UNCLASSIFIED

AD NUMBER
AD828649
LIMITATION CHANGES
TO: **Approved for public release; distribution is unlimited.**
FROM: **Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational Use; 20 FEB 1968. Other requests shall be referred to Air Force Technical Appliations Center, Washington, DC 20330.**
AUTHORITY
AFTAC/USAF ltr 25 Jan 1972

THIS PAGE IS UNCLASSIFIED

BEST LINEAR UNBIASED ESTIMATION FOR MULTIVARIATE STATIONARY PROCESSES

20 FEBRUARY 1968

Prepared For

AIR FORCE TECHNICAL APPLICATIONS CENTER
Washington, D. C.

By

William C. Dean
TELEDYNE, INC.

Robert H. Shumway
The George Washington University

Charles S. Duris
TELEDYNE, INC.

Under

Project VELA UNIFORM

Sponsored By

ADVANCED RESEARCH PROJECTS AGENCY
Nuclear Test Detection Office
ARPA Order No. 624

BEST LINEAR UNBIASED ESTIMATION FOR MULTIVARIATE STATIONARY PROCESSES

SEISMIC DATA LABORATORY REPORT NO. 207

AFTAC Project No.:	VELA T/6702
Project Title:	Seismic Data Laboratory
ARPA Order No.:	624
ARPA Program Code No.:	5810
Name of Cotntractor:	TELEDYNE, INC.
Contract No.:	F 33657-67-C-1313
Date of Contract:	2 March 1967
Amount of Contract:	$ 1,736,617
Contract Expiration Date:	1 March 1968
Project Manager:	William C. Dean (703) 836-7644

P. O. Box 334, Alexandria, Virginia

AVAILABILITY

This document is subject to special export controls and each transmittal to foreign governments or foreign national may be made only with prior approval of Chief, AFTAC.

Wash. D.C. 20310

This research was supported by the Advanced Research Projects Agency, Nuclear Test Detection Office, under Project VELA-UNIFORM and accomplished under the technical direction of the Air Force Technical Applications Center under Contract F 33657-67-C-1313.

Neither the Advanced Research Projects Agency nor the Air Force Technical Applications Center will be responsible for information contained herein which may have been supplied by other organizations or contractors, and this document is subject to later revision as may be necessary.

TABLE OF CONTENTS

ABSTRACT

The general linear hypothesis is formulated for a multivariate stationary stochastic process. The best (minimum variance) linear unbiased estimates are derived for the regression functions and it is shown that many signal estimation problems are special cases of the general linear model. Several examples are presented illustrating the technique for particular multivariate processes.

INTRODUCTION

Many problems in the area of applied time series analysis can be formulated and solved by extending and generalizing well known techniques from the classical theory of the multivariate linear hypothesis. The analogies between some of the physical models expressed in terms of signals propagating across arrays and the usual analysis of variance models involving various kinds of effects are striking, particularly when it can be assumed that the multivariate stochastic process in question is weakly stationary. In the stationary case, simplicity in exposition as well as economy in computation result from performing the analysis in the frequency domain using the properties of the Fourier transform.

As an example of a simple estimation problem suppose that a multivariate stochastic process $\{Y_j(t),\ j=1,2,\ldots,n,\ -\infty<t<\infty\}$ consists of a signal which is identical for each j and a weakly stationary multivariate noise process. More specifically we assume that

$$Y_j(t) = s(t) + n_j(t) \qquad \begin{matrix} j=1,2,\ldots,n \\ -\infty<t<\infty \end{matrix} \tag{1}$$

In this model $s(t)$ is regarded as a fixed unknown signal with $n_j(t)$ a weakly stationary zero mean multivariate noise process with a correlation function given by

$$R_{jk}(t-t') = En_j(t)n_k(t') = \int_{-\infty}^{\infty} e^{i\omega(t-t')}\ \sigma_{jk}(\omega)\ \frac{d\omega}{2\pi} \tag{2}$$

where we assume that $\sigma_{jk}(\omega)$ is the cross spectral density matrix of the noise process. The notation is simplified if we use continuous parameter processes and the results apply equally well to the discrete parameter sampled data version. By a best linear unbiased estimate of the signal, say $\hat{s}(t)$, is meant the unbiased linearly filtered version of the data ($E\hat{s}(t) = s(t)$) which has the smallest variance ($E(\hat{s}(t) - s(t))^2 = \text{min.}$). This problem is analogous to estimating the mean of a set of correlated random variables and has been previously considered by Kelly and Levin 5, and Capon, Greenfield, and Kolker, 1. The inherent advantage of this model over the Wiener approach (for example see Papoulis, 8) is that it is not necessary to know in advance the functional form of the signal in order to minimize the mean square error or variance of the signal estimate.

