Artificial Intelligence Techniques applied to Seismic Signal Analysis

by **Manfred Joswig**

Lehrstuhl für Meßtechnik, RWTH Aachen, D-52056 Aachen

Abstract

A survey on different approaches of Artificial Intelligence is given for the task of automated seismogram analysis. Expert systems and pattern recognition are based on conscious knowledge descriptions, they extend some known approaches for seismic signal processing like STA/LTA detection and the voting scheme. Without AI, the main difference was in network versus array processing. Now we must also differenciate between dense and sparse layouts as the quantity of stations will determine the quality of best suited AI techniques. By many stations we get a good S/N ratio for simple feature extraction, but complex rule systems are necessary to avoid a combinatoric explosion. A network or array with just a few stations needs sophisticated single-trace analysis based on PR while a rudimentary expert system is sufficient.

More recently a new branch of 'Bottom-Up' techniques got attention of AI. Besides artificial neural networks, it's made up by simulated annealing and genetic algorithms. Applications to seismology are still rare and lack the full power of unattended self-adaptation as convergence in training cycles is still too low. However, the available examples promise to extend our capabilities of recognition and optimization to tasks that lack sufficient theory for analytical solutions.

Introduction

Over the last 50 years, the research on Artificial Intelligence has produced many different approaches to copy the human reasoning and recognition capabilities. Starting with the early vision of the Perceptron, methods range from Expert Systems and Pattern Recognition to speech understanding, computer vision and autonomous learning. They copy mechanisms from biology in Artificial Neural Networks and Genetic Algorithms or from physics in the crystal growth of Simulated Annealing. Each approach is unique in its features and has its own history of success and failure. And history often repeats under new brand like for the revival of the Perceptron in ANN's or for the vitalizing of Expert Systems by Fuzzy Logic.

Where can we use this treasure for our problems in seismology? To answer, first we should stress the principal distinction of AI into the two branches 'Top-Down' and 'Bottom-Up' as in Fig. 1. Both ways reflect completely different ways for knowledge representation as one of the crucial entities of any AI system. For the 'Top-Down' branch, AI can be seen as a collection of programming techniques to copy concious human reasoning by computer - like numerics is a collection of algorithms to solve analytical equations. One paradigm of AI is the separation between knowledge and inference engine to ease the repeated update by knowledge incrementals. Important aspects are the details in knowledge representation (frames, scripts, semantic nets, grammers, images, ..), the

Explicit Knowledge Representation 'Top-Down'	Self-Adaptation of Internalities ('Bottom-Up')
Knowledge-based System ('Expert System') - Truth Maintenance System - Blackboard System	Artificial Neural Network - Backpropagation Network - Kohonen Feature Maps
Fuzzy Logic	Simulated Annealing
Pattern Recognition - Pathfinding on Syntactic Description - Adaptation of Mental Images	Genetic Algorithm

Fig. 1 Knowledge Representation in AI and Related Techniques

uncertainty in reasoning and its inheritance through different levels of the reasoning chain and finally the restriction of search domain to bypass the combinatoric explosion.

For 'Bottom-Up', we follow a different path. It starts with the specification of an atomic cell like one artificial neuron or one member of population classes in genetic algorithms. Then we define rules of interdependence and of adaptation to environmental conditions. To get the system running, we accumulate a reasonable number of atomic cells and expose it to the environment. Then we see - and hope on 'intelligence'. But what's with knowledge representation? Like a child that counts 2 and 2, we can't say how it does and we don't even care, as long as it counts 4. In 'Bottom-Up' the representation of knowledge is hidden to internal weights, maps or character strings - nothing what would make sense for us. So this method of AI is best for unexplored situations without sufficient theory or rule systems.

The research in AI may be as fascinating as will, we are looking for working applications in seismology. AI is still in its infant state, it lacks the matured and powerful tools to handle the full complexity of real world situations. So one conclusion is that we can't expect AI to work as general replacement of humans - even in the simple job of routine scientific evaluations. To be more specific for our application, each greater earthquake is an 'individuum' with new characteristics. There will be no chance for computer solutions based purely on comparison with known situations. On the other hand, we must handle many small and similar looking events, aftershock series or active noise sources. This could be the framework for automated solutions: situations are repetive (i.e., suited for comparison techniques), frequent (i.e., there is a need to off-load humans) and less important. So no extreme reliability is necessary and the dependence on unattended computer decisions might seem tolerable.

