AEDC-TR-74-53

cy. 3

THE DEVELOPMENT OF NEW DIGITAL DATA PROCESSING TECHNIQUES FOR TURBULENCE MEASUREMENTS WITH A LASER VELOCIMETER

W. T. Mayo, Jr., M. T. Shay, and S. Riter
Texas A&M University
College Station, Texas

August 1974

Final Report for Period December 1, 1972 — January 1, 1974

Approved for public release; distribution unlimited.

Property of U. S. Air Force
AEDC LIBRARY
F40600-72-C-0001

Prepared for

DEPARTMENT OF THE AIR FORCE
HEADQUARTERS ARNOLD ENGINEERING DEVELOPMENT CENTER (DYFS)
AIR FORCE SYSTEMS COMMAND
ARNOLD AIR FORCE STATION, TENNESSEE 37389

NOTICES

When U. S. Government drawings specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, or in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Qualified users may obtain copies of this report from the Defense Documentation Center.

References to named commercial products in this report are not to be considered in any sense as an endorsement of the product by the United States Air Force or the Government.

APPROVAL STATEMENT

This technical report has been reviewed and is approved.

Melvin L. Guiou

MELVIN L. GUIOU
Captain, USAF
Research and Development
Division
Directorate of Technology

Robert O Dietz

ROBERT O. DIETZ
Director of Technology

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AEDC-TR-74-53	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE DEVELOPMENT OF NEW DIGITAL DATA PROCESSING TECHNIQUES FOR TURBULENCE MEASUREMENTS WITH A LASER VELOCIMETER		5. TYPE OF REPORT & PERIOD COVERED Final Report-December 1, 1972 to January 1, 1974 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) W. T. Mayo, Jr., M. T. Shay, and S. Riter Texas A&M University College Station, Texas		8. CONTRACT OR GRANT NUMBER(s) F40600-73-C-0003
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University College Station, Texas 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Program Element 65802F
11 CONTROLLING OFFICE NAME AND ADDRESS Arnold Engineering Development Center (DYFS) Arnold Air Force Station, Tennessee 37389		12 REPORT DATE August 1974 13. NUMBER OF PAGES 111
14 MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15 SECURITY CLASS. (of this report) UNCLASSIFIED 15a DECLASSIFICATION/DOWNGRADING SCHEDULE N/A

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

Available in DDC

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

data processing	turbulence
digital systems	measurements
engineering	laser velocimeter

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

This research has developed a laser velocimeter data processing technique that may be used for on-line estimation of turbulence power spectra from randomly timed samples at low data rates. First, an estimate of the turbulence autocovariance is made at each lag by discretizing the time axis into lag units, accumulating the products of pairs of sample magnitudes occurring at each lag, and dividing the accumulated sums by the histogram of

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

lag products. Then, the discrete Fourier transform of the turbulence autocovariance is computed, giving an estimate of the turbulence power spectra. Theory has been developed for prediction of the spectrum estimate error in terms of the algorithm and LV system-flow parameters, showing that the optimum data rate, for efficient estimation of the major distribution of turbulence power, is twice the equivalent power bandwidth of the spectrum. Experimental evidence is presented supporting the theoretical error prediction equations. A preliminary design guideline is presented, with examples, and alternate implementations of the autocovariance estimator are discussed.

PREFACE

This technical report was prepared by Texas A&M University, College Station, Texas, under U. S. Air Force Contract No. F40600-73-C-0003, for the Arnold Engineering Development Center (AEDC) (DYR), Air Force Systems Command (AFSC), Arnold Air Force Station, Tennessee. The work was conducted by the Remote Sensing Center, Texas A&M University, from December 1, 1972 to January 1, 1974 under Program Element 65802F. The manuscript was submitted in April 1974. The program was under the direction of Lt Col J. R. Taylor of AEDC.

The reproducibles used in the reproduction of this report were supplied by the authors.

CONTENTS

	Page
PREFACE	i
LIST OF ILLUSTRATIONS	vi
SYMBOLS AND NOTATION	vii
I. INTRODUCTION	1
1.1 The Problem	1
1.2 Background for Statistical Estimation	1
1.2.1 Conventional Spectrum Analysis	1
1.2.2 Randomly Sampled Signals	2
1.2.3 Estimation of Broadband Spectra from Random Samples	4
1.3 Scope	6
II. THEORY	7
2.1 Introduction	7
2.1.1 Assumptions	7
2.1.2 Mathematical Background	8
2.2 Mean and Intensity Estimators	9
2.2.1 Introduction	9
2.2.2 Mean Flow	10
2.2.3 Mean-Square Intensity	11
2.2.4 Normalized Turbulence Intensity	12
2.3 Autocovariance	13
2.3.1 The Estimators	13
2.3.2 Bias Error Due to Time Increment	14
2.3.3 Additional Bias Error	16
2.3.4 The Histogram	16
2.3.5 Variance of the Autocovariance Estimate	17
2.4 Power Spectrum	18
2.4.1 Computation from the Autocovariance	18

2.4.2 Bias Effect of $\Delta\tau$ Selection 21

2.4.3 Effect of Non-Zero Mean and Linear Trends 22

2.4.4 Variance of the Spectral Estimate 23

2.4.5 Post Estimate Smoothing 25

2.5 Processor Dead Time 26

2.6 Segmented Data - Nonstationarity 27

III. EXPERIMENTAL 29

3.1 Introduction 29

3.2 Results 30

3.2.1 Preliminary Results 30

3.2.2 Parametric Variability Study 35

3.2.3 Parametric Study of Finite Time Correlation Effects on Mean and Turbulence Intensity Variability 42

3.2.4 Histogram Error 42

IV. PRELIMINARY DESIGN 46

4.1 Introduction 46

4.2 Preliminary Design Procedure 46

4.2.1 Bandwidth - Selection of $\Delta\tau$ 47

4.2.2 Resolution, Frequency Spacing, and Number of Estimates 47

4.2.3 Data Segment Length 48

4.2.4 Variability Error 51

4.2.5 Amplitude and Time Error 52

4.2.6 Computation Time and Memory Considerations 52

4.3 Examples 56

4.3.1 Problem Specification 56

4.3.2 Low Frequency, High Resolution 57

4.3.3 High Frequency Spectrum Tails 60

V. DISCUSSION 61

5.1 Implementation Alternatives 61

5.1.1 Batch Processing - Segmented Data 61
5.1.2 On Line Processing - Segmented Data 62
5.1.3 On Line Processing - Continuous 63
5.1.4 Hardware 63
5.2 Relation Between Theory and Experiment 64
5.3 Comments on Future Research 66
5.3.1 Sample Amplitude Error 66
5.3.2 Cross-Correlations and Cross-Spectra 68
5.3.3 Correlations Between Mean Sample Rate and the Velocity Process 70
VI. SUMMARY AND CONCLUSIONS 72
REFERENCES 74

APPENDIXES

I. TWO SPECTRAL ESTIMATORS BASED ON THE PERIODIOGRAM APPROACH 77
II. DERIVATIONS OF ESTIMATOR ERRORS 85
III. EXPERIMENTAL FACILITIES AND DATA PROCESSING 91

LIST OF ILLUSTRATIONS

Figure		Page
1	Variation of Spectrum Estimates with Number of Data Segments	31
2	Variation of Spectrum Estimates with Mean Data Rate	33
3	Spectrum Estimates Exhibiting Aliasing	34
4	Analytical Butterworth Spectra and "High Accuracy" Spectrum Estimates	37
5	Spectrum Estimate Error in the Passband - Theoretical and Experimental	40
6	Spectrum Estimate Error in the Stopband - Theoretical and Experimental	41
7	Normalized RMS Error of the Mean Estimate vs α	43
8	Normalized RMS Error of the Turbulence Intensity Estimate vs α	44
9	Histograms of the Number of Lag Products	45
10	Normalized RMS Spectrum Error with α vs β	50
11	Normalized RMS Spectrum Error with ε_1 vs β	51
12	Normalized RMS Spectrum Error with α vs γ	53
13	Normalized RMS Spectrum Error with ε_1 vs γ	54

SYMBOLS AND NOTATION

B	- equivalent power bandwidth
$C(\tau)$	- autocovariance of $U(t)$
$\hat{C}(k)$	- estimate of $C(\tau)$ at $\tau = k\Delta\tau$
$\hat{C}'(k)$	- estimate of $C(\tau)$ at $\tau = k\Delta\tau$ using alternate formulation
$H(k)$	- histogram of the number of lag products
I	- turbulence intensity
$\hat{I}$	- estimate of turbulence intensity
M	- maximum number of $\Delta\tau$ time increments
N	- the number of samples in a data segment
N'	- number of uniform time intervals $\Delta\tau$ in T
Ns	- the number of data segments used to form a spectrum estimate
R	- frequency resolution $\neq \Delta f$
$R(\tau)$	- autocorrelation of $U(t)$
$S(f)$	- power spectrum (Fourier transform of $C(\tau)$)
$\hat{S}(i\Delta f)$	- estimate of $S(f)$ at $f = i\Delta f$
T	- time duration of data collection
$U(t)$	- velocity
$u(t)$	- time-varying velocity (zero mean)
$\overline{U}$	- mean of $U(t)$
$\hat{U}$	- estimate of $\overline{U}$

$w(k)$ - lag window function at lag k

$W(f)$ - Fourier transform of $w(k)$

α - normalized mean rate parameter, $\frac{\lambda}{2B}$

β - normalized number of data samples, $\frac{NNs}{M}$

γ - normalized data collection time, $\frac{T}{M\Delta\tau}$

Δf - estimate spacing in frequency

$\Delta\tau$ - uniform increments in time

$M\Delta\tau$ - maximum delay

ε_1 - normalized RMS deviation of spectrum estimate

$\hat{\varepsilon}_1$ - experimental measure of ε_1

λ - mean data rate

σ^2 - variance of $U(t)$ (mean-square intensity)

$\hat{\sigma}^2$ - estimate of σ^2

σ_f^2 - variance of $\hat{S}(i\Delta f)$

$\sigma_{\hat{u}}$ - standard deviation of $\hat{U}$

τ_c - turbulence time scale

ψ - rate at which lag products occur in real-time

SECTION I

INTRODUCTION

1.1 THE PROBLEM

Within many subtle limitations, laser velocimeter (LV) systems provide non-interfering point measurements of a velocity field when suspended scatterers pass through a small optical probe volume [1]. For wind tunnel measurements with little or no particulate seeding of the gas, the LV systems of concern here make essentially instantaneous sample measurements from sparsely distributed scatterers which are detected at random intervals in time. The primary objective of the research reported herein has been the development of digital statistical data processing techniques for the estimation of turbulence intensity, autocorrelation, and power spectrum, with low mean data rates which minimize the particulate seeding requirement. An additional objective has been the reduction of system costs by elimination of the need for expensive mass memory through the use of real-time minicomputer processing.

1.2 BACKGROUND FOR STATISTICAL ESTIMATION

1.2.1 Conventional Spectrum Analysis

Many engineers who are familiar with the use of analog spectrum analyzers in measuring periodic data are unaware of the many subtleties in the measurement of power

spectra of continuous noise processes. However, theory and practices are well developed, at least for stationary Gaussian processes [2,3].

In many applications the need for greater accuracy or the digital nature of the original data format has led to digital power spectrum analysis of periodically sampled data. There is a clear distinction between the requirements of periodic or nearly periodic signals and those for broadband random signals, but the techniques and limitations are well understood for stationary Gaussian signals [2-9]. An excellent introduction to this subject has been provided in the very readable paper by Richards [4]. Free use of this background material is utilized in this report.

1.2.2 Randomly Sampled Signals

Fourteen years of rigorous mathematical research with statistical averages or infinite time averages indicates that not only is it possible to make mean, correlation, and spectral estimates using randomly timed samples but also that there is a significant advantage in doing so. For estimating second order statistical parameters such as correlation and spectrum it is well known that with periodic sampling the sample rate must exceed twice the signal bandwidth to avoid aliasing error [5]. This Nyquist criterion, which is also required for time history reconstruction from periodic samples, often imposes difficult constraints on the data processing when spectral resolution requirements are added. It has been shown mathematically that time history reconstruction from random samples also requires the Nyquist criterion; but practi-

cal reconstruction with low error requires approximately three times the Nyquist rate even for nonrealizable, nonlinear, straight-line interpolation and much higher rates for optimum linear filters or sample-and-hold interpolators [10,11]. In contrast, alias-free spectrum estimation may theoretically be obtained with any mean sample rate, however small, when the random sampling process and the signal are both stationary and independent [12-14].

The economical implications of the mathematical theory are very significant for many applications in addition to the LV problem which naturally provides randomly sampled data: we need only to reduce the mean sample rate to reduce the required electronic computation speed and memory to obtain economical real-time processing.

Unfortunately, in contrast to the rigorous mathematical theory based on infinite averages, the theory of practical spectral estimation from finite data sets is in its infancy and is almost non-existant. First we must distinguish between a few different classes of problems. Singleton [15] has considered estimation of parameters of randomly samples signals known to be periodic and Shaw [16] has considered periodic sampling with random time jitter; but these are different problems. In particular, our experiments have shown that randomly sampled periodic data imposes minor requirements compared with randomly sampled broadband noise processes. Jones [17] has investigated another class of problems, which is more closely related to our concern, where independent random omissions of periodic samples occur. This type of sampling becomes similar to Poisson random sampling as the periodic sampling frequency increases and the probability of a given sample occurring goes to zero [11].

1.2.3 Estimation of Broadband Spectra from Random Samples

In the literature search only a few recent papers were found which relate directly to the problem. Adegbola [18] has theoretically analyzed one approach in detail in which the data are used to form a rectangular approximation to the Fourier integral formulas applicable to continuous data. He presents no experimental work and the estimator equation is not amenable to rapid computation. We have derived a related estimator which involves less approximation and computation. This estimator and an example of the poor experimental results which were obtained with it for subNyquist mean sampling rate are presented in Appendix I as a warning to others. We did not investigate Adegbola's estimator experimentally because it appears slow computationally.

In a very recent research study concerning spectral analysis from single-particle LV data, Asher, Scott, and Wang [19] attempted a straightforward, economical analog approach in which continuous sample and hold signals were analyzed by conventional spectral analysis instrumentation. Their unfortunate results should be considered by those who would apply the output of a frequency-tracker LV system which incorporates hold circuits during signal drop out to conventional instrumentation. It is not difficult to see by hindsight that the use of sample-and-hold is mathematically equivalent to our rectangular numerical approximation of the Fourier integrals previously mentioned. For practical single-particle LV data rates, or for trackers with appreciable drop out, we end up spectrally analyzing a random step-

function whose low-pass spectrum obliterates the desired spectrum.

Another digital approach was taken by Thompson [20] whose derivation was inadequate to predict the additive white spectral level he obtained in simulations, and whose simulations with deteriministic data are misleading in relation to the actual difficulties encountered with signals which are broadband random processes. Through a different and more exact reasoning process we have derived an estimator similar to Thompson's which correctly subtracts the additive white level. Our estimator, which is based on the known relationship between the signal spectrum and the spectrum of the samples signal [11, 21] and its experimental demonstration are also displayed in Appendix I. This estimator was not pursued because it appears less amenable to real-time computation than the approach taken.

The spectrum estimator which we derived and found most useful during the exploratory phase of this research consists of determining an autocovariance estimate from the data after discretizing the occurrence time intervals as though the sampling were periodic with most of the samples randomly missing. Our conclusion that this approach is optimum at the present time is supported by additional references by Jones [22,23] in which he describes three estimators including the discretizing autocovariance approach and concludes that this approach is the most practical of the three. Jones unfortunately makes the erroneous conclusion that once the time discretizing has been accomplished, the error analysis problem reduces to that of randomly missed

samples and references his earlier analysis [17]. This is unfortunate because the analysis referred to does not correctly predict the spectrum estimate error, as we have shown by experiment.

Further support of the discretizing autocovariance approach may be found in experimental development results currently being obtained at Lockheed-Georgia Co. under Contract No. F-33615-73-C-2032. At the time of this writing, good agreement between hot wire anemometry and LV spectrum measurements has been obtained, but the results are not yet published.*

1.3 SCOPE

In Section 2, we present equations for estimation and error prediction of turbulence parameters measurable by a single-point, single-component LV system. Section 3 presents experimental evidence which supports the error prediction equations of Section 2. Section 4 provides a guideline procedure for experiment and implementation design with further interpretation of the error prediction equations. Section 5 is a brief discussion of the results, and Section 6 provides conclusions and recommendations for further work.

---

* The results have been described in presentations by D. M. Meadows at the "Workshop on Laser Doppler Velocimetry," held at Purdue University, West Lafayette, Ind., March 27-29, 1974. Papers will appear in the Conference Proceedings.

SECTION II

THEORY

2.1 INTRODUCTION

2.1.1 Assumptions

For purposes of analysis, we consider an idealized model of the fluid flow, the LV system, and the mathematical computations. The practical significance of these assumptions is discussed later.

The velocity component U(t) is an ergodic Gaussian random process with mean[1]

$$\overline{U(t)} = \overline{U} \tag{2.1}$$

and variance or mean-square turbulence intensity,

$$\overline{u^2(t)} = \overline{(U(t) - \overline{U})^2} = \sigma^2 \tag{2.2}$$

The autocorrelation of U(t) is a real, even function

$$R(\tau) = \overline{U(t)\ U(t+\tau)} = C(\tau) + \overline{U}^2 \tag{2.3}$$

where the autocovariance $C(\tau)$ is

$$C(\tau) = \overline{u(t)\ u(t+\tau)} \tag{2.4}$$

The two sided spectrum, S(f), and the autocovariance form a Fourier transform pair

$$S(f) = \int_{-\infty}^{\infty} C(\tau) \exp\,[-j2\pi f\tau]\, d\tau \tag{2.5}$$

[1]The horizontal bar denotes statistical expectation or infinite time average interchangeably.

The LV system is assumed to detect error-free instantaneous samples U_i of the velocity at time instants t_i which are the realization of a stationary Poisson point process, statistically independent of U(t), with mean rate λ.

2.1.2 Mathematical Background

For the Poisson process the expected number of samples occurring in any interval of length T is λT, with variance λT, and the probability of m samples in T is

$$P\ (m,T) = \frac{e^{-\lambda T}}{m!}\ (\lambda T)^m \tag{2.6}$$

It proves convenient at times to divide the time axis into uniform increments of length $\Delta\tau$ which are small enough that $\lambda\Delta\tau << 1$. In this case, expansion of (2.6) as a power series shows that the probability of samples occurring in any $\Delta\tau$ interval is

$$\begin{aligned} P(0,\Delta\tau) &\simeq 1-p \\ P(1,\Delta\tau) &\simeq p = \lambda\Delta\tau \\ P(m>1,\ \Delta\tau) &\simeq 0 \end{aligned} \tag{2.7}$$

For sufficiently small $\Delta\tau$, a negligibly small timing error occurs in replacing the exact occurrence time t_i with $j\Delta\tau$, the location of the nearest uniform grid element. Using this approximation, the data sequence, discretized in time, is given by U(j) I(j) where U(j) is the value of U(t) at $j\Delta\tau$, and I(j) is an independent binary random variable with

$$\begin{aligned} P\{I=1\} &\simeq p \\ P\{I=0\} &\simeq 1-p \end{aligned} \tag{2.8}$$

The use of this notation is only justified when $p = \lambda\Delta\tau$ is small.

The concepts of equivalent power bandwidth B and turbulence time scale τ_c will prove useful:

$$2B = \frac{\int_{-\infty}^{\infty} S(f)\ df}{S(f_{peak})} = \frac{\sigma^2}{S\ (f_{peak})} \tag{2.9}$$

$$\tau_c \doteq \frac{1}{\sigma^2} \int_0^{\infty} C(\tau)\ d\tau \tag{2.10}$$

where $S(f_{peak})$ is the maximum value of the power spectrum. For a typical low-pass spectrum with maximum value at zero frequency, we note that $C(\tau)$ is even and

$$S(0) = \int_{-\infty}^{\infty} C(\tau) d\tau \tag{2.11}$$

to obtain

$$B\tau_c = 1/4 \tag{2.12}$$

2.2 MEAN AND INTENSITY ESTIMATORS

2.2.1 Introduction

Well-known statistical formulas exist for estimation of mean and variance, and for prediction of the mean-square error of these estimates, when uncorrelated samples

are used as data. However, it is contradictory to assume uncorrelated samples obtained in a finite time with random times of occurrence when we require sufficiently high data rates to estimate correlation functions!

The detrimental effects of correlated samples are as follows: the standard unbiased estimate of variance becomes biased, and the variance of the estimates increases due to the reduced information content of the data. We have obtained approximate formulas for the estimate errors which include finite-time correlation effects. These approximate formulas were obtained by assuming that in the limiting case of high λ the error could be no less than that which would be obtained from a continuous time segment of U(t). The errors for continuous estimators were derived from results in Papoulis [21] and added to the standard formulas. The details of the derivations are omitted for brevity.

2.2.2 Mean Flow

Subject to our initial assumptions of section 2.1.1 the usual sample mean

$$\hat{U} = \frac{1}{N} \sum_{i=1}^{N} U_i \tag{2.13}$$

is unbiased and has rms error given approximately by

$$\sigma_{\hat{u}} \simeq \sigma \left(\frac{1+2\lambda\tau_c}{N}\right)^{\frac{1}{2}} = \frac{\sigma}{\sqrt{N}} \left(1+\frac{\lambda}{2B}\right)^{\frac{1}{2}} \tag{2.14}$$

where the $\lambda/2B$ term results from finite time effects.

2.2.3 Mean-Square Intensity

Subject to the assumptions of section 2.1.1, the conventional variance estimator may be used if limitations due to finite-time effects are considered:

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (U_i - \hat{U})^2 \tag{2.15}$$

With $N \simeq \lambda T$ we find the continuous estimate of variance over the same duration would be biased and assume we can expect no better:

$$\overline{\hat{\sigma}^2} \simeq \sigma^2 \left(1 - \frac{2\tau_c}{T}\right) \simeq \sigma^2 \left(1 - \frac{\lambda}{2BN}\right) \tag{2.16}$$

The bias error should generally not be significant but could easily become so under some of the implementation schemes discussed later in which data is broken into small segments for processing. In such an implementation, care should be taken.

The derivation of the added finite-time variance of the mean-square intensity estimate is simplified by assuming the bias is negligible. When combined with the standard result given by Jenkins and Watts [9], the result is

$$\overline{(\hat{\sigma}^2 - \sigma^2)^2} \simeq \frac{2\sigma^4}{(N-1)} (1+2\lambda\tau_c') \leq \frac{2\sigma^4}{N-1} \left(1+\frac{\lambda}{2B}\right) \tag{2.17}$$

where

$$\tau_c' = \frac{1}{\sigma^4} \int_0^\infty C^2(\tau)d\tau \leq \tau_c \tag{2.18}$$

The inequality is valid for monotonically decreasing positive functions $C(\tau)$ but should be checked in more general cases.

2.2.4 Normalized Turbulence Intensity

We may estimate $I = \sigma/\overline{U}$ using the previous formulas as

$$\hat{I} = \frac{\sqrt{\hat{\sigma}^2}}{\hat{U}} \tag{2.19}$$

Assuming the errors in $\hat{\sigma}^2$ and $\hat{U}$ are small, Taylor series expansion shows

$$\overline{\hat{I}} \simeq I\left(1 - \frac{\lambda}{4BN}\right) \tag{2.20}$$

Assuming that the bias error is negligible, and that the errors in $\hat{\sigma}^2$ and $\hat{U}$ are uncorrelated, we find the variance approximately as

$$\overline{(\hat{I} - I)^2} \simeq \left(\frac{I^4}{N} + \frac{I^2}{2(N-1)}\right)\left(1 + \frac{\lambda}{2B}\right) \tag{2.21}$$

In low turbulence measurements, the I^4 term will be negligible in comparison with the I^2 term.

