2003

Statistical signal processing for mechanical systems

Li Wen
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd
Part of the Applied Mechanics Commons, and the Statistics and Probability Commons

Recommended Citation
Wen, Li, "Statistical signal processing for mechanical systems" (2003). Retrospective Theses and Dissertations. 752.
https://lib.dr.iastate.edu/rtd/752

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Statistical signal processing for mechanical systems

by

Li Wen

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Engineering Mechanics

Program of Study Committee:
Peter J. Sherman, Major Professor
   Julie A. Dickerson
   Adin Mann III
   Lester W. Schmerr
   Yuhong Yang

Iowa State University
Ames, Iowa
2003
Copyright © Li Wen, 2003. All rights reserved.
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
This is to certify that the doctoral dissertation of

Li Wen

has met the dissertation requirements of Iowa State University

Committee Member

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Committee Member

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program
TABLE OF CONTENTS

LIST OF FIGURES ........................................ vii
LIST OF TABLES ........................................... xiv
NOMENCLATURE ........................................... xvi
ACKNOWLEDGEMENTS ..................................... xx
ABSTRACT .................................................. xxi

CHAPTER 1 GENERAL INTRODUCTION ..................... 1

1.1 Signals from Mechanical Systems .................... 1
  1.1.1 Why Should Measurements of Signals from Mechanical Systems
        Be Treated as Sample Realizations of Random Processes .......... 3

1.2 Categorization of Random Processes ................... 4

1.3 Spectral Analysis of Random Processes ................. 5
  1.3.1 Fourier Series for a Deterministic Periodic Signal ............ 6
  1.3.2 Fourier Integral of a Transient Signal ...................... 7
  1.3.3 Spectral Analysis of Stationary Random Processes with Transient
        Autocorrelation Functions ...................................... 7
  1.3.4 Spectral Analysis of Stationary Processes with Periodic Autocor-
        relation Functions ........................................ 8
  1.3.5 Mixed Random Processes with Both Discrete and Continuous Spec-
        trum ..................................................... 9

1.4 Continuous-Time and Discrete-Time Signals ............ 11
CHAPTER 2 USING VARIABILITY RELATED TO FAMILIES OF SPECTRAL ESTIMATORS FOR MIXED RANDOM PROCESSES

2.1 Abstract ...................................................... 16
2.2 Introduction ................................................. 17
2.3 The Structure of P.s.d. Estimator for Mixed Spectrum Processes ......... 19
2.4 Arithmetic Variability of Theoretical FT(n), PER(n) and AR(n) Spectral Families ...................................................... 25
  2.4.1 Variability of the Theoretical FT(n) Spectra ........................ 25
  2.4.2 Variability of the Theoretical PER(n) Spectra ...................... 31
  2.4.3 Variability of the Theoretical AR(n) Spectra ...................... 34
2.5 Variability of the Estimated FT(n), PER(n) and AR(n) Spectra ........... 37
  2.5.1 Statistical Variability of the FT(n) Family of P.s.d. Estimators . 38
  2.5.2 Statistical Variability of the PER(n) Family of P.s.d. Estimators . 41
  2.5.3 Statistical Variability of the AR(n) Family of P.s.d. Estimators . 43
2.6 Statistical Properties of Averages of Families of P.s.d. Estimators ....... 44
  2.6.1 The FT(n) Spectra and Their Averages ........................... 46
  2.6.2 The PER(n) Spectra and Their Averages ........................... 47
  2.6.3 The AR(n) Spectra and Their Averages ........................... 49
2.7 Application to the Westland Helicopter Vibration Data .................... 51
2.8 Summary and Conclusions ....................................... 56

CHAPTER 3 ON THE INFLUENCE OF SAMPLING AND OBSERVATION TIMES ON ESTIMATION OF THE BANDWIDTH PARAMETER OF A GAUSS-MARKOV PROCESS ........................ 59
3.1 Abstract ...................................................... 59
3.2 Introduction ................................................. 60
3.3 Estimation Procedure ............................................... 64
3.4 Autocorrelation Function Estimation .............................. 66
3.5 Statistics of $\alpha$ Estimator, $\hat{\alpha}$ ............................ 73
3.6 Statistics of $\beta$ Estimator, $\hat{\beta}$ .............................. 77
3.7 Guidelines in Selecting $\Delta$ ....................................... 82
3.8 Application in Prediction ............................................ 85
  3.8.1 Prediction of a Time Invariant GM Process for Various Prediction steps ....................................... 85
3.9 Conclusion .................................................................... 102
3.10 Appendix ..................................................................... 104
  3.10.1 Derivation of Bias($\hat{\alpha}$) Based on a 2$^{nd}$ Order Taylor Expansion ...................... 104
  3.10.2 Derivation of the Variance of $\hat{\alpha}$ Based on a 2$^{nd}$ Order Taylor Expansion ............. 105

CHAPTER 4 APPLICATION OF PREVIOUS RESULTS ON SOUND PRESSURE DATA ANALYSIS ........................................ 107
  4.1 The Objective of the Measurement ................................. 107
  4.2 Description of the Experiment and Measurement Data Set ................................. 109
  4.3 Noise Generation Mechanism of the Engine Cooling System ............................... 112
    4.3.1 Air Flow Across the Cooling Radiator ................................ 112
    4.3.2 Fan Noise .......................................................... 115
    4.3.3 Influence of the Engine Block ................................... 117
  4.4 Preliminary Analysis of the Sound Pressure Measurement ............................. 118
  4.5 Octave Band Analysis of Sound Pressure Process with and without Mock Engine ........................................ 123
    4.5.1 Using Bandpass Filter to Obtain the Octave Band Level ................................. 123
LIST OF FIGURES

Figure 2.1 Arithmetic mean and $2-\sigma$ curves corresponding to use of a family of theoretical FT(n) spectra. [Note: Where not shown, the lower $2-\sigma$ curve is $-\infty$.] ........................................ 21

Figure 2.2 (a) $2-\sigma$ curves for an average of 5 AR(n) spectra .................. 23

Figure 2.3 (b) $2-\sigma$ curves for an average of 80 AR(n) spectra. ................. 24

Figure 2.4 Example of Dirichlet kernel for n=10,20, and 30 ....................... 27

Figure 2.5 Average of FT(n) theoretical spectra for the range 1 $\rightarrow$ $n_{\text{max}}$ for $n_{\text{max}} = 100, 200$ and 500. .................................................. 28

Figure 2.6 Example of Fejer kernel for n=10, 20 and 30 in dB .................... 30

Figure 2.7 Comparison of the $n^{th}$ and the average of the first $n$ Fejer's kernels for n =100 and 1000 ................................................. 33

Figure 2.8 Arithmetic mean (top) and standard deviation (bottom) of collection of AR(k): $k = 2, \cdots, m$ spectra for m = 20, 40, 80, and 160 ................................................................. 36

Figure 2.9 Comparison of simulation results against predictions in (2.30) at tone's frequency $f=0.25$Hz .................................................. 40

Figure 2.10 Comparison of simulation results against predictions in (2.31) at noise frequency $f=0.1$Hz ................................................. 40
Figure 2.11 Evaluation of the variance expression in (2.36) as a function of model order at a noise frequency. Also shown is the estimated variability of the average of AR spectra. (See Section 2.6 for related discussion). ........................................... 44

Figure 2.12 Comparison of predicted (using (refeq:sardist0)) $2 - \sigma$ regions associated with an AR(40) model, and an average of AR(p) model for $p = 2 : 160$. Also shown is the estimated variability of the average of AR spectra (see Section 2.5 for related discussion) . . . 45

Figure 2.13 (a) (Top) Comparison of statistical standard deviation of single order PER(n) and the averaged PER(n) spectral estimates for selected order n at non-tone frequency $f = 0.1 \text{ Hz}$. (b) (Bottom) Comparison of statistical standard deviation of single order PER(n) and the averaged PER(n) spectral estimates for selected order n at tone frequency $f = 0.3 \text{ Hz}$. .................................................. 48

Figure 2.14 (a) (Top) Comparison of statistical standard deviation of single order AR(n) and the averaged AR(n) spectral estimates for selected order n at non-tone frequency $f = 0.1 \text{ Hz}$. (b) (Bottom) Comparison of statistical standard deviation of single order AR(n) and the averaged AR(n) spectral estimates for selected order n at tone frequency $f = 0.3 \text{ Hz}$. .................................................. 50

Figure 2.15 Comparison of 95% (or $2 - \sigma$) confidence interval of PER(n) for n = 256 and 1024 for Westland data ......................... 52

Figure 2.16 Comparison of $2 - \sigma$ of PER(n) for n = 256 and 1024 for Westland data. .............................................................. 53

Figure 2.17 Comparison of 95% (or $2 - \sigma$) confidence interval of AR(n) for n = 20 and 40 for Westland data ............................ 55
Figure 2.18 Comparison of standard deviation of AR(n) for n = 20 and 40 for Westland data. ................................. 55

Figure 3.1 Relative error of $\text{cov}\{\hat{R}_r, \hat{R}_{r+v}\}$ for $(r, v) = (0, 0) (0, 1) (1, 0)$ and $\beta T = 20$ ............................................. 67

Figure 3.2 The predicted variance from (3.8) of autocorrelation function of $X(t)$ sampling at different time interval $\Delta$. ......................... 69

Figure 3.3 The sample and gamma and normal fitted distribution of $\hat{R}_1$ at $\beta \Delta = 0.0098$, $n = 32$ ........................................... 71

Figure 3.4 The sample and gamma and normal fitted distribution of $\hat{R}_0$ at $\beta \Delta = 0.0098$, $n = 32$ ........................................... 72

Figure 3.5 The sample and predicted bias $\hat{\alpha}$ from 1st and 2nd expansion at different sampling rates. ............................... 75

Figure 3.6 The sample and predicted variance of $\hat{\alpha}$ from 1st and 2nd expansion of $\hat{\alpha}$ at different sampling rates with $T = 30/\beta$. .......... 76

Figure 3.7 The absolute relative difference of predicted variance of $\hat{\alpha}$ from 1st and 2nd expansion from the sample variance at different sampling rates with data length $T = 30/\beta$. ................................. 77

Figure 3.8 The sample, normal and beta fitted pdf of $\hat{\alpha}$ for $\beta = 10$, $\beta T = 20$ $N=2037$, $\beta \Delta = 0.0098$. ................................. 78

Figure 3.9 Sample and predicted Bias($\hat{\beta}$), based on 1st and 2nd expansion for $\beta = 10$, $\beta T = 30$ ........................................ 80

Figure 3.10 Sample and predicted Var($\hat{\beta}$), based on 1st in (3.32) and 2nd expansion in (3.35) with approximation and in (3.34) without approximation, for $\beta = 10$, $\beta T = 30$. ................................. 80

Figure 3.11 Same as figure 3.10 except for $\beta = 10$, $\beta T = 60$. ............... 81
Figure 3.12 Sample and Normal $L_{Norm} = 53775$ and Gamma $L_{Gamma} = -52609$ fitted pdf of $\hat{\beta}$ for $\beta = 10$, $\beta T = 20$, $\Delta = 0.00098$, sample size $N = 2037$ from 20000 simulations. 82

Figure 3.13 Plots of the $drrmse_0$, (3.49) as a function the number of prediction steps $m$, and of $\delta \beta/\beta$ 90

Figure 3.14 Plots of the $drrmse_1$ in (3.50) as a function the number of prediction steps $m$, and of $\delta \beta/\beta$ 90

Figure 3.15 The relative error (%) for $(\beta, \hat{\beta})$ and for $(\hat{x}_m(\beta), \hat{x}_m(\hat{\beta}))$ for a GM process with $\beta = 1$, $\Delta = 0.1$, $m = 1$ and $T = 25$s. 91

Figure 3.16 Same as figure 3.15 except for $\Delta = 0.01$, $m = 10$. 92

Figure 3.17 Plots of the $srrmse_0$ versus $\beta T$ at different $\beta \Delta$ values for $m \beta \Delta = 0.1$. 94

Figure 3.18 Plots of the $srrmse_0$ versus $\beta \Delta$ at different $m \beta \Delta$ values for $\beta T = 100$. 95

Figure 3.19 Plots of the $srrmse_0$ versus $\beta \Delta$ at different $m \beta \Delta$ values for $\beta T = 20$. 95

Figure 3.20 Plots of the $srrmse_0$ versus $m \beta \Delta$ at different $\beta T$ values for $\beta \Delta = 0.04$. 96

Figure 3.21 Plots of the $srrmse_1$ versus $m \beta \Delta$ at different $\beta T$ values for $\beta \Delta = 0.04$. 97

Figure 3.22 Contour plot of the $srrmse_0$ versus $\beta T$ and $m \beta \Delta$, with $\beta T \in (20, 200)$. 98

Figure 3.23 Same as figure 3.22, except that $\beta T \in (20, 40)$. 99

Figure 3.24 An active noise cancellation system. 100

Figure 3.25 Sample prediction $mse(\hat{\beta})$ versus $\beta t_0$ for a GM process with $\beta = 10$, $\beta T = 200$. 102
Figure 4.1 The front view of the radiator ........................................... 108
Figure 4.2 The right side view of experimental stand including the radiator,
fan and shroud, mock engine, shaft and shield ................................. 110
Figure 4.3 The sketch of airflow passing through a single tube and fin of the
radiator ......................................................................................... 113
Figure 4.4 A detailed kalman vortex formed from airflow passing an cylinder
when $Re > 300.$ ................................................................. 114
Figure 4.5 Strouhal number for vortex shedding frequency behind a circular
cylinder ...................................................................................... 114
Figure 4.6 The time signal of the sound pressure $x(t)$ in mock16000.dat . . 118
Figure 4.7 The estimated PSD(4096) and its 95% C.I. of the measure sound
pressure process $x(t)$ in mock16000.dat ................................. 120
Figure 4.8 The 0 – – 300 Hz estimated PSD(4096) and its 95% C.I. of the measured sound pressure process $x(t)$ in mock16000.dat ....... 122
Figure 4.9 The octave band sound pressure dB(A), with mock engine. . . . 125
Figure 4.10 The 1/3 octave band sound pressure dB(A), with no mock engine
installed ................................................................. 126
Figure 4.11 Comparison of estimated unweighted power of each octave band
and 1/3 octave band with that of the potential harmonics of BPF
shown in original magnitude ...................................................... 128
Figure 4.12 Comparison of octave band dB(A) of $x(t)$ with and without mock
engine ................................................................. 130
Figure 4.13 Comparison of 1/3 octave band dB(A) of $x(t)$ with and without
mock engine ................................................................. 131
Figure 4.14 Comparison of the standard deviation of p.s.d. estimator from
matlab and that from our method (2.35) .................................. 137
Figure 4.15  Comparison of the 95% C.I. of p.s.d. estimator from matlab and
that from our method (2.35). ........................................... 137

