Spectral Analysis of a Univariate Process With Bad Data Points, Via Maximum Entropy, And Linear Predictive Techniques

ALBERT H. NUTTALL
*Office of the Director of
Science and Technology*

26 March 1976

NAVAL UNDERWATER SYSTEMS CENTER

New London Laboratory

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

PREFACE

REVIEWED AND APPROVED: 26 March 1976

W. A. Von Winkle
Director, Science & Technology

The author of this report is located at the New London
Laboratory, Naval Underwater Systems Center,
New London, Connecticut 06320.

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
NUSC-TR-5303		

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
Spectral Analysis of a Univariate Process With Bad Data Points, Via Maximum Entropy and Linear Predictive Techniques.	Technical rept.
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Albert H. Nuttall	

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Naval Underwater Systems Center New London Laboratory New London, CT 06320	A-752-05 A-758-02 ZR 000 01 SF 11 121 701

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
Chief of Naval Material (MAT 03521) Washington, DC 20310	26 Mar 1976
	13. NUMBER OF PAGES
	132

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
NUSC-A-752-05, ZR000-01	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spectral Analysis	Linear Prediction	Burg Technique
Bad Data Points	Forward and Backward	Yule-Walker Equations
Maximum Entropy	Averaging	Resolution, Bias,
Autoregression	All-Pole Filter Model	Variability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A comparison of several methods for spectral estimation of a univariate process with equi-spaced samples, including maximum entropy, linear predictive, and auto-regressive techniques, is made. The comparison is conducted via simulation for situations both with and without bad (or missing) data points. The case of bad data points required extensions of existing techniques in the literature and is documented fully here in the form of processing equations and FORTRAN programs. It is con-cluded that the maximum entropy (Burg) technique is as good as any of the methods

next page

DD FORM 1473 1 JAN 73   EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

20. (Cont'd):

considered, for the univariate case. The methods considered are particularly advantageous for short data segments.

This report also reviews several available techniques for spectral analysis under different states of knowledge and presents the interrelationships of the various approaches in a consistent notation. Hopefully, this non-rigorous presentation will clarify this method of spectral analysis for readers who are nonexpert in the field.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

MESA	Maximum Entropy Spectral Analysis
t	Time
$x(t)$	Random process
Δ	Sampling interval in time
x_n	Sample value $x(n\Delta)$
$\equiv$	Defined as
R_k	Correlation of $x(t)$ at delay $k\Delta$
$*$	Conjugate
N	Number of samples available
f	Frequency
$G_x(f)$	Spectrum of process $\{x_n\}$
i	$\sqrt{-1}$
$\hat{x}_k$	Predicted value of x_k
a_n	n-th coefficient of predictive filter
ϵ_k	Instantaneous error at time $k\Delta$
Overbar	Ensemble average
E	Ensemble average magnitude-squared error
$\tilde{a}_n$	n-th optimum predictive filter coefficient
$\tilde{\epsilon}_k$	Minimum-error sequence
E_j	Correlation of minimum-error sequence

LIST OF ABBREVIATIONS AND SYMBOLS (Cont'd)

$G_{\tilde{\epsilon}}(f)$	Spectrum of minimum-error sequence
$A(f)$	Transfer function of whitening filter
z	Complex variable
O	Unit circle in complex z-plane
C_j	Crosscorrelation between $\tilde{r}_k$ and x_k at delay $j\Delta$
$G_{\tilde{\epsilon}x}(f)$	Cross-spectrum between $\tilde{\epsilon}_k$ and x_k
p	Order to which R_k is known; assumed order of predictive filter
$\int_{1/\Delta}$	Integration over $\left(-\frac{1}{2\Delta}, \frac{1}{2\Delta}\right)$
$G_o(f)$	Approximation to $G_x(f)$
μ_k, λ	Lagrange multipliers
$\gamma(f)$	Auxiliary function; (37)
$B(z)$	Auxiliary polynomial; (38), (56)
α_k, b_ℓ	Constants
$\oint$	Counterclockwise integration around unit circle O in complex z-plane
$\delta_{k\ell}$	Kronecker delta; $= 1$ if $k = \ell$; $= 0$ otherwise
R	Correlation matrix; (47)
$\mathbf{s}, \mathbf{a}$	Column matrices; (48)
Superscript T	Transpose
$c_{k\ell}$	Element of inverse of matrix R; (51)

LIST OF ABBREVIATIONS AND SYMBOLS (Cont'd)

θ	Arbitrary real constant
Superscript H	Conjugate transpose
a	Column matrix of $\{a_n\}$
Superscript caret	Estimate of quantity under caret
w_k	White noise
y_k	All-pole filter output process
$H(z)$	Digital filter transfer function
$G_w(f)$	Spectrum of $\{w_k\}$
β_n	Digital filter coefficient
$G_y(f)$	Spectrum of $\{y_k\}$
det	Determinant
F, F_0, $F^{(p)}$, $F_0^{(p)}$	Average magnitude-squared error for a member function
S_{n-m}, S_{nm}	Summations
D_ℓ	Sample crosscorrelation between γ_k and x_k
$\check{x}_k$	Backward-predicted value of x_k
$\check{e}_k$	Backward error
$\hat{R}_k$	Estimate of R_k
$p^{(p)}$	Auxiliary variable; (147)
$f_n^{(p)}$, $b_n^{(p)}$	Forward and backward residuals
g_p	Cross-gain

LIST OF ABBREVIATIONS AND SYMBOLS (Cont'd)

Num(p), Den(p)	Numerator and denominator, respectively, of (155) or (193)
$\mathcal{J}^{(p)}(z)$, $\mathcal{G}^{(p)}(z)$	Transfer functions
$G_f^{(p)}(f)$, $G_b^{(p)}(f)$	Spectra of residuals
B	Number of bad or missing points
M_j	Location of j-th bad point in data sequence
$\sum_{+}$	Summation over good data points; (170A)
I_p	Set of integers to be skipped; (173)
B_p	Number of distinct integers in I_p which are $\geq p+1$ and $\leq N$
$\not\subset$	Not contained in
$\hat{F}$, $\check{F}$	Average forward and backward errors, respectively
AIC	Akaike's Information Criterion

SPECTRAL ANALYSIS OF A UNIVARIATE PROCESS WITH
BAD DATA POINTS, VIA MAXIMUM ENTROPY AND
LINEAR PREDICTIVE TECHNIQUES

1. INTRODUCTION

The analysis of power density spectra of random processes via maximum entropy, linear predictive, and autoregressive techniques has attracted much attention recently, especially for short data segments. In particular, a good review article (reference 1) recently appeared in which 115 references are listed on the topic of linear prediction. Another good paper on this method of spectral analysis (including a comparison of techniques) is available in reference 2, where 66 references are cited. Additional related references, that this author is aware of, are given in references 3 through 15 of this report. The close links that exist between maximum entropy spectral analysis (MESA), autoregressive spectral analysis, predictive error filters, noise-whitening filters, and least-squares model building are pointed out very well in reference 14.

The purposes of this report are to review and interrelate several available techniques for spectral analysis under different states of knowledge, for equi-spaced samples, in a consistent notation; collect and compare the techniques via simulation in order to determine the best available technique(s); and extend the best technique(s) to handle the case of bad (or missing) data points and compare them via simulation. The only detailed comparison of techniques for no missing data points available thus far in the literature is that in reference 2, where the Burg technique and the Yule-Walker approach are compared. Here we will extend the comparison to include the Burg technique, the Yule-Walker approach, an unbiased version of the Yule-Walker approach, the approximate maximum likelihood and least-squares approaches of reference 16, the auto-correlation and covariance approaches of reference 1, and an extended version of the covariance approach. (A comparison with the maximum likelihood technique is reserved for a future report.) Also, we will compare the best of these approaches for the case of bad (or missing) data points and present FORTRAN programs for the recommended techniques.

Throughout this report, we assume we are dealing with equispaced samples of a stationary zero-mean random process x(t); that is, $x_n \equiv x(n\Delta)$, where Δ is the

sampling interval in time. In section 2, we will assume that the correlation function of the sampled process, $\{x_n\}$, namely,*

$$R_k \equiv \overline{x_n \, x^*_{n-k}} = R^*_{-k} \, , \tag{1}$$

is known exactly for all k, and shall present two alternative equations to determine the spectrum of $\{x_n\}$; the latter of the two equations serves as a guide to the MESA, linear predictive, and autoregressive approaches. In section 3, it will be assumed that R_k is known only for a limited range of values of k, and three alternative approaches will be considered and shown to lead to identically the same spectral approximation. Next, in sections 4 and 5, the practical problem of an unknown correlation function and only a finite data set of $\{x_n\}$, $n = 1, 2, \ldots, N$, some of which may be bad, will be addressed, and several candidate techniques for spectral estimation will be presented. Finally, a comparison of the techniques, via simulation, will be conducted and conclusions drawn regarding the best available technique, both with and without bad data points. FORTRAN programs for the best technique for both situations will also be presented.

---

*The case of complex samples is treated, so that we can handle complex envelope or complex demodulated processes. Specialization to real processes is immediate, and (1) becomes $R_k = R_{-k}$. An overbar indicates an ensemble average.

2. CORRELATION KNOWN EXACTLY FOR ALL ARGUMENT VALUES

Suppose the correlation function in (1) of process $\{x_n\}$ is known for all k. The standard (double-sided) definition of the spectrum of $\{x_n\}$ is then (see, for example, reference 14, equation (10))

$$G_x(f) = \Delta \sum_{k=-\infty}^{\infty} R_k \exp(-i2\pi fk\Delta), \quad |f| < \frac{1}{2\Delta}. \tag{2}$$

$G_x(f)$ is real and nonnegative, but need not be even in frequency f for complex $\{R_k\}$.

2.1 LINEAR PREDICTION BASED ON INFINITE PAST

Suppose that sample values x_{k-1}, x_{k-2}, ... are available and are used to linearly predict the value of x_k. Then the one-step predicted value, based on the infinite past, is (for a zero-mean process)

$$\hat{x}_k \equiv \sum_{n=1}^{\infty} a_n x_{k-n}. \tag{3}$$

The values of the complex predictive filter coefficients $\{a_n\}_1^{\infty}$ are chosen such that the one-step prediction error

$$\epsilon_k \equiv \hat{x}_k - x_k = \sum_{n=0}^{\infty} a_n x_{k-n} \quad (a_o = -1) \tag{4}$$

has minimum ensemble average magnitude-squared value. Figure 1 depicts the interrelationships.

The ensemble average magnitude-squared error is, employing (1), given by

$$E \equiv \overline{|\epsilon_k|^2} = \sum_{m,n=0}^{\infty} a_m a_n^* R_{n-m}. \tag{5}$$

For a minimum, we first compute (see reference 17, appendix A)

3
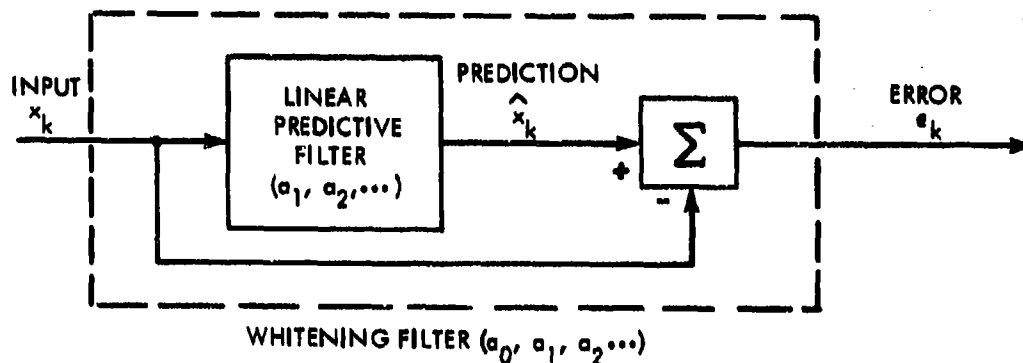
Figure 1.   Block Diagram of Predictive and Whitening Operations

$$\frac{\partial E}{\partial a_\ell^*} = \sum_{m=0}^{\infty} R_{\ell-m}\, \varepsilon_m \,, \quad \ell \geq 1 \tag{6}$$

and set it equal to zero, obtaining the optimum predictive filter coefficients $\{\tilde{a}_m\}_1^{\infty}$ as the solution of the set of equations*

$$\sum_{m=0}^{\infty} R_{\ell-m}\, \tilde{a}_m = 0,\ \ell \geq 1 \qquad (\tilde{a}_0 = a_0 = -1)\,. \tag{7}$$

The minimum-error sequence $\{\tilde{\varepsilon}_k\}$ then possesses correlation

$$E_j \equiv \overline{\tilde{\varepsilon}_k \tilde{\varepsilon}_{k-j}^*} = \sum_{m,n=0}^{\infty} \tilde{a}_m \tilde{a}_n^* \, \overline{x_{k-m} x_{k-j-n}^*}$$

$$= \sum_{m,n=0}^{\infty} \tilde{a}_m \tilde{a}_n^* \, R_{j+n-m} = \sum_{n=0}^{\infty} \tilde{a}_n^* \sum_{m=0}^{\infty} R_{j+n-m}\, \tilde{a}_m \,, \tag{8}$$

---

*The same result, (7), can be obtained by setting the partial derivatives of E, with respect to the real and imaginary parts of a_ℓ, equal to zero.

4

where we have employed (4) and (1). Now the innermost sum on m in (8) is 0 for $j + n \geq 1$, by (7). And if $j \geq 1$ in (8), then $j + n \geq 1$ since $n \geq 0$ in the outermost sum in (8). Therefore, $E_j = 0$ for $j \geq 1$. Also since $E_{-j} = E_j^*$, we have

$$E_j = 0 \text{ for } j \neq 0; \tag{9}$$

that is, the minimum-error sequence $\{\widetilde{\epsilon}_k\}$ is uncorrelated and therefore possesses a white spectrum. The linear filter characterized by coefficients $\{\widetilde{a}_n\}_0^\infty$ is a whitening filter; see figure 1.

The correlation of $\{\widetilde{\epsilon}_k\}$ for zero time delay is the power of the minimum error and is given by

$$E_0 = \overline{\left|\widetilde{\epsilon}_k\right|^2} = \sum_{n=0}^\infty \widetilde{a}_n^* \sum_{m=0}^\infty R_{n-m} \widetilde{a}_m$$

$$= \widetilde{a}_0^* \sum_{m=0}^\infty R_{-m} \widetilde{a}_m = R_0 - \sum_{m=1}^\infty R_m^* \widetilde{a}_m, \tag{10}$$

where we have used (8), (7), and (1). The spectrum of $\{\widetilde{\epsilon}_k\}$ is therefore (using (9))

$$G_{\widetilde{\epsilon}}(f) = \Delta \sum_{j=-\infty}^\infty E_j \exp(-i2\pi f j \Delta) = \Delta E_0, \quad |f| < \frac{1}{2\Delta}, \tag{11}$$

which is white, as mentioned above.

But since the error sequence is given by a linear transformation of process $\{x_k\}$ according to (4) and figure 1, the spectrum of $\{\widetilde{\epsilon}_k\}$ is given by the standard linear filter relation

$$G_{\widetilde{\epsilon}}(f) = \left|A(f)\right|^2 G_x(f), \tag{12}$$

where

$$A(f) \equiv \sum_{n=0}^\infty \widetilde{a}_n \exp(-i2\pi f n \Delta) \tag{13}$$

5

is the transfer function of the whitening filter and is assumed to be stable.* Combining (11)-(13), we obtain an alternative expression to (2) for the spectrum of $\{x_k\}$ as

$$G_x(f) = \frac{G_{\widetilde{\epsilon}}(f)}{|A(f)|^2} = \frac{\Delta E_o}{\left| \sum_{n=0}^{\infty} \widetilde{a}_n \exp(-i2\pi fn\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \qquad (14)$$

Given the correlation values $\{R_k\}$, utilization of (14) requires solution of the set of equations in (7) for the filter coefficients $\{\widetilde{a}_n\}$ and subsequent substitution in (10) and (14). Although this is not a practical alternative to (2) in this case, it does serve to indicate that there is possibly some potential in the idea of determining predictive filter coefficients to minimize the average magnitude-squared one-step prediction error and thereby obtain a white spectrum; this idea will prove to be quite fruitful later on.

As an aside, if we allow $a_{-1} \neq 0$ in (3) and minimize $\overline{|\epsilon_k|^2}$, we find $E_1 \neq 0$, although $E_j = 0$ for $j \geq 2$. Thus, the minimum-error sequence would not be white, and a convenient expression like (14) would not result.

It should also be noted that the crosscorrelations between the minimum-error sequence $\{\widetilde{\epsilon}_k\}$ and all past values of the input, $\{x_k\}$, are zero; this follows by use of (4), (1), and (7).

2.2 LINEAR PREDICTION BASED ON INFINITE FUTURE

If sample values x_{k+1}, x_{k+2}, ... are available and are used to linearly "predict" the value of x_k according to a backward regression (that is, combine future values),

$$\widehat{x}_k = \sum_{n=1}^{\infty} a_n^* x_{k+n}, \qquad (15)$$

---

*That is, $\sum_{n=0}^{\infty} \widetilde{a}_n z^{-n}$ has all its poles inside the unit circle, O, in the complex z-plane.

then the one-step error

$$\epsilon_k \equiv \hat{x}_k - x_k = \sum_{n=0}^{\infty} a_n^* x_{k+n} \qquad (a_o = -1) \tag{16}$$

has average magnitude-squared value

$$\overline{|\epsilon_k|^2} = \sum_{m,n=0}^{\infty} a_m a_n^* R_{n-m} , \tag{17}$$

which is identical to (5). Thus, the same optimum filter coefficients in (7) that minimized (5) would also minimize (17). The minimum-error sequence in (16) would also be white, and an expression for the spectrum of $\{x_k\}$ identical to (14) would result. The point of this result is that an equivalent expression for the spectrum of $\{x_k\}$ is obtained by the backward regression (15), rather than the forward regression (3) of the preceding subsection. This idea will prove useful later when we have to deal with finite data sets and unknown correlation functions.

The crosscorrelations between the minimum-error sequence and all <u>future</u> values of the input are zero; this follows by use of (16), (1), and (7).

2.3 LINEAR INTERPOLATION BASED ON INFINITE PAST AND FUTURE

If we attempt to combine the approaches of the previous two subsections, we are led into considering linear interpolation according to

$$\hat{x}_k \equiv \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} a_n x_{k-n} . \tag{18}$$

The error

$$\epsilon_k \equiv \hat{x}_k - x_k = \sum_{n=-\infty}^{\infty} a_n x_{k-n} \qquad (a_o = -1) \tag{19}$$

has average magnitude-squared value

$$E \equiv \overline{\left|\epsilon_k\right|^2} = \sum_{m,n=-\infty}^{\infty} a_m a_n^* R_{n-m}, \tag{20}$$

using (1). Setting $\partial E/\partial a_\ell^* = 0$ for $\ell \neq 0$, we obtain for the optimum filter coefficients

$$\sum_{m=-\infty}^{\infty} R_{\ell-m} \tilde{a}_m = 0, \; \ell \neq 0 \qquad (\tilde{a}_0 = a_0 = 1). \tag{21}$$

There follows, by use of (1),

$$\tilde{a}_{-j} = \tilde{a}_j^*. \tag{22}$$

The correlation of the minimum-error sequence $\{\tilde{\epsilon}_k\}$ is now

$$E_j \equiv \overline{\tilde{\epsilon}_k \tilde{\epsilon}_{k-j}^*} = \sum_{m,n=-\infty}^{\infty} \tilde{a}_m \tilde{a}_n^* R_{j+n-m}$$

$$= \sum_{n=-\infty}^{\infty} \tilde{a}_n^* \sum_{m=-\infty}^{\infty} \tilde{a}_m R_{j+n-m} = \tilde{a}_{-j}^* \sum_{m=-\infty}^{\infty} \tilde{a}_m R_{-m} = -\tilde{a}_j E_0, \tag{23}$$

where we have employed (19), (1), (21), and (22). It is generally nonzero for $j \neq 0$. The spectrum of the minimum-error sequence is therefore

$$G_{\tilde{\epsilon}}(f) = -\Delta E_0 \sum_{j=-\infty}^{\infty} \tilde{a}_j \exp(-i2\pi f j \Delta) = -\Delta E_0 A(f), \; |f| < \frac{1}{2\Delta}, \tag{24}$$

where we have used (23) and assumed A(f) to be stable. This spectrum is not white; in fact, employing (12), (24) can be expressed as

$$G_{\tilde{\epsilon}}(f) = \frac{\Delta^2 E_0^2}{G_x(f)}, \; |f| < \frac{1}{2\Delta}, \tag{25}$$

which is the _inverse_ of the input spectrum.

If we instead eliminate $G_{\tilde{\epsilon}}(f)$ from (12) and (24), we obtain an expression for the input spectrum in terms of filter A(f) in (24) as

$$G_x(f) = -\frac{\Delta E_o}{A^*(f)} = -\frac{\Delta E_o}{A(f)}, \quad |f| < \frac{1}{2\Delta}; \quad (26)$$

the realness of A(f) follows from (22).

There is an uncorrelated property between the minimum error and the input in the present case also. Namely, the crosscorrelation between the minimum-error sequence and the input is

$$C_j \equiv \overline{\tilde{\epsilon}_k x^*_{k-j}} = \sum_{n=-\infty}^{\infty} \tilde{a}_n \overline{x_{k-n} x^*_{k-j}} = \sum_{n=-\infty}^{\infty} \tilde{a}_n R_{j-n} = E_o \delta_{oj}, \quad (27)$$

using (19), (1), (21), and (23). Thus, the minimum-error sequence is uncorrelated with all past and future values of the input except at the same time instant. The cross-spectrum is

$$G_{\tilde{\epsilon}x}(f) = \Delta \sum_{j=-\infty}^{\infty} C_j \exp(-i2\pi fj\Delta) = \Delta E_o, \quad |f| < \frac{1}{2\Delta}, \quad (28)$$

which is white.

Although (26) and (21) offer an alternative to (14) and (7) in the present case of known correlation function $\{R_k\}$, it suffers in the practical case of unknown correlation and a finite data set, by virtue of the estimate of the real denominator of (26) going through zero (or being complex if (22) is ignored) at some values of f. This is not a significant problem for (14) since both the real and imaginary parts of the estimate of (13) must simultaneously equal zero there, in order to constitute a problem.

Another important practical drawback of this interpolation approach is that ensemble average $\overline{|\epsilon_k|^2}$ would probably be approximated by $\sum_k |\epsilon_k|^2$, where the sum is conducted over those values of k at which a meaningful value of error ϵ_k can be formed for a segment of a single member function of an ensemble. But since the minimum-error sequence $\{\tilde{\epsilon}_k\}$ is not uncorrelated in this case (see (23)), minimization of $\sum_k |\epsilon_k|^2$ for a single member function segment is not synonymous with minimization of $\overline{|\epsilon_k|^2}$; rather, the minimiza-

tion of $\sum_k |\epsilon_k|^2$ will spuriously involve correlation between adjacent terms which are not included in $\overline{|\epsilon_k|^2}$ and which will bias the filter coefficients. Several simulation runs (on real data) confirmed this conclusion by yielding severely biased (and negative) estimates of spectrum $G_x(f)$, even when (22) was taken into account. Accordingly, the interpolation approach was dropped from further consideration.

3. CORRELATION KNOWN EXACTLY FOR A LIMITED RANGE OF ARGUMENT VALUES

In this section, R_k of (1) is assumed to be known exactly for $|k| \leq p$ and unknown for $|k| > p$. Since we are unable to compute the exact spectrum $G_x(f)$, given by (2), in this case, a different approach involving approximation to $G_x(f)$ is required. Three different techniques will be considered and shown to yield identically the same approximation to $G_x(f)$.

3.1 MAXIMUM ENTROPY SPECTRAL ANALYSIS (MESA)

The method in this subsection was originally given in reference 18 and elaborated upon in reference 19. We begin with (2) and note that

$$\int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} df\, G_x(f)\, \exp(i2\pi fk\Delta) = \int_{1/\Delta} df\, G_x(f)\, \exp(i2\pi fk\Delta) = R_k. \tag{29}$$

We wish to approximate $G_x(f)$ by a real nonnegative function $G(f)$ such that its entropy (reference 18, equation (1))

$$\Delta \int_{1/\Delta} df\, \ell n\, G(f) \tag{30}$$

is maximized, subject to the integral constraints

$$\int_{1/\Delta} df\, G(f)\, \exp(i2\pi fk\Delta) = R_k, \quad |k| \leq p. \tag{31}$$

To this aim, we form the quantity

$$Q \equiv \int_{1/\Delta} df\, \ell n\, G(f) - \sum_{k=-p}^{p} \mu_k \int_{1/\Delta} df\, G(f)\, \exp(i2\pi fk\Delta), \tag{32}$$

where Lagrange multipliers $\mu_{-k} = \mu_k^*$, because of the restriction $R_{-k} = R_k^*$, as shown in (1). We perform a variation of (32) according to

$$Q + \delta Q = \int_{1/\Delta} df \, \ln\left[G_o(f) + \epsilon \eta(f)\right] - \sum_{k=-p}^{p} \mu_k \int_{1/\Delta} df \left[G_o(f) + \epsilon \eta(f)\right] \exp(i2\pi fk\Delta), \tag{33}$$

where $G_o(f)$ is the "optimum" approximation to $G_x(f)$ under criterion (30), and obtain, upon setting

$$\frac{\partial(Q + \delta Q)}{\partial \epsilon} = 0 \text{ at } \epsilon = 0, \tag{34}$$

the relation

$$G_o(f) = \frac{1}{\sum_{k=-p}^{p} \mu_k \exp(i2\pi fk\Delta)}, \quad |f| < \frac{1}{2\Delta}. \tag{35}$$

$G_o(f)$ is real since $\mu_{-k} = \mu_k^*$. Since it is also to be nonnegative, we can express

$$G_o(f) = \frac{1}{\left|\gamma(f)\right|^2}, \quad |f| < \frac{1}{2\Delta}, \tag{36}$$

where

$$\gamma(f) \equiv \sum_{k=0}^{p} \alpha_k \exp(i2\pi fk\Delta), \quad |f| < \frac{1}{2\Delta}, \tag{37}$$

and where $\gamma(f)$ has no zeros in the upper-half complex f-plane; that is, polynomial

$$B(z) \equiv \sum_{k=0}^{p} \alpha_k z^k \tag{38}$$

has no zeros inside the unit circle, O, in the complex z-plane. A proof that B(z) in (38) has no zeros inside O is given in reference 11, page 7, for example.* Specifically, it is shown that B(1/z) has all its poles and zeros inside O; that is, B(1/z) is minimum phase.

In order to determine the constants $\{\alpha_k\}_0^p$ in (37), we express (36) as

$$G_0(f)\gamma^*(f) = \frac{1}{\gamma(f)}, \quad |f| < \frac{1}{2\Delta}. \tag{39}$$

(We could equally well have multiplied by $\gamma(f)$ in the following.) Therefore, for all values of ℓ,

$$\int_{1/\Delta} df \, G_0(f)\gamma^*(f) \exp(i2\pi f\ell\Delta) = \int_{1/\Delta} df \, \frac{1}{\gamma(f)} \exp(i2\pi f\ell\Delta). \tag{40}$$

But using (37), this can be expressed as

$$\sum_{k=0}^{p} \alpha_k^* \int_{1/\Delta} df \, G_0(f) \exp(i2\pi f(\ell-k)\Delta) = \int_{1/\Delta} df \, \frac{\exp(i2\pi f\ell\Delta)}{\sum_{k=0}^{p} \alpha_k \exp(i2\pi fk\Delta)}, \quad \text{all } \ell. \tag{41}$$

Now if ℓ is an integer in the range $[0, p]$, the integral on the left side of (41) is equal to $R_{\ell-k}$ (via (29)) for any value of k in its range $[0, p]$; this is where the constraints are employed. Therefore, we have for integer ℓ,

$$\sum_{k=0}^{p} R_{\ell-k} \, \alpha_k^* = b_\ell, \quad 0 \leq \ell \leq p, \tag{42}$$

where

$$b_\ell = \int_{1/\Delta} df \, \frac{\exp(i2\pi f\ell\Delta)}{\sum_{k=0}^{p} \alpha_k \exp(i2\pi fk\Delta)}, \quad 0 \leq \ell \leq p. \tag{43}$$

---

*The proof is couched in terms of the recursive solution of (46) presented in appendix A.

In (43), letting $z = \exp(i2\pi f\Delta)$ and using (38), we have

$$b_\ell = \frac{1}{i2\pi\Delta} \oint \frac{dz}{z} \frac{z^\ell}{B(z)}, \quad 0 \le \ell \le p, \tag{44}$$

where $\oint$ denotes counterclockwise integration around the unit circle O in the complex z-plane. Now $B(z)$ has no zeros inside O by construction. Furthermore, $B(z)$ can have no zeros on O, for then $\gamma(f)$ would be zero for some f, and $G_0(f)$ would possess infinite power, contradicting $R_0 < \infty$. Then (44) yields

$$b_\ell = \frac{1}{\Delta\alpha_0} \delta_{\ell 0}, \quad 0 \le \ell \le p, \tag{45}$$

and (42) becomes

$$\sum_{k=0}^{p} R_{\ell-k} \alpha_k^* = \frac{1}{\Delta\alpha_0} \delta_{\ell 0}, \quad 0 \le \ell \le p. \tag{46}$$

This is p+1 linear equations in p+1 unknowns.*

Now let correlation matrix R be defined as

$$R = \begin{bmatrix} R_0 & R_{-1} & \cdots & R_{-p} \\ R_1 & R_0 & & \\ \cdot & & \cdot & \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ R_p & & & R_0 \end{bmatrix} \tag{47}$$

and define two column matrices

$$\mathbf{s} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}^T, \quad \mathbf{a} = \begin{bmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_p \end{bmatrix}^T. \tag{48}$$

---

*The recursive solution of (46) is presented in appendix A.

R is Hermitian, Toeplitz, and nonnegative definite. Then (46) can be expressed as

$$R \alpha^* = \frac{1}{\Delta \alpha_0} \delta \tag{49}$$

with solution

$$\alpha^* = \frac{1}{\Delta \alpha_0} R^{-1} \delta . \tag{50}$$

Now let the inverse matrix

$$R^{-1} \equiv \begin{bmatrix} c_{00} & c_{01} & \cdots & c_{0p} \\ c_{10} & c_{11} & & \\ . & & & \\ . & & & \\ . & & & \\ c_{p0} & & & c_{pp} \end{bmatrix} . \tag{51}$$

Then (50) and (48) yield

$$\alpha_0^* = \frac{1}{\Delta \alpha_0} c_{00} , \quad \left| \alpha_0 \right|^2 = \frac{1}{\Delta} c_{00} , \quad \alpha_0 = \left(\frac{c_{00}}{\Delta} \right)^{1/2} \exp(i\theta) , \tag{52}$$

where θ is an arbitrary real constant. (c_{00} is always real.) Utilizing this result and (15) in (50), there follows

$$\alpha_k^* = -\frac{c_{k0}}{\sqrt{\Delta c_{00}}} \exp(-i\theta), \quad 0 \le k \le p , \tag{53}$$

and (37) becomes

$$\gamma(f) = \frac{\exp(i\theta)}{\sqrt{\Delta c_{oo}}} \sum_{k=0}^{p} c_{ko}^{*} \exp(i2\pi fk\Delta), \quad |f| < \frac{1}{2\Delta}. \tag{54}$$

Finally, using (36), the "optimum" spectrum (called the maximum entropy spectrum) is

$$G_{o}(f) = \frac{\Delta c_{oo}}{\left| \sum_{k=0}^{p} c_{ko} \exp(-i2\pi fk\Delta) \right|^{2}}, \quad |f| < \frac{1}{2\Delta}. \tag{55}$$

Equation (55) gives the maximum entropy spectrum in terms of the first column of the inverse of the correlation matrix R of available known correlation values; see (47). The forms of (55) and (46) are similar to those encountered earlier in (14) and (7), respectively; see also appendix A. The maximum value of the entropy defined in (30) is evaluated in appendix B and is given by $\ln(\Delta/c_{oo})$.

