

ADVANCES IN MIXED SIGNAL PROCESSING FOR REGIONAL AND TELESEISMIC ARRAYS

Robert H. Shumway

Department of Statistics, University of California, Davis

Sponsored by Air Force Research Laboratory

Contract No. FA8718-04-C-3010

ABSTRACT

This project is aimed at applying recently modified array signal processing techniques to the problem of detecting and estimating mixtures of signals observed on teleseismic and regional arrays. We are developing new techniques for enhancing both signal detection and estimation of azimuth and phase velocity parameters. In particular, we seek an automated detection procedure that sequentially isolates signals in an unknown mixture and provides estimators and confidence intervals for both the propagation parameters and the number of signals.

The methodology is an extension of the standard F-detector which is well known to be based on a nonlinear regression model that tests for the presence of a single signal as a function of slowness parameters. Formulating the multiple signal model as a test of hypothesis for the presence of the most recently added signal, we are able to develop a sequential procedure analogous to stepwise multiple regression by adding signals until no further additions are statistically significant. Both the sequential F-tests and versions of AIC and squared error are monitored to arrive at the final model. A frequency domain bootstrap procedure provides estimators for the standard errors of the estimated velocities and azimuths of the component signals.

Current tests of the methodology have focused on a verifiable mixture of two simultaneously occurring earthquakes observed at the United States Atomic Energy Detection System (USAEDS) long-period seismic array in Korea and a contrived short-period mixture of two regional events. Conventional methods such as single-signal F, MUSIC and high resolution detectors arrive at the incorrect azimuths for this data whereas the multiple-signal F-detector finds the correct number of signals and identifies their propagation characteristics.

OBJECTIVES

This project is aimed at applying recently modified array signal processing techniques to problems involving single and multiple signals observed on teleseismic and regional arrays. We are focused on Topic 3 (Seismic Detection, Location and Discrimination) with particular emphasis on proposing “new techniques to enhance signal detection and parameter estimation (e.g., azimuth, phase velocity) in strongly heterogeneous media”.

Specifically, we are developing the sequential F-statistic as a method for off-line or on-line processing of signals in the presence of possible interfering signals or noises. We are evaluating the statistical performance of the detectors as well as the accuracy of estimated velocities and azimuths of the component signals. Deconvolutions of the component signals will be included. The sequential method as well as current methods will be tested on a test-bed of data obtained from AFTAC to determine the best procedures for on-line detection and off-line analysis.

RESEARCH ACCOMPLISHED

Several algorithms such as the sequential F-detector considered here and the multiple signal characteristic (MUSIC) algorithm are available that offer promise for handling array data with low signal-to-noise ratios and contamination from interfering signals. In this project, we are investigating the performance of currently available algorithms on teleseismic and regional data containing mixed signals in order to demonstrate the superior performance of the sequential F-statistic. A sequential analysis of power using the F-statistic is employed that estimates the correct number of signals and their velocities and azimuths. This is contrasted with results using conventional f-k estimators that do not handle the mixed signal case.

Approaches to detecting signals on arrays all focus on the basic model that expresses the observed channel as sums of delayed signals and a unique noise process. The delays are functionally dependent on velocity and azimuth if the signals are propagating plane waves and this is the assumption that is usually made. Methods that are commonly in use for analyzing such data when a single signal is assumed to be present can be roughly categorized as (1) beam-forming and plotting the power as a function of slowness, which can be converted to estimators of velocity and azimuth, (2) beam-forming converted to an F-statistic by dividing by an estimator of the noise power (see Shumway, 1983, 1999, Shumway et al., 1999, Blandford, 2002,a,b) (3) Capon's estimator (see Capon, 1969), (4) Multiple Signal Characteristic (MUSIC)(Schmidt, 1979, Stoica and Nehorai, 1989) and (5) cross correlation (Tribuleac and Herrin, 1997). Only the MUSIC estimator listed above seems to be at all appropriate for analyzing the mixed signal case.

In the first year of this contract, we have concentrated on developing the multiple signal F-detector as an improvement over conventional detectors such as those given above or the currently favored ratio of short term to long term mean squares (STA/LTA). Technical difficulties in applying the Cramer lower bound approach to getting the variances has necessitated the use of the frequency domain bootstrap (Paparoditis, E. and D.N. Politis, 1999) as an alternative. Software (*mul_sig_sl*, *mul_boot_sl*) now computes the detection results along with standard deviations and confidence intervals for velocity and azimuth from the bootstrap distribution for data containing an arbitrary number of signals. Empirical results indicate that the distribution is approximately normal but this is not necessary for the validity of the bootstrap confidence intervals. Analysis capabilities have been expanded to include the ability to generate simulated data containing multiple signals at varying signal-to-noise ratios using the software *array_sim*.