Figure 4.16  2 - σ Confidence Interval of A-weighted one-third octave band
level in mock16000.dat and nomock16000.dat .......................... 138

Figure 4.17  Estimated APER(512) and APER(1024) of the decimated pro-
cess $X_4$, in mock16000.dat, with a frequency range of (0, 2016)
Hz ................................................................. 146

Figure 4.18  AR spectrum of decimated mock16000.dat, $X_4$, at order $p = 40$,
80 and 160. ......................................................... 149

Figure 4.19  MV spectrum of decimated mock16000.dat, $X_4$, at order $p = 40$,
80, 160 and 320. ..................................................... 149

Figure 4.20  Zoomed in figure of Confidence Interval of AR spectrum of dec-
imated mock16000.dat, $X_4$, at order $p = 40$, 80, 160 at the po-
tential HFs of BPF. ............................................. 151

Figure 4.21  AR spectrum of decimated mock16000.dat, $X_{80}$ with a frequency
range of (0, 1000)Hz, at order $p = 40$, 80, 160. ...................... 152

Figure 4.22  MV spectrum of decimated mock16000.dat, $X_{80}$ with a frequency
range of (0, 1000)Hz, at order $p = 40$, 80, 160. ..................... 152

Figure 4.23  AR spectrum of heterodyned mock16000.dat, $X_{81}$ with a fre-
quency range of (800, 1600)Hz, at order $p = 40$, 80, 160. ....... 153

Figure 4.24  MV spectrum of heterodyned mock16000.dat, $X_{81}$ with a fre-
quency range of (800, 1600)Hz, at order $p = 40$, 80, 160. ....... 154

Figure 4.25  AR spectrum of heterodyned mock16000.dat, $X_{16}$ with a frequency
range of (707, 880)Hz, at order $p = 16$, 32, 64, 128 .............. 155

Figure 4.26  MV spectrum of heterodyned mock16000.dat, $X_{16}$ with a fre-
quency range of (707, 880)Hz, at order $p = 16$, 32, 64, 128 ....... 155
Figure 4.27  APER(512) and APER(1024) of decimated nomock16000.dat, 
X4 with a frequency range of (0,2016)Hz . . . . . . . . . . . . 156
Figure 4.28  AR spectrum of decimated nomock16000.dat, X4 with a fre-
quency range of (0,2016)Hz, at order p = 40, 80, 160 . . . . . 157
Figure 4.29  MV spectrum of decimated nomock16000.dat, X4 with a fre-
quency range of (0,2016)Hz, at order p = 40, 80, 160, 320 . . . . 158
Figure 4.30  AR spectrum of decimated nomock16000.dat, X80 with a fre-
quency range of (0,1000)Hz, at order p = 40, 80, 160 . . . . . 158
Figure 4.31  MV spectrum of decimated nomock16000.dat, X80 with a fre-
quency range of (0,1000)Hz, at order p = 40, 80, 160 . . . . . 159
Figure 4.32  AR spectrum of heterodyned nomock16000.dat, X81 with a fre-
quency range of (800,1600)Hz, at order p = 40, 80, 160 . . . . . 159
Figure 4.33  MV spectrum of hetordyned nomock16000.dat, X81 with a fre-
quency range of (800,1600)Hz, at order p = 40, 80, 160 . . . . . 160
LIST OF TABLES

Table 3.1 Parameter values to be used in subsequent investigations .... 87
Table 4.1 Description of measurement data sets and conditions of the measure­ment ................................................................. 111
Table 4.2 Description of each column in the measurement data sets .... .... 111
Table 4.3 The most important octave and 1/3 octave bands as well as the contained HFs of the BPF of the process in mock16000.dat .... 127
Table 4.4 The percentage of the power of tone to corresponding octave and 1/3 octave level of \( x(t) \) in mock16000.dat ............... 128
Table 4.5 Comparison of the A-weighed octave and 1/3 octave band and overall sound pressure level with and without mock engine .... 131
Table 4.6 Comparison of the estimated power at HFs of BPF of the sound pressure process with and without mock engine ............ 133
Table 4.7 Comparison of the estimated tonal power of the noise with mock engine from MV spectra and that from p.s.d. estimator .... 134
Table 4.8 Comparison of 1/3 octave and overall sound pressure level with and without mock engine from the C.I. information .......... 139
Table 4.9 Estimated parameters and their standard errors of \( x_h(t) \) in the bands (440, 565) Hz and (283, 353) Hz with and without mock engine. ................................................................. 142
Table 4.10  Test of hypothesis of $x_h(t)$ in the bands (440, 565) Hz and (283, 353) Hz under the two conditions.  
Table 4.11  The increase in dB of PER(1024) from PER(512)  

142

147
## NOMENCLATURE

### Lower Case

- \( \alpha \): continuous time AR(1) parameter
- \( \beta \): discrete time bandwidth parameter
- \( t \): time
- \( \Delta \): sampling interval
- \( f \): frequency in Hz
- \( f_v \): vortex shedding frequency in Hz
- \( f_b \): blade passing frequency in Hz
- \( \tau \): the lag of the autocorrelation function
- \( \omega \): angular frequency in rad/s
- \( \infty \): infinity

### Upper Case

- \( A \): amplitude
- \( L \): diameter of a cylinder
- \( Re \): Reynolds number
St  Strouhal number
T  observation window length
U  free speed of airflow passing a cylinder
X(t)  a continuous-time random process (r.p.), the parameter $t \in (0, T)$ is time
X_k  a discrete time r.p. with time index $k$
R  universal gas constant ($R = 1716.56 \text{ ft}^2/(\text{s}^2\text{o}^R$))
$R_x(\tau)$  autocorrelation function of the continuous time r.p. $X(t)$
$R_r$  autocorrelation sequence of lag $r$ of a discrete-time r.p. $X_k$, $r$ is an integer
$R_0$  autocorrelation sequence of lag 0
$R_1$  autocorrelation sequence of lag 1
$S_x(\omega)$  spectrum of the r.p. $x(t)$ at frequency $\omega$
$T_{avg}$  average temperature for both runs ($^\circ\text{R}$)
$V_e$  equivalent velocity (ft/s)
$E(\ )$  expectation of a r.v.
$F(\ )$  Fourier Transform operator

Acronyms

AR  autoregressive
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW</td>
<td>bandwidth</td>
</tr>
<tr>
<td>cdf</td>
<td>cumulative distribution function</td>
</tr>
<tr>
<td>CFT</td>
<td>Continuous Fourier Transform</td>
</tr>
<tr>
<td>DTFT</td>
<td>Discrete-time Fourier Transform</td>
</tr>
<tr>
<td>dB(A)</td>
<td>A-weighted power in decibel</td>
</tr>
<tr>
<td>DSP</td>
<td>Digital Signal Processing</td>
</tr>
<tr>
<td>EKF</td>
<td>Extended Kalman Filter</td>
</tr>
<tr>
<td>FT</td>
<td>Fourier Transform</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>HF</td>
<td>Harmonic Frequency</td>
</tr>
<tr>
<td>mse</td>
<td>mean squared error</td>
</tr>
<tr>
<td>mmse</td>
<td>minimum mean squared error</td>
</tr>
<tr>
<td>MSP</td>
<td>Mechanical Signal Processing</td>
</tr>
<tr>
<td>MSA</td>
<td>Mechanical Signature Analysis</td>
</tr>
<tr>
<td>p.s.d.</td>
<td>power spectral density</td>
</tr>
<tr>
<td>pdf</td>
<td>probability density function</td>
</tr>
<tr>
<td>pmf</td>
<td>probability mass function</td>
</tr>
<tr>
<td>rmse</td>
<td>root mean squared error</td>
</tr>
<tr>
<td>PER</td>
<td>periodogram</td>
</tr>
</tbody>
</table>
r.v. random variable
r.v.s random variable
r.p. Random process
r.p.es Random processes
rpm rotation per minute
WSS Wide sense stationary
WCS Wide cyclostationary
ACKNOWLEDGEMENTS

It is my pleasure to express my sincere gratitude to my major professor, Dr. Peter Sherman, who guided and encouraged me through the completion of this entire research and dissertation. It’s also my pleasure to express my appreciation to Dr. Lester Schmerr, Dr. Adin Mann, Dr. Julia Dickerson and Dr. Yuhong Yang, who served as my committee members and provided valuable suggestions in the completion of this research. In addition, I would like to thank Dr. Adin Mann for his generous help in my experimental work. I also wish to thank the Department of Aerospace Engineering who made this Ph.D. program possible. Last, but not least, my thanks go to all the peer graduate students, friends and families, for all their encouragement and support toward completion of the Ph.D. program.

The research in Chapter 2 was supported under AFOSR Grant # F4962098.
ABSTRACT

Random processes such as vibration, temperature, pressure and acoustic noise are found in all types of mechanical systems. Knowledge of these processes can lead to improved design and to improved detection methods related to faulty operation. The goal of this dissertation is to contribute to the knowledge base of such processes. Specifically, we address statistical signal processing methods that are appropriate and consistent relative to the physics associated with these systems. Two generic problems associated with random signal measurements from mechanical systems are addressed in this dissertation. Measurements from mechanical systems are treated as sample realizations of random processes which contain information about the systems.

Random processes associated with mechanical systems usually have complex spectral structure containing both continuous and line spectral components. Accordingly, the random process is called mixed random process. Therefore, one problem we address is to use variability related to families of spectral estimators for a mixed random process, rather than traditionally to use a single best one, to better characterize its spectral information. It is shown that tones are a significant source of bias and variability of families of spectral estimators, which are reported in terms of order dependence rates. Expressions for estimating statistical and arithmetic variability of three families of spectral estimators including FT(n), PER(n) and AR(n) are provided. An important and immediate application of these results is the problem of detecting tones.

A second problem addressed is the statistical problem of estimating the bandwidth parameter of a Gauss-Markov (GM) process from a realization of fixed and finite duration
at selectable sampling interval $\Delta$. The motivation is that when continuous-time processes are sampled, they are often sampled at a rate far higher than the dynamics of the underlying random process. It is commonly assumed that a faster sample rate is better. But in many real world situations, such as in design of adaptive feedback control schemes, or in fault detection, short time changes demand that only a limited amount of time be utilized as the sampling interval. Potential numerical and statistical problems also arise with fast sampling. Thus the problem of parameter estimation in relation to sampling rate is investigated here. The bias and variance expressions of the bandwidth parameter estimator are provided based on a second order expansion. Three sample rate regions, which are finite, large and very large ones, corresponding to substantial, gradual, and very slight decrease in variance respectively, are quantitatively identified. Guidelines in choosing sampling rate based on estimator performance requirement are provided.

Results from study on these two problems are used to characterize the stochastic structure of the sound pressure process from an engine cooling fan with and without the mock engine, and to arrive at a hypothesis test for deciding whether a given design change has a significant effect on the sound.
CHAPTER 1 GENERAL INTRODUCTION

The goal of this research is to place elements of mechanical signal processing on a more sound statistical foundation. In other words, it is the author’s purpose to use statistics to better characterize random signals from mechanical systems.

1.1 Signals from Mechanical Systems

Signals are detectable physical quantities or variables by which information can be carried. A wide variety of signals include human voices, radio waves, vibrations of a machinery, the blow down pressure of a turbine engine, wind speed, stock market, EEG brain waves, and so on. Before the signals can be analyzed, they need to be measured by first being transformed to time varying current or voltage signals using transducers [1] and then sampled digitally. Digital signal processing (DSP) has been developed to reveal information from digital signals and has found applications in many different areas related to mechanical systems.

Mechanical signal processing (MSP) usually deals with measured signals from mechanical systems. The measured signals transmit useful information about the mechanical systems. MSP has very broad applications. One application of MSP is in machinery condition monitoring, which is based on early warning of developing failures. If applied successfully, MSP could be extremely valuable for mission critical machinery, such as helicopter transmission, where unplanned shutdown could lead to unacceptable losses [2]. Another application is in noise and vibration abatement. MSP can be used to identify a
multiple-input/single-output model from the measured vibration and noise signals. This model can then help to determine noise sources and their propagation paths, so that an abatement method can be developed[3]. Other applications include identification of structures, production quality control, environmental testing and so forth.

Signals from mechanical systems are complex because of the complicated structures of the mechanical systems, which may contain many shafts, gears and bearings. The signals can usually be viewed in two domains – frequency domain and time domain. In the frequency domain, they contain complicated spectral structures, with both continuous broad band spectrum and many narrow peaks centered on the integral times of fundamental frequency. Those peaks could be narrow band peaks corresponding to a resonance frequency or they could be tones. Such signals can be modeled as mixed random processes. A mixed random process is one which is the sum of a regular colored stationary process plus sinusoids having different frequencies. If \( x(t) \) denotes the signal, it could be written as

\[
x(t) = \sum_{i=1}^{k} s_i(t) + n(t),
\]

where each \( s_i(t) = A_i \cos(\omega_i t + \phi_i) \) is a sinusoid on angular frequency \( \omega_i \) with amplitude \( A_i \) and initial phase \( \phi_i \). \( n(t) \) are regular colored noise.

The method of data acquisition can determine whether \( s_i(t) \) can be considered as stationary or not. The measured process would, in general, be non-stationary if it is collected synchronously with the shaft speed, i.e., every time the measurement starts from the same shaft position such that the initial phase \( \phi_i \) is constant rather than a random variable (r.v.). On the other hand, asynchronous data collection will usually result in a stationary process, since the starting position for any realization is randomly selected such that the initial phase \( \phi_i \) is a r.v. uniformly distributed over the period of shaft rotation.
1.1.1 Why Should Measurements of Signals from Mechanical Systems Be Treated as Sample Realizations of Random Processes

First of all, let’s understand the difference between a random signal and a deterministic one.

A sinusoid with a non-random phase is an example of a deterministic signal.