Substitution of (53) in (38) yields

$$B(z) = \frac{\exp(-i\theta)}{\sqrt{\Delta c_{oo}}} \sum_{k=0}^{p} c_{ko} z^{k}. \tag{56}$$

Thus, investigation of the zeros of $B(z)$ depends on the polynomial $\sum_{k=0}^{p} c_{ko} z^{k}$; it must have no zeros inside the unit circle O. But if we combine (46) and (53), we can write that

$$\sum_{k=0}^{p} R_{\ell-k} \frac{c_{ko}}{c_{oo}} = -R_{\ell}, \quad 1 \le \ell \le p. \tag{57}$$

Now reference 1, page 567, declares that all the zeros of $\sum_{k=0}^{p} c_{ko} z^{-k}$ must lie inside O since R is a correlation matrix. Therefore, polynomial $B(z)$ has no zeros inside O.

3.2 LINEAR PREDICTIVE FILTERING

Here, as in the previous subsection, the available information is knowledge of R_k for $|k| \leq p$. A linear one-step prediction of x_k, based on the past p values, $x_{k-1}, \ldots, x_{k-p}$, is to be accomplished with minimum average magnitude-squared error; see figure 1. Now, however, instead of (3), we have for the predicted value the finite sum*

$$\hat{x}_k \equiv \sum_{n=1}^{p} a_n x_{k-n} \, . \tag{58}$$

The instantaneous error is

$$\epsilon_k \equiv \hat{x}_k - x_k = \sum_{n=0}^{p} a_n x_{k-n} \qquad (a_o = -1) \, . \tag{59}$$

(Equations (58) and (59) constitute stable digital filters regardless of the choice of coefficients.) The ensemble average magnitude-squared error is

$$E \equiv \overline{\left| \epsilon_k \right|^2} = \sum_{m,n=0}^{p} a_m a_n^* R_{n-m} = a^H R a \, , \tag{60}$$

where we have used (1) and (47) and defined

$$a = [a_o \ a_1 \ \ldots \ a_p]^T \, . \tag{61}$$

We now wish to minimize E by choice of filter coefficients $\{a_n\}$. However, we have the constraint on a_o in (59); this can be expressed mathematically as

$$a^H \delta = -1 \, , \tag{62}$$

where δ is defined in (48). In order to minimize (60) subject to (62), we form the quantity

---

*The more general form including $\sum_{n=1}^{p} b_n x_{k-n}^*$ is not considered here.

$$\mathbf{a}^H R \mathbf{a} - \lambda \mathbf{a}^H \mathbf{s} - \lambda^* \mathbf{s}^T \mathbf{a}^*$$

$$= (\mathbf{a} - \lambda R^{-1} \mathbf{s})^H R (\mathbf{a} - \lambda R^{-1} \mathbf{s}) - |\lambda|^2 \mathbf{s}^H R^{-1} \mathbf{s}, \qquad (63)$$

where R^{-1} is defined in (51). Since R is nonnegative definite, being a correlation matrix, (63) is minimized by the choice of coefficients

$$\widetilde{\mathbf{a}} = \lambda R^{-1} \mathbf{s}. \qquad (64)$$

The Lagrange multiplier λ is obtained by substituting (64) in constraint (62), and using (51) and (48):

$$\lambda^* \mathbf{s}^H R^{-1} \delta = -1, \quad \lambda = -\frac{1}{c_{oo}}. \qquad (65)$$

Then (64) yields

$$\widetilde{a}_k = -\frac{c_{ko}}{c_{oo}}, \ 0 \le k \le p. \qquad (66)$$

The minimum value of the error power is found by utilizing (64) and (65) in (60):

$$E_o \equiv \overline{\left| \widetilde{\epsilon}_k \right|^2} = |\lambda|^2 \mathbf{s}^H R^{-1} R R^{-1} \mathbf{s} = |\lambda|^2 c_{oo} = \frac{1}{c_{oo}}, \qquad (67)$$

where $\{\widetilde{\epsilon}_k\}$ is the minimum-error sequence obtained by employing (66) in (59). (A recursion relation for $E_o^{(p)}$ is presented in (A-7); it can be started with $1/c_{oo}^{(0)} = R_o$.) Notice from (67) that c_{oo} must be positive, for non-negative definite R.

The transfer function of the optimum error filter from input x to output $\widetilde{\epsilon}$ in figure 1 is, from (59) and (66),

$$A(f) \equiv \sum_{k=0}^{p} \hat{a}_k \exp(-i2\pi fk\Delta)$$

$$= -\frac{1}{c_{oo}} \sum_{k=0}^{p} c_{ko} \exp(-i2\pi fk\Delta), \quad |f| < \frac{1}{2\Delta} . \tag{68}$$

Furthermore, the spectra in figure 1 are related by

$$G_{\tilde{\epsilon}}(f) = \left| A(f) \right|^2 G_x(f) . \tag{69}$$

Now let us __assume__ that the spectrum of the minimum-error sequence is white over the band $(-\frac{1}{2\Delta}, \frac{1}{2\Delta})$; this is in line with the property (11) which held for the case when the infinite past was available. Then we say

$$\hat{G}_{\tilde{\epsilon}}(f) \equiv \frac{E_o}{1/\Delta} = \frac{\Delta}{c_{oo}}, \quad |f| < \frac{1}{2\Delta} , \tag{70}$$

where we have used (67). Substitution of (68) and (70) in (69) yields the linear predictive spectrum approximation to the input spectrum according to the definition

$$\hat{G}_x(f) \equiv \frac{\hat{G}_{\tilde{\epsilon}}(f)}{\left| A(f) \right|^2} = \frac{\Delta c_{oo}}{\left| \sum_{k=0}^{p} c_{ko} \exp(-i2\pi fk\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \tag{71}$$

This is identical to the approximation (55) obtained by MESA. It is critically dependent on the __assumption__ that the spectrum of the minimum-error $\tilde{\epsilon}$ in figure 1 is white.

Since (71) is identical with the maximum entropy spectrum, (55), it must follow that

$$\int_{1/\Delta} df \, \hat{G}_x(f) \, \exp(i2\pi fk\Delta) = R_k \text{ for } |k| \leq p ; \tag{72}$$

that is, although not specified in the current approach, the correlation function formed from the linear predictive spectrum $\hat{G}_x(f)$ in (71) has the same values at $k\Delta$ for $|k| \leq p$ as the known correlation values $\{R_k\}$.

The implications of the assumption (70) of a white spectrum for the minimum error are investigated in appendix C. It is shown that the crosscorrelation function between input x and output Υ of figure 1,

$$C_\ell \equiv \overline{\Upsilon_k x^*_{k-\ell}},\tag{73}$$

must then satisfy

$$C_\ell = \left\{ \begin{array}{ll} 1/c_{oo}, & \ell = 0 \\ \\ 0, & \ell \geq 1 \end{array} \right\};\tag{74}$$

that is, minimum-error sequence $\{\Upsilon_k\}$ is assumed uncorrelated with all the past values of the input. It is also shown that the unknown correlation values R_k for $k > p$ can be <u>approximated</u> according to

$$R_k = -\sum_{n=1}^{p} \frac{c_{no}}{c_{oo}} R_{k-n} = \sum_{n=1}^{p} \tilde{a}_n R_{k-n}, \ k \geq p+1.\tag{75}$$

This recursion relation, starting with known values $R_1, \ldots, R_p$, can be considered to be an extrapolation of the known correlation values into regions where they are unknown. Equation (75) is shown in appendix D to be a stable recursion when B(z) of (56) has no zeros inside O; this property has been discussed under (38), (56), and (A-9). It can also be shown that Fourier transformation of the extrapolated correlation approximants yields precisely (71). It is interesting to note that (75) has the same form as the predictive equation (58) for individual data values.

Since

$$\int_{1/\Delta} df \ \hat{G}_x(f) \ \exp(i2\pi fk\Delta)\tag{76}$$

is the autocorrelation at delay $k\Delta$, it is given by (72) for $|k| \leq p$, and by (75) for $k \geq p + 1$, where the latter correlations are extrapolated values. This follows

from setting $\hat{G}_{\bar{e}}(f)$ white and choosing $\hat{G}_x(f)$ by (71), according to the analysis in appendixes C and D.

If sample values $x_{k+1}, \ldots, x_{k+p}$ were used to linearly "predict" x_k according to backward regression

$$\hat{x}_k = \sum_{n=1}^p a_n^* x_{k+n} , \qquad (77A)$$

the one-step error $\epsilon_k = \hat{x}_k - x_k$ has average magnitude-squared value

$$E \equiv \overline{\left| \epsilon_k \right|^2} = \sum_{m,n=0}^p a_m a_n^* R_{n-m} \qquad (a_0 = -1) , \qquad (77B)$$

which is identical to (60). Thus, the same optimum filter coefficients in (66) that minimized (60) would also minimize (77B), and an approach similar to that above would yield a spectral approximation identical to (71). The equivalence of the results of this backward regression to that of the forward regression in (58) will prove useful later when we deal with finite data sets and unknown correlation functions.

3.3 ALL-POLE DIGITAL FILTER MODEL

The available information about process $\{x_k\}$ is the same as in the previous two subsections, namely, knowledge of R_k for $|k| \leq p$. Consider a sampled autoregressive process $\{y_k\}$ in steady state generated by a stable all-pole digital filter, $H(z)$, excited by discrete white noise $\{w_k\}$; see figure 2. The noise is characterized by correlation

$$\overline{w_k w_{k-n}^*} = \delta_{no} , \text{ all } n , \qquad (78A)$$

with no loss of generality, and has spectrum

$$G_w(f) = \Delta , \quad |f| < \frac{1}{2\Delta} . \qquad (78B)$$

21
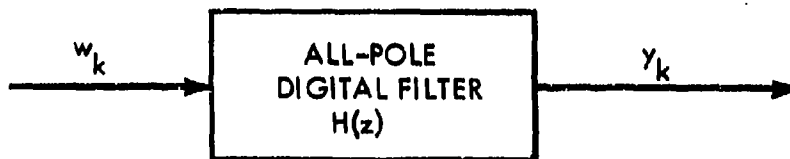
Figure 2. Generation of All-Pole Process

The digital filter is characterized by a p-th order autoregressive relationship,

$$\sum_{n=0}^{p} \beta_n \, y_{k-n} = w_k \, , \tag{79}$$

with transfer function

$$H(z) = \frac{1}{\sum_{n=0}^{p} \beta_n \, z^{-n}} \, . \tag{80}$$

We are going to choose digital filter coefficients $\{\beta_n\}_0^p$ so that autoregressive process $\{y_k\}$ has the same correlation values as process $\{x_k\}$, up through order p; that is, we will set

$$\overline{y_k \, y_{k-n}^*} = R_n \quad \text{for} \quad |n| \leq p \, . \tag{81}$$

Then the spectrum of autoregressive process $\{y_k\}$, given by

$$G_y(f) = G_w(f) \left| H\left(\exp(i2\pi f \Delta)\right) \right|^2$$

$$= \frac{\Delta}{\left| \sum_{n=0}^{p} \beta_n \, \exp(-i2\pi f n \Delta) \right|^2} \, , \quad |f| < \frac{1}{2\Delta} \, , \tag{82}$$

will be used as an approximation to the spectrum of $\{x_k\}$. The spectral relation in (82) holds only if H(z) is stable; that is, all the zeros of the denominator of (80) must lie inside O.

In order to evaluate the filter coefficients $\{\beta_n\}_0^p$, we notice that

$$\overline{w_k \, y^*_{k-n}} = 0 \text{ for } n > 0 \tag{83}$$

since noise samples $\{w_k\}$ are uncorrelated (see (78)) and filter H(z) is realizable (see (79)). The first step we take is to express (79) as

$$y_k = \frac{1}{\beta_0} \left[w_k - \sum_{n=1}^{p} \beta_n \, y_{k-n} \right]. \tag{84}$$

Then using (78) and (83),

$$\overline{w_k \, y^*_k} = \frac{1}{\beta^*_0}. \tag{85}$$

Now multiply both sides of (79) by $y^*_{k-\ell}$ and average; there follows

$$\sum_{n=0}^{p} \beta_n R_{\ell-n} = \frac{1}{\beta^*_0} \delta_{\ell 0}, \; 0 \leq \ell \leq p, \tag{86}$$

upon use of (81), (83), and (85). Now if we let $\beta_n = \sqrt{\Delta} \; \alpha^*_n$, (86) becomes identical to (46). Therefore, we can use solution (53) to obtain for the filter coefficients

$$\beta_n = \frac{c_{00}}{\sqrt{c_{00}}} \exp(-i\theta), \; 0 \leq n \leq p, \tag{87}$$

where θ is an arbitrary real constant.

Substitution of (87) in (82) yields the autoregressive spectrum approximation to the input spectrum as

$$\hat{G}_x(f) \cong G_y(f) = \frac{\Delta c_{oo}}{\left| \sum_{n=0}^{p} c_{no} \exp(-i2\pi f n\Delta) \right|^2} , \quad |f| < \frac{1}{2\Delta} . \qquad (88)$$

This is identical to the maximum entropy spectrum (55) and the linear predictive spectrum (71). The discussion surrounding (76) is relevant here also.

Substitution of (87) in digital filter (80) yields

$$H(z) = \frac{\sqrt{c_{oo}} \ \exp(i\theta)}{\sum_{n=0}^{p} c_{no} z^{-n}} . \qquad (89)$$

This is stable if the denominator contains all its zeros within O; that is, H(z) is stable if and only if B(z) of (56) has no zeros inside O. This property has already been shown true in the discussions under (38), (56), and (A-9).

The relationship in (86) can be extended to $\ell = p + 1$ with the result that

$$\sum_{n=0}^{p} \beta_n R_{p+1-n} = 0, \qquad (90)$$

where R_{p+1} is now interpreted as the value of $\overline{y_k y^*_{k-p-1}}$, and was never specified. If we combine (90) with the last p equations of (86), we obtain

$$\sum_{n=0}^{p} \beta_n R_{\ell-n} = 0 , \ 1 \le \ell \le p+1 . \qquad (91)$$

In order for this set of p + 1 linear equations to possess a nonzero solution for $\left\{ \beta_n \right\}_0^p$ (as it did above), we must have

$$\det \begin{bmatrix} R_1 & R_0 & R_{-1} & \cdots & R_{1-p} \\ R_2 & R_1 & R_0 & \cdots & R_{2-p} \\ \vdots & & & & \\ R_{p+1} & R_p & R_{p-1} & \cdots & R_1 \end{bmatrix} = 0. \tag{92}$$

This can be solved* for R_{p+1}. But since this is identical with reference 19, equation (1), we see that the all-pole digital filter model is identical to choosing R_{p+1} such that

$$\det \begin{bmatrix} R_0 & R_{-1} & \cdots & R_{-p} & R_{-p-1} \\ R_1 & R_0 & & & \\ \vdots & & \ddots & & \\ R_p & R_{p-1} & & & R_{-1} \\ R_{p+1} & R_p & \cdots & R_1 & R_0 \end{bmatrix} \tag{93}$$

is maximized. Additional interpretations of (93) in terms of maximum uncertainty and entropy are presented in references 20 and 14.

---

*Of course, a far more practical method is given by (90) and (87), and more generally by (75).

4. CORRELATION UNKNOWN; FINITE DATA SET

In this section, the correlation values $\{R_k\}$ are unknown, and the only information available about the random process $x(t)$ is a finite set of N samples $x_1, \ldots x_N$, from which we remove the sample mean. From these N samples, we desire an estimate of the spectrum $G_x(f)$. Yet we can not minimize or utilize any ensemble averages as was done in sections 2 and 3, since we have only a finite segment of one member function to work with.

The MESA and autoregressive methods of subsections 3.1 and 3.3 are not easily directly extended to the case of unknown correlation, because they make explicit use of this correlation information; see (31) and (81), respectively. A direct extension of these two methods would require us to decide on the form of the correlation estimates a priori, and could unnecessarily restrict the quality of the spectral estimate we finally obtain. The linear predictive method of sections 2 and 3.2, on the other hand, requires that the ensemble average magnitude-squared error be replaced by some estimating quantity that can be readily calculated from the available data $\{x_n\}_1^N$; as a by-product, we may get estimates of the correlation. Several candidate processing techniques based on subsection 3.2 will therefore be considered, and their processing equations derived. Also, some of the results of subsection 3.1 on MESA will be adapted and combined with the linear predictive approach to form a viable approach to spectral estimation; this technique was originally presented by Burg in reference 21. In section 6, all the techniques will be compared by means of simulation.

4.1 YULE-WALKER EQUATIONS

We begin by defining in this subsection

$$x_k = 0 \text{ for } k < 1, \ k > N, \tag{94}$$

since these samples are unavailable. Taking (58) in subsection 3.2 as a guide, we attempt a linear prediction according to

$$\hat{x}_k = \sum_{n=1}^{p} a_n x_{k-n}, \text{ all } k, \tag{95}$$

where the choice of p is arbitrary for the moment. It should be noticed that although $\hat{x}_k$ is defined for all k, it is not expected to have a good chance of accurately predicting x_k for $k \leq p$ or $k \geq N + 2$ since some zero values of x_k have

been utilized in those regions, according to (94). Nevertheless, we define an instantaneous error

$$\epsilon_k = \hat{x}_k - x_k = \sum_{n=0}^{p} a_n x_{k-n}, \text{ all } k \qquad (a_o = -1);$$

(96)

it is expected to be valid or meaningful, however, only if $k \geq p + 1$ and $k \leq N$ (error ϵ_{N+1} must utilize a zero value for x_{N+1}). Digital filtering operations (95) and (96) are stable for any choice of coefficients $\{a_n\}$.

Since we cannot compute an ensemble average magnitude-squared error now, an average magnitude-squared error is defined for the available data of the single member function as

$$F = \frac{1}{N} \sum_k |\epsilon_k|^2 = \sum_{m,n=0}^{p} a_m a_n^* \frac{1}{N} \sum_k x_{k-m} x_{k-n}^*,$$

(97)

where $\sum_k$ denotes summation over <u>all</u> nonzero values of the summand $|\epsilon_k|^2$, regardless of how meaningful they are. The normalizing factor $1/N$ is somewhat arbitrary; there are N+p nonzero terms in the first sum in (97), but only N-p meaningful terms.

We define, for all n, m

$$S_{n-m} \equiv \frac{1}{N} \sum_k x_{k-m} x_{k-n}^* = S_{m-n}^*,$$

(98)

in which case (97) yields

$$F = \sum_{m,n=0}^{p} a_m a_n^* S_{n-m}.$$

(99)

This relation uses S_ℓ only for $|\ell| \leq p$. In order to minimize F by choice of filter coefficients $\{a_n\}_1^p$, we compute

$$\frac{\partial F}{\partial a_\ell^*} = \sum_{m=0}^{p} S_{\ell-m} a_m, \quad 1 \leq \ell \leq p.$$

(100)

28

The optimum coefficients $\{\tilde{a}_n\}_1^p$ are therefore solutions of the p linear equations

$$\sum_{m=0}^p S_{\ell-m}\,\tilde{a}_m = 0\,, \quad 1 \le \ell \le p \qquad (\tilde{a}_o = a_o = -1) \tag{101}$$

or

$$\sum_{m=1}^p S_{\ell-m}\,\tilde{a}_m = S_\ell\,, \quad 1 \le \ell \le p\,. \tag{102}$$

These are the Yule-Walker equations for the optimum filter coefficients. The method here is called the autocorrelation method in reference 1. (As an aside, in analogy to subsections 2.2 and 3.2, identically the same equations (102) result when x_k is predicted on the basis of p future values, rather than p past values as was done here in (95); see (5) and (17) et seq. and (77) et seq.)

The minimum value of average error F is obtained by substituting (101) in (97) and (99):

$$F_o \equiv \frac{1}{N} \sum_k \left|\tilde{e}_k\right|^2 = \sum_{n=0}^p \tilde{a}_n^* \sum_{m=0}^p S_{n-m}\,\tilde{a}_m = \tilde{a}_o \sum_{m=0}^p S_{-m}\,\tilde{a}_m$$

$$= -\sum_{m=0}^p S_m^*\,\tilde{a}_m = S_o - \sum_{m=1}^p S_m^*\,\tilde{a}_m\,, \tag{103}$$

where we have employed (98) and (101).

The $p \times p$ matrix $[S_{\ell-m}]_1^p$ on the left side of (102) has the form of a legal correlation matrix in that it is Hermitian, Toeplitz, and nonnegative definite. The last property follows from

$$\sum_{\ell,m=1}^p S_{\ell-m}\alpha_m \alpha_\ell^* = \sum_{\ell,m=1}^p \alpha_m \alpha_\ell^* \frac{1}{N} \sum_k x_{k-m} x_{k-\ell}^*$$

$$= \frac{1}{N} \sum_k \left| \sum_{m=1}^p \alpha_m x_{k-m} \right|^2 \ge 0 \tag{104}$$

for any $\{\alpha_m\}_1^p$. Since (104) is greater than zero with probability one, (102) will possess a solution with probability one.

A convenient method of obtaining this solution is to combine (101) and (103) to get

$$- \sum_{m=0}^{p} S_{l-m} \widetilde{a}_m = F_o \delta_{lo}, \ 0 \leq l \leq p. \tag{105}$$

Written out in detail, this is

$$\begin{bmatrix} S_o & S_{-1} & \cdots & S_{-p} \\ S_1 & S_o & & \\ \vdots & & \ddots & \\ S_p & & & S_o \end{bmatrix} \begin{bmatrix} 1 \\ -\widetilde{a}_1 \\ \vdots \\ -\widetilde{a}_p \end{bmatrix} = \begin{bmatrix} F_o \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{106}$$

(The (p+1) $\times$ (p+1) matrix in (106) is nonnegative definite, as a simple extension of (104) shows.) But (106) is identical in form to (A-3), and the recursive solution presented in (A-4) through (A-7) applies directly.

The spectral estimate we adopt follows from results (68) through (71) in subsection 3.2 on linear predictive filtering for known correlation values: first, the optimum transfer function leading from $\{x_k\}$ to minimum-error sequence $\{\widetilde{\tau}_k\}$ in (96) is

$$A(f) = \sum_{n=0}^{p} \widetilde{a}_n \exp(-i2\pi f n \Delta). \tag{107}$$

However, we have a problem in trying to accurately estimate the average minimum-error power that would be used in the numerator of the assumed white spectrum for the error in (70). Although minimum average error F_o of (103) could be used, it is not recommended because not all the error terms in the sum in definition (97) are meaningful. Therefore, because of our inability to

accurately estimate the average minimum-error power in this case, we shall adopt as our spectral estimate

$$\hat{G}_x(f) \equiv \frac{\Delta \cdot 1}{\left| A(f) \right|^2} = \frac{\Delta}{\left| \sum_{n=0}^{p} \tilde{a}_n \exp(-i2\pi fn\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \qquad (108)$$

This is tantamount to assuming the average minimum-error power equal to unity (in addition to assuming the minimum-error spectrum white). This procedure also eliminates level perturbations in the spectral estimate (108) due to random fluctuations in the absolute level of the sample set $\{x_n\}_1^N$; that is, from (102) and (98), it is seen that the optimum values of the filter coefficients, $\{\tilde{a}_n\}_1^p$, would be the same if $\{Kx_n\}_1^N$ were the available samples, for any K. Therefore, estimate $\hat{G}_x(f)$ in (108) is also independent of the absolute level of the available samples. The choice (108) allows for convenient comparisons of the spectral estimates obtained by the various methods presented here.

As an alternative, (108) could be numerically integrated over $(-\frac{1}{2\Delta}, \frac{1}{2\Delta})$, and then (108) could be scaled so that the area under the estimated spectrum is equal to the sample power, $\frac{1}{N} \sum_{n=1}^{N} |x_n|^2$, if desired.

The implications of the assumption in (108) that the minimum-error sequence has a white spectrum are investigated in appendix E. It is shown that the sample crosscorrelation between input sequence $\{x_k\}$ and minimum-error sequence $\{\tilde{\tau}_k\}$, defined for the available data of the single member function as

$$D_\ell = \frac{1}{N} \sum_k \tilde{\tau}_k x_{k-\ell}^*, \quad \text{all } \ell, \qquad (109)$$

is assumed to satisfy

$$D_\ell = 0, \quad 1 \le \ell; \qquad (110)$$

that is, the minimum-error sequence is uncorrelated (on a single member function basis) with all the past input. It is also shown that the quantities S_ℓ de-

fined in (98) (of which only S_ℓ for $|\ell| \leq p$ were used in (99) et seq.) can be estimated for $\ell \geq p + 1$ according to

$$S_\ell = \sum_{n=1}^{p} \widetilde{a}_n S_{\ell - n} , \, \ell \geq p + 1 . \tag{111}$$

This relation, (111), which may not be true for the quantities S_ℓ actually obtained from data $\{x_n\}_1^N$ via (98), is due directly to the assumption that the sample spectrum of the minimum-error is white; see appendix E. The recursion relation (111) is stable, according to appendix D, if

$$1 - \sum_{n=1}^{p} \widetilde{a}_n z^{-n} \tag{112}$$

possesses all its zeros within O. But since matrix $[S_{\ell - m}]$ in (102) has the form of a legal correlation matrix, we appeal directly to reference 1, page 567, to state that this property does indeed hold. Therefore, recursion (111) is stable.

It is worthwhile noting that no direct estimation of unknown correlation values $\{R_k\}$ was attempted in this approach; rather, we minimized the average error defined in (97) and solved directly for the filter coefficients in (102). However, if we rewrite (105) in the form

$$\sum_{m=0}^{p} S_{\ell - m} \widetilde{a}_m = - F_o \, \delta_{\ell o} , \, 0 \leq \ell \leq p , \tag{113}$$

and compare with (C-3), we see that the quantity S_ℓ could be adopted as an estimate of R_ℓ for $|\ell| \leq p$; that is, using (98), we could say

$$\widehat{R}_\ell \equiv S_\ell = \frac{1}{N} \sum_{k} x_k x^*_{k - \ell} , \, |\ell| \leq p , \tag{114}$$

(and then (111), with $\widehat{R}$ replacing S, could be used to estimate R_ℓ for $|\ell| \geq p + 1$, rather than (98)). This is in fact the approach adopted by some authors; see, for example, reference 2, equation (19). However, (114) yields biased estimates because

$$\overline{\hat{R}_\ell} = \overline{\hat{S}_\ell} = \left\{ \begin{array}{c} \dfrac{N-|\ell|}{N}\ R_\ell\ ,\ \ |\ell| \le N \\ \\ 0,\ \text{otherwise} \end{array} \right\}. \tag{115}$$

It is interesting to note that if (114) were adopted a priori as estimates of the unknown correlation values $\{R_k\}$, then the MESA and autoregressive approaches of subsections 3.1 and 3.3 could be utilized directly, if the right sides of (31) and (81) were replaced by $\{\hat{R}_k\}$. The spectral estimates would then be given by results identical to (108), except for a scale factor. The major drawback of this approach is the need to commit oneself to a specific form for the correlation estimates, such as (114), rather than letting the technique itself yield alternative estimates. The specific form used for the correlation estimates could limit the quality of the spectral estimate obtained; this contention is proven true by simulation in section 6.

4.2 UNBIASED VERSION OF YULE-WALKER EQUATIONS

One method of obtaining unbiased estimates of the correlation values $\{R_\ell\}$ is to define estimators

$$\hat{R}_\ell \equiv \frac{1}{N-\ell} \sum_k x_k x_{k-\ell}^* = \frac{1}{N-\ell} \sum_{k=\ell+1}^{N} x_k x_{k-\ell}^* \quad \text{for } 0 \le \ell \le p. \tag{116}$$

Of course $\hat{R}_{-\ell} \equiv \hat{R}_\ell^*$. These could then be used in (102) in the form

$$\sum_{m=1}^{p} \hat{R}_{\ell-m}\, \tilde{a}_m = \hat{R}_\ell\ ,\ 1 \le \ell \le p\ , \tag{117}$$

to solve for the filter coefficients $\{\tilde{a}_m\}_1^p$. And (108) could again be adopted for the spectral estimate. The solution for the coefficients in (117) minimizes no error criterion; it merely utilizes unbiased correlation estimates. The discussion under (115) is relevant to this approach; how good the technique is will be ascertained in section 6.

The matrix $[\hat{R}_{\ell-m}]_1^p$ of estimated correlation values on the left side of (117) is Hermitian and Toeplitz; however, it is not necessarily nonnegative definite. (This last property is shown by considering the example $p = 2$, $N = 3$,

with $x_1 = 2$, $x_2 = x_3 = 3$, for then $\hat{R}_0 = 22/3$, and $\hat{R}_1 = 15/2$.) The recursive solution of appendix A could again be applied to a modified form of (117); see (105) and (106). If the recursive technique in (111) were utilized to extrapolate $\hat{R}$, according to

$$\hat{R}_\ell = \sum_{n=1}^{p} \tilde{a}_n \hat{R}_{\ell-n}, \quad \ell \geq p + 1 \tag{118}$$

and (116), it need not be stable unless $[\hat{R}_{\ell-m}]_1^p$ is nonnegative definite. Even if (118) were unstable, (108) could still be used as a spectral estimate of $G_x(f)$; there would, however, be a greater tendency of some pole-pairs of (108) to drift close to the unit circle, O, in the z-plane and give rise to spurious large peaks in the spectral estimate. This tendency is reduced for stable recursions (118), that is, if (112) possesses all its zeros within O.

4.3 LEAST-SQUARES ESTIMATES OF BOX AND JENKINS

In reference 16, appendix A7.5, a likelihood function approach to estimation of the coefficients in an all-pole (that is, autoregressive) filter model for generation of the process $\{x_n\}$ is considered. The end result (in our notation) is given in (A7.5.7) for real data by

$$S_{ij} \equiv \frac{1}{N} D_{i+1, j+1} = \frac{1}{N} \sum_{k=1}^{N-i-j} x_{i+k} x_{j+k}, \quad 0 \leq i, j \leq p \tag{119}$$

and in (A7.5.15) by

$$\sum_{j=1}^{p} S_{ij} \tilde{a}_j = S_{io}, \quad 1 \leq i \leq p. \tag{120}$$

This constitutes p linear equations in the p unknowns $\{\tilde{a}_j\}_1^p$. The matrix $[S_{ij}]_1^p$ occurring in (120) is symmetric, not necessarily Toeplitz, and not necessarily nonnegative definite. (The last property is shown by considering the example $N = 5$, $p = 2$, with $x_2 = x_3 = x_4 = 1$, for then $S_{11} = 3/5$, $S_{12} = S_{21} = 2/5$, $S_{22} = 1/5$, and the determinant is $-1/25$.) The quantities $\{S_{ij}\}$ also yield biased estimates of $\{R_{i-j}\}$, because

$$\overline{S_{ij}} = \frac{N - i - j}{N} R_{i-j}.$$ (121)

Nevertheless we will adopt (108) for our spectral estimate in this case. The fact that we encounter a non-Toeplitz matrix in (120) disallows the use of the recursive technique for solution in appendix A.