Frequently in the analysis of the multivariate process $Y_j(t)$ it is necessary to consider more general models for the signal such as

$$Y_j(t) = s_1(t) + s_2(t-T_j) + n_j(t) \tag{3}$$

where the output of the j^{th} process is represented as the sum of two signals one of which has been delayed by a known amount T_j. Equation (3) resembles the usual analysis of variance model with $s_1(t)$ corresponding to a row effect and $s_2(t)$ corresponding to a column effect. Delaying $s_2(t)$ by T_j makes the separation possible and corresponds to multiplying by $e^{-i\omega T_j}$ in the frequency domain. Equation (3) suggests the utility of a more general linear model of the form

$$Y_j(t) = \sum_{m=1}^{p} \int_{-\infty}^{\infty} X_{jm}(t-u)\, \beta_m(u)\, du + n_j(t) \tag{4}$$

$$j=1,\ldots,n \qquad -\infty<t<\infty$$

which includes p fixed unknown functions (signals) ($\beta_1(t), \beta_2(t),\ldots, \beta_p(t)$) to be estimated with $X_{jk}(t)$ an nxp matrix of fixed functions which are determined by the model in (4). For example, equation

(4) reduces to equation (1) by the choices $\beta_1(t) = s(t)$, $X_{j1}(t) = \delta(t)$ the Dirac delta function and $p = 1$. Equation (2) may be obtained by taking $X_{j1}(t) = \delta(t)$, $X_{j2}(t) = \delta(t-T_j)$ with $p = 2$, and $\beta_1(t) = s_1(t)$, $\beta_2(t) = s_2(t)$. The Fourier transforms of the observables $X_{jk}(t)$ are often well known for estimation problems and the analysis proceeds easily in the frequency domain. It should be noted that linear models of the form (4) have been developed for the single dimensional case, Grenander and Rosenblatt, 4, and the multivariate case Rosenblatt, 10, for some trigonometric and polynomial regression models or with special assumptions about the autoregressive nature of the residuals $n_j(t)$. Parzen, 9, also gives the minimum variance solution for a single dimensional stochastic process. The approach here takes advantage of formulas relating convolution in time to multiplication in frequency to develop sets of filters which can be applied either in time or frequency. In addition we show how the general linear model can be specialized to include a majority of the signal estimation problems encountered in practical situations. Two examples are presented using the minimum variance unbiased solutions for equations (3) and (4) when it can be assumed that the noises are uncorrelated for $j \neq k$.

GENERAL SOLUTION FOR BLUE (BEST LINEAR UNBIASED) ESTIMATORS

We consider first the derivation of the best linear unbiased (BLUE) estimates for the p regression functions $(\beta_1(t), \beta_2(t), \ldots, \beta_p(t))$ in equation (4) where $(X_{jk}(t), j=1, \ldots, n\ k = 1, \ldots, p)$ is an nxp matrix of fixed known functions. The Fourier transforms of $X_{jk}(t)$ and $\beta_j(t)$ are assumed to exist for all j and k. Consider a class of linear estimates of the form

$$\hat{\beta}_j(t) = \sum_{k=1}^{n} \int_{-\infty}^{\infty} h_{jk}(\tau)\, Y_k(t-\tau)\, d\tau \tag{5}$$

where $h_{jk}(t)$ is a pxn matrix of filter functions to be determined. Note that determining $h_{jk}(t)$ is sufficient for estimating $\beta_j(t)$ as the filters need only be convolved with the multivariate process $Y_j(t)$ and summed to produce $\hat{\beta}_j(t)$.

To introduce the unbiased property we assume that

$$\beta_j(t) = \sum_k\sum_m \int\int_{-\infty}^{\infty} h_{jk}(\tau) X_{km}(t-\tau-u)\, \beta_m(u)\, du\, d\tau \tag{6}$$

which implies by equations (4) and (5) that

$$\hat{\beta}_j(t) = \beta_j(t) + \sum_{k=1}^{n} \int_{-\infty}^{\infty} h_{jk}(\tau)\, n_k(t-\tau)\, d\tau \tag{7}$$

Since $En_k(t) = 0$, equation (7) implies that $E\hat{\beta}_j(t) = \beta_j(t)$ if equation (6) is satisfied for all j and t. Then, by taking Fourier transforms in equation (6) we arrive at the unbiased condition in frequency

$$\beta_j(t) = \sum_m\sum_k \int_{-\infty}^{\infty} H_{jk}(\omega) X_{km}(\omega) B_m(\omega)\, e^{i\omega t} \frac{d\omega}{2\pi} \tag{8}$$

where the Fourier transforms of the elements of (7) appear in the same order in (8). Now equation (8) implies that we must have

$$\sum_{k=1}^{n} H_{jk}(\omega)\, X_{km}(\omega) = \delta_{jm} \tag{9}$$

where δ_{jm} is the ordinary delta function. Equation (9) can be written in matrix notation as

$$HX(\omega) = I \tag{10}$$

with I the pxp identity matrix. The equation is an identity at each frequency ω.

Since (7) is a vector of functions we determine the estimate which minimizes the variance of

$$z(t) = \sum_{j=1}^{p} \int_{-\infty}^{\infty} a_j(u)\, \hat{\beta}_j(t-u)\, du \tag{11}$$

an arbitrary linearly filtered combination of the estimates $\hat{\beta}_j(t)$. Note first that

$$(z(t) - Ez(t)) = \sum_{j,k} \int\int_{-\infty}^{\infty} a_j(u) h_{jk}(\tau) n_k(t-u-\tau)\, du\, d\tau$$

so that by squaring and taking expectations we obtain the expression for the variance. The final result after taking Fourier transforms is expressed in the frequency domain as

$$E(z(t) - Ez(t))^2 = \sum_{j,k} \sum_{j',k'} \int_{-\infty}^{\infty} A_j(\omega) H_{jk}(\omega) \sigma_{kk'}(\omega) H^*_{j'k'}(\omega)\, A_{j'}^*(\omega) \frac{d\omega}{2\pi}$$

