NORSAR

ROYAL NORWEGIAN COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

Scientific Report 1-78/79

FINAL TECHNICAL SUMMARY

1 April - 30 September 1978

D. Rieber-Mohn (ed.)

Kjeller, October 1978

VI.6 General Purpose Program for Seismic Discrimination

One of the current projects at NORSAR is to investigate the discrimination power of the Lg-phase for near-field observations. In this connection a general program has been completed which is based upon a feature-extraction procedure combined with classification statistics. The idea is to extract as few parameters as possible from the records and still preserve the main information (information pertaining to the second order statistics of the time series like the autocorrelation function or equivalently the power spectrum. The program is general in the sense that all the available information about an event is loaded into the computer and then the information carrying parameters is extracted and subsequently used for classification (for references, see Sandvin and Tjøstheim, 1978).

The input data to the program should be single parameters like the m_b-parameter for body waves, M_s-parameter for surface waves and/or some L_g-parameters considered to have a substantial discrimination potential. Combined with these parameters different time windows with amplitude registration from the seismogram starting from the onset of the chosen phase, may be incorporated in the input data file.

With registrations of the actual phases/wave trains for the same event but from different stations, the various time series may be combined into one multidimensional time series. In case of different frequency response of the seismometers, the traces should be filtered to remove the instrument response from the registrations. The feature extraction procedure consists of two steps. The first step is accomplished by fitting a multivariate autoregressive model of order P to the combined phase registrations described above.

$$\underline{x}(t) - A_1\underline{x}(t-1) \ldots - A_p\underline{x}(t-p) = \underline{z}(t)$$

Here $\underline{x}(t)$ denotes the vector registration at time t, A_i, i=1,2,...p are n x n matrices where n is the number of individual records, and $\underline{z}(t)$ an n-dimensional white noise vector.

The appropriate order p of the model should be determined from a criterion given by Akaike (1971). The fit of the model to the observations is evaluated by checking the whiteness of the residual process $\underline{z}(t)$. If the model is found appropriate, the second order statistics of the multidimensional time series as given by the individual spectra and the cospectra are completely specified from the matrices A_i and the variance matrix V of the white noise process.

The second step of the feature extraction procedure is to combine the parameters contained in the matrices A_i with the single input parameters (m_b, M_s, etc.) from the different stations into a vector $\underline{Y}$ and then apply a principal component analysis to this vector. The idea in the principal component analysis of $\underline{Y}$ is to pick vectors $\underline{h}_i$, i=1,2,...,m in such a way that the main part of the information as expressed by the variation of $\underline{Y}$ is decomposed along a few of the vectors $\underline{h}_i$. The basis vector h_i are given by the eigenvalue problem:

$$R\, \underline{h}_i = \gamma_i \underline{h}_i$$

where R is the covariance matrix $E\{\underline{Y}\,\underline{Y}^T\}$. The estimation of the covariance matrix for earthquakes, R_{EQ} and for explosions R_{EX} requires a data base of presumed earthquakes and of presumed explosions.

Now for each event the estimated vector $\underline{Y}$ is decomposed along M_{EQ} principal vectors $h_{i_{EQ}}$ for the earthquake data base and along M_{EX} principal vectors $h_{i_{EX}}$ for the explosion data base with components

$$Z_{EQ}(i) = \underline{Y}^T \cdot \underline{h}_{i,EQ}, \qquad i = 1,2,\ldots, M_{EQ}$$

and

$$Z_{EX}(j) = \underline{Y}^T \cdot \underline{h}_{j,EX}, \qquad j = 1,2,\ldots, M_{EX}$$

respectively.

The components are finally combined into one vector

$$\underline{Z}^T = [Z_{EQ}(1),\ldots,Z_{EQ}(M_{EQ}),\ Z_{EX}(1),\ldots,Z_{EX}(M_{EQ})]$$

The vector $\underline{Z}$ is then regarded as a stochastic variable with distribution function F_{EQ} or F_{EX} depending on whether the event is an earthquake or an explosion. It is assumed that F_{EQ} and F_{EX} are the multivariate Gaussian distribution with the mean value and covariance matrix for each population determined from the earthquake data base and the explosion data base respectively. The discrimination is then accomplished by a classification procedure where the event is assigned to the population having the highest probability of $\underline{Z}$ occurring. As pointed out (Azen et al, 1975), the classification procedure is relatively robust to deviations from normality. Added flexibility to this particular discrimination is needed in order to handle missing observations and also changes in the number of reporting stations with changing source regions.

At present an extensive analysis of the Lg-phase from near-field (within 20^o) observations obtained from WWSSN is going on. It is concluded that the Lg-phase observations as well as those of Sn are evident from these readings and the amplitudes should be included in the input data to the program described above.

Finally, if the Lg-phase may turn out to be a potential discrimination parameter, SRO-recordings should be provided in order to apply the feature extraction procedure to that section of the records where the Lg-phase is found. 'Identical Classification Procedures' should be applied to different two-dimensional discriminants, like the well-established m_b:M_s criterion, to have the opportunity to give a precise comparison of the different discriminants.

O.A. Sandvin

References

Akaike, H. (1971): Autoregressive model fitting for control, Ann. Inst Stat. Math. 23, 163-180.

Azen, S.P., L. Breiman and W.S. Meisel (1975): Modem approaches to data analysis, Course notes. Technology Service Corporation, Santa Monica, California.

Panza, G.F., and G. Calcagnile (1974): Lg, Li and Rg from Rayleigh Modes, Geophys. J.R. Astron. Soc., 40, 475-487.

Robinson, E.A. (1967): Multichannel time series analysis with digital computer programs, Holden-Day, San Francisco.

Sandvin, O.A., and D. Tjøstheim (1978): Multivariate autoregressive representation of seismic P-wave signals with application to short-period discrimination, Bull. Seism. Soc. Amer., 68, 735-756.

Seismic Discrimination, 31 March 1978. Semiannual Technical Summary, M.I.T. Lincoln Laboratory, Cambridge, Mass.