A scalar random process (r.p.) \(\{X(t), t \in T\}\) \(^1\) is a set of random variables indexed by the parameter set \(T\) \([4]\). The parameter \(t\) usually refers to time, although it could also refer to other variables, such as space. The r.p. is also called spatial process if \(t\) refers to space. In this thesis, only processes with the parameter time are dealt with.

A finite-time measurement of a random signal is a partial realization of the r.p.. The measurement of a signal from a mechanical system will, almost invariably, be a partial realization of a r.p., as opposed to a deterministic signal.

When we measure any signal, noises are inevitably added to the measured signal. Although experimenters have done their best to minimize all the noise producing mechanisms, the noise can not be completely eliminated. Noise by itself is random, which makes the measured signal random, so that the measured signal has to be viewed with probabilistic description.

Also, the chosen physical variable in a mechanical system very often has elements of uncertainties. For example, in the case of measuring the vibration signal on a roller-bearing housing, dynamic loads and disturbance on the housing may be random for a number of reasons. These reasons may include, but not limited to:

1. The input load of the shaft may not be constant.

2. The shaft rotational speed may be slowly time varying because of the shaft unbalance.

\(^1\)The notation \(X(t)\) does not mean there is a functional relationship between \(X\) and \(t\) like a deterministic signal. Instead, it means, at each time \(t_0\), \(X(t_0)\) is a random variable.
3. The dynamic load applied on the roller bearing may change with different contact angles of the roller.

4. If there are developing faults on the roller or ridge, the intensity of the impulse generated by the impact of the rolling element on the faults may change slowly with time [5].

Because of the randomness, the signals from mechanical systems should be analyzed carefully based on the theory of probability and random processes. Even though practitioners often treat measurements as realizations of a r.p., e.g. in p.s.d. analysis and system identification, (Bendat and Piersol [6]), these tools are not usually well-suited for signals associated with machinery; due to their mixed spectral nature (Lou [7], Lau and Sherman [8]). Furthermore, statistical tools, such as p.s.d. confidence intervals and mean squared error (mse) associated with estimation of system parameters, also become flawed.

One purpose of this dissertation is to bring the necessity of a careful statistical treatment of mechanical signals to the awareness of practical mechanical engineers. It is also the author’s desire to develop more advanced statistical tools that are specially suitable for analyzing signals from mechanical systems, given that most present DSP algorithms are designed for signals from electrical engineering.

1.2 Categorization of Random Processes

A r.p. can be categorized as continuous-time or discrete-time process depending on whether the parameter set $T$ is continuous or not. Even though signals from machines are continuous-time in nature, they are usually sampled. In this thesis, we will focus primarily on sampled (i.e. discrete-time) r.p.es. Furthermore, we will assume the sampling procedure has been properly conducted to avoid the aliasing problem.
A process is strictly stationary if the joint probability distribution of the set of r.v.s 
\{X_1 = X(t_1), X_2 = X(t_2), \cdots X_n = X(t_n)\} is identical to that of a translated set 
\{X'_1 = X(t_1 + \tau), X'_2 = X(t_2 + \tau), \cdots X'_n = X(t_n + \tau)\}, for all \(n\) and \(\tau\).

A process is wide sense stationary (WSS)\(^2\) if it is stationary up to second order. Specifically, the first moment or mean \(E(X(t))\) and second moment \(E(X(t))(X(t + \tau))\) of the process are invariant with respect to time \(t\). For the rest of this dissertation, the term “stationary” will refer to “WSS”. The second moment of a r.p. is also called autocorrelation, expressed as \(R_X(t, t + \tau) = E(X(t))(X(t + \tau))\). If the process is stationary, then \(R_X(t, t + \tau)\) is usually written as \(R_X(\tau)\) where the time difference \(\tau\) is called lag.

If averages of both \(X(t)\) and \((X(t))(X(t+\tau))\) for any fixed lag \(\tau\) taken along a single time record of a stationary process converges in the mean square sense to \(E(X(t)) = \mu_X\) and \(R_X(\tau)\) as \(t \to \infty\), then the process is ergodic in the wide sense. Ergodicity is a very important property of a stationary process, since it implies that a single sample time signal of the process contains all possible statistical variation of the process. Note that ergodicity requires that the r.p. be stationary. But stationarity is not sufficient to assume ergodicity. In this work, we will assume all r.p. es, are both stationary and ergodic.

1.3 Spectral Analysis of Random Processes

Even though signals are acquired in the time domain, they are often transferred into the frequency domain.

Spectral analysis is used to analyze random signals in the frequency domain. The essence of the spectral estimation is to estimate how the total power of the process is distributed over a collection of frequency bands [9].

The Fourier transform is the basis for p.s.d. analysis. The continuous (CFT) and

---

\(^2\)also called weakly stationary or covariance stationary.
discrete-time Fourier transform (DTFT) is applied to continuous-time and discrete-time signals respectively. The basic CFT $X(f)$ of a signal $x(t)$ has the form:

$$X(f) = F(x(t)) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} \, dt \quad (1.1)$$

However, in the above transformation, there should be a scale factor accommodating different types of signals. There are two types of typical deterministic signals—periodic signals and aperiodic transient signals.

### 1.3.1 Fourier Series for a Deterministic Periodic Signal

Deterministic periodic signals are commonly observed in mechanical systems. The mean pressure signature of the pressure process acquired synchronously with the shaft speed of an axial compressor which runs at steady-state condition is a deterministic periodic signal [10], [11]. The sound pressure process of an axial fan running at constant fan speed in free field contains harmonics of blade passing frequency (BPF). This is another example of a deterministic periodic signal.

A continuous-time periodic signal $x(t)$ with period $T$ can be expanded in a Fourier series which is a linear combination of sine and cosine functions.

$$x(t) = \sum_{n=-\infty}^{\infty} C_n e^{j\omega_0 nt} \quad (1.2)$$

where $\omega_0 = \frac{2\pi}{T}$ is the fundamental angular frequency. The Fourier coefficients, $c_n$ in (1.2) are given by

$$C_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t)e^{-j\omega_0 nt} \, dt \quad (1.3)$$

For a periodic signal, Parseval’s Theorem states that

$$\int_{-T/2}^{T/2} x^2(t) \, dt = T \left( \sum_{n=0}^{\infty} c_n^2 \right) \quad (1.4)$$

(1.4) shows that the total energy of a periodic function is proportional to the period $T$. If the period approaches $\infty$, the energy over the time interval $(-\infty, \infty)$ will be infinite. Thus the scale factor of Fourier series associated with a periodic signal is $\frac{1}{T}$. 
1.3.2 Fourier Integral of a Transient Signal

Transient signals are also commonly found in mechanical systems. For example, when a turbine is brought up to speed, a main valve that opens allows hot nitrogen gas to be released from a supply tank, a transient pressure signal is then produced. In the case of a roller-bearing with a fault on the ridge, when the roller hits the fault, it will also produce a transient signal.

If a deterministic aperiodic signal \( x(t) \) is absolutely integrable, i.e., \( \int_{-\infty}^{\infty} |x(t)| \, dt < \infty \), then the signal has finite energy and is called a transient signal. Typically, the absolute integrability of \( x(t) \) also means that \( x(t) \) will decay to zero as \( t \to \pm \infty \). A transient signal could be considered as a periodic signal with a period \( T \to \infty \), such that a Fourier integral representation of the signal \( x(t) \) exists.

\[
X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} \, dt \tag{1.6}
\]

Where \( X(\omega) \) is called the Fourier transform of \( x(t) \). Equation (1.6) shows that for a deterministic transient signal, the scale factor is 1. The aperiodic transient signal has a distribution of finite energy over a continuous range of frequencies, while a periodic signal has a distribution of finite power over a discrete set of frequencies. The essential difference between the Fourier series and the Fourier transform is that the spectrum in the latter case is continuous, so that the synthesis of a transient signal from its spectrum is accomplished by means of integration instead of summation.

1.3.3 Spectral Analysis of Stationary Random Processes with Transient Autocorrelation Functions

Spectral analysis is almost always performed on WSS processes. A regular stationary r.p. is one that has a transient autocorrelation function. A broad band colored noise
process is a regular stationary r.p. For a regular stationary r.p., the Fourier’s transform of \( x(t) \) should be defined as:

\[
X(\omega) \equiv \lim_{T \to \infty} \frac{1}{\sqrt{T}} \int_{0}^{T} x(t) e^{-j\omega t} \, dt
\]  

(1.7)

To see this, we appeal to the Wiener-Khinchine Theorem:

\[
\lim_{T \to \infty} E \left[ \left| \frac{1}{\sqrt{T}} \int_{0}^{T} x(t) e^{-j\omega t} \, dt \right|^2 \right] = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega \tau} \, d\tau = S_x(j\omega)
\]  

(1.8)

Therefore the scale factor associated with a regular stationary process should be \( \frac{1}{\sqrt{T}} \). This scale factor is used in many p.s.d. analysis software packages such as Matlab. In fact, the left side of (1.8) is the basis for almost all p.s.d. estimation softwares. Specifically, \( T \) is fixed as the observation time window length and the expectation \( E() \) is replaced by the time average operation. Such a stationary process with transient autocorrelation function is called a regular WSS r.p. with continuous spectrum since it has a distribution of finite power over a continuous range of frequencies. That’s why the spectrum in (1.8) is called p.s.d. function.

1.3.4 Spectral Analysis of Stationary Processes with Periodic Autocorrelation Functions

There exist stationary processes which have periodic autocorrelation functions. A tone\(^3\) is one example. The sound pressure process of a free speed fan rotating at constant speed contains harmonics which are a series of tones on multiple times of the blade passing frequency (BPF). Define a tone \( x(t) \) as:

\[
x(t) = A \sin(\omega_0 t + \phi)
\]  

(1.9)

where \( A \) is a constant amplitude and \( \phi \) is the random initial phase uniformly distributed on \( (0, 2\pi) \). Then \( x(t) \) in (1.9) could be proved to be a WSS r.p.. The autocorrelation of

\(^3\)A tone is also called a harmonic process.
\( x(t) \) is a periodic function described as:

\[
R_x(\tau) = \frac{A^2}{2} \cos(\omega_0 \tau)
\]

Since the autocorrelation function \( R_x(\tau) \) does not decay to 0 as \( \tau \to \infty \), using the definition of a p.s.d. in (1.8) if this fact is ignored, the p.s.d. of \( x(t) \) will be:

\[
S_x(j\omega) = \int_{-\infty}^{\infty} R_x(\tau) e^{-j\omega \tau} d\tau
\]

\[
= \lim_{T \to \infty} \int_{-T/2}^{T/2} \frac{A^2}{2} \cos(\omega_0 \tau) e^{-j\omega \tau} d\tau
\]

\[
= \frac{A^2}{4} \delta(\omega \pm \omega_0)
\]

(1.11) shows the theoretical p.s.d. of a tone is a Dirac-\( \delta \) function, whose integral over the frequency range is finite, but the spectral density magnitude at the tone’s frequency \( \pm \omega_0 \) is \( \infty \).

If the spectral analysis is implemented with window length \( T \), then the p.s.d. of a tone will be on the order of \( T \). Thus the peak magnitude of a tone in the p.s.d. is meaningless since it will approach \( \infty \). It is the area under the Dirac-\( \delta \) function that indicates the power of the tone, \( \frac{A^2}{2} \).

A more appropriate approach is to introduce a scale factor \( 1/T \) in (1.11) such that a power spectrum instead of p.s.d. is obtained. The power spectrum of a tone will be a line spectrum with a magnitude of \( \frac{A^2}{4} \) on \( \pm \omega_0 \) and 0 elsewhere.

### 1.3.5 Mixed Random Processes with Both Discrete and Continuous Spectrum

As mentioned at the beginning of this chapter, most signals from rotating machineries contain a mixture of regular WSS r.p. with continuous spectrum and tones centered on integer times of fundamental frequency. The sound pressure process from a tractor

\footnote{\( \frac{A^2}{2} \) is the one-sided power of the tone and is the sum of the power of the tone on \( \pm \omega_0 \).}
engine cooling system studied in Chapter 5 is a typical example of a mixed random process.

As discussed above, a WSS r.p. with pure continuous spectrum should have a scale factor of \(1/\sqrt{T}\), while a WSS r.p. with discrete spectrum need a scale factor of \(1/T\). This causes a dilemma since neither scale factor will fit both types of r.p. es.

The p.s.d. of a mixed r.p. is an analog to a probability density function (pdf) of a mixed r.v. which is the sum of both discrete and continuous r.v.s. The continuous r.v. has continuous pdf while the discrete r.v. only has probability mass function (pmf) with mass probability at each discrete possible values of the r.v.. Neither pdf or pmf will be suitable to describe the probability distribution of the mixed r.v in that the pdf of a discrete r.v. will be a set of Dirac-\(\delta\) functions. It is the area under the pdf that means the probability. The height of the Dirac-\(\delta\) functions is meaningless since it goes to \(\infty\).

In the implementation of spectral analysis, we are only given the sample realization of the underlying r.p. from a mechanical system. The measurement is often a sampled version in discrete-time even though the original process may be a continuous-time one. Let’s express the length of the sample realization is \(N\). Usually, the sample realization is divided into non-overlapping segments of size \(n\). Different window functions and algorithms may be applied to each segments such that Different p.s.d. estimates may be obtained based on different window functions and algorithms. The p.s.d. estimates from all segments will then be averaged to decrease the variance of the p.s.d. estimator.

Among different p.s.d. estimator algorithms, there are two approaches– non-parametric and parametric one. Periodogram is a commonly used non-parametric p.s.d. estimator. And autoregressive (AR) spectrum is a parametric one.

Because the theoretical p.s.d. of a tone is a Dirac-\(\delta\) function, it can only be approximated in a sense such that it will go to \(\infty\) as the window length approaches \(\infty\). While the mean of the p.s.d. estimator of a regular WSS r.p. will be independent of the window size. The mean of each p.s.d. estimator of a mixed r.p. will exhibit different degrees of
variability depending on the estimators at the tone's frequency. While at non-tone’s frequency, the mean of p.s.d. estimators should be the same.

However this variability of p.s.d. estimators of different window size of a mixed r.p. is frequently ignored in the implementation of spectral analysis. The p.s.d. estimators of different window size is also called a family of p.s.d. estimators. It is this variability of a family of p.s.d. estimators of mixed r.p. that lead to the interest of Chapter 2.