2.3 AUTOCOVARIANCE

2.3.1 The Estimators

The method we have found most useful for spectrum analysis begins with a discretized autocovariance estimate similar to that reported by Jones [22, 23]. The exact time differences between pairs of samples are represented approximately as integral binary numbers with an arbitrary time increment $\Delta\tau$. For each pair of samples for which the discretized time difference does not exceed the preselected maximum lag number M, the product of the two samples (lag product) is accumulated in a 1xM matrix SUM(k) and 1 is added to a corresponding histogram H(k). For purposes of analysis we assume that the true mean has been previously subtracted from the data. After N samples are processed the resulting estimate $\hat{C}(k)$ of the value of $C(\tau)$ at $\tau=k\Delta\tau$, $k\neq 0$, is

$$\hat{C}(k) = \frac{1}{H(k)} \; SUM(k) \tag{2.22}$$

which the summation includes only the H(k) products for

$$\left| \frac{|t_i - t_j|}{\Delta t} - k \right| < 0.5 \tag{2.23}$$

For k=0, the practical advisability of including products for which $|t_i - t_j| < \Delta\tau/2$ along with the u_i^2 terms depends on the implementation; without these terms, H(0) = N and

$$\hat{C}(0) = \frac{1}{N} \sum_{i=1}^{N} u_i^2 \tag{2.24}$$

A very similar autocovariance estimate may also be expressed in the alternate notation which assumes assignment of t_i to the nearest value of $j\Delta\tau$ prior to computation as

$$\hat{C}'(k) = \frac{\sum_{j=0}^{N'-k} u(j)\ u(j+k)\ I(j)\ I(j+k)}{\sum_{j=0}^{N'-k} I(j)\ I(j+k)} \qquad (2.25)$$

when $N \simeq \lambda T = \lambda N' \Delta t$ and $I(j)$ is the binary indicator sequence. The estimator is denoted with a prime to remind us that (2.25) differs from (2.23) and (2.24) because $\pm\Delta\tau/2$ errors in the discretization of t_i and t_j prior to subtraction allows maximum errors in time differences of $\pm\Delta\tau$. The maximum time error for $\hat{C}(k)$ is $\pm\Delta\tau/2$. Also $\hat{C}(k)$ uses a fixed number of samples, while $\hat{C}'(k)$ uses a fixed length of time. The relation $N = \lambda T$ is only true on the average.

2.3.2 Bias Error Due to Time Increment

Jones dismisses the analysis of the discretized autocovariance estimate by reference to his analysis of periodic sampling with random omissions [17]. If, in fact, $\Delta\tau$ is assumed small enough to make the time errors negligible in both $\hat{C}(k)$ and $\hat{C}'(k)$ then the application of conditional expectation techniques to the random omission model shows that the estimates are unbiased provided that N is sufficiently large so that each lag value has at least one lag product. We note, however, that practical limitations

will sometimes require $\Delta\tau$ to be as large as is permissible without bias error. In the case of periodic sampling with random omissions $\Delta\tau$ must satisfy the Nyquist criterion.

In Appendix II we demonstrate that the expected value of $\hat{C}(k)$ is

$$\overline{\hat{C}(k)} = \int_{-\infty}^{+\infty} C(\tau)\, p_k(\tau)\, d\tau \qquad (2.26$$

where $p_k(\tau)$ is the conditional probability density for occurrence of lag delays and is nonzero only on the interval $(k\Delta\tau - \Delta\tau/2,\ k\Delta\tau + \Delta\tau/2)$. The function $p_k(\tau)$ has unit area and is not given by a simple formula. The somewhat complicated exact expression for $p_k(\tau)$ is given in Appendix II for the case of uniform Poisson random sampling.

As we show in section 2.3.4 below, in many practical cases $p_k(\tau)$ may be replaced by a constant $1/\Delta\tau$ over the interval of definition with the result that the expected value of $\hat{C}(k)$ is the average value of $C(\tau)$ on the surrounding $\pm\Delta\tau/2$ interval.

The limitation imposed by selection of $\Delta\tau$ is best understood by its effect on the spectrum estimate and will be discussed further in section 2.4. The resulting maximum value of $\Delta\tau$ is of the same order of magnitude as the Nyquist criterion for periodic sampling. The bias error for $\hat{C}'(k)$ is expected to be worse than that for $\hat{C}(k)$, but on the same order of magnitude.

2.3.3 Additional Bias Error

For the present we assume the true mean $\bar{U}$ has been subtracted from the samples before formation of the estimate $\hat{C}(k)$. In implementation it is actually necessary to subtract sample means to avoid low-frequency distortions of the spectrum estimate. This procedure is expected to produce small bias errors in the autocovariance estimate which are associated with removal of the very low-frequency components, but the effect has not yet been analyzed in detail.

2.3.4 The Histogram

The histogram H(k) of the number of lag products obtained at each lag value k plays a significant role in practical application of the estimator $\hat{C}(k)$. In Appendix II we have evaluated its expectation approximately under the assumption that $\lambda\Delta\tau$ is small and N is large with the result that with only u_i^2 terms used in $\hat{C}(0)$,

$$H(k=0) = N$$

$$\bar{H}(k) \simeq N\lambda\Delta\tau - k(\lambda\Delta\tau)^2 \simeq \left(\frac{T}{\Delta\tau} - k\right)(\lambda\Delta\tau)^2 \tag{2.27}$$

The result shows that the number of products at each lag is generally much less than the number of u_i^2 terms. The expected number of products at each lag decreases linearly with k, but if the data collection time, T, for each data segment, is much greater than the maximum delay, i.e., if

$$T \gg M\Delta\tau, \quad N \gg M\lambda\Delta\tau \tag{2.28}$$

then the value of $\bar{H}(k)$ is nearly constant. This condition is an important consideration in implementation which must be observed when segmented data methods are used.

The variance of H(k) has not been theoretically evaluated, but we expect it to behave in the manner of a Poisson process for acceptably large values; i.e.

$$\overline{(H(k) - \bar{H}(k))^2} = \bar{H}(k) \tag{2.29}$$

We may obtain an approximate expression for the function $p_k(\tau)$ utilized in section 2.3.2 from $\bar{H}(k)$ by assuming that H(k) is an unnormalized estimate of the probability density for occurrences of lag values on the range $0<\tau<M\Delta\tau$. The result is

$$p_k(\tau) \simeq \frac{N\lambda - \lambda^2\tau}{N\lambda\Delta\tau - k\lambda^2\Delta\tau^2} \tag{2.30}$$

It is expected, therefore, that $p_k(\tau)$ may be replaced in (2.26) by a unit area rectangular function of width $\Delta\tau$ and height $1/\Delta\tau$ when the data collection time is much greater than the maximum delay.

2.3.5 Variance of the Autocovariance Estimate

When the system parameters are correctly chosen to avoid bias error, the important remaining problem is the determination of the statistical error of the estimate in terms of the amount of data collected and processed.

In Appendix II we derive the variance of $\hat{C}(k)$ under the assumptions that the normalized rate parameter $\lambda/2B$ is small compared with unity and the variance of H(k) is small.

$$\sigma^2_{ck} = \overline{(\hat{C}(k) - C(k\Delta\tau))^2}$$

$$\simeq \frac{1}{\bar{H}(k)} [\sigma^4 + C^2(k\Delta\tau)] \qquad (2.31)$$

$$k \neq 0$$

The value of $\hat{C}(0)$ is identical with the previously given mean-square intensity estimate if the same sample mean is subtracted, and the previous result with low mean sample rate and large N reduces to the usual formula:

$$\sigma^2_{co} = \frac{2\sigma^4}{N} \qquad (2.32)$$

2.4 POWER SPECTRUM

2.4.1 Computation From the Autocovariance

Once the discrete autocovariance estimate $\hat{C}(k)$ has been computed, the two-sided power spectrum estimate follows using a discrete Fourier transform [4,7] as

$$\hat{S}(i\Delta f) = \Delta\tau \left[\hat{C}(0) + 2\sum_{k=1}^{M-1} w(k)\ \hat{C}(k)\ \cos\left(\frac{ik\pi}{M}\right)\right] \qquad (2.33)$$

$$i = 1, 2, \ldots M$$

where

$$M = \frac{\tau_{max}}{\Delta\tau}$$

$$\Delta f = \frac{1}{2M\Delta\tau} = \frac{1}{2\tau_{max}}$$

$$f = i\Delta f$$

and w(k) is a smoothing window function which is unity at k=0 and which decreases to zero at k=M. The discrete transform may be implemented to a fast Fourier transform with certain precautions regarding the addition of zeros as discussed by Brumbach [7].

Several observations are appropriate. Details concerning these observations may be found in the literature [2,4,7,9,17] since they are common to conventional digital spectrum estimation. First, $\hat{S}(f)$ is an even periodic function with period $1/\Delta\tau$. Since the estimate replicates itself for $f>1\Delta\tau$, the usual Nyquist criterion is a necessary condition on $\Delta\tau$ to avoid aliasing distortion, just as it is in periodic sampling. The additional constraint on $\Delta\tau$ mentioned in section 2.3.2 is discussed further below.

Optimum selection of the shape of the window function is an art involving knowledge of the spectrum shape (either *a priori* or by trial and error procedure) and the measurement objectives. For smooth broadband turbulence spectra without spikes due to periodic components the simple Bartlett window given by

$$w(k) = 1 - \frac{k}{M}, \quad k < M \qquad (2.34)$$
$$= 0 \qquad \text{otherwise}$$

should be adequate. The effect of the window is a running weighted average, or convolution, in the frequency domain which reduces the error variation of the estimate. At the same time, bias error is introduced in the form of loss of spectral resolution and the expected value of the estimate, assuming $\hat{C}(k)$ is unbiased, is

$$\bar{S}(f) = S(f)*W(f) \tag{2.35}$$

where * denotes convolution and W(f) is the Fourier transform of $w(\tau)$. For the Bartlett window given in (2.34)

$$W(f) = M\Delta\tau \left[\frac{\sin(\pi f M\Delta\tau)}{\pi f M\Delta\tau} \right]^2 \tag{2.36}$$

the "frequency resolution", R, may be taken as $1/M\Delta\tau$, the spread between the peak and the first zero of W(f).

Several other window functions having better resolution and/or side-lobe characteristics are described in the literature. The "Hamming" window is a common compromise favorite which performs better than the Bartlett window in most cases. It is given by

$$W(k) = 0.54 + 0.46 \cos\left(\frac{\pi k}{M}\right), \quad k < M \tag{2.37}$$

$$= 0 \quad \text{otherwise}$$

A final observation concerns the temptation on the part of some to compute S(f) at frequency intervals less than $1/(2\Delta\tau)$ because the result is smoother and more pleasing looking than the somewhat jagged appearance of the result of using (2.33). Giving into this temptation does not produce more useful information or resolution since the width of W(f) generally exceeds $1/(2M\Delta\tau)$. On the other hand, natural spectra are usually "smooth", and the jagged appearance resulting from errors is not disguised when the frequency spacing is maintained large enough for the estimates to remain nearly independent.

2.4.2 Bias Effect of $\Delta\tau$ Selection

Under the conditions $\lambda\Delta\tau << 1$ and $N >> M\lambda\Delta\tau$, result (2.26) becomes

$$\bar{\hat{C}}(k) = \int_{-\infty}^{\infty} C(\tau) \frac{1}{\Delta\tau} \text{Rect}\left(\frac{k\Delta\tau - \tau}{\Delta\tau}\right) d\tau \tag{2.38}$$

where

$$\begin{aligned} \text{Rect}(\tau) &= 1, \quad -1/2 < \tau < 1/2 \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{2.39}$$

Equation (2.38) is a convolution integral which produces the effect in the spectrum estimate of multiplication by the transform of the Rect function. Manipulation with Fourier transform relations yields

$$\bar{\hat{S}}(f) = \sum_{-\infty}^{\infty} S\left(f - \frac{n}{\Delta\tau}\right) \frac{\sin\left[\pi\Delta\tau\left(f - \frac{n}{\Delta\tau}\right)\right]}{\pi\Delta\tau\left(f - \frac{n}{\Delta\tau}\right)} \tag{2.40}$$

Equation (2.40) reveals the usual nature of aliasing error wherein components of $S(f) \geq 1/2\Delta\tau$ fold and add to the estimate. The equation also shows the attenuation of all spectral components by a (sin x)/x function which is down to a factor of 0.637 at the $1/2\Delta\tau$ folding frequency.

2.4.3 Effect of Non-Zero Mean and Linear Trends

The finite resolution imposed by the finite width of the spectral window function ($M<\infty$) means that any discrete frequency component in the data, which ideally would produce an impulse in the spectrum, will result in a replication of the spectral window at the location of the frequency spike with bias distortion of nearby estimates. Even in the absence of periodic velocity fluctuations, the steady mean velocity component has the same effect at the frequency origin. The presence of a large linear trend in the data which results from a sinusoidal component too low in frequency to be resolved also has the same effect. For this reason, it is common spectral estimation practice to remove both the sample mean from the data and to further correct for linear trends [4].

The necessity for linear trend removal in turbulence measurements does not seem apparent, and the theory concerning such a step in the case of random sampling has not been considered. In addition, the exact nature of the effect of removing the sample mean is also slightly unclear when short data segments are used. We believe that, when the data segments are sufficiently long to satisfy other requirements, any low frequency bias errors

resulting from subtraction of the sample mean instead of the true mean will be confined primarily to the zero frequency estimate with some leakage to the first frequency estimate ($f=\Delta f$) through the greater than $1/(2M\Delta\tau)$ width of the spectral window.

2.4.4 Variance of the Spectral Estimate

In Appendix II we derive the variance of the spectrum estimate under the assumption that $\lambda/2B << \lambda M\Delta\tau < 1$. The derivation also assumes that $\Delta\tau$ is small enough, M and N are large enough, and the true mean has been subtracted so that the resulting estimate is unbiased and the histogram H(k) is approximately a constant given by $N\lambda\Delta\tau$. The result is

$$\sigma_f^2 = \overline{(\hat{S}(i\Delta f)-S(i\Delta f))^2}$$

$$= \frac{2\Delta\tau}{N\lambda} \sum_{k=0}^{M} w^2(k) \; [\sigma^4 + C^2(k\Delta\tau)] \; [1 + \cos\left(\frac{2\pi ik}{M}\right)] \tag{2.41}$$

This result shows the mean square error has a maxima at the zero frequency and the folding frequency with values which decrease to a value approximately one half as large for frequencies away from i=(0,M).

If the summation is viewed as a numerical integration, the variance away from the frequency origin may

be evaluated approximately by neglecting the small contribution from the $C^2(k\Delta\tau)$ term as

$$\sigma_f^2 \simeq \frac{2\sigma^4}{N\lambda} \int_0^{M\Delta\tau} w^2(\tau)d\tau \tag{2.42}$$

which for the Bartlett window gives

$$\sigma_f^2 \simeq \frac{2\sigma^4 M\Delta\tau}{3N\lambda} \tag{2.43}$$

This result shows clearly the price paid for increasing $M\Delta\tau$ to obtain better spectral resolution.

We may compare our error prediction result with that given by Jones [17] for spectrum estimation with missing observations. Putting his results in the same notation and using our result for the value of the histogram H(k) we obtain

$$\sigma_{fj}^2 \simeq \frac{2S^2(f)}{N\lambda\Delta\tau} \sum_0^M w^2(k) \tag{2.44}$$

where replacement by a continuous integral gives

$$\sigma_{fj}^2 \simeq \frac{2S^2(f)}{N\lambda\Delta\tau^2} \int_0^{M\Delta\tau} w^2(\tau)d\tau \tag{2.45}$$

This result agrees exactly with ours under the condition that the true spectrum is white so that $\sigma^2 = S(f)/\Delta\tau$. Our result, however, indicates that the error level is not proportional as a function of frequency to the true spectrum as Jones' result implies.

The usual practice of indicating a constant confidence interval on a logarithmic spectrum plot is obviously not appropriate when the error level is not

proportional to the spectrum itself. It now becomes more appropriate to indicate a constant confidence interval on a linear spectrum plot. Supporting evidence for this statement in addition to our own may be found in careful observation of the digital simulation results in Jones' previously referenced paper. Unfortunately our analysis is supported by experiment with rather grim consequences for the estimation of low level spectral tails.

2.4.5 Post Estimate Smoothing

Nature has provided some assistance in the dilemma which results when the constraints of resolution, error, and measurement time are all considered. For many phenomena, a constant resolution on a log-frequency plot is more meaningful than is the constant resolution which results in the preceding discussion. Various computation-saving methods of achieving resolution at least appoximately proportional to frequency which reduce error variation at high frequencies have been devised for conventional spectral analysis [4,24].

Computationally efficient methods of achieving non-uniform spatial resolution, appropriate for random sampling should be pursued, perhaps by extension of low-pass digital filtering to randomly sampled data [25], but in their absence the error benefits may still be obtained by additional post-estimation smoothing. This is accomplished by replacing each spectral estimate with an average over a number of the original neighboring estimates

proportional to the frequency of the estimate. For example, if Δf is 100 Hz, the resolution is 200 Hz, and there are 500 estimates from 100 Hz to 50 KHz, it might be appropriate to retain no better than 10% resolution. This would result in leaving the first 10 estimates unchanged, averaging ten estimates on either side along with the 10 KHz estimate, and averaging 45 estimates on either side with the 45 KHz estimate.

If all of the spectral estimates were uncorrelated, the variance of the post-smoothed estimate would be reduced by the inverse of the number of estimates averaged. Since the resolution generally exceeds the Δf spacing, the actual reduction in variance is less due to correlation between adjacent estimates. This effect has not been analyzed in detail, but a factor of 2 seems conservative when the shape of the Bartlett window is considered.

2.5 PROCESSOR DEAD TIME EFFECT

Our idealized assumptions do not provide for the significant fact that all LV burst-counter processors exhibit dead time, d, during which the processors make the "instantaneous" sample measurement and verifies the validity of the measurement with various data checks which have been devised. The significance lies in the fact that the minimum[2] value of $\Delta\tau$ is 2d if $\hat{C}(1)$ is not to be biased. It would thus seem that the electronic signal processor dead time limits the highest spectral frequencies which may be observed.

---

2 This constraint could be reduced slightly by restricting the band of delays to less than $\pm\Delta\tau/2$ about each value $k\Delta\tau$ at the expense of loss of data and more complex data processing requirements.

The minimum possible processor dead time is bounded by the LV system parameters: it is the time required for passage of a scattering particle through the probe volume.

2.6 SEGMENTED DATA - NONSTATIONARITY

The ergodic assumption includes statistical stationarity of both the turbulence and the random sampling process. In many practical situations one or both stationarity assumptions will be violated. The analysis of nonstationary random processes is difficult, and generally requires characterization models and parameters which are not available. Evans and McCarty [26] have discussed such problems at length for the case of periodically sampled data.

The problem we face is that in implementation we must choose between obtaining, storing and processing one large amount of data N as one continuous record; or computing many estimates from short data segments and averaging the results; or doing something that is a compromise. We may further subdivide the choices by pointing out that the results will be different if data segments are chosen to be of equal length in time, $N'\Delta\tau$, or equal length in number of data words N. Under the stationarity assumptions our theory is directly applicable to individual segment estimates with identical numbers of samples N. The succeeding averaging of Ns segments will not affect the predicted bias errors, but it will reduce variance in a manner simply obtained by replacing N by NNs in the equations. It is expected that differences in implementations

will be minor when the random process are stationary, and that approximately correct predictions will be obtained by replacing $N \approx \lambda T$ when it is desirable to do so for design purposes.

In the absence of theoretical guidelines, it appears advantageous to make use of the fact that in many nonstationary situations the statistics may be "slowly varying" in such a manner that the processes are quasi-stationary over short time segments. Under these ill defined conditions the use of statistical averages may provide valid answers in terms of the time dependent statistics, although the variance of the individual estimates may be unusably large and the effect of averaging the individual estimates to obtain a long-time estimate is not known.

SECTION III

EXPERIMENTAL

3.1 INTRODUCTION

An experimental investigation was conducted in parallel with theoretical studies to guide the direction of the theoretical research and to verify results as they were derived. A hybrid electronic simulation facility was designed and constructed to produce randomly sampled experimental data for the investigation and computer programs were written to process the data in records (segments) of 250 samples each.

Several stages of experimental work followed. The first stage involved a comparison of the performance of different spectrum estimators (see Appendix I) and the latter stages were studies of a particular estimator (see section 2.4) with emphasis on how its error behaved as a function of average sample rate and bandwidth as well as the estimator parameters.

A detailed description of the simulation facility, the data processing procedures and a copy of the subroutine RASPEC implementing the spectrum estimator of Section 2.4 are all included in Appendix III. A description of the main experimental results follows.

3.2 RESULTS

3.2.1 Preliminary Results

Initially, three spectrum estimators were programmed and tested with the same randomly sampled data. Brief descriptions of the first two estimators and the test results are given in Appendix I. The third estimator, which gave the least qualitative error with less computational effort, is described in Section 2.4.

Other experimental work followed using the third estimator. A low frequency sinusoid (200 Hz) was randomly sampled at two different average sample rates and this data produced clearly discernable spikes at the fundamental frequency with only 250 data samples. Both bandpass and lowpass random processes, generated by passing white noise through controllable known filters, were then randomly sampled for a number of average sample rates, ranging from below the Nyquist rate to above the Nyquist rate.

