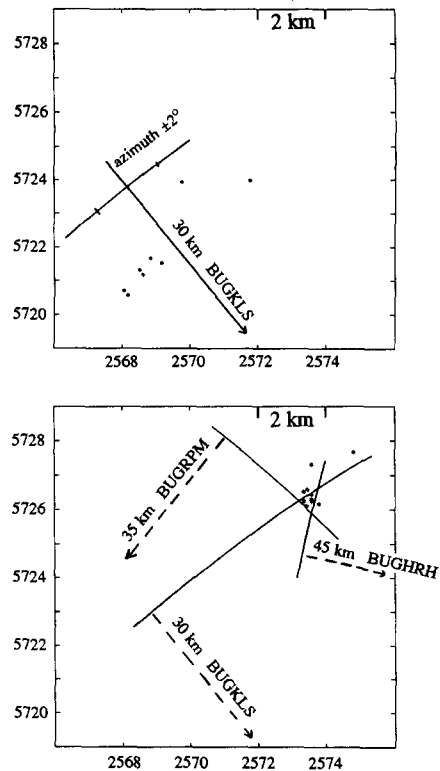


Figure 9. Epicenter maps for manual and automated location of events from Figure 8. Region has same distance to all three stations. Epicenter location by azimuth and distance from KLS—although performed by humans—shows greater deviation and systematic mis-steering compared to automated evaluation which is based on two S-P circles.

the wrong S-onset whereas for 140 events or 20% the polarization image was not stable against azimuthal variations of  $\pm 30^\circ$ . We defined this situation as unstable and suppressed their doubtful results by an additional clause in the phase picker program.

When comparing both maps, significant differences are determined for the regions SW of RPM (bad S/N ratio) and NW of HRH, SW of KLS where the interactive analysis yields more distance clustering. This is true but reflects our additional knowledge by the adjacent station HRH; in former bulletins of 1986 and earlier, this station was missing and human analysis of KLS data did also not resolve the clustering. In the area SW of KLS, however, our routine analysis got too familiar with the knowledge of quarry blast activities in that 'Velbert' region. Instead of the one distance (smeared by azimuthal uncertainty), there are some distinct quarries at different distances (Joswig and Schulte-Theis, 1993), so the automated analysis yields a more unbiased image of the true deviations in the data.

Another comparison is based on 10 'Dorsten' events recorded by two of our three 3-component stations. Figure 8 shows P- and S-phase on the rotated P-traces, Figure 9 gives two epicenter maps. The upper one is based on array azimuths and

distances from KLS (no difference for automated and interactive analysis, two events coincide with previous ones). The lower map is achieved by automated analysis of the 3-component stations, the results have significantly less deviation and indicate either an azimuthal mis-steering of the array beam or a yet to be introduced path dependence for our ground model. Nonetheless, the automated analysis once again yields unbiased results with near human-like performance.

### CONCLUSIONS

Pattern recognition on polarization images has proven to be a practical tool to neutralize the fine-structure perturbations of upper crust in the high-frequency registrations of particle motion. The approach works although we can not model the effects by a complete theory. In this sense pattern recognition is best suited to be concerned with the complexity of real-life data on a heuristic base.