If the solution to (120) is substituted in (112), the zeros need not all lie inside O. Therefore, there would be a greater tendency for some pole-pairs of (108) to drift close to O than when all the zeros must lie inside O, as for subsection 4.1.

4.4 APPROXIMATE MAXIMUM LIKELIHOOD ESTIMATES OF BOX AND JENKINS

This technique is a slight modification of the previous one in subsection 4.3. Namely, in reference 16, under (A7.5.18), the coefficients are solutions of

$$\sum_{j=1}^{p} S_{ij} \, \widehat{a}_j = S_{io}, \quad 1 \leq i \leq p,$$ (122)

where (see (119))

$$S_{ij} = \frac{1}{N-i-j} D_{i+1, j+1} = \frac{1}{N-i-j} \sum_{k=1}^{N-i-j} x_{i+k} x_{j+k}, \quad 0 \leq i, j \leq p.$$ (123)

The matrix $[S_{ij}]_1^p$ occurring in (122) is symmetric, not necessarily Toeplitz, and not necessarily nonnegative definite. (The last property is shown by considering the example $N = 5$, $p = 2$, with $x_2 = 2$, $x_3 = 1$, $x_4 = 2$, for then $S_{11} = 3$, $S_{12} = S_{21} = 2$, $S_{22} = 1$, and the determinant of these coefficients is -1). The quantities $\{S_{ij}\}$ yield unbiased estimates of $\{R_{i-j}\}$; however, every element in a particular diagonal can be different, even though they are estimating the same quantity. Also, the number of terms (in the sum in (123)) along a particular diagonal varies with the position of the element, thereby yielding differing degrees of stability. Equation (108) can be used with (122) to obtain the spectral estimate. Recursive solution of (122) is not allowed because of the non-Toeplitz character of the matrix $[S_{ij}]_1^p$. The comments at the end of subsection 4.3 are relevant here also.

4.5 PREDICTION USING VALID ERROR
POINTS

The method of subsection 4.1 utilized an average error measure defined over all nonzero error terms; see (97). However, as noted under (96), instantaneous error ϵ_k is meaningful only if $k \geq p+1$ and $k \leq N$. Here we define an average magnitude-squared error by summing only over the set of valid error points:

$$F \equiv \frac{1}{N-p} \sum_{k=p+1}^{N} |\epsilon_k|^2 .$$ (124)

There are $N-p$ terms in this sum. This procedure is tantamount to not running off the edges of the available data $\{x_n\}_1^N$. Employing (96), (124) can be written as

$$F = \sum_{m,n=0}^{p} a_m a_n^* S_{nm} ,$$ (125)

where

$$S_{nm} \equiv \frac{1}{N-p} \sum_{k=p+1}^{N} x_{k-m} x_{k-n}^* = S_{mn}^* .$$ (126)

This quantity always contains $N-p$ terms for $0 \leq n, m \leq p$. In order to minimize F, we compute

$$\frac{\partial F}{\partial a_\ell^*} = \sum_{m=0}^{p} S_{\ell m} a_m , \quad 1 \leq \ell \leq p .$$ (127)

The optimum predictive coefficients are therefore solutions of

$$\sum_{m=0}^{p} S_{\ell m} \tilde{a}_m = 0, \quad 1 \leq \ell \leq p \qquad (\tilde{a}_o = a_o = -1) ,$$ (128)

or

$$\sum_{m=1}^{p} S_{\ell m} \widetilde{a}_m = S_{\ell o}, \quad 1 \le \ell \le p. \tag{129}$$

The method here is called the covariance method in reference 1.

The minimum value of the average error F is obtained by substituting (128) in (124) and (125):

$$F_o = \frac{1}{N-p} \sum_{k=p+1}^{N} \left| \widetilde{\tau}_k \right|^2 = \sum_{n=0}^{p} \widetilde{a}_n^* \sum_{m=0}^{p} S_{nm} \widetilde{a}_m = \widetilde{a}_o^* \sum_{m=0}^{p} S_{om} \widetilde{a}_m$$

$$= -\sum_{m=0}^{p} S_{mo}^* \widetilde{a}_m = S_{oo} - \sum_{m=1}^{p} S_{mo}^* \widetilde{a}_m, \tag{130}$$

where we have used (126) and (128).

The $p \times p$ matrix $[S_{\ell m}]_1^p$ on the left side of (129) is a legal correlation matrix in that it is Hermitian and nonnegative definite. The last property follows from

$$\sum_{\ell,m=1}^{p} S_{\ell m} \alpha_m \alpha_\ell^* = \sum_{\ell,m=1}^{p} \alpha_m \alpha_\ell^* \frac{1}{N-p} \sum_{k=p+1}^{N} x_{k-m} x_{k-\ell}^*$$

$$= \frac{1}{N-p} \sum_{k=p+1}^{N} \left| \sum_{m=1}^{p} \alpha_m x_{k-m} \right|^2 \ge 0 \tag{131}$$

for any $\{\alpha_m\}_1^p$. Since $[S_{\ell m}]$ is not necessarily Toeplitz, however, the recursive solution in appendix A is not applicable. Numerical computation of $[S_{\ell m}]$ is eased by taking advantage of a recursive relation between $S_{\ell+1, m+1}$ and $S_{\ell m}$.

The spectral estimate we adopt is given by (108). However, note that we could, if desired, get an estimate here of the average minimum error power E_o, used in (70), according to

$$\hat{E}_0 = F_0. \tag{132}$$

This quantity is meaningful because (130) involves only the valid error terms.

Equation (129) is similar to, but not identical with, the form of (117). The quantities $\{S_{nm}\}$, defined in (126), yield unbiased estimates of $\{R_{n-m}\}$; however, every element in a particular diagonal can be different, even though they are estimating the same quantity.

If the solution to (129) is substituted in (112), the zeros need not lie inside O, despite the nonnegative definite property demonstrated in (131). (The example

$$p = 1, \ N = 2, \ \text{yields} \ \tilde{a}_1 = x_2/x_1 \tag{133}$$

and gives a zero location of (112) which can lie anywhere in the z-plane.) Therefore, the comments at the end of subsection 4.3 are relevant here also.

4.6 FORWARD AND BACKWARD PRE-DICTION USING VALID ERROR POINTS

It was noted in subsections 2.2, 3.2, and 4.1 that "prediction" based on future values of the input $\{x_k\}$ yielded an equivalent spectral estimate to that obtained by prediction based on past values. Here we combine the two techniques. The forward-predicted value of x_k is

$$\hat{x}_k = \sum_{n=1}^{p} a_n \, x_{k-n}, \ p + 1 \le k \le N, \tag{134}$$

where we limit k to the range $[p + 1, N]$, in anticipation of the fact that we can only measure valid errors in that range; see (96) et seq. The backward-predicted value of x_k is

$$\check{x}_k = \sum_{n=1}^{p} a_n^* \, x_{k+n}, \ 1 \le k \le N-p, \tag{135}$$

where we again limit the range of k. (See, for example, (15), (22), and (77).) The forward and backward errors are, respectively,

$$\hat{\epsilon}_k \equiv \hat{x}_k - x_k = \sum_{n=0}^{p} a_n x_{k-n}, \quad p+1 \le k \le N,$$

(136)

$$\check{\gamma}_k = \check{x}_k - x_k = \sum_{n=0}^{p} a_n^* x_{k+n}, \quad 1 \le k \le N-p,$$

where $a_0 = -1$.

An overall average magnitude-squared error is defined as

$$F = \frac{1}{2(N-p)} \left(\sum_{k=p+1}^{N} \left| \hat{\epsilon}_k \right|^2 + \sum_{k=1}^{N-p} \left| \check{\gamma}_k \right|^2 \right) = \sum_{m,n=0}^{p} a_m a_n^* S_{nm}, \quad (137)$$

where, in this subsection,

$$S_{nm} = \frac{1}{2(N-p)} \left(\sum_{k=p+1}^{N} x_{k-m} x_{k-n}^* + \sum_{k=1}^{N-p} x_{k+n} x_{k+m}^* \right).$$

(138)

This quantity always contains $2(N-p)$ terms for $0 \le n, \ m \le p$. Two useful properties of S_{nm} are immediately available:

$$S_{mn} = S_{nm}^*, \quad S_{p-n, \ p-m} = S_{nm}^*.$$

(139)

These properties, plus a recursive relation relating $S_{n+1, \ m+1}$ and S_{nm}, ease the numerical computation of matrix $[S_{nm}]$.

We minimize F by choice of $\{a_m\}_1^p$, getting (see (127) - (129))

$$\sum_{m=1}^{p} S_{nm} \tilde{a}_m = S_{no}, \quad 1 \le n \le p.$$

(140)

The minimum value of F is (see (130))

$$F_o = S_{oo} - \sum_{m=1}^{p} S_{mo}^* \tilde{a}_m . \tag{141}$$

The method here is an extended version of the covariance approach in reference 1.

The matrix $[S_{\ell m}]_1^p$ is Hermitian and nonnegative definite:

$$\sum_{\ell, m=1}^{p} S_{\ell m} \alpha_m \alpha_\ell^* = \frac{1}{2(N-p)} \left(\sum_{k=p+1}^{N} \left| \sum_{m=1}^{p} \alpha_m x_{k-m} \right|^2 + \sum_{k=1}^{N-p} \left| \sum_{m=1}^{p} \alpha_m^* x_{k+m} \right|^2 \right) \geq 0 \tag{142}$$

for any $\{\alpha_m\}_1^p$. However, this matrix is not necessarily Toeplitz; therefore, we cannot apply the recursive solution of appendix A.

The spectral estimate we adopt is obtained by substituting the solution of (140) in (108). An estimate of the average minimum error power E_o, used in (70), is available here according to

$$\hat{E}_o = F_o , \tag{143}$$

if desired, where F_o is given by (141). This is meaningful because (137) utilized only the valid error terms.

In analogy to (126), the quantities $\{S_{nm}\}$ in (138) yield unbiased estimates of $\{R_{n-m}\}$. Nevertheless, if the solution to (140) is substituted in (112), the zeros need not lie within O, despite the nonnegative definite property shown in (142). For $p = 1$, we find

$$\tilde{a}_1 = \frac{x_2 x_1^* + x_3 x_2^* + \dots + x_N x_{N-1}^*}{\frac{1}{2}|x_1|^2 + |x_2|^2 + |x_3|^2 + \dots + |x_{N-2}|^2 + |x_{N-1}|^2 + \frac{1}{2}|x_N|^2} . \tag{144}$$

And since

$$\frac{\left| x_k x_{k-1}^* \right|}{\frac{1}{2}|x_{k-1}|^2 + \frac{1}{2}|x_k|^2} \leq 1 , \tag{145}$$

it follows that

$$\left|\widetilde{a}_1\right| \leq 1.$$ (146)

So for p = 1, the zero of (112) must lie within O, (unless $x_k = A \exp(ikB)$ for all k, in which case it lies on O). However for p = 2, N = 3, and real data, the zeros of (112) lie at $r \pm \sqrt{r^2 - 1}$, where $r = \dfrac{x_1 + x_3}{2x_2}$. So if $|r| \leq 1$, both zeros lies on O, whereas if $|r| > 1$, one zero lies outside O. Therefore, the comments at the end of subsection 4.3 are relevant here also.

4.7 BURG TECHNIQUE

The key to this technique, first presented in reference 21, is the observation from equation (A-6) in appendix A that if the particular p-th order coefficient $a_p^{(p)}$ can be evaluated, the rest of the p-th order predictive filter coefficients, $a_k^{(p)}$, $1 \leq k \leq p - 1$, could be evaluated from (p - 1)-th order coefficients. This relation (A-6) holds true for the solution of normal equations (A-3) even if $\{R_k\}$ are replaced by estimated values. Explicitly, if estimates $\hat{R}_o$, $\hat{R}_1$, ..., $\hat{R}_{p-1}$, and $a_p^{(p)}$ are considered known in the matrix equation

$$\begin{bmatrix} \hat{R}_o & \hat{R}_{-1} & \cdots & \hat{R}_{-p} \\ \hat{R}_1 & \hat{R}_o & & \\ \vdots & & \ddots & \\ \vdots & & & \ddots \\ \hat{R}_p & & & \hat{R}_o \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(p)} \\ \vdots \\ \vdots \\ -a_p^{(p)} \end{bmatrix} = \begin{bmatrix} P^{(p)} \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix},$$ (147)

then we have p + 1 linear equations in the p + 1 unknowns $a_1^{(p)}$, ..., $a_{p-1}^{(p)}$, $\hat{R}_p$, $P^{(p)}$. (Notice that whereas R_p was known and $a_p^{(p)}$ unknown in (A-3), the situation is reversed here for these two variables.) The solutions are given, for $p \geq 1$, by

$$a_k^{(p)} = a_k^{(p-1)} - a_p^{(p)} a_{p-k}^{(p-1)*}, \quad k = 1, 2, \ldots, p - 1 \text{ (no terms if } p = 1) \text{ (148)}$$

$$\hat{R}_p = \sum_{k=1}^{p} \hat{R}_{p-k} \, a_k^{(p)}, \quad \hat{\rho}_p = \sum_{k=1}^{p} \hat{\rho}_{p-k} \, a_k^{(p)}, \tag{149}$$

$$P^{(p)} = P^{(p-1)} \left(1 - \left|a_p^{(p)}\right|^2\right). \tag{150}$$

The quantities $\left\{\hat{\rho}_k\right\}$ in (149) are the estimated normalized correlation coefficients $\left\{\hat{R}_k/\hat{R}_o\right\}$. The recursion (150) is started with

$$P^{(0)} = \hat{R}_o = \frac{1}{N} \sum_{n=1}^{N} \left|x_n\right|^2, \tag{151}$$

which is the sample power of the available samples. A method of evaluating $a_p^{(p)}$ for $p \geq 1$ is treated below.

The method presented here is a combination of references 21 and 7. It begins by defining zero-th order forward and backward sequences according to

$$f_n^{(o)} = x_n, \quad b_n^{(o)} = x_n, \quad 1 \leq n \leq N. \tag{152}$$

The p-th order forward and backward sequences (residuals) for $p \geq 1$ are defined according to

$$\left. \begin{array}{l} f_n^{(p)} \equiv f_n^{(p-1)} - g_p \, b_{n-1}^{(p-1)} \\[3mm] b_n^{(p)} \equiv b_{n-1}^{(p-1)} - g_p^* \, f_n^{(p-1)} \end{array} \right\} \quad \text{for } p + 1 \leq n \leq N. \tag{153}$$

(These can be interpreted as one-step forward and backward prediction errors.) A chain interpretation of (153) is presented in figure 3. (From the known correlation results in subsection 3.2, if we define

$$f_n^{(p)} = x_n - \sum_{k=1}^{p} a_k^{(p)} \, x_{n-k}, \quad b_n^{(p)} = x_{n-p} - \sum_{k=1}^{p} a_k^{(p)^*} x_{n-p+k},$$

we find that figure 3 results, with g_p replaced by $a_p^{(p)}$.)
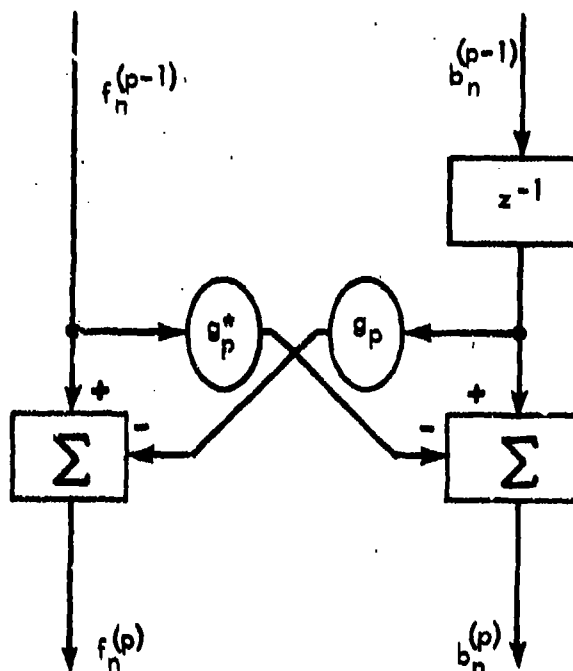
Figure 3. Chain Interpretation of Burg Technique

The average magnitude-squared value of the p-th order forward and backward sequences is

$$F^{(p)} = \frac{1}{2(N-p)} \sum_{n=p+1}^{N} \left(\left| f_n^{(p)} \right|^2 + \left| b_n^{(p)} \right|^2 \right)$$

$$= \frac{1}{2(N-p)} \sum_{n=p+1}^{N} \left(\left| f_n^{(p-1)} - g_p \, b_{n-1}^{(p-1)} \right|^2 + \left| b_{n-1}^{(p-1)} - g_p^* \, f_n^{(p-1)} \right|^2 \right), \; p \geq 1. \tag{154}$$

We wish to minimize this average power at the p-th stage by choice of cross-gain g_p. We find the optimum choice to be

$$g_p = \frac{2 \sum\limits_{n=p+1}^{N} f_n^{(p-1)} \, b_{n-1}^{(p-1)*}}{\sum\limits_{n=p+1}^{N} \left(\left| f_n^{(p-1)} \right|^2 + \left| b_{n-1}^{(p-1)} \right|^2 \right)} = \frac{\text{Num}(p)}{\text{Den}(p)}, \; p \geq 1. \tag{155}$$

43

When (155) is substituted in (153), the results are called the residuals. The minimum value of the residual power at the p-th stage is obtained by substituting (155) in (154) and is expressible as

$$F_o^{(p)} = \left(1 - \left|g_p\right|^2\right) \frac{\text{Den}(p)}{2(N-p)}, \ p \geq 1.$$

(156)

The quantities necessary for this evaluation are available when (155) is evaluated. The value of (156) will never be smaller than (141), since the procedure here is a step-by-step procedure, not a simultaneous procedure as used in subsection 4.6.

An immediate recursion for the transfer functions of the p-th stage in figure 3 is

$$\left.\begin{array}{l} \mathfrak{Z}^{(p)}{}_{(z)} = \mathfrak{Z}^{(p-1)}{}_{(z)} - g_p z^{-1} \mathcal{B}^{(p-1)}{}_{(z)} \\[2mm] \mathcal{B}^{(p)}{}_{(z)} = z^{-1} \mathcal{B}^{(p-1)}{}_{(z)} - g_p^* \mathfrak{Z}^{(p-1)}{}_{(z)} \end{array}\right\}, \ p \geq 1,$$

(157)

with starting values

$$\mathfrak{Z}^{(o)}{}_{(z)} = \mathcal{B}^{(o)}{}_{(z)} = 1.$$

(158)

If we let transfer functions

$$\mathfrak{Z}^{(p)}{}_{(z)} = 1 - \sum_{k=1}^{p} a_k^{(p)} z^{-k},$$

$$\mathcal{B}^{(p)}{}_{(z)} = - \sum_{k=0}^{p-1} a_{p-k}^{(p)*} z^{-k} + z^{-p} = z^{-p} \mathfrak{Z}^{(p)*}{}_{(z^{-1})},$$

(159)

the solution is

$$a_p^{(p)} = g_p, \ p \geq 1,$$

(160)

with the lower order coefficients given by (148). Thus, the only remaining quantity, $a_p^{(p)}$, that was necessary for solution of (147) - (150) is given by (160)

and (155), along with (152) and (153). To the three lowest orders, the solutions are given by

$$p = 0, \quad P^{(0)} = \hat{R}_0 = \frac{1}{N} \sum_{n=1}^{N} |x_n|^2 \tag{161}$$

$$p = 1, \quad a_1^{(1)} = g_1$$

$$\hat{R}_1 = \hat{R}_0 \, a_1^{(1)}$$

$$P^{(1)} = P^{(0)} \left(1 - \left| a_1^{(1)} \right|^2 \right) \tag{162}$$

$$p = 2, \quad a_2^{(2)} = g_2$$

$$a_1^{(2)} = a_1^{(1)} - a_2^{(2)} \, a_1^{(1)*}$$

$$\hat{R}_2 = \hat{R}_1 \, a_1^{(2)} + \hat{R}_0 \, a_2^{(2)}$$

$$P^{(2)} = P^{(1)} \left(1 - \left| a_2^{(2)} \right|^2 \right). \tag{163}$$

It will be observed that for $p = 1$, $a_1^{(1)}$ is identical to (144); in fact, the procedures are identical in this case. It should also be noted that at each stage, an estimate, $\hat{R}_p$, of the true correlation value R_p becomes available via (149), and is unchanged by the addition of any further stages (larger p).

It was demonstrated in (A-9) that the magnitude of $a_p^{(p)}$ was bounded by unity if the known correlation matrix R was nonnegative definite. The same property,

$$\left| a_p^{(p)} \right| \le 1 , \tag{164}$$

is true here in the case of unknown correlation when $a_p^{(p)}$ is determined by (160) and (155); see appendix F. This is sufficient to show that all zeros of (A-10)

lie inside O; see reference 11, for example. Therefore, the recursion (149) can be used in the form

$$\hat{R}_{\ell} = \sum_{k=1}^{p} \hat{R}_{\ell-k}\, a_k^{(p)}, \quad \ell \geq p + 1 , \tag{165}$$

to extrapolate the estimated correlation values beyond p-th order, with the p-th order coefficients $\{a_k^{(p)}\}_1^p$, and is guaranteed to be stable. Division of (165) by $\hat{R}_0$ yields the normalized correlation coefficients. Recursion (165) is similar in form to those encountered in (75), (111), and (118).

The quantity $P^{(p)}$ that results as the solution (150) of matrix equation (147) is not the minimum average magnitude-squared error as it was for known correlation; see (A-3), (A-7), and (67). In fact, $P^{(p)}$ has no direct physical significance; it is merely the variable left over in that position in the normal equations (147) when modified from the case of known correlation values, (A-3). Rather, $F_0^{(p)}$ in (154) and (156) is the minimum average magnitude-squared error of the forward and backward residuals, (153), of the available data. Thus, as far as picking an "optimum" value of p at which to terminate the recursion in (147) - (150) is concerned, the latter quantity has more physical significance. However, the two quantities are very close to each other for no bad data points, especially for N-p large; see appendix G. Both quantities are readily calculated at any stage via (150) and (156), respectively.

The transfer functions from input x to the p-th order residuals are given in (159). Therefore, the spectra of the residuals are given by

$$G_f^{(p)}(f) = G_b^{(p)}(f) = \left| \mathcal{F}^{(p)}\left(\exp(i2\pi f\Delta)\right) \right|^2 G_x(f) . \tag{166}$$

Now if the chain in figure 3 has been carried to the stage where further values of cross-gain g_p would be substantially zero, then the residuals are approximately white. Therefore, an estimate of the input spectrum is available from (166) and (159) according to

$$\hat{G}_x(f) = \frac{\Delta}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta} , \tag{167}$$

46

where the residual power has been set at unity; see the discussion under (108). Two alternatives to this scale factor are discussed in appendix H; namely, it is shown that $P^{(p)}$ and $F_0^{(p)}$ are both meaningful scale factors that could be applied to (167).

The estimated correlation values in (149) are generally biased. This may be anticipated from the complicated forms of (149), (148), and (155), since additional statistics than simply $\overline{x_{k+p}x_k}$ need to be known in order that $\hat{R}_p$ be capable of evaluation; that is, $\hat{R}_p$ depends on much more than just $\overline{x_{k+p}x_k}$, for the Burg method. This biasedness may be proven for a simple example with $p = 1$, $N = 3$. ($\hat{R}_0$ in (151) is unbiased; and for $p = 1$, $N = 2$, we find $\hat{R}_1 = x_2 x_1^*$, which is unbiased.) For real data, with random variables $\{x_k\}_1^3$ being zero-mean unit-variance Gaussian, and $\overline{x_2 x_1} = \overline{x_3 x_2} = \pm\frac{1}{\sqrt{2}}$, $\overline{x_3 x_1} = 0$, we find (in appendix I) that $\hat{\overline{R}}_1 = \pm\frac{1}{\sqrt{2}} \frac{12 - 2\sqrt{3}}{9} = \pm\frac{1}{\sqrt{2}} (.9484)$. The bias is slight but non-zero.

In summary, the Burg algorithm for data processing consists of initialization (152); followed by the cross-gain calculation in (155); filter coefficients via (160) and (148); and normalized correlation coefficients (149) (if desired) at every stage. The update required at each stage is given by (153), and the extrapolated normalized correlation coefficients at any stage are available from (185), upon division by $\hat{R}_0$.

4.8 SUMMARY OF PROPERTIES OF TECHNIQUES

The solution for the filter coefficients in the techniques considered above can be put in the form

$$\hat{R}(-\tau) = F_0 \; \mathbf{s} . \tag{168}$$

The properties of the estimated correlation matrix $\hat{R}$ (if desired) are tabulated in table 1. (Actually, several of the "No" entries should be "Not Necessarily.")

It will be seen that none of the techniques possesses a "Yes" for all the properties. The Yule-Walker and Burg techniques possess everything but the unbiased property; however, the unbiased property, per se, of the correlation estimates is not necessarily a desirable feature for spectral estimation, as will

Table 1. Properties of Estimated Correlation Matrices

Technique	Correlation Estimates	Unbiased	Hermitian	Toeplitz	Nonnegative Definite	Stable Recursion
Yule-Walker	(114)	No	Yes	Yes	Yes	Yes
Unbiased Yule-Walker	(116)	Yes	Yes	Yes	No	No
Least Squares of Box and Jenkins	(119)	No	Yes	No	No	No
Approximate maximum likelihood of Box and Jenkins	(123)	Yes	Yes	No	No	No
Prediction	(126)	Yes	Yes	No	Yes	No
Forward and Backward Prediction	(138)	Yes	Yes	No	Yes	No
Burg	(149)	No	Yes	Yes	Yes	Yes

be seen by later simulation results. On the other hand, simultaneous satisfaction of the three properties of Hermitian, Toeplitz, and nonnegative definite guarantees that a stable recursion and nonspiky spectral estimates result; see reference 1, page 567.

5. CORRELATION UNKNOWN: FINITE DATA SET WITH BAD DATA POINTS

In some applications, some data values can be bad as a result of malfunctioning equipment or human errors in reading or recording, for example. Also, some data values can be missing as a result of equipment being inadvertently or intermittently turned off for calibration purposes, for example; or some sections of data can be contaminated by strong burst-like noise and be virtually useless in those sections. All of these problems can be characterized mathematically by saying that of the available data set $\{x_n\}_1^N$, the values x_n for the distinct integers

$$n = M_1, \ M_2, \ \ldots \ M_B \tag{169}$$

are known to be bad (or missing). The B bad locations $\{M_j\}_1^B$ are presumed to be known. The bad data points can be regularly spaced, or randomly spaced, or a combination, depending on the application, it will make no difference to the techniques to be developed here.

In this section, we wish to estimate the input spectrum despite the presence of known bad points. The last two methods in subsections 4.6 and 4.7 will be extended to cover this case. The reason we do not extend the other methods in section 4 will become clear when we compare the various techniques by simulation in section 6.

5.1 FORWARD AND BACKWARD PREDICTION USING VALID ERROR POINTS

The method to be presented here is very similar to that given earlier in subsection 4.6; accordingly the treatment will be briefer. For convenience and to enable a better estimation of the true spectrum with a limited number, p, of parameters, we subtract the sample mean of the N-B good data points so that

$$\frac{1}{N-B} \sum_{n=1}^{N} x_n = 0 , \tag{170A}$$

where $\sum$ denotes that we skip those values of n in the set (169); that is, we simply ignore the bad data points -- this is, in fact, the main theme of the methods to be presented. We attempt no interpolation on the bad points, nor do

we set them equal to zero or the sample mean. We also scale the good points so that the sample variance is unity:

$$\frac{1}{N-B-1} \sum_{n=1}^{N} \left| x_n \right|^2 = 1 . \tag{170B}$$

This helps avoid overflow and underflow problems in the numerical manipulation of large arrays encountered for large p.

A forward prediction of x_k is afforded by

$$\hat{x}_k \equiv \sum_{n=1}^{p} a_n x_{k-n} , \, p + 1 \leq k \leq N , \tag{171}$$

provided that $k-1, \ k-2, \ \ldots, \ k-p \neq M_1, \ M_2, \ \ldots, \ M_B$. Then a valid forward error can be defined as

$$\hat{\epsilon}_k \equiv \hat{x}_k - x_k = \sum_{n=0}^{p} a_n x_{k-n} \qquad (a_0 = -1) , \, p + 1 \leq k \leq N , \tag{172}$$

provided that $k, \ k-1, \ \ldots, \ k-p \neq M_1, \ M_2, \ \ldots, \ M_B$; that is, $\hat{\epsilon}_k$ is defined for $p+1 \leq k \leq N$ except for k in the set of integers

$$I_p : \left\{ \begin{array}{cccc} M_1, & M_1 + 1, & \ldots, & M_1 + p \\[1em] M_2, & M_2 + 1, & \ldots, & M_2 + p \\[0.5em] \vdots & & & \\[0.5em] M_B, & M_B + 1, & \ldots, & M_B + p \end{array} \right\} \tag{173}$$

If any numbers in set I_p are $< p+1$ or $> N$, they are not encountered in the error definition (172). Let B_p denote the number of distinct integers in I_p which are $\geq p+1$ and $\leq N$; this is the number of gaps (bad points) in the error sequence (172).

We now define an average forward error over the valid error points as

$$\hat{F} = \frac{1}{N-p-B_p} \sum_{\substack{k=p+1 \\ k \notin I_p}}^{N} \left| \hat{\epsilon}_k \right|^2 , \tag{174}$$

where $\notin$ denotes "not contained in," and $N-p-B_p$ is the number of terms in the sum. Substituting (172) in (174), we obtain

$$\hat{F} = \sum_{m,n=0}^{p} a_m a_n^* \frac{1}{N-p-B_p} \sum_{\substack{k=p+1 \\ k \notin I_p}}^{N} x_{k-m} x_{k-n}^* . \tag{175}$$

A backward prediction of x_k is available as

$$\breve{x}_k \equiv \sum_{n=1}^{p} a_n^* x_{k+n} , \quad 1 \leq k \leq N - p , \tag{176}$$

provided that $k+1, k+2, \ldots, k+p \neq M_1, M_2, \ldots, M_B$. And a backward error

$$\breve{\epsilon}_k \equiv \breve{x}_k - x_k = \sum_{n=0}^{p} a_n^* x_{k+n} \qquad (a_o = -1) , \quad 1 \leq k \leq N - p , \tag{177}$$

is available if $k, k+1, \ldots, k+p \neq M_1, M_2, \ldots, M_B$. Letting $\ell = k+p$ in (177), we can write

$$\breve{\epsilon}_{\ell-p} = \sum_{n=0}^{p} a_n^* x_{\ell-p+n} , \quad p + 1 \leq \ell \leq N , \tag{178}$$

if ℓ is not contained in the set L_p defined in (173). The we can define an average backward error over the valid error points as

$$\check{F} = \frac{1}{N-p-B_p} \sum_{\substack{l=p+1 \\ l \notin I_p}}^{N} \left| \check{\iota}_{l-p} \right|^2 \qquad (179)$$

$$= \sum_{m,n=0}^{p} a_m a_n^* \frac{1}{N-p-B_p} \sum_{\substack{l=p+1 \\ l \notin I_p}}^{N} x_{l-p+n} x_{l-p+m}^* , \qquad (180)$$

where we have substituted (178).