1.4 Continuous-Time and Discrete-Time Signals

Another generic problem in signal estimation is the sampling problem. Most observable physical variables are continuous-time in nature, but have to be sampled in discrete time. Often, the process is sampled in uniform time interval $\Delta$. Sometimes, it is sampled at a uniform angle interval in rotating machinery, in that case, the random process will be considered as indexed by angle parameter. In view of the ubiquitous presence of digital computers for both data acquisition and analysis, the interest in the sampling aspect should be self-evident. We are particularly interested in the issue of selection of the sample rate.

The most common rationale of the choice of sampling rate is the famous sampling theorem. It states, that the minimum sampling frequency should be at least twice of the maximum bandwidth of the signal in order that the original signal can be fully reconstructed. This minimum sampling rate is also called Nyquist rate of the signal. Any sampling rate lower than the Nyquist rate would result in aliasing errors.

However, the availability of high speed computer and mass storage hardware made it readily performed to sample a process at a much higher rate relative to the dynamics of the process. A common false perception is that faster sampling is better. Although faster sampling may have more detailed visualization of the continuous-time process, sampling the process at an unnecessarily high rate has disadvantages. First of all,
it will result in enormous amounts of data, which will increase the computation and storage cost. Second, it will limit the sophistication of real time control algorithms. For example, in real-time feedback control, the computation time has to be no greater than the sampling interval $\Delta$. A fast sampling rate will thus limit the sophistication of the algorithm by limiting the computation time. Third, it will degenerate the discrete-time stochastic models for WSS r.p.es. This would in turn cause problems such as stability, round-off error, convergence problems in the discrete-time algorithms used in estimation and control since they will become ill-conditioned when applied to processes sampled at very fast rates. The ill-conditioning of the stochastic model and the algorithm mainly comes from the ill-conditioning of the covariance matrix $R_n$ of the discrete-time process sampled at a very rapid rate. As could be shown, the covariance of all the finite lags will become a constant, resulting in $R_n$ becoming singular when the sampling interval approaches zero. This in turn renders the autoregressive parameters to converge to the $(-1)^j \binom{n}{j}$, where $n$ is the highest order of differential equation of the underlying continuous-time process. As a result, the discrete-time process becomes independent of the underlying continuous-time process as $\Delta \to 0$.

On the other hand, in the state space model of using extended Kalman filter (EKF) in signal tracking, the additive measurement noise will become more and more correlated at a very rapid sampling rate. The correlated noise will violate the assumption of white noise in the EKF. All this can reduce the potential of tracking the process itself, or of tracking slowly time varying parameters associated with a model for the underlying continuous-time process.

There are some literature trying to attenuate problems associated with fast sampling. In Kalman-Bucy filtering, the discrete-time algebraic Riccati equation (DARE) does not converge smoothly to a meaningful limiting (continuous-time) equation as the sampling rate increases. Similarly, in discrete-time pole placement problems, the condition num-
ber of the Sylvester matrix increases substantially with the sampling rate. These two particular problems have been overcome in [12], [13] respectively, by transforming the conventional sampled-data problems, which use a shift-operator representation of data dynamics, into problems based on divided-difference operator. Motivated by this problem, Vijayan et al. [14] developed a model based on an incremental difference operator rather than the shift operator. A detailed review of algorithms based on this divided-difference operator in the regime of fast sampling is provided in [15]. However, only numerical issues are discussed, the statistical issues related with fast sampling is not yet addressed.

Statistical performance of parameter estimation related to the sampling rate is also known in literature. For instance, it is known that, in relation to digital systems, discrete parameter estimator bias can increase significantly when such frequencies are very close to zero, relative to the analysis bandwidth (BW) (Stoica & Söderström [16], Kay [17], Marple [18]). In the problem of estimating the frequency of a sinusoid corrupted by white Gaussian noise, the Cramér-Rao lower bound derived in [19] for unbiased maximum likelihood frequency estimator reveals that estimating very low and high frequency is more difficult than other frequencies. In the same problem, Lau and Sherman [20] noticed that both the bias and variance of the constrained AR(2)-based frequency estimator are minimized when the frequency of the sinusoid is at the center of the linear-frequency analysis BW. In fact, the bias is zero.

This statistical problem of estimating a BW parameter will be addressed in detail in Chapter 3. The particular interest of Chapter 3 will be limited to the parameter estimation problem of a First-order Gauss-Markov (GM) process. The statistical influence of the sampling interval $\Delta$ and observation duration $T$ of a GM process on the BW parameter estimator is studied. The case of allowing the sampling interval $\Delta$ to be selectable while the duration time $T$ of the observed process is fixed is of particular interest. This is different from a more commonly considered case where the observation
time duration is allowed to approach infinity with the sample rate fixed.

There are also numerous asymptotic results for this more common case, such as [20], [21], [22], [23], [24], and so on. These results rely heavily on the condition that the estimator is consistent; in other words, that its variance approaches zero as the number of samples approaches infinity. The former case receives little attention in literature though it reflects common real application situation. For example, the observation duration of the processes could be constrained on account of the cost and time of the measurement. In addition, in the case of slowly time varying process, the duration of the process used to estimate the locally time invariant parameter is limited by its time varying property.

The prediction problem of a time invariant GM process, another important problem in stochastic processes, is then studied to show the value of parameter estimation results. The prediction performance of the m-step prediction that uses an assumed deterministic BW parameter $\beta$ and that uses an estimated one from previous process of length $T$ is investigated. The relationship of the prediction performance to the variables—$\beta$, sampling interval $\Delta$, prediction step $m$ and window length $T$ are explicitly expressed, which can be used to design a prediction filter such as the one in an active noise cancellation system. In particular, we show that prediction performance is highly robust with respect to estimation accuracy of $\beta$. This is important, because it allows one to use a surprisingly small observation time $T$ and still achieve nearly optimal performance with perfect knowledge of $\beta$.

1.5 Organization of the Dissertation

Following this introduction chapter will be Chapter 2 which discusses the idea of using the variability related to families of spectral estimators for mixed random processes. Chapter 3 studies the choice of the sampling rate and sampling duration in the parameter estimation of a first order GM process. To show the value of the results from Chapter 2
and 3, in Chapter 4, these results will be applied to the problem of characterizing noise from an engine cooling system, in particular to understand stochastic structure of the sound pressure process and the effect of the engine block on the sound pressure process at chosen position. Finally Chapter 5 will summarize this dissertation and discuss the direction of possible future research.
CHAPTER 2 USING VARIABILITY RELATED TO FAMILIES OF SPECTRAL ESTIMATORS FOR MIXED RANDOM PROCESSES

A paper published in the Journal of the Transaction of ASME in Dynamics, Systems, Measurement, and Control.\textsuperscript{1}

Li Wen\textsuperscript{2}, Changxue Wang\textsuperscript{3} and Peter Sherman\textsuperscript{4}

2.1 Abstract

Traditionally, characterization of spectral information for wide sense stationary processes has been addressed by identifying a single best spectral estimator from a given family. If one were to observe significant variability in neighboring spectral estimators then the level of confidence in the chosen estimator would naturally be lessened. Such variability naturally occurs in the case of a mixed random process, since the influence of the point spectrum in a spectral density characterization arises in the form of approximations of Dirac-$\delta$ functions. In this work we investigate the nature of the variability of the point spectrum related to three families of spectral estimators: Fourier

\textsuperscript{1}Reprinted with permission of Trans. of ASME, in Dynamics, Systems, Measurement, and Control, Vol. 123, No. 4, December 2001, 572-584,[25]. There is minor revision.
\textsuperscript{2}Graduate Student and primary researcher and author, Department of Aerospace Engineering, Iowa State University.
\textsuperscript{3}Graduate student, secondary author for algorithm and paper review, Department of Aerospace Engineering, Iowa State University.
\textsuperscript{4}Major professor, Department of Statistics and Department of Aerospace Engineering, Iowa State University.
transform of the truncated unbiased correlation estimator, the truncated periodogram, and the autoregressive estimator. We show that tones are a significant source of bias and variability. This is done in the context of Dirichlet and Fejer kernels, and with respect to order rates. We offer some expressions for estimating statistical and arithmetic variability. Finally, we include an example concerning helicopter vibration. These results are especially pertinent to mechanical systems settings wherein harmonic content is prevalent.

2.2 Introduction

Spectral estimation has played a major role in a wide variety of theoretical and application areas of science and engineering since the advent of modern computing and the FFT in the mid-1960's. Traditionally, the idea has been to find the "best" spectral estimator. Often the desire was to balance resolution and variability. This is the idea behind both averaging of periodograms and autoregressive (AR) order selection methods. Perhaps because of the limitations and expense of computing resources in the early years it was natural to rely on such selection methods. But it is also natural to question this entire approach if the variability within the family of spectra under consideration is significant. It may well be that neighboring spectra exhibit measurable variability at certain frequencies, while not at others. In fact, this is exactly the case at and near frequencies corresponding to the point spectrum, when the random process includes a deterministic as well as regular component. Sinusoids are the most common source of point spectrum. Given an infinite number of correlation lags, they would appear as Dirac delta functions. But if the spectral family is indexed by the number of correlation lags used, as the case in periodogram, AR and other methods, then the influence of the point spectrum will be seen as peaks whose values are, in and of themselves meaningless, and as spectral leakage.
With the advances in computing resources it is now far easier to both compute and analyze a large family of spectral estimators than it was even 15 years ago. Even so, this family-based approach to statistically reliable spectral estimation has received very little attention; in spite of the fact that it has been suggested for over 15-20 years now. For example, in [17] and [26], the use of periodograms with successively larger windows is proposed. The idea is that if the spectral information remains insensitive to the window size changes then one can have greater confidence in it. This work is intended to contribute a better understanding of the variability of spectral estimator families, with particular attention to the cases that caused by the presence of point spectrum. Specifically, we address three families: the truncated Fourier transform, the averaged periodogram, and the AR spectra. This variability will be addressed in two stages. In Section 2.4 we will investigate the variability of these families when the autocorrelation information is exact. This will reflect the order-dependent theoretical spectral variability. It is also valuable in its own right, since there are many applications where the amount of available data far exceeds the range of reliable correlation lags that one might consider. In Section 2.5, we address the statistical variability associated with lagged-product estimates of the correlation information. The value of the sample mean and corresponding variance is one way of using a family of spectra, as suggested in [17] and [26]. The value of this information is the subject of Section 2.6. In Section 2.7, we apply the results of the previous sections to vibration data from a helicopter drive train. Our summary and conclusions are given in Section 2.8. We now proceed to motivate our investigation and describe the types of processes we are concerned with.
2.3 The Structure of P.s.d. Estimator for Mixed Spectrum Processes

We consider wide sense stationary (WSS) random processes of the form

\[ x(t) = s(t) + \epsilon(t) \]  \hspace{1cm} (2.1)

where the signal, \( s(t) \), is composed of sinusoids with deterministic amplitudes and frequencies \( \{A_k, \omega_k\} \), and with independent phases each distributed uniformly over \([0, 2\pi)\). The noise, \( \epsilon(t) \), is regular, and is assumed to have a continuous power spectral density (p.s.d.), \( S_c(\omega) \). The p.s.d. of (2.1) is given by

\[ S_x(\omega) = \sum_k \frac{A_k^2}{4} \delta(\omega \pm \omega_k) + S_c(\omega) \]  \hspace{1cm} (2.2)

where \( \delta(\omega) \) is the well known Dirac-\( \delta \) function. Consequently, (2.2) is only defined in the sense of a generalized function, in that only its integral, the cumulative p.s.d., is well defined with jumps at the signal frequencies. The model (2.1) is fundamental to mechanical systems, such as rotating machinery. Typically, processes associated with such systems include harmonics as well as highly colored spectral components.

Let \( \{R_x(\tau)\}_{\tau=-n+1}^{n-1} \) be the theoretical correlation information through the \( nth \) lag. Then the theoretical Fourier transform (FT) spectral estimator is given by

\[ S_{FT(n)}(\omega) = \sum_{\tau=-n+1}^{n-1} R_x(\tau)e^{-i\omega\tau} \]  \hspace{1cm} (2.3)

We remark that in (2.3) and throughout the remainder of this work it is assumed that the sampling interval is 1 second, so that all frequencies are in the interval \([0, \pi)\). It is commonly assumed that (2.3) will converge to (2.2) when the number of lags, \( n \), approaches infinity. In the absence of tones this will generally be true. But when tones are present it is not true, as will be shown. One solution to this problem turns out to be to use an average of \( FT(n) \) spectra for a range of values of \( n \). In addition to solving
this problem, the use of such a family offers information that the use of any single p.s.d. cannot offer, namely spectral variability with respect to the number of lags. This is in addition to the statistical variability associated with the use of estimated correlation lags. The theoretical truncated periodogram, $PER(n)$, is one such spectrum. Another common p.s.d. estimator which uses the same $n$ correlation lags is the $AR(n)$ spectrum. The specific form of the theoretical AR($n$) spectrum is well known (e.g. [18]), and so it will not be repeated here. In contrast to the FT($n$) spectrum, (2.3), the theoretical PER($n$) and AR($n$) spectra converge at almost every frequency (except at the point spectrum frequencies) to the continuous spectrum as $n \rightarrow \infty$. The FT($n$), PER($n$) and AR($n$) theoretical spectra all exhibit order-dependent variability due to the presence of tones, and become unbounded at the tone frequencies as $n \rightarrow \infty$. The use of estimated correlation information introduces statistical variability, in addition to the arithmetic variability that we will investigate in the next section. Before doing so, however, we offer the following example to provide more motivation for our investigation of the utility of a family of spectral estimators.

**Example 1.** In this example we consider a process (2.1) consisting of a single sinusoid, plus a regular component. The theoretical p.s.d. is given in figure 2.1. It includes the $\delta$-function associated with the tone. The peak in the continuous spectrum was selected to simulate a strong system resonance, while the dip corresponds to an anti-resonance. This structure is commonplace in mechanical systems settings.

Assuming that a sufficiently large number of measurements is available (as the case with rotating machinery operating at constant speed) allows us to justify the use of theoretical correlation information. Figure 2.1 includes the $\pm 2\sigma$ arithmetic variability (dashed lines) of the family of theoretical FT($n$) spectra for $n = 32, 33, \cdots, 1024$. While not immediately obvious from figure 2.1, this variability reflects the fact that all of these spectra accurately capture the spectral resonance region, while none of them capture either the anti-resonance or the tone. It is well known that the spectral leakage associated
with the tone is the source of local variability. But here it is also responsible for non-local bias and variability in regions where the p.s.d. magnitude is not significant. In the region near the tone both bias and variability are meaningless, since in the p.s.d. domain a tone is a \( \delta \) function. The practical implications of this are that the estimated amplitude will converge to infinity as \( n \to \infty \), and, consequently, so will the variability of the family of estimators.