$$= \int_{-\infty}^{\infty} \underline{A} H \Sigma H^* \underline{A}^*(\omega) \frac{d\omega}{2\pi} \tag{12}$$

where by $\underline{A}$ we mean a 1xp row vector consisting of the Fourier transforms of $a_k(t)$ and by $\underline{A}^*$ we mean a px1 column vector which is the complex conjugate transpose of $\underline{A}$ with * signifying complex conjugate. Also, from equation (2), $\Sigma = \{\sigma_{jk}(\omega)\}$ is the nxn matrix of spectra and cross spectra at the frequency ω. In the Appendix we show that the inequality

$$\underline{\alpha} B^{-1} \underline{\alpha}^* \geq \underline{\alpha} X (X^* B X)^{-1} X^* \underline{\alpha}^* \tag{13}$$

holds with B an nxn positive definite matrix and X a pxn matrix, $p \leq n$, of rank p and $\underline{\alpha}$ any 1xn complex row vector. Now with $\underline{\alpha} = \underline{A}H$ and $\Sigma = B^{-1}$ we may derive a lower bound for the variance given in equation (12). For

$$\text{var } z(t) \geq \int_{-\infty}^{\infty} \underline{A} H X (X^* \Sigma^{-1} X)^{-1} X^* H^* \underline{A}^*(\omega) \frac{d\omega}{2\pi}$$

$$= \int_{-\infty}^{\infty} \underline{A} (X^* \Sigma^{-1} X)^{-1} \underline{A}^*(\omega) \frac{d\omega}{2\pi} \tag{14}$$

when the unbiased condition (10) is taken into account. Comparing (12) and (14) shows that equality is achieved when

$$H(\omega) = (X^*\Sigma^{-1}X)^{-1}X^*\Sigma^{-1}\ (\omega) \tag{15}$$

Hence, the time version of the matrix of filters is given by calculating equation (15) separately at each frequency and then transforming the result back into time. The resemblance of (15) to the matrix equations for the usual weighted multivariate hypothesis is obvious.

We may compute the variance-covariance matrix of the estimators $\hat{\beta}_i(t)$ and $\hat{\beta}_j(t)$ using (7). Again, squaring, taking expectations and transforming the results into the frequency domain yields

$$\begin{aligned}\left[\operatorname{cov}(\hat{\beta}_i(t),\ \hat{\beta}_j(t)\right] &= \int_{-\infty}^{\infty} H\Sigma H^*\ (\omega)\ \frac{d\omega}{2\pi} \\ &= \int_{-\infty}^{\infty} (X^*\Sigma^{-1}X)^{-1}\ (\omega)\ \frac{d\omega}{2\pi}\end{aligned} \tag{16}$$

which corresponds to the usual formulas in the multivariate linear hypothesis.

The general solution given in (15) requires that the spectral matrix of the noises be known and although this is seldom the case, good estimates for Σ can sometimes be obtained by using a sample of noise when the regression coefficients $\beta_j(t)$ are not present. In the next section we show how the general linear model through an appropriate choice for $X_{jk}(t)$ can be specialized to a number of interesting cases occurring in practical applications.

SOME SPECIAL CASES

Case (i) Estimation of the signal for uncorrelated processes

Consider the model

$$Y_j(t) = s(t-T_j) + n_j(t) \tag{17}$$

which assumes that the signal appears the same of each process except that it is delayed by a known amount T_j. We assume also that

$$R_{jk}(t-t') = \int_{-\infty}^{\infty} \delta_{jk}\sigma^2(\omega) e^{i\omega(t-t')} \frac{d\omega}{2\pi} \qquad (18)$$

which means that the noise is uncorrelated for $j \neq k$ and has the same spectrum (autocorrelation) at each level. In this case $X_{j1}(t) = \delta(t-T_j)$ and $X_{j1}(\omega) = e^{-i\omega T_j}$ with $\beta_1(t) = s(t)$. Hence $X^*(\omega) = (e^{i\omega T_1}, \ldots, e^{i\omega T_n})$ and $\sum^{-1}(\omega) = (1/\sigma^2(\omega))\,I$. Thus, equation (15) yields

$$H(\omega) = (1/n)\;(e^{i\omega T_1}, e^{i\omega T_2}, \ldots, e^{i\omega T_n}) \qquad (19)$$

so that

$$h_k(t) = (1/n)\;\delta(t+T_k) \qquad (20)$$

or from (5)

$$\hat{s}(t) = (1/n) \sum_{k=1}^{n} Y_k(t+T_k) \qquad (21)$$

Hence, the BLUE estimate for this case is just the average of the process accounting for the delay time T_k and the variance of the signal estimate is, from (16)

$$\text{var } \hat{s}(t) = \int_{-\infty}^{\infty} \sigma^2(\omega)/n \frac{d\omega}{2\pi} = R(0)/n \qquad (22)$$

Case (ii) Estimation of the signal for correlated processes

Assume that the basic model is

$$Y_j(t) = s(t) + n_j(t) \qquad (23)$$

where the correlation structure and power spectral matrix of the noise is specified by (2). In this case $X_{j1}(t) = \delta(t)$ so that $X_{j1}(\omega) = 1$ and denoting the inverse of the spectral matrix by $\sum^{-1} = \{\sigma^{jk}(\omega)\}$ we have by (15)

$$H_k(\omega) = \frac{\sum_j \sigma^{jk}(\omega)}{\sum_{j,k} \sigma^{jk}(\omega)} \tag{24}$$

with

$$\operatorname{var} \hat{s}(t) = \int_{-\infty}^{\infty} [1/\sum_{j,k} \sigma^{jk}(\omega)] \frac{d\omega}{2\pi} \tag{25}$$

This corresponds to the case considered in Kelly and Levin, 5, and Capon, Greenfield, and Kolker, 1.