We are now in a position to define an overall average error as

$$F = \frac{1}{2} (\hat{F} + \check{F}) = \sum_{m,n=0}^{p} a_m a_n^* S_{nm} \qquad (181)$$

where, from (175) and (180),

$$S_{nm} = \frac{1}{2(N-p-B_p)} \sum_{\substack{k=p+1 \\ k \notin I_p}}^{N} \left(x_{k-m} x_{k-n}^* + x_{k-p+n} x_{k-p+m}^* \right). \qquad (182)$$

It should be noticed that (182) does not tell us merely to sum over the "good" products, but rather to exclude set I_p. The number of terms in the sum (182) is the same for $0 \leq n$, $m \leq p$ and is $N-p-B_p$. (For no bad points, (182) reduces to (138).) Two useful properties of S_{nm} are

$$S_{mn} = S_{nm}^*, \quad S_{p-n, \, p-m} = S_{nm}^*. \qquad (183)$$

The quantity S_{nm} is an unbiased estimate of R_{n-m}; however, the presence of bad points will increase the variance of S_{nm}; see reference 5. The matrix $[S_{nm}]_1^p$ is Hermitian and nonnegative definite.

The optimum predictive filter coefficients $\{\mathfrak{A}_m\}_1^p$ are obtained by mini-mizing (181):

$$\sum_{m=1}^{p} S_{\ell m} \, \widetilde{a}_m = S_{\ell o} \, , \quad 1 \le \ell \le p \, . \tag{184}$$

The minimum average error is

$$F_o = S_{oo} - \sum_{m=1}^{p} S_{mo}^* \, \widetilde{a}_m \, . \tag{185}$$

And since the sample variance of the good data points was set equal to unity in (170B), (185) is a relative error measure that can be used to decide what value of p should be used in (171) and (176); see reference 1, equations (41) and (89) et seq. The spectral estimate we adopt is obtained by substituting the solution of (184) into (108), as usual. The quantity F_o in (185) could be used as a scale factor, if desired, according to

$$\widehat{E}_o = F_o \, . \tag{186}$$

5.2 BURG TECHNIQUE

The problem setting is the same as that for the previous subsection, including (169) – (170). The solution is identical to that for subsection 4.7, up to (150). Now we define zero-th order forward and backward sequences as

$$f_n^{(0)} = x_n \, , \quad b_n^{(0)} = x_n \, , \quad 1 \le n \le N \, , \quad n \notin I_o \, , \tag{187}$$

where we again employ the definition (173). The first-order sequences are defined as

$$\left. \begin{array}{l} f_n^{(1)} = f_n^{(0)} - g_1 \, b_{n-1}^{(0)} \\[2ex] b_n^{(1)} = b_{n-1}^{(0)} - g_1^* \, f_n^{(0)} \end{array} \right\} \quad \text{for } 2 \le n \le N \, , \quad n \notin I_1 \, , \tag{188}$$

where the restriction of set I_1 is due to the fact that the first-order sequences cannot be formed (evaluated) in set I_1. We choose cross gain g_1 to minimize the average error

$$F^{(1)} \equiv \frac{1}{2(N-1-B_1)} \sum_{\substack{n=2 \\ n \notin I_1}}^{N} \left(\left| f_n^{(1)} \right|^2 + \left| b_n^{(1)} \right|^2 \right), \qquad (189)$$

where $N-1-B_1$ is the number of terms in the sum. The solution is given by

$$g_1 = \frac{2 \sum_{\substack{n=2 \\ n \notin I_1}}^{N} f_n^{(0)} b_{n-1}^{(0)*}}{\sum_{\substack{n=2 \\ n \notin I_1}}^{N} \left(\left| f_n^{(0)} \right|^2 + \left| b_{n-1}^{(0)} \right|^2 \right)}. \qquad (190)$$

With this value of g_1, we can now compute values for residuals $f_n^{(1)}$, $b_n^{(1)}$ in (188) and continue the procedure.

At stage p, we have

$$\left. \begin{aligned} f_n^{(p)} &= f_n^{(p-1)} - g_p \, b_{n-1}^{(p-1)} \\ b_n^{(p)} &= b_{n-1}^{(p-1)} - g_p^* \, f_n^{(p-1)} \end{aligned} \right\} \quad \text{for } p + 1 \leq n \leq N, \, n \notin I_p . \qquad (191)$$

The choice of cross-gain g_p that minimizes average error

$$F^{(p)} \equiv \frac{1}{2(N-p-B_p)} \sum_{\substack{n=p+1 \\ n \notin I_p}}^{N} \left(\left| f_n^{(p)} \right|^2 + \left| b_n^{(p)} \right|^2 \right) \qquad \left(F^{(0)} = 1 \right) \quad (192)$$

is

$$g_p = \frac{2 \sum\limits_{\substack{n=p+1 \\ n \notin I_p}}^{N} f_n^{(p-1)} b_{n-1}^{(p-1)*}}{\sum\limits_{\substack{n=p+1 \\ n \notin I_p}}^{N} \left(\left| f_n^{(p-1)} \right|^2 + \left| b_{n-1}^{(p-1)} \right|^2 \right)} \equiv \frac{Num(p)}{Den(p)}, \qquad (193)$$

and the minimum value of (192) can be expressed as

$$F_o^{(p)} = \left(1 - \left| g_p \right|^2 \right) \frac{Den(p)}{2(N-p-B_p)} \qquad \left(F_o^{(0)} = 1 \right). \qquad (194)$$

This is also a relative error, due to the normalization (170B), and can be used as an indicator when to terminate the recursion procedure in (191).

It may be seen from (192) and (193) that the sums are merely taken over those values of n where the summands are defined. The number of terms in all the sums is $N-p-B_p$.

As in subsection 4.7, the filter coefficients are given by

$$a_p^{(p)} = g_p, \quad p \geq 1, \qquad (195)$$

and for $p \geq 2$, by

$$a_k^{(p)} = a_k^{(p-1)} - a_p^{(p)} a_{p-k}^{(p-1)*}, \quad 1 \leq k \leq p - 1. \qquad (196)$$

Equations (147) through (150) still hold true. The starting value of $P^{(0)}$ is now 1, by virtue of normalization (170B). Recursion (165) for $\ell \geq p+1$ is still valid and is stable since

$$\left| a_p^{(p)} \right| = \left| g_p \right| \leq 1, \qquad (197)$$

as may be seen from (193) and appendix F. The spectral estimate is again given by (167). The discussions in appendixes G and H are relevant here also.

6. COMPARISONS

All the techniques considered in section 4 will now be compared in terms of their resolution capability, bias, and statistical stability, by means of a simulation approach. In particular, the fourth-order autoregressive process which was intensively investigated in reference 2 (see figures 4a and 5a) will be the basic process of interest here also. It is characterized by

$$x_k = \sum_{n=1}^{4} \alpha_n x_{k-n} + w_k,$$ (198)

where

$$\alpha_1 = 2.7607, \ \alpha_2 = -3.8106, \ \alpha_3 = 2.6535, \ \alpha_4 = -0.9238,$$ (199)

and where $\{w_k\}$ is Gaussian white noise. We restrict consideration to real processes here. We will not address the problem of how best to pick the value of p used in the techniques of sections 4 and 5, but shall instead set p equal to the known value, 4, and concentrate on the ability of the various techniques to estimate the parameters in (199), and thereby the spectrum of $\{x_k\}$, from a finite set of N data points.

The simulation method consists of the generation of 100 independent realizations of the process in (198) in steady state. The coefficients in (198) are estimated for each of the 100 realizations, and the corresponding 100 estimated spectra are computed by means of (108), for every technique in sections 4 and 5. The examples to be considered are summarized in table 2, where N is the number of data points in each realization (trial), and B is the number of bad points in each realization. The corresponding figures are collected together at the end of this section.

6.1 NO BAD DATA POINTS

In figure 4A, the 100 different estimated spectra, one for each of the 100 independent trials, are plotted for the Yule-Walker approach, and for N = 40 data points. In figure 4B, the (power) average spectrum of the 100 estimated spectra is plotted, along with the true spectrum of process (198) and (199). The true spectrum is scaled so that its area is equal to that of the average spectrum. It will be seen from figure 4A that there is a great deal of variability in the individual spectral estimates. From figure 4B, we observe that the

Table 2. Simulation Examples

Figure Number	Number of Data Points N	Number of Bad Points B	Technique
4	40	0	Yule-Walker
5	40	0	Yule-Walker, Unbiased
6	40	0	Least Squares of Box and Jenkins
7	40	0	Approximate Maximum Likelihood of Box and Jenkins
8	40	0	Prediction, Valid Error Points
9	40	0	Forward & Backward Prediction
10	40	0	Burg
11	40	0	Burg, Uniform Noise
12	40	4	Forward & Backward Prediction
13	40	4	Burg
14	100	0	Forward & Backward Prediction
15	100	0	Burg
16	100	10	Forward & Backward Prediction
17	100	10	Burg
18	100	20	Forward & Backward Prediction
19	100	20	Burg
20	100	30	Forward & Backward Prediction
21	100	30	Burg

average spectrum does not resolve the two narrowband peaks of the true spectrum*; in fact, this same conclusion is true for the individual spectra in figure 4A. A severe bias exists in the skirts of the average spectrum, which gives a gross overestimate of the power in bands away from the peaks. Thus, the Yule-Walker approach has poor resolution, severe bias, and substantial variability.

The corresponding results for the unbiased version of the Yule-Walker approach are displayed in figure 5. Rather than improving the situation, it is found that the spectral estimates are worse in every regard. The spectral estimates with strong spikes near $f = \frac{1}{8\Delta}$ are manifestations of pole-pair locations of estimate (108) that are very close to the unit circle O. Recall from subsection 4.2 that the zeros of (112) need not lie inside O; see the discussion below (118).

The unbiased correlation estimates utilized above in the normal equations are of the same form as those suggested in reference 5 for missing data, when spectral estimation is attempted directly via (2). But since the performance of these unbiased correlation estimates is so poor here, they are not considered worthwhile in the presence of bad data points, when spectral estimation is accomplished via (108). Whether they are worthwhile for use in (2) is not known.

Results for the least-squares approach of Box and Jenkins are given in figure 6. The variability is less than that for the Yule-Walker estimates in figure 4A. And some resolution is achieved in figure 6B, in addition to good skirt selectivity. There is still, however, a large number of spiky spectral estimates, as anticipated in the discussion under (121).

Conditions are not much improved for the approximate maximum likelihood method of Box and Jenkins presented in figure 7. There happens to be one particular spectral estimate with a very large spike (a zero virtually on O) that severely influences the average power level. The variability in the estimated skirt level is quite small for this technique (as well as for the previous one).

In figure 8, the results for prediction using only the valid error points are presented. The resolution and bias in figure 8B are observed to be very good,

---

*This same conclusion is reached in reference 2, figure 5b, for the same number of data points. Increasing p (above 4) does recover some of the resolution of the two narrowband peaks, but it does not reduce the severe bias of the Yule-Walker approach.

and except for some spiky estimates in figure 8A, the variability of the individual estimates is fairly small.

The situation is still better when we consider forward and backward prediction, using only the valid error points, in figure 9. There are a couple of spiky estimates, but they are not excessively large, as they were previously. The bias and resolution are very good in figure 9B. Although the zeros of (112) need not remain inside O for this technique, it was found that in all 100 trials, no zeros were ever located outside of O. The presence of the two spiky estimates, however, indicates that on two occasions, a zero came close to the periphery of O.

One of the major drawbacks of this technique is the need to invert a non-Toeplitz matrix (or an equivalent operation) in order to evaluate the optimum filter coefficients; see (140). For large p, this is a significant numerical problem. We therefore attempted to convert the matrix $[S_{\ell m}]$ in (140) to a Toeplitz matrix, so that the recursive solution in appendix A could be utilized. We first averaged $[S_{\ell m}]_1^p$ down the diagonals and left the right-hand side of (140) as is; however, we lost resolution and got badly biased and more variable spectral estimates. Next we diagonally averaged $[S_{\ell m}]_0^p$ and left the right-hand side of (140) alone, but got the same bad effects. Finally, we diagonally averaged $[S_{\ell m}]_0^p$ and also replaced the terms on the right-hand side of (140) by the appropriate averages, but again to no avail. Thus, we are unable to significantly modify (140) without dire effects on the spectral estimate.

Finally, when the Burg technique is considered in figure 10, we observe the complete absence of spiky estimates; this is due mainly to the guaranteed location of the zeros of (A-10) inside O. In other respects, the results of figures 9 and 10 are very similar. There is a small bias in figure 10B, with the peaks being rounded off and the valley filled in; this is similar to figure 5 in reference 2.

All the results above have been conducted for Gaussian white noise $\{w_k\}$ in (198). To see the effect of the statistics of $\{w_k\}$ upon the spectral estimates, we changed to a uniform distribution. The results in figure 11 are virtually identical to those in figure 10. Accordingly, Gaussian statistics are kept for the remainder of the simulation.

6.2 BAD DATA POINTS

By virtue of the results of the preceding subsection, further consideration is limited to the forward and backward prediction technique and the Burg technique. The first example we consider is $B = 4$ bad data points out of a total of $N = 40$ data points; that is, in each of the 100 realizations of 40 data points, 4 points (no more, no less) were randomly selected as being bad, and the corresponding values of x_k were suppressed. In some of the realizations, the four data points may have been close together (for example, 10, 12, 14, 15), whereas in other realizations, they might have been far apart (for example, 1, 14, 27, 40).

The resulting spectral estimates are given in figures 12 and 13. The variability in the skirts is less for the forward and backward prediction technique than for the Burg technique. However, the spiky nature of the former technique is quite evident in comparison with the latter technique. Both techniques have suffered a significant loss of resolution near the narrowband peaks.

The reason for the significant degradation in performance of both techniques is that although only $B/N = 4/40$ (10%) of the points are bad, the number of valid error points, $N-p-B_p$ in (174) and (192), can decrease significantly. For example, for $p = 4$ and spaced bad points at $M_1 = 11$, $M_2 = 16$, $M_3 = 21$, $M_4 = 26$ (see (169)), we have

$$B_p = 20, \quad N - p - B_p = 16. \tag{200}$$

On the other hand, for contiguous bad points at $M_1 = 1$, $M_2 = 2$, $M_3 = 3$, $M_4 = 4$, we have

$$B_p = 4, \quad N - p - B_p = 32. \tag{201}$$

Thus, anywhere from 16 to 32 valid error points can be achieved. The stability of the spectral estimate for (200) will be less than that for (201). Generally, contiguous bad points are less damaging than spaced bad points, because more valid error points can be formed when the bad points are contiguous.

One of the points of the above example is that 4 bad data points out of 40 is rather detrimental. We consider now $N = 100$ data points. The first example of interest will serve as a comparison case and is $B = 0$. The results of spectral estimation for the forward and backward prediction technique and Burg

technique are given in figures 14 and 15, respectively. The results are virtually identical; there is excellent resolution and almost no bias for both techniques.

When B is increased to 10, the results in figures 16 and 17 are obtained. Despite 10% bad points, good performance in terms of stability, bias, and resolution is attained. The number, $N-p-B_p$, of valid data points can vary from 46 to 86; however, the likelihood of realizing as few as 46 on a random basis is very remote. The Burg technique has less-spiky estimates near the narrowband peaks, as expected; however, it is more variable in the skirts than the forward and backware prediction technique.

When B is increased to 20, the results in figures 18 and 19 indicate that the Burg technique has more variability, but is less spiky and has better resolution. The same conclusions hold true for B = 30 in figures 20 and 21; however, neither technique resolves the two narrowband peaks for this many bad data points.
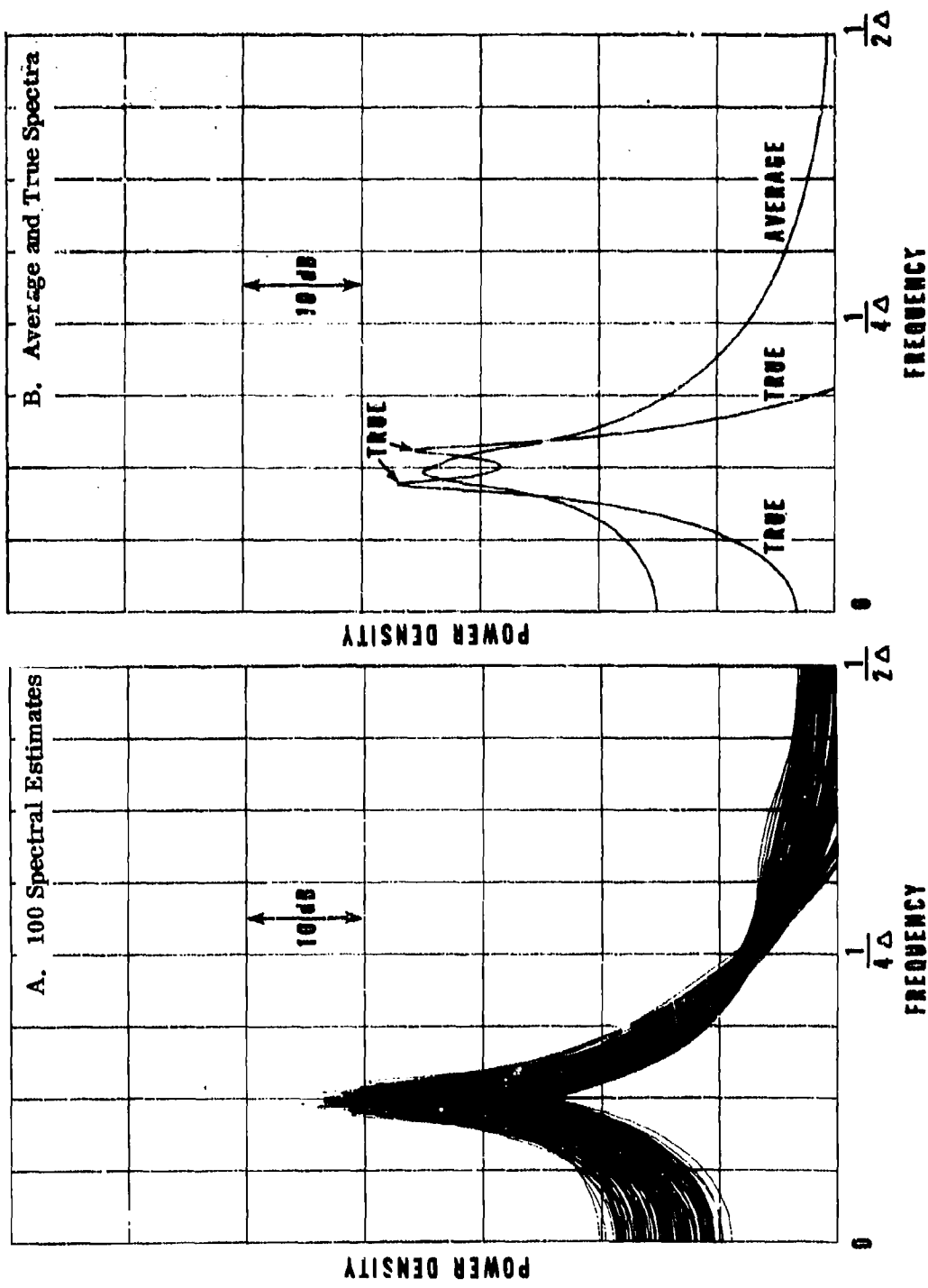
Figure 4. Yule-Walker; N = 40, B = 0
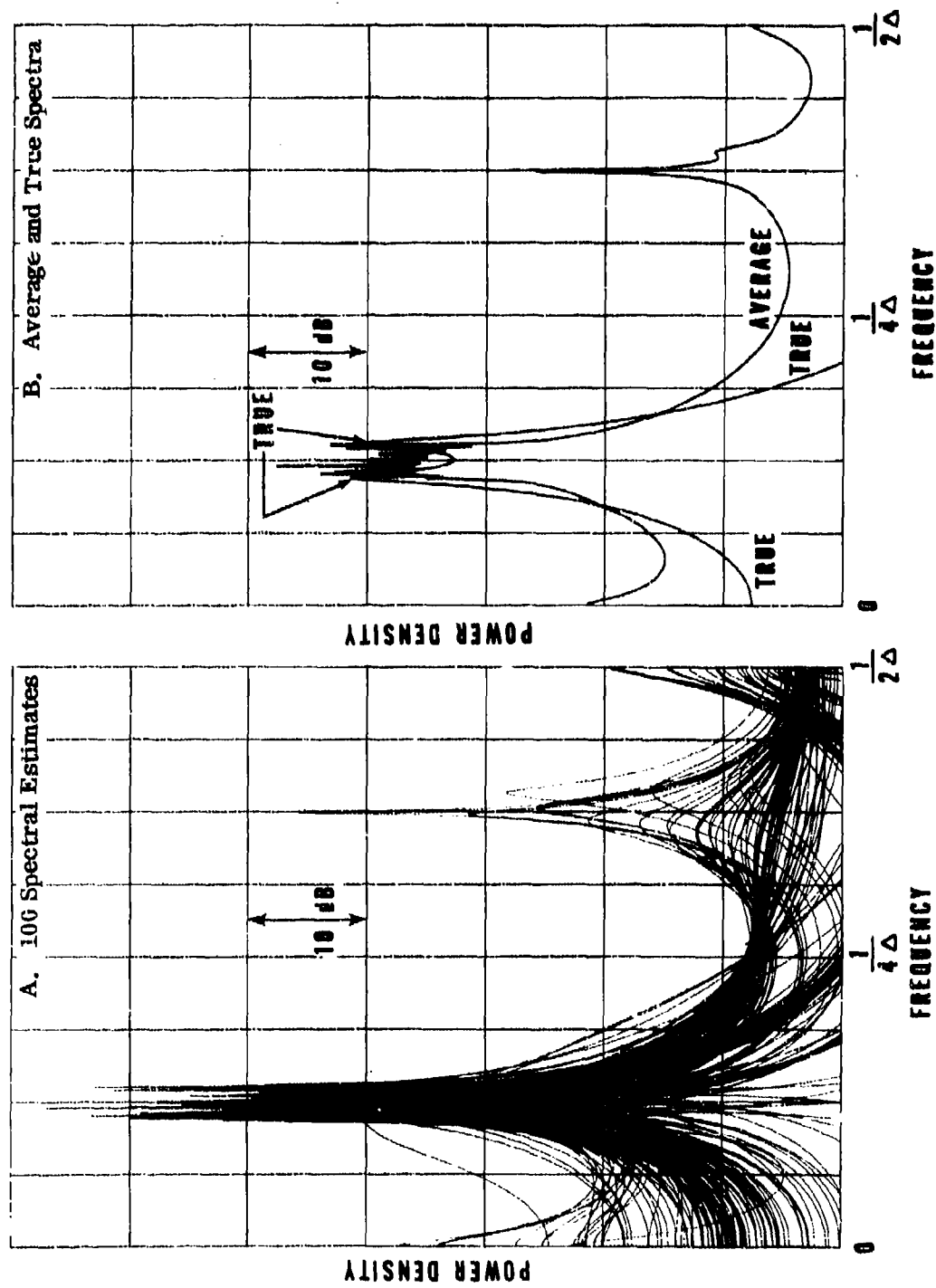
Figure 5. Yule-Walker, Unbiased; N = 40, B = 0
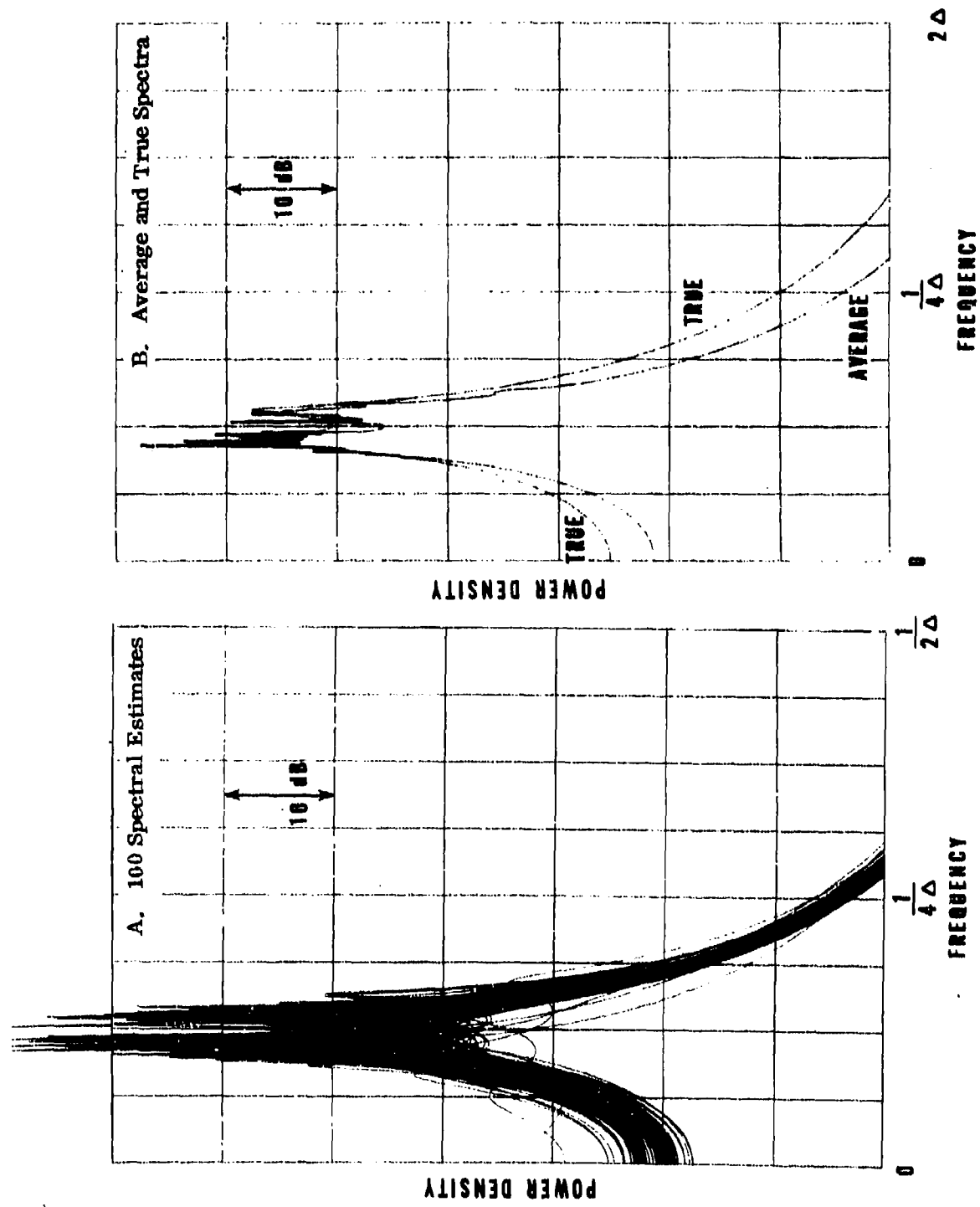
Figure 6. Least Squares of Box and Jenkins; N = 40, B = 0
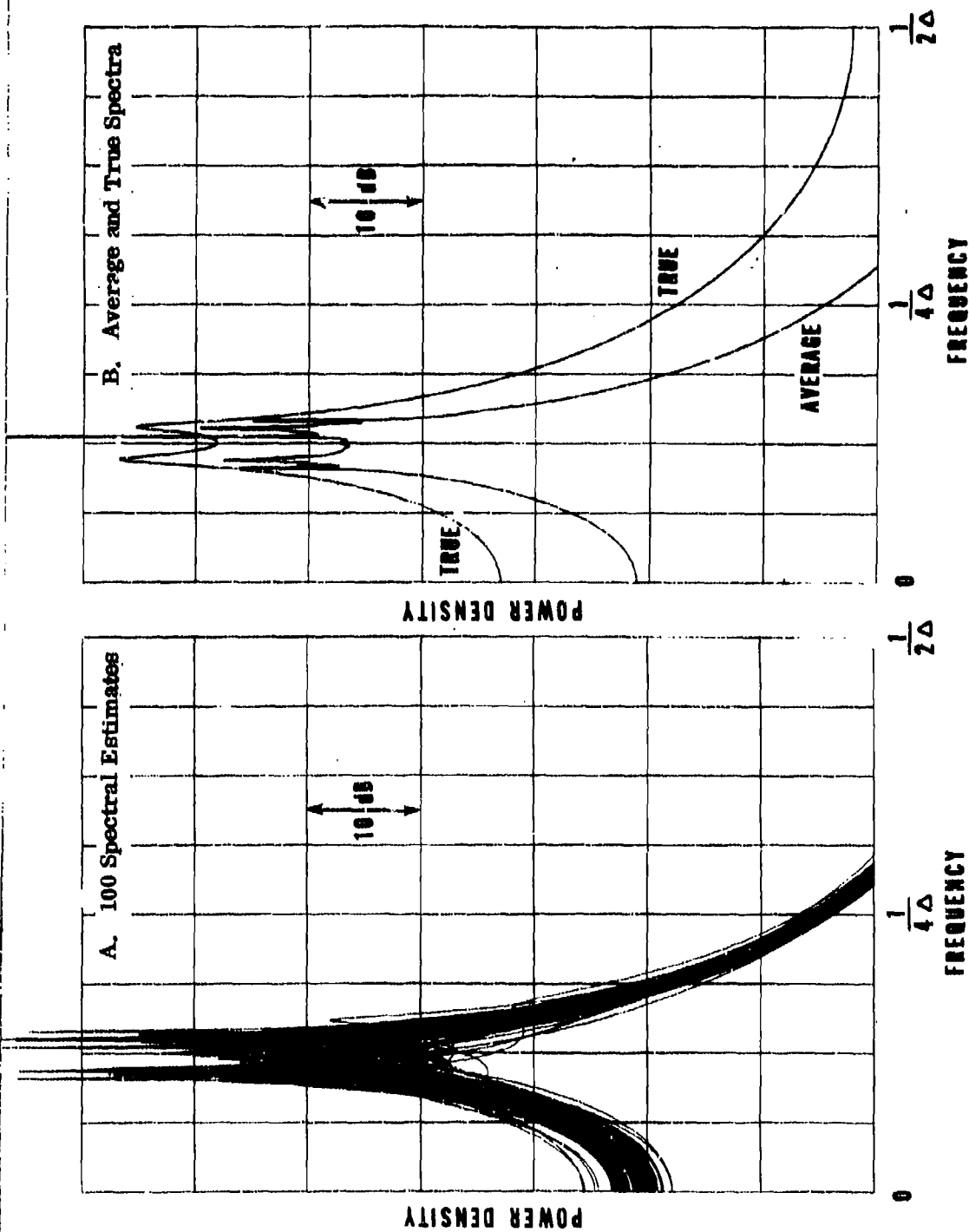
Figure 7. Approximate Maximum Likelihood of Box and Jenkins; N = 40, B = 0
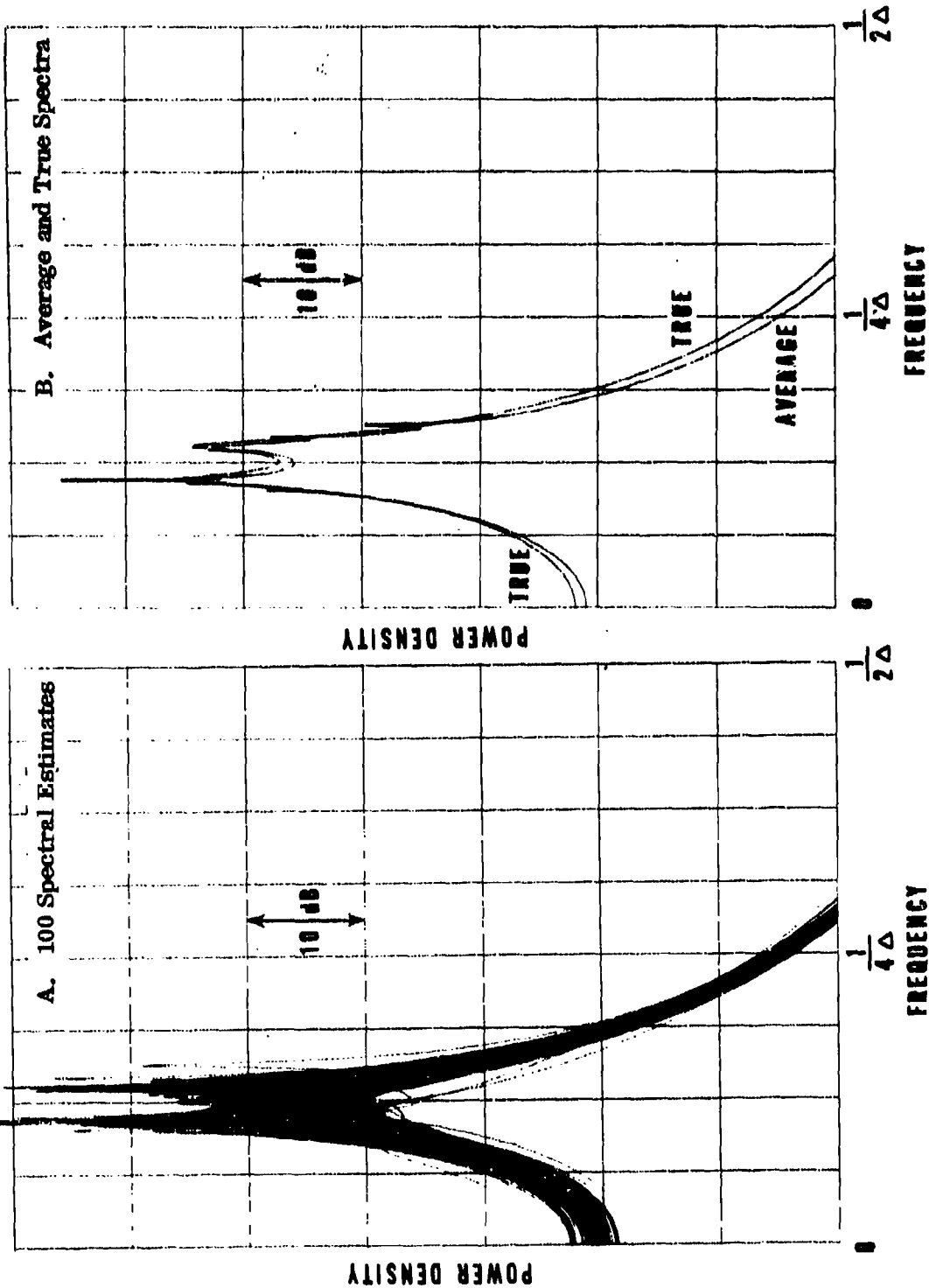
Figure 8. Prediction, Valid Error Points; N = 40, B = 0
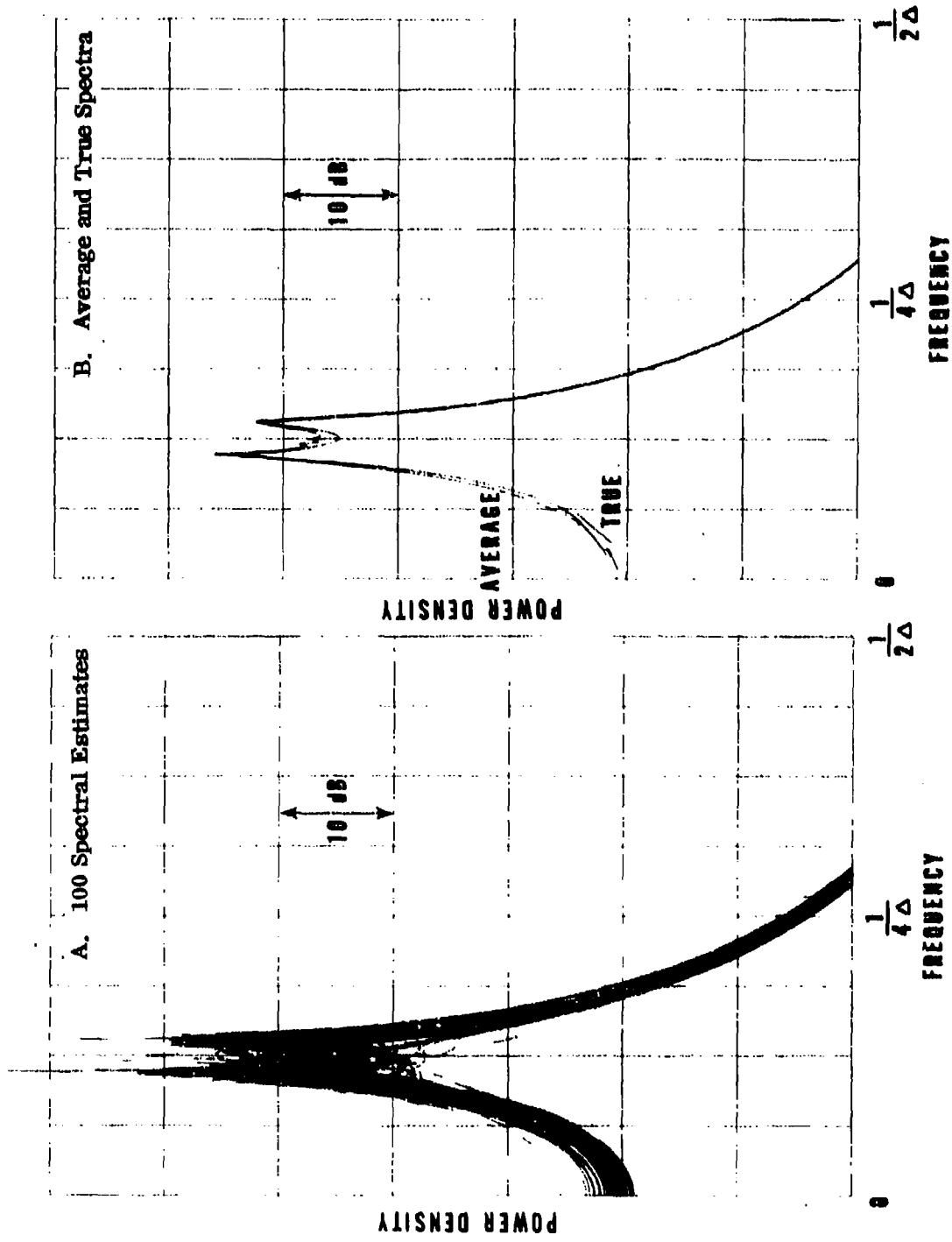
Figure 9. Forward and Backward Prediction; $N = 40$, $B = 0$
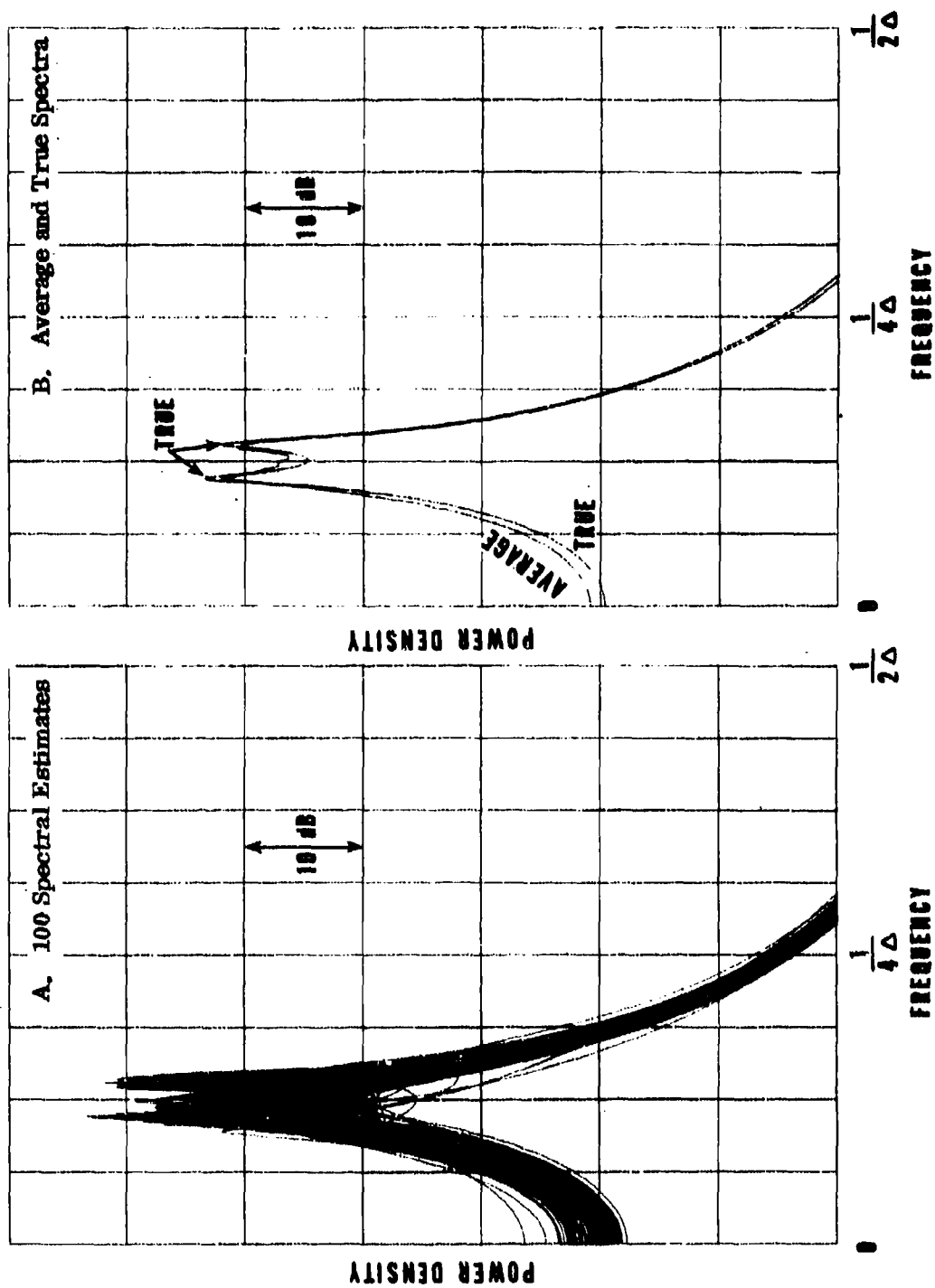
Figure 10. Burg; N = 40, B = 0
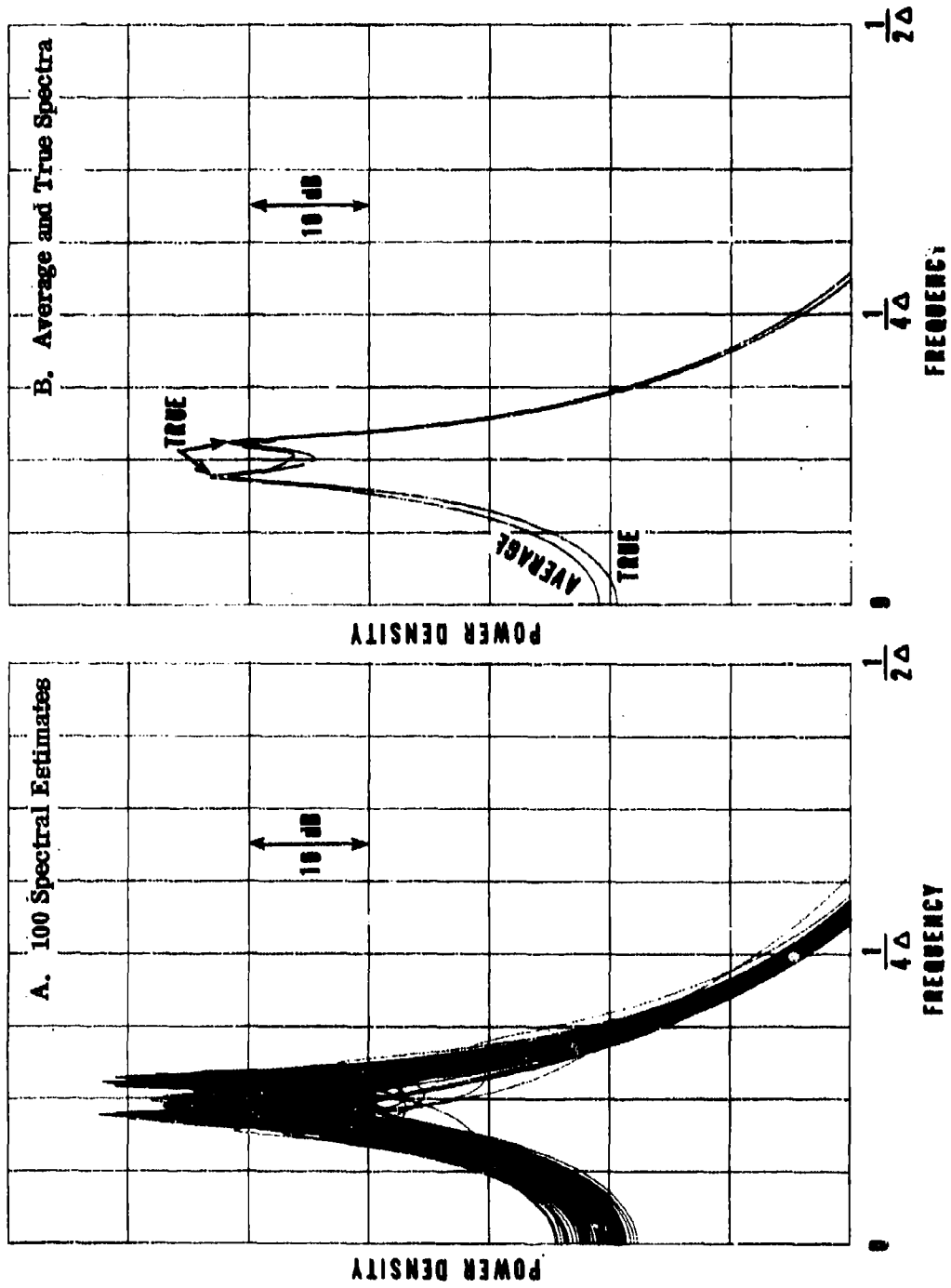
Figure 1L. Burg, Uniform Noise; N = 40, B = 0