A major point of the following sections is to examine the above behavior in detail. Hopefully, this example has hinted at the potential value of using a family of spectral estimators, as opposed to a single "best" estimator, as is traditionally done. By observing the behavior of variability over increasingly larger ranges of \( n \) (termed window closing in [17] and [26], it is possible to gain greater confidence of the spectral structure. For example, in the resonance region there is very little variability, so that one can presume that this region is well characterized without concern for any order selection rules. In
the region of the tone the variability range increases, suggesting that this region is not appropriate for characterization by a FT(n) spectrum for any value of n. This suggestion requires clarification. In many situations, such as developing noise and vibration specifications for mechanical systems, the window size, n, is required to be a specific value. In such situations where everyone uses the same window size, type, number of averages, etc. the FT(n), PER(n) and AR(n) p.s.d. estimators can provide proper spectral distribution information over frequency intervals. But just as often, if not more so, the value for n is not fixed. As n grows so does the peak of any tone associated with a p.s.d.. This behavior does not appear to have bothered many people, since it has been demonstrated routinely in most of the high resolution spectral research conducted over the past 35 years, in the context of the two-sinusoid plus white noise setting (e.g. [26]). But in the context of using a family of spectra, as suggested in [17] and [26], one would conclude that any region involving tones should be viewed as unreliable. In the realm of mechanical systems, and in particular, rotating machinery, this would adversely affect spectral analysis, as a whole. This could lead one to apply spectral decomposition tools such as [27] to eliminate this problem. But we will not address this approach in this work, since we are concerned here with the common procedure of analysis of the mixed spectrum, as it is. The FT(n) spectrum is perhaps not as popular as the PER(n) spectrum. Traditional reasons for this range from the fact that it can lead to negative p.s.d. estimates, to the fact that the side-lobe behavior associated with the rectangular windowing operation results in excessive local spectral smearing. The above example suggests that the FT(n) family is not well-suited for accurate characterization of anti-resonance structure when tones are present anywhere in the spectrum.

To further motivate the following sections we offer the performance of a family of AR(n) spectra. Using a minimum number of orders $n = 5, 6, \cdots, 10$ in figure 2.2 produces less variability in the anti-resonance region than the FT(n) family does. Furthermore, by using orders $n = 20, \cdots, 100$ (far fewer and lower than the FT(n)) figure 2.3
indicates not only lack of bias, but minimal variability everywhere except at the tone region. Also, the size of that region is lessening with the use of higher lags (in contrast to the FT(n) spectra). Thus, one can conclude that for accurate anti-resonance characterization the AR(n) family is more appropriate than the FT(n) family. In the region near the tone the variability is much more localized than that of the FT(n) family. Even so, since the AR(n) family is a family of p.s.d. estimators, at the tone frequency the amplitude will converge to infinity as \( n \to \infty \), as was the case with the FT(n) family. Our analysis will reveal, however, that the rate of convergence is markedly faster than that of the FT(n) family. This suggests yet another use of spectral families; namely to use their convergence properties to identify tone components. We will discuss this point in greater detail.

![Figure 2.2](image)

**Figure 2.2** (a) \( 2 - \sigma \) curves for an average of 5 AR(n) spectra

The above example utilized theoretical correlation information. Hence, the bias and variability may be said to be arithmetic in nature, as opposed to statistical variability related to using estimates of the correlation lags. We now proceed to a more detailed
discussion of the arithmetic variability associated with the theoretical FT(n), PER(n) and AR(n) spectra, which utilize perfect correlation information. The justification for this is in the fact that often in mechanical system analysis one has access to extremely large amounts of data, in relation to the number, n, of estimated correlation lags used for analysis. In such cases it may be reasonable to presume that the correlation information is highly reliable. We will, at times during this discussion, restrict our attention to the situation of a single tone plus white noise. The reason for this is twofold. First, the presence of a tone can have a significant effect on spectral variability. Second, by use of band pass filtering it is sometimes possible to restrict the region of interest such that within that region the noise spectrum is relatively flat.
2.4 Arithmetic Variability of Theoretical FT(n), PER(n) and AR(n) Spectral Families

In this section we investigate the arithmetic variability associated with the use of theoretical correlation information for three spectral families. Two of these, namely the FT(n) and AR(n), have been discussed above. The third is the family of truncated periodograms, which we denote as the PER(n) family. This is by far the most popular family of spectral estimators in use in practically all areas of science and engineering. We will obtain quantitative expressions for both the bias and variance. These will entail order-dependent terms, which will provide growth rate information in relation to tones.

We restrict our attention here to the case of the model (2.1) with a single sinusoid:

\[ x(t) = A\sin(\omega_0 t + \theta) + \epsilon(t) \]  

(2.4)

To be sure, the two-tone problem is an important and common one. But such a setting would significantly complicate the analysis, possibly to the point of distraction from our main goal, which is to gain a better understanding of the variability of a spectral family in relation to a mixed spectrum setting. So little attention has been paid to this problem that we believe it is appropriate here to restrict our investigation to the more simple setting (2.4), in order to achieve our goal.

2.4.1 Variability of the Theoretical FT(n) Spectra

For the model (2.4), equation (2.3) takes the well known form (e.g. [17]):

\[ S_{\text{FT}(n)}(\omega) = \frac{A^2}{4} D_n(\omega + \omega_0) + \frac{A^2}{4} D_n(\omega - \omega_0) + S_\epsilon \otimes D_n(\omega) \]  

(2.5)

where \( D_n(\omega) = \frac{\sin((2n-1)\omega/2)}{\sin(\omega/2)} \) is the Dirichlet kernel [18] associated with the \( 2n - 1 \) point rectangular window. Now,

\[ \lim_{n \to \infty} S_{\text{FT}(n)}(\omega) \neq \frac{A^2}{4} \delta(\omega \pm \omega_0) + S_\epsilon(\omega) \]  

(2.6)
Notice that the middle relation in (2.6) is not an equality, but rather a non-equality. This reflects the fact that the family of Dirichlet functions does not converge (anywhere) as \( n \to \infty \). This is exemplified in figure 2.4. There, we see that because the Dirichlet function is proportional to \( n \), even though the major side lobes move closer to the tone frequency for increasing \( n \), they are also increasing in size. Consequently, we see that at frequencies sufficiently far from this frequency the Dirichlet peak values do not decrease with increasing \( n \). This fact, while certainly not new, seems to have been ignored in the vast majority of books and papers on the subject where the spectral density is defined via the limiting Fourier transform of the correlation function, using a finite number of lags, as in figure 2.4. Even though, relatively speaking, the energy at the origin will overwhelm that in other regions, so that it may appear that the sequence is converging to a Dirac-\( \delta \) function, it is not. In fact, at any fixed frequency other than that of the tone, the sequence of functions will neither converge nor diverge as \( n \to \infty \). Rather it will oscillate with bounded variation. This is of sufficient importance when assessing variability, and is so often ignored in systems and signals publications, that we now present a formalized statement of this for the family of rectangular windows.

**Result 1.** The Fourier transform pair: \( f(t) = 1 \iff F(\omega) = \delta(\omega) \) is not necessarily true. It depends on the selection of functions of which \( f(t) \) is the pointwise limit. In particular, let \( w_{(n)}(t) = 1 \) for \(- (n-1) \leq t \leq n - 1\), and let it equal zero otherwise. Let \( W_{(n)}(\omega) = \sum_{t=-n+1}^{n-1} w(t)e^{j\omega t} \) be the Fourier Transform of \( w_{(n)}(t) \). Then \( \lim_{n \to \infty} W_n(\omega) = \lim_{n \to \infty} D_n(\omega) \) exists nowhere. Specifically, \( W_{(n)}(0) \) is \( O(n) \), while for \( |\omega| >> 0 \), \( W_{(n)}(\omega) \) oscillates (as a function of \( n \)) between \( \pm 1/\sin(\omega/2) \).

For the more mathematically inclined reader, we remark that a rigorous definition of the Dirac-\( \delta \) function as the limit of a family of functions requires the use of a family which is suitably well-behaved ([28], [29]). The discontinuities at the ends of the rectangular function family are well known to be not well-behaved, to the extent that they yield what is commonly known as the Gibb's phenomenon. The fact is that one can never
collect an infinite amount of information, be it data or correlation information. And so the form of truncation becomes important if one desires to correctly infer the outcome, were all the information available.

Recall, that this work is concerned with families of spectra. Because of the undesirable properties associated with the collection of FT(n) spectra which utilize the expected value of the unbiased correlation estimator, it may be of interest to investigate whether averaging them can offer any advantage. For a collection of theoretical spectra \( \{S_{FT(n)}(\omega)\}_{n=n_0}^{n_1} \), we now formally define the arithmetic average and variability of the collection of FT(n) spectra over the indices \( n_0 \) to \( n_1 \), respectively, as

\[
S_{FT(n_0,n_1)}(\omega) = \frac{1}{n_1 - n_0 + 1} \sum_{n=n_0}^{n_1} S_{FT(n)}(\omega) \tag{2.7}
\]

\[
\gamma_{FT(n_0,n_1)}(\omega) = \frac{1}{n_1 - n_0 + 1} \sum_{n=n_0}^{n_1} [S_{FT(n)}(\omega) - S_{FT(n_0,n_1)}(\omega)]^2 \tag{2.8}
\]

It was just noted that the FT(n) spectrum has the undesirable property that the leak-
age influence associated with a tone will persist independent of \( n \). This is illustrated in figure 2.5, which is related to Example 1. We see that not only is this family of averaged FT(n) spectra converging at the anti-resonance location, it also yields progressively localized leakage in the vicinity of the tone. Thus, as suggested previously by [17], [26], [27] and others, the use of a family can provide an advantage over the use of any single spectra. One could proceed to use this averaging procedure, as opposed to the use of a single FT(n) spectrum, even though it would involve more computation. The following result shows that this averaging procedure may be implemented without the need to perform the averaging computation.

\[
S_{FT(n)}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(\nu) K_n(\omega - \nu) d\nu = S_{PER(n)}(\omega)
\]  

(2.9)

Figure 2.5  Average of FT(n) theoretical spectra for the range 1 \( \rightarrow n_{\text{max}} \) for \( n_{\text{max}} = 100, 200 \) and 500.

**Result 2.** ([30] p.16). For \( n_0 = 1 \) and \( n_1 = n \) equation (2.7) may be expressed as

Here, \( K_n(\omega) = \frac{\sin(n(\omega/2))}{n \sin(\omega/2)} \) is known as Fejer's kernel [18]. Also, \( S_{PER(n)}(\omega) \) is the expected value of the popular periodogram spectral estimator, and uses the biased lagged-
product correlation estimates, as opposed to the unbiased ones used in equation \((2.3)\). To be precise, equation \((2.9)\) is not the expected value of the periodogram unless the order, \(n\), is identical to the data record size, \(N\). More generally, for \(n\) less than \(N\) it is referred to as the truncated periodogram. For convenience we will not make such a distinction unless it is necessary. While perhaps not evident from equation \((2.7)\), this kernel is exactly the average of the collection of Dirichlet kernels from 1 to \(n\). The fact that the rightmost equality in equation \((2.9)\) corresponds to the Fourier transform of the theoretical correlation function that has been windowed using a triangular, or Bartlett window is well known. What the leftmost equality shows is that in equation \((2.7)\), when summation indices range from 1 to \(n\), we obtain the expected value of the periodogram estimator. This observation is a long known result, but one that is seldom noted in most books on signals and systems. Traditionally, the Bartlett window is used to reduce the intensity of the spectral side lobes associated with leakage. Our observation suggests that it should also be used to ensure proper behavior of the family of spectra in the case of a mixed process such as equation \((2.4)\); namely, convergence as \(n \to \infty\).

Both the Fejer and Dirichlet kernels grow at a rate \(O(n)\) at a tone frequency. But in contrast to the Dirichlet family, the Fejer family converges to a Dirac-\(\delta\) function; that is, the leakage away from the tone frequency goes to zero as \(n \to \infty\). This is illustrated in figure 2.6. In the absence of any tones, if the noise p.s.d. is continuous, then equation \((2.7)\), which is a Cesaro mean, will converge uniformly (in \(w\)) to the true p.s.d. as \(n_1 \to \infty\). ([30] p.16). Since the (generalized) p.s.d. limit in equation \((2.9)\) is absolutely integrable, it follows [30] that even though convergence will fail at the tone frequency, it will take place elsewhere.

When \(n\) is sufficiently large, the expression for the arithmetic variance, equation \((2.8)\),
of the FT(n) family of theoretical spectra can be approximated as the following:

\[
\gamma_{FT}(\omega) \approx \frac{A^4}{16} \cdot \frac{1}{\sin \left( \frac{\omega - \omega_0}{2} \right)^2} \quad \text{for} \quad |\omega - \omega_0| \gg 0 \quad (2.10)
\]

\[
\gamma_{FT}(\omega_0) \approx \frac{A^4}{16} \cdot O(n^2) \quad \text{for} \quad \omega = \omega_0 \quad (2.11)
\]

As mentioned above, at a tone frequency the variability of any p.s.d. family will grow with increasing order, \(n\), as is the case in equation (2.11). The fact that the variability of the FT(n) family persists at frequencies removed from the tone, as given in equation (2.10), make that family undesirable for use in the case of a mixed process.

We now proceed to a discussion of the variability of the more desirable and commonly used PER(n) family.
2.4.2 Variability of the Theoretical PER(n) Spectra

The theoretical PER(n) spectrum, which is exactly the expected value of the truncated periodogram, is given by

\[ S_{\text{PER}(n)}(\omega) = \sum_{\tau=-\frac{n-1}{2}}^{\frac{n-1}{2}} R_\epsilon(\tau)B_n(\tau)e^{-i\omega\tau} \] (2.12)

where \( B_n(\omega) = \frac{(n-|\tau|)}{n} \) is the triangular, or Bartlett window, whose Fourier transform is the \( n \)-th Fejer kernel, \( K_n(\omega) \). In the manner of equation (2.5), it may be expressed as

\[ S_{\text{PER}(n)}(\omega) = \frac{A^2}{4} K_n(\omega + \omega_0) + \frac{A^2}{4} K_n(\omega - \omega_0) + S_\epsilon \otimes K_n(\omega) \] (2.13)

Notice that if \( \omega_0 = \frac{2\pi k_0}{n} \) for any integer \( k_0 \), then the first two terms in equation (2.13) vanish at all discrete computation frequencies other than \( \omega_0 \); that is, the tone spectral leakage will not distort the continuous spectrum information at those frequencies. Unfortunately, in practice one seldom has control over such precise placement of the tones in relation to the sampling frequency.