Case (iii) Estimation of multiple signals for uncorrelated processes.

If in the general model given by equation (4) it is assumed that the noise processes are uncorrelated for $j \neq k$ as in case (i) then $\sum(\omega) = \sigma^2(\omega)\, I$ where I is the nxn identity matrix and the equations resemble those arising in the multivariate linear hypothesis. For (15) becomes

$$H(\omega) = (X^*X)^{-1}X^*(\omega) \tag{26}$$

and (16) yields

$$\operatorname{cov}(\hat{\beta}_i(t),\hat{\beta}_j(t)) = \int_{-\infty}^{\infty} \sigma^2(\omega)\ (X^*X)^{-1}\ (\omega)\frac{d\omega}{2\pi} \tag{27}$$

As a particular example the multiple signal model represented by equation (3) yields

$$X^*(\omega) = \begin{pmatrix} 1, 1, \ldots, 1 \\ e^{i\omega T_1}, \ldots, e^{i\omega T_n} \end{pmatrix} \tag{28}$$

and

$$(X^*X)^{-1}(\omega) = \frac{1}{\Delta(\omega)} \begin{pmatrix} n & -z(\omega) \\ -z^*(\omega) & n \end{pmatrix} \tag{29}$$

with

$$z(\omega) = \sum_{k=1}^{n} e^{-i\omega T_k} \qquad \Delta(\omega) = n^2 - |z(\omega)|^2$$

when $\Delta(\omega) \neq 0$. When $\Delta(\omega) = 0$ as, for example, at $\omega = 0$, adjustments are needed in the procedure. The next sections illustrate several examples constructed to test the computational procedures and theory.

EXAMPLES

The theoretical results of the previous sections are easily extended to discrete time parameter processes by replacing integrals with summations and the infinite frequency range by $(-\pi<\omega<\pi)$ but in the application to finite time sampled data certain approximations are necessary. We must decide first whether the analysis proceeds more reasonably and economically in the time domain or the frequency domain. We have chosen the frequency domain for several reasons. First, the restrictions imposed by signal models such as equation (3) must allow for time delays as long as 300 digital points. This means that the time domain analogue of the matrix product $(X^*\Sigma^{-1}X)^{-1}$ becomes extremely large if a 512 point time delay preserving filter is required. Hence, the frequency domain approach becomes much faster as the analysis can be performed separately at each frequency. In addition, for the signal model discussed as Case (ii), it has been noted in Capon, Greenfield, and Kolker, 1, that short filters designed in the time domain are more sensitive to slight stationarity and signal model perturbations. Reference 1 also contains several examples illustrating time and frequency domain computations and concludes that the loss due to using a two sided infinite lag filter rather than the physically realizable filter obtained in the time domain is small.

The approximations used in the examples below are based on the use of the finite Fourier transform and involve replacing the data and filter by aliased versions of the optimum infinite two sided Fourier transforms. References 2 and 7 contain excellent material describing the theory and application of the finite Fourier transform. The effects of aliasing are reduced by choosing an appropriate sampling interval in time based on the highest frequency observable in the data. In the two examples given below the signals are limited to the band from 0 to 10 cycles per second (cps) with the major frequency content less than 5 cps. Hence, the data can be sampled with $\Delta t = .05$ seconds yielding an aliasing frequency of 10 cps. The data samples are long with the first being 150 seconds (3,000 pts) and the second 60 seconds (1200 points) so that $-\infty<t<\infty$ is approximately valid. Finally, the accuracy of the various approximations employed can be evaluated from simulated examples similar to those given below.

<u>Example 1 (Case (iii))</u>

Here we consider the model given by equation (3)

$$Y_j(t) = s_1(t) + s_2(t-T_j) + n_j(t) \qquad j = 1,2,\ldots,20 \tag{30}$$

Figure 1 shows the first five channels of data constructed to conform with Equation (30). Two seismic signals and smoothed white noise uncorrelated from channel to channel were added in the same proportions to produce twenty channels. Five of the twenty channels are displayed in Figure 1. The general signal estimation procedure was to compute the matrix product $(X^*X)^{-1}X^*(\omega)$ at each of 256 frequencies using Equations (28) and (29). The singularity at $\omega=0$ was eliminated by tapering the frequency response functions down to zero at 0 cps. A fast Fourier transform subroutine (see McCowan,7) applied to the frequency response matrix $H(\omega)$ produced the impulse response functions shown in Figure 2. Note that the net results of the filters are to reinforce and sum the aligned signal $s_1(t)$ while canceling the second signal $s_2(t)$ which appears at the given time delays. The reverse holds when estimating $s_2(t)$ with the unaligned signal reinforced and the aligned signal canceled. The signal estimates are formed using a high speed convolution subroutine which calculates equation (5) as a matrix product in the frequency domain and then re-transforms the result to obtain time domain estimates for the regression functions. Note that in the application of the convolution each $Y_j(t)$ is available only for the finite interval T and we assume that $Y_j(t) = 0$ for $t \notin T$. The true signals $s_1(t)$ and $s_2(t)$ are displayed along with the estimates $\hat{s}_1(t)$ and $\hat{s}_2(t)$ in Figure 1. The generation of the filters at 256 frequencies or 512 time points for 20 channels took eight minutes of CDC 1604 time with the subsequent convolution requiring thirty minutes. However, by combining the operations in a different order the total running time has been reduced to eight minutes.