3.2.1.1 Experimental Error as a Function of the Number of Data Words

The decrease in experimental spectrum estimate error, $\hat{\varepsilon}_{1s}$, with an increase in the number of data words processed is shown in Figure 1, a. and b., where

$$\hat{\varepsilon}_{1s} = \frac{\left[\dfrac{\sum_{i=50}^{200} \hat{S}^2(i\Delta f)}{151}\right]^{\frac{1}{2}}}{\hat{S}(f_{peak})} \tag{3.1}$$

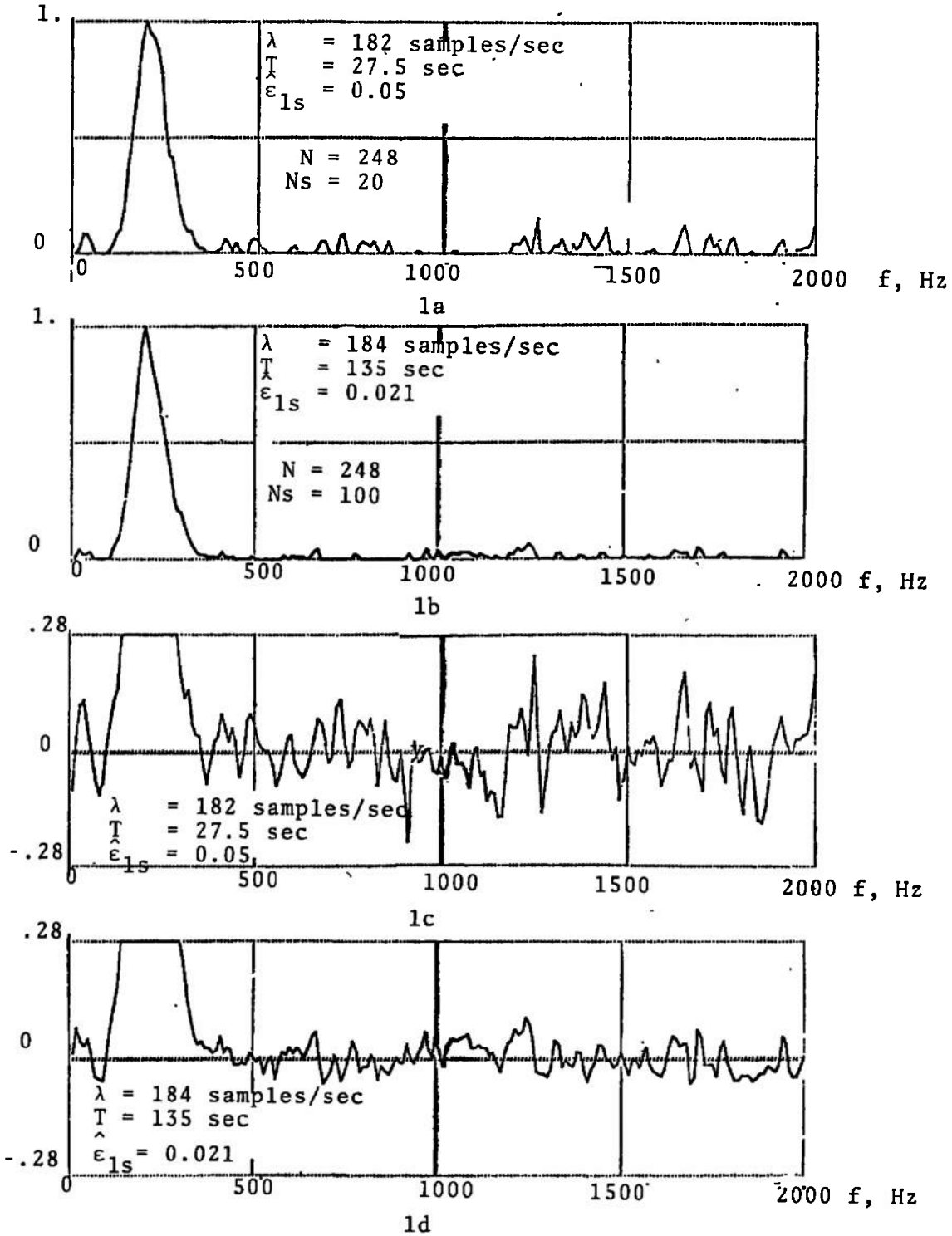


Figure 1. Variation of Spectrum Estimates with Number of Data Samples.

The subscript s indicates that this is the error in the stopband which, in this case, is the band from 500 Hz to 2000 Hz where the true spectrum is assumed to be zero. There are 151 estimates in this band at 10 Hz intervals. Qualitatively, $\hat{\varepsilon}_{1s}$ is the square root of the average error squared, normalized to the peak of the spectrum estimate.

The same pair of estimates, plotted on a different scale (Figure 1, c. and d.), indicates that the estimator produced negative and positive errors that are symmetrical about the true spectrum. This is clearly seen in the stopband.

3.2.1.2 Experimental Error as a Function of Mean Sample Rate

In the series of estimates shown in Figure 2, the number of data words (N=248) and the number of data segments (Ns=20) are held constant and the mean sample rate, λ, is increased. The reduction in $\hat{\varepsilon}_{1s}$ is indicated, where $\hat{\varepsilon}_{1s}$ is defined just as in the previous section. Note that in the main lobe of the bandpass spectrum, the estimate error qualitatively appears to increase for the higher values of λ where $\alpha = \lambda/2B >> 1$.

3.2.1.3 Aliasing Error and the Effect of Smoothing

Aliasing error is shown in Figure 3. In Figure 3a., the parameter $\Delta\tau$ was purposefully selected large enough to cause the spectrum estimate to replicate itself within the scale of the plot. A further increase in $\Delta\tau$ results in the increased aliasing of Figure 3b. The error reduction

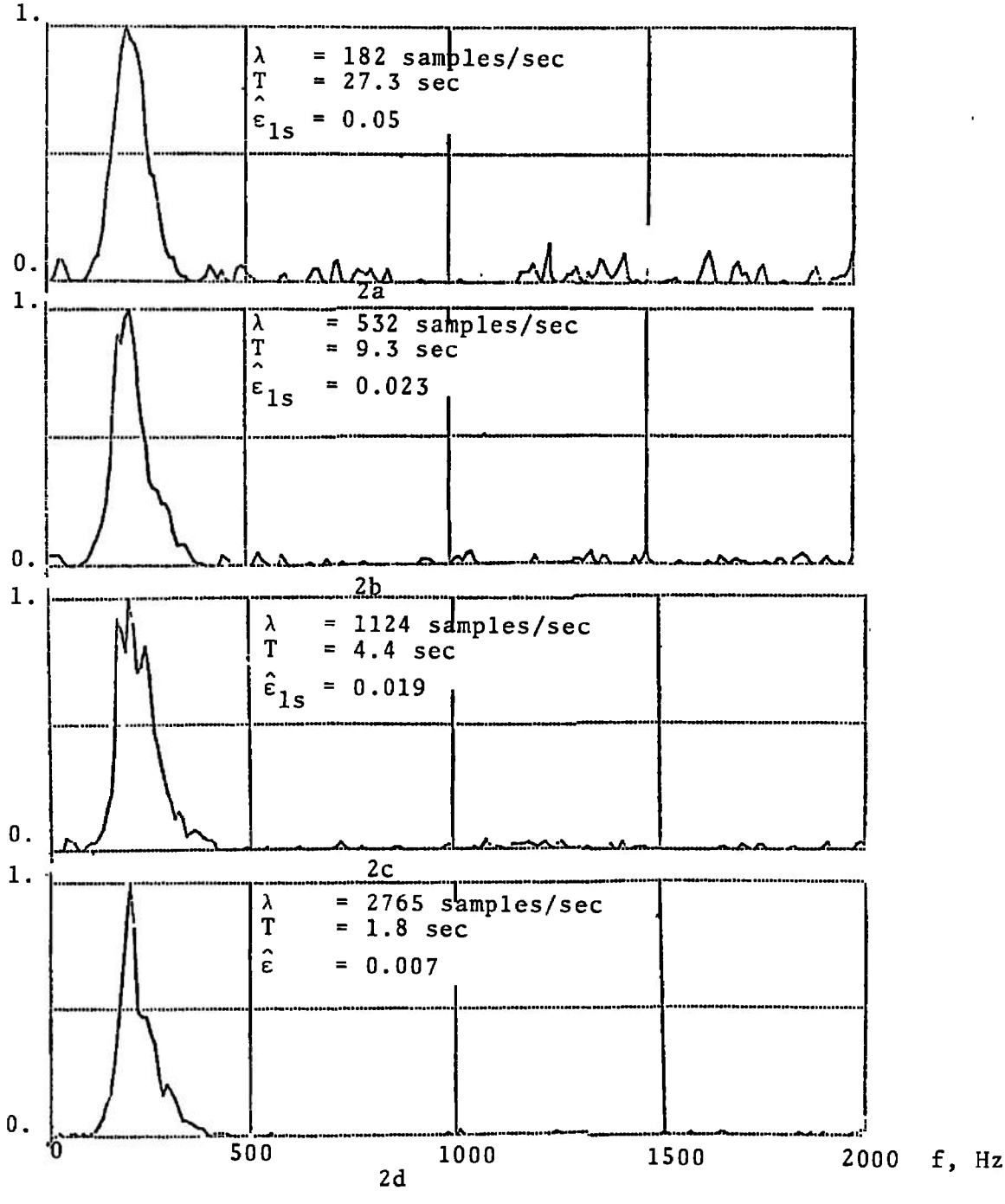
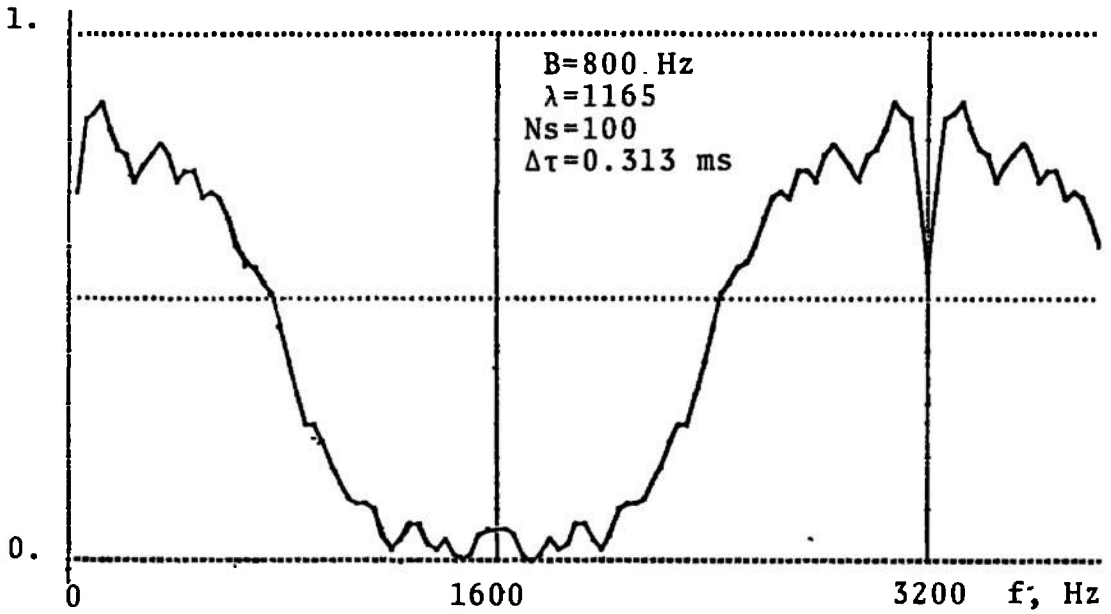


Figure 2. Variation of Spectrum Estimates with Mean Data Rate.



3a

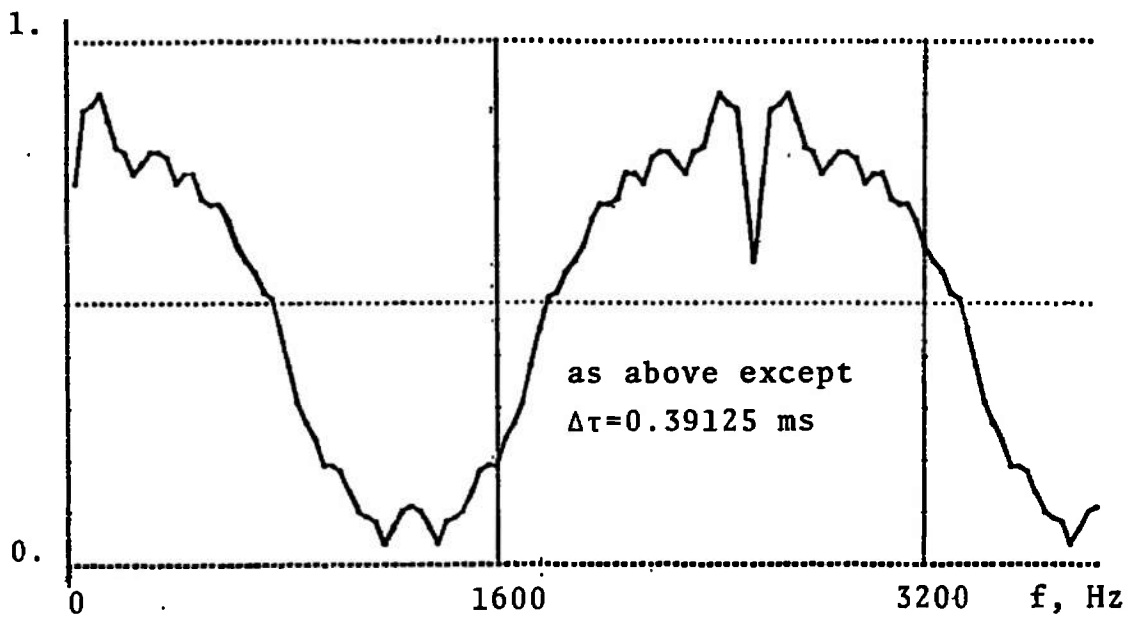


Figure 3. Spectrum Estimates Exhibiting Aliasing.

effect of decreasing the width of the window w(k) is shown in Appendix I, Figure I-4.

3.2.2 Parametric Variability Study

One of the main theoretical results was the derivation of an expression for the variance of the spectrum estimator of Section 2.4. Accordingly, an experimental effort was undertaken to verify this expression (equation (2.43)) by plotting theoretical error and experimental error as a function of the mean rate parameter, $\alpha = \frac{\lambda}{2B}$.

To obtain a range of values for α, random processes with two different lowpass spectra were used, one with B=100 Hz and one with B = 800 Hz (nominal values). Four different average sampling rates were used for each of the two spectra to give eight values of the parameter α. This experimental data was then used to compute a "high accuracy" spectrum estimate for each bandwidth. For this, four 20 data segment spectrum estimates were computed for each value of λ. These estimates were averaged at each point in frequency to form an averaged estimate for each value of λ. These four averaged estimates (from the four different values of λ) were again averaged at each point in frequency to form the overall "high accuracy" estimate for that bandwidth. The "high accuracy" estimates were then fitted with the ideal Butterworth spectrum (fourth order) for each bandwidth using a least-mean squares program to determine the

level of the Butterworth curve in the passband. Wave analyzer data showed that frequencies below about 20 Hz were attenuated in the actual spectrum and so it was not a true Butterworth lowpass spectrum. Consequently, the first six points of the estimates were not used in the least-mean squares program. The plots of Figure 4 show the "high accuracy" estimates and least-mean square fit Butterworth curves for each bandwidth.

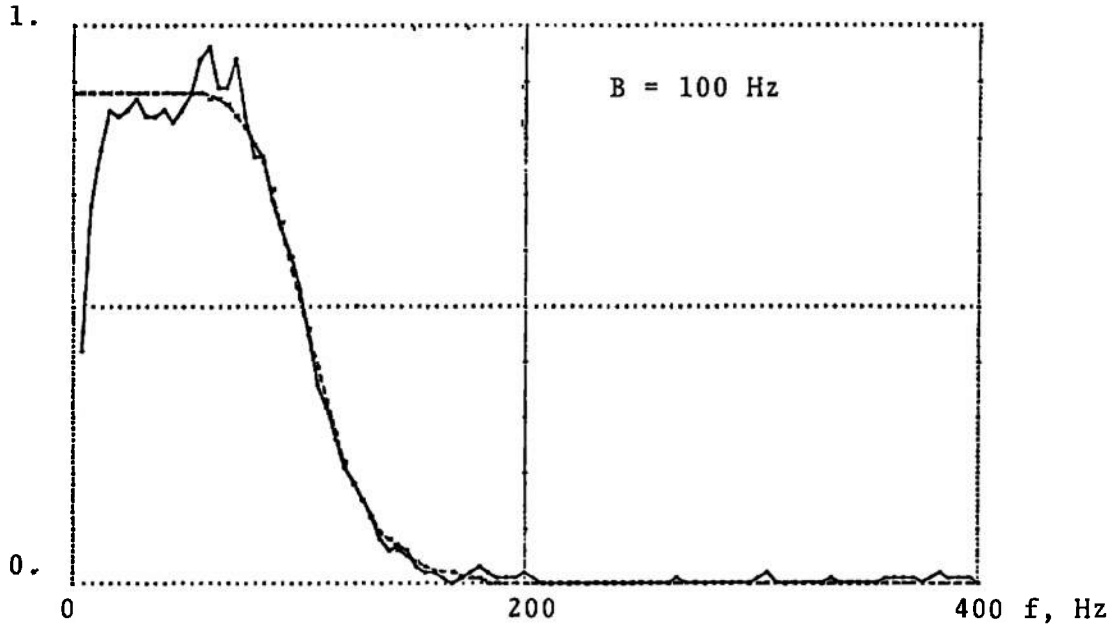
The next step involved was to compute experimental error by computing error between a number of spectral estimates and the Butterworth curves described above. For each value of α, 80 data segments were used in computing estimates; one estimate from the first 20 data segments, and so forth for a total of four estimates. This resulted in 32 spectral estimates. The experimental error was defined over two frequency bands for each bandwidth. The bands were:

Passbands

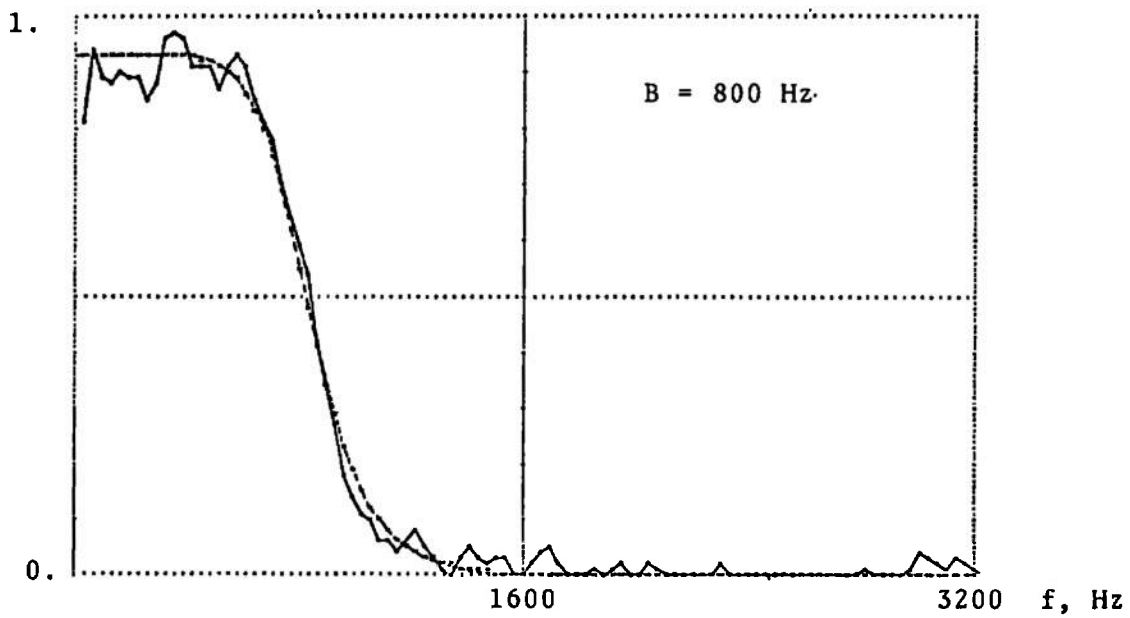
$24 \text{ Hz} \leq f \leq 104 \text{ Hz}$	100 Hz filter
$204 \text{ Hz} \leq f \leq 800 \text{ Hz}$	800 Hz filter

Stopbands

$204 \text{ Hz} \leq f \leq 00 \text{ Hz}$	100 Hz filter
$1632 \text{ Hz} \leq \quad \leq 6400 \text{ Hz}$	800 Hz filter



4a



4b

Figure 4. Analytical Butterworth Spectra and "High Accuracy" Spectrum Estimates.

where the upper and lower limits were determined by counting off a discrete number of points of frequency spacing Δf.

The general expression for the experimental error in the passband, $\hat{\varepsilon}_{1p}$, was

$$\hat{\varepsilon}_{1p} = \frac{\left[\frac{1}{n} \sum_{i=m}^{n+m-1} (\hat{S}(i\Delta f) - S(i\Delta f))^2\right]^{\frac{1}{2}}}{S(0)} \tag{3.2}$$

where $\hat{S}(i\Delta f)$ is the value of a spectrum estimate at frequency $f = i\Delta f$.

For the 100 Hz filter,

$\hat{S}(i\Delta f)$ = value of the analytical model (100 Hz filter) at frequency $f = i\Delta f$.

$S(0)$ = value of analytical model (100 Hz filter) in the low frequency portion of the passband

n = number of points in the passband over which the summation is taken = 20

m = first point in the passband = 6

Similar definitions apply for the experimental error in the passband of the 800 Hz filter.

The general expression for the experimental error in the stopband, $\hat{\varepsilon}_{1s}$, is defined similarly to the experimental

error in the passband, except the summations are taken over n = 149 and m = 51 in the stopbands of both filters.

By applying these definitions to the 32 spectrum estimates, two plots were generated, one for $\hat{\varepsilon}_{1s}$ vs α, the normalized mean rate parameter, (Figure 5) and one for $\hat{\varepsilon}_{1s}$ vs α (Figure 6). For each value of α, four values of $\hat{\varepsilon}_{1p}$ were computed, one from each of the four spectrum estimates. These were then averaged to form $\hat{\varepsilon}_{1p\text{-}av}$ and this is plotted as a solid line in Figure 5. The values of $\hat{\varepsilon}_{1p}$ from the first 20 records, $\hat{\varepsilon}_{1p\text{-}1}$, and the second 20 records, $\hat{\varepsilon}_{1p\text{-}2}$, are shown as dashed lines in Figure 5 to show typical variability of the $\hat{\varepsilon}_{1p}$ values about their average.

To compare theory and experimental results, equation (2.43) was plotted in normalized rms form:

$$\varepsilon_{1p} = \left(\frac{2}{3}\right)^{\frac{1}{2}} 2B \left(\frac{M\Delta\tau}{N\lambda}\right)^{\frac{1}{2}} \tag{3.3}$$

where $\sigma^2 = 2BS(f_{peak})$ has been substituted using the normalized mean rate parameter $\alpha = \frac{\lambda}{2B}$. The values of the parameters used in the experiment were substituted in the equation for computations.

The description of Figure 6, the plot of $\hat{\varepsilon}_{1s}$ vs α, is essentially the same as for Figure 5 but the results are somewhat different. The plot shows much less variability between the empirical average $\hat{\varepsilon}_{1s\text{-}av}$ and $\hat{\varepsilon}_{1s\text{-}1}$ or $\hat{\varepsilon}_{1s\text{-}2}$ and also shows much better agreement between ε_{1s} (the theoretical curve) and $\hat{\varepsilon}_{1s\text{-}av}$. The agreement between the empirical and theoretical is good even for relatively large values of α.

