NORSAR ROYAL NORWEGIAN COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

NORSAR Scientific Report No. 4-75/76

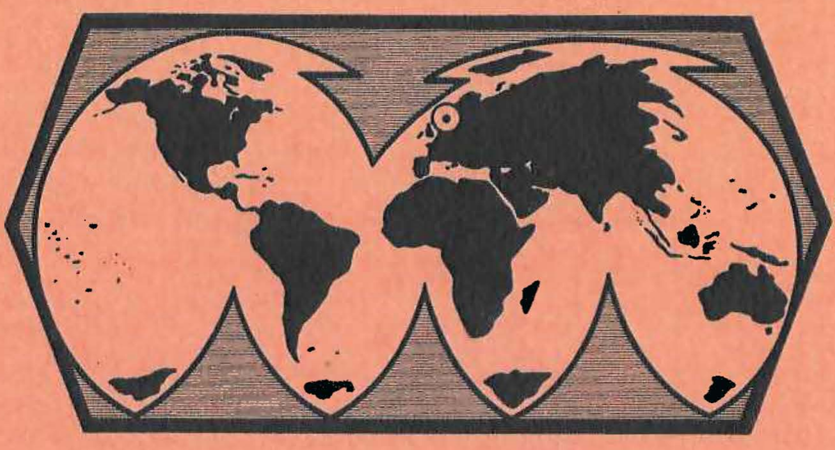
SEMIANNUAL TECHNICAL SUMMARY

1 January – 30 June 1976

Prepared by
K. A. Berteussen

Kjeller, 23 July 1976

Sponsored by
Advanced Research Projects Agency
ARPA Order No. 2551



VII. SUMMARY OF SPECIAL TECHNICAL REPORTS/PAPERS PREPARED

VII.1 A Pattern Recognition Approach to Seismic Discrimination

The task of discriminating between earthquakes and underground nuclear explosions can be formulated as a problem in pattern recognition: On the basis of an observational raw data vector $\underline{X} = [X_1, \ldots, X_N]$ which may represent the digitized short period and long period wave traces from one or more seismological stations, the task is to recognize the vector and to decide which of two populations it belongs to. As a problem in pattern recognition it may be separated into two stages, feature extraction and classification. The feature extraction stage consists of reducing the original data vector $\underline{X}$ to a feature vector $\underline{Z} = [Z(1), \ldots, Z(M)]$ where it is desirable that M is small compared to N while $\underline{Z}$ is still preserving as much information as possible from the original vector $\underline{X}$. The classification then proceeds on the vector $\underline{Z}$.

In the literature on pattern recognition a variety of techniques for feature extraction and classification have been discussed. Curiously enough these methods have not received much attention in seismic discrimination. Motivated by this fact we have initiated a two-stage pattern recognition study of seismic discrimination. Up to now the emphasis has been on feature extraction and some preliminary results are reported in Tjøstheim and Husebye (1976). From a raw data vector $\underline{X}$ with the total number of long period and short period data samples ranging between 3000 and 5000 (depending on epicenter distance) we have constructed a primary feature vector $\underline{Y}$ of dimension 37. The short period features consist of m_b and 9 autoregressive parameters characterizing the signal, coda and the preceding noise. Contrary to common usage we have extracted long period features from Love waves and horizontal Rayleigh waves as well as from vertical Rayleigh waves. Altogether we have used 3 x 9 = 27 long period power spectral estimates computed within various group velocity windows.