An explicit expression for the arithmetic variability of the PER(n) family can be obtained in the same manner as was done for the FT(n) family. We begin by noting that since the Fejer kernels are \( O(n) \) at the tone frequency, then so is the arithmetic average of the theoretical PER(n) spectra. And since, unlike the Dirichlet family, the Fejer family is \( O(1/n) \) at frequencies removed from the tone frequency, then the arithmetic variability of the PER(n) family is \( O(1/n) \). What follows is a more quantitative description of this behavior.

For a collection of theoretical spectra \( \{S_{\text{PER}(k)}(\omega)\}_{k=1}^{n} \), we define the arithmetic average and variability of the collection of PER(n) spectra over the indices 1 to \( n \) in the usual way. The average over the theoretical periodogram family is simply the average
over the Fejer kernels, which for large $n$ is approximately

$$
\psi_{(n)}(\omega) = \frac{1}{n} \sum_{k=1}^{n} K_{(n)}(\omega)
$$

$$
\cong \begin{cases} 
\frac{1}{1-\cos(\omega)} \left[ \frac{\ln n}{n} + \frac{1}{2n^2} - \frac{1}{2n} \ln \frac{1}{2(1-\cos \omega)} \right] & \text{for } \omega \neq 0 \\
\frac{(n+1)}{2} & \text{for } \omega = 0 
\end{cases}
$$

for $\omega \in [0, 2\pi)$. Therefore the averaged PER(n) spectra is

$$
S_{\overline{PER(n)}}(\omega) = \frac{A^2}{4} \psi_n(\omega + \omega_0) + \frac{A^2}{4} \psi_n(\omega - \omega_0) + S_e \otimes \psi_n(\omega)
$$

From equation (2.14), as expected, the $\overline{PER(n)}$ spectrum is growing at a rate of $O(n)$ at the $\omega = \pm \omega_0$. At frequencies away from the sinusoid, the $\overline{PER(n)}$ spectrum will converge to true spectrum at a rate of $O(\ln n/n)$. It follows that the kernel $\psi_{(n)}(\omega)$ will converge to the Dirac-\(\delta\) function.

For the case of white noise, since the last term in equation (2.15) is simply $\sigma_e^2$, equation (2.15) is controlled entirely by equation (2.14). The 3-dB bandwidth of the average of the first $n$ Fejer kernels is two to three times greater than that of a single PER(n) spectrum. Figure 2.7 shows the comparison of the $n^{th}$ and the average of the the first $n$ Fejer’s kernels for values of $n = 100$ and 1000. In contrast to the $n^{th}$ Fejer’s kernel, the average of Fejer’s kernels does not have side lobes. At frequencies far away from $\omega = 0$ the average kernel converges to zero as $n \to \infty$. 
For a process consisting of white noise plus a sinusoid, it is also straightforward to show that for large $n$ the arithmetic variance of the collection $\{S_{PER}(\omega)\}_{k=1}^{n}$ is approximately:

$$Var_{PER(n)}(\omega) = \frac{A^4}{16} \left[ \gamma_{K(n)}(\omega - \omega_0) + \gamma_{K(n)}(\omega + \omega_0) \right]$$

(2.16)
where

\[
\gamma_{K(n)}(\omega) = \frac{1}{n} \sum_{k=1}^{n} (K_k(\omega) - \psi_n(\omega))^2 \\
\geq \begin{cases} 
\frac{1}{n(1-\cos\omega)^2} \times \left[ \frac{n^2}{24} + \frac{n\omega}{2} - \frac{\omega^2}{8} - \frac{1}{n} \left( \ln n + \frac{1}{2n} - \frac{1}{2} \ln \frac{1}{2(1-\cos\omega)} \right)^2 \right] & \text{for } \omega \neq 0 \\
\frac{n^2 - 1}{12} & \text{for } \omega = 0 
\end{cases}
\] (2.17)

In particular, at the tone frequency the square root of equation (2.16) becomes

\[
\sigma_{\text{PER}(n)}(\omega_0) = \frac{A^2}{4} \times \frac{n}{3.464} 
\] (2.18)

Thus, at the tone frequency the standard deviation of Fejer kernel family increases by 3 dB as \( n \) doubles (at rate \( O(n) \)), while at frequencies removed from the tone it decreases at a rate of 1.5 dB per doubling (at rate \( O(1/n) \)).

**2.4.3 Variability of the Theoretical AR(n) Spectra**

We now turn to the arithmetic mean and variability of the AR(n) family of spectra for a mixed process. These two quantities are defined exactly as they were in equation (2.7) and (2.8) for the FT(n) family. Because the AR(n) spectrum is based on prediction of the correlation lags of orders greater than \( n \), we do not have the situation where a kernel function (which is independent of the noise spectral structure) may be analyzed. For this reason we will restrict our attention in this section to processes of the form equation (2.4), where the noise is white. In this case, it can be shown that the theoretical AR(n) spectrum for single (complex) sinusoid plus white noise is given by

\[
S_{\text{AR}(n)}(\omega) = \sigma_z^2 / \left| 1 - \frac{\rho}{1 + \rho n} D_n(\omega - \omega_0) \right|^2 
\] (2.19)

Here \( \rho \approx (A^2/4)/\sigma_z^2 \). From equation (2.19) it follows trivially that for \( \rho n \gg 1 \):

\[
S_{\text{AR}(n)}(\omega_0) \approx \sigma_z^2 |1 + \rho n|^2 \approx \frac{(A^2/4)^2 n^2}{\sigma_z^2} 
\] (2.20)
and
\begin{equation}
S_{AR(n)}(\omega) \approx \frac{\sigma^2_e}{|1-(1/n)D_n(\omega-\omega_0)|^2} \quad |\omega-\omega_0| \gg 0 \tag{2.21}
\end{equation}

Equation (2.20) states that at the tone frequency the theoretical AR(n) spectrum is proportional to \(n^2\), and to the local SNR, defined as \((A^2/4)/(\sigma_e^2/n)\) for \(\rho n \gg 1\). At the tone frequency the AR(n) spectrum is \(O(n^2)\). Equation (2.21) states that for \(\rho n \gg 1\), at frequencies sufficiently far from the tone frequency the AR(n) spectrum will be close to the noise spectrum, and will, in fact, converge to it as \(n \to \infty\). This is due to the fact that the Dirichlet kernel is scaled by a factor of \(1/n\), so that it converges to one at the tone frequency and to zero elsewhere. It is possible to gain more insight into the rate of convergence of equation (2.21) by expressing the difference between it and the limit \(\sigma^2_e\):

\begin{equation}
\sigma^2_e - S_{AR(n)}(\omega) = \sigma^2_e \left( \frac{(1/n)K_n(\omega-\omega_0) + (2/n)D_n(\omega-\omega_0)}{1 + (1/n)K_n(\omega-\omega_0) + (2/n)D_n(\omega-\omega_0)} \right) \tag{2.22}
\end{equation}

We know that at frequencies removed from the tone frequency the Fejer kernel converges to zero at rate \(O(1/n)\). Thus, the error (2.22) is dominated by the Dirichlet term in the numerator. Since the error is \(O(1/n)\), it follows that both the arithmetic average and standard deviation of the collection \({\text{AR}}(k)\}_{k=1}^n\) are \(O(1/n)\).

To evaluate the utility of these results for the case of colored noise, we consider figure 2.8 corresponding to Example 1. They include the arithmetic mean \(s_{AR(n)}(\omega)\) and standard deviation \(\sigma_{AR(n)}(\omega)\) of AR spectral family for order ranges 2 to \(n = 20, 40, 80\) and 160. Figure 2.8(top) shows that the average converges to the continuous spectrum at all frequencies removed from the tone frequency, while at the tone frequency the rate of growth is \(O(n^2)\). Figure 2.8(bottom) shows that at frequencies removed from the tone frequency the variance decreases at a rate \(O(1/n)\), while at the tone frequency it increases at a rate of \(O(n^4)\).
To summarize this section, we have provided order-related rates of behavior for the FT(n), PER(n) and AR(n) theoretical spectra, as well as rates related to their arithmetic means and standard deviations. This was in the context of mixed spectrum processes of
the form (2.4). At the tone frequency it was noted that the FT(n) and PER(n) spectra grow at a rate \( O(n) \), while the AR(n) grows at a rate \( O(n^2) \). At frequencies removed from the tone, the variability of the FT(n) family is \( O(1) \); that is, it never converges to the true spectrum. The variability of the PER(n) and AR(n) spectra are \( O(1/n) \). We now proceed to address the statistical variability of the families.

2.5 Variability of the Estimated FT(n), PER(n) and AR(n) Spectra

There is a wealth of literature on the statistical properties of the marginal FT(n) and AR(n) spectral estimators in the case of regular random processes (see e.g. [21] for references). One can argue that in view of the orthogonal property of the frequency decomposition those results should hold in all but the local regions associated with sinusoids. However, we saw in the last section that those local regions can extend over a significant area. In this section we summarize results, having to do with the statistical variability of FT(n), PER(n) and AR(n) spectral estimators in the case of random processes with mixed spectrum. Both of these rely on the statistics of the lagged-product correlation estimator. The following result [31] provides this. Let

\[
\hat{R}_x(\tau) = \frac{1}{N} \sum_{n=1}^{N-\tau} x(n)x(n+\tau) \tag{2.23}
\]

\( \hat{R}_x \) denote the biased lagged-product estimator \( R_x(\tau) \). Let \( R_x = [R_x(0), \ldots, R_x(n-1)]^T \) and let \( \hat{R}_x \) denote the estimator of it. Then

Result 3. [31] \( \sqrt{N}(\hat{R}_x - R_x) \longrightarrow^d N(0, \Sigma) \) as \( N \rightarrow \infty \).

The form of the covariance, \( \Sigma \), given in [31] is lengthy and does not offer much insight. The following expression of [8] affords insight in the frequency domain:

\[
\Sigma = \frac{2}{2\pi} \int_{-\pi}^{\pi} \{4\pi(A^2/4)\delta(\omega \pm \omega_0) + S_i(\omega)\}S_e(\omega)\gamma(\omega)\gamma(\omega)^T d\omega \tag{2.24}
\]
where \( \gamma(\omega)^{tr} = [1, \cos(\omega), \cdots, \cos((n - 1)\omega)] \). This expression will be useful in characterizing the statistical variability of both the FT\((n)\) and AR\((n)\) families. It reveals the direct contribution of the tone at the tone frequency, as well as how it contributes to variability at other frequencies.

### 2.5.1 Statistical Variability of the FT\((n)\) Family of P.s.d. Estimators

In keeping with equation 2.3, we define the FT\((n)\) spectral estimator as

\[
\hat{S}_{FT(n)}(\omega) = \sum_{\tau=-(n-1)}^{n-1} \hat{R}_x(\tau)e^{-j\omega\tau}
\] (2.25)

Recall that equation 2.23 is biased. The bias factor equals \(1 - \tau/N\), so that for \(\tau \ll N\) the bias will be small. Thus, in the case where \(n \ll N\) the mean of (2.25) is approximately equal to the theoretical FT\((n)\) spectrum in (2.3). In many applications involving mechanical systems one has access to a very large amount of data relative to the number of correlation lags selected for spectral analysis. Proceeding under this assumption, and then to compute the variance of equation (2.25), notice that it may be expressed as

\[
\hat{S}_{FT(n)}(\omega) = 2\gamma(\omega)^{tr}\hat{R}_x - \hat{R}_x(0)
\] (2.26)

Substituting equation (2.24) into (2.27) gives

\[
Var\{\hat{S}_{FT(n)}(\omega)\} = (4/N)E\{|\gamma(\omega)^{tr}(\hat{R}_x - R_x)|^2\}
\] (2.27)

Substituting equation (2.24) into (2.27) gives

\[
Var\{\hat{S}_{FT(n)}(\omega)\} = \frac{4}{2\pi N} \int_{-\pi}^{\pi} \{4\pi(A^2/4)\delta(\nu \pm \omega_0) + S_i(\nu)\}S_i(\nu)|\gamma(\nu)^{tr}\gamma(\nu)|^2 d\nu
\] (2.28)

To simplify (2.28) notice that

\[
|\gamma(\nu)^{tr}\gamma(\omega)|^2 \approx (n/4)\Delta_n^2(\omega, \nu)K_n(\omega - \nu) - \Delta_n(\omega, \nu)D_n(\omega - \nu) + 1
\] (2.29)
where $\Delta_n(\omega, \nu) = \cos[(n - 1)(\omega - \nu)/2]$. The approximate equality in (2.29) relies on the fact that $n$ is sufficiently large so that for a given $\omega$ the above kernels contribute a negligible amount at negative frequencies. Substituting (2.29) into (2.26) yields the following new result.