Figure 12. Forward and Backward Prediction; N = 40, B = 4

Figure 13. Burg; N = 40, B = 4

Figure 14. Forward and Backward Prediction; N = 100, B = 0

Figure 15. Burg; N = 100, B = 0

Figure 16. Forward and Backward Prediction; N = 100, B = 10
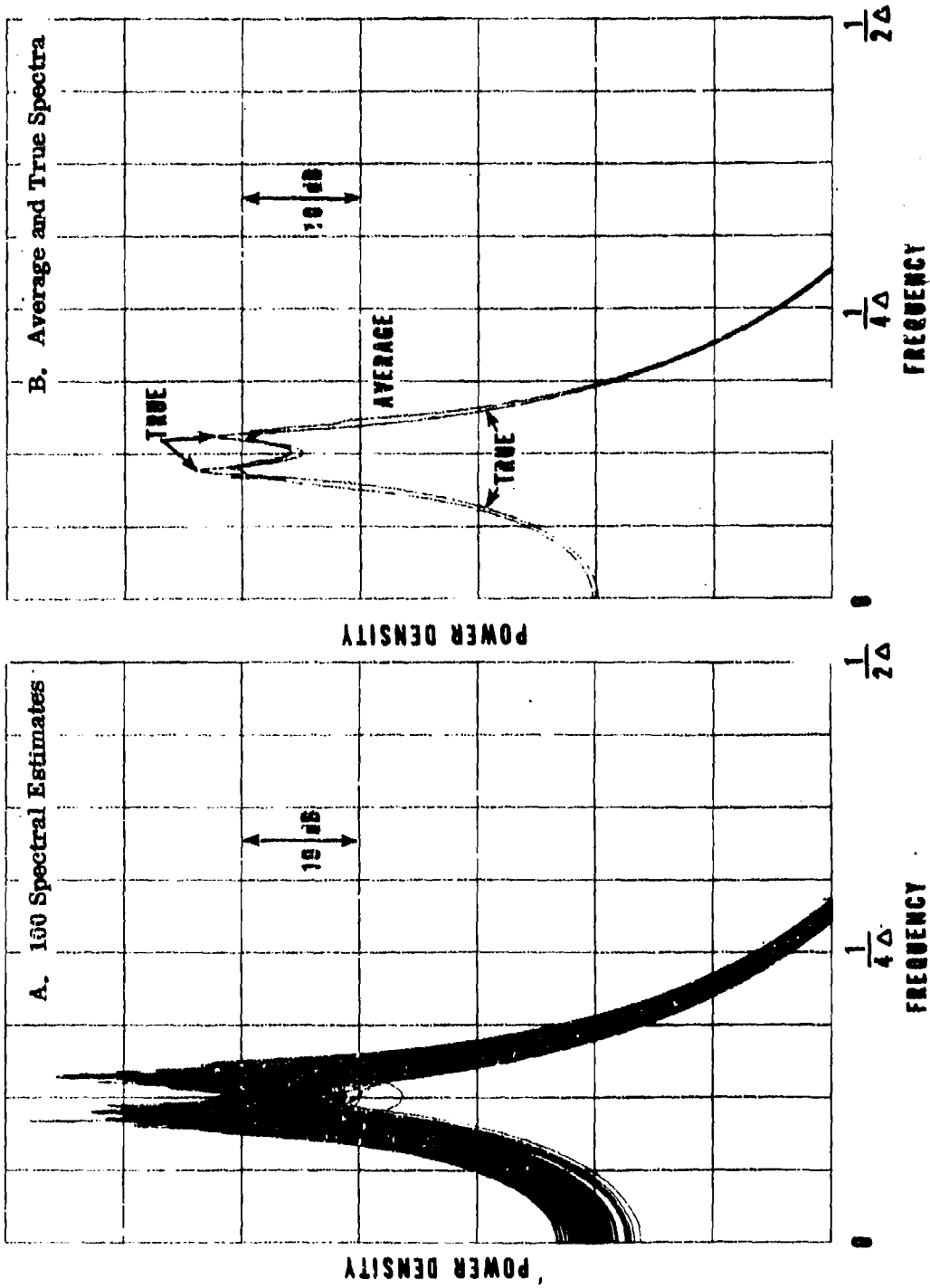
Figure 17. Burg; N = 100, B = 10
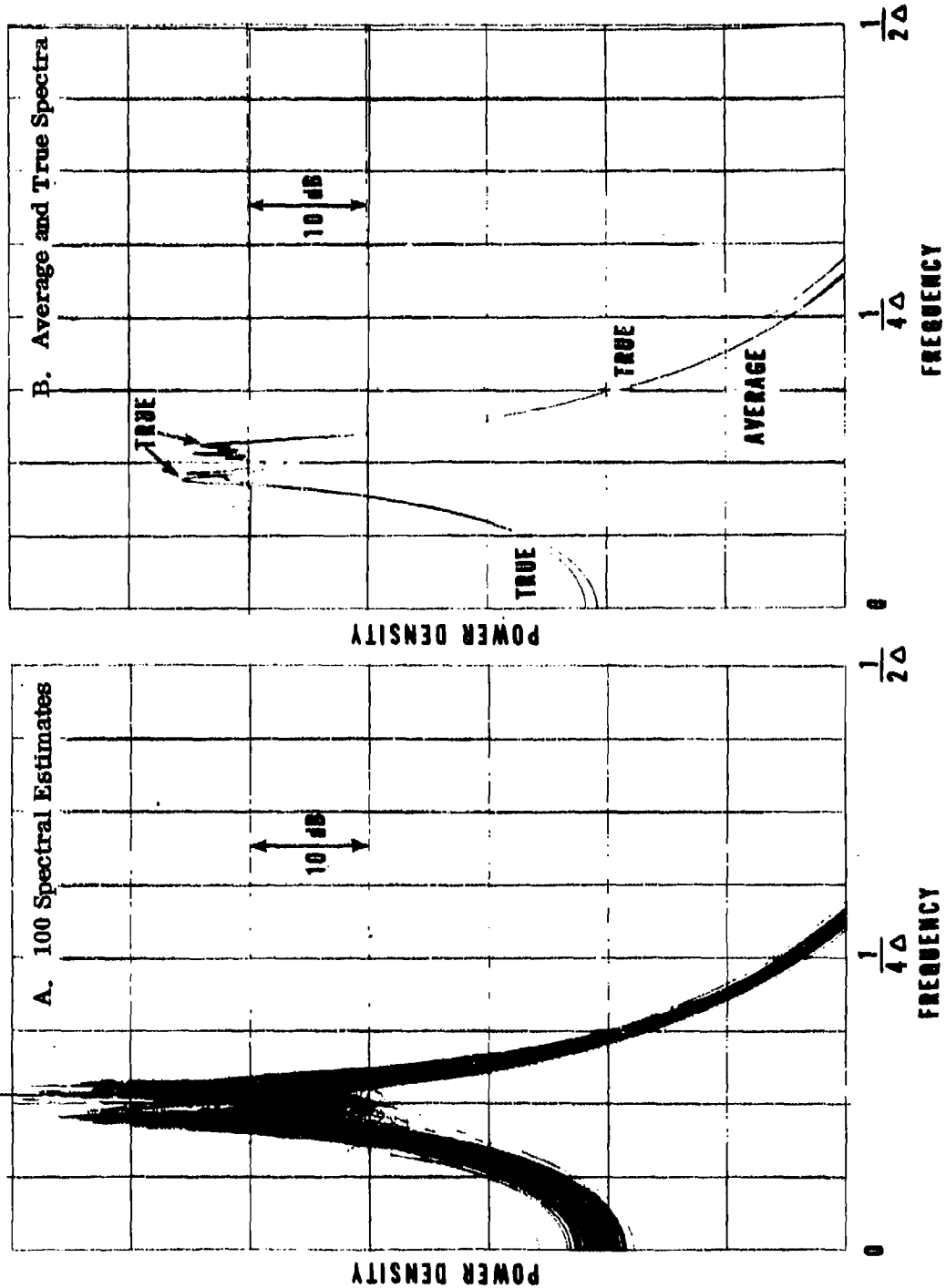
Figure 18. Forward and Backward Prediction; N = 100, B = 20
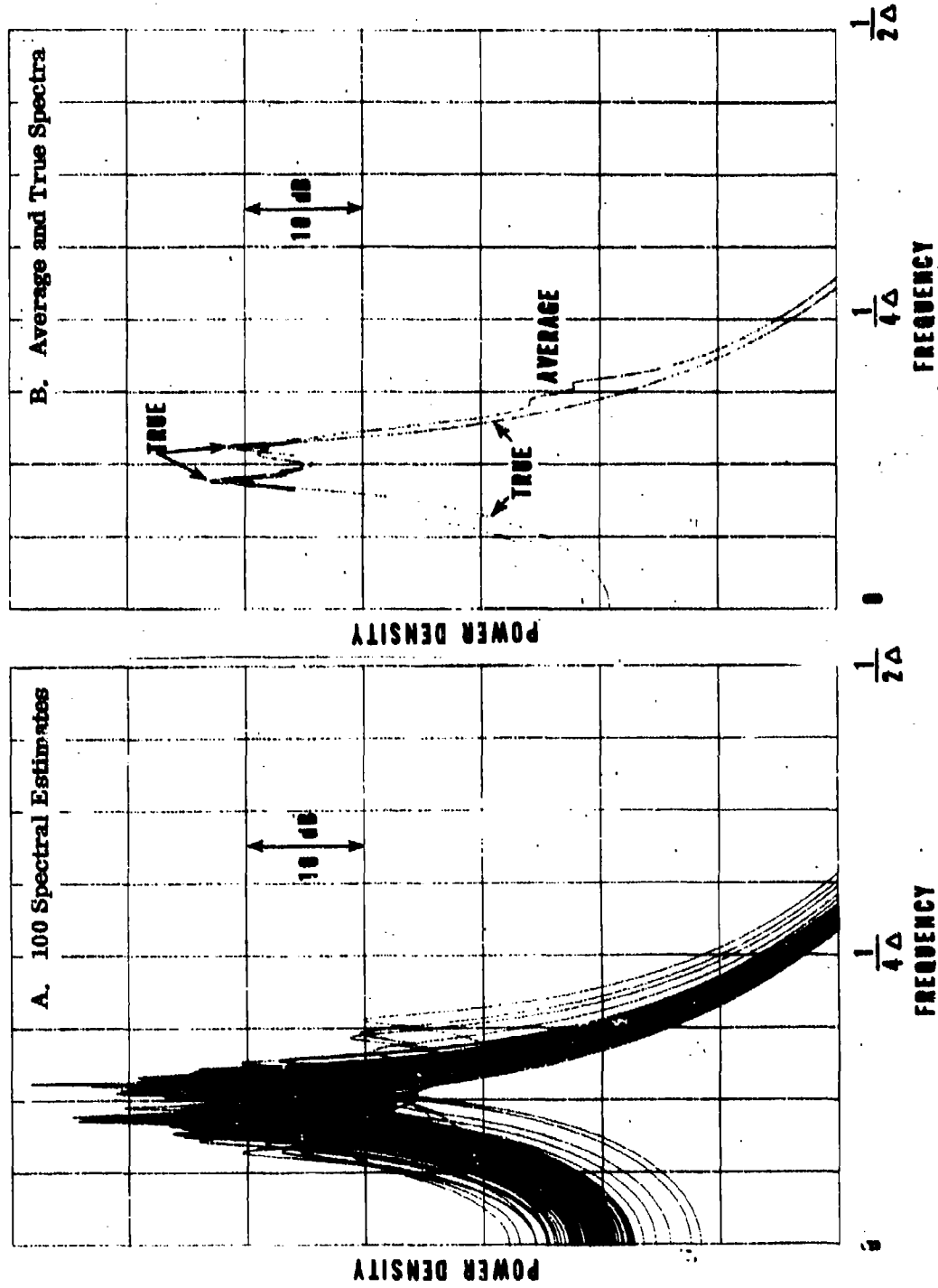
Figure 19. Burg; N = 100, B = 20
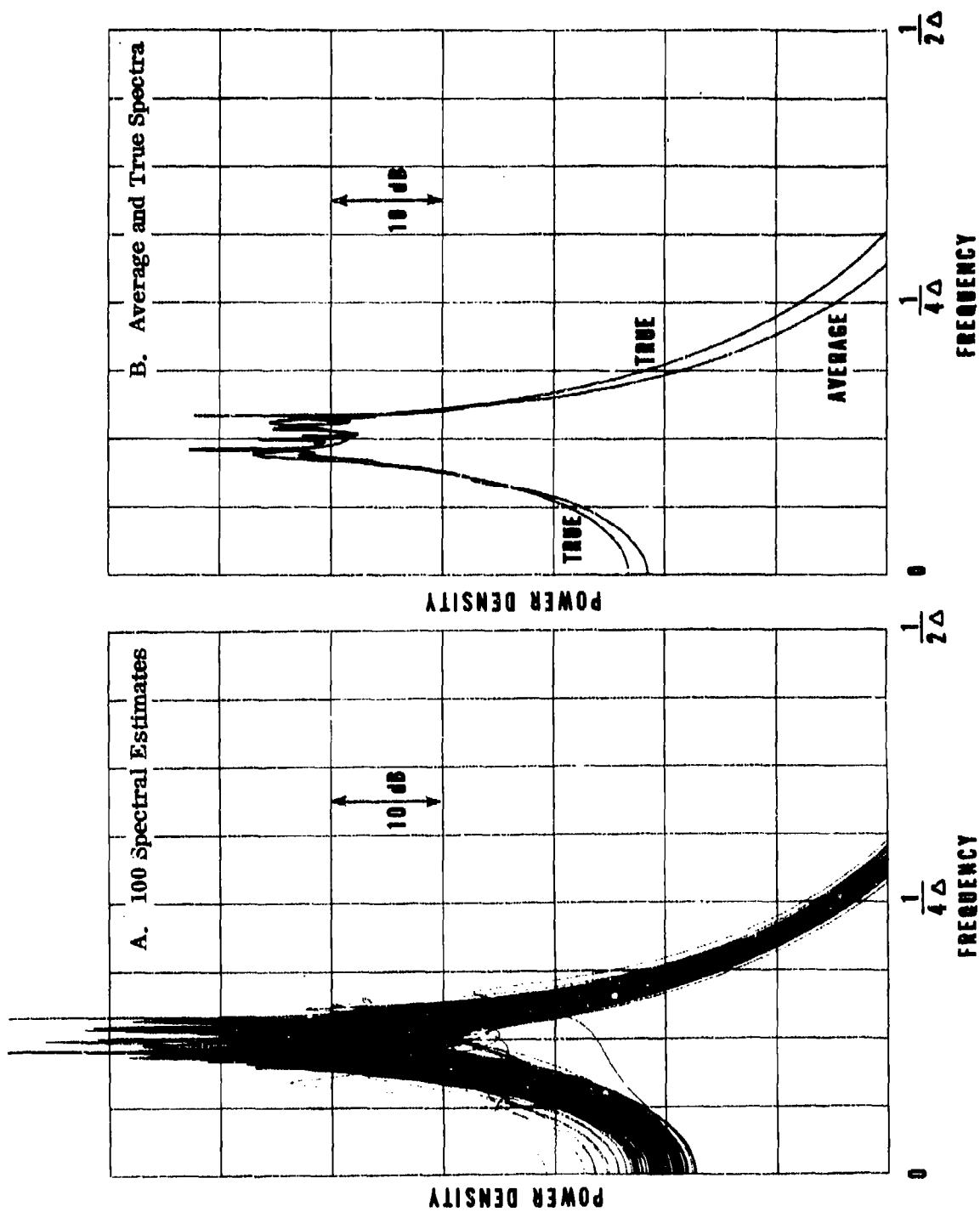
Figure 20. Forward and Backward Prediction; N = 100, B = 30
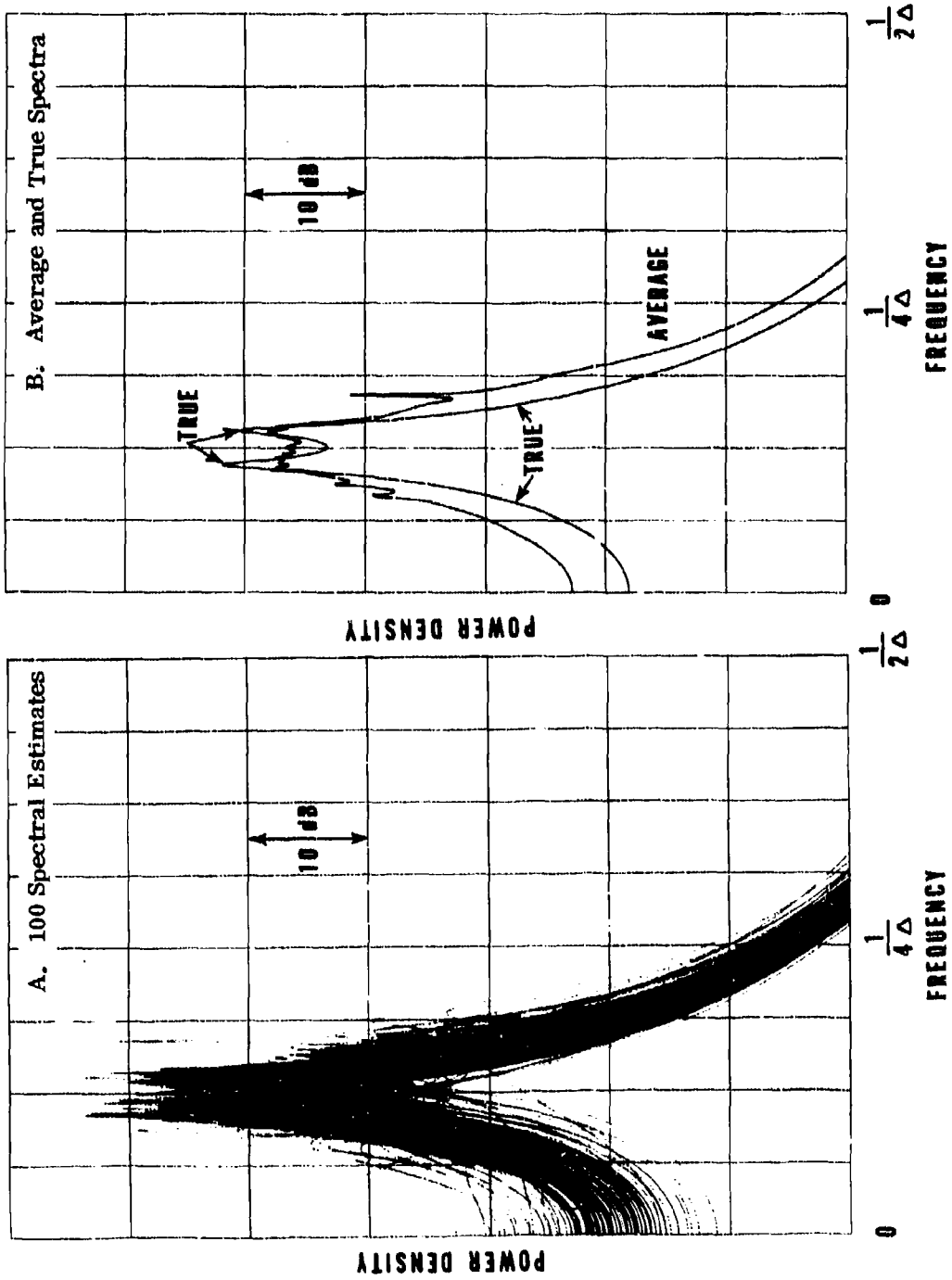
Figure 21. Burg; N = 100, B = 30