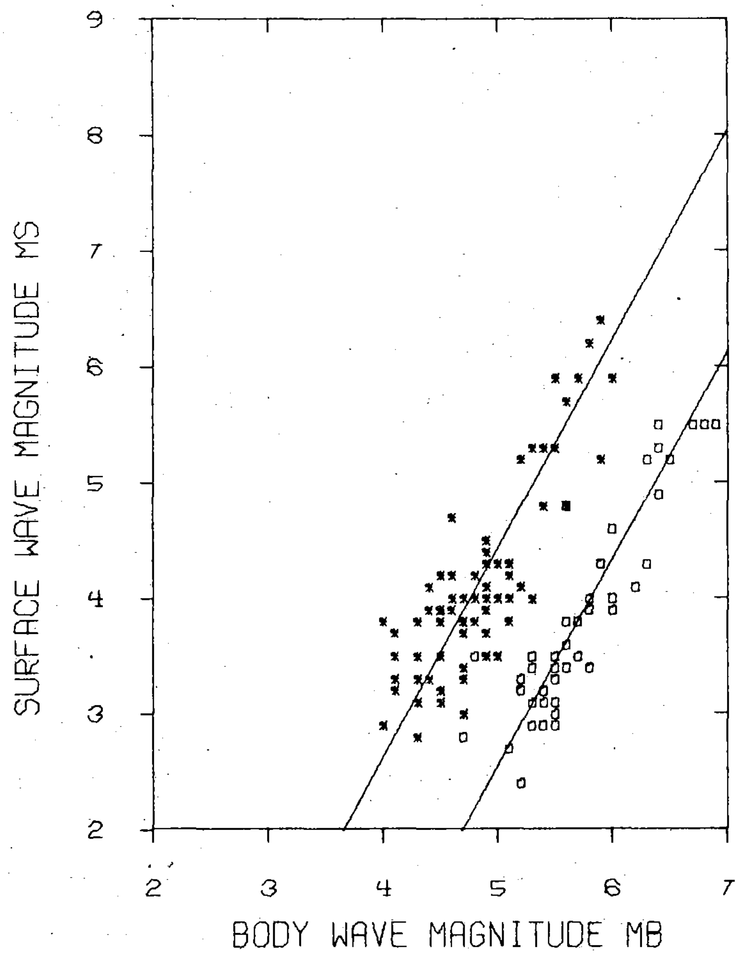


Fig. VII.1.1 $m_b:M_s$ diagram for the Eurasian data set of 52 explosions and 73 earthquakes. PDE m_b and NORSAR M_s values have been used.

We have tested the feature extractors on a data set of Eurasian events containing 52 explosions and 73 earthquakes. An $m_b:M_s$ diagram of the data set is shown in Fig. VII.1.1. To get a rough indication of the quality of the feature extractors, the following generalization of the X1:X2 discriminant of Tjøstheim and Husebye (1976) was studied:

$$X1(A,B) = m_b - B\,\hat{a}_1(S)$$
$$X2(A,B) = E_{20}^{(1)} + A(E_{20}^{(2)} - E_{20}^{(3)}) + B(\hat{a}_1(C) - \hat{a}_1(N)) \qquad \text{(VII.1.1)}$$

Here A and B are scaling parameters and $E_{20}^{(i)}$ are long period energy estimates as defined by Tjøstheim and Husebye (1976). We evaluated the X1(A,B):X2(A,B) discriminant separately for vertical and horizontal Rayleigh waves and Love waves. The results are shown in Fig. VII.1.2, which gives the false alarm rate for the various cases. The figure indicates that the Love wave feature extractors $E_{20}^{(i)}$ are more useful than the corresponding vertical and horizontal Rayleigh features. Also, it is seen that the combination of short period and long period features as in formula (1) is superior to the m_b:M_s discriminant over a wide range of values for the scaling factors A and B, this being true for all three categories of surface waves.

We have also done some experiments to test the appropriateness of a 5th order autoregressive model when computing the E_{20} estimates. Fig. VII.1.3 shows the values of Akaike's (1970) FPE criterion for deciding the "optimal" order for an autoregressive fit to a long period time series generated by an Eastern Kazakh explosion which occurred on 30 Dec 1971. The optimal order is obtained by choosing the order corresponding to the minimum FPE. It is seen that this is close to 20 for the horizontal Rayleigh wave and close to 30 for the Love and vertical Rayleigh wave. However, most of the variation in FPE is from order 1 to 5, so using a 5th order model as an approximation should not have too large effect on discrimination.

The dimension of the vector $\underline{Y}$ is still a little too high for an efficient application of the standard multivariate statistical classification procedures. The next stage therefore consists of reducing the vector $\underline{Y}$ to a secondary feature vector $\underline{Z}$. This can be done using for example the technique of principal components. The resulting vector $\underline{Z}$ can then be