Two Examples

Conventional methods such as (1)-(5) above, have met with varying degrees of success when they have been applied in practice to cases where there are known interfering signals. We illustrate some of the pitfalls by considering the two events in Figure 1. The left panel of Figure 1 shows three of six channels containing a mixture of two simultaneously occurring earthquakes, one from the south of Africa and the other from the Philippine, observed at the USAEDS long-period seismic array in Korea. The correct back-azimuths for these events are 226 degrees and 198 degrees. Yet, simple time delay estimation for this event gives a back-azimuth of 203 degrees, which is close to

the second signal. A second example illustrates the simulation capability of the current software and shows a contrived mixture of two regional events, also observed at the Korean Seismic Array (KSAR). The right panel of Figure 1 shows the two component signals and the mixture of the two at azimuths 198 and 231 degrees with additive white Gaussian noise with standard deviation .05.

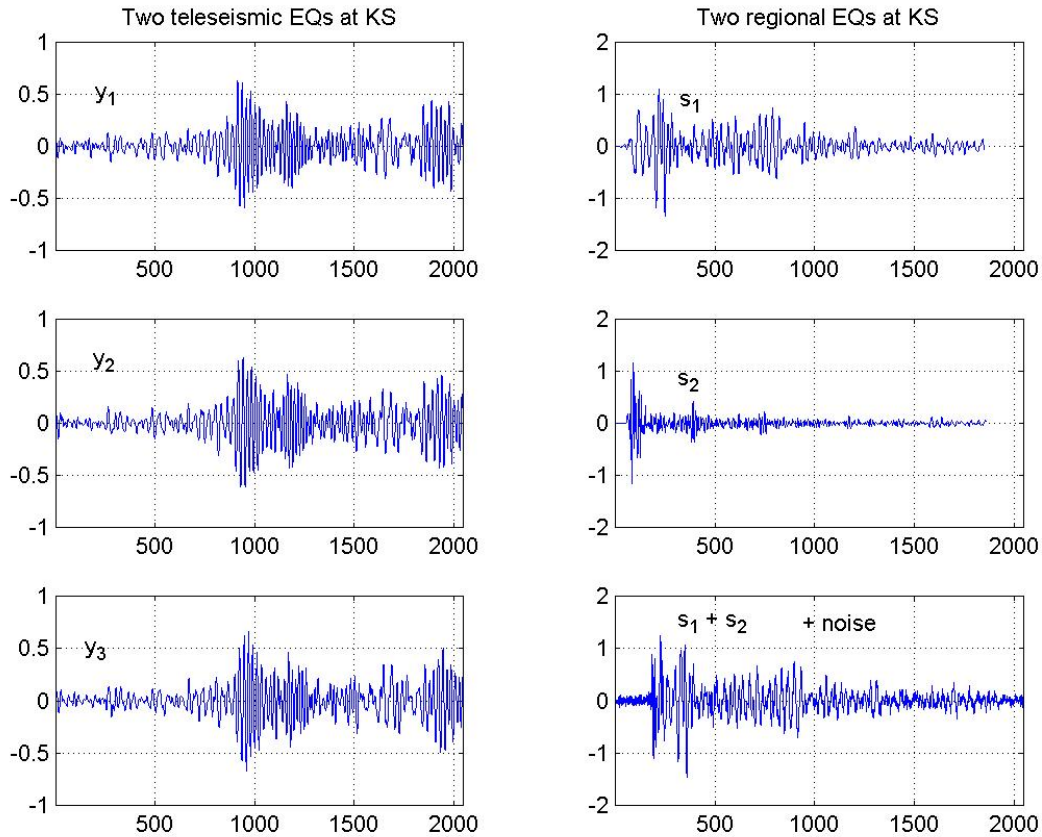


Figure 1. Left column contains three of six channels from a known mix of two EQ's at teleseismic distances. Right column has a single simulated channel of 18 total channels with a known mixture of two closely spaced regional events (s_1 and s_2) with added white noise.

Analysis with conventional detectors

Figure 2 shows estimated velocities and azimuths for the teleseismic mixture using the methods 1)-5). Note that methods (1) and (2), shown as (a) and (c), are based on the slowness coordinates that maximize beam power. The F-detector divides the beam power by an estimator of the noise power computed over the signal window and has that advantage as well of being distributed as an F-statistic that is independent of signal and noise nuisance parameters. The beam power, however, depends on the unknown noise spectrum and is only distributed proportionally to a chi-squared statistic, making it less suitable as a detector. Both statistics lead to an azimuth of 203 degrees which is well off the azimuths (226, 198) of the two known components. In this case the estimated azimuth is midway between the two known azimuths although this is not typical. Equating measured time delays from cross correlations to the theoretical slowness parameters and solving yields similar results, namely, an azimuth of 204 degrees and a velocity of 3.7 km/sec.