7. DISCUSSION AND CONCLUSIONS

Several methods of spectral estimation via linear predictive techniques have been considered for a univariate process, both with and without bad data points; the bad points can be regularly spaced, randomly spaced, or a combination. Two particular methods have been found to have better performance than the remainder, 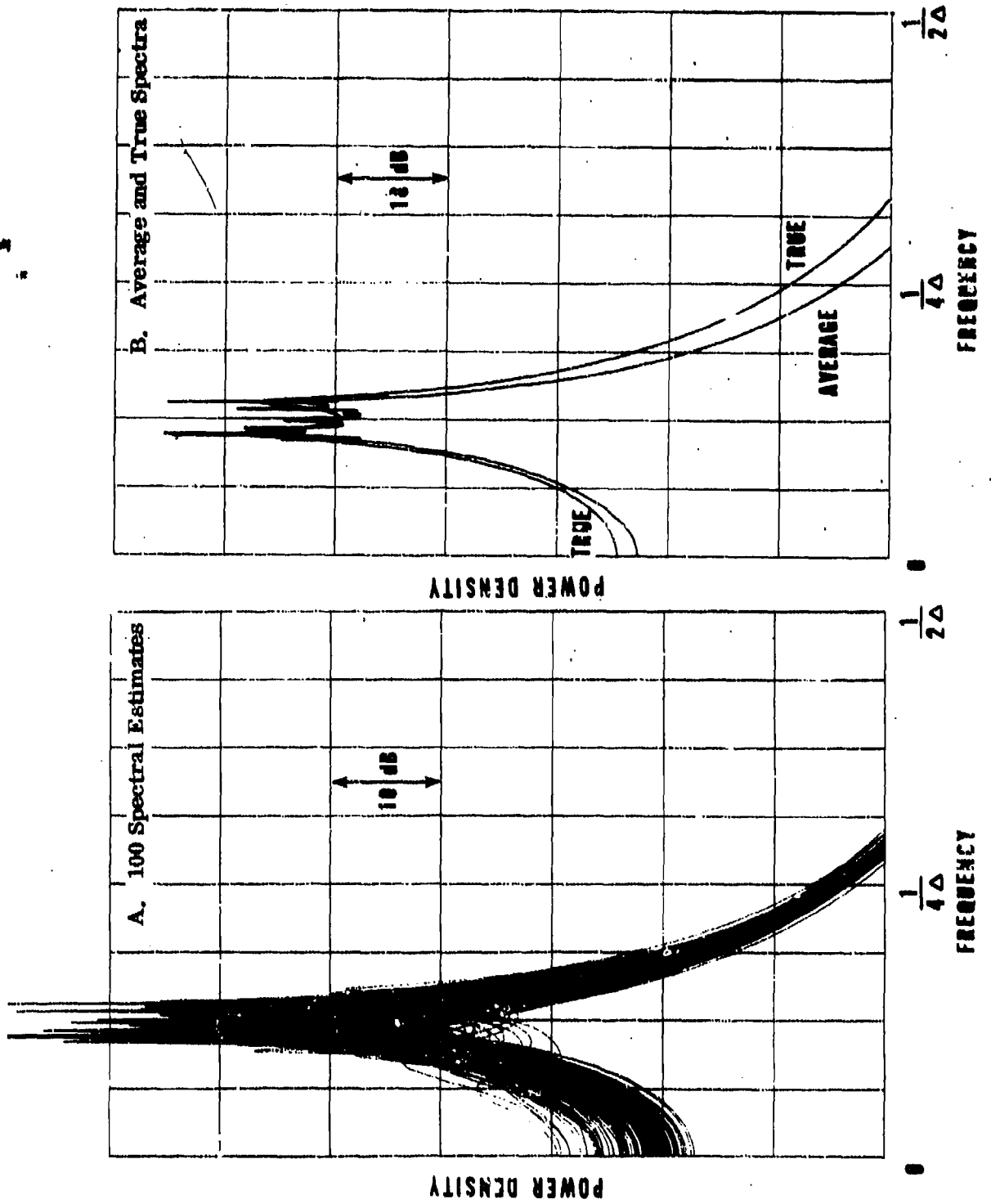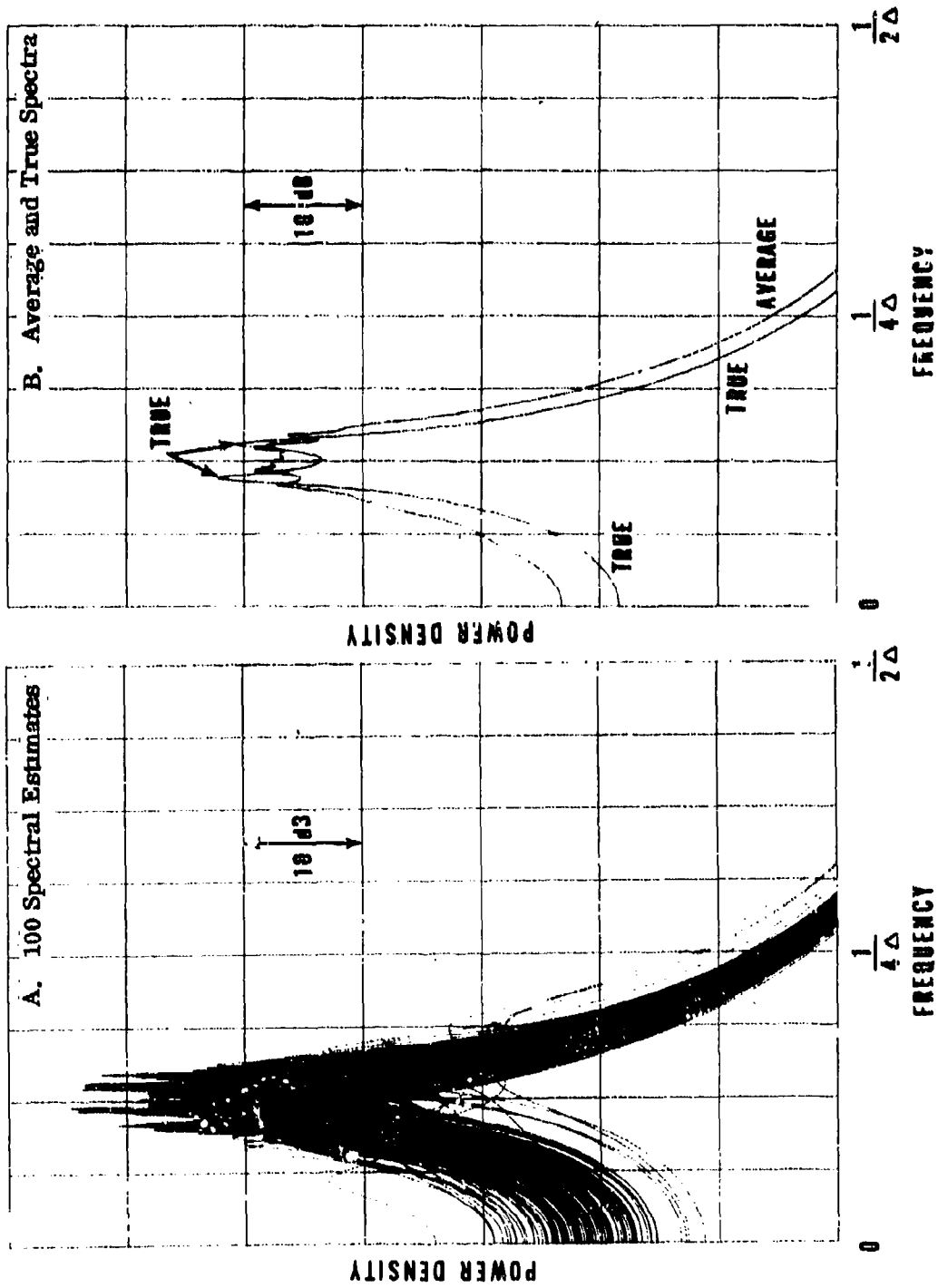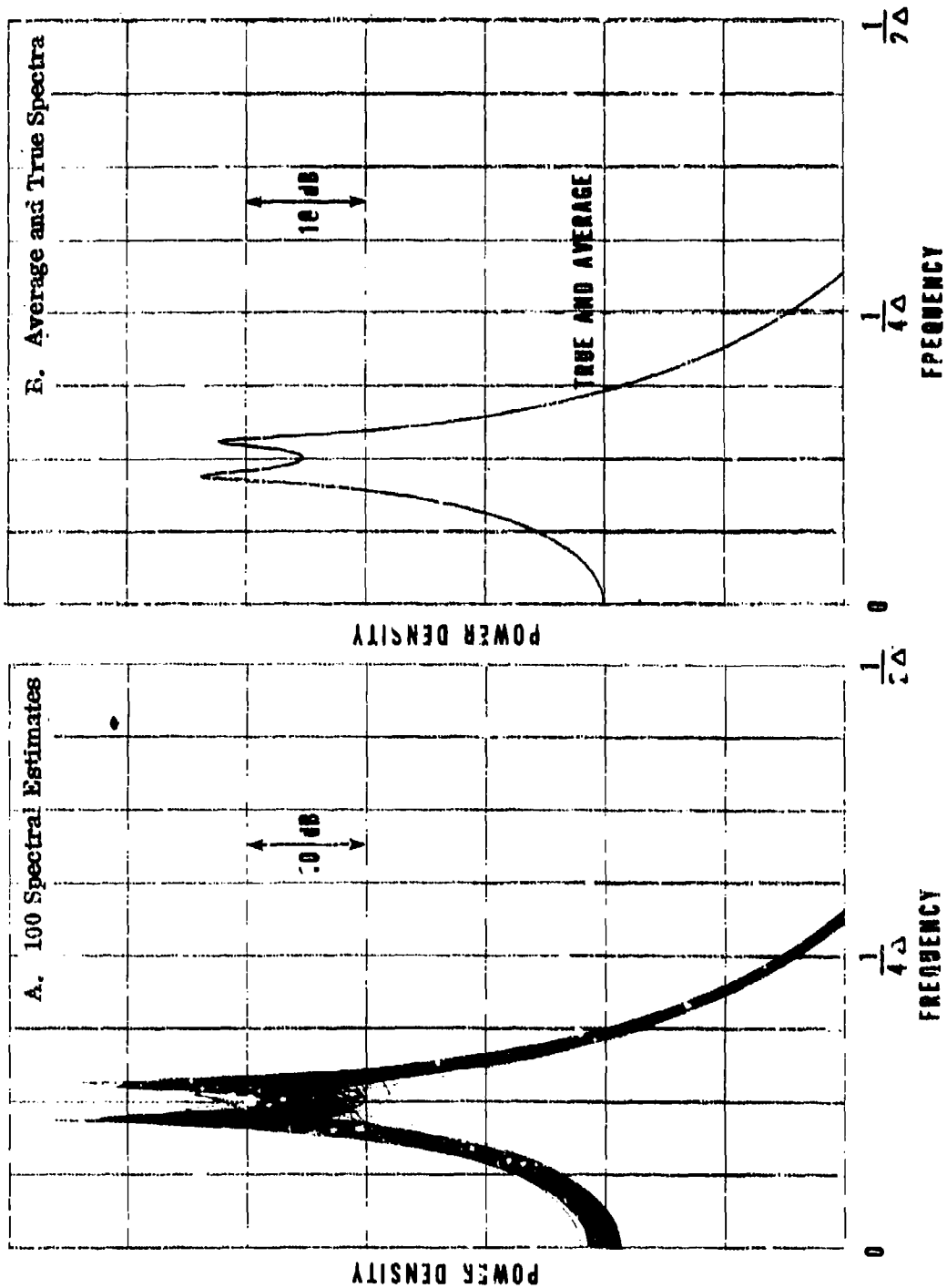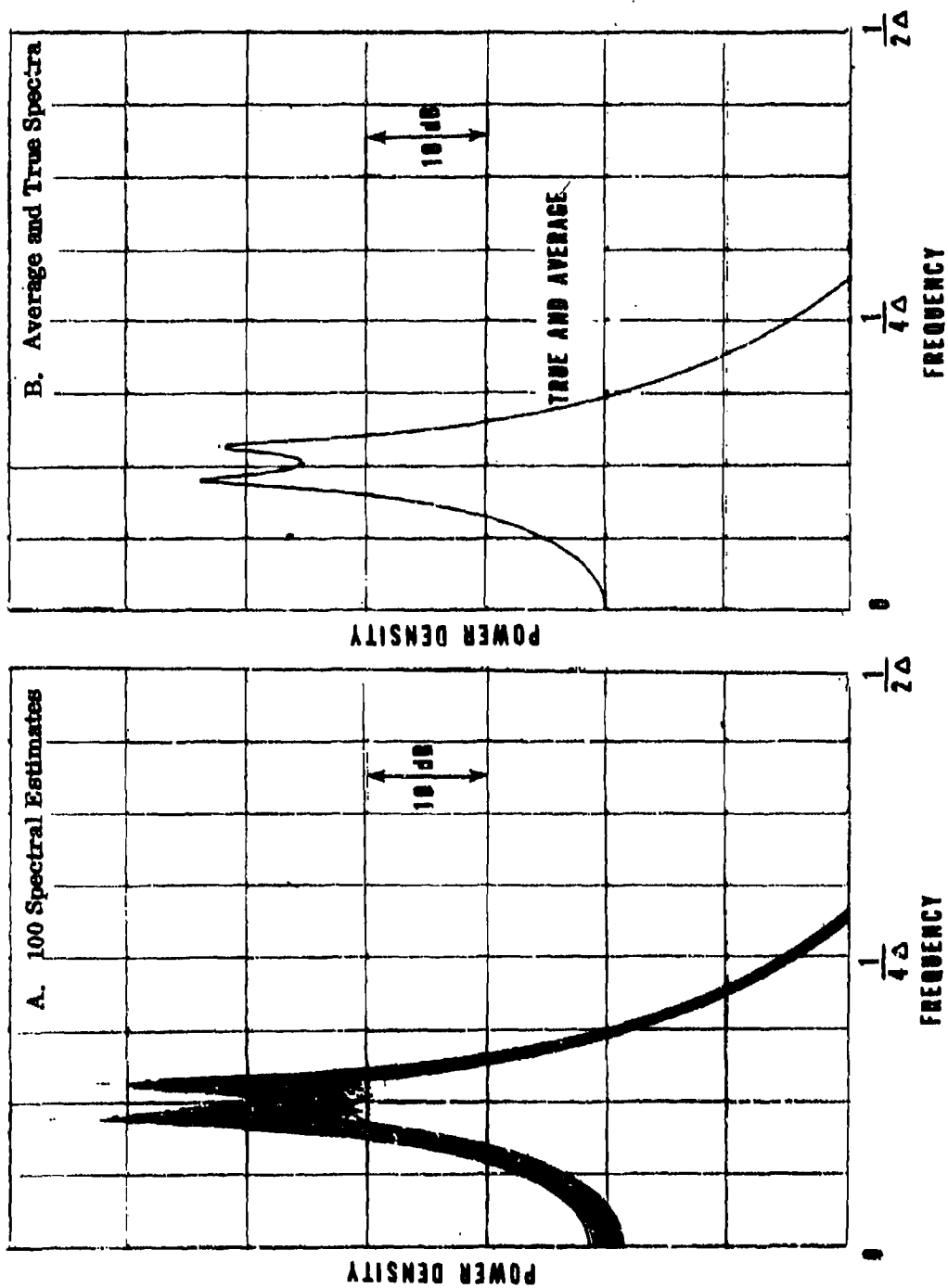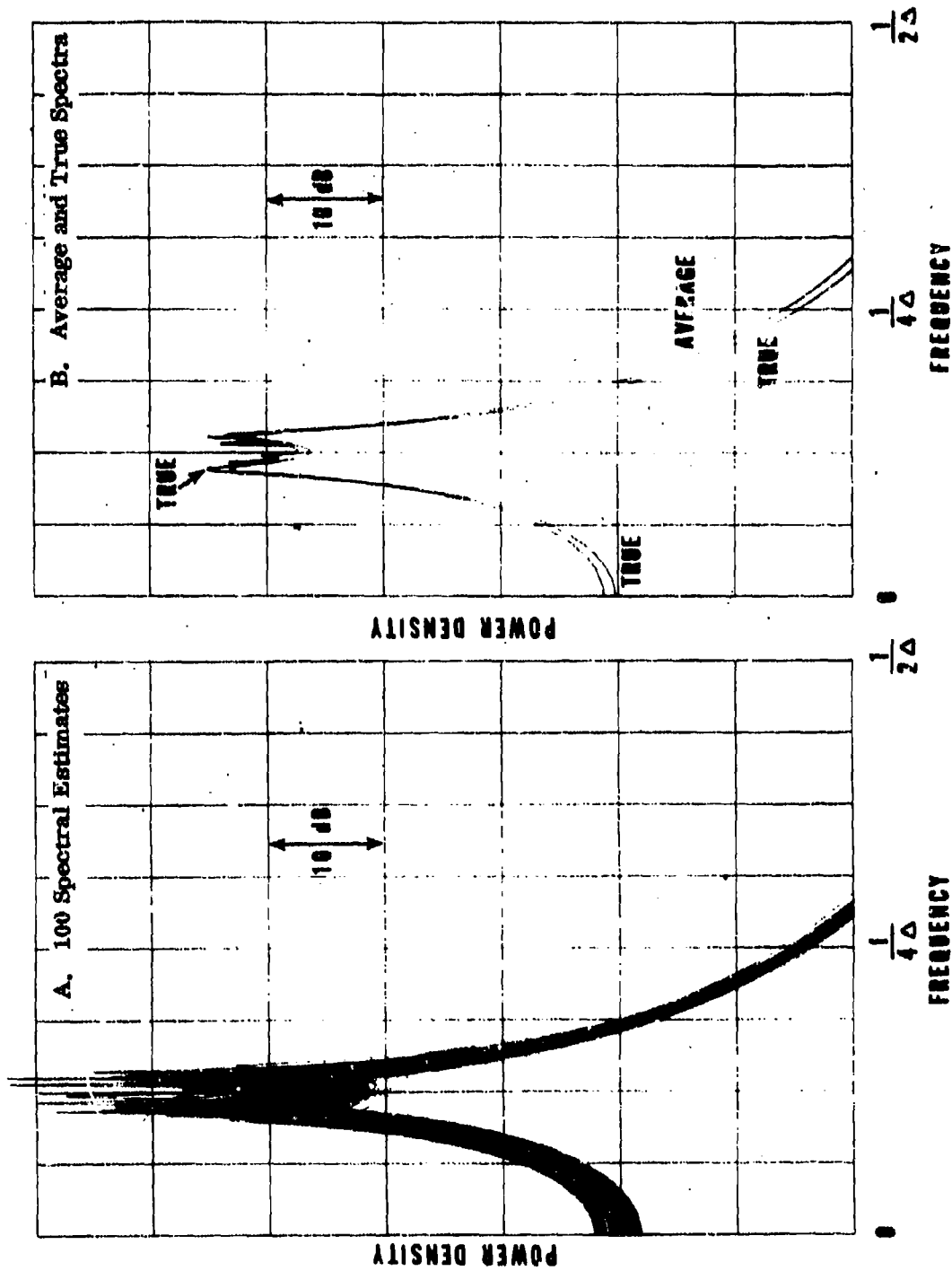namely, the forward and backward prediction technique and the Burg technique. The former technique tends to have less variability on the skirts, but has more spiky estimates near the peaks of the spectrum; the latter technique has very few spiky estimates. Both techniques have comparable resolution and bias.

Since the best choice of filter order, p, is not known a priori, it is necessary in practice to make several guesses at this parameter and compute some error criterion that indicates when to terminate the recursion. In particular, Akaike's Information Criterion (reference 22) is often adopted as a termination procedure; it takes the form (reference 1, equations (91) and (41) or reference 22, page 719)

$$AIC = \ell n \ (\text{Relative Error}) + \frac{2p}{N_e} \qquad (AIC \ (p = 0) = 0) \ , \quad (202)$$

where N_e is the "effective" number of data points, and is taken as $N-p$ (or $N-p-B_p$ for bad points) here, at the p-th stage. The value of p at which (202) is a minimum is taken as the best estimate of this parameter; however, criterion (202) is not absolute, and the user can adjust it to fit his application (reference 1, page 575). A wide range of values of p may have to be investigated if little is known about the true spectrum a priori; an upper bound on p is given by Akaike as $3N^{1/2}$ (Ibid).

One of the ramifications of this successive guessing procedure is that for the forward and backward prediction technique, a different $p \times p$ matrix $[S_{nm}]_1^p$ must be inverted (or an equivalent operation conducted) at each stage (see (140) and (138)) in order to determine the filter coefficients and minimum error, (141). Although the matrix terms can be updated according to

$$S_{nm}^{(p+1)} = \frac{N-p}{N-p-1} S_{nm}^{(p)} - \frac{x_{p+1-m} x_{p+1-n}^* + x_{N-p+n} x_{N-p+m}^*}{2(N-p-1)} \ , \quad (203)$$

in addition to the relations in (139), the size of the matrix $[S_{nm}]_1^p$ grows with p, and the solution of (140) can be a time-consuming procedure, if many large values of p must be investigated. This fact, coupled with the fact that this spectral estimation technique can yield spiky estimates and an unstable recur-

sion relation (149), leads to the conclusion that, of the methods considered, the Burg technique is the recommended procedure for spectral analysis of univariate processes. A comparison with the maximum likelihood technique (reference 23) is underway and will be documented in a future report.

The solution for the filter coefficients in the Burg technique is accomplished recursively as shown in subsection 4.7 and automatically progresses through successively larger values of p at which error measures (150) and (156) are readily calculated. There is, of course, the need to update the forward and backward residuals via (153), and the calculation of cross-gain g_p in (155), both of which take time to effect. But the effort required actually decreases as p increases, since fewer terms are involved in (153) and (155); in exchange, the stability of the estimates also decreases.

FORTRAN programs for the Burg technique, both with and without bad data points, are given in appendix J. Some representative execution times on the Univac 1108 for the computation of the filter coefficients (SUBROUTINE BURG) are given in table 3, where N is the number of data points and PMAX is the maximum order of filter considered. The times are approximately linearly proportional to N and PMAX. The execution time for the evaluation of the power density estimate itself is governed by the FFT technique employed to evaluate (167) (SUBROUTINE POWERS).

Table 3. Execution Times; No Bad Data Points

N	PMAX	Time (sec)
100	10	0.038
100	20	0.073
1000	10	0.33
1000	50	1.78
10000	50	17.9
10000	150	48.4

The presence of bad data points is easily accommodated in the Burg technique, as shown in subsection 4.7. If the bad data points are contiguous, the loss in stability of the estimates is not as great as when the bad data points are spaced. The worst possible locations of bad data points occur when the closest spacing is $\geq p + 1$, since each bad data point causes the loss of p + 1 valid error points. Interpolation of spaced bad data points has proven poorer than the technique utilized here (of ignoring bad points) when the spectral content of the input process extends fairly close to the Nyquist frequency $(2\Delta)^{-1}$. Since the exact extent of the input spectrum is unknown a priori, interpolation can be a damaging procedure in some cases.

The spectral estimation technique investigated here is particularly advantageous for short data segments, where other methods are inapplicable. For example, if a piece of equipment fails frequently, short disjointed pieces of data may be all that are available. Or if a process is nonstationary, it may be necessary to cut the total data record into small segments in each of which it is believed that conditions are substantially stationary. For longer data records, where standard FFT techniques can be applied, it has been recommended that both spectral estimation procedures be applied and the results plotted together to glean maximum information about the true spectrum (see reference 12). This seems particularly useful when some pure tones are present in the input data; the standard FFT technique is ideally suited for the analysis of pure tones or very narrowband components.

Appendix A

RECURSIVE SOLUTION

If we employ (52) in (46), there results

$$\sum_{k=0}^{p} R_{\ell - k} \frac{\alpha_k^*}{\alpha_0^*} = \frac{1}{c_{00}} \delta_{\ell 0}, \quad 0 \le \ell \le p.$$
(A-1)

Now define

$$\frac{\alpha_k^*}{\alpha_0^*} = -a_k^{(p)}, \quad 0 \le k \le p,$$
(A-2)

where the dependence of the coefficients on the order p in (31) is indicated explicitly. Then (A-1) becomes

$$\begin{bmatrix} R_0 & R_{-1} & \cdots & R_{-p} \\ R_1 & R_0 & & \\ \vdots & & \ddots & \\ R_p & & & R_0 \end{bmatrix} \begin{bmatrix} 1 \\ -a_1^{(p)} \\ \vdots \\ -a_p^{(p)} \end{bmatrix} = \begin{bmatrix} 1/c_{00}^{(p)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
(A-3)

where the matrix R is Hermitian and where we have also indicated that the real quantity c_{00} is dependent on p; see (47) and (51). Equation (A-3) constitutes p+1 linear equations in the p+1 unknowns $a_1^{(p)}, \ldots, a_p^{(p)}, 1/c_{00}^{(p)}$.

The solution to (A-3) can be obtained recursively as follows (see, for example, reference 11 or reference 24, appendix B):

$$a_1^{(1)} = R_1/R_0, \quad \frac{1}{c_{00}^{(1)}} = R_0 - R_{-1} a_1^{(1)} = R_0 \left(1 - \left| a_1^{(1)} \right|^2 \right);$$
(A-4)

for $p \geq 2$:

$$a_p^{(p)} = \frac{R_p - \sum_{k=1}^{p-1} a_k^{(p-1)} R_{p-k}}{R_0 - \sum_{k=1}^{p-1} a_k^{(p-1)} R_k^*} = \left[R_p - \sum_{k=1}^{p-1} a_k^{(p-1)} R_{p-k} \right] c_{oo}^{(p-1)}, \quad (A-5)$$

$$a_k^{(p)} = a_k^{(p-1)} - a_p^{(p)} a_{p-k}^{(p-1)^*}, \quad k = 1, 2, \ldots, p-1, \quad (A-6)$$

$$\frac{1}{c_{oo}^{(p)}} = R_0 - \sum_{k=1}^{p} a_k^{(p)} R_k^* = R_0 - a_p^{(p)} R_p^* - \sum_{k=1}^{p-1} a_k^{(p)} R_k^* = \frac{1}{c_{oo}^{(p-1)}} \left(1 - \left| a_p^{(p)} \right|^2 \right). \quad (A-7)$$

The last step in (A-7) is obtained by substituting (A-6) and employing (A-5). It is very important to notice from (A-6) that once $a_p^{(p)}$ is specified, **all** the p-th order filter coefficients can be calculated from (p-1)th order coefficients. The same is true of (A-7).

If we use (A-2) and (53), the maximum entropy spectrum in (55) can be expressed as

$$G_o(f) = \frac{\Delta / c_{oo}^{(p)}}{\left| \sum_{k=0}^{p} a_k^{(p)} \exp(-i2\pi f k \Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \quad (A-8)$$

The similarity in form to (14) will be complete when it is shown (in (67)) that $1/c_{oo}^{(p)}$ is the minimum value of the average magnitude-squared error for a p-th order predictive filter; therefore $c_{oo}^{(p)}$ must be positive for all p, for non-negative definite R. Equation (A-7) offers a recursive calculation of the average error; it can be started with $\frac{1}{c_{oo}^{(p)}} = R_0$. (In fact, (A-5) through (A-7) can be used for $p \geq 1$ when that starting value is used.)

A-2

Since $c_{00}^{(p)}$ must be positive for all p, (A-7) indicates that

$$\left| a_k^{(k)} \right| \leq 1 \text{ for } k = 1, 2, \ldots, p. \tag{A-9}$$

This is equivalent to having all the zeros of

$$\sum_{k=0}^{p} a_k^{(p)} z^{-k} , \tag{A-10}$$

(where the remaining coefficients are determined via (A-6)) inside the unit circle, O, in the complex z-plane; see reference 1, page 567. Therefore

$$B^{(p)}(z) = \sum_{k=0}^{p} a_k^{(p)} z^k \tag{A-11}$$

has no zeros inside O.