One last remark before we start the description of specific applications. AI techniques should solve our urgent needs in seismic data processing. They have been existent long before AI got our attention. So there will always be a history of prior solutions based on classical processing schemes and programming techniques. If successful, AI should improve performance to lower S/N ratio and higher reliability.

Detection on Single Traces

The history of automated seismogram processing started with the STA/LTA detector based on the theoretical work of Freiberger (1963). It was the first time that seismology entered the new field of unmaned data interpretation and was confronted with criteria like false alarms, missed events or the receiver operating characteristic (ROC). While everybody knew that human work is by no means without error, they were usually taken as god-given. Automated processing was different as it demands a conscious discussion and decision about performance versus security trade-offs. Like in a burning glass, all the discussions about automatization have focused on the first step of single-trace detectors. This is in sharp contrast to their actual simplicity, similarity and predictable performance. As explained for the STA/LTA detector in Fig. 2, even with the additional step of postpro-

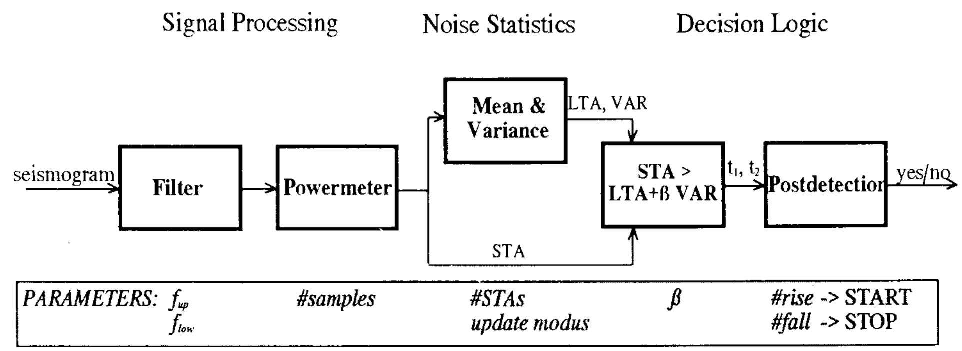

Fig. 2 Basic Elements of STA/LTA-like Detectors

cessing it demands just 8 parameters for optimum tuning. But experience has shown that - by not understanding their impact - we can choose these parameters so to miss any event or to get it masked in numerous false alarms as in Fig. 3. The principle processing units - filter, powermeter, statistics and decision logic - are found in all variants of STA/LTA like the WALSH-detector (Goforth & Herrin, 1981), the SRO-detector (Murdock & Hutt, 1983) or the approaches from Stewart (1977) and Allen (1978) - for a more detailed comparison, see Joswig (1990). The inherent paradigm of all approaches - any deviation from the permanent noise process is detected as event - is also the cause of their common and principle weakness: They *must* trigger for every sonic bang, traffic noise, explosion or spiky interference to produce the false alarms.

Coincidence Reasoning

This inherent weakness of single-trace detectors is only tolerable if additional processing can compensate for it. The most simple and earliest approach was voting, i.e., the test on coincidence in detection timing at different sites. It is quite a powerful criterion since noise sources are local but propagation of seismic waves is spatial. Only explosions and casual correlation of noise bursts will