The Capon (1969) estimator is based on the inverse of a quadratic form that matches the unknown wave-number vector with the inverse of the estimated spectral matrix. Both the Capon and multiple signal estimator (MUSIC) of

Schmidt (1979, see also Stoica and Nehorai, 1989) expressed in terms of the eigen vectors and eigen values of the spectral matrix and have expectations that get large when the correct slowness vector is matched. They are both

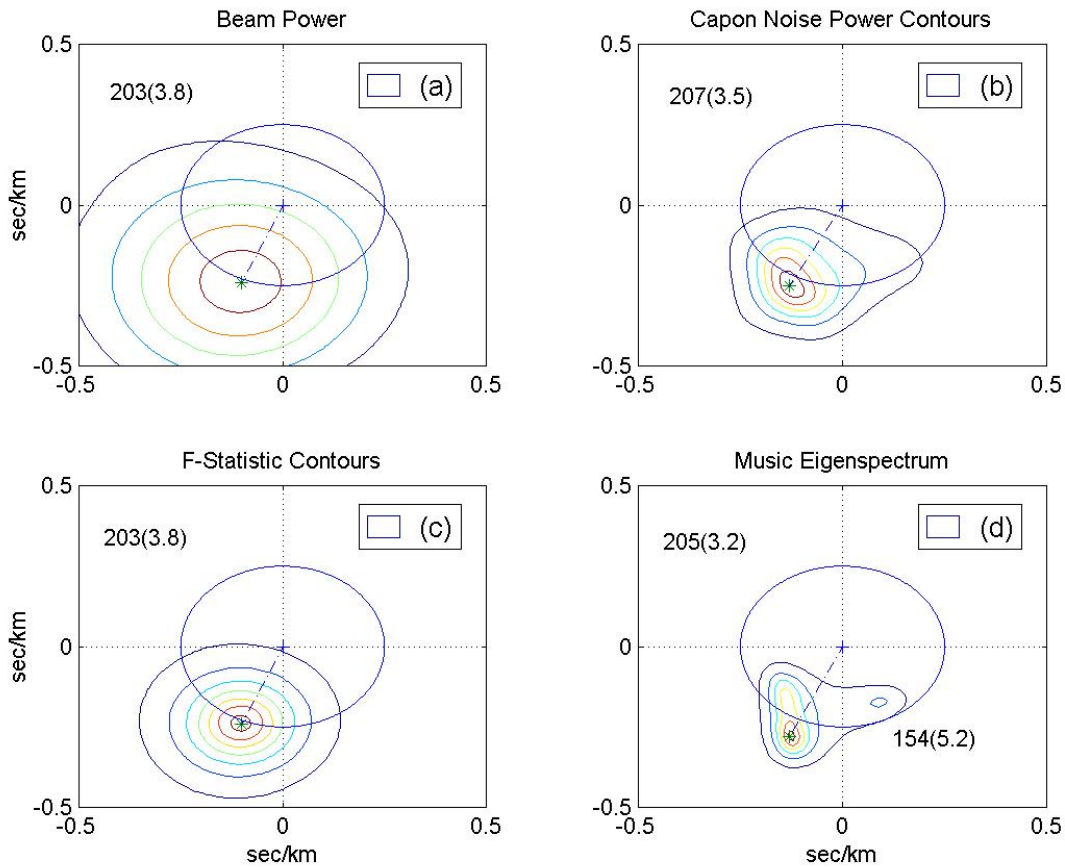


Figure 2. Slowness plots for four conventional detectors applied to the mixture of two long period signals known to be at azimuths 226 and 198 degrees. Estimated azimuth (velocity) pairs are given on each plot.

based on the assumption that the signal is stochastic and perfectly correlated with an additive uncorrelated noise. Again the estimators, shown as (b) and (d), are rather far off at azimuths of 205 and 207 degrees, a gain, essentially midway between the two correct values. Both show mild evidence of a second signal in the 150 degree range. This is of interest even though it is at the wrong azimuth because the sequential F-detector shown later on finds the two correct azimuths plus a third azimuth at 130 degrees, indicating that azimuth as a third statistically significant contributor. The MUSIC estimator here was based on assuming a two-signal source and was adapted to arrays by Shumway (2002) for the infrasound problem.