**Result 4.** For a process of the form (2.4) where the noise is white with variance $\sigma_z^2$, the variance of the $FT(n)$ p.s.d. estimator (2.25) is given approximately as

$$
\text{Var}\{\hat{S}_{FT(n)}(\omega_0)\} \cong \frac{[(A^2/n)n^2 + 2\sigma_z^4]}{N} \quad (2.30)
$$

$$
\text{Var}\{\hat{S}_{FT(n)}(\omega)\} \cong \frac{(2 + n\pi/2)\sigma_z^4}{N} \quad \text{for } \omega \neq \omega_0 \quad (2.31)
$$

If $n$ and $N$ are selected such that $n/N \to 0$ as $n, N \to \infty$, then equation (2.31) yields the well-known variance value of $2\sigma_z^4/N$ at frequencies removed from the tone location. At the tone location we see from (2.30) that the statistical variability for large $n$ is dominated by the tone arithmetic variability, and is $O(n^2)$. Figure 2.9 and 2.10 compare (2.30) and (2.31) with variances obtained from running 500 simulations. The record size was $N = 2000$, and for each record correlation lag estimates up to order $n = 400$ were computed using (2.23). The agreement is reasonable, but the predicted quantities in (2.30) and (2.31) are slightly less than the observed values. We believe that this is, in part, due to neglecting the influence of the Dirichlet kernels, in favor of the Fejer’s kernels in obtaining (2.29). It is not unexpected that at the tone frequency the variance becomes dominated by the arithmetic variance of the tone as $n \to \infty$. This will also be seen to be the case with the PER(n) and AR(n) estimators.
Figure 2.9  Comparison of simulation results against predictions in (2.30) at tone’s frequency $f=0.25\text{Hz}$

Figure 2.10  Comparison of simulation results against predictions in (2.31) at noise frequency $f=0.1\text{Hz}$
2.5.2 Statistical Variability of the PER(n) Family of P.s.d. Estimators

The expected value of the PER(n) spectral estimator associated with (2.4) is given by (2.13). For regular random processes (i.e. without tones) the statistical variance of this estimator is well known. At this point we must recall that we are considering the truncated periodogram, where the number of utilized correlation lags, \( n \), is less than the record size, \( N \). Rather than using (2.23) in (2.12) it is more common to compute the truncated periodogram as an average of \( N/n \) periodograms associated with the contiguous data. Since the single periodogram variance approximately equals to \( S^2(\omega) \) for a colored noise process, the variance of this average of \( N/n \) periodogram estimators will converge to zero as \( N/n \rightarrow \infty \), and becomes approximately

\[
\text{Var}\{\hat{S}_{\text{PER}(n)}(\omega)\} \approx nS^2(\omega)/N \quad \text{for } |\omega - \omega_0| \gg 0 \quad (2.32)
\]

Notice that (2.32) is constant in the case of the non-truncated periodogram \( (n = N) \). This is the well known inconsistency property of the periodogram. When a tone is present then one can show that, because of the nature of the Fejer kernel, for sufficiently large order, \( n \), the tone influence will be localized about \( \omega_0 \). In that region the statistical variance will be dominated by the arithmetic variance, which is \( O(n^2) \) regardless of \( N \). It should be expected that when \( n \) is sufficiently large then (2.32) will hold approximately at frequencies removed from the tone frequency. But there is one more very important point to mention. Commonly, it has been speculated that (2.32) will hold for reasonably narrow band processes \([26],[17],[21]\). When conducting p.s.d. analysis it is often presumed that one is dealing with a purely regular process. In this situation the estimate of (2.32) is obtained by replacing the theoretical, and unknown spectrum, with the estimated one. While not shown here, it turns out that (2.32) holds very well even in the case of a tone, which is the limit of a narrow band process. If one were to replace
the noise spectrum in (2.32) by (2.13), then one obtains the variance expression

$$\text{Var}\left\{ \hat{S}_{PER(n)}(\omega) \right\} = n \left[ \frac{A_0^2}{4} K_n(\omega - \omega_0) + \frac{A_0^2}{4} K_n(\omega + \omega_0) + S_\varepsilon \otimes K_n(\omega) \right]^2 / N \quad (2.33)$$

Clearly, away from the tone frequency (2.33) yields a value close to the correct variance (2.32) for large \( n \). At the tone frequency, if we ignore the negative frequency contribution, (2.33) becomes

$$\text{Var}\left\{ \hat{S}_{PER(n)}(\omega_0) \right\} = n \left[ \frac{n A_0^2}{4} + \sigma_\varepsilon^2 \right]^2 / N \quad (2.34)$$

However, in actuality, for a mixed process of sinusoid plus normal white noise with form (2.2), where \( S_\varepsilon(\omega) = \sigma_\varepsilon^2 \), if the tone's frequency \( \omega_0 \) is exactly at a bin frequency, it is straightforward to show that the variance of the PER(n) estimators at the tone's frequency \( \omega_0 \) is:

$$\text{Var}\{ \hat{S}_{PER(n)}(\omega_0) \} = \frac{n S_\varepsilon^2(\omega_0)}{N} \left( 1 + \frac{n A_0^2}{2 S_\varepsilon(\omega_0)} \right) \quad (2.35)$$

where, \( A_0^2/(2 S_\varepsilon(\omega_0)/n) \) is the local SNR at the tone frequency. Comparing (2.34) with (2.35), we see that the presumed variance (obtained for example, using Matlab) will behave similar to the true variance as a function of \( N \), which is \( O(1/N) \). However, there are notable differences. For example, as a function of order, \( n \), the presumed variance will behave as \( O(n^3) \), while the true variance will behave as \( O(n^2) \). For a large number of averages, \( N/n \), the chi-squared distribution used in the Matlab estimation of a specified \((1 - \alpha)\%\) confidence interval (C.I.) can be approximated using a normal distribution via the central limit theorem. In this case the presumed \( 2 - \sigma \) C.I. will use (2.34), while the actual one will use (2.35). In either case, it will still be a factor of \( 2N/n \) above and below the estimated p.s.d., as would be the case for noise alone. However, at tone frequency it is ill-defined in the sense that if \( N \) and \( n \) are increased in a way such that \( N/n \) is held constant, then the C.I. will change accordingly at the tone frequency, while remaining the same at other frequencies. Simply, this is because the p.s.d. estimate of the tone itself is ill-defined, as discussed above.
2.5.3 Statistical Variability of the AR(n) Family of P.s.d. Estimators

Result 3 above was significant in that it was the first characterization of the statistics of the lagged-product autocorrelation estimator for processes with mixed spectrum. It was the essential ingredient for obtaining the following result for the family of AR(n) spectral estimators. Let $\hat{S}_{AR(n)}(\omega)$ denote the AR(n) p.s.d. estimator based on the estimated autocorrelation lags given by (2.19). The following result is from [8]:

Result 5 [8]

$$\sqrt{N}[\hat{S}_{AR(n)}(\omega) - S_{AR(n)}(\omega)] \rightarrow^d N(0, \Omega_n)$$  \hspace{1cm} (2.36)

The form of $\Omega_n$ is quite involved, and so is not repeated here for the sake of brevity. The interested reader may refer to [8]. In order to illustrate (2.36) we offer figure 2.11, which illustrates how the standard deviation depends on the AR order, $n$, at a non-tone's frequency. We see that at frequencies removed from the tone frequency, it increases sharply for values of the AR order, $n$, typically used in spectral analysis, while at higher orders the rate becomes more constant, and is approximately 1.5 dB per order doubling. We also notice that it is very insensitive to the SNR. As noted above, the complex nature of $\Omega_n$ in (2.36) precludes its use to predict this rate. However, in view of the orthogonalizing role that kernels such as the Dirichlet and Fejer type play, it is possible to obtain a simple approximation for it. This expression is from [22], for the case of a process with no tones. It is given by

$$\Omega_n \cong 2n \frac{2S^2(\omega)}{N} \quad |\omega - \omega_0| \gg 0$$  \hspace{1cm} (2.37)

The variance expression in (2.37) is extremely simple relative to that in [10] for the variance in (2.36). Furthermore, it does predict the noted rate. A comparison of (2.32) and (2.37) shows that for a given order, $n$, the AR estimator variance is three times greater than the periodogram. The fact is, however, that due to the poor resolution of the latter in favor of the former, the value of $n$ used in periodogram analysis is usually
orders of magnitude larger than that used in AR analysis. In the context of Example 1, figure 2.12 illustrates the variability of an AR(40) p.s.d. estimate based on a record of size $N = 5,000$.

2.6 Statistical Properties of Averages of Families of P.s.d. Estimators

The last two sections addressed the general behavior of families of FT(n), PER(n) and AR(n) spectra, in terms of arithmetic and statistical variability. We will investigate the behavior of averages of a given family, and in particular, what advantages might be gained. In this section we investigate the possible advantages of using an average.

Figure 2.11 Evaluation of the variance expression in (2.36) as a function of model order at a noise frequency. Also shown is the estimated variability of the average of AR spectra. (See Section 2.6 for related discussion).
of p.s.d. estimators of a given type, as opposed to a single one. As noted in Example 1 of section 2.3, it can provide some level of increased confidence in the order selection process. In section 2.4, we discovered that by averaging FT(n) spectra, one arrives at an estimator which has more desirable properties. The main difficulty with conducting an analytical study of statistical properties of an average of estimators is that of obtaining all of the joint statistics. For this reason, we will here resort to the use of simulations. Specifically, to estimate the mean and variance information associated with a family, we will use 200 realizations. Each realization includes 10,000 samples of the process (2.4). We investigate three data sets of different SNR=0.1, 1.0 and 10. We keep the noise power constant while changing the power of the sinusoid to change SNR. These simulations will be used for the FT(n), PER(n) and AR(n) investigations.
2.6.1 The FT(n) Spectra and Their Averages

From (2.9), we have that the average of the \( \{FT(k)\}_{k=1}^{n} \) spectra is simply the PER(n) spectrum. Hence, here we will denote this average by the latter.

Comparison of the Means of the FT(n) and PER(n) Spectra. In this subsection there is no need to resort to simulations, since we have expressions for the means of both spectra. The means of these two spectra are given more generally by (2.5) and (2.13), respectively. Because our current investigation focuses on white noise, only the last term in each of these equations is altered. Specifically, the terms are simply replaced by the noise variance, \( \sigma_n^2 \). Thus, the advantage offered by averaging is simply that the Dirichlet kernel is replaced by the more desirable Fejer’s kernel. They are equal at the tone frequency, and away from it the latter will be closer to the true noise spectrum than the former.

Comparison of the Standard Deviations of the FT(n) and PER(n) Spectra. For \( n \) sufficiently large, the arithmetic standard deviations of these collections of theoretical spectra can be obtained from (2.10)', (2.11) and (2.16). The statistical variances of each of these single spectra for a given \( n \) can be obtained from (2.30), (2.31) and (2.32), (2.35). But since the PER(n) spectrum is exactly the average of the \( \{FT(k)\}_{k=1}^{n} \) spectra, clearly, it includes the arithmetic variance of this collection as a portion of its statistical variance. So the advantage of averaging, in terms of variance reduction at frequencies removed from the tone (recall, at the tone both the mean and variance, being functions of \( n \), are both converging to infinity with increasing \( n \)) is obtained by simply comparing (2.31) to (2.32). The averaging procedure offers only 1 dB of reduction in the standard deviation.
2.6.2 The PER(n) Spectra and Their Averages

The average of a collection \( \{PER(k)\}_{k=1}^n \) spectra does not, to our knowledge, have a well known closed form, as was the case in the last subsection. Hence, here we are forced to conduct simulations in order to estimate the statistical variability of the \( \overline{PER(n)} \) spectral estimator.

Comparison of the Means of the \( PER(n) \) and \( \overline{PER(n)} \) Spectra. It follows from (2.14) that at non-tone frequencies, the mean of \( \overline{PER(n)} \) estimate converges to the true spectrum at rate \( O(\ln(n)/n) \), which is slower than the convergence rate of the mean of single order \( PER(n) \) estimate, which is \( O(1/n) \). At the tone frequency, it follows from (2.14) and (2.13) that the mean of \( \overline{PER(n)} \) estimate is half that of \( PER(n) \), or 3dB smaller. Because the joint statistics have no bearing on the mean of the \( \overline{PER(n)} \) estimator, we were able to compute the mean, (2.15), without the need for simulations.

Comparison of the Standard Deviations of the \( PER(n) \) and \( \overline{PER(n)} \) Spectra. To evaluate the potential advantage in terms of variance reduction, we offer figure 2.13(a) which shows that the statistical standard deviations of both \( \overline{PER(n)} \) and \( PER(n) \) spectral estimates are nearly independent of SNR and both increase by 1.5dB per order doubling at non-tone frequencies. But we gain about 2dB or 37% decrease in standard deviation by averaging for each selected order. At the tone frequency, according to (2.35), the statistical variance of \( PER(n) \) estimate is approximately proportional to and to SNR when the noise power are constant. At the tone frequency, figure 2.13(b) illustrates (2.35), and in particular, that the statistical standard deviation of single order \( PER(n) \) estimate increases by 3dB as \( n \) doubles and by 5dB as SNR increases by 10 times. From this figure we observe that the statistical standard deviation of \( \overline{PER(n)} \) estimate has similar behavior to that of the single order \( PER(n) \) estimate, while offering a reduction of 3dB or 50% in standard deviation by averaging at tone frequency.
Figure 2.13 (a) (Top) Comparison of statistical standard deviation of single order PER(n) and the averaged PER(n) spectral estimates for selected order n at non-tone frequency f = 0.1 Hz. (b) (Bottom) Comparison of statistical standard deviation of single order PER(n) and the averaged PER(n) spectral estimates for selected order n at tone frequency f = 0.3 Hz.
2.6.3 The AR(n) Spectra and Their Averages

The simulations were run for \( n = 20, 40, 80, 160 \). While (2.19) or (2.20) or (2.21) may be used to arrive at a form for the mean of the \( \overline{AR(n)} \) spectral estimator, we will forego this exercise, and simply note that at all frequencies except that of the tone the average will converge in the mean to the true spectrum, while at the tone frequency it will approach to infinity at a rate of \( O(n^2) \). This is the same behavior as that of the mean of the single order AR(n) estimator for large \( n \).

Comparison of the Standard Deviations of the AR(n) and \( \overline{AR(n)} \) Spectra. At \( \omega \neq \omega_0 \), (2.37) predicts that the statistical variance of AR(n) spectral estimate is approximately proportional to the noise power square and to the order \( n \). Thus, we see about 1.5dB increase as \( n \) doubles in figure 2.14(a) and the standard deviation is the same for the three data sets with different SNRs while the noise power keeps constant. The statistical standard deviation of the estimate is about 2dB smaller than that of the corresponding single order AR(n) spectral estimate. At \( \omega = \omega_0 \), the standard deviation of AR(n) spectral estimate is, as mentioned in (2.36), quite complex, and not amenable to analysis. The simulation results in figure 2.14(b) show that it increases by about 7.5dB as \( n \) doubles and for large \( n \), it increases by 20dB as SNR increases by 10 times which could also be approximately predicted from (2.37), if we replace the noise power spectrum with the sinusoid's power spectrum at the tone's frequency in (2.37). Thus the statistical variance at \( \omega = \omega_0 \) is approximately proportional to the square of SNR and to \( O(n^5) \). We gain more than 2dB decrease in statistical standard deviation by averaging and the gain will be higher by increasing SNR and the averaging order \( n \).
Figure 2.14 (a) (Top) Comparison of statistical standard deviation of single order AR(n) and the averaged AR(n) spectral estimates for selected order n at non-tone frequency f = 0.1 Hz. (b) (Bottom) Comparison of statistical standard deviation of single order AR(n) and the averaged AR(n) spectral estimates for selected order n at tone frequency f = 0.3 Hz.
2.7 Application to the Westland Helicopter Vibration Data

The last section suggested that there might exist little gain, in terms of variance reduction, by averaging either $\overline{\text{PER}(n)}$ or $\overline{\text{AR}(n)}$ collections, with the exception being that averaging the former can eliminate spectral oscillations in the side lobe leakage. So, in this section our focus will be limited to the variability of the collections, as was the case in the example in section 2.3. Furthermore, we will not include the FT(n) family here, as its properties are, in our opinion, not sufficiently attractive for use in mixed spectral analysis using spectral families.