Example 2

As an example of a more complicated model consider the multivariate process shown in Figure 3. Each level or channel conforms artificially to the model.

$$Y_j(t) = \beta_1(t) + \sum_{m=2}^{4} \int_{-\infty}^{\infty} X_{jm}(t-u)\, \beta_m(u)\, du + n_j(t) \qquad (30)$$

$$(j = 1, \ldots, 5)$$

The data was generated by adding noise and the known functions $\beta_j(t)$ in Figure 3 to the data using the transfer functions $X_{jk}(t)$. The real valued frequency responses ($X_{jk}(f)$, $f = \omega/2\pi$ cps) are shown in Figure 4 for the range 0 to 2 cps. The response functions were low pass filtered with a cutoff at 2 cps: hence, there is no significant frequency response above that value. The optimum filters in this case were 200 points or 10 seconds long and were computed by evaluating the matrix product $(X^*X)^{-1}X^*(\omega)$ at 200 frequencies. (The fast Fourier transform was not used here so the number of data points is not a power of two.) The resulting matrix of filter coefficients was convolved with the multivariate process using equation (5) to generate the estimates shown in Figure 3. Comparison of these estimated regression functions with the true regression functions indicates that the procedure again produces reasonable estimates.

REFERENCES

1. Capon, J., Greenfield, R.J., and Kolker, R.J., 1967, Multidimensional Maximum-Likelihood Processing of a Large Aperture Seismic Array, Proceedings of the IEEE,Vol,55,#2.

2. Cooley, J. W., P. A. W. Lewis and P. D. Welch, 1967, The Fast Fourier Transform and Its Applications, IBM Research Paper, RC-1743.

3. Goodman, N. R., 1963, Statistical Analysis Based on A Certain Multivariate Complex Gaussian Distribution, Ann. Math. Statist., 34, 152-177.

4. Grenander, U. and Rosenblatt, M., 1957, Statistical Analysis of Stationary Time Series, John Wiley.

5. Kelly, E.J., Jr., and Levin, M. J., 1964, Signal Parameter Estimation for Seismometer Arrays, MIT Lincoln Lab., Tech. Report 339.

6. Kullback, S., 1959, Information Theory and Statistics, John Wiley.

7. McCowan, D. W., 1966, Finite Fourier Transform Theory and Its Application to the Computation of Correlations, Convolutions and Spectra, Seismic Data Laboratory Report No. 168 (Revised), Teledyne, Inc., Alexandria, Virginia.

8. Papoulis, A. 1965, Probability, Random Variables, and Stochastic Processes, McGraw-Hill.

9. Parzen, E., 1961, An Approach to Time Series Analysis, Ann. Math. Statist., 34, 951-989.

10. Rosenblatt, M., 1956, Some Regression Problems in Time Series Analysis, Proceedings of the Third Berkeley Symp. on Math. Statist., and Probability., Univ. of California Press, Berkeley.

APPENDIX

PROOF OF BASIC INEQUALITY

In order to prove the inequality (13) we use the n variate complex normal distribution defined by Goodman, 3. An n-variate complex normal random variable $\underline{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$ is an n-tuple of complex random variables with a probability density function given by

$$P_1 (\underline{\xi}) = |\pi\Sigma|^{-n} \exp - (\underline{\xi} - \underline{\alpha})B^{-1} (\underline{\xi} - \underline{\alpha})^* \tag{31}$$

with

$$B = E(\underline{\xi} - \underline{\alpha})^* (\underline{\xi} - \underline{\alpha})$$

and B a Hermitian positive definite complex covariance matrix. The mean of the complex vector $\underline{\xi}$ is assumed to be the complex vector $\underline{\alpha}$.

We shall need to calculate the discrimination information measure as developed in Kullback, 6. If by $P_2(\underline{\xi})$ is meant a zero mean complex multivariate normal random variable with covariance matrix B, the discrimination information is defined to be

$$I(I:2;\underline{\xi}) = \int_{\underline{\xi}} P_1 (\underline{\xi}) \log \frac{P_1(\underline{\xi})}{P_2(\underline{\xi})} d\underline{\xi} \tag{32}$$

which for $P_1 (\underline{\xi})$ and $P_2(\underline{\xi})$ above becomes

$$I(I:2;\underline{\xi}) = \underline{\alpha} B^{-1} \underline{\alpha}^* \tag{33}$$

Now if X is an nxp ($p\leq n$) matrix of rank p, define the transformation

$$\underline{\eta} = \underline{\xi}X \tag{34}$$

with the px1 vector $\underline{\eta}$ having a complex multivariate normal distribution with mean $\underline{\alpha}X$ and variance-covariance matrix X*BX. Then X*BX is non-singular and Hermitian positive definite and

$$I(1:2;\underline{\eta}) = \underline{\alpha}X(X^*BX)^{-1} X^*\underline{\alpha}^* \tag{35}$$

Now since (34) is a measurable transformation we have (for the proof of this inequality in the real case, see Kullback, 6, p.57) $I(1:2;\underline{\eta}) \leq I(1:2;\underline{\xi})$ which implies the inequality

$$\underline{\alpha}B^{-1}\underline{\alpha}^* \geq \underline{\alpha}X(X^*BX)^{-1}X^*\underline{\alpha}^* \qquad (36)$$