Analysis of the regional mixture on the right-hand side of Figure 1 showed similar results although the single signal F-statistic focused on the first signal at 198 degrees, the correct azimuth of the second signal. The Capon and MUSIC detectors showed the primary maximum at 201 degrees and 197 degrees respectively, with some distortion indicating a possible additional signal at around the correct azimuth of the second signal at 226 degrees; the estimated velocities of 9.6 and 6.1 obtained with the MUSIC estimator were somewhat off. Hence, the simulated data gave a better result for the MUSIC detector and we conclude that this estimator offers some promise. Measured time delays gave an azimuth of 199 degrees and a velocity of 7.9 km/sec, focusing again on the first signal.

Analysis with the multiple signal F-detector

The analysis of multiple signals involves considering a succession of nonlinear regression models written in the frequency domain with parameters expressed in terms of slowness. Beginning with the single signal model with an estimated set of slowness coordinates, we consider an alternative model with two signals. The likelihood ratio test for a two-signal model against a single-signal model yields a monotone function of an F-statistic. The numerator is the difference between the error power under the two-signal model and the error power under the single signal model and represents the reduction in power possible from the added signal. This reduction is scaled by the noise power under the full model and a function of the number of parameters and the error degrees of freedom. The sequential fitting of more signals continues one at a time until no more added signals are statistically significant. The final estimated velocities and azimuths are those that obtain under the best model.

Standard errors and confidence intervals for velocities and azimuths are computed using the frequency domain bootstrap of Paparoditis and Politis (1999) adapted to the non-linear regression case under the multiple signal model. This involves reconstructing the frequency domain observations from the regression model evaluated at the maximum likelihood estimators for slowness. The residuals from this model will be roughly independent and constitute the basic re-sampling population. To reconstruct a bootstrap sample of the data, draw a sample of these residuals with replacement and use the non-linear regression model to reconstitute a pseudo-sample of the observed data. The estimated and velocity and azimuth computed from this pseudo-sample constitute the first pair of estimated parameters. Repeat the above procedure a large number of times (500) and retain the estimators. The sampling distribution of these estimators yields the standard deviations and the 95% confidence intervals shown in Tables 1-4 below.

Table 1. Analysis of Power for long period mixture. Sequential F-Tests.

Source	Added Power	F-Statistic	P-Value	% Power
First Signal	396	35.8	0	86
Second Signal	43	4.9	.0000	95
Third Signal	14	2.5	.0003	98
Fourth Signal	3	0.5	.1183	99

Table 2. Estimated velocities and azimuths for single-signal and best model for long period mixture. True azimuths are 226 and 198 degrees respectively. Confidence intervals are from the bootstrap distribution.

Model	Azimuth(Velocity)	Standard Errors	95% Confidence Intervals
Single Signal	203(3.85)	1(.04)	(200, 206) (3.8, 3.9)
Signal 1	200(3.5)	2(.05)	(196, 203) (3.4, 3.6))
Signal 2	223(3.9)	2(.05)	(218, 228) (3.8, 4.0)
Signal 3	130(4.0)	2(.05)	(125, 133) (3.8, 4.1)

27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies

Tables 1 and 2 show the results of the sequential F-tests applied to the long period event and give the confidence intervals resulting from the best model. Note that the first signal identified at azimuth 203 degrees accounts for 85% of the total power and still gives an estimated azimuth midway between the two known azimuths of the mixture. The F-detector is highly significant. Adding a second signal to the model increases the percentage of power accounted for substantially and still yields a highly significant F. Testing for the third signal again produces a highly significant F-statistic and increases the power accounted for to 98%. Finally, testing for a fourth potential component accounts for a minimal increase in power and the P-value of about .12 is not significant at any useful level. Table 2 shows the estimated azimuths for the three component signals as 200, 223, and 130 degrees. The first two match up well with the known values but the third has not been identified from alternate records as a real seismic signal. Possibly, this third signal is a coherent noise source, and we have not investigated pure seismic noise sources for this phenomenon. A second concern is the small size (6 elements) of the array in relative to the large number of potential signals.

Tables 3 and 4 below give the comparable results for the regional mixture. Here, the two signals known to be present come in strongly at reasonable azimuths (199, 229) and the confidence intervals include the true values. The first two signals account for 99% of the power and the addition of a potential third signal leads to a non-significant F-statistic with a P-value of .32.

Table 3. Analysis of Power for regional mixture. Sequential F-Tests.

Source	Added Power	F-Statistic	P-Value	% Power
First Signal	482	378	0	96
Second Signal	17	28	.0000	99
Third Signal	.3	.3	.32	99

Table 4. Estimated velocities and azimuths for single-signal and best model for contrived regional mixture. True azimuths are 198 and 231 degrees respectively. Estimators shown are for a single signal model and for a double signal model.