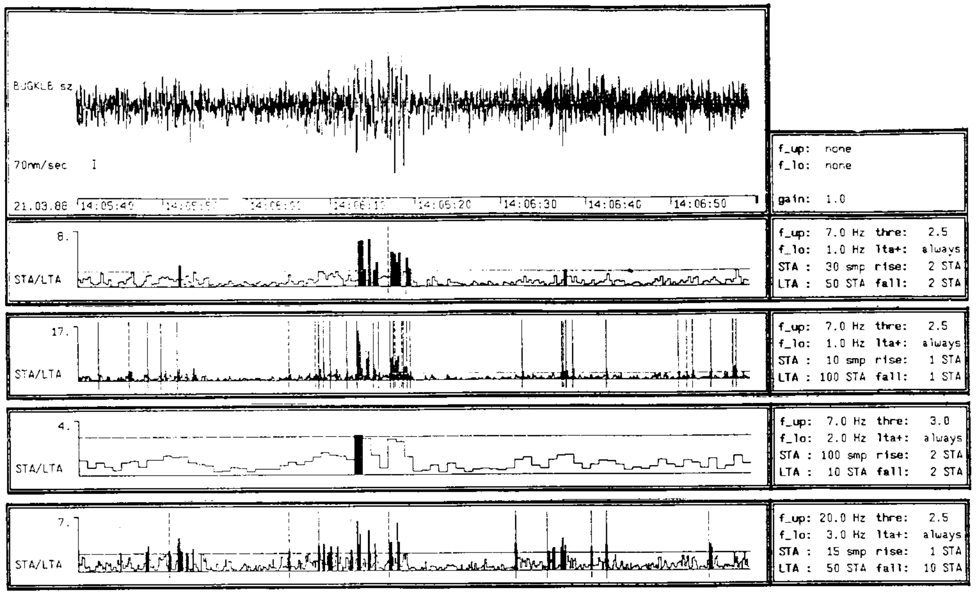

Fig. 3 Performance of STA/LTA by different Parameter Values

pass this test unjustified. If still there is demand for more sophisticated reasoning, then it comes from the needs of postprocessing. The price we payed for the achieved robustness in the initial detection and voting process is fuzzyness. Our final aime is the automated bulletin, but up to now the results are unsufficient for any event analysis, e.g., the location of hypocenters. Most urgently, we need a more accurate picking of onset times. It can be realized by just another tuning of the STA/LTA parameters to yield more sensitivity. In the example of Fig. 3, the first trace (STA=30, ..) was the best detector; it triggers the event without significant false alarms. But for location, we missed the correct P-onset time at 14:06:08. It is found by the 'phase picker' in the last row with (STA=15, ..) - on cost of its much higher false alarm rate. The efficiency of phase picking could be enhanced very much if we guide it by prior knowledge. One way is the restriction to predefined, very narrow detection windows. To derive them, we could start our analysis on a subset of stations with good S/N ratio, i.e., around the actual event. Once we have the intial hypocenter guess, we can proceed to weaker traces by tests on plausibility with standard travel time tables (Anderson, 1978).

This kind of reasoning is performed best by expert systems. Either they follow the scheme of truth maintenance (Johnson, 1986; Johnson et al., 1987; Searfus, 1989) or they implement a 'blackboard' intercommunication model (Chiaruttini et al., 1989; Chiaruttini, 1991, 1992; Chiaruttini & Salemi, 1993; Roberto & Chiaruttini, 1992). Once the rule base is adopted to the specific situation, we can expect a rate of 80 - 90% success for the automated bulletin.

But why did we choose the knowledge-based system (KBS) technique? Is it best suited for the given task? Which implications for problem solving did we accept? One answer is history: KBS was first of all AI techniques that gained acceptance outside its own community. Famous applicati-

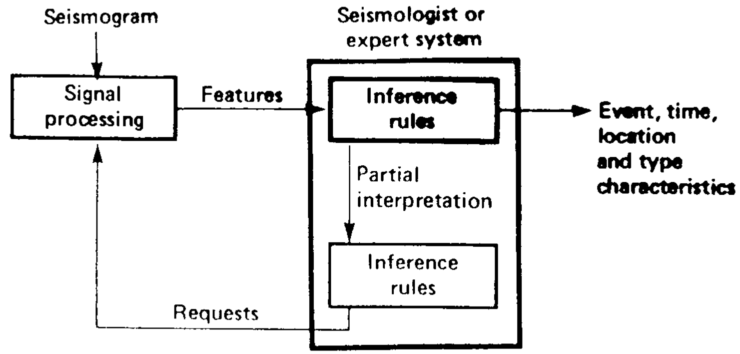

Fig. 4 Buildings Blocks of an Expert System in Seismology (after Johnson, 1986)

ons are MYCIN in medicine or the 'General Problem Solver' in logics. Another reason is in the operations that we want to perform on entities like events or phases. KBS rules match well to our needs of *symbolic* reasoning. However, the strength of KBS is also its weakness: it *must* operate on symbolic entities, so we need the additional modules of signal processing like in Fig. 4. They derive *features* from our initial seismograms being *subsymbolic* entities. These modules can also perform the 'guided' postprocessing based on requests of the expert system. In this scheme, the power of KBS will compensate for the simple frontends in detection and phase picking.