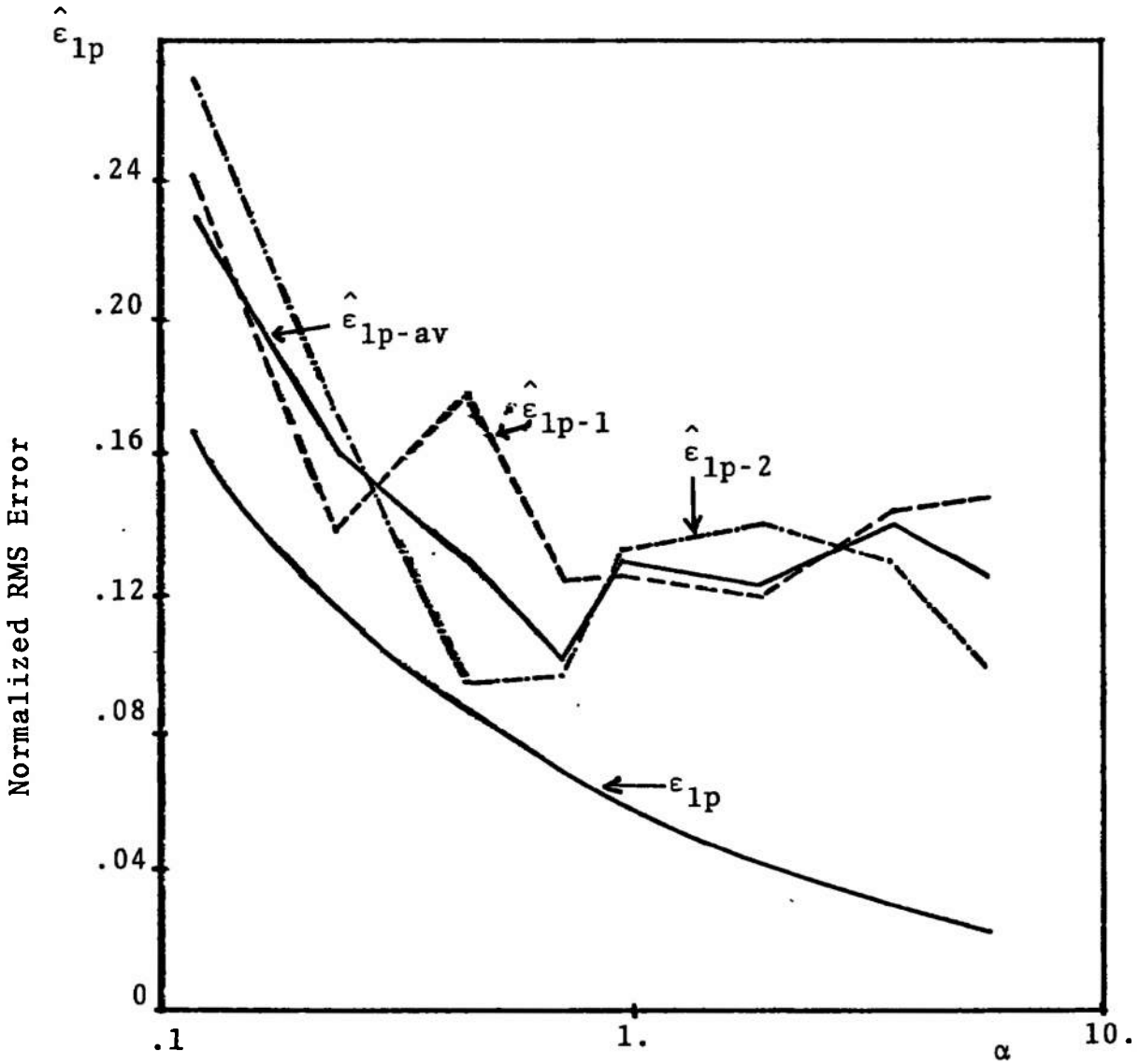


Figure 5. Spectrum Estimate Error in the Passband - Theoretical and Experimental.

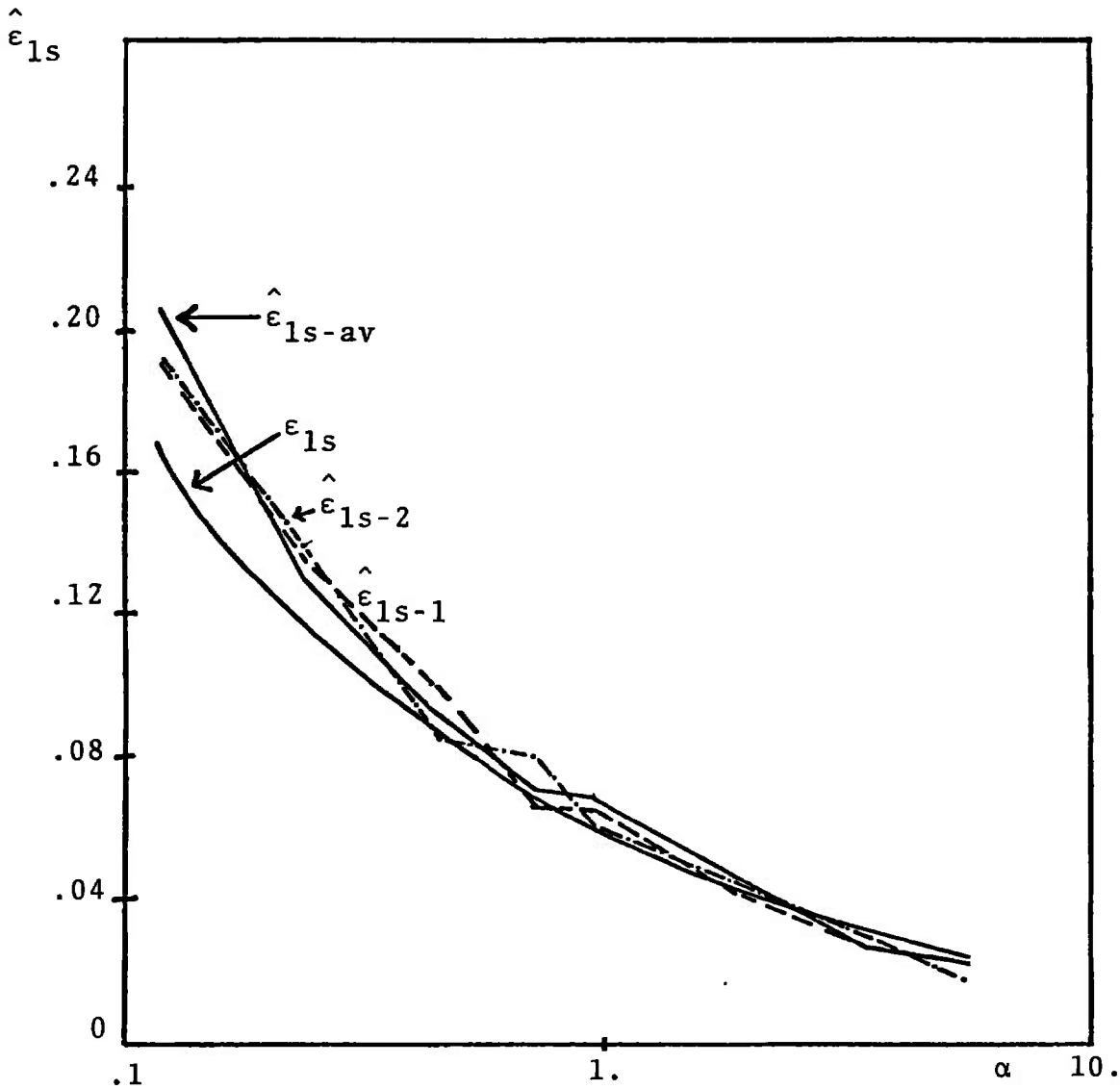


Figure 6. Spectrum Estimate Error in the Stopband - Theoretical and Experimental.

3.2.3 Parametric Study of Finite Time Correlation Effects on Mean and Turbulence Intensity Variability

Experiments similar to those described in Section 3.2.2 were conducted to determine whether or not the increase of varability due to finite time correlation effects were important for fixed N. The results are shown in Figures 7 and 8 where experimental measures of the normalized RMS deviations of the mean and turbulence intensity estimates are plotted versus α for N=248 and Ns=400. The theoretical curves are plots of equation (2.14) (mean estimate) and equation (2.21) (turbulence intensity estimate) in normalized form. It is seen that estimate error did increase for $\alpha > 1$ but was less than predicted. The theoretical equations appear conservative and further work is required for exact results.

3.2.4 Histogram Error

Figure 9 shows the finite time effects on the histogram H(k) of lag products as α increases. Figure 9a shows the histogram H(k) and its expected value, equation (2.27), for comparison. Figure 9b shows the histogram H(k) and equation (2.27) under the condition of equation (2.28).

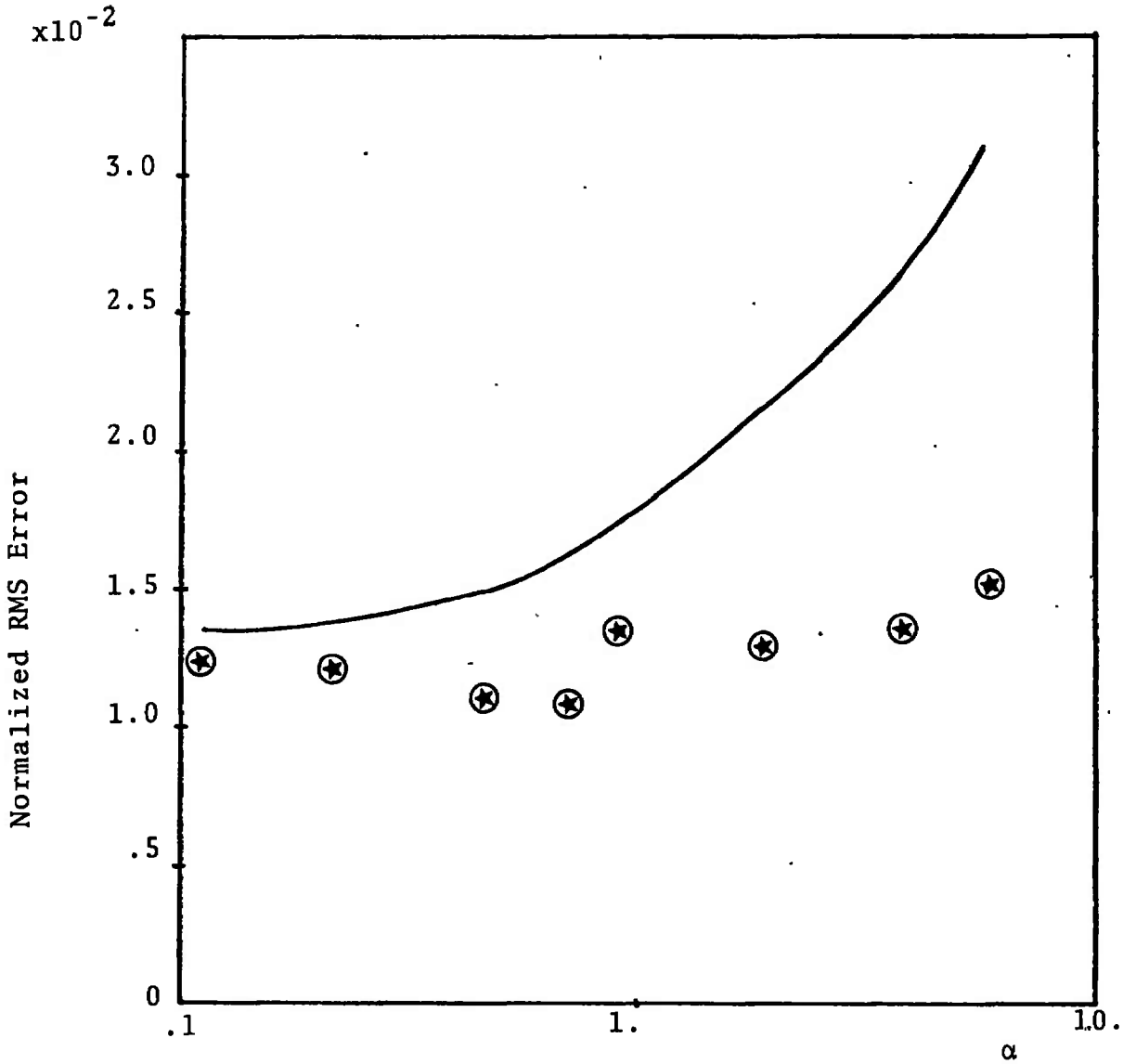


Figure 7. Normalized RMS Error of the Mean Estimate vs α.

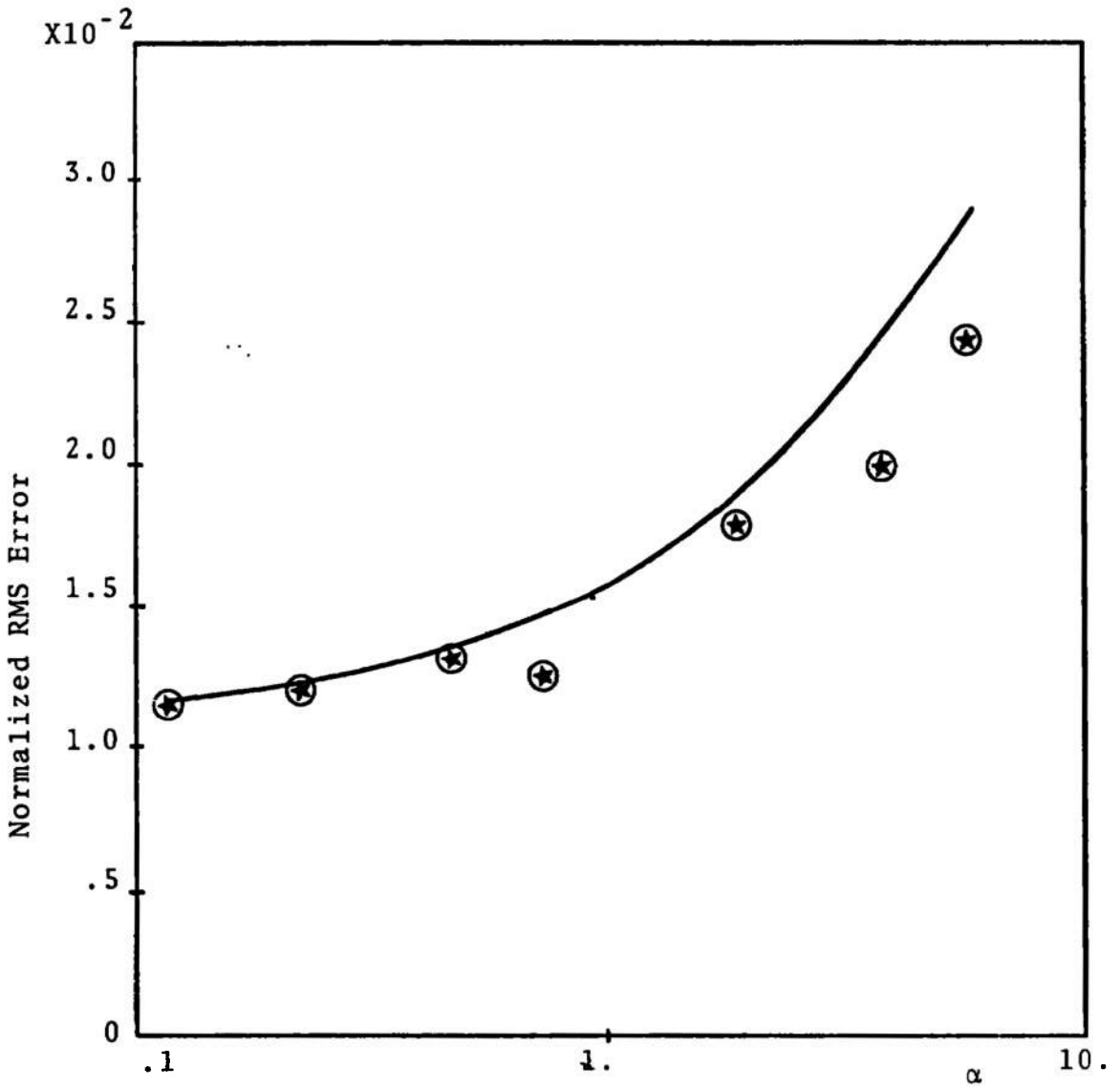
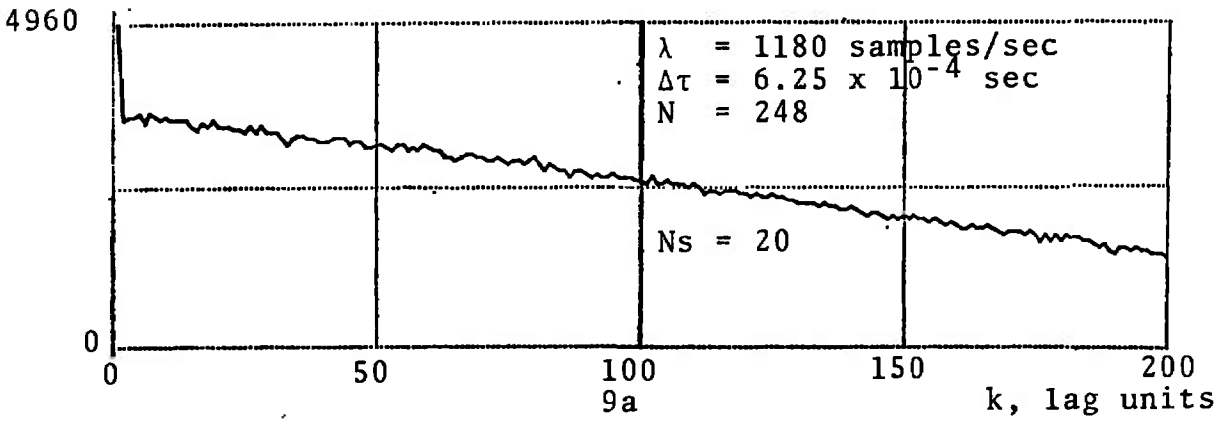
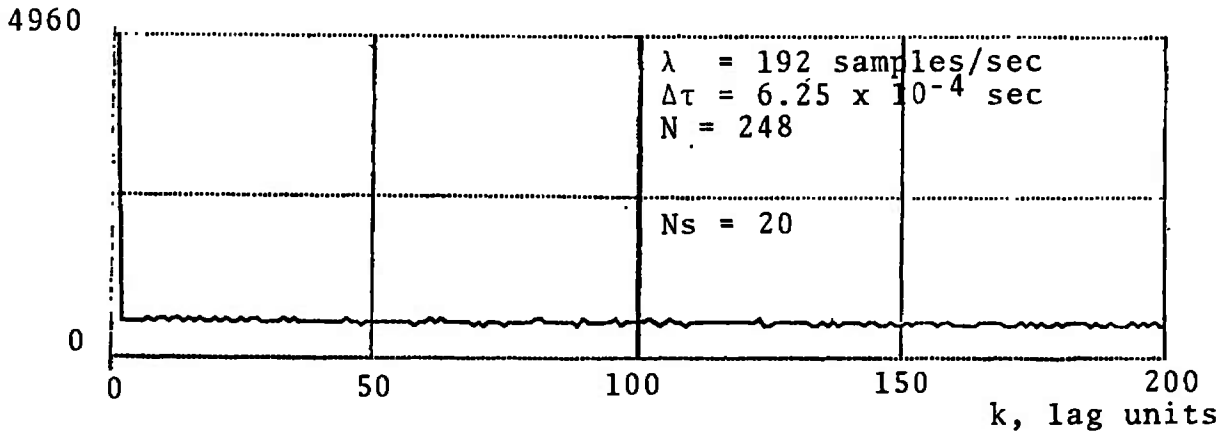


Figure 8. Normalized RMS Error of the Turbulence Intensity Estimate vs α.



9a



9b

Figure 9. Histograms of the Number of Lag Products.

SECTION IV

PRELIMINARY DESIGN

4.1 INTRODUCTION

In this section we provide some extension and further interpretation of the theoretical results of Section II in the form of a preliminary systems design guidelines and provide two illustrative examples. The discussion is restricted to the requirements of spectral analysis since implementation of mean and turbulence intensity estimates is straightforward under the assumptions of lack of correlation between λ and $U(t)$.

4.2 PRELIMINARY DESIGN PROCEDURE

There are many interrelated system variables whose proper relationship must be maintained to achieve results which satisfy error objectives while minimizing data collection and processing time. The same procedures must be used in system design, experiment planning, and final data analysis. The procedure is applied iteratively because of the many tradeoffs and the general fact that *a priori* knowledge of the spectrum to be estimated is essential in the analysis. With the exception of the alteration of the relationships due to random sampling, the design process is similar to that involved in conventional spectral estimation. Detailed design discussion for conventional estimation has been provided by Blackman and Tukey [2] and Parzen [28].

4.2.1 Bandwidth - Selection of $\Delta\tau$

Naturally produced spectra are not sharply band-limited and some aliasing bias error will generally have to be tolerated. When designing a system for turbulence spectrum estimation, electrical engineers accustomed to the use of 3 db bandwidth must exercise caution and communicate clearly with the aerodynamicists about the desired high frequency response. By way of example we note that a simple one-pole lowpass spectrum with 3 db bandwidth of 2.5 KHz and log-spectrum tail slope of -2 has an equivalent power bandwidth of 3.92 KHz and is only 26 db down at 50 KHz. For this example, the factor C_2 used in 4.2.4 below is 0.0785.

The aliasing error equation (2.40) indicates that *careful* consideration must be given to both the usual additive aliasing error and the multiplicative error due to random sampling. In a typical case, the maximum error occurs at a frequency less than the folding frequency $1/2\Delta\tau$ because the multiplicative error and the additive errors may cancel at the folding frequency.

4.2.2 Resolution, Frequency Spacing, and Number of Estimates

The frequency resolution R is an ambiguous quantity related to window shape, method of definition, and the lack or presence of discrete frequency spikes. For broadband spectra with R defined as the half width of the spectral window measured to the first zero and with Δf selected as $1/2M\Delta\tau$ where M is the maximum lag number and also the number of spectral estimates, then $R \approx 2\Delta f$ when a Bartlett or Hamming window is selected.

The value of M is given by

$$M = \frac{1}{2\Delta f \Delta\tau} \tag{4.1}$$

with typical values being 256, 512, 1024. It is important to note that it is a simple matter to reduce the value of M after the autocovariance estimate has been made if it proves desirable to sacrifice resolution for reduced variability error. In other circumstances, post-estimate smoothing may prove desirable. However, it M is chosen inadequately small, no simple remedy is available.

4.2.3 Data Segment Length

In some implementations it is desirable to process data in segments and subtract the sample mean for the segment from each sample. If the segment length, N, is too small the turbulence intensity estimate may be biased low according to equation (2.16). The bias effects on the spectrum are unknown, but are believed to be confined to the zero frequency estimate with leakage through the window function to the estimate at $f=\Delta\tau$.

An additional constraint on N is imposed by equation (2.28) to prevent excessive variance of the spectral estimate due to low values of the histogram H(k) for the larger values of k. This constraint is generally satisfied by

$$N \geq M \tag{4.2}$$

for sample rates $\lambda/2B < 1$, $\lambda\Delta\tau << 1$. Note that this constraint may be violated after system implementation by an over-eager experimenter who increases λ to decrease variability error. The inclusion of a comparison of H(1) and H(M-1) as a check is recommended.

4.2.4 Variability Error

Once bias error effects are included properly in the system design and experimental execution, the primary concern is the statistical error due to finite data collection. Equation (2.42) may be expressed logarithmically for worst case error analysis in terms of the normalized RMS spectrum error

$$\varepsilon_1 = \frac{(\sigma_f^2)^{\frac{1}{2}}}{S(f_{peak})} \tag{4.3}$$

where $S(f_{peak})$ is the peak spectrum value. The worst case (largest value of) equivalent power bandwidth B

$$B = \frac{\sigma^2}{2S(f_{peak})} \tag{4.4}$$

is used to obtain

$$\log \varepsilon_1 = \tfrac{1}{2} \log (2C_1C_2) - \tfrac{1}{2} \log \alpha - \tfrac{1}{2} \log (\beta) \tag{4.5}$$

where

$$C_1 = \frac{1}{M\Delta\tau}\int_0^{M\Delta\tau} w^2(\tau)d\tau \qquad \alpha = \frac{\lambda}{2B}$$

$$C_2 = 2\Delta\tau B \qquad \beta = \frac{NNs}{M}$$

This result is plotted in Figures 10 and 11 in db using typical values of $C_1 = 1/3$ (Bartlett window) and $C_2 = 0.0785$.