Appendix B

EVALUATION OF MAXIMUM ENTROPY

The optimum spectrum is given by (36) and (37). The maximum entropy then follows from (30) as

$$\text{Ent} = \Delta \int_{1/\Delta} df \, \ln G_o(f) = -\Delta \int_{1/\Delta} df \, [\ln \gamma(f) + \ln \gamma^*(f)] \equiv \mathcal{E}_1 + \mathcal{E}_2 \, . \tag{B-1}$$

Consider

$$\mathcal{E}_1 = -\Delta \int_{1/\Delta} df \, \ln \left\{ \sum_{k=0}^{p} \alpha_k \, \exp(i2\pi f k \Delta) \right\} . \tag{B-2}$$

Letting $z = \exp(i2\pi f \Delta)$ and using (38), (B-2) becomes

$$\mathcal{E}_1 = -\frac{1}{i2\pi} \oint \frac{dz}{z} \, \ln \, B(z) \, , \tag{B-3}$$

where $\oint$ denotes counterclockwise integration around the unit circle, O, in the complex z-plane. Now

$$B(z) = \sum_{k=0}^{p} \alpha_k \, z^k = \alpha_p \prod_{k=1}^{p} (z - o_k) \tag{B-4}$$

where, from (A-14), zero locations $\{o_k\}$ satisfy

$$|o_k| > 1, \text{ all } k; \tag{B-5}$$

that is, all the zeros of $B(z)$ lie outside O. (There can be multiple-order zeros in (B-4).) Also assume $p \geq 1$ for now. Then (B-3) can be expressed as

$$\mathcal{E}_1 = -\frac{1}{12\pi} \oint \frac{dz}{z} \left[\ln \alpha_p + \sum_{k=1}^{p} \ln (z - o_k) \right]$$

$$= -\left[\ln \alpha_p + \sum_{k=1}^{p} \frac{1}{12\pi} \oint \frac{dz}{z} \ln (z - o_k) \right]. \qquad (B-6)$$

But

$$\ln (z - o_k) = \ln (- o_k) + \ln (1 - \frac{z}{o_k})$$

$$= \ln (- o_k) - \frac{z}{o_k} - \left(\frac{z}{o_k}\right)^2 - \ldots \text{ for } \left| \frac{z}{o_k} \right| < 1 ; \qquad (B-7)$$

that is, expansion (B-7) converges for $|z| < |o_k|$. But since $|o_k| > 1$, the region of integration in (B-6) remains in the convergence region of (B-7). Therefore, the integral in (B-6) is

$$\frac{1}{12\pi} \oint \frac{dz}{z} \ln (z - o_k) = \frac{1}{12\pi} \oint \frac{dz}{z} \left\{ \ln (- o_k) - \frac{z}{o_k} - \left(\frac{z}{o_k}\right)^2 - \ldots \right\} = \ln (- o_k). (B-8)$$

Then from (B-6) and (B-4)

$$\mathcal{E}_1 = -\left[\ln \alpha_p + \sum_{k=1}^{p} \ln (- o_k) \right] = -\ln \left[\alpha_p \prod_{k=1}^{p} (- o_k) \right]$$

$$= -\ln B(0) = -\ln \alpha_o. \qquad (B-9)$$

And from (B-1) and (B-4)

$$\mathcal{E}_2 = -\Delta \int_{1/\Delta} df \ln \gamma^*(f) = -\Delta \int_{1/\Delta} df \ln \left\{ \sum_{k=0}^{p} \alpha_k^* \exp(12\pi f k \Delta) \right\}$$

$$= -\frac{1}{12\pi} \oint \frac{dz}{z} \ln\left\{ \sum_{k=0}^{p} \alpha_k^* z^{-k} \right\} = -\frac{1}{12\pi} \oint \frac{dz}{z} \ln B^*\!\left(\frac{1}{z^*}\right)$$

$$= -\frac{1}{12\pi} \oint \frac{dz}{z} \ln\left\{ \alpha_p^* \prod_{k=1}^{p} \left(\frac{1}{z} - o_k^*\right) \right\}$$

$$= -\frac{1}{12\pi} \oint \frac{dz}{z} \left[\ln \alpha_p^* + \sum_{k=1}^{p} \ln\!\left(\frac{1}{z} - o_k^*\right) \right]$$

$$= -\left[\ln \alpha_p^* + \sum_{k=1}^{p} \frac{1}{12\pi} \oint \frac{dz}{z} \ln\!\left(\frac{1}{z} - o_k^*\right) \right]. \qquad \text{(B-10)}$$

Now

$$\ln\!\left(\frac{1}{z} - o_k^*\right) = \ln\left(- o_k^*\right) + \ln\left(1 - \frac{1}{o_k^* z}\right)$$

$$= \ln\left(- o_k^*\right) - \frac{1}{o_k^* z} - \left(\frac{1}{o_k^* z}\right)^2 - \dots \text{ for } \left|\frac{1}{o_k^* z}\right| < 1; \qquad \text{(B-11)}$$

that is, expansion (B-11) converges for $|z| > \frac{1}{|o_k^*|}$. But since $|o_k| > 1$, the region of integration in (B-10) remains in the convergence region of (B-11). Therefore, the integral in (B-10) is

$$\frac{1}{12\pi} \oint \frac{dz}{z} \ln\!\left(\frac{1}{z} - o_k^*\right) = \frac{1}{12\pi} \oint \frac{dz}{z} \left\{ \ln\left(- o_k^*\right) - \frac{1}{o_k^* z} - \left(\frac{1}{o_k^* z}\right)^2 - \dots \right\} = \ln\left(- o_k^*\right). \qquad \text{(B-12)}$$

Then from (B-10) and (B-4)

$$\mathcal{E}_2 = -\left[\ln \alpha_p^* + \sum_{k=1}^{p} \ln\left(- o_k^*\right) \right] = -\ln\left[\alpha_p^* \prod_{k=1}^{p} \left(- o_k^*\right) \right]$$

$$= -\ln B^*(0) = -\ln \alpha_0^*. \qquad \text{(B-13)}$$

Combining (B-9) and (B-13) in (B-1), there follows for the maximum entropy

$$\text{Ent} = -\ln |\alpha_0|^2 = \ln (\Delta/c_{00}), \qquad (B-14)$$

where we have also employed (52). (For $p = 0$, a separate derivation yields (B-14) also.) Recall from (51) that c_{00} is the upper-left corner element of R^{-1}, where R is defined by (47).

Appendix C

IMPLICATIONS OF ASSUMPTION OF WHITE SPECTRUM
FOR MINIMUM ERROR; KNOWN CORRELATION

We define the crosscorrelation function between minimum error $\tilde{\tau}$ and input x in figure 1 as

$$C_\ell = \overline{\tilde{\tau}_k \, x^*_{k-\ell}}, \quad \text{all } \ell . \tag{C-1}$$

Substituting (59) and utilizing (1), this becomes

$$C_\ell = \sum_{n=0}^{p} \tilde{a}_n R_{\ell-n}, \quad \text{all } \ell . \tag{C-2}$$

Now from (64) and (65), we can express

$$R\,\tilde{a} = -\frac{1}{c_{00}}\,\delta . \tag{C-3}$$

Thus, (C-2) immediately yields

$$C_\ell = \left\{ \begin{array}{l} -1/c_{00}, \; \ell = 0 \\[2ex] 0, \; 1 \le \ell \le p \end{array} \right\} ; \tag{C-4}$$

that is, minimum error value $\tilde{\tau}_k$ is uncorrelated with the past p inputs x_{k-1}, ..., x_{k-p}.

Now using (59) and (C-1), the autocorrelation function of the minimum error is

$$E_\ell = \overline{\tilde{\tau}_k \, \tilde{\tau}^*_{k-\ell}} = \sum_{n=0}^{p} \tilde{a}^*_n \overline{\tilde{\tau}_k \, x^*_{k-\ell-n}} = \sum_{n=0}^{p} \tilde{a}^*_n C_{\ell+n}, \quad \text{all } \ell . \tag{C-5}$$

In particular, using (C-4),

$$E_1 = \tilde{a}_p^* \, C_{p+1} \, . \tag{C-6}$$

But from (C-2) and (66),

$$C_{p+1} = \sum_{n=0}^{p} \tilde{a}_n \, R_{p+1-n} = - R_{p+1} \quad \sum_{n=1}^{p} \frac{c_{nc}}{c_{oo}} \, R_{p+1-n} \, . \tag{C-7}$$

Therefore, assuming $E_1 = 0$ is equivalent to assuming $C_{p+1} = 0$, (that is, minimum error $\tilde{r}_k$ uncorrelated with input x_{k-p-1}), which in turn is equivalent to requiring

$$R_{p+1} = - \sum_{n=1}^{p} \frac{c_{no}}{c_{oo}} \, R_{p+1-n} = \sum_{n=1}^{p} \tilde{a}_n \, R_{p+1-n} \, . \tag{C-8}$$

This relation, which may not be true for the actual process $\{x_k\}$, is a direct result of assumption (70); the quantity R_{p+1} in (C-8) is really an approximation to the true (unknown) correlation value.

Next from (C-5),

$$E_2 = \tilde{a}_{p-1}^* \, C_{p+1} + \tilde{a}_p^* \, C_{p+2} \, . \tag{C-9}$$

Assuming $E_2 = 0$ (in addition to $E_1 = 0$) is equivalent to also assuming $C_{p+2} = 0$, which in turn from (C-2) and (66) requires that we approximate according to

$$R_{p+2} = - \sum_{n=1}^{p} \frac{c_{no}}{c_{oo}} \, R_{p+2-n} = \sum_{n=1}^{p} \tilde{a}_n \, R_{p+2-n} \, . \tag{C-10}$$

Continuing in this way, it follows that assuming white noise for $\{\tilde{r}_k\}$, that is, assuming

$$E_\ell = 0 \text{ for } \ell \geq 1 \, , \tag{C-11}$$

is equivalent to assuming that $C_\ell = 0$ for $\ell \geq p + 1$; that is, the minimum error is uncorrelated with <u>all</u> past inputs. There follows the approximations

$$R_\ell = - \sum_{n=1}^{p} \frac{c_{no}}{c_{oo}} R_{\ell-n} = \sum_{n=1}^{p} \tilde{a}_n R_{\ell-n} \text{ for } \ell \geq p+1. \qquad (C-12)$$

This recursion relation (starting with known values R_1, R_2, ..., R_p) can be considered to be an extrapolation of the known correlation values into regions where they are unknown.

If we augment (C-12) according to

$$R_{-\ell} = R_\ell^* \text{ for } \ell \geq p+1 , \qquad (C-13)$$

then it can be shown that the spectrum defined by

$$\Delta \sum_{\ell=-\infty}^{\infty} R_\ell \exp(-i2\pi f\ell\Delta) = \frac{\Delta/c_{oo}}{\left| 1 - \sum_{n=1}^{p} \tilde{a}_n \exp(-i2\pi fn\Delta) \right|^2} , \qquad (C-14)$$

which is identical to (71). The transform in (C-14) converges if $|R_\ell|$ decays with increasing $|\ell|$, that is, if $B(z)$ of (56) has no zeros inside O.

Appendix D

STABILITY OF RECURSION RELATION

The recursion relation for approximated correlation value $\widetilde{R}_\ell$, is given in (C-12) and (75) as

$$R_\ell = \sum_{k=1}^{p} \widetilde{a}_k \, R_{\ell-k} \text{ for } \ell \geq p + 1 . \tag{D-1}$$

Therefore,

$$U(z) \equiv \sum_{\ell=p+1}^{\infty} R_\ell \, z^{-\ell} = \sum_{k=1}^{p} \widetilde{a}_k \, z^{-k} \sum_{\ell=p+1}^{\infty} R_{\ell-k} \, z^{-\ell+k} . \tag{D-2}$$

But

$$\sum_{\ell=p+1}^{\infty} R_{\ell-k} \, z^{-\ell+k} = \sum_{j=p+1-k}^{\infty} R_j \, z^{-j} = \sum_{j=p+1-k}^{p} R_j \, z^{-j}$$

$$+ \sum_{j=p+1}^{\infty} R_j \, z^{-j} \equiv V_k(z) + U(z) , \tag{D-3}$$

where

$$V_k(z) = R_{p+1-k} \, z^{-(p+1-k)} + V_{k-1}(z) , \quad k \geq 2 ; \quad V_1(z) = R_p \, z^{-p} . \tag{D-4}$$

$V_k(z)$ involves the starting values $R_{p+1-k}, \ldots, R_p$ for $1 \leq k \leq p$. Employment of (D-3) in (D-2) yields

$$U(z) = \sum_{k=1}^{p} \widetilde{a}_k \, z^{-k} \, V_k(z) + U(z) \sum_{k=1}^{p} \widetilde{a}_k \, z^{-k} , \tag{D-5}$$

D-1

or

$$U(z) = \frac{\sum\limits_{k=1}^{p} \tilde{a}_k \, z^{-k} \, V_k(z)}{1 - \sum\limits_{k=1}^{p} \tilde{a}_k \, z^{-k}} = - \frac{\sum\limits_{k=1}^{p} c_{ko} \, z^{-k} \, V_k(z)}{\sum\limits_{k=0}^{p} c_{ko} \, z^{-k}} , \qquad (D-6)$$

where we have utilized (66). In order that recursion (D-1) be stable, the denominator of (D-6) must possess all its zeros within the unit circle O in the complex z-plane. Therefore, B(z) of (56) must possess all its zeros outside O if recursion (D-1) is to be stable. This is guaranteed by the results in (A-9) et seq.

Appendix E

IMPLICATIONS OF ASSUMPTION OF WHITE SPECTRUM;
UNKNOWN CORRELATION

The minimum error sequence is given by (96) and (101) as

$$\tau_k = \sum_{n=0}^{p} \tilde{a}_n x_{k-n}, \text{ all } k. \qquad (E-1)$$

The sample autocorrelation of $\{\tilde{\tau}_k\}$ is defined here as

$$F_\ell \equiv \frac{1}{N} \sum_k \tau_k \tau_{k-\ell}^* = \sum_{m,n=0}^{p} \tilde{a}_m \tilde{a}_n^* S_{\ell+n-m} \qquad (E-2)$$

using (E-1) and (98). The sample spectrum of $\{\tilde{\tau}_k\}$ is defined here as

$$H_{\tilde{\tau}}(f) \equiv \Delta \sum_{\ell=-\infty}^{\infty} F_\ell \exp(-i2\pi f \ell \Delta) = H_x(f) \left| A(f) \right|^2, \ |f| < \frac{1}{2\Delta}, \qquad (E-3)$$

where we have employed (E-2) and (107) and defined the sample spectrum of $\{x_k\}$ as

$$H_x(f) \equiv \Delta \sum_{\ell=-\infty}^{\infty} S_\ell \exp(-i2\pi f \ell \Delta), \ |f| < \frac{1}{2\Delta}. \qquad (E-4)$$

Therefore, (E-3) yields

$$H_x(f) = \frac{H_{\tilde{\tau}}(f)}{\left| A(f) \right|^2}, \ |f| < \frac{1}{2\Delta}. \qquad (E-5)$$

Now we will assume that the sample spectrum of $\{\tilde{\tau}_k\}$ is white; that is, we set

$$\hat{H}_{\tilde{\tau}}(f) \equiv K\Delta, \quad |f| < \frac{1}{2\Delta}, \tag{E-6}$$

where K is a constant. We then adopt an estimate of the sample spectrum of sequence $\{x_n\}_1^N$ according to

$$\hat{H}_x(f) \equiv \frac{\hat{H}_{\tilde{\tau}}(f)}{|A(f)|^2} = \frac{K\Delta}{\left| \sum_{n=0}^{p} \tilde{a}_n \exp(-i2\pi f n \Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}, \tag{E-7}$$

and adopt a scaled version of this quantity as a spectral estimate of process $\{x_n\}$:

$$\hat{G}_x(f) \equiv \frac{\Delta}{\left| \sum_{n=0}^{p} \tilde{a}_n \exp(-i2\pi f n \Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \tag{E-8}$$

The white assumption in (E-6) forces us to assume that

$$F_\ell = 0 \text{ for } \ell \neq 0, \tag{E-9}$$

as (E-3) shows. In order to see what this implies, we utilize the definition of the sample crosscorrelation in (109), along with (96) and (98), to obtain

$$D_\ell \equiv \frac{1}{N} \sum_k \tilde{\tau}_k x_{k-\ell}^* = \sum_{n=0}^{p} a_n S_{\ell-n}, \text{ all } \ell. \tag{E-10}$$

Use of (101) then shows that

$$D_\ell = 0 \text{ for } 1 \leq \ell \leq p. \tag{E-11}$$

Meanwhile, the sample autocorrelation in (E-2) can be written in the form

$$F_\ell = \sum_{n=0}^{p} \tilde{a}_n^* D_{\ell+n}, \text{ all } \ell, \tag{E-12}$$

upon employment of (E-10). There immediately follows from (E-9), (E-11), and (E-12)

$$F_1 = \tilde{a}_p^* \, D_{p+1} = 0 . \tag{E-13}$$

But then (E-9) and (E-10) indicate that

$$S_{p+1} = \sum_{n=1}^{p} \tilde{a}_n \, S_{p+1-n} , \tag{E-14}$$

where $\{\tilde{a}_n\}_1^p$ are the solutions of (102). But relation (E-14) may not be true for the quantity S_{p+1} actually obtained from data $\{x_n\}_1^N$ via (98). Thus, assumption $F_1 = 0$ is forcing us to assume that S_{p+1} can be obtained via (E-14) and (102), when $\{S_\ell\}_{-p}^p$ are obtained from (98).

Next from (E-12) and (E-11),

$$F_2 = \tilde{a}_{p-1}^* \, D_{p+1} + \tilde{a}_p^* \, D_{p+2} . \tag{E-15}$$

Assuming $F_2 = 0$ (in addition to $F_1 = 0$) is equivalent to also assuming $D_{p+2} = 0$, which in turn from (E-10) requires that

$$S_{p+2} = \sum_{n=1}^{p} \tilde{a}_n \, S_{p+2-n} . \tag{E-16}$$

Continuing in this way, it follows that assuming

$$F_\ell = 0 \quad \text{for} \quad \ell \geq 1 \tag{E-17}$$

is equivalent to assuming $D_\ell = 0$ for $\ell \geq p+1$; that is, the minimum-error sequence is uncorrelated (on a single member function basis) with all past inputs. There follows the estimates

$$S_\ell = \sum_{n=1}^{p} \tilde{a}_n \, S_{\ell-n} , \quad \ell \geq p+1 . \tag{E-18}$$

Stability is discussed in (111) et seq.

Appendix F

BOUND ON CROSS-GAIN

The value of the cross-gain g_p in (155) can be written as

$$g_p = \frac{\displaystyle\sum_{n=p+1}^{N} f_n^{(p-1)} b_{n-1}^{(p-1)*}}{\left(\displaystyle\sum_{n=p+1}^{N} \left|f_n^{(p-1)}\right|^2 \sum_{n=p+1}^{N} \left|b_{n-1}^{(p-1)}\right|^2\right)^{1/2}} \cdot \frac{2\left(\displaystyle\sum_{n=p+1}^{N} \left|f_n^{(p-1)}\right|^2 \sum_{n=p+1}^{N} \left|b_{n-1}^{(p-1)}\right|^2\right)^{1/2}}{\displaystyle\sum_{n=p+1}^{N} \left(\left|f_n^{(p-1)}\right|^2 + \left|b_{n-1}^{(p-1)}\right|^2\right)} \tag{F-1}$$

The first factor in (F-1) is of the form of a correlation coefficient of the (p-1)-th order forward and backward sequences and can never exceed unity in magnitude (by Schwarz's inequality). The second factor in (F-1) is almost always very close to 1: let the pair of sums

$$\left\{\sum_{n=p+1}^{N} \left|f_n^{(p-1)}\right|^2 \text{ and } \sum_{n=p+1}^{N} \left|b_{n-1}^{(p-1)}\right|^2\right\} = \left\{A \text{ and } A(1+r)\right\}, \tag{F-2}$$

where $r \geq 0$ without loss of generality. The second factor in (F-1) then equals $\dfrac{(1+r)^{1/2}}{1+r/2}$, which is never larger than 1 and is tabulated below. Thus, g_p in (F-1) is virtually identical to the correlation coefficient of the forward and backward sequences, since r is near zero with high probability.

Table F.1 Second Factor in (F-1).

r	0	.1	.2	.3	.4	.5
$\dfrac{(1+r)^{1/2}}{1+r/2}$	1	.999	.996	.991	.986	.980

Appendix G

CLOSENESS OF ERROR MEASURES

Two possible error measures for the Burg technique were presented in (150) and (156). For $p = 0$, employing (154) and (152),

$$F^{(0)} = \frac{1}{N} \sum_{n=1}^{N} \left| x_n \right|^2 . \tag{G-1}$$

Comparing this result with (151), we find

$$F^{(0)} = P^{(0)} . \tag{G-2}$$

Thus, the two error measures are identical for $p = 0$.

Next from (150) and (151)

$$P^{(1)} = P^{(0)} \left(1 - \left| a_1^{(1)} \right|^2 \right) = \left(1 - \left| a_1^{(1)} \right|^2 \right) \frac{1}{N} \sum_{n=1}^{N} \left| x_n \right|^2 , \tag{G-3}$$

whereas from (156), (160), (155), and (152),

$$F_o^{(1)} = \left(1 - \left| a_1^{(1)} \right|^2 \right) \frac{\text{Den}(1)}{2(N-1)}$$

$$= \left(1 - \left| a_1^{(1)} \right|^2 \right) \frac{1}{2(N-1)} \sum_{n=2}^{N} \left(\left| f_n^{(0)} \right|^2 + \left| b_{n-1}^{(0)} \right|^2 \right)$$

$$= \left(1 - \left| a_1^{(1)} \right|^2 \right) \frac{\frac{1}{2}\left| x_1 \right|^2 + \left| x_2 \right|^2 + \left| x_3 \right|^2 + \ldots + \left| x_{N-2} \right|^2 + \left| x_{N-1} \right|^2 + \frac{1}{2}\left| x_N \right|^2}{N-1} . \tag{G-4}$$

But now reference to (151) and (G-3) reveals that, for N-1 large,

$$F_o^{(1)} \cong \left(1 - \left|a_1^{(1)}\right|^2\right) P^{(0)} = P^{(1)} . \qquad (G-5)$$

Continuing with (150),

$$P^{(2)} = P^{(1)} \left(1 - \left|a_2^{(2)}\right|^2\right). \qquad (G-6)$$

And (156), (160), and (155) combine to yield

$$F_o^{(2)} = \left(1 - \left|a_2^{(2)}\right|^2\right) \frac{\mathrm{Den}(2)}{2(N-2)} = \left(1 - \left|a_2^{(2)}\right|^2\right) \frac{1}{2(N-2)} \sum_{n=3}^{N} \left(\left|f_n^{(1)}\right|^2 + \left|b_{n-1}^{(1)}\right|^2\right). \qquad (G-7)$$

But from (154),

$$F_o^{(1)} = \frac{1}{2(N-1)} \sum_{n=2}^{N} \left(\left|f_n^{(1)}\right|^2 + \left|b_{n-1}^{(1)}\right|^2\right). \qquad (G-8)$$

Comparing (G-7) and (G-8), we see that, for N-2 large,

$$F_o^{(2)} \cong \left(1 - \left|a_2^{(2)}\right|^2\right) F_o^{(1)} . \qquad (G-9)$$

Then employing (G-5) and (G-6), we have

$$F_o^{(2)} \cong \left(1 - \left|a_2^{(2)}\right|^2\right) P^{(1)} = P^{(2)} , \qquad (G-10)$$

which is the desired relationship. In general, for no bad data points, we have

$$F_o^{(p)} \cong P^{(p)} \text{ for } N-p \text{ large} . \qquad (G-11)$$

Numerical computations have borne this result out, with the two quantities not having any ordered relationship; that is, either quantity can be larger (or smaller) at different stages, p. (G-9) generalizes to

$$F_o^{(p)} \cong \left(1 - \left|a_p^{(p)}\right|^2\right) F_o^{(p-1)} \text{ for } N-p \text{ large} . \qquad (G-12)$$

Appendix H

SCALE FACTORS IN SPECTRAL ESTIMATES

Instead of using a unity value for the average minimum error or residual power in the numberator of (167), we could use the value given by (156). Then our spectral estimate would be

$$\hat{G}_x(f) = \frac{\Delta F_o^{(p)}}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \qquad (H-1)$$

An alternative approach is to use an arbitrary scale factor K and choose it so that the area under the spectral estimate is equal to the sample power (151), as suggested under (108); that is, set

$$\hat{G}_x(f) = \frac{\Delta K}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}, \qquad (H-2)$$

and force

$$\int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} df\, \hat{G}_x(f) = P^{(0)} = \frac{1}{N} \sum_{n=1}^{N} |x_n|^2. \qquad (H-3)$$

Substituting (H-2) in (H-3), and using (159), we have

$$P^{(0)} = \int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} df\, \frac{\Delta K}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2} = \frac{K}{\prod_{m=1}^{p} \left\{ 1 - \left| a_m^{(m)} \right|^2 \right\}}. \qquad (H-4)$$

The last step in (H-4) is proven as follows: from (A-8) and (29), we know that

$$\int_{-\frac{1}{2\Delta}}^{\frac{1}{2\Delta}} df \frac{\Delta}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2} = R_o c_{oo}^{(p)} . \qquad (H-5)$$

But from (A-7),

$$R_o c_{oo}^{(p)} = \frac{R_o c_{oo}^{(p-1)}}{1 - \left| a_p^{(p)} \right|^2} = \frac{1}{\prod_{m=1}^{p} \left\{ 1 - \left| a_m^{(m)} \right|^2 \right\}} , \qquad (H-6)$$

where we have employed $R_o c_{oo}^{(0)} = 1$. The relationship in (H-4) holds when the filter coefficients are determined via (148).

Therefore, (H-4) yields, with the aid of (150).

$$K = P^{(0)} \prod_{m=1}^{p} \left\{ 1 - \left| a_m^{(m)} \right|^2 \right\} = P^{(p)} , \qquad (H-7)$$

and the estimate (H-2) becomes

$$\hat{G}_x(f) = \frac{\Delta P^{(p)}}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2} , \quad |f| < \frac{1}{2\Delta} . \qquad (H-8)$$

The very close similarity of values between the alternatives (H-1) and (H-8) is made evident by the results of appendix G, in particular (G-11). Thus, there is virtually no difference between estimates (H-1) and (H-8), for no bad data points.

Appendix I

BIASEDNESS OF BURG'S CORRELATION ESTIMATE

For the Burg technique with $p = 1$, $N = 3$, we find from (162) and (144) that (for real data)

$$\hat{R}_1 = \frac{2}{3} \frac{x_2 \left(x_1 + x_3\right)\left(x_1^2 + x_2^2 + x_3^2\right)}{x_1^2 + 2x_2^2 + x_3^2} \qquad (I-1)$$

The mean of this random variable depends on more than just $\overline{x_2 x_1}$ $(= \overline{x_3 x_2})$; in fact, it depends on the third-order joint density of (x_1, x_2, x_3). As an example, let

$$x_1 = u, \ x_2 = \pm \frac{1}{\sqrt{2}} (u + v), \ x_3 = v, \qquad (I-2)$$

where u and v are independent, zero-mean, unit-variance, Gaussian random variables. Then x_1, x_2, x_3 are zero-mean, unit-variance, Gaussian random variables with

$$\overline{x_2 \, x_1} = \overline{x_3 \, x_2} = \pm \frac{1}{\sqrt{2}}, \ \overline{x_3 \, x_1} = 0. \qquad (I-3)$$

Employing (I-2) in (I-1), we obtain

$$\hat{R}_1 = \pm \frac{1}{\sqrt{2}} \ \frac{1}{6} \ \frac{(u + v)^2 (3u^2 + 2uv + 3v^2)}{u^2 + uv + v^2}. \qquad (I-4)$$

Therefore,

$$\overline{\hat{R}_1} = \pm \frac{1}{\sqrt{2}} \ \frac{1}{6} \ \frac{1}{2\pi} \iint\limits_{-\infty}^{\infty} du \, dv \, \exp\left(-\frac{u^2 + v^2}{2}\right) \frac{(u + v)^2 (3u^2 + 2uv + 3v^2)}{u^2 + uv + v^2}$$

$$= \pm \frac{1}{\sqrt{2}} \ \frac{1}{6} \ \frac{1}{2\pi} \int_0^\infty dr \ r^3 \ \exp(-r^2/2) \int_{-\pi}^{\pi} d\theta \ \frac{(C + S)^2 (3C^2 + 2CS + 3S^2)}{C^2 + CS + S^2}, \quad (I-5)$$

where we have changed to polar coordinates and let $C \equiv \cos\theta$, $S \equiv \sin\theta$. The integral on r in (I-5) is 2, and the integral on θ is $4\pi\left(2 - \frac{1}{\sqrt{3}}\right)$. Therefore,

$$\overline{\widehat{R}_1} = \pm \frac{1}{\sqrt{2}} \ \frac{12 - 2\sqrt{3}}{9} = \pm \frac{1}{\sqrt{2}} \ (.9484), \quad (I-6)$$

which is not equal to

$$\overline{x_2 \, x_1} = \overline{x_3 \, x_2} = \pm \frac{1}{\sqrt{2}} . \quad (I-7)$$

Appendix J

FORTRAN PROGRAMS

The programs in this appendix are written for real data, but may be readily generalized to complex data by means of the general equations in the main text. From (H-8) and (H-7), for real data, the spectral estimate is given by

$$\hat{G}_x(f) = \frac{\Delta \prod_{k=1}^{p} \left\{ 1 - a_k^{(k)2} \right\} P^{(0)}}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi fk\Delta) \right|^2}, \quad |f| < \frac{1}{2\Delta}. \qquad (J-1)$$

Let frequency increment

$$\Delta_f \equiv \frac{1}{J\Delta} = \frac{1}{J/2} \frac{1}{2\Delta} = \frac{f_N}{J/2}, \qquad (J-2)$$

where f_N is the Nyquist frequency, and J is an integer. Then, using (H-3) and the real behavior of the data,

$$P^{(0)} = 2 \int_0^{\frac{1}{2\Delta}} df\ \hat{G}_x(f) \cong 2\Delta P^{(0)} \prod_{k=1}^{p} \left\{ 1 - a_k^{(k)2} \right\} \sum_{m=0}^{J/2} \Delta_f \cdot$$

$$\frac{\epsilon_m}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi m\Delta_f k\Delta) \right|^2}$$

$$= \frac{2}{J} P^{(0)} \prod_{k=1}^{p} \left\{ 1 - a_k^{(k)2} \right\} \sum_{m=0}^{J/2} \frac{\epsilon_m}{\left| 1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi mk/J) \right|^2} \equiv \sum_{m=0}^{J/2} \epsilon_m P_m,$$

$$(J-3)$$

J-1

where $\{\epsilon_m\}$ is a set of integration weights (for example, trapezoidal). So we can compute (independent of time increment Δ) the quantity

$$\frac{P_m}{P^{(0)}} = \frac{2}{J} \prod_{k=1}^{p} \left\{1 - a_k^{(k)^2}\right\} \frac{1}{\left|1 - \sum_{k=1}^{p} a_k^{(p)} \exp(-i2\pi mk/J)\right|^2} \quad \text{for } 0 \le m \le \frac{J}{2}, \tag{J-4}$$

which represents the fractional power in the frequency band

$$\left(\frac{m - \frac{1}{2}}{J\Delta}, \frac{m + \frac{1}{2}}{J\Delta}\right); \tag{J-5}$$

that is,

$$\sum_{m=0}^{J/2} \epsilon_m \frac{P_m}{P^{(0)}} \cong 1 \tag{J-6}$$

if estimate $\hat{G}_x(f)$ in (J-1) has been sampled finely enough (that is, large J). The denominator of (J-4) is recognized as a J-point FFT of p+1 nonzero numbers; hence, J should be chosen as a power of 2 for speed purposes. The programs below yield the fraction of power in frequency bands of width $(J\Delta)^{-1}$, if J is an integer large enough that the spectral estimate (167) or (H-8) is adequately sampled to keep track of its peaks.