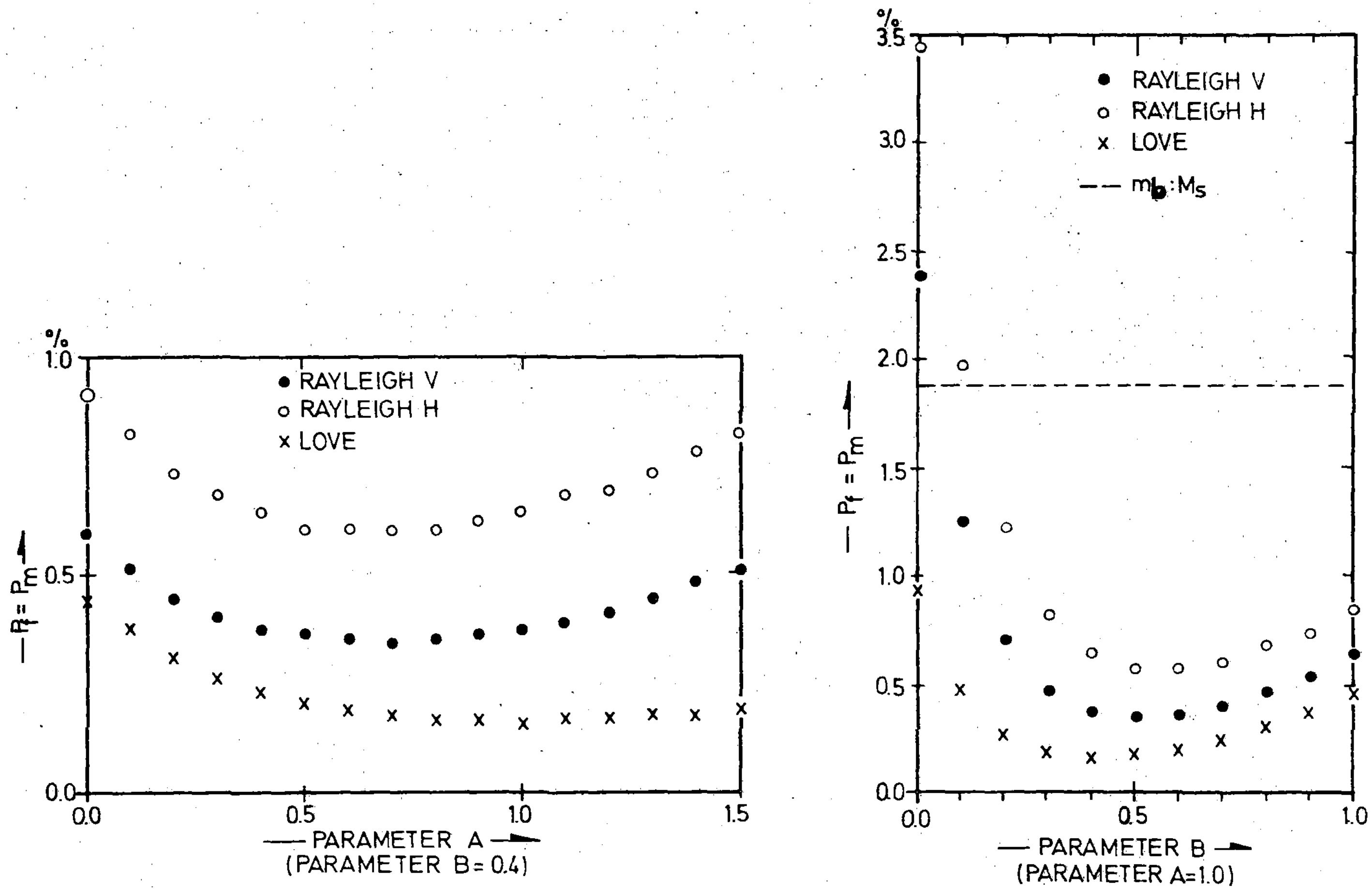


Fig. VII.1.2 (a) The false alarm rate P_f (when this equals the probability P_m of missing an explosion) as a function of the LP scaling factor A of Eq. (VII.1.1) when the SP scaling factor B equals 0.4. (b) $P_f = P_m$ as a function of the SP scaling factor B for a fixed value A=1.0 of the LP scaling factor. The dashed line represents the $m_b:M_s$ discriminant.