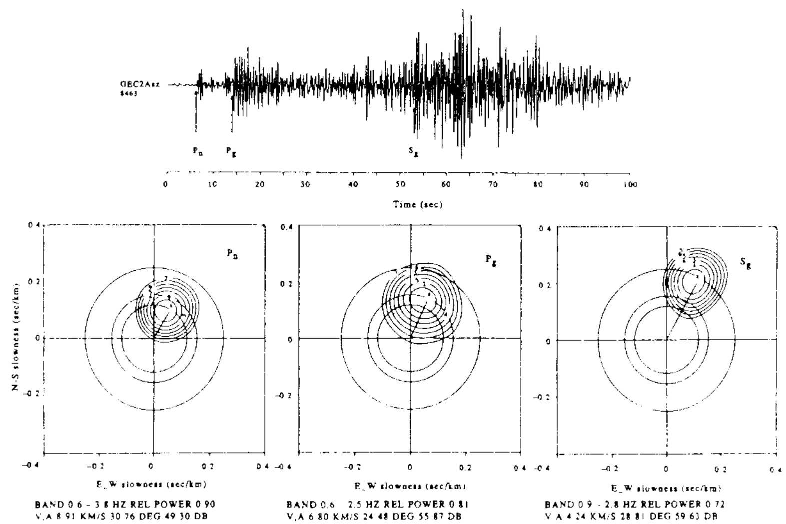

Fig. 5 Test on Common Azimuth in Array Analysis (after Harjes et al., 1993)

Parameter Extraction for Seismic Arrays

The situation for automated seismogram processing gets completely different in case of seismic arrays. Now we can utilize the coherence in station signals to enhance the S/N ratio by array beamforming. This processing scheme is superior to voting (Wirth et al., 1976). But we can benefit from the spatial sampling of the wavefield in a more significant way. Imagine, we must group a mixture of different phase onsets to reasonable sequences from possibly distinct earthquakes. Based on the test of common azimuths via f-k analysis, this gets quite an easy task (see Fig. 5). The f-k analysis can also be used to distinguish between P- and S-phases by their significantly different slownesses. In fact, there is no simpler S-onset picking as in Fig. 6.

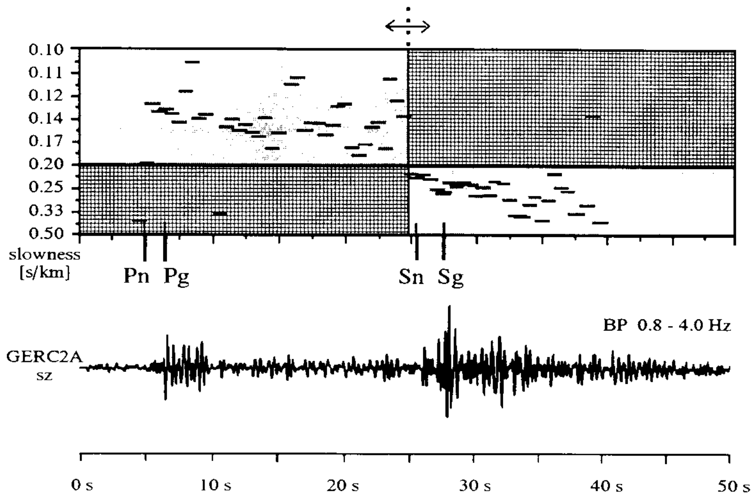

Fig. 6 S-onset Determination by Slowness Matching

To select the most suited AI approach for these powerful frontends, we once again choose symbolic reasoning. So the scheme of KBS as in Fig. 4 stays applicable. The most advanced example is IMS at NORSAR (Bache et al., 1990, 1993; Bratt et al., 1990). Its automated seismogram processing is performed by RONAPP (Mykkeltveit & Bungum, 1984).