Equation (2.42) may also be expressed in terms of the total data collection time T (not including any processing time between data segments) by replacing NNs with the expected number of samples λT:

$$\log \varepsilon_1 = \tfrac{1}{2} \log (2C_1) - \log (\alpha) - \tfrac{1}{2} \log (\gamma) \tag{4.6}$$

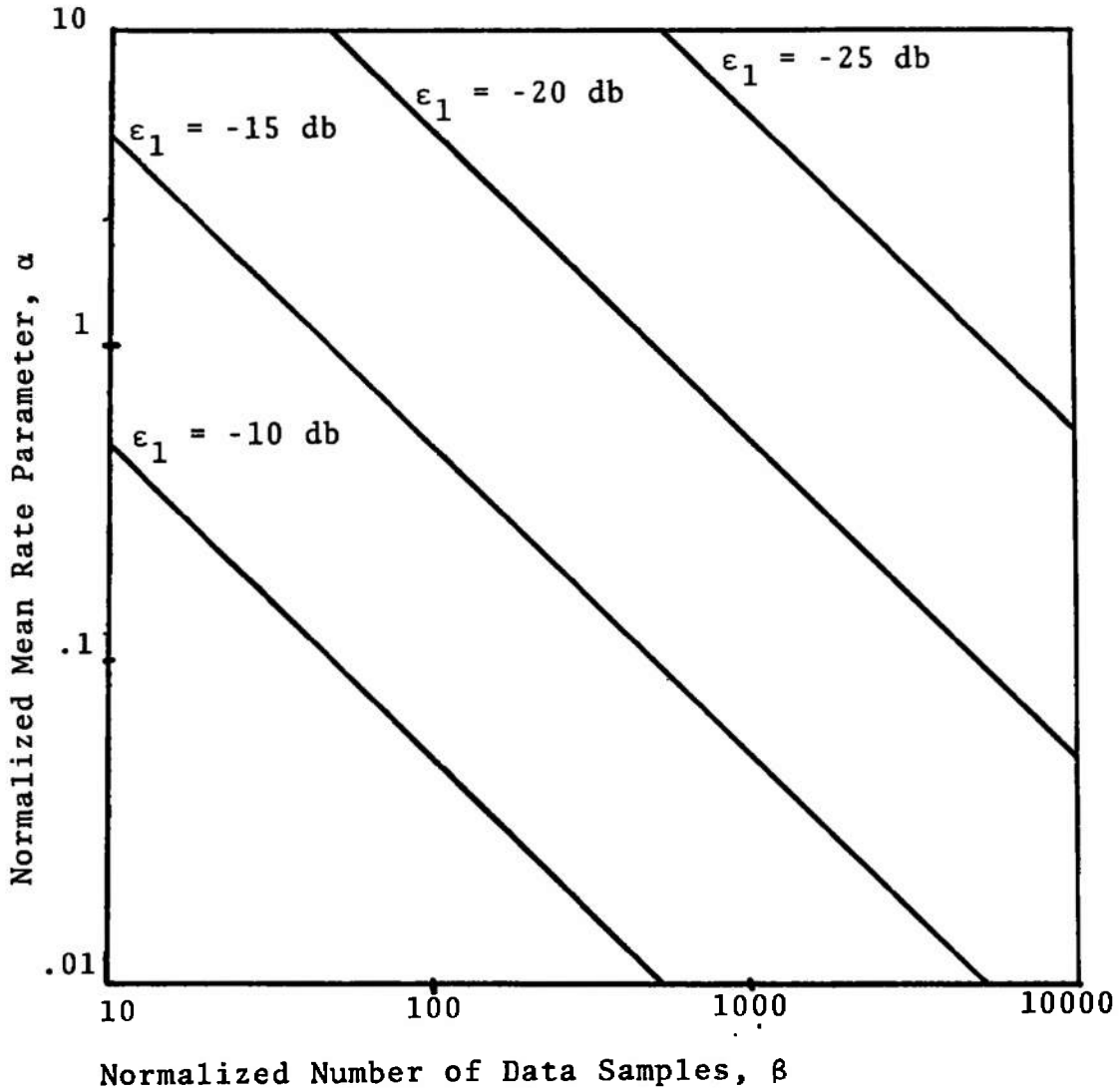


Figure 10. Normalized RMS Spectrum Error α vs β.

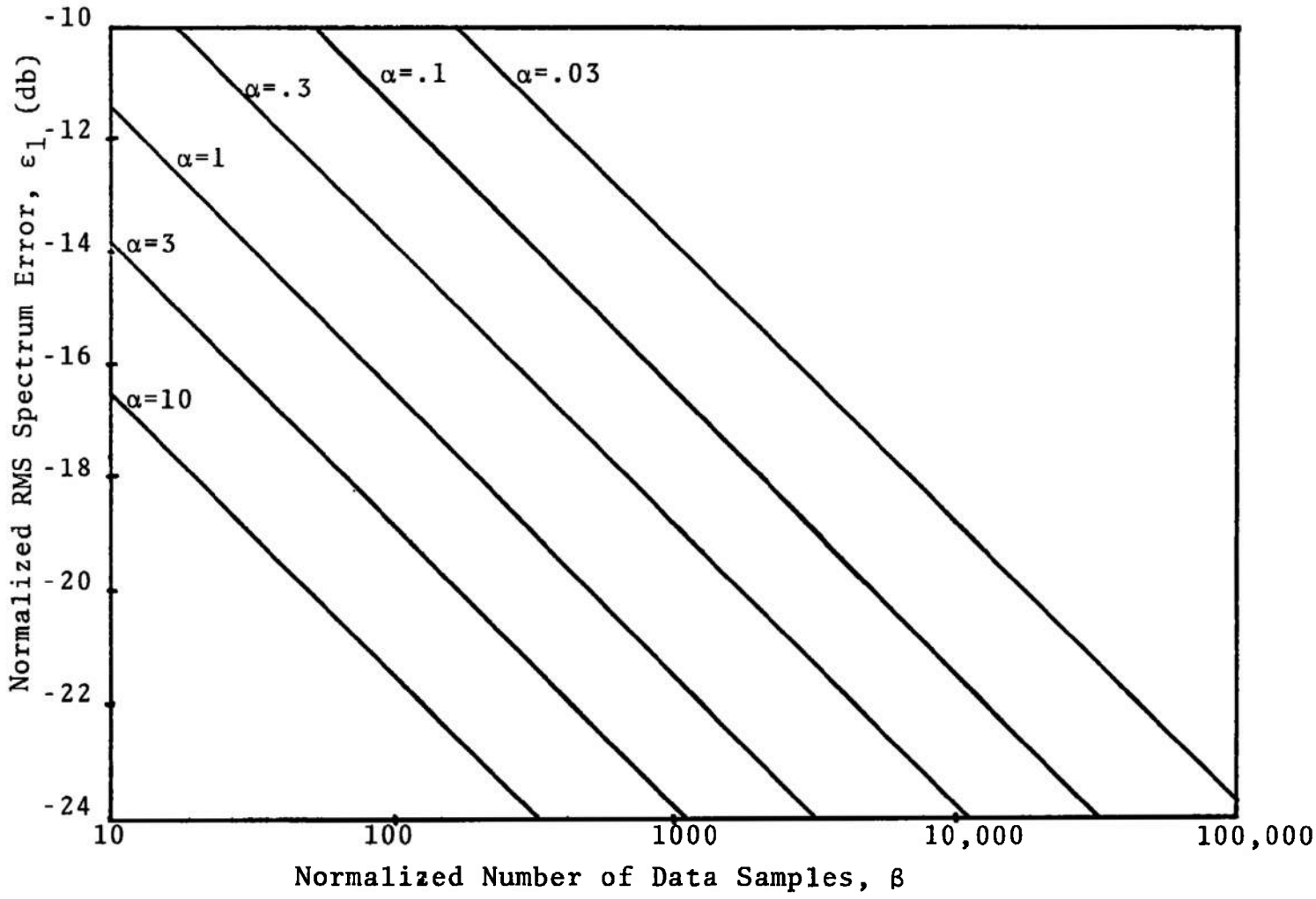


Figure 11. Normalized RMS Spectrum Error ε_1 vs β.

where C_1 and α are the same as in (3.6) and

$$\gamma = \frac{T}{M\Delta\tau} \tag{4.7}$$

The result is plotted in Figures 12 and 13.

We examine the significance of the error equations in examples in Section 4.3 below. First there are two more primary considerations.

4.2.5 Amplitude and Time Error

The present analysis has not included the effects of amplitude errors due to quantum and electronic noise, velocity gradients and particle lag effects, and timing errors in both the time-between-samples and the signal period counters. In addition, fast hardware reciprocation (period to scaled velocity) conversion has been assumed without regard to error. None of these effects were observed in the experimental results reported herein because they were negligible or non-existent in the simulation. However, they should be expected in LV measurements of high velocity gas flow. Further discussion is presented in Section 5.

4.2.6 Computation Time and Memory Considerations

The theory section does not concern itself with the time required to compute the autocovariance estimate. The time required to compute the required discrete Fourier transform to obtain the spectral estimate is small and is of secondary importance. The critical parameter is the rate ψ at which lag products must be formed, indexed, and accumulated. This rate parameter determines what form of implementation is possible in terms of avaliable software

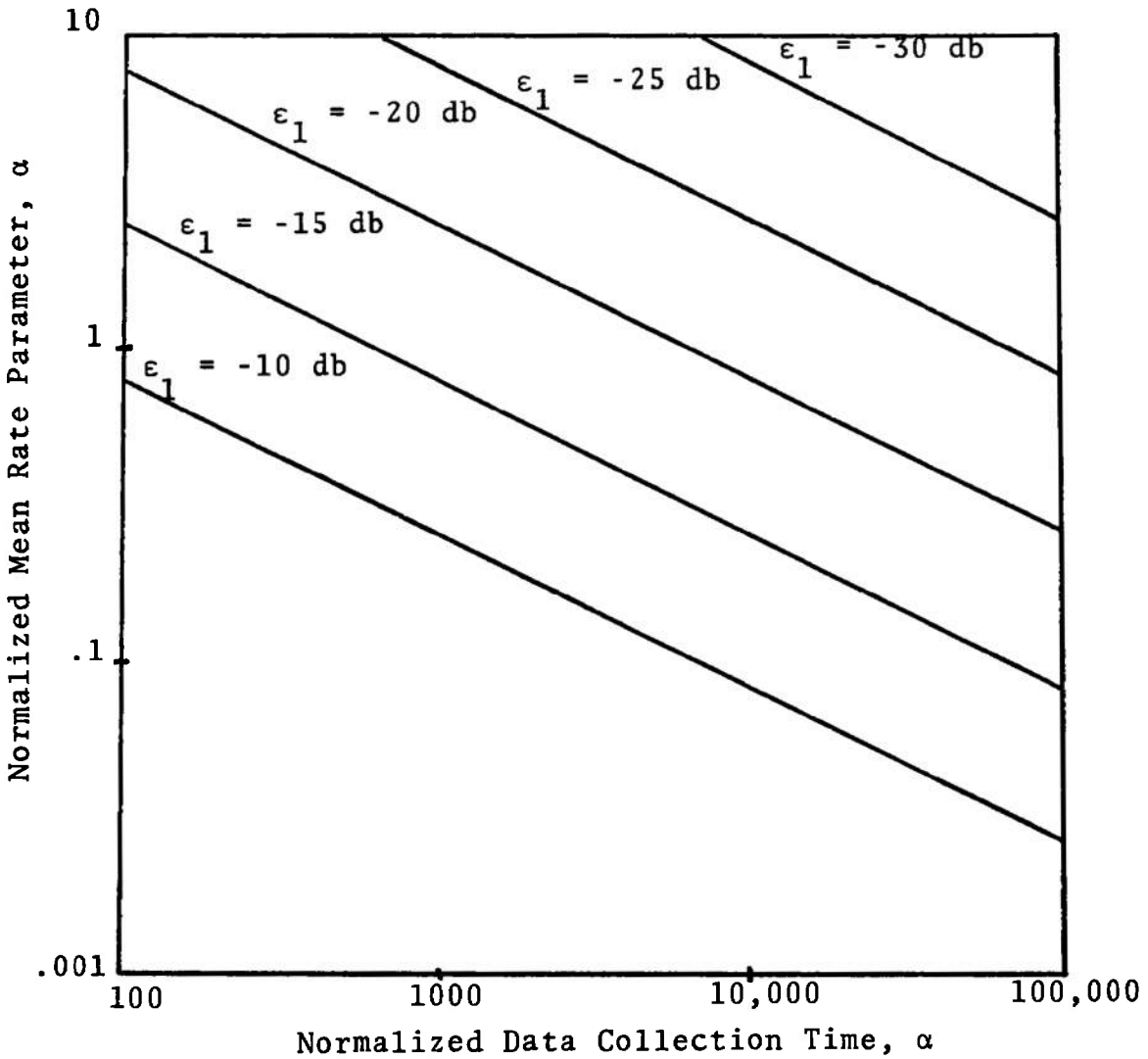


Figure 12. Normalized RMS Spectrum Error α vs γ.

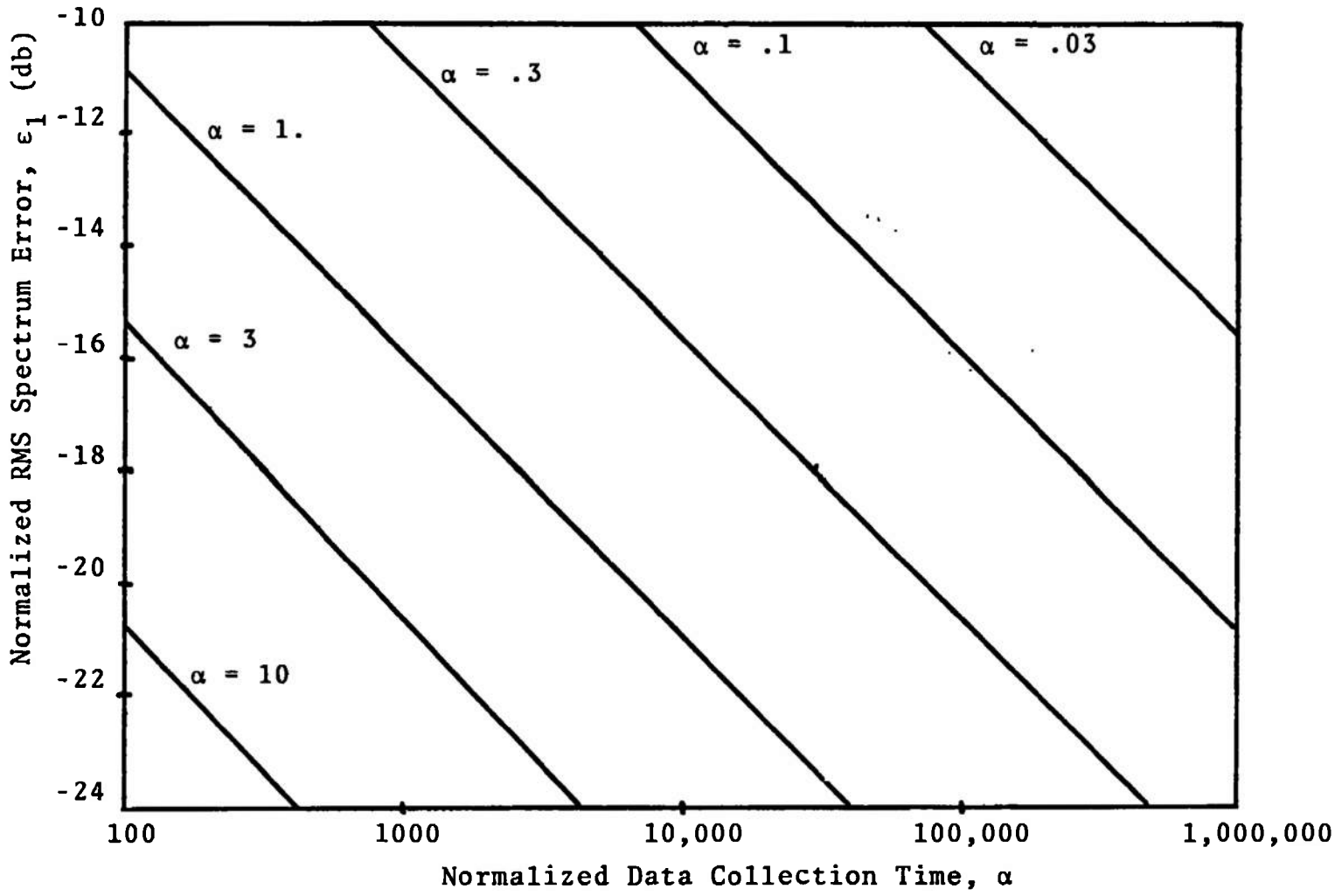


Figure 13. Normalized RMS Spectrum Error ϵ_1 vs γ.

and hardware.

The factors which determine ψ are the mean data rate λ and the maximum frequency resolution selection of $M\Delta\tau$. The average lag product rate is obtained simply as the product of λ and the expected number of samples $\lambda M\Delta\tau$ available at any instant in the preceding $M\Delta\tau$ seconds. Adding the rate of u_i^2 terms gives

$$\psi = \lambda^2 M\Delta\tau + \lambda \tag{4.8}$$

For segmented data implementations, the average number of product computations per data segment $N\psi$ may be obtained from the expected value of the histogram by summation:

$$N_\psi = N + N\lambda\Delta\tau - \frac{M}{2}(\lambda\Delta\tau)^2 \tag{4.9}$$

The $(\lambda\Delta\tau)^2$ term is small in (3.4) when the suggested condition $\lambda\Delta\tau << 1$ is satisfied. Division of (3.9) by N/λ, the average time required to obtain one data segment, produces equation (3.8) except for the $(\lambda\Delta\tau)^2$ term which results from the loss of available lag products for which the time pairs cross a segment boundary.

The system memory requirements vary considerably with implementation. The minimum possible data memory occurs when only the last JMAX samples and occurrance times $(U_i, U_{i-1}, \ldots U_{i-JMAX})$ are retained in real time processing. The quantity JMAX is also useful in segmented data implementations. This number is chosen statistically to insure that most of the sample pairs with time differences less than $M\Delta\tau$ will be retained while most of those with excessive time differences will not even be tested. This benefit is obtained by processing lag product pairs along diagonals of the segment product matrix ($U_i\ U_{i+j}$, $U_{i+1}\ U_{i+1+j}$, etc.)

and stopping when JMAX is obtained in the second subscript.

The value of JMAX is obtained by noting that for a Poisson sampling process the expected number of samples which will occur in time $M\Delta\tau$ is $\lambda M\Delta\tau$, and the *variance* of this number is also $\lambda M\Delta\tau$, for values of JMAX >> 1, the Poisson probability law has an envelope which is nearly Gaussian and the ratio of the RMS deviation to the mean becomes small. The mathematics are far from exact, but we may use the usual 3σ law for Gaussian statistics to obtain the following condition which assures that very few valid product pairs are missed:

$$JMAX = \lambda M\Delta\tau + 3\,(\lambda M\Delta\tau)^{\frac{1}{2}} \qquad (4.10)$$

When $\lambda M\Delta\tau > 9$, equation (4.10) gives a value of JMAX less than $2\lambda M\Delta\tau$.

4.3 EXAMPLES

4.3.1 Problem Specification

The following example is artificial but illustration. We assume that the spectrum to be measured will be bounded from above in the high frequency portion by

$$S_1(f) = \frac{S_o}{1 + (\frac{f}{f_o})^2}, \quad f > f_o = 2.5 \text{ KHz} \qquad (4.11)$$

The LV signal process is assumed to provide accurate measurements of the velocity samples and time differences with a system dead time of 2.5 μsec maximum. The measurement objectives are given as:

a) determine details of S(f) to 50 KHz

b) resolution = 100 Hz
c) data collection time = 30 seconds
d) use on-line real-time minicomputer program without external high speed memory
e) provide maximum error at each frequency less than 5% of the true spectrum
f) use mean data rate less than 1000 samples/sec.

It is impossible for any system to be designed which will meet the measurement objective specifications since the estimate error is statistical and cannot be specified by a maximum value. The usual spectrum estimation practice of specifying confidence interval using chi-squared statistics is not appropriate since the random estimate error is not a positive random variable. At the present time only the RMS error may be specified. With the change of e) to "provide bias error of less than 5% and RMS error of 2% of the true spectrum at each point" the specifications become legitimate in form but unattainable. This may be verified by applying the previously provided equations and guidelines.

The next step is to compromise on the specification. In what follows we arbitrarily relax some of the constraints to obtain two more realistic examples. The results are arbitrary design selections which cannot be further optimized without priority design objective tradeoff specifications. The difference in the results for the two examples clearly illustrates the need for priority assignment.

4.3.2 Low Frequency, High Resolution

The specifications are to be applied only to the frequency range 100 Hz to 2500 Hz. The measurement time

objective of 30 seconds is to be relaxed if necessary. It is assumed that no strong single-frequency spikes or steep steps are present which would provide excessive bias error at near-by estimates through the convolution of the spectral window function with the true spectrum.

The first selection of parameters is given in Table 1 under "Trial One". In the first trial all objectives are met except for T. The selection of ε_1 as 23 db is 17 db below the 2500 Hz 3 db value with an additional 1.5 db assumed for the unknown shape of the low-frequency portion of the spectrum and 1.5 db for experimentally indicated difference of actual error and theoretical error in the passband. The folding frequency $M\Delta f$ is sufficiently high that the additive aliasing error ε_2^+ from the 17.5 KHz portion of the assumed tail is 4%. This is adequate even if the predicted 2.5% negative error ε_2^- from random sampling effects does not materialize. Unfortunately the required γ and T are too large for consideration.

Table 1 also presents a second iteration in which the resolution requirement is reduced a factor of two by decreasing $\Delta\tau$ and increasing the folding frequency. This change makes ε_2 negligible but does not reduce T by a sufficient amount. The second trial also shows the effect of increasing the data rate to 7.85 K/sec to give $\alpha = 1$. Experiment indicates that there would be little or no advantage in increasing λ beyond the $\alpha = 1$ value for passband error. Unfortunately, the mean lag product rate ψ is too high for continuous online processing but other alternatives requiring only a digital tape recorder are possible as described in Section V.

	Example 1		Example 2
	Trial One	Trial Two	
Δf(Hz)	50	100	2000
R(Hz)	100	200	4000
M	200	200	50
$M\Delta f$(Hz)	10K	20K	100K
$\Delta\tau$(sec)	50μ	25μ	5μ
λ(/sec)	1K	7.85K	23.5K
f_o(Hz-3db)	2.5K	2.5K	2.5K
B(Hz)	3.93K	3.93K	3.93K
α	0.127	1.0	3.00
ψ(/sec)	11K	316K	138K
$1/\psi$(sec)	90μ	3.2μ	7.24μ
ϵ_1(db)	-23	-23	-35.5
ϵ_2^+	4%	negligible	11%
ϵ_2^-	2.5%	negligible	-10%
ϵ_2	1.5%	negligible	1%
γ	excessive	26K	900K
T	excessive	130	225
NN_s	excessive	1020K	5288K

Table 1. Trial Parameters for Example Problems

4.3.3 High Frequency Spectrum Tails

In order to obtain 2% error at 50 KHz the value of ε_1 relative to the constant lowpass level of the assumed spectrum $S_1(f)$ would be -43 db since $S_1(f)$ is -26 db at 50 KHz. A few trial parameter selections shows this error level is very difficult to reach. As a demonstration, Example 2 provides a parameter set which gives -35 db for ε_1 with no theory compensation deemed necessary. The value of $\Delta\tau$ = 5μsec is the smallest allowed by the signal processor dead time without additional bias errors. We note that in this example the lag product rate ψ is less than in the previous example because of the considerable loss of resolution in the reduction of $M\Delta\tau$. The total number of computations required is larger, however. The bias errors ε_2^+ and ε_2^- are the values at 50 KHz and appear to nearly cancel. The theory concerning ε_2^- has not been verified experimentally and the value of ε_2 has not been checked at lower values of f. The bias error could therefore be as high as 11% at 50 KHz which would be of the same order of magnitude as ε_1 (35 db).