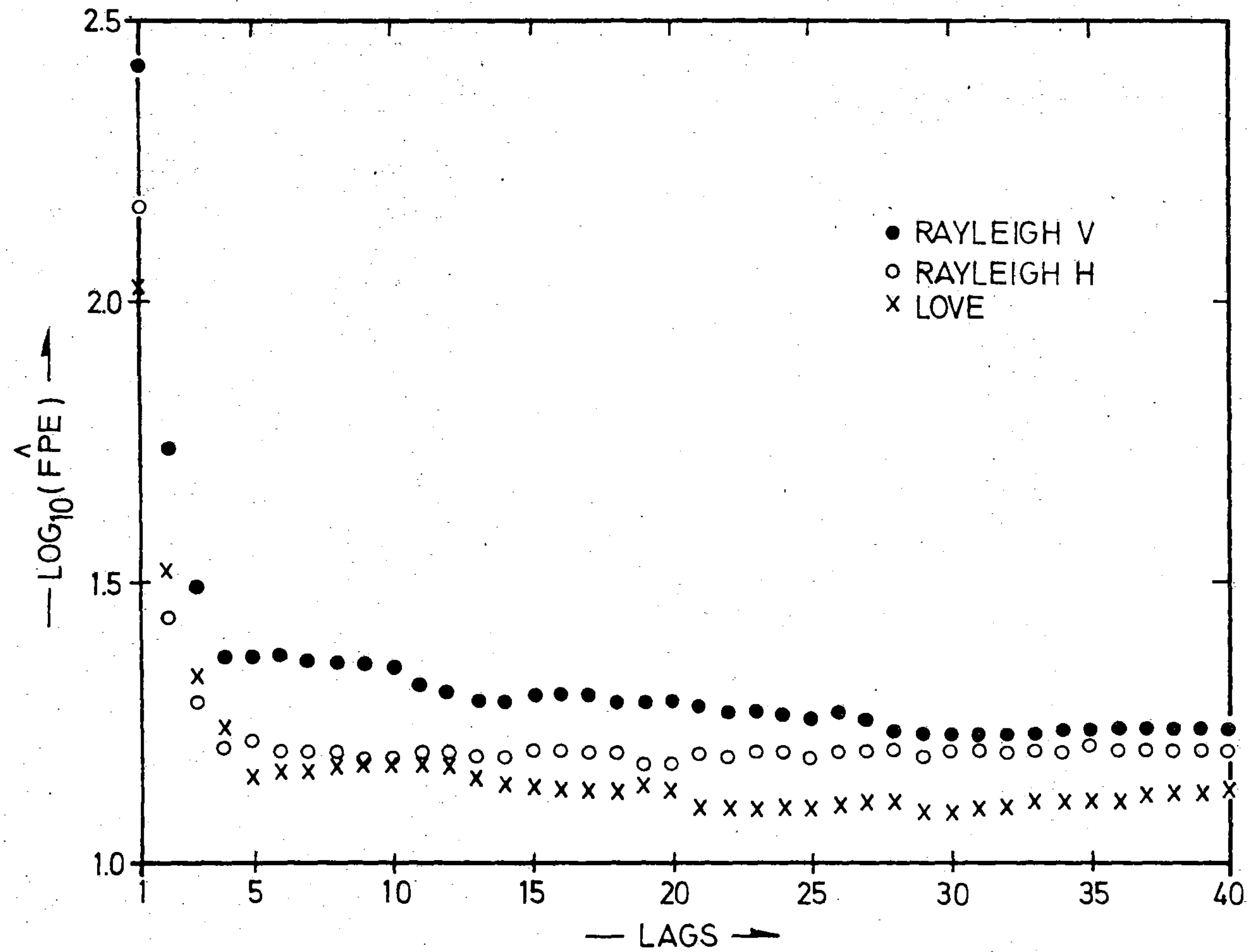


Fig. VII.1.3 Estimated values of Akaike's FPE criterion. The corresponding long period time series data are from an Eastern Kazakh explosion which occurred 06.20.57.7 on Dec 30 1971.

classified by approximating the distribution of earthquake and explosion $\underline{Z}$ vectors by multivariate normal distributions and using the nonlinear version of the so-called Fisher discriminant (see Anderson, 1968).

D. Tjøstheim

References

Akaike, H. (1970): Statistical predictor identification, Ann. Inst. Statist. Math., 22, 203-217.

Anderson, T.W. (1968): An Introduction to Multivariate Statistical Analysis, Wiley, New York.

Tjøstheim, D., and E.S. Husebye (1976): An improved discriminant for test ban verification using short and long period spectral parameters, Geophys. Res. Lett., in press.