Monitoring by Sparse Layouts

The situation changes once again if we can't base our evaluation on many seismic stations. Either a small event is recorded only by its adjacent stations. Or we just face the always restricting limits on funding resulting in the minimum investment of a tripartite array site (see Fig. 7). It is important to note that this simple change in *quantity* of sites converts to a distinct *quality* in automated seismogram processing (Joswig, 1992). In case of a sparse network, no redundancy is left to identify and correct wrong phase pickings. For local seismicity, especially S-onset times are very hard to determine. The beam trace of a sparse array is often worse than the best single station, so

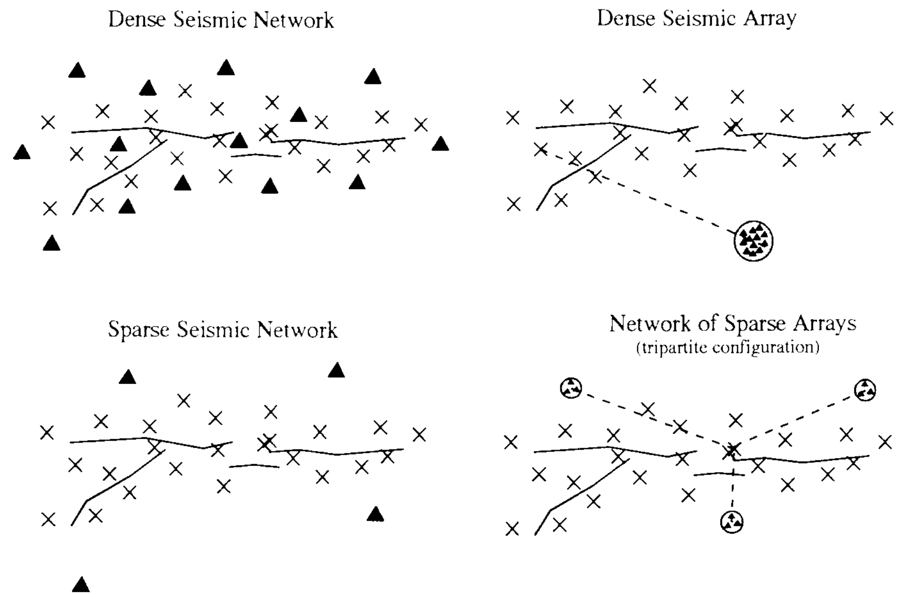

Fig. 7 Seismic Monitoring by Different Geographical Layouts

coherence processing can not help. Then we must rely on voting will yields many false alarms because of the limited aperture and station number. Finally, postprocessing can't be based on f-k analysis due to the unsufficient spatial sampling.

As a result of these restrictions, the efficiency of parameter extraction is much more significant for sparse layouts than for dense station distributions. This demands more intelligent approaches than just detection by STA/LTA or any of its relatives. The fundamental alternative is pattern recognition. It is based on suited descriptions of known signals and noise bursts, either by grammars (Anderson, 1982) or by clustered segments in syntactic PR (Liu & Fu, 1983; Roberto et al., 1990; Zhizhin et al., 1993). Another alternative is based on knowledge representation by images (Joswig, 1994a); its application to event detection utilizes sonograms (Joswig, 1990, 1993b, 1994b). As we work on seismograms, symbolic reasoning by KBS is not applicable. However, PR could also be performed by ANN's but we'll delay their discussion to a separate chapter.

The sparse station layout does not only influence the initial task of event detection, it determines all aspects of automated processing. The goal for detection was robustness but it's achieved on cost of fuzzyness in event characterization and onset timing. So subsequent system layers - KBS approaches and further knowledge-based waveform processing like PR - are necessary to derive the parameters suited for an automated bulletin. Their design to a complete system like in Fig. 8 must be well balanced to benefit from complementary strengths and weaknesses (Joswig, 1993a).