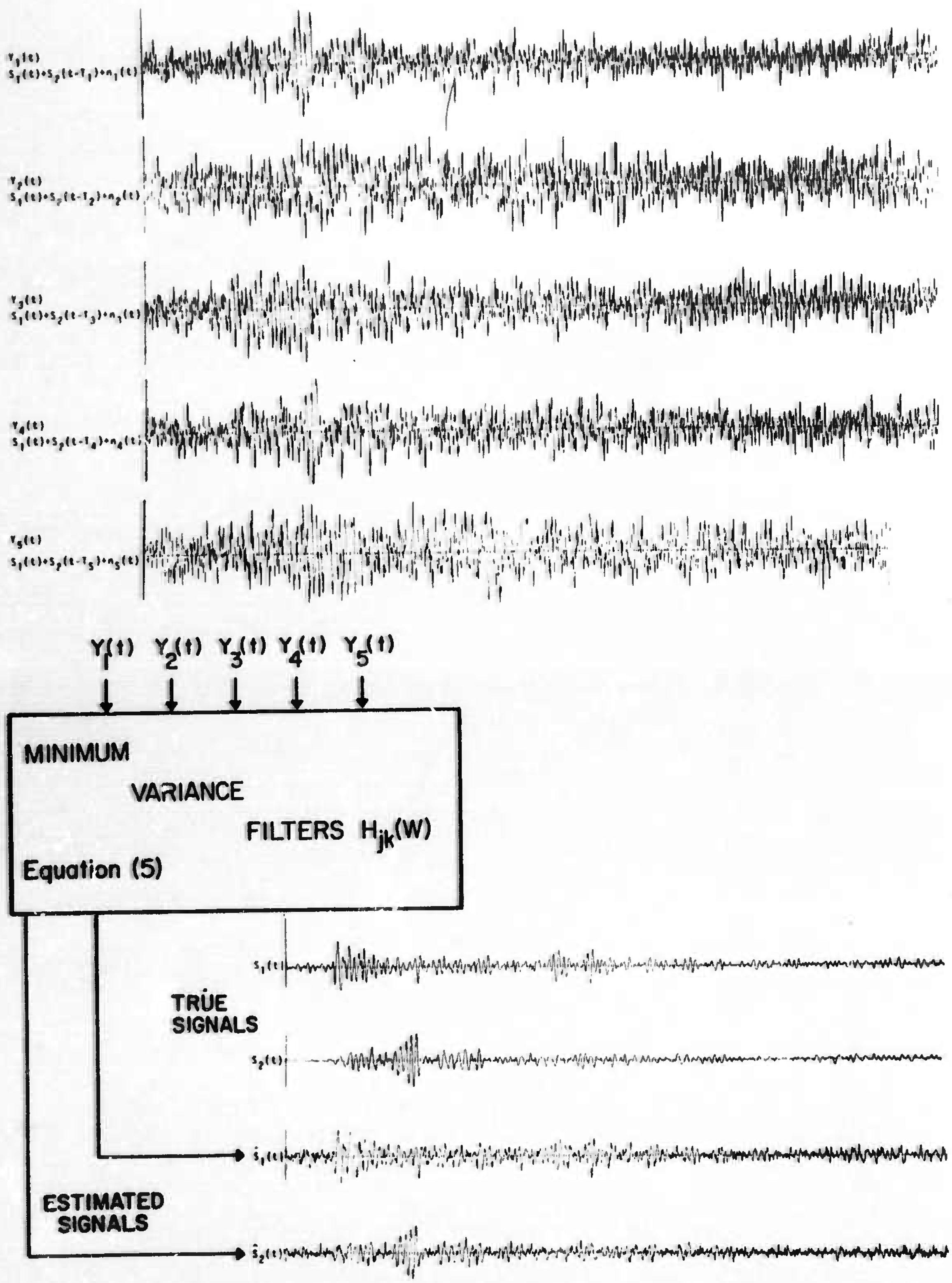

Figure 1. Separation of Two Mixed Signals
By Minimum Variance Filters

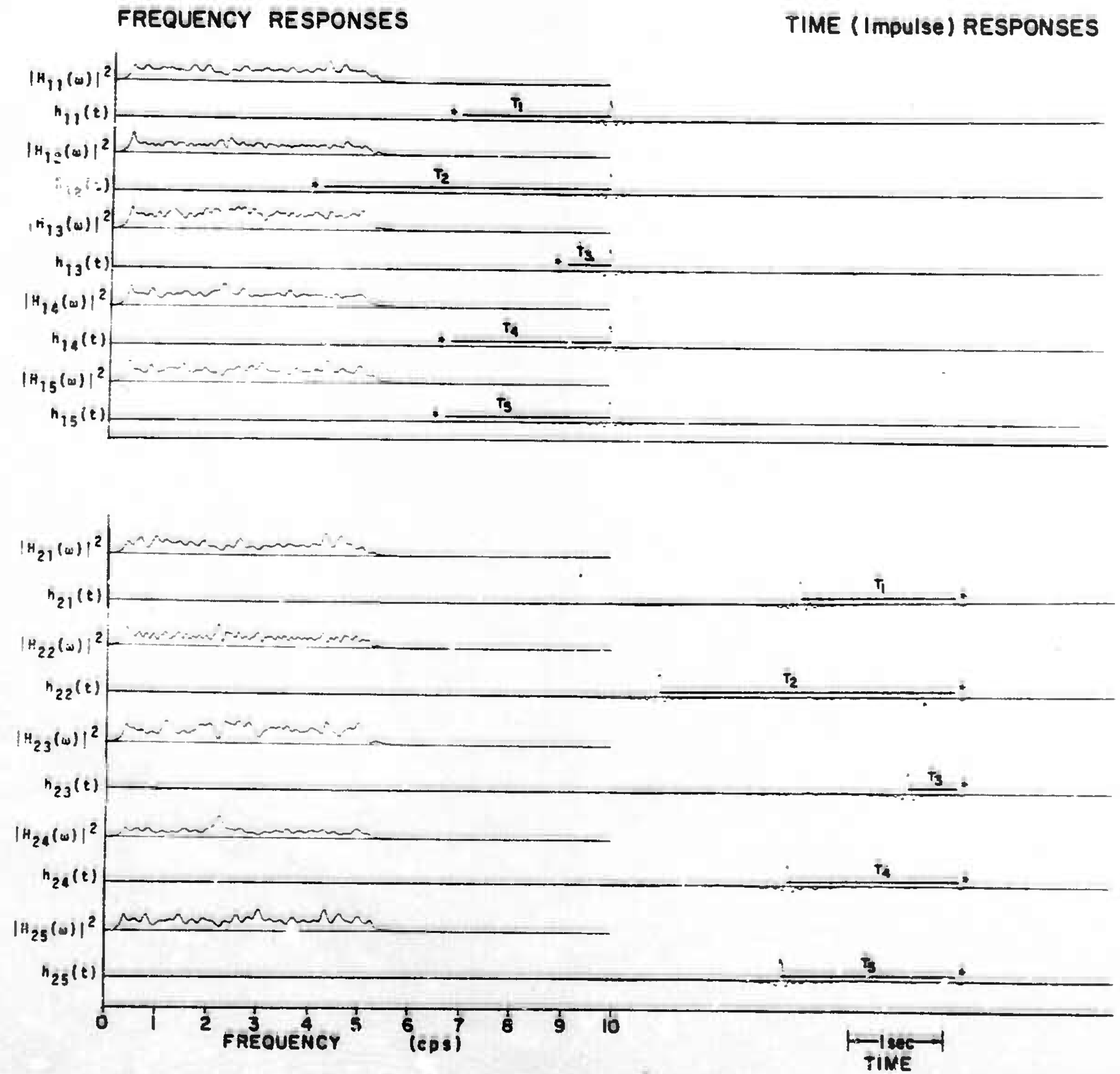

Figure 2. Multiple Channel, Minimum Variance Filters For Estimating Two Mixed Signals