Model	Azimuth(Velocity)	Standard Errors	95% Confidence Intervals
Single Signal	198.3(8.1))	.40(.02)	(197.5, 199.1) (8.04, 8.10)
Signal 1	199.2(8.2)	.2(.01)	(198.8, 199.6) (8.09, 8.12))
Signal 2	229.9(8.0)	.8(.04)	(228.4, 231.3) (7.94, 8.09)

CONCLUSIONS AND RECOMMENDATIONS

We have refined the statistical properties of the proposed sequential F-statistic testing procedure and shown that it will give correct velocities and azimuths in both known and contrived mixtures. The frequency domain bootstrap as applied to the nonlinear regression model will yield satisfactory estimates for variances and confidence intervals. Recommendations for additional research are summarized in 1-5 below.

1. **Deconvolution of component signals and testing on limited data:** Deconvolving the signal mixtures into their component parts should lead to improved estimators for magnitudes of the separate events. The software *mul_decon* is being debugged and tested for this purpose.
2. **Further simulations:** We are continuing to apply the sequential F-detector and the velocity and azimuth estimation procedures to signals embedded in noise at different signal-to-noise ratios.
3. **Analysis of other events:** We are planning on applying the sequential F-statistic to larger test bases of teleseismic and regional data.
4. **Coherent noise sources.** Of interest is testing the single and multiple signal detectors on coherent noise sources and on mixtures of signals in the presence of coherent noise. The assumption at present is that most noise sources propagate at some velocity and are directional in nature. In this case, they will appear as additional signals and produce *false signals* similar to those plaguing the single-signal detectors.
5. **Online monitoring:** The sequential procedure is somewhat sensitive to the locations of the start values for nonlinear optimization of the multiple signal slowness vector. We need to investigate parameters such as these start values as well as lengths of time windows for a time varying online procedure. An automatic version of the current procedure would begin in the slowness quadrant corresponding to the likely location of the test and proceeds by searching adjacent quadrants first.

ACKNOWLEDGEMENTS

Gene Smart and Jon Clauter of AFTAC have graciously provided the data used in this analysis as well as valuable insights into potential methods of analysis. In particular, the previous work of Smart (1972, 1976) on FK analysis forms the basis for the sequential F-detector.

27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies

REFERENCE(S)

- Blandford, R. R. (2002a). Detection and azimuth estimation by infrasonic arrays as a function of array aperture and signal coherence. *AFTAC-TR-02-007*, Air Force Technical Applications Center, Patrick AFB, FL 32925-002.
- Blandford, R. R. (2002b). A plan of development for detection systems for seismic and infrasound arrays. *AFTAC-TR-02-005*, Air Force Technical Applications Center, Patrick AFB, FL 32925-002.
- Capon, J. (1969). High-resolution frequency-wavenumber spectrum analysis. *Proc. IEEE* 57: 1408-1418.
- Paparoditis, E. and D. N. Politis (1999), The local bootstrap for periodogram statistics, *J. Time Series Analysis* 20: 193-222.
- Schmidt, R.O (1979), Multiple emitter location and signal parameter estimation. *Proc. RADC Spectral Estimation Workshop*, 243-258. Rome, Italy.
- Shumway, R.H.(1983), Replicated time series regression: An approach to signal estimation and detection. *Handbook of Statistics Vol. 3*, Chapt. 18, 383-408, Time Series in the Frequency Domain. D.R. Brillinger and P.R. Krishnaiah ed., North Holland.
- Shumway, R.H., S.E. Kim and R.R. Blandford (1999). Nonlinear estimation for time series observed on arrays. Chapter 7, Ghosh ed. *Asymptotics, Nonparametrics and Time Series*, 227-258. New York: Marcel Dekker.
- Shumway, R.H. (2002), Detection and location capabilities of multiple infrasound arrays. Final Scientific Report. *DTRA01-00-C0082*, Preprint available from author.
- Smart, E. (1972), FKCOMB, A Fast General-Purpose Array Processor, *Seismic Array Analysis Center*, Teledyne Geotech, 20 December, 1972.
- Smart, E. (1976), Linear high-resolution frequency-wavenumber analysis. PhD Dissertation, *Southern Methodist University*.
- Stoica, P. and A Nehorai (1989), Music, maximum likelihood, and Cramer-Rao lower Bound. *IEEE Trans. Acoustics, Speech and Signal Processing*, 37, 720-741.
- Tribuleac, I. M. and E. T. Herrin (1997), Calibration studies at TXAR. *Seis. Res. Ltrrs* 68: 353-365.