SECTION V

DISCUSSION

5.1 IMPLEMENTATION ALTERNATIVES

The results show that there are many implementation alternatives, depending on the measurement objectives and the values of λ, the mean data rate, and ψ, the mean lag product rate.

5.1.1 Batch Processing - Segmented Data

In many cases where only the principal turbulence power distribution is of interest and measurement time is not critical, the simplest implementation will be an arrangement similar to that which we used in which a minicomputer acts as a time buffer and pre-processor between the LV electronics and a digital tape recorder. Additional features could be incorporated such as segment mean computation and subtraction, segment turbulence intensity estimate, and elimination of data points lying outside 3 or 4 σ deviation (assumed to be LV errors). This mode of operation would allow immediate access to first order statistics. It would also even allow for software period-to-velocity inversion and scaling at the data rates of 5 to 10 K.

For higher values of λ, such as may be required for tail investigations or for multiple channel processing, several batch mode alternatives are possible. The fastest methods would include either a large core or a high speed

disc.[3] The inexpensive method would be to simply retain the tape drive, use slightly modified pre-processing and transfer programs, but let the computer wait between segments the additional time required to write a segment on tape. Possible complications due to any low frequency sampling effect are not expected to be serious due to the random length of time required for acquiring a segment. With slightly more waiting time these same arguments apply to replacement of the tape unit with a telephone connection to a large time shared machine.

The batch processing mode is recommended at the present time not only for reasons of cost and simplicity but also because enough unknowns still remain to make saving the data for future reevaluation worthwhile if the test is an important one. For example, after seeing the results the investigator may wish to change $\Delta\tau$ a week later and reprocess the data.

5.1.2 On Line Processing - Segmented Data

This alternative is similar to the Batch approach. The only difference is that after a segment is read, the processing continues until the lag products and additions to the histogram have been completed before another segment is read. At low values of λ and ψ, this approach could be made nearly real-time if two memory segments are used with priority interrupt for reading and writing in between computations. The differences between this approach and Batch processing would be most apparent at higher data rates where ψ is increased by a λ^2 factor, particularly in cases where multiple channels of data would be processed. In any event,

[3]This is currently being done at the Lockheed Georgia Company -

the memory requirements would be greater because of the additional programming. Since the output data rate would be quite low in this approach, the division of the SUM matrix by the histogram, the multiplication by the window function, and the transform operations could all be done remotely on a time-shared computer.

5.1.3 On Line Processing - Continuous

This approach is feasible for sufficiently low lag product rates. It could even be used in Batch processing if no data is skipped in the recording process because it has the advantage of requiring less data memory in core. The data memory is indexed in a circular manner with the most recent sample replacing the oldest. The length of the memory array need be no longer than JMAX as described in Section 4.2.6. Unfortunately, in most cases the achievable ψ rate would be inadequate unless long times T are used.

Another difference in this approach is that a preestimate of the mean would be required for subtraction from each sample as it occurs. This has the advantage of not removing low frequency components in the first frequency estimate since the same constant is subtracted from all data. It has the disadvantage of leaving a small spike of unknown amplitude at zero frequency due to error in the mean estimate used in the subtraction.

5.1.4 Hardware

This is a class of alternatives based on the second formulation of the autocovariance estimate given by

equation (2.25). In this class of alternatives a buffer containing the latest sample value would be periodically read with reset to zero. The data would then consist of many zeros with sample values sparsely distributed. The result would be very inefficient with memory utilization. The alternate formulation[4] also has differences in the required error analysis as previously noted. The approach is designated "hardware" since it is conceivable that two commercial correlators could be modified to produce the autocorrelation of the sampled data sequence and the indicator sequence separately for division and Fourier processing on a time shared computer. The general approach could also be implemented as software, and has the advantage that time differences need not be computed. The probable disadvantages include poorer error performance, the need for hardware period-to-velocity conversion, and the lack of analysis at the present time of other factors which make recording the data worthwhile.

5.2 RELATION BETWEEN THEORY AND EXPERIMENT

The experimental results of Section III show excellent agreement with the theory of Section II with regard to spectral estimate varability error prediction

[4]This formulation is attributed to an interpretation by P. Scott, General Electric Co., of the results from this contract presented by Mayo, Shay, and Riter at the Oklahoma State University Workshop on "Theory and Application of the Laser Doppler Anemometer", June 11, 1973.

in the tail region. That this agreement persists at values of $\lambda/2B > 1$ is somewhat surprising in view of the violation of the derivation assumptions of low data rate.

The experimental results also show the theory is inadequate for error prediction in the spectrum passband at rates $\lambda/2B > 1$. In this case, the experiments substantiate intuitive reasoning based on information theory relevant to periodic sampling: sampling at rates in excess of $\lambda/2B$ is redundant; the only way to increase the information transfer concerning the random process being sampled is to increase T. Continuation of this intuitive logic shows that for a fixed number of samples, increasing λ above 2B is detrimental becauses it *reduces* T. The quantitative experimental results of Figure 5 are inadquate to prove that ε_1 actually exhibits a minimum at $\lambda/2B = 1$, but the possibility is certainly not eliminated. This possibility is qualitatively substantiated by the sequence of Figure 2 where the values of $\lambda/2B$ attained were greater.

We conclude from our results that spectrum estimation with mean sample rates much less than twice the *highest* frequency are not only possible but also desirable when we are concerned with the primary turbulence power distribution. (The quantity B is generally much less than the highest signal frequency). On the other hand, attainment of usably low estimate error in the spectrum tails will generally require many compromises in terms of measurement time, resolution, the use of attainable but relatively high data rates, and possibly in the use of expensive high speed memory.

5.3 COMMENTS ON FUTURE RESEARCH

This investigation has raised many questions which require further work both theoretically and experimentally. We include here brief descriptions of a few of these questions with intuitive comments on their implications.

5.3.1 Sample Amplitude Error

The velocity samples U_i will contain "small" errors from sources which include photon and electronic noise, threshold period measurement error, ± 1 clock count error, and inversion round-off error. In addition, there will be occasional "large" errors when a noise burst succeeds in satisfying the check logic of the LV processor. Finally, there is a different type of effect due to particle lag, velocity gradients, and finite measurement volume size.

As a first approximation, we may assume that the "small errors" collectively form a zero-mean wide-band random noise process, n(t) randomly sampled at instants t_i. If n(t) is independent of the velocity U(t) then the actual signal

$$s(t) = n(t) + U(t)$$

has true autocovariance given by

$$C_s(\tau) = C(\tau) + C_n \delta(\tau)$$

where C_n is a constant and $\delta(\tau)$ is the unit impulse function. The resulting true spectrum is

$$S_s(f) = S(f) + C_n$$

i.e., the value of C_n is the white noise level added by the* error process n(t).

The bias error effect of n(t) in the estimate $\hat{C}(k)$ will be confined largely to C(0). This means that the noise will appear in the measurement process as turbulence intensity which is not actually present. We observe, however, that with the simplified model assumed for the noise n(t) the value of the white level C_n may be *measured* independently through spectrum estimation by using the fact that the true spectrum must go to zero at the higher frequencies. Unfortunately in the widest bandwidth estimation problems where the ± 1 clock errors are most likely to be serious, the noise spectrum is least likely to be white. The immediate result will probably be a small signal-dependent broad-band spectral bias level which will add further difficulties to the tail estimation problem.

The second class of errors, "large errors", occur relatively seldomly, but the much larger magnitude implies that these errors could also produce a non-negligible additive white spectrum. Partial remedies include the data-check circuits and preprocessing steps of estimation of σ and throwing out excessively large or small estimates. These practices are presently being incorporated in LV processors and will aid significantly in the spectrum estimation problem.

The third type of error is associated with the minor particle lag and velocity gradient effects. These are the most fundamental in that they will be present even if the LV system is error free. It is common to be worried about the nontrivial problem of the low-pass effect of particle inertia on spectrum estimation. We have left that one

* Further analysis has revealed that this white level is also proportional to $\Delta\tau$. See "A Discussion of Limitations and Extensions of Power Spectrum Estimation with Burst Counter LDV Systems.", W. T. Mayo, Jr., to be published in the Proceedings of the 1974 Purdue LDV Workshop.

for the particle dynamicists. A more subtle effect of minor particle lag is the fact that two particles entering the measurement volume at different velocities and closely spaced in time will appear spectrally as part of a wide-band random process. The same type of effect will result from velocity gradients when particles enter opposite sides of the probe volume closely in time. These effects will not only increase the apparent turbulence intensity but also contribute difficult-to-characterize high-frequency spectral error. They will be most insidious in the measurement of nearly steady flows where the actual turbulence levels are quite low.

A related effect will accompany high turbulence and small turbulence scale: while the finite measurement volume size acts as an averager or low-pass filter, it also will artificially create high frequency spectra due to random timing effects. Thus some of the difficulties discussed by W. George [29] under the assumptions of high density seeding and Gaussian statistics also exist in a different form in the single-particle LV systems.

5.3.2 Cross-Correlations and Cross-Spectra

We wish to point out here that there is no fundamental reason for not extending the concepts we have developed to two-channel cross-correlations and cross-spectra and that this is, in fact, being done at the Lockheed-Georgia Company. In this regard, however, we note that

there are a few subtleties in measuring cross-correlations between two separate probe locations which are different than simply measuring two components at the same point; and further there is a difference in a two component measurement with 45° crossed fringes and a one-component measurement with normal fringes.

First, the matter is straight-forward for two orthogonal measurements at the same point if it is required that both channels produce "good data" simultaneously and the same time interval measurement applies to both. The only problem would seem to be the tolerances on the "simultaneous" requirement. However, we note that this requirement may bias the selection of "good data" toward direction vectors along the bisector of the orthogonal component directions. This results when Bragg-cell systems are not used because off-axis direction vectors produce a higher probability of "good data" in one channel than in the other.

If simultaneity is not required, then the algorithms may become more complex as they will in separate-point measurements where each channel has a partially correlated or independent random sampling process. The correlation, or lack of it, between the sampling processes is irrelevant if they are jointly independent of the velocity process. The algorithms for time difference computations are more complicated because both backward and forward time differences must be computed.

Finally we note that with space correlations involving zero delay in time the concept of simultaneity must once more be considered. Finite tolerances are required to obtain any data points, and the joint data rate will be similar to

that of one element of the autocovariance histogram: generally much lower than the individual channel data rate. This difficulty cannot be overcome by indiscriminately opening up the simultaneity tolerance. It must be similar to the restrictions on $\Delta\tau$ in the autocovariance estimation to prevent the averaging to zero of small scale components.

5.3.3 Correlations Between Mean Sample Rate and the Velocity Process.

In informal monthly progress reports, we derived the fact that if scatterers have a uniform Poisson random distribution in volume, then the expectation $\lambda(t)$ of the sampling process is not constant in time. Instead, it is proportional to velocity and the estimate of the mean produced by equation (2.13) is biased due to the correlation between the sampling process and the velocity process. This work is not reproduced, because McClaughlin and Tiederman have since reported similar conclusions, with extension of the same concepts to the turbulence intensity estimate, in detail [30]. We note that subsequently we pointed out that the bias of the mean estimate vanishes in first order approximation when the average of the burst period measurements is computed before inversion and scaling to velocity. Barnett and Bentley [31] have extended this discussion which has now become a topic of current debate.

It is not unreasonable to suspect that correlation between velocity fluctuations and mean data rate will also

adversely affect the spectrum estimation problem under high turbulence intensity conditions. Unfortunately we are now dealing with non-stationary statistics for which the use of interchangable time and statistical averages is not appropriate. A fallacy underlying the whole discussion, however, is the lack of experimental evidence concerning the nature of the correlations involved. None of the theory mentioned thus far even remotely relates to the interpretation of jet mixing data, for example, in which the mean data rate depends on which air source each small volume increment originated in. Considerably more definitive work is needed in this area.

SECTION VI

SUMMARY AND CONCLUSIONS

The objectives of this research have been accomplished. It has been experimentally demonstrated that, with randomly timed sampled data, power spectra may be estimated regardless of the average data rate. This means that, in principle, turbulence power spectrum measurements are possible with unseeded air at very low data rates with real-time on line minicomputer data processing, without the use of large high speed memory. The algorithms for accomplishing this have been identified.

The principal difficulty lies in the amount of data collection time required to determine the low level spectral tails with practical error tolerances. Theory has been developed, with excellent agreement with experiment, for prediction of the spectrum estimate error in terms of the algorithm and LV system-flow parameters. The results indicate that the optimum mean data rate, for efficient estimation of the major distribution of turbulence power, is twice the equivalent power bandwidth of the spectrum; (not twice the "highest" frequency). However, the attainment of practical error levels in spectrum tail estimation may impose very stringent requirements including low resolution and excessively long data collection and processing time. For estimation of the spectrum tail, the most significant error and time reduction is accomplished by increase of the mean data rate to values exceeding twice the equivalent power bandwidth with no apparent limit. An example calculation

in which the spectrum tail is down 26 db at 50 KHz indicates that a mean data rate of 23.5 K/sec may provide practical measurement times and error levels. In the same example, real-time continuous minicomputer processing would not be feasible, but the use of a high-speed disc could be avoided through recording the data in segments with gaps.

A preliminary design guideline is presented and alternate implementations with differing tradeoffs in cost and time are discussed. At the present time, the recommended implementation would use a minicomputer as a buffer and preprocessor for recording data segments on digital tape or for transmission to a large time-shared computer. Later, when the remaining unknowns which have been identified have been more fully explored, more efficient implementations which destroy the data as it is obtained may be appropriate.

REFERENCES

1. Stevenson, W. H., Pedigo, M. K., and Zammit, R. E., "Bibliography on Laser Doppler Velocimeters: Theory, Design, and Application." RD-TR-72-8, U.S. Army Missile Command, Redstone Arsenal, Alabama, July 1972.

2. Blackman, R. B. and Tukey, J. W., The Measurement of Power Spectra from the Viewpoint of Communications Engineering. Dover, New York, 1954.

3. Bendat, J. W. and Piersol, A. G., Measurements and Analysis of Random Data. John Wiley and Sons, New York, 1966.

4. Richards, P. I., "Computing Reliable Power Spectra." IEEE Spectrum, Vol. 4, No. 1, pp. 83-90, January 1967.

5. Lee, Y. W., Statistical Theory of Communication. John Wiley and Sons, New York, 1960.

6. Bingham, C., Godfrey, M. D., and Tukey, J. W., "Modern Techniques of Power Spectrum Estimation." IEEE Trans. on Audio and Electroacoustics, Vol. AU-15, No. 2, pp. 56-66, June 1967.

7. Brumbach, R. P., "Digital Computer Routines for Power Spectral Analysis." TR68-31 (AD673859), AC Electronics Defense Research Laboratories, General Motors Corp., July 1968.

8. Jones, R. H., "Spectrum Estimation and Time Series Analysis - A Review." International Symposium on Probability and Statistics in the Atmospheric Sciences, June 1-4, 1971, AFOSR-TR-71-0725.

9. Jenkins, G. M. and Watts, D. G., Spectral Analysis and Its Applications. Holden-Day, San Francisco, 1968.

10. Leneman, O. A. Z. and Lewis, J. B., "Random Sampling of Random Processes." IEEE Trans. on Auto. Control, Vol. AC-11, Nol 3, pp. 396-403, July 1966.

11. Lorens, C. S., "Recovery of Randomly Sampled Time Sequences." IRE Trans. on Communication Systems, Vol. CS-10, No. 2 pp. 214-216, June 1962.

12. Shapiro, H. S. and Silverman, R. A., "Alias-Free Sampling of Random Noise." J. Soc. Indust. Appl. Math. Vol. 8, No. 2, pp. 225-248, June 1960.

13. Beutler, F. J., "Alias-Free Randomly Timed Sampling of Stochastic Processes." IEEE Trans. on Information Theory, Vol. IT-16, No. 2, pp. 147-152, March 1970.

14. Masry, E., "Random Sampling and Reconstruction of Spectra." Information and Control, Vol. 19, pp. 275-288, 1971.

15. Singleton, R. C. and Larson, A. G., "Spectral Analysis of Unequally Spaced Data Samples." (AD74347), Presented at IEEE International Symposium on Circuit Theory, North Hollywood, California, April 18-21, 1972.

16. Shaw, L., "Spectral Estimates from Nonuniform Samples." IEEE Trans. on Audio and Electroacoustics, Vol. AU-19, No. 1, pp. 24-31, March 1971.

17. Jones, R. H., "Spectrum Estimation with Missing Observations." Annals of the Institute of Statistical Mathematics, Vol. 23, No. 3, pp. 387-398, 1971. AFOSR-TR-72-0513.

18. Adegbola, M. O., Alias-Free Spectral Estimation of Stochastic Processes, Dissertation, Ph.D., California Institute of Technology, 1971. (Available from University Microfilms)

19. Asher, J. A., Scott, P. F., and Wang, J. C., "Parameters Affecting Laser Velocimeter Turbulence Power Spectrum Measurements." AEDC-TR-74-54, Contract F40600-72-C-00013.

20. Thompson, Rory O. R. Y., "Spectral Estimation from Irregularly Spaced Data." IEEE Trans. Geoscience Electronics, Vol. GE-9., No. 2, pp. 107-110, April 1971.

21. Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill, New York, 1965.

22. Jones, R. H., "Aliasing with Unequally Spaced Observations." Journal of Applied Meteorology, Vol. 11, No. 2, pp. 245-254, March 1972, AFOSR-TR-72-1548.

23. Jones, R. H., "Spectrum Estimation with Unequally Spaced Observations." Proceedings of the Kyoto International Conference on Circuit and System Theory, September 9-11, 1970, AFOSR-70-2747TR.

24. Oppenheim, A. B. and Johnson, D. H., "Discrete Representation of Signals." Proc. IEEE, Vol. 60, No. 6, pp. 681-691, June 1972.

25. Kan, E. P. F. and Aggarwal, J. K., "Randomly Sampled Digital Filters." IEEE Trans. on Audio and Electroacoustics, Vol. AU-20, No. 1, pp. 52-57, March 1972.

26. Evans, G. W., II, and McCarty, R. C., "Applied Spectral Analysis of Mixed Random and Deterministic Processes." (AD 691302) Stanford Research Institute, March 1969.

27. Welch, P. D., "The Use of the Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short, Modified, Periograms." IEEE Trans. on Audio and Electcoacoustics, Vol. AU-15, No. 2, pp. 70-73, June 1967.

28. Parzen, E., "Mathematical Considerations in Estimation of Spectra." Technometrics, Vol. 3, pp. 167-189, May 1961.

29. George, W. K., Jr., and Lumley, John L., "The Laser Doppler Velocimeter and its Application to the Measurement of Turbulence," Journal of Fluid Mechanics, 60, part 2, pp. 321-362, 1973.

30. McLaughlin, D. K., and Tiederman, W. G., "Statistical Biasing in Individual Realization Laser Anemometry,: Oklahoma State University School of Mechanical and Aerospace Engineering Report ER-73-J-19.

31. Barnett, D. O., and Bentley, H. T., "Statistical Bias of Individual Realization Laser Velocimeters," to be published in the Proceedings of the "Workshop on Laser Doppler Velocimetry," Purdue University, West Lafayette Indiana, March 27-29, 1974.

APPENDIX I

TWO SPECTRAL ESTIMATORS BASED ON THE PERIODOGRAM APPROACH

I.1 INTRODUCTION

One general approach to the estimation of broad-band spectra from random samples is to approximate the periodogram, where the raw periodogram is defined as

$$\hat{S}_T(\omega)=\frac{1}{T}\left|\int_{-T/2}^{+T/2} u(t)\ \exp(-j\omega t)dt\right|^2 \tag{I-1}$$

with u(t) assumed to be a zero mean random process and T is the record length. The periodogram estimate is made to converge by averaging many estimates from short segments of data [27].

I.2 ESTIMATOR 1

The most straightforward approach is to use a rectangular approximation of the raw periodogram integral, equation (I.1). After some manipulation, the result is

$$\hat{S}_1(\omega)=\frac{1}{T}\left[\left(\sum_{i=1}^{N} u_i\ \cos(\omega t_i)\Delta_i\right)^2 + \left(\sum_{i=1}^{N} u_i\ \sin(\omega t_i)\Delta_i\right)^2\right] \tag{I-2}$$

where $\Delta_i = t_i - t_{i-1}$

$T = t_N - t_0$

I.3 ESTIMATOR 2

Another approach is to first use the periodogram integral to estimate the spectrum of the process

$$Z_2(t) = u(t)\ Z_1(t) = \sum_{i=-\infty}^{+\infty} u_i\ \delta(t-t_i)$$

$$\text{where}\quad Z_1(t) = \sum_{i=-\infty}^{+\infty} \delta(t-t_i). \tag{I-3}$$

It is assumed $Z_1(t)$ is a Poisson impulse process with sample rate λ. With this estimate, it is possible to compute an estimate of $S(\omega)$, the power spectrum of $U(t)$, because it is known that

$$S_r(\omega) = \lambda^2\ S(\omega) + \lambda\sigma^2 \tag{I-4}$$

where $S_r(\omega)$ is the power spectrum of $Z_2(t)$ and σ^2 is the variance of $U(t)$. The periodogram estimate of $S_r(\omega)$ is

$$\hat{S}_r(\omega) = \frac{1}{T}\left|\int_0^T Z_2(t)\ \exp(-j\omega t)dt\right|^2$$

$$= \frac{1}{T}\left|\sum_{i=1}^{N} u_i\ \exp(-j\omega t_i)\right|^2 \tag{I-5}$$

and the estimate of $S(\omega)$ is

$$\hat{S}_2(\omega) = \frac{1}{\lambda^2} \hat{S}_r(\omega) - \frac{\sigma^2}{\lambda} \tag{I-6}$$

For real computation, it is necessary to use estimates of λ and σ^2 which are

$$\hat{\lambda} = \frac{N}{T}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} u_i^2 \ . \tag{I-7}$$

Substitution of these estimates and some algebra leads to

$$\hat{S}_2(\omega) = \frac{T}{N^2} \left[\left(\sum_{i=1}^{N} u_i \cos \omega t_i\right)^2 + \left(\sum_{i=1}^{N} u_i \sin \omega t_i\right)^2 - \sum_{i=1}^{N} u_i^2 \right] \tag{I-8}$$

I.4 EXPERIMENTAL COMPARISON

The two estimators were programmed and debugged using both uniformly and randomly sampled data produced by the electronic simulation facility. To compare the estimators when used with the same data, a zero mean bandpass process was randomly sampled at λ = 166 samples/sec. The same 20 records of this data were then processed with Estimators 1 and 2 and with the autocovariance estimator given by equation 2.33 of the text (noted as Estimator 3 in Figure I-3). The resulting estimates are shown in Figures I-1, I-2, and I-3 where the asterisks represent

measurements of the true spectrum as made with a wave analyzer.

Figure I-4 is the result of reducing M by a factor of 4 with a decrease in resolution (which was four times more than required in Figure I-3) to $\Delta f = 1/(2M\Delta\tau)$ and corresponding reduction of estimate variability.

The result with Estimator 1 was not satisfactory, while the results with Estimators 2 and 3 were comparable. The significant differences were that Estimator 2 required nearly twice as much computer time as Estimator 3 and depended heavily on the stationarity assumption of the sampling process for the prediction of the white level.