For KBS, the restricted number of stations limits the variety of situations that must be handled. While this will simplify the system, PR detectors yield more parameters than STA/LTA to perform the reasoning on coincidence. So KBS for sparse layouts will significantly differ from dense

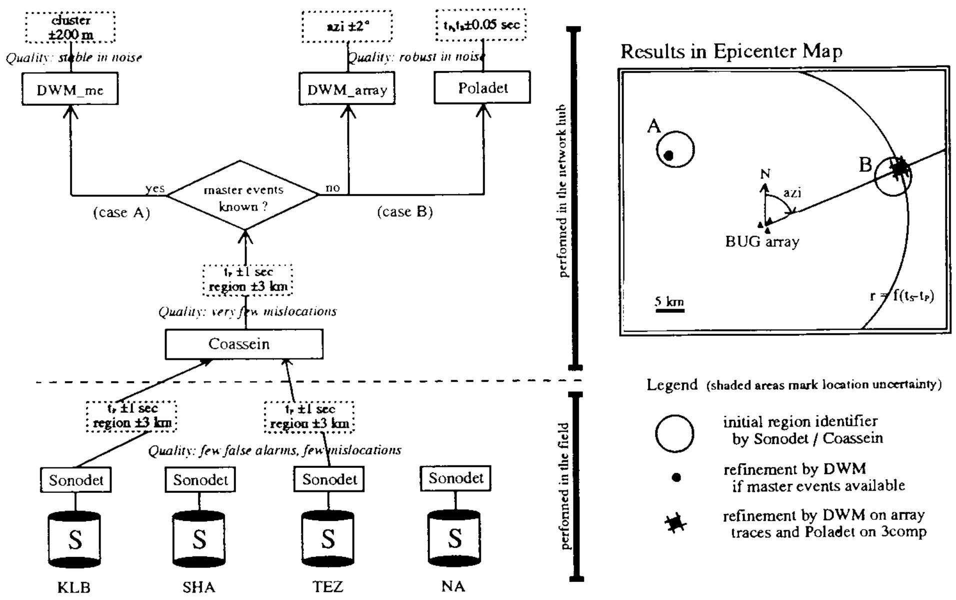

Fig. 8 Modules for Automated Seismogram Processing in the BUG Small Array

network or array systems (Joswig, 1993b, 1994c). Then the further path of refinement depends on the availability of additional knowledge. If we have a set of master events for the identified source region, non-linear correlation by DWM is the most accurate way (Schulte-Theis & Joswig, 1993; Joswig & Schulte-Theis, 1993). Else we must determine the array azimuth via DWM and the distance by pattern-based phase picking on three-component data (Klumpen & Joswig, 1993).

To sum it up, the trade-off between dense and sparse layouts is simple parameter extraction and complex expert system versus sophisticated parameter recognition and rudimentary reasoning. Other criteria are in common like a multi-layer approach starting by fuzzy recognition and then proceeding to hypothesis-guided postprocessing. For both approaches, the expectable rate of success ranges from 80 - 90% in routine observatory situations.

'Bottom-up' in Seismology

All approaches discussed so far - either expert systems or PR on syntax or images - fall into the class of 'Top-Down' methods. They utilize the explicit coding of knowledge in consciously performed descriptions. The principle alternative is 'Bottom-Up', i.e., constructing a self-learning algorithm that independently adapts itself to the environment. In consequence we can't expect to reason, justify or modify on its internal knowledge representation. Either the system works or we must continue in its 'training phase'. A very delicate question is the prediction of performance and reliablity of such an independent system.

Promissing applications for artificial neural networks (ANN) are any kind of recognition task. However, the severe weakness to date is the slow convergence in training cycles. This excludes direct application to the high dimensionality of real world data. Instead we must break the problem to a set of managable parameters. This is achieved by very carefully chosen preprocessing units like decimation filters, determination of spectral weights or other *intendedly* defined signal processing (Dowla et al., 1990; Dysart & Pulli, 1990; Falsaperla et al., 1993). In general, these prefilters will dominate the problem solving while our ANN is just one of several possible PR units. Instead of a truely self-adapting system in the spirit of 'Bottom-Up', we find ANN as the top of cream on the cage of 'Top-Down' processing. However, this comment should by no means disqualify 'Bottom-Up' approaches. They really are very promissing tools for handling situations which are beyond our specification capabilities - once the self-organization is accelerated sufficiently.

Simulated annealing and genetic algorithms are well suited to approximate the inversion of multi-dimensional performance functions even if no local derivative can be given. So these tools can extend our capabilites in seismic network optimization (Scherbaum & Hardt, 1993) and hypocenter location (Sambridge & Gallagher, 1993) to a new level of complexity.