NO BAD DATA POINTS (SUBSECTION 4.7)

The data generation is accomplished via function IRAND, which generates integers uniformly distributed over $(0, 2^{35}-1)$; by RAND, which generates numbers uniformly distributed over $(0,1)$; and by TINORM, which generates zero-mean unit-variance Gaussian variables. The FFT used below is that presented in reference 25.

```
C   SPECTRAL ESTIMATION     USER: CHANGE LINE 13 AND REPLACE LINES 17-31
C   N = NUMBER OF DATA POINTS
C   X(1),....,X(N) = INPUT DATA
C   PMAX = MAXIMUM ORDER OF FILTER
C   PBEST = BEST ORDER OF FILTER
C   A(1),....,A(PBEST) = PREDICTIVE FILTER COEFFICIENTS
C   PROD = PRODUCT(1-A(P)**2) FOR P=1 TO PBEST
C   RHO(1),....,RHO(PMAX) = NORMALIZED CORRELATION COEFFICIENTS
C   J = SIZE OF FFT   (MUST BE A POWER OF 2)
C   XX(1),....,XX(J/2+1) = FRACTIONAL POWERS, FROM DC TO NYQUIST FREQUENCY
C   CO(1),....,CO(J/4+1) = QUARTER COSINE TABLE
C   Y AND YY ARE REQUIRED AUXILIARY ARRAYS
        PARAMETER N= 100, PMAX=10, J=2048, J41=J/4+1
        INTEGER PBEST
        DIMENSION X(N),Y(N),A(PMAX),RHO(PMAX),XX(J),YY(J),CO(J41)
C   INPUT DATA IN X(1),....,X(N)
        DEFINE IRAND=I*5**15+((1-SIGN(1,I*5**15))/2)*34359738367
        DEFINE RAND=FLOAT(I)/34359738367.
        I=5281
        NSTART=N+400       G WILL DISCARD INITIAL 400 POINTS
        XX(1)=0.
        XX(2)=0.
        XX(3)=0.
        XX(4)=0.
        DO 11 L=5,NSTART
        I=IRAND
        XX(L)=2.7607*XX(L-1)-3.8106*XX(L-2)+2.6535*XX(L-3)-
       +0.9238*XX(L-4)+TINORM(RAND,$11)
11      CONTINUE
        DO 12 I=1,N
12      X(I)=XX(I+NSTART-N)
        PRINT 1,
1       FORMAT(/' INPUT DATA!')
        PRINT 4, (X(I),I=1,N)
C   EVALUATE PREDICTIVE FILTER COEFFICIENTS
        CALL BURG(N,PMAX,X,Y,PBEST,A,PROD,RHO)
        PRINT 9, X(N)
9       FORMAT(/' MEAN =',E14.8)
        PRINT 10, Y(N)
10      FORMAT(' STANDARD DEVIATION =',E13.8)
        PRINT 2, PBEST
2       FORMAT(/' PBEST =',I3)
        PRINT 3,
3       FORMAT(/' PREDICTIVE FILTER COEFFICIENTS!')
        PRINT 4, (A(I),I=1,PBEST)
4       FORMAT(5E20.8)
        PRINT 5, PROD
5       FORMAT(/' PRODUCT(1-A(P)**2) =',E13.8)
        PRINT 6,
6       FORMAT(/' NORMALIZED CORRELATION COEFFICIENTS!')
        PRINT 4, (RHO(I),I=1,PMAX)
        CALL QTRCOS(CO,J)
C   EVALUATE FRACTIONAL POWERS
        CALL POWERS(PBEST,A,PROD,J,XX,YY,CO)
        PRINT 7,
7       FORMAT(/' FRACTIONAL POWERS!')
        L=J/2+1
        PRINT 8, (XX(I),I=1,L)
8       FORMAT(2X,10E13.6)
        END
```

```
      SUBROUTINE BURG(N,PMAX,X,Y,PBEST,A,PROD,RHO)      G 2 FEB 1976
C  THIS SUBROUTINE COMPUTES THE PREDICTIVE FILTER COEFFICIENTS
C  N = NUMBER OF DATA POINTS; INTEGER INPUT
C  PMAX = MAXIMUM ORDER OF FILTER; INTEGER INPUT
C  X(1),X(2),....,X(N) = DATA ARRAY ON INPUT; ALTERED ON OUTPUT
C  ON OUTPUT, X(1),X(2),....,X(PMAX) = A(1,PMAX),A(2,PMAX),....,A(PMAX,PMAX)
C  Y(1),Y(2),....,Y(N) = AUXILIARY ARRAY; SCRATCH INPUT
C  ON OUTPUT, Y(1),Y(2),....,Y(PMAX) = A(1,1),A(2,2),....,A(PMAX,PMAX)
C  ON OUTPUT, X(N) = MEAN, AND Y(N) = STANDARD DEVIATION OF INPUT DATA
C  PBEST = BEST ORDER OF FILTER; INTEGER OUTPUT
C  A(1),A(2),....,A(PBEST) = PREDICTIVE FILTER COEFFICIENT ARRAY =
C  A(1,PBEST),A(2,PBEST),....,A(PBEST,PBEST); OUTPUT
C  PROD = PRODUCT(1-A(P,PBEST)**2) FOR P=1 TO PBEST OUTPUT
C  RHO(1),....,RHO(PMAX) = NORMALIZED CORRELATION COEFFICIENTS; OUTPUT
C  DIMENSION X(N),Y(N),A(PMAX),RHO(PMAX)      IS REQUIRED IN MAIN PROGRAM
      INTEGER PMAX,PBEST,P
      DOUBLE PRECISION SA,SB
      DIMENSION X(1),Y(1),A(1),RHO(1)
      IF(PMAX.GT.3.*SQRT(N)) PRINT 2, PMAX,N
2     FORMAT(/,' PMAX =',I4,' IS TOO LARGE FOR NUMBER OF DATA POINTS N ='
     8,I5)
C  COMPUTE MEAN
      S1=0.
      DO 1 I=1,N
1     S1=S1+X(I)
      S1=S1/N
C  SUBTRACT MEAN, AND SCALE TO UNIT VARIANCE
      S2=0.
      DO 3 I=1,N
      X(I)=X(I)-S1
3     S2=S2+X(I)**2
      S2=SQRT(S2/(N-1.))
      T=1./S2
      DO 5 I=1,N
      X(I)=X(I)*T
5     Y(I)=X(I)
C  BEGIN RECURSION
      P=0
      PRODUC=1.
      AICMIN=0.
      PBEST=0
      PROD=1.
6     P=P+1
C  CALCULATE CROSS-GAIN; EQ. 155
      SA=0.D0
      SB=0.D0
      L=P+1
      DO 7 I=L,N
      SA=SA+X(I)*Y(I-1)
7     SB=SB+X(I)**2+Y(I-1)**2
      G=2.*SA/SB
      PRODUC=PRODUC*(1.-G*G)
C  CALCULATE FILTER COEFFICIENTS; EQS. 146&148, STORE IN X(1),....,X(P)
      X(P)=G
      IF(P.EQ.1) GO TO 8
      L=P/2
      DO 9 I=1,L
      T=X(I)-G*X(P-I)
      X(P-I)=X(P-I)-G*X(I)
9     X(I)=T
C  CALCULATE NORMALIZED CORRELATION COEFFICIENT; EQ. 149
8     T=X(P)
      IF(P.EQ.1) GO TO 14
      L=P-1
      DO 15 I=1,L
15    T=T+X(I)*RHO(P-I)
14    RHO(P)=T
```

```
C   CALCULATE AKAIKE'S INFORMATION CRITERION; EQS. 156&202
        RELERR=(1.-G*G)*SNGL(S6)/(2.*(N-P))
        AIC=LOG(RELERR)+2.*FLOAT(P)/(N-P)
        IF(AIC.GE.AICMIN) GO TO 10
        AICMIN=AIC
        PBEST=P
        PROD=PRODUC
        DO 11 I=1,P
11      A(I)=X(I)
10      IF(P.EQ.PMAX) GO TO 16
C   UPDATE FORWARD AND BACKWARD SEQUENCES; EQ. 153
        L=P+1
        DO 12 I=N,L,-1
        T=X(I)-G*Y(I-1)
        Y(I)=Y(I-1)-G*X(I)
12      X(I)=T
        Y(P)=G
        GO TO 6
16      Y(PMAX)=G
        IF(PBEST.EQ.PMAX) GO TO 4
C   COMPUTE EXTRAPOLATED NORMALIZED CORRELATION
C   COEFFICIENTS FROM PBEST+1 TO PMAX; EQ. 165
        L=PBEST+1
        DO 17 P=L,PMAX
        A(P)=0.
        T=0.
        DO 18 I=1,PBEST
18      T=T+A(I)*RHO(P-I)
17      RHO(P)=T
4       X(N)=S1
        Y(N)=S2
        RETURN
        END


        SUBROUTINE POWERS(PBEST,A,PROD,J,XX,YY,CO)
C   THIS SUBROUTINE COMPUTES THE FRACTIONAL POWERS IN BANDS 1/(J*DELTA); EQ. J-4
C   PBEST = BEST ORDER OF FILTER; INTEGER INPUT
C   A(1),....,A(PBEST) = FILTER COEFFICIENT ARRAY; INPUT
C   PROD = PRODUCT(1-A(P)**2)  FOR P=1 TO PBEST; INPUT
C   J = SIZE OF FFT   (J/2+1=NUMBER OF FREQUENCY POINTS); INTEGER INPUT
C   XX = AUXILIARY ARRAY ON INPUT
C   XX(1),....,XX(J/2+1) = FRACTIONAL POWERS ON OUTPUT
C   YY = AUXILIARY ARRAY; SCRATCH INPUT
C   CO(1),....,CO(J/4+1) = QUARTER COSINE TABLE FOR FFT; INPUT
C   DIMENSION XX(J),YY(J),CO(J/4+1)  IS REQUIRED IN MAIN PROGRAM
C   DIMENSION A(PMAX)  IS REQUIRED IN MAIN PROGRAM, WHERE PMAX.GE.PBEST
        INTEGER PBEST
        DIMENSION A(1),XX(1),YY(1),CO(1)
        F=PROD*2./J
        XX(1)=1.
        YY(1)=0.
        DO 1 I=1,PBEST
        XX(I+1)=-A(I)
1       YY(I+1)=0.
        L=PBEST+2
        DO 2 I=L,J
        XX(I)=0.
2       YY(I)=0.
        L=1.4427*LOG(J)+.5       @ LOG2(J)
        CALL MKLFFT(XX,YY,CO,L,-1)
        L=J/2+1
        DO 3 I=1,L
3       XX(I)=F/(XX(I)**2+YY(I)**2)
        RETURN
        END
```

```
INPUT DATA:
    -.65987096+01           .11340011+02           .15902418+02           .74923898+01          -.22727143+01
    -.21771812+01           .67804896+01           .20620527+02           .19559856+02           .10803774+01
    -.26660023+02          -.41166628+02          -.30993711+02           .25941087+01           .38769337+02
     .64401053+02           .36074435+02          -.72542098+01          -.48348869+02          -.61963475+02
     .39646678+02           .67681677+01           .46871204+02           .60280199+02           .39106104+02
    -.24300807+01          -.39391616+02          -.62051208+02          -.37437309+02          -.58956462+01
     .25681209+02           .43658428+02           .42103511+02           .23748203+02          -.36984812+01
    -.28371239+02          -.40389613+02          -.36354962+02          -.15176634+02           .13645936+02
     .38701371+02           .47343318+02           .33784445+02           .18979633+01          -.34617383+02
    -.56673562+02          -.50217300+02          -.18068573+02           .22344941+02           .48678704+02
     .46966025+02           .21067918+02          -.11789432+02          -.32019099+02          -.32151017+02
    -.17956020+02           .14407900+01           .17644500+02           .44891572+02           .29231678+02
     .24659073+02           .67850477+01          -.14015273+02          -.32436075+02          -.37618856+02
    -.24832260+02          -.45756262+00           .43166870+02           .35915088+02           .31645702+02
     .14791357+02          -.11089472+02          -.28190744+02          -.31179497+02          -.19281728+02
    -.16106005+02           .12430834+02           .18994507+02           .19214284+02           .13325080+02
     .39338522+00          -.18632998+02          -.35072248+02          -.38184243+02          -.21167069+02
     .13023073+02           .46910409+02           .64112801+02           .44770043+02          -.37641446+01
    -.56600650+02          -.60467312+02          -.67783633+02          -.11811747+00           .50741565+02
     .86403726+02           .63360344+02           .60409774+01          -.32971808+02          -.79348459+02
```

MEAN = .11733613+00
STANDARD DEVIATION = .30066007+02

PBEST = 4

PREDICTIVE FILTER COEFFICIENTS:
    .27035131+01   -.37823616+01   .26406127+01   -.93677752+00

PRODUCT(1-A(P)**2) = .88956827-03

NORMALIZED CORRELATION COEFFICIENTS:
    .73302390+00   .81104860-01   -.59619489+00   -.93601970+00   -.77701127+00
   -.23157107+00   .39061750+00   .76830663+00   .73634707+00   .36043824+00

BAD DATA POINTS (SUBSECTION 5.2)


```
C  SPECTRAL ESTIMATION FOR BAD DATA POINTS       USER: CHANGE
C  LINE 17 AND REPLACE LINES 22-36 AND 41-46
C  N = NUMBER OF DATA POINTS
C  X(1),....,X(N) = INPUT DATA
C  BMAX = MAXIMUM NUMBER OF BAD DATA POINTS
C  B = ACTUAL NUMBER OF BAD DATA POINTS   (MUST HAVE B.LE.BMAX)
C  M(1),....,M(B) = LOCATIONS OF BAD DATA POINTS
C  PMAX = MAXIMUM ORDER OF FILTER
C  PBEST = BEST ORDER OF FILTER
C  A(1),....,A(PBEST) = PREDICTIVE FILTER COEFFICIENTS
C  PROD = PRODUCT(1-A(P)**2) FOR P=1 TO PBEST
C  RHO(1),....,RHO(PMAX) = NORMALIZED CORRELATION COEFFICIENTS
C  J = SIZE OF FFT   (MUST BE A POWER OF 2)
C  XX(1),.....,XX(J/2+1) = FRACTIONAL POWERS, FROM DC TO NYQUIST FREQUENCY
C  CO(1),.....,CO(J/4+1) = QUARTER COSINE TABLE
C  Y, YY, AND IP ARE REQUIRED AUXILIARY ARRAYS
      PARAMETER N= 100, BMAX= 25, PMAX=10, J=2048, J41=J/4+1
      INTEGER B,PBEST
      DIMENSION X(N),Y(N),A(PMAX),RHO(PMAX),XX(J),YY(J),CO(J41)
      DIMENSION M(BMAX),IP(N)
C  INPUT DATA IN X(1),.....,X(N)
      DEFINE IRAND=I*5**15+((1-SIGN(1,I*5**15))/2)*34359738367
      DEFINE RAND=FLOAT(I)/34359738367.
      I=5281
      NSTART=N+400      @ WILL DISCARD INITIAL 400 POINTS
```

```
      XX(1)=0.
      XX(2)=0.
      XX(3)=0.
      XX(4)=0.
      DO 11 L=5,NSTART
      I=IRAND
      XX(L)=2.7607*XX(L-1)-3.8106*XX(L-2)+2.6536*XX(L-3)-
     $0.9238*XX(L-4)+TINORM(RAND,$11)
11    CONTINUE
      DO 12 I=1,N
12    X(I)=XX(I+NSTART-N)
      PRINT 1,
1     FORMAT(/' INPUT DATA!')
      PRINT 4, (X(I),I=1,N)
C  ENTER B, AND ENTER BAD DATA LOCATIONS IN M(1),....,M(B)
      B=5
      M(1)=3
      M(2)=7
      M(3)=11
      M(4)=12
      M(5)=19
C  EVALUATE PREDICTIVE FILTER COEFFICIENTS
      CALL BURGBD(N,PMAX,X,B,M,IP,Y,PBEST,A,PROD,RHO)
      PRINT 9, X(N)
9     FORMAT(/' MEAN =',E14.6)
      PRINT 10, Y(N)
10    FORMAT(' STANDARD DEVIATION =',E13.8)
      PRINT 2, PBEST
2     FORMAT(/' PBEST =',I3)
      PRINT 3,
3     FORMAT(/' PREDICTIVE FILTER COEFFICIENTS!')
      PRINT 4, (A(I),I=1,PBEST)
4     FORMAT(5E20.8)
      PRINT 5, PROD
5     FORMAT(/' PRODUCT(1-A(P)**2) =',E13.8)
      PRINT 6,
6     FORMAT(/' NORMALIZED CORRELATION COEFFICIENTS!')
      PRINT 4, (RHO(I),I=1,PMAX)
      CALL QTRCOS(CO,J)
C  EVALUATE FRACTIONAL POWERS
      CALL POWERS(PBEST,A,PROD,J,XX,YY,CO)
      PRINT 7,
7     FORMAT(/' FRACTIONAL POWERS!')
      L=J/2+1
      PRINT 8, (XX(I),I=1,L)
8     FORMAT(2X,10E13.6)
      END


      SUBROUTINE BURGBD(N,PMAX,X,B,M,IP,Y,PBEST,A,PROD,RHO)  W 2 FEB 1976
C  THIS SUBROUTINE COMPUTES THE PREDICTIVE FILTER COEFFICIENTS FOR B BAD POINTS
C  N = NUMBER OF DATA POINTS! INTEGER INPUT
C  PMAX = MAXIMUM ORDER OF FILTER! INTEGER INPUT
C  X(1),X(2),....,X(N) = DATA ARRAY ON INPUT! ALTERED ON OUTPUT
C  ON OUTPUT, X(1),X(2),....,X(PMAX) = A(1!PMAX),A(2!PMAX),....,A(PMAX!PMAX)
C  B = NUMBER OF BAD DATA POINTS! INTEGER INPUT
C  M(1),M(2),....,M(B) = LOCATIONS OF BAD DATA POINTS! INTEGER INPUTS
C  THESE LOCATIONS MUST BE DISTINCT AND LIE IN THE RANGE [1,N]
C  IP(1),IP(2),....,IP(N) = AUXILIARY ARRAY! SCRATCH INPUT
C  Y(1),Y(2),....,Y(N) = AUXILIARY ARRAY! SCRATCH INPUT
C  ON OUTPUT, Y(1),Y(2),....,Y(PMAX) = A(1!1),A(2!2),....,A(PMAX!PMAX)
C  ON OUTPUT, X(N) = MEAN, AND Y(N) = STANDARD DEVIATION OF INPUT DATA
C  PBEST = BEST ORDER OF FILTER! INTEGER OUTPUT
C  A(1),A(2),....,A(PBEST) = PREDICTIVE FILTER COEFFICIENT ARRAY =
C  A(1!PBEST),A(2!PBEST),....,A(PBEST!PBEST)! OUTPUT
```

```
C   PROD = PRODUCT(1-A(P,PBEST)**2) FOR P=1 TO PBEST; OUTPUT
C   RHO(1),....,RHO(PMAX) = NORMALIZED CORRELATION COEFFICIENTS; OUTPUT
C   DIMENSION X(N),Y(N),A(PMAX),RHO(PMAX)      IS REQUIRED IN MAIN PROGRAM.
C   DIMENSION M(BMAX),IP(N)    IS REQUIRED IN MAIN PROGRAM
        INTEGER PMAX,B,PBEST,P,BP
        DOUBLE PRECISION SA,SB
        DIMENSION X(1),M(1),IP(1),Y(1),A(11),RHO(1)
        IF(B.GT.0) GO TO 21
        CALL BURG(N,PMAX,X,Y,PBEST,A,PROD,RHO)
        RETURN
21      L=N-B
        IF(PMAX.GT.3.*SQRT(L)) PRINT 2, PMAX,L
2       FORMAT(/,' PMAX =',I4,' IS TOO LARGE FOR NUMBER OF GOOD DATA POINTS
     3 N-B =',I5)
C   SET UP IP ARRAY FOR P=0; EQ. 173
        DO 22 I=1,N
22      IP(I)=1
        DO 23 L=1,B
        I=M(L)
23      IP(I)=0
C   COMPUTE MEAN OF GOOD DATA POINTS
        S1=0.
        DO 1 I=1,N
        IF(IP(I).EQ.0) GO TO 1
        S1=S1+X(I)
1       CONTINUE
        S1=S1/(N-B)
C   SUBTRACT MEAN, AND SCALE TO UNIT VARIANCE, FOR GOOD DATA POINTS
        S2=0.
        DO 3 I=1,N
        IF(IP(I).EQ.0) GO TO 3
        X(I)=X(I)-S1
        S2=S2+X(I)**2
3       CONTINUE
        S2=SQRT(S2/(N-B-1.))
        T=1./S2
        DO 5 I=1,N
        IF(IP(I).EQ.0) GO TO 5
        X(I)=X(I)*T
        Y(I)=X(I)
5       CONTINUE
C   BEGIN RECURSION
        P=0
        PRODUC=1.
        AICMIN=0.
        PBEST=0
        PROD=1.
6       P=P+1
C   UPDATE IP ARRAY; EQ. 173
        DO 24 L=1,B
        I=M(L)+P
        IF(I.GT.N) GO TO 24
        IP(I)=0
24      CONTINUE
        BP=0
        L=P+1
        DO 25 I=L,N
25      BP=BP+1-IP(I)
        K=N-P-BP
        IF(K.LT.25) PRINT 26, K,P
26      FORMAT(/,' NUMBER OF VALID ERROR POINTS IS ONLY',I3,' FOR P=',I3)
C   CALCULATE CROSS-GAIN; EQ. 193
        SA=0.D0
```

```
      SB=0.D0
      L=P+1
      DO 7 I=L,N
      IF(IP(I).EQ.0) GO TO 7
      SA=SA+X(I)*Y(I-1)
      SB=SB+X(I)**2+Y(I-1)**2
7     CONTINUE
      G=2.*SA/SB
      PRODUC=PRODUC*(1.-G*G)
C  CALCULATE FILTER COEFFICIENTS: EQS. 195&196. STORE IN X(1),...,X(P)
      X(P)=G
      IF(P.EQ.1) GO TO 8
      L=P/2
      DO 9 I=1,L
      T=X(I)-G*X(P-I)
      X(P-I)=X(P-I)-G*X(I)
9     X(I)=T
C  CALCULATE NORMALIZED CORRELATION COEFFICIENT: EQ. 149
8     T=A(P)
      IF(P.EQ.1) GO TO 14
      L=P-1
      DO 15 I=1,L
15    T=T+X(I)*RHO(P-I)
14    RHO(P)=T
C  CALCULATE AKAIKE'S INFORMATION CRITERION: EQS. 194&202
      RELERR=(1.-G*G)*SNGL(SB)/(2.*K)
      AIC=LOG(RELERR)+2.*FLOAT(P)/K
      IF(AIC.GE.AICMIN) GO TO 10
      AICMIN=AIC
      PBEST=P
      PROD=PRODUC
      DO 11 I=1,P
11    A(I)=X(I)
10    IF(P.EQ.PMAX) GO TO 16
C  UPDATE FORWARD AND BACKWARD SEQUENCES: EQ.191
      L=P+1
      DO 12 I=N,L,-1
      IF(IP(I).EQ.0) GO TO 12
      T=X(I)-G*Y(I-1)
      Y(I)=Y(I-1)-G*X(I)
      X(I)=T
12    CONTINUE
      Y(P)=G
      GO TO 6
16    Y(PMAX)=G
      IF(PBEST.EQ.PMAX) GO TO 4
C  COMPUTE EXTRAPOLATED NORMALIZED CORRELATION
C  COEFFICIENTS FROM PBEST+1 TO PMAX: EQ. 165
      L=PBEST+1
      DO 17 P=L,PMAX
      A(P)=0.
      T=0.
      DO 18 I=1,PBEST
18    T=T+A(I)*RHO(P-I)
17    RHO(P)=T
4     X(N)=S1
      Y(N)=S2
      RETURN
      END
```

```
      SUBROUTINE POWERS(PBEST,A,PROD,J,XX,YY,CO)
C THIS SUBROUTINE COMPUTES THE FRACTIONAL POWERS IN BANDS 1/(J*DELTA); EQ. J-4
C PBEST = BEST ORDER OF FILTER; INTEGER INPUT
C A(1),...,A(PBEST) = FILTER COEFFICIENT ARRAY; INPUT
C PROD = PRODUCT(1-A(P)**2)  FOR P=1 TO PBEST; INPUT
C J = SIZE OF FFT   (J/2+1=NUMBER OF FREQUENCY POINTS); INTEGER INPUT
C XX = AUXILIARY ARRAY ON INPUT
C XX(1),...,XX(J/2+1) = FRACTIONAL POWERS ON OUTPUT
C YY = AUXILIARY ARRAY; SCRATCH INPUT
C CO(1),.....,CO(J/4+1) = QUARTER COSINE TABLE FOR FFT; INPUT
C DIMENSION XX(J),YY(J),CO(J/4+1)  IS REQUIRED IN MAIN PROGRAM
C DIMENSION A(PMAX)  IS REQUIRED IN MAIN PROGRAM, WHERE PMAX.GE.PBEST
      INTEGER PBEST
      DIMENSION A(1),XX(1),YY(1),CO(1)
      F=PROD*2./J
      XX(1)=1.
      YY(1)=0.
      DO 1 I=1,PBEST
      XX(I+1)=-A(I)
1     YY(I+1)=0.
      L=PBEST+2
      DO 2 I=L,J
      XX(I)=0.
2     YY(I)=0.
      L=1.4427*LOG(J)+.5          LOG2(J)
      CALL MKLFFT(XX,YY,CO,L,-1)
      L=J/2+1
      DO 3 I=1,L
3     XX(I)=F/(XX(I)**2+YY(I)**2)
      RETURN
      END
```

REFERENCES

1. J. Makhoul, "Linear Prediction : A Tutorial Review," <u>Proceedings of the IEEE</u>, vol. 63, no. 4, April 1975, pp. 561-580.

2. T. J. Ulrych and T. N. Bishop, "Maximum Entropy Spectral Analysis and Autoregressive Decomposition," <u>Reviews of Geophysics and Space Physics</u>, vol. 13, no. 1, February 1975, pp. 183-200.

3. H. Akaike, "On the Use of a Linear Model for the Identification of Feedback Systems," <u>Annals of the Institute of Statistical Mathematics</u>, vol. 20, no. 3, 1968, pp. 425-439.

4. J. P. Burg, "New Concepts in Power Spectra Estimation," <u>40th Meeting of Society of Exploration Geophysicists</u>, New Orleans, Louisiana, November 1970.

5. R. H. Jones, "Spectrum Estimation with Missing Observations," <u>Annals of the Institute of Statistical Mathematics</u>, vol. 23, no. 3, 1971, pp. 387-398.

6. E. Parzen, "Some Recent Advances in Time Series Analysis," <u>Statistical Models and Turbulence</u>, edited by M. Rosenblatt, Springer-Verlag, New York, 1971, pp. 470-492.

7. D. C. Riley and J. P. Burg, "Time and Space Adaptive Deconvolution Filters," <u>42 Meeting of Society of Exploration Geophysicists</u>, Anaheim, Calif., 1972.

8. B. Veltman, et al., "Some Remarks on the Use of Auto-correlation Functions with the Analysis and Design of Signals," <u>Signal Processing</u>, edited by J. W. R. Griffiths, P. L. Stocklin, and C. Van Schooneveld, Academic Press, London and New York, 1973.

9. R. H. Jones, "Autoregressive Spectrum Estimation," <u>Third Conference on Probability and Statistics in Atmospheric Science</u>, Boulder, Colo., 19-22 June 1973.

10. W. R. King, W. H. Swindell, and L. J. O'Brien, <u>Final Report on Development of a Curvilinear Ray Theory Model and Maximum Entropy Spectral</u>

Analysis, Texas Instruments Inc., Report ALEX(03)-FR-74-01, 28 February 1974.

11. E. M. Hofstetter, "An Introduction to the Mathematics of Linear Predictive Filtering as Applied to Speech Analysis and Synthesis," Lincoln Lab, Mass. Inst. of Tech., Technical Note 1973-36, Rev. 1, 12 April 1974.

12. R. H. Jones, "Identification and Autoregressive Spectrum Estimation," Proceedings of Conference on Decision and Control, Phoenix, Arizona, 20-22 November 1974.

13. E. Parzen, "Some Recent Advances in Time Series Modeling," IEEE Transaction on Automatic Control, vol. AC-19, no. 6, December 1974, pp. 723-730.

14. R. N. McDonough, "Maximum-Entropy Spatial Processing of Array Data," Geophysics, vol. 39, no. 6, December 1974, pp. 843-851.

15. F. J. Harris, A Maximum Entropy Filter, Naval Undersea Center Report NUC TP 442, January 1975.

16. G. E. Box and G. M. Jenkins, Time Series Analysis; Forecasting and Control, Holden-Day, San Francisco, Calif., 1970.

17. A. H. Nuttall, "Trigonometric Smoothing and Interpolation of Sampled Complex Functions, via the FFT," NUSC Technical Memorandum TC-94-71, 19 April 1971.

18. J. P. Burg, "Maximum Entropy Spectral Analysis," 37th Meeting of Society of Exploration Geophysicists, Oklahoma City, Oklahoma, 31 October 1967.

19. J. A. Edward and M. M. Fitelson, "Notes on Maximum-Entropy Processing," IEEE Transactions on Information Theory, vol IT-19, March 1973, pp. 232-234.

20. A van den Bos, "Alternative Interpretation of Maximum Entropy Spectral Analysis," IEEE Transactions on Information Theory, vol. IT-17, no. 4, July 1971, pp. 493-494.

21. J. P. Burg, "A New Analysis Technique for Time Series Data," NATO Advanced Study Institute on Signal Processing, Enschede, Netherlands, volume 1, August 1968.

22. H. Akaike, "A New Look at Statistical Model Identification," IEEE Transactions on Automatic Control, vol. AC-19, December 1974, pp. 716-723. (This entire volume is devoted to Systems Identification and Time Series Analysis.)

23. R. T. Lacoss, "Data Adaptive Spectral Analysis Methods," Geophysics, vol. 36, no. 4, August 1971, pp. 661-675.

24. N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, John Wiley and Sons, Inc., New York, May 1950.

25. J. F. Ferrie, G. C. Carter, and C. W. Nawrocki, "Availability of Markel's FFT Algorithm," NUSC Technical Memorandum TC-1-73, 15 January 1973.

INITIAL DISTRIBUTION LIST

Addressee	No. of Copies
ASN (R&D)	1
ONR	1
ONR, Code 412-3	1
Code 480	1
CNO, Op-098	
Op-098T	1
Op-23T	1
Op-96	1
Op-96C	1
Op-96C1	1
CNM, MAT-03	1
MAT-0302	1
PM-4 (ASW-132)	1
NRL	1
NRL (W. Morrough)	1
SUBASELANT	1
NAVOCEANO, Code 02	1
Code 7200	1
NAVELEX, Code 03	1
PME-124 (D. Hyde)	1
NAVSEA, SEA-03C	1
SEA-09G3	4
SEA-06H1 (J. Neely, D. Porter)	2
SEA-06H2 (D. Early	1
SEA-660	1
NWC	1
DTNSRDC	1
NCSL	1
NCEL	1
NELC	1
NUC (D. Hanna, R. Hally, R. Bolam, C. Chabries, R. Albreight, G. Byram, P. Moose, J. Stewart, R. Hern)	9
NAVSEC, SEC-6034	1
NAVPGSCOL	1
ARL/PENN STATE	1
Center for Naval Analyses (Acquisition Unit)	1
DDC, Alexandria	12

INITIAL DISTRIBUTION LIST (Cont'd)

Addressee	No. of Copies
Defense Intelligence Agency, DT2-C	1
IRIA	1
Marine Physical Lab., Scripps	1
National Research Council (Committee on Undersea Warfare)	1
Woods Hole Oceanographic Institution	1
NADC (T. Castaldi)	1
ARL-UT	1
NSWC (M. Stripling)	1
NISC, Code 243 (F. Huntsinger)	1
ARPA (CAPT H. Cox)	1
SACLANT Research Centre (E. Hug)	1