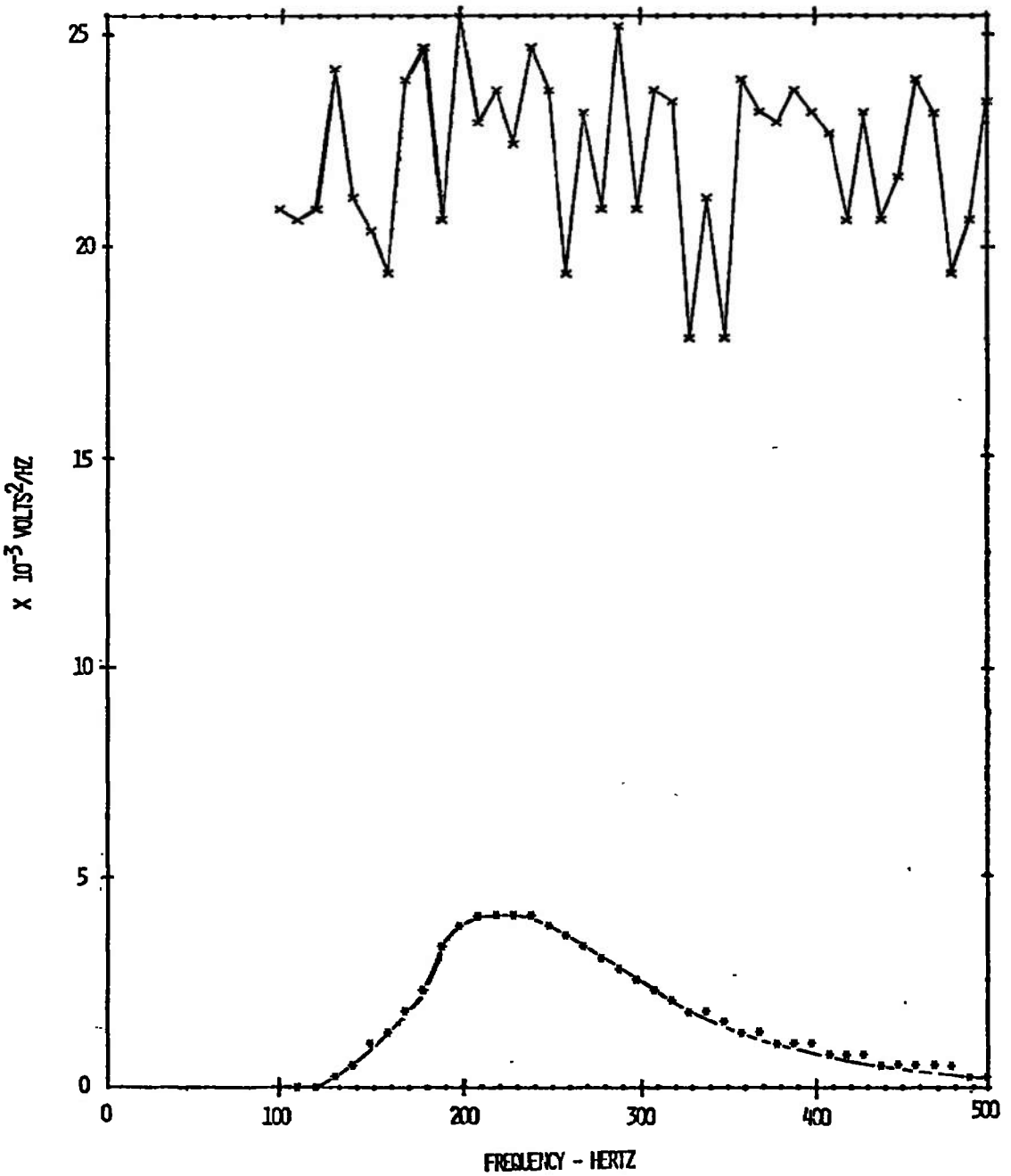


Figure I-1. Spectrum Estimate by Estimator 1
(λ = 166 samples/sec)
X - Estimate
* - Scaled Wave Analyzer Data

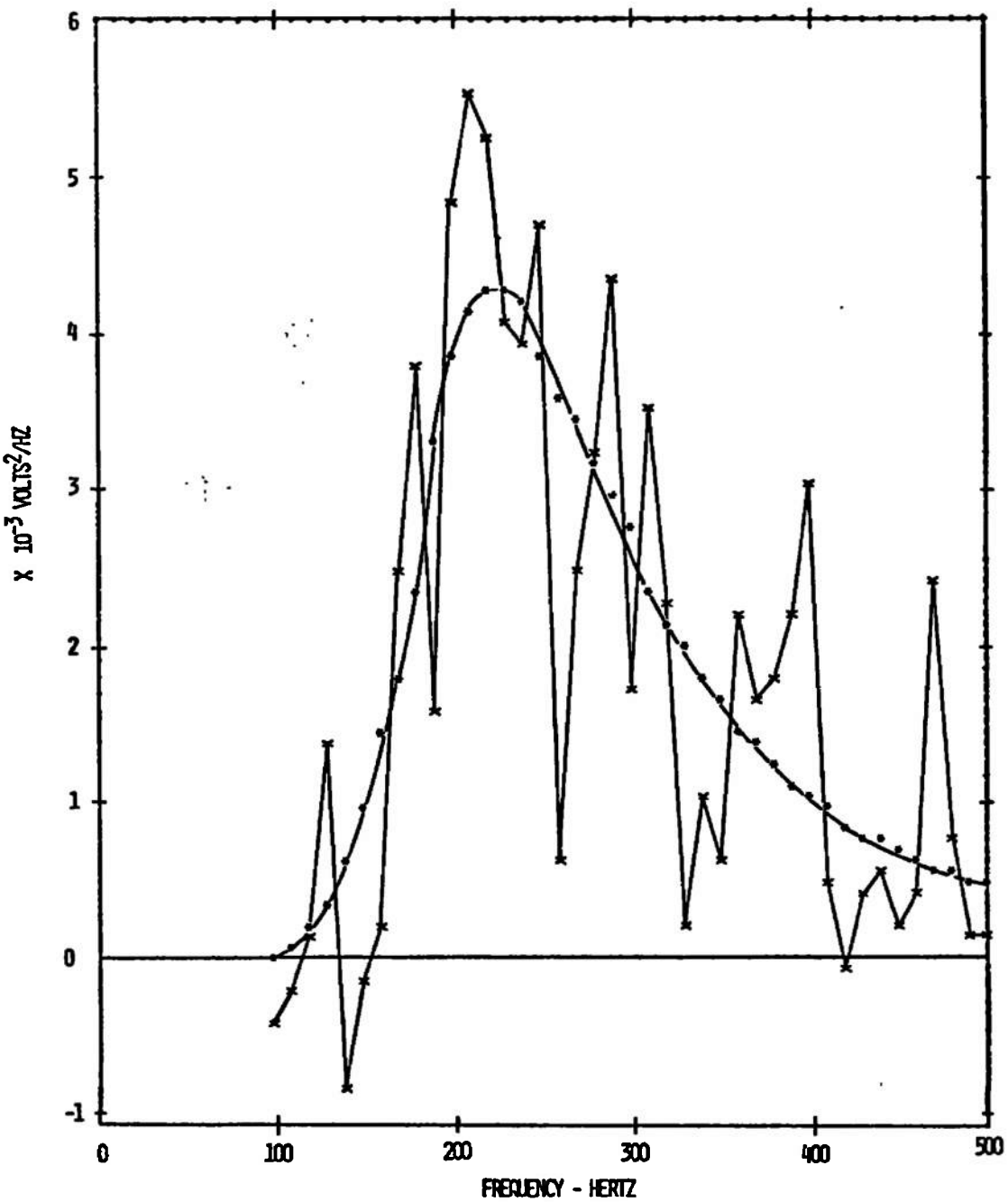


Figure I-2. Spectrum Estimate by Estimator 2
(λ = 166 samples/sec)
X - Estimate
* - Scaled Wave Analyzer Data

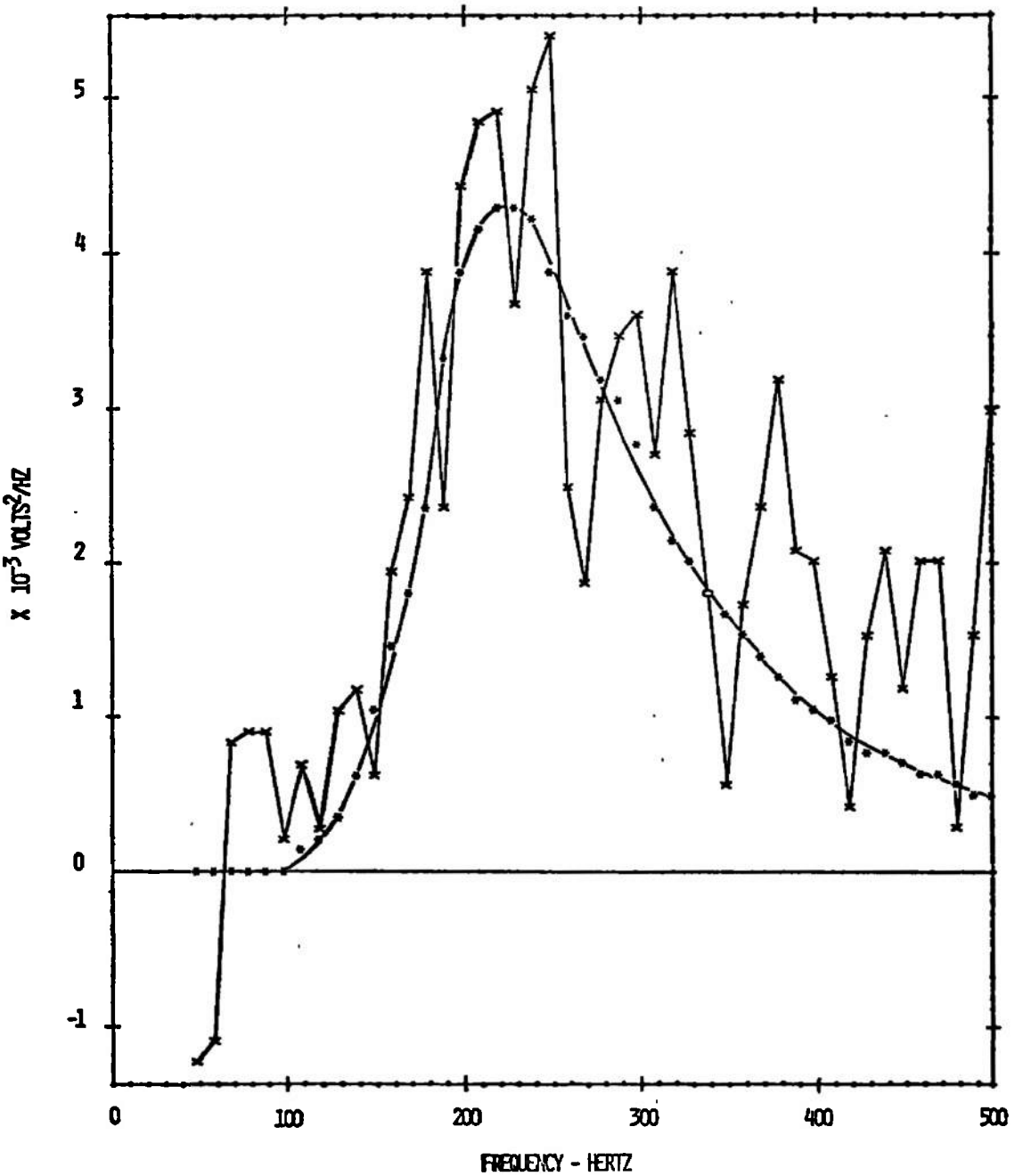


Figure I-3. Spectrum Estimate by Estimator 3
(λ = 166 samples/sec)
X - Estimate
* - Scaled Wave Analyzer Data

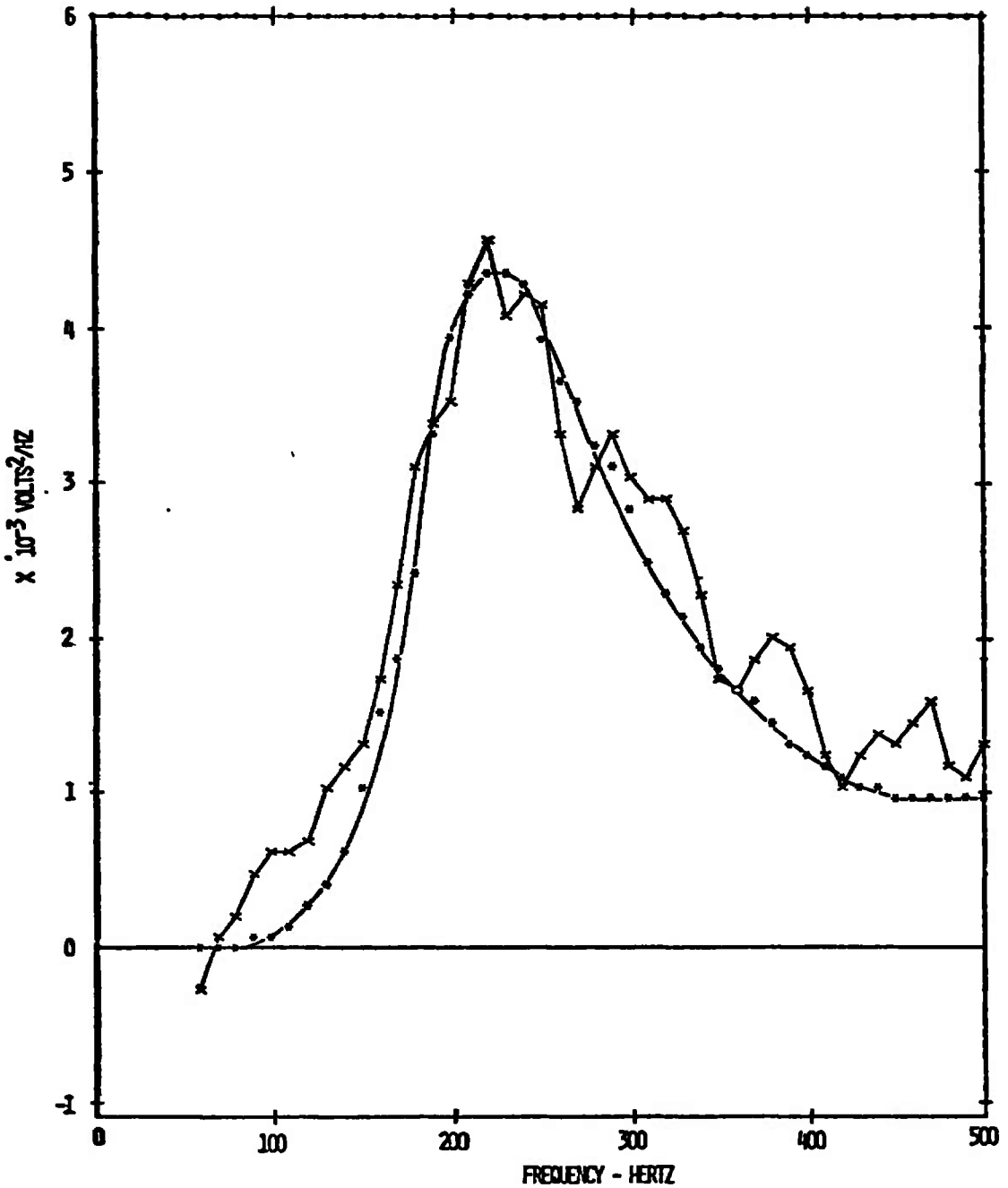


Figure I-4. Spectrum Estimate by Estimator 3 with Additional Smoothing (λ=166 samples/sec)
X - Estimate
* - Scaled Wave Analyzer Data

APPENDIX II

DERIVATIONS OF ESTIMATOR ERRORS

In this appendix we provide the details concerning the random sampling bias error of the autocovariance function, the histogram effects, and the autocovariance and spectrum estimator variability errors.

II.1 RANDOM SAMPLING BIAS ERROR

Equation (2.24) for $\hat{C}(0)$ is unbiased. Equation (2.22) for $\hat{C}(k)$, $k \neq 0$, may be expressed as

$$\hat{C}(k) = \frac{1}{H(k)} \sum_{m=1}^{H(k)} C_m \tag{II-1}$$

where C_m is a $u_i u_j$ product for which

$$\tau_m = |t_i - t_j| \tag{II-2}$$

satisfies the constraint of equation (2.23). Using bracket notation for expectation and using conditional expectation, we obtain

$$\left\langle \hat{C}(k) \right\rangle = \left\langle \frac{1}{H(k)} \sum_{m=1}^{H(k)} \cdot \left\langle C_m | \tau_m, H(k) \right\rangle \right\rangle_{\tau_m, H(k)} \tag{II-3}$$

But the expectation of C_m depends only on the velocity process $U(t)$ and the time difference τ_m. The result is

$$\left\langle C_m | \tau_m, H(k) \right\rangle = \left\langle C_m | \tau_m \right\rangle = C(\tau_m) \tag{II-4}$$

Taking the expectation with respect to τ_m inside the summation gives

$$\left\langle \hat{C}(k) \right\rangle = \left\langle \frac{1}{H(k)} \sum_{m=1}^{H(k)} \int_{-\infty}^{\infty} C(\tau)\ p_k(\tau)\ d\tau \right\rangle_{H(k)} \qquad \text{(II-5)}$$

where $p_k(\tau)$ is the conditional probability density for the value of τ_m, given that τ_m lies in the $\pm\Delta\tau/2$ interval about $k\Delta\tau$. Since the integral does not depend on H(k), the summation produces H(k) identical terms and the result is as given in (2.26).

The function $p_k(\tau)$ may be obtained as

$$p_k(\tau) = \frac{p_N(\tau)}{\left[\int_{k\Delta\tau - \Delta\tau/2}^{k\Delta\tau + \Delta\tau/2} p_N(\tau)\ d\tau\right]} \qquad \text{(II-6)}$$

where $p_N(\tau)$ is the conditional probability density function of τ_m given that N consecutive samples were obtained. The time differences τ_m may be divided into N-1 exclusive sets according to the order of the time difference: there are N-1 first differences, $t_i - t_{i-1}$; N-2 second differences, $t_i - t_{i-2}$; ...; and one N-1 difference, $t_n - t_1$. The total set of time differences is the set union of N-1 mutually exclusive sets. We may therefore express the density $p_N(\tau)$ as a probability weighted sum over the respective densities for first, second, third, etc., time differences. The required functions $f_i(\tau)$ are known for the case of stationary Poisson sampling.* The result is

$$p_N(\tau) = \sum_{i=1}^{N-1} P_i\ f_i(\tau) \qquad \text{(II-7)}$$

---

*Modifications of the Erlang functions. See page 558 of Papoulis [21].

where P_i = probability of ith order difference and

$$f_i(\tau) = \begin{cases} \frac{\lambda^i \tau^{i-1}}{(i-1)!} \exp[-\lambda\tau] & \tau > 0 \\ 0 & \tau \leq 0 \end{cases} \tag{II-8}$$

and where P_i can be obtained by dividing the number of ith order differences (N-1) by the total number of differences

$$\sum_{i=1}^{N-1} (N-i) = \frac{N^2-N}{2} \tag{II-9}$$

with the result

$$P_i = \frac{2(N-i)}{N^2-N} \tag{II-10}$$

Finally we see that an exact expression for $p_k(\tau)$ may be obtained as a finite sum of integrals involving the $f_i(\tau)$ functions if (II-10), (II-8), and (II-7) are substituted in (II-6). The result is a complicated expression which could be approximately evaluated by power series or numerical integration; but this is not yet been undertaken.

II.2 Histogram H(k)

Using the same type of derivation as in the previous section we would obtain

$$H(k) = \frac{N^2-N}{2} \int_{k\Delta\tau - \Delta\tau/2}^{k\Delta\tau + \Delta\tau/2} p_N(\tau)\, d\tau \tag{II-11}$$

This expression has not been evaluated and compared with experiment.

Using the approximate alternate formulation as in equations (2.25) with the conditions for (2.7), we obtain the result given in equation (2.27) after noting that

$$\langle H(k) \rangle = \sum_{i=0}^{N'-k} \langle I_i \; I_{i+k} \rangle = (N'-k) \; p^2 \qquad \text{(II-12)}$$

by replacing $N' = T/\Delta\tau \approx N/(\lambda\Delta\tau)$. The extent of validity of the approximate result is not known, but good agreement with experiment was obtained in the experiments performed.

II.3 VARIANCE OF THE AUTOCOVARIANCE ESTIMATE

In addition to the assumptions of Section 2.1.1 several more assumptions are used.

a) $\Delta\tau$ is small enough to prevent bias error
b) RMS deviation of H(k) is small (large N)
c) H(k) = constant, $k < M$ (large N)
d) individual product pairs are independent events
e) the true mean is known ($\hat{U} \approx \overline{U}$)

Assumption d) is not strictly ever valid; however, for sufficiently low values of λ there are negligibly few instances in which two product pairs with the same lag value occur close enough together in time to be appreciably correlated. It was initially assumed that $\lambda/2B \ll 1$ would be required to obtain d). The question has not been resolved analytically, but the following results appear to be valid experimentally for values of λ up to 2B (and even higher for the stopband frequency components.)

Under our assumptions $\hat{C}(k)$ is unbiased so that

$$\langle \varepsilon_\tau(k) \rangle = \langle \hat{C}(k) - C(k) \rangle = 0 \tag{II-13}$$

and the mean square error is

$$\sigma^2_{ck}(k) = \left\langle \frac{1}{H^2(k)} \sum_{m=1}^{H(k)} \sum_{n=1}^{H(k)} \langle C_m \, C_n | \tau_m, \tau_n \rangle \right\rangle_{H(k)} - C^2(k) \tag{II-14}$$

with C_m and C_n each being product pairs, as in (II-1), which are independent of $H(k)$. Expanding one term of the summation we have $H(k)$ terms where $m = n$:

$$\langle u^2(t) \; u^2(t + \tau_m) \; | \tau_m \rangle = \sigma^4 + 2 \; C^2(\tau_m) \tag{II-15}$$

and $H^2(k) - H(k)$ terms $m \neq n$ which with the assumed independence gives

$$\begin{aligned} & \langle u(t) \; u(t+\tau_m) \; u(t') \; u(t'+\tau_n) \rangle \\ = \; & \langle u(t) \; u(t+\tau_m) \rangle \; \langle u(t') \; u(t'+\tau_n) \rangle \\ = \; & C(\tau_m) \; C(\tau_n) \end{aligned} \tag{II-16}$$

Assumption a) allows replacement of τ_m, τ_n by $k\Delta\tau$ with negligible effect so that (II-14) becomes

$$\begin{aligned} \sigma^2_{ck} = & \left\langle \frac{1}{H^2(k)} \; \{H(k) \; [\sigma^4 + 2C^2(k)] + \right. \\ & \left. [H^2(k) - H(k)] \; C^2(k)\} \right\rangle_{H(k)} - C^2(k) \\ = & \left\langle \frac{1}{H(k)} \right\rangle_{H(k)} \; [\sigma^4 + C^2(k)] \end{aligned} \tag{II-17}$$

We observe that equation (2.31) results if the RMS deviation of H(k) is small so that

$$\left\langle \frac{1}{H(k)} \right\rangle \simeq \frac{1}{\langle H(k) \rangle} \tag{II-18}$$

II.4 VARIANCE OF THE SPECTRAL ESTIMATE

With the additional assumption that $M\Delta\tau$ is large enough so that negligible smoothing bias error exists, the error $\varepsilon_f(i)$ in the spectral estimate at $i\Delta f$ is zero mean. The mean square error is obtained by inserting equation (II.13) in equation (II-33) squaring the result, expanding as a double sum as

$$\sigma_f^2(i) = 4\Delta\tau^2 \sum_{m=0}^{M} \sum_{n=0}^{M} w'(m)\ w'(n) \left\langle \varepsilon_\tau(m)\ \varepsilon_\tau(n) \right\rangle \cos\left[\frac{im\pi}{M}\right] \cdot \cos\left[\frac{in\pi}{M}\right] \tag{II-19}$$

where

$$w'(k) = w(k),\ k \neq 0;\ w'(k) = 1/2,\ k=0 \tag{II-20}$$

But since $\hat{S}(I)$ is unbiased and $\varepsilon_\tau(m)\varepsilon_\tau(n) = 0$ for $m \neq n$ because of the assumed independence the result becomes

$$\sigma_f^2(i) = 2\Delta\tau^2 \sum_{k=0}^{M} w'^2(k)\ \sigma_{ck}^2(k)\ \left(1 + \cos\left[\frac{2ik\pi}{M}\right]\right) \tag{II-21}$$

neglecting the fact that $w'^2(0) = 1/4$ and that $w^2(0) = 1$, equation (2.41) results from inserting equation (2.27) into equation (2.31), with (2.28) satisfied, then using equation (2.31) in (II.21).