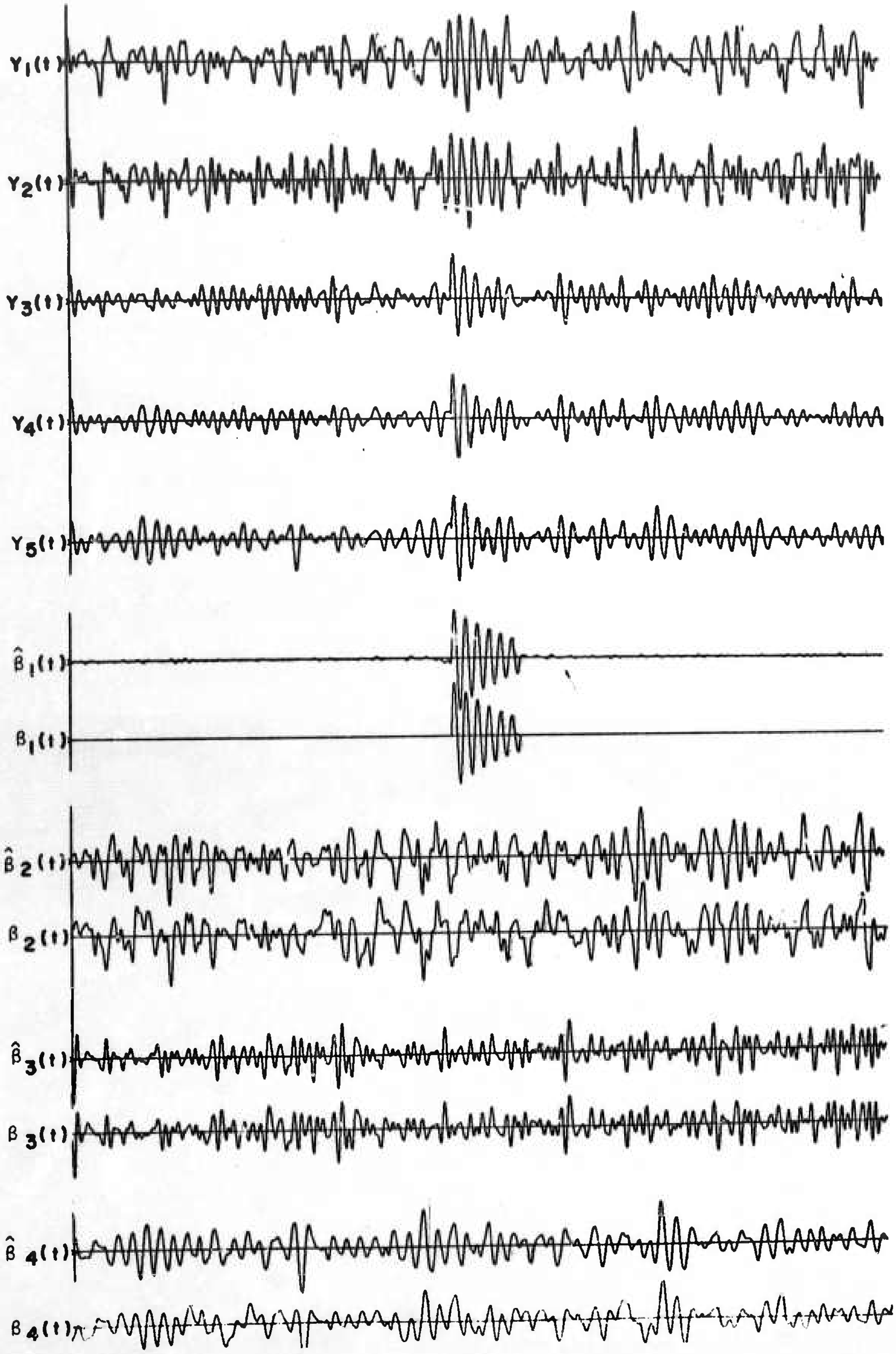

Figure 3 - Basic Data and Estimates Compared With True Values