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

### REFERENCES

- Allen, R., 1982, Automatic phase pickers: their present use and future prospects: *Bull. Seism. Soc. America*, v. 72, no. 6, pt. B, p. S225-S242.
- Buchbinder, G. G. R., and Haddon, R. A. W., 1990, Azimuthal anomalies of short period p-wave arrivals from Nahanni aftershocks: *Bull. Seism. Soc. America*, v. 80, no. 5, p. 1272-1283.
- Christoffersson, A., Husebye, E. S., and Ingate, S. F., 1988, Wavefield decomposition using ML-probabilities in modelling single-site-3-component records: *Jour. Geophysics*, v. 93, no. 2, p. 197-214.
- Dziewonski, A., Bloch, S., and Landisman, M., 1969, A technique for the analysis of transient seismic signals: *Bull. Seism. Soc. America* v. 59, no. 1, p. 427-444.
- Ebel, J. E., 1989, The effect of crustal scattering on observed high-frequency earthquake seismograms: *Geophys. Jour. Int.* v. 98, no. 2, p. 329-342.
- Flinn, E. A., 1965, Signal analysis using rectilinearity and direction of particle motion teleseismic signals: *Proc. IEEE*, v. 53, no. 12, p. 1874-1876.
- Joswig, M., 1990a, Pattern recognition for earthquake detection: *Bull. Seism. Soc. America*, v. 80, no. 1, p. 170-186.
- Joswig, M., 1990b, Wissensbasierte Erdbebenerkennung mit Sonogrammen als metalen Bilder, in Marburger, M., ed., *GWAI-90: Springer-Verlag, IFB 251, Berlin*, p. 144-149.
- Joswig, M., 1991, Automated detection and interpretation of earthquake seismograms by adaptive pattern recognition, in Krönig, D. and Lang, M., eds., *Physik und Informatik—Informatik und Physik, Springer-Verlag, IFB 306, Berlin*, p. 153-161.
- Joswig, M., 1993a, Single-trace detection and array-wide coincidence association of local earthquakes and explosions: *Computers & Geosciences*, v. 19, no. 2, p. 207-221.
- Joswig, M., 1993b, Automated seismogram analysis for the tripartite BUG array: an introduction: *Computers & Geosciences*, v. 19, no. 2, p. 203-206.
- Joswig, M., and Schulte-Theis, H., 1993, Master event correlation of weak local earthquakes by Dynamik Waveform Matching: *Int. Geophys. Jour.*, in press.
- Jurcevic, A., 1988, Polarization analysis of three-component array data: *Bull. Seism. Soc. America*, v. 78, no. 5, p. 1725-1743.
- MacBeth, C. D., and Redmayne, D. W., 1989, Source study of local coalfield events using the modal synthesis of shear and surface waves: *Geophys. Jour. Int.* v. 99, no. 1, p. 155-172.
- Menke, W., Lerner-Lam, A. L., Dubendorff, B., and Pachero, J., 1990, Polarization and coherence of 5 to 30 Hz seismic wave fields at a hard-rock site and their relevance to velocity heterogeneities in the crust: *Bull. Seism. Soc. America*, v. 80, no. 2, p. 430-449.
- Montalbetti, J., and Kanasewich, E. R., 1970, Enhancement of teleseismic body phases with a polarization filter: *Royal Astr. Soc., Geophys. Jour.*, v. 21, no. 2, p. 119-129.
- Plesinger, A., Hellweg, M., and Seidl, D., 1986, Interactive high-resolution polarization analysis of broad-band seismograms: *Jour. Geophysics*, v. 59, no. 2, p. 129-139.
- Rioul, O., and Vetterli, M., 1991, Wavelets and signal processing: *IEEE SP Magazine*, v. 8, no. 4, p. 14-38.
- Roberts, R. G., and Christoffersson, A., 1990, Decomposition of complex single-station three-component seismograms: *Geophys. Jour. Int.* v. 103, no. 1, p. 55-74.
- Roberts, R. G., Christoffersson, A., and Cassidy, F., 1989, Real-time event detection phase identification and source location estimation using single station three-component seismic data: *Geophys. Jour. Int.* v. 97, no. 3, p. 471-480.
- Ruud, B. O., Husebye, E. S., Ingate, S. F., and Christoffersson, A., 1988, Event location at any distance using seismic data from a single, three-component station: *Bull. Seism. Soc. America*, v. 78, no. 1, p. 308-325.
- Ruud, B. O., and Husebye, E. S. 1992, A new three-component detector and automatic single station bulletin production: *Bull. Seism. Soc. America*, v. 82, no. 1, p. 221-237.



APPENDIX

This is a numerical example intended to explain the phase picking on polarization images. The data matrices represent two frequencies and 10 time increments (5 samples each). All vector data, for example Z-NS-EW or P-SH-SV are framed into one box. The initial data  $l_{max}$  represent log energy to ease the calculation of  $thres = mean + variance$  for the log normal distributed noise window. The rotation operator is not commutative, here it is given for the correct sequence of incidence after azimuth. Our example is given for SH-onset picking, thus we must select the SH combination rule and the SH pattern.

**DETECTION ON POLARIZATION IMAGES EXAMPLE (SH-Onset)**

Select Largest Eigenvalue  $l_{max}[f][t]$  with its  
 Eigenvector  $e_{Z,NS,EW}[f][t]$   
 (notice:  $e_z^2 + e_{NS}^2 + e_{EW}^2 = 1$ )

$l_{max}$	6	6	4	6	4	5	8	6	8	7
	2	3	4	6	5	9	7	8	6	7
$e_z$	.9	.7	.9	.9	.9	.5	.7	.1	.0	
	.4	.7	.9	.9	.9	.6	.8	.8	.6	.4
$e_{EW}$	.2	.2	.5	.1	.3	.1	.3	.5	.7	.8
	.7	.5	.4	.2	.1	.1	.4	.4	.6	.7
$e_{NS}$	.1	.7	.1	.5	.1	-.3	.8	-.5	-.7	-.6
	-.6	.5	.1	.2	.2	.8	-.5	-.4	-.5	-.6