Conclusions

The task of automated seismogram processing shows - besides its necessity to be solved for any future progress in seismic monitoring - considerable complexity that excludes the application of simple solutions. As our problems proceed from detection to suppression of false alarms, then location and finally identification, so does the diversification of algorithms. Starting with optimal filters for STA/LTA, we separate into pattern recognition versus expert systems, then we subdivide by knowledge representation schemes with images or syntax, frames, scripts or fuzzy logic. The more details our coded knowledge includes, the less portable the system gets.

To date the only firm ground are preliminary recommendations: Field stations should rely on STA/LTA-like detectors as long as we can't forsee the desired signals. Observatories with many station - either array or network design - will focus on expert systems based on a more primitive parameter extraction for detection and guided postprocessing. Sparse layouts must rely on sophisticated single-trace algorithms complemented by more simple rules and a knowledge-based parameter refinement once again by PR.

While it may be true that AI enlarges our knowledge about nature, its application to seismology is driven just by the opposite situation. Once we have the framework of complete theories, we can calculate all desired effects. Obviously nobody would expect to replace the formula of Love and Rayleigh waves by some fuzzy recognition of underived patterns. For seismology, the power of AI is in handling our uncomplete knowledge in the tentative processing of real world data.

Literature

Allen, R., 1978. Automatic earthquake recognition and timing from single traces, *Bull. seism. Soc. Am.*, **68**, 1521-1532.

Anderson, K. R., 1978. Automatic analysis of microearthquake network data, *Geoexploration,* **16**, 159-175.

Anderson, K. R., 1982. Syntactic analysis of seismic waveforms using augmented transition network grammers, *Geoexploration,* **20**, 161-182.

Bache, T. C., Bratt, S. R., Wang, J., Fung, R. M., Kobryn C. & Given, J. W., 1990. The intelligent monitoring system, *Bull. seism. Soc. Am.*, **80**, 1833-1851.

Bache, T. C., Bratt, S. R., Swanger, H. J., Beall, G. W. & Dashiell, F. K., 1993. Knowledge-based interpretation of seismic data in the Intelligent Monitoring System, *Bull. seism. Soc. Am.*, **83**, 1507-1526.

Bratt, S. R., Swanger, H. J., Stead, R. J., Ryall, F. & Bache, T. C., 1990. Initial results from the intelligent monitoring system, *Bull. seism. Soc. Am.*, **80**, 1852-1873.

Chiaruttini, C., Roberto, V. & Saitta, F., 1989. Artificial intelligence techniques in seismic signal interpretation, *Geophys.J. Int.*, **98**, 223-232.

Chiaruttini, C., 1991. Focus-of-attention techniques in the automatic interpretation of seismograms, *Pure Appl. Geophys.*, **135**, 61-75.

Chiaruttini, C., 1992. Perceptual and conceptual aspects in seismogram understanding, *Cahiers du Centre Europeen de Geodynamique et de Seismologie*, **6**, 3-12.

Chiaruttini, C. & Salemi, G., 1993. Artificial intelligence techniques in the analysis of digital seismograms, *Computers and Geosciences*, **19**, 149-156.

Dowla, F. U., Taylor, S. R. & Anderson, R. W., 1990. Seismic dicrimination with artificial neural networks: preliminary results with regional spectral data, *Bull. seism. Soc. Am.*, **80**, 1346-1373.

Dysart, P. S. & Pulli, J. J., 1990. Regional seismic event classification at the NORESS array: seismological measurements and the use of trained neural networks, *Bull. seism. Soc. Am.*, **80**, 1910-1933.

Falsaperla, S., Graziani, S., Nunnari, G. & Spampinato, S., 1993. An application of neural networks for the classification of the seismic events at Stromboli volcano (Sicily, Italy), *Cahiers du Centre Europeen de Geodynamique et de Seismologie*, **6**, 13-21.

Freiberger, W. F., 1963. An approximate method in signal detection, *Quarterly Appl. Math.*, **20**, 373-378.

Goforth, T. & Herrin, E., 1981. An automatic seismic signal detection algorithm based on the Walsh transform, *Bull. seism. Soc. Am.*, **71**, 1351-1360.

Harjes, H.-P., Jost, M., Schweitzer, J. & Gestermann, N., 1993. Automatic seismogram analysis at GERESS, *Computers and Geosciences*, **19**, 157-166.