APPENDIX III

EXPERIMENTAL FACILITIES AND DATA PROCESSING

III.1 ELECTRONIC SIMULATION FACILITY

III.1.1 Introduction

The electronic simulation facility used the random emissions from a radioactive source to control an analog-to-digtal converter (ADC) and a special purpose time-interval counter. The ADC produced sample magnitudes and the time-interval counter counted the number of clock pulses between samples to indicate the time between consecutive samples. The digital data from the ADC and the time-interval counter was read by a minicomputer and stored on magnetic tape for off-line processing. The system block diagram in Figure III-1 shows the general arrangement of equipment. Each component is discussed below.

III.1.2 Radioactive Source and Proportional Counter

The radioactive source was a laboratory sample of carbon 14 of .001 μc activity. The gas proportional counter detected the emission of alpha and beta particles and generated a -1.5 v pulse with duration 1 μs for each detection event. The dead time between events was of the order of a few microseconds. The number of events per unit time was assumed to have a Poisson distribution with mean rate λ. The mean rate was controlled by covering the

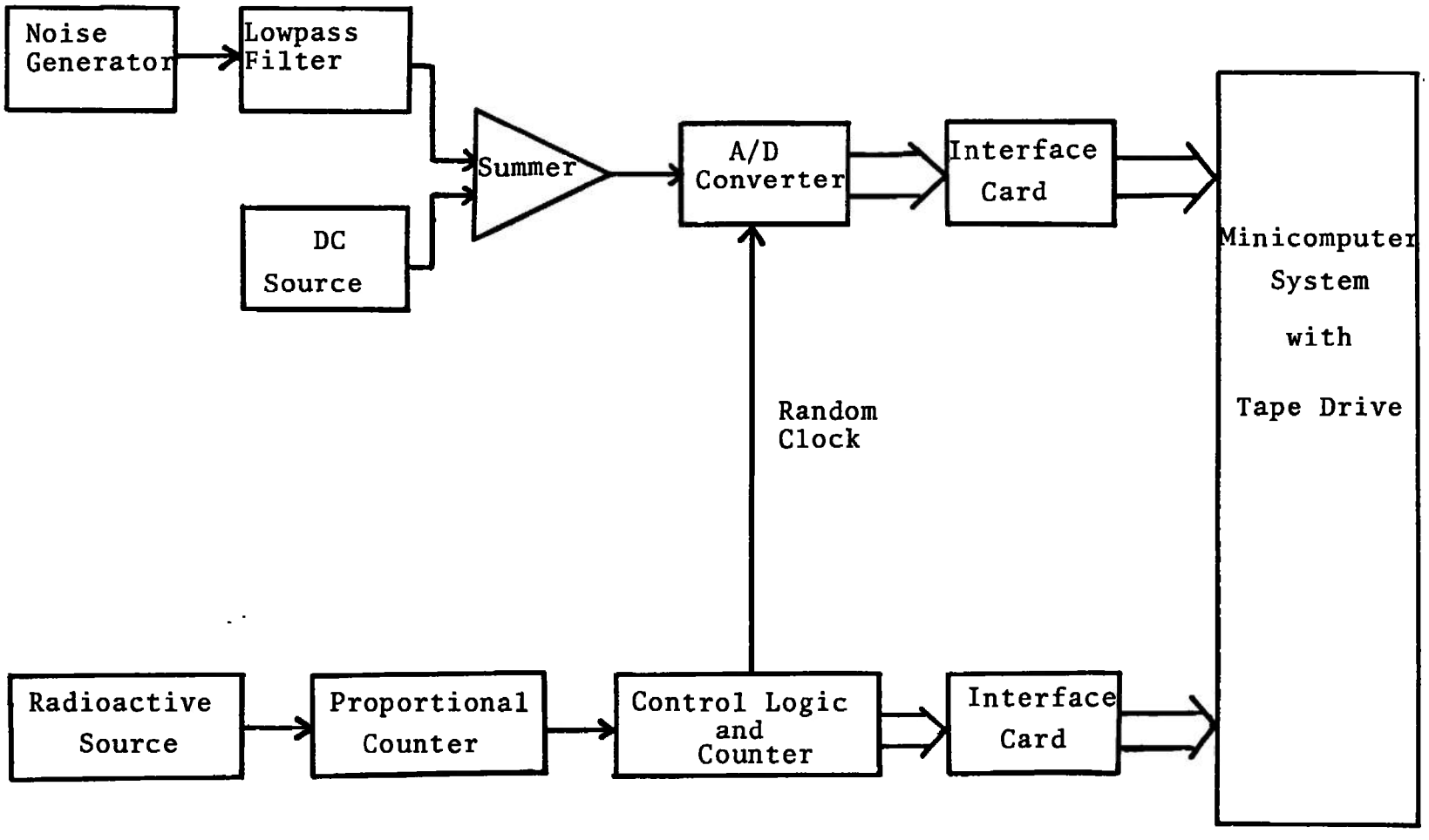


Figure III-1. System Block Diagram.

sample with various thicknesses of aluminum foil, while maintaining a constant voltage on the proportional counter electrode of approximately 2000 volts. The mean rate was stable under these conditions.

III.1.3 Control Logic and Counter

This component contained three sections:

1) circuitry to interface the output pulse of the gas proportional counter to TTL digital logic,

2) TTL logic to generate control pulses for the ADC and the time-interval counter,

3) a time-interval counter consisting of standard TTL packages and a Schmitt trigger clock to drive the counter.

The time-interval counter was 16 bits (15 magnitude bits and a sign bit). The sign bit was included solely for convenience in interfacing with Hewlett-Packard logic. The time resolution of the counter was approximately 2 μs (the period of the clock) and the upper range was $(2^{15}-1) \times 2\ \mu s = .066$ seconds.

III.1.4 Noise Generator and Filter

The noise generator provided a zero mean Gaussian signal which was approximately "white" over the range 5 Hz to 20 KHz. The filter was an adjustable fourth order Butterworth active filter that could be used in a lowpass, bandpass or highpass mode. The 3 db cut-off frequencies were variable from 10 Hz to 1.1 MHz. The roll-off rate was 24 db/octave. In the lowpass mode, experimental measurements

with a wave analyzer showed that the response of the filter was not flat below approximately 20 Hz. This was verified in the spectral estimates. The wave analyzer measurements also showed that the cutoff frequencies indicated by the filter dial settings were low by 5%. Figure III-2 shows a plot of wave analyzer data (dashed line) and a plot of the theoretical Butterworth curve (solid line) for comparison.

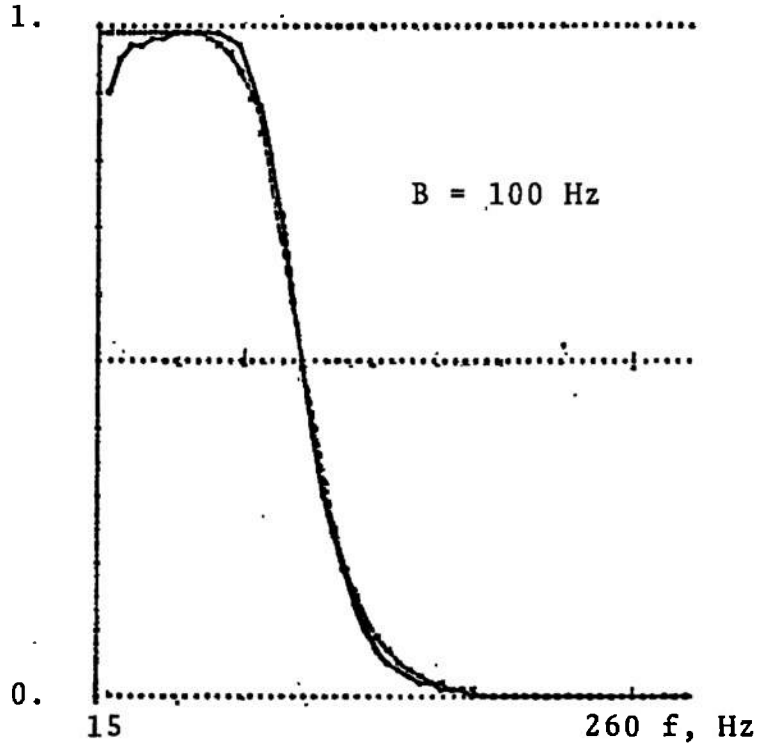
III.1.5 DC Source and Summer

The DC Source was simply a DC power supply. The DC level was added to the output of the filter to provide a non-zero mean signal. To perform the addition, a summer was built with a general-purpose operational amplifier. The summer provided a gain of approximately 100. Adjustment of the DC power supply and the output level of the noise generator provided a nonzero mean signal, with variable turbulence intensity and known spectrum, in the input range of the ADC.

III.1.6 Analog to Digital Converter

The ADC had an input range of ± 5 volts and the digital output consisted of nine magnitude bits and a sign bit representing a decimal range of -511 to +512. Negative magnitudes were in two's complement form. The ADC required a 12 μs trigger pulse which was generated by the control logic described earlier.

The minimum sampling interval was approximately 60 μs, consisting of 50 μs dead time for the ADC (maximum)



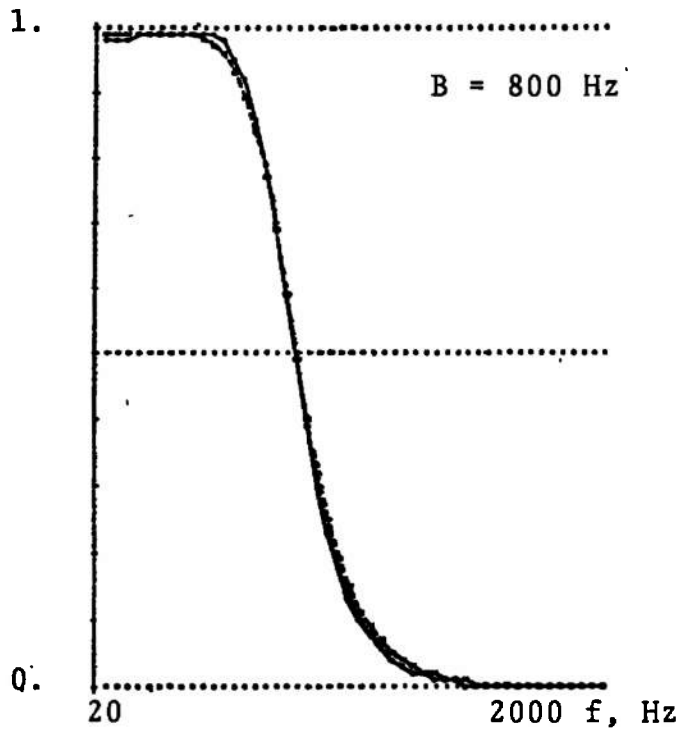


Figure III-2. Wave Analyzer Data and Analytical Butterworth Spectra.

and a 10 μs allowance for execution time of computer instructions. Randomly spaced samples occurring at smaller time intervals were ignored.

III.1.7 Minicomputer

The data was collected with a Hewlett-Packard 2114B minicomputer (8K memory) and recorded on a Hewlett-Packard 7970B Digital Tape Unit in integer format in data segments consisting of 500 words (250 samples) per segment. An existing program (ADCMT) written at Texas A&M was modified for use with the ADC and the time interval counter.

III.1.8 Instrumentation

A laboratory time-interval counter was used to count the number of random pulses in a given interval to determine the average sample rate. A laboratory wave analyzer was used to measure the distribution of power vs frequency at the output of the analog filter. The magnitude of the time-varying component was measured with a true rms voltmeter.

III.2 DATA PROCESSING

III.2.1 Data Files

The data for each experimental run, recorded on a Hewlett-Packard Minicomputer system, was taken to an IBM 360/65 facility (with IBM 3420 tape drives) for transfer

to a Master Data Tape. The systems were compatible for integer data, except for word length (16 bits/word for Hewlett-Packard; 32 bits/word for IBM 360/65). Each experimental run was assigned a separate file on the Master Data Tape.

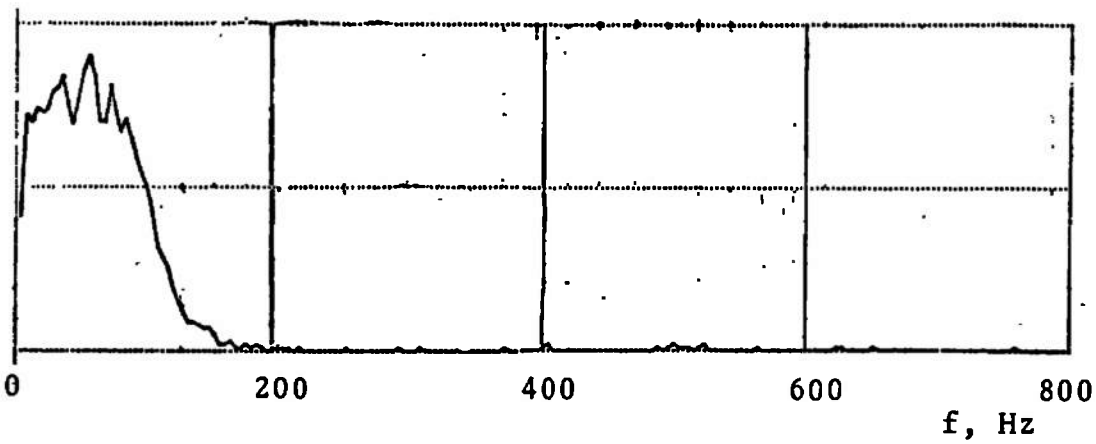
III.2.2 Software

Programs were written in Fortran IV to batch process the experimental data in data segments of 500 words/segment, where 1 word = 16 bits. The autocovariance estimator described in 2.3.1, eq. (2.22), was implemented as a subroutine by keeping running accumulations in the arrays H(k) and SUM(k) until a predetermined number of data segments had been processed. The estimate $\hat{C}(k)$ was then formed, followed by the spectral estimate $S(i\Delta f)$ computed from eq. (2.33). The spectral estimation subroutine used is reproduced on the following pages.

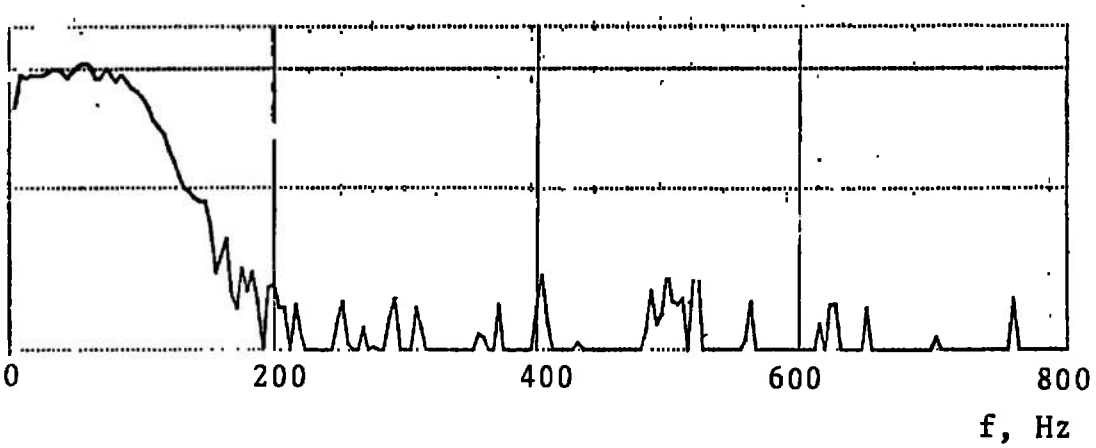
Other subroutines were written to read data segments from the Master Data Tape, to compute the sample time of each sample assuming t = 0 at the beginning of each data segment, to compute segment mean, and to plot the various arrays and estimates on a line printer. Execution times of the order to 45 sec were encountered when processing 20 data segments with this package, but this could be decreased considerably by reprogramming with use of integer arithmetic operations.

III.2.3. Typical Output

Figure III-3, a and b, illustrate the difference between a linear plot and a logarithmic plot. Because of the symmetrical type of error produced by the estimator of Section 2.4, (not positive definite) linear plot routines have been used during this study. The two figures illustrate the difference in appearance for normalized RMS error level of -17 db.



III-3 a. Linear



III-3 b. Logarithmic

Figure III-3. Linear and Logarithmic Plots of a Spectrum Estimate.

```
      SUBROUTINE RASPEC (NUM,NUM1,M,DELTAU,AMPL,T,SUM,H,C,PS,NF,
     1KLM,KGM,JMAX,STEP)
C     MODEL 5   REENTRANT MOD 5      JANUARY 28,1974
C     SPECTRUM ESTIMATION
C     FROM RANDOMLY SAMPLED RANDOM PROCESS
C         NUM - NUMBER OF DATA WORDS IN EACH RECORD
C         NUM1 - NUM + 1
C         M - NUMBER OF DELTA TAU SLOTS USED IN HISTOGRAM
C         DELTAU - DELTA TAU - TIME LENGTH OF EACH SLOT
C         AMPL - INPUT DATA AMPLITUDE MATRIX
C                (RECORD MEAN ALREADY SUBTRACTED
C         T - TIME MATRIX - ELAPSED TIME CORRESPONDING TO EACH AMPL(I)
C         SUM - MATRIX OF SUBTOTALS OF AMPLITUDES CORRESPONDING TO EACH
C               HISTOGRAM SLOT
C         H - MATRIX OF SUBTOTAL NUMBER OF ELEMENTS IN EACH SLOT
C               CORRESPONDS TO SUM MATRIX
C         C -  AUTOCOVARIANCE MATRIX
C         PS - POWER SPECTRUM MATRIX OF LENGTH NF
C         NF - NUMBER OF FREQUENCIES TO BE EVALUATED FOR POWER SPECTRUM
C         KLM - NUMBER OF LAGS LESS THAN OF EQUAL TO M
C         KGM - NUMBER OF LAGS GREATER THAN M
      CONTINUE
C     THE FULL SUBROUTINE MUST BE CALLED FIRST TO INITIALIZE ALL STORAGE
C     LOCATIONS.  CONTROL IS RETURNED IMMEDIATELY
C     PSTEP2 IS CALLED EACH TIME FOR SOME NUMBER OF RECORDS OF INPUT
C     CONTROL IS RETURNED AFTER THE HISTOGRAM AND SUM MATRIX ARE COM-
C     PLETED FOR THAT RECORD
C     PSTEP3 IS CALLED TO CREATE THE C MATRIX FROM THE ACCUMULATED SUM
C     AND H MATRICES.  RETURNS C
C     PSTEP4 MULTIPLIES C BY THE W RAMP.  THE W RAMP STARTS AT 1 AND
C     FALLS TO 0 IN IH STEPS AND REMAINS 0 FOR THE REST OF TH C MATRIX
C     PSTEP5 CREATES THE POWER SPECTRUM ESTIMATE FOR NF FREQUENCIES
C         ISTART - STARTING FREQUENCY DIVIDED BY 2.5
C         STEP - MULTIPLE OF 2.5 TO BE STEP IN FREQUENCY
C     POWER SPECTRUM IS RETURNED AND SUBROUTINE IS TERMINATED
```

```
      INTEGER JMAX,H(M),CO,STEP
      REAL SUM(M),C(M),PS(NF),T(NUM1),AMPL(NUM)
C
C     INITIALIZE CONSTANTS AND MATRICES
C
      SUMO=0.
      HO=0
      KLM=0
      KGM=0
      DO 5 I=1,M
      SUM(I)=0.
    5 H(I)=0
      RETURN
      ENTRY PSTEP2 (NUM,AMPL,T,SUM,H,M,NUM1,KLM,KGM,JMAX)
C
C     PRODUCE SUM AND H ARRAYS, KEEPING TRACK OF KLM AND KGM
C
      DO 10 I=1,NUM
   10 SUMO=SUMO+AMPL(I)*AMPL(I)
      HO=NUM+HO
      DO 20 J=1,JMAX
      NP=NUM-J
      DO 30 I=1,NP
      I1=I+1
      DELTA=T(I1+J)-T(I1)
      K=DELTA*(1./DELTAU)+1.5
      IF(K.GT.M)GO TO 25
   29 H(K)=H(K)+1
      SUM(K)=SUM(K)+AMPL(I+J)*AMPL(I)
      KLM=KLM+1
      GO TO 30
   25 KGM=KGM+1
   30 CONTINUE
   20 CONTINUE
      SUM(1)=SUMO
      H(1)=HO
```

```
      RETURN
      ENTRY PSTEP3 (SUM,H,C,M)
C
C     PRODUCE AUTOCOVARIANCE MATRIX
C
      MM1=M-1
      DO 40 K=2,MM1
      IF(H(K).NE.0) GO TO 40
C
C     IF H(K)=0, THEN STRAIGHT LINE INTERPOLATE BETWEEN H(K-1) AND H(K+1)
C
      SUM(K)=SUM(K-1)+SUM(K+1)
      H(K)=H(K-1)+H(K+1)
   40 CONTINUE
      IF(H(M).NE.0) GO TO 46
C
C     IF H(M)=0, THEN H(M)=H(M-1)
C
      SUM(M)=SUM(M-1)
      H(M)=H(M-1)
   46 DO 50 K=1,M
   50 C(K)=SUM(K)/H(K)
      RETURN
      ENTRY PSTEP4 (C,IH,M)
C
C     PRODUCE W RAMP (BARTLETT WINDOW) AND MULTIPLY C MATRIX BY THIS
C     WEIGHTING FUNCTION
C
      DO 60 K=1,IH
      W=1.-FLOAT(K-1)/FLOAT(IH)
   60 C(K)=W*C(K)
      IH1=IH+1
      DO 61 K=IH1,M
   61 C(K)=0.
      RETURN
      ENTRY PSTEP5 (NF,ISTART,STEP,PS)
```

```
C     ESTIMATE POWER SPECTRUM FOR DESIRED FREQUENCIES
C
      DO 70 J=1,NF
      I=ISTART+STEP*(J-1)
      SUMT=0.
      DO 80 K=2,M
   80 SUMT=SUMT+C(K)*COS(3.14159*I*FLOAT(K-1)/FLOAT(M))
      SUMT=SUMT+SUMT+C(1)
   70 PS(J)=DELTAU*SUMT
      RETURN
      END
```