Rotate to  $e_{P,SH,SV}[f][t]$  with Azimuth  $a$  and Incidence  $i$   
 for  $(f,t) \{ e_p = e_z * \cos(i) - e_{NS} * \sin(i) * \cos(a) - e_{EW} * \sin(i) * \sin(a);$   
 $e_{SH} = -e_{NS} * \sin(a) + e_{EW} * \cos(a);$   
 $e_{SV} = e_z * \sin(i) + e_{NS} * \cos(i) * \cos(a) + e_{EW} * \cos(i) * \sin(a); \}$   
 Example:  $a=52^\circ, i=10^\circ$

$e_p$	.9	.6	.8	.8	.8	.9	.6	.7	.1	-.0
	.3	.6	.9	.9	.9	.5	.8	.8	.6	.3
$e_{SH}$	.2	.4	.2	.3	.1	.3	.8	.7	.9	.9
	.9	.0	.3	.3	.0	-.5	.6	.6	.8	.9
$e_{SV}$	.3	.7	.6	.5	.5	.6	-.2	.2	.1	.3
	.2	.8	.4	.2	.4	.7	.1	.2	.3	.2

Phase Criterion: Reduce to Main Component  
 for  $(f,t) \{ P_p [f][t] = (e_p [f][t] < 0.5) ? 0 : 1;$   
 $P_{SH}[f][t] = (e_{SH}[f][t] < 0.5) ? 0 : 1;$   
 $P_{SV}[f][t] = (e_{SV}[f][t] < 0.5) ? 0 : 1; \}$

$P_p$	1	1	1	1	1	1	1	1	0	0
	0	1	1	1	1	1	1	1	1	0
$P_{SH}$	0	0	0	0	0	0	1	1	1	1
	1	0	0	0	0	1	1	1	1	1
$P_{SV}$	0	1	1	1	1	1	0	0	0	0
	0	1	0	0	0	1	0	0	0	0

Energy Criterion: Reduce to Significant Amplitudes (6)  
 for  $(f,t) \{ E[f][t] = (l[f][t] < thres[f]) ? 0 : 1; \}$  (5)

$E$	1	1	0	1	0	0	1	1	1	1
	0	0	0	1	1	1	1	1	1	1

Combine both Criteria to the Data Image  $D[f][t]$

Detection of P-Onset			Detection of SH-Onset		
D	P=0	P=1	D	P=0	P=1
E=0	-1	-1	E=0	-1	0
E=1	0	+1	E=1	-1	+1

$D_p$	1	1	-1	1	-1	-1	1	1	0	0
	-1	-1	-1	1	1	1	1	1	1	0
$D_{SH}$	-1	-1	-1	-1	-1	-1	1	1	1	1
	0	-1	-1	-1	-1	1	1	1	1	1
$D_{SV}$	-1	1	0	1	0	0	-1	-1	-1	-1
	-1	0	-1	-1	-1	1	-1	-1	-1	-1

Take Appropriate Generic Polarization Pattern  $M[f][t]$

P-Onset			SH-Onset		
$-1/3$	$-1/3$	$-1/3$	0	0	0
$-1/3$	$-1/3$	1	0	0	$-1/3$
$-1/3$	$-1/3$	0	$-1/3$	$-1/3$	$-1/3$
$-1/3$	$-1/3$	0	$-1/3$	$-1/3$	1
$-1/3$	$-1/3$	0	$-1/3$	$-1/3$	1
$-1/3$	$-1/3$	0	0	0	0
$-1/3$	$-1/3$	0	0	0	$-1/3$

SH-Onset		
0	0	0
0	0	$-1/3$
$-1/3$	$-1/3$	$-1/3$
$-1/3$	$-1/3$	1
$-1/3$	$-1/3$	1
0	0	0
0	0	$-1/3$

Calculate  $fit[t]$  by Cross-Correlation

(detection is maximum of  $fit[t]$  if above threshold)  
 $acm=0;$  for  $(f,t)$  for  $(P,SH,SV) acm+=pow(M_x[f][t],2);$   
 for  $(t) \{ ccf=0;$  for  $(f,t)$  for  $(P,SH,SV)$   
 $ccf+=M_x[f][t]*D_x[f][t+tt]; fit[t]=ccf/acm; \}$

$fit[t]$	-	-	-	$-10/3$	$-1/3$	$7/3$	18/3	17/3	15/3	-
----------	---	---	---	---------	--------	-------	------	------	------	---