Johnson, R. R., 1986. An expert system for interpreting NORSAR events, *Proc. RSTN/NORESS research symposium*, ed. K. K. Nakanishi & D. R. Breding, 160-168, LLNL, Livermore CA.

Johnson, R. R., Canales, T. W., Lager, D. L., Mason, C. L. & Searfus, R. M., 1987. Interpreting signals with an assumption-based truth maintenance system, *Proc. Int. Soc. Opt. Eng.*, **786**, 332-337.

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

Joswig, M., 1992. System architecture of seismic networks and its implications to network automatization, *Cahiers du Centre Europeen de Geodynamique et de Seismologie*, **5**, 75-84.

Joswig, M. & Schulte-Theis, H., 1993. Master event correlation of weak local earthquakes by Dynamik Waveform Matching, *Geophys. J. Int.*, **113**, 562-574.

Joswig, M., 1993a. Automated seismogram analysis for the tripartite BUG array: an introduction, *Computers and Geosciences*, **19**, 203-206.

Joswig, M., 1993b. Single-trace detection and array-wide coincidence association of local earthquakes and explosions, *Computers and Geosciences*, **19**, 207-221.

Joswig, M., 1994a. Knowledge-based seismogram processing by mental images, *IEEE trans. System Man Cybern.*, **24**, (in press).

Joswig, M., 1994b. Automated classification of local earthquake data in the BUG small array Part I: single-trace recognition of sonograms, *Geophys. J. Int.* (in press).

Joswig, M., 1994c. Automated classification of local earthquake data in the BUG small array Part II: rule-based coincidence evaluation, *Geophys. J. Int.* (in press).

Klumpen, E. & Joswig, M., 1993. Automated reevaluation of local earthquake data by application of generic polarization patterns for P- and S-onsets, *Computers and Geosciences*, **19**,223-231.

Liu, Hsi-Ho & Fu, King-Sun, 1983. An application of syntactic pattern recognition to seismic discrimination", *IEEE trans. Geosci. Remote Sens.*, **21**, 125-132.

Murdock, J. N. & Hutt, C. R., 1983. A new event detector designed for the Seismic Research Observatories, USGS Open-file report 83-785, Albuquerque NM.

Mykkeltveit, S. & Bungum, H., 1984. Processing of regional seismic events using data from small-aperture arrays, *Bull. seism. Soc. Am.*, **74**, 2313-2333.

Roberto, V. & Chiaruttini, C., 1992. Seismic signal understanding: a knowledge-based recognition system, *IEEE trans Signal Proc.*, **40**, 1787-1806.

Roberto, V., Paglietti, P. & Chiaruttini, C., 1990. Syntactic filtering and recognition of wide-band noise waveforms, *Signal Processing*, **19**, 43-60.

Sambridge, M. & Gallagher, K., 1993. Earthquake hypocenter location using genetic algorithms, *Bull. seism. Soc. Am.*, **83**, 1467-1491.

Scherbaum, F. & Hardt, M., 1993. Optimizing the station distribution of seismic networks for after-shock recordings by simulated annealing, *Cahiers du Centre Europeen de Geodynamique et de Seismologie*, **6**, 237-251.

Schulte-Theis, H. & Joswig, M., 1993. Clustering and location of mining induced seismicity in the Ruhr basin by automated master event comparison based on Dynamic Waveform Matching (DWM), *Computers and Geosciences*, **19**, 233-241.

Searfus, R. M., 1989. Using expert systems in treaty verification, *Energy and Techn. Review,* No 1, 26-33, Livermore CA.

Stewart, S. S., 1977. Real-time detection and location of local seismic events in Central California, *Bull. seism. Soc. Am.*, **67**, 433-452.

Wirth, M. H., Blandford, R. R. & Shumway, R. H., 1976, Automatic seismic array and network detection, *Bull. seism. Soc. Am.*, **66**, 1375-1380.

Zhizhin, M., Gvishiani, A., Bottard, S., Mohammadioun, B. & Bonnin, J., 1993. Classification of strong motion waveforms from different geological regions using syntactic pattern recognition scheme, *Cahiers du Centre Europeen de Geodynamique et de Seismologie*, **6**, 33-42.