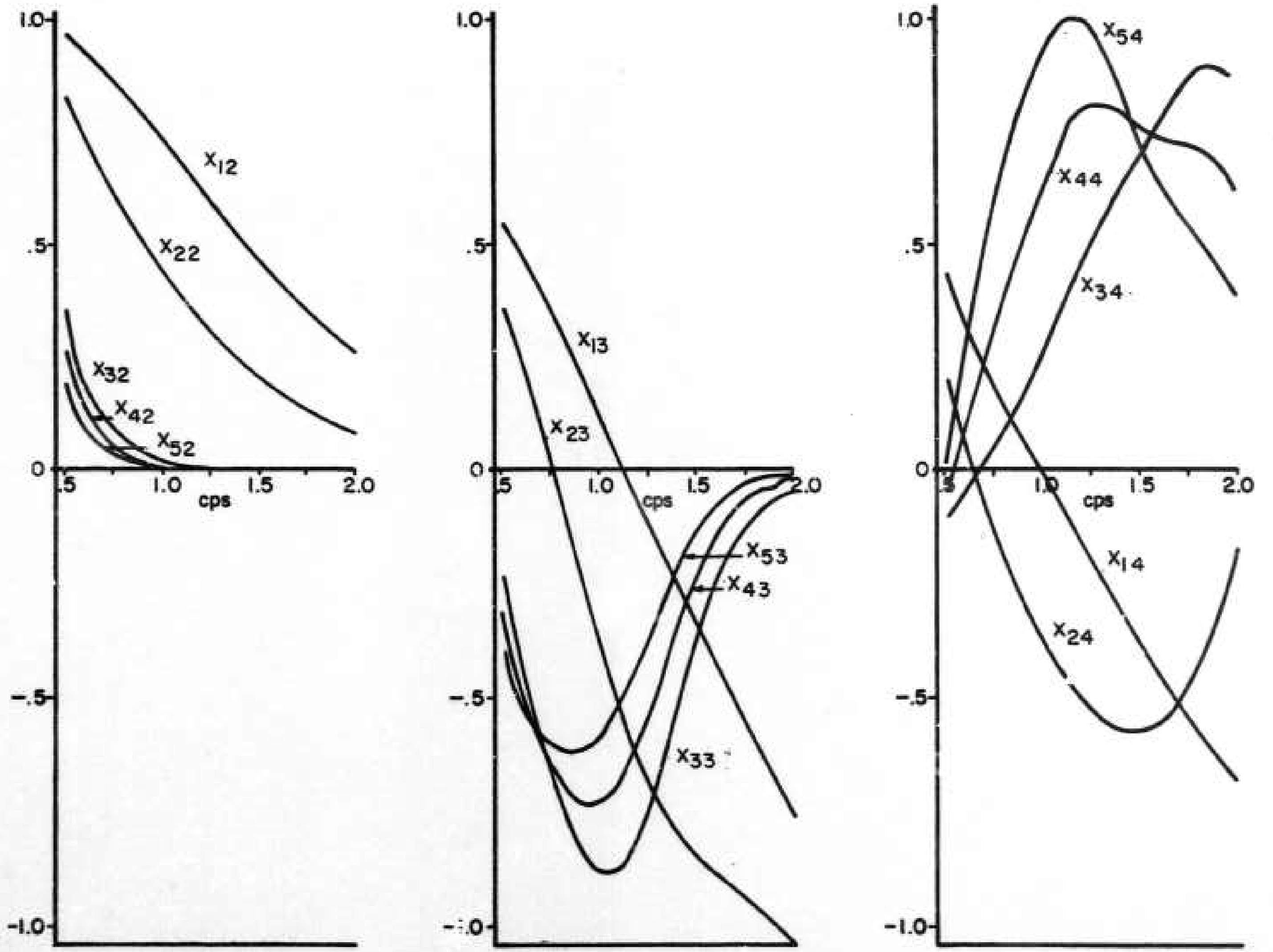

Figure 4 - Matrix of Observations $X_{jk}(f)$, $\omega=2\pi f$, f=cps