In order to illuminate the value of the variability of the collections, we make a comparison of 95% (or $2 - \sigma$) confidence intervals of the PER(n) estimators for $n = 256$ and 1024 and of the AR(n) estimators for $n = 20$ and 40 spectral in relation to real vibration data from Westland Helicopter data set (file w3003001.bin). The file size is 412,464. The data were sampled at 103,116.8Hz, which we have normalized to 1 Hz. The data time duration is thus, $T=4$ seconds, which corresponds to a maximum $(1/T)$ frequency resolution of 0.25 Hz. Since we utilize a normalized the sampling frequency, 1 Hz is actually 103,116.8Hz. The spectral structure is complicated and contains many sinusoids plus highly colored noise. For this reason we heterodyned the data (including decimation by factor of 10) to restrict the range of interest to the normalized frequency range of $[0.16 - 0.21]$ Hz.

*Comparison of 95% (or $2 - \sigma$) confidence intervals corresponding to the PER(n) estimates.* Figure 2.15 and 2.16 shows that the 95% C.I. (figure 2.15) and standard deviation (figure 2.16) of PER(256) and PER(1024) p.s.d. estimators calculated using Matlab, which applies Kay's formula ([17]). The orders 256 and 1024 is chosen because people traditionally use these two orders in periodogram-based spectral analysis. As $N$ is large enough, by the central limit theorem, the variance of the PER(n) from estimate can be predicted by (2.32) at non-tone frequencies, and by and (2.33) at tone frequencies.
Figure 2.15  Comparison of 95% (or $2 - \sigma$) confidence interval of PER(n) for $n = 256$ and $1024$ for Westland data
Figure 2.16  Comparison of $2 - \sigma$ of PER(n) for $n = 256$ and 1024 for Westland data.

Figure 2.15 and 2.16 show that at 0.1671Hz, 0.1729Hz, 0.1761Hz and 0.1834Hz, the standard deviations of the PER(1024) estimate are approximately 8dB, 7dB, 9dB, and 8dB, respectively, larger than those of the PER(256) estimate. At other frequencies, we only see about 3dB increase from PER(256) to Per(1024). Now, (2.33) indicates that for large fixed $N$, if the local SNR is large, the variance of PER(n) at tone's frequency would be $O(n^5)$, corresponding to a 9dB increase of standard deviation from PER(256) to PER(1024). Thus, from figure 2.15 and 2.16, we can be reasonably confident that there is a sinusoid at 0.1761Hz. At the other three frequencies, this increase of standard deviation is close to 9dB. We need to note that first, at these three frequencies, the local SNR is much smaller than that at 0.1761Hz. Second, if the tone frequency is not exactly at bin frequency, the extent of the increase would be a little less than 9dB. So, the spectra suggest that each of the three frequencies may well also correspond to tones. It is shown in (2.32) that at non tone's frequency, the variance of PER(n) is just $O(n)$,
which is 3dB increase from the standard deviation of PER(256) to that of PER(1024). Therefore, at other frequencies, it is just regular process. If we compare the behavior of the means of the PER(256) and PER(1024) estimates with (2.13), the same conclusion follows. At the four tone frequencies the mean increases by a approximately 6dB from the PER(256) to the PER(1024) estimate, while remaining essentially unchanged at other frequencies.

**Comparison of 95% (or 2 − σ) confidence intervals corresponding to the AR(n) estimates.** The AR(20) and AR(40) spectral estimates were selected since their orders are in the range that would be obtained using most of the popular order selection rules (e.g. [17]). The 95% C.I. of AR(20) and AR(40) estimates are plotted in figure 2.17. They were obtained by using (2.37) to estimate the standard deviation, along with a normality assumption. According to (2.37), for large N, at a tone frequency, the standard deviation of AR(n) spectral estimator should be $O(n^5)$. While at a frequency away from the tone, it should be $O(n)$. Figure 2.18 shows that at 0.1761Hz, the standard deviation of AR(40) spectral estimate is 7.5dB larger than that of the AR(20) spectral estimate. However, at frequencies removed from this frequency [outside of the interval (0.17,0.18)] the standard deviation increase is only about 1.5 dB. While at several frequencies between 0.167Hz and 0.183Hz, there is more than 3dB difference between the standard deviations of the two spectral estimators, their uniform spacing suggests that they are a consequence of the distribution of extraneous model poles. In figure 2.17, we see a 6dB increase in the mean at the frequency, 0.176 Hz. By (2.20), this amount of increase is consistent with the presence of a tone. At other frequencies the two confidence intervals are nearly identical. Hence, in those regions one might assume the spectrum is not only devoid of tones, but is well characterized independent of the model order.
Figure 2.17  Comparison of 95% (or 2 - σ) confidence interval of AR(n) for n = 20 and 40 for Westland data.

Figure 2.18  Comparison of standard deviation of AR(n) for n = 20 and 40 for Westland data.
2.8 Summary and Conclusions

Though the idea of using a family of spectral estimators to characterize the spectral information of a random process has been proposed over 10 years, for example in [17] and [26], this family-based approach to statistically reliable spectral estimation has received very little attention, in particular for the case of a mixed random process which is commonly found in mechanical systems. The current literature for this family based approach is only done for the case of a regular random process with purely continuous spectrum. The major contribution of this chapter relative to present literature is a detailed statistical analysis of the statistics of certain families of spectral estimators in particular to the case of a mixed random process. The families included the FT(n), PER(n) and AR(n) p.s.d. estimators, where n denotes the number of correlation lags used.

The interest was to identify what factors control both arithmetic and statistical variability within a family. The processes considered were those composed of tones and colored noise. For very low orders, the arithmetic variability will be closely related to the noise color. For this reason, we elected to focus predominantly on the case of white noise.

In this setting the arithmetic variability associated with spectral bias is caused by the presence of the tone. The derived formulae of the arithmetical mean and variability of families of theoretical spectra of a mixed random process as a function of frequency and order was one of our key results. The limiting behavior of these families of theoretical spectra was also derived for the three families as convergence rate. The undesirable properties of the FT(n) spectrum, stemming from the Dirichlet kernel, were such that it was decided to see if averaging could improve matters. This average FT(n) spectra is, in fact, the tapered (i.e. averaged) periodogram which is so popular. The AR family was noted to have $O(n^2)$ behavior in both mean and standard deviation at tone’s frequency.
This is in contrast to the tapered periodogram, whose behavior is $O(n)$.

These arithmetic mean and variance results are valuable in the cases where the amount of available data of the mixed random process far exceeds the range of reliable correlation lags that one might consider, such that the estimated autocorrelation can be considered to be almost exact. Then our formulae can be used to estimate the mean and variance of the average of the family of estimated spectra.

There is a wealth of literature on the statistical properties of the marginal FT(n) and AR(n) spectral estimators in the case of regular random processes (see [21] for reference). In this chapter, we extended the marginal distribution of the families of FT(n) and PER(n) spectral estimator in the case of random processes with mixed spectrum. Our results were based on the statistics of the lagged-product autocorrelation estimator in [31] and [8]. Our results agreed well with sample results both at tone’s and non-tone’s frequencies. For the FT(n) spectra estimator, we found that at the tone location, the statistical variability for large $n$ is dominated by the tone arithmetical variability and is $O(n^2)$. A notable new result is that at the tone’s frequency, the variance of the PER(n) spectral estimator is $O(n^2)$ while it would be $O(n^3)$ if the mixed spectral nature is ignored, for example in Matlab. This would be particularly a problem when $n/N$ is held as constant. This means the actual variance in the former would be much smaller than the latter. The ratio of the latter to the former is asymptotically proportional to $n$, which means the difference of the two would be more substantial with bigger window length. The marginal distribution of AR(n) spectral estimator was from [8] for a mixed random process and from [22] for a regular random process.

We also studied the statistical variability of the average of the three families of spectral estimators as a method to use the family. The statistical properties of averages of families of p.s.d. estimators were obtained from simulation because the joint distribution of p.s.d. estimators of different order is not known. The results were then compared with the predicted variance of any single order p.s.d. estimator. It was shown that, a con-
vergence PER(n) estimator is obtained by averaging a non-convergence FT(n) spectrum estimators. The averaged spectral estimators also have notably reduced statistical variability than any of the single counterpart both at the tone's frequency and at non-tone's frequency. It was also noted that the use of a family, as opposed to a single spectrum, can reduce sensitivity of results to order selection factors.

An important and immediate application of these results is the problem of detecting tones. For example, of the three spectral estimators it was noted that only the AR(n) estimator converged to infinity at a rate $O(n^2)$ at a tone frequency. Both the FT(n) and the PER(n) estimators had a rate of only $O(n)$. Thus, even though the vast majority of tone detection algorithms are based on the latter, it is quite possible that the former could offer significant improvement. This was, in fact, the basis for [8].

Future works could be on the joint distribution of the three families of spectral estimators. This would be a more mathematically difficult problem, but the solution of it will lead to the use of any functions of a family of spectral estimators.
CHAPTER 3 ON THE INFLUENCE OF SAMPLING AND
OBSERVATION TIMES ON ESTIMATION OF THE
BANDWIDTH PARAMETER OF A GAUSS-MARKOV
PROCESS

This chapter addresses the sampling problem associated with the bandwidth (BW) parameter estimation of a Gauss-Markov (GM) process. Given a Gauss Markov continuous-time process of fixed time window length $T$, while sampling the process more rapidly or allowing the sampling interval $\Delta$ to approach 0, the statistical influence of the sampling interval on the parameter estimation is studied. The motivation of this chapter is to sample the process at a low rate while still obtaining near optimal statistical performance in parameter estimation as well as in prediction.

3.1 Abstract

The statistical problem of estimating the BW parameter of a GM process from a realization of fixed and finite duration $T$ at selectable sampling interval $\Delta$ is addressed in this paper. As the observation time, $T$, is fixed and finite, the variance of estimated autocorrelation and continuous-time parameter does not vanish as $\Delta$ approaches 0. This necessitate a second order Taylor expansion in deriving the parameter estimator bias and variance. The 2nd order Taylor expansion produces better bias and variance results than a first order one does. The distribution of the estimator is also discussed. According to
the gradient change of the variance, our key result is that three sample regions, which are finite, large and very large ones, corresponding to substantial, gradual, and very slight decrease in variance respectively, are quantitatively proposed. The trade off between the decrease in variance and increase in the sampling rate is also analyzed. A practical guideline of choosing sampling rate is then established. These results are then applied to the prediction problems of a time invariant GM processes to show the value of them.

3.2 Introduction

In this chapter we address the statistical problem of estimating the BW parameter of a GM process from a finite duration realization. In view of the ubiquitous presence of digital computers for both data acquisition and analysis, the interest in the sampling aspect should be self-evident. We are particularly interested in the issue of selection of the sample rate. One reason for this interest is a common false perception that faster sampling is better. It may be better for a more detailed visualization of the continuous-time process. But an unnecessarily high sampling rate can have notable drawbacks. Clearly, hardware and processing costs are higher. It can result in very large data sets, contributing to the information overload phenomenon. Another drawback is that, in relation to, for example, implementation of feedback control algorithms, it limits the sophistication of the algorithms that can be used. The computation time of such algorithms must be no greater than the sampling interval, in order to avoid lag accumulation that can result in instability.

Recently, algorithms based on the divided difference operator rather than the usual shift operator in [15], [13], [12], [14] have emerged to lessen this problem. However they only attenuated the numerical problems resulting from fast sampling, the statistical issue is not yet solved. Yet another drawback, and in fact, the one of concern in this paper, is related to parameter estimation performance. It is commonly held that the best perfor-
mance of estimators of linear system parameters such as break and natural frequencies is achieved when these frequencies are near the center of the log-frequency analysis BW, associated with the system Bode plot representation. Thus, for a continuous-time first order system with a break frequency, \( \beta \) in rad/s, the analysis BW might extend from \( 0.1\beta \) to \( 10\beta \), which corresponds to the system -23 dB crossover frequency. (Recall that first order systems have a high frequency spectral amplitude decay rate of 20 dB per decade.) If we define the analysis BW to be one half of the sampling frequency, then aliasing of energy at frequencies above the analysis BW will occur, no matter what sampling frequency is chosen. Clearly, however, in view of the high frequency roll-off behavior, aliasing up to the frequency \( \beta \) will be minor if the sampling frequency is chosen to be suitably high. In fact, it is this aliasing concern that is, perhaps, the most common rationale for selection of the lowest acceptable sampling frequency.

As noted above, slow sampling is desirable for implementation of sophisticated feedback control and tracking applications. But, if one samples slowly, then for a given duration of time, there will be fewer samples with which to use to estimate slowly time varying parameters that may be needed in an adaptive algorithm. This can result in poor parameter estimation performance. A typical solution to this problem is to sample at multiple rates: a slower rate for control and tracking purposes, and a higher one for parameter estimation. Not only does this require more hardware and computational expense, but also the question arises: how fast should sampling be conducted to good estimator performance? As mentioned above, a standard answer is: the faster the better. There is little rigorous support for this answer. It is known that, in relation to digital systems, discrete parameter estimator bias can increase significantly when such frequencies are very close to zero, in relation to the analysis BW (Stoica & Söderström [16], Kay [17], Marple [18]). In the problem of estimating the frequency of a sinusoid that is corrupted by white noise, it turns out (Lau and Sherman [20]) that both the bias and variance of the estimator are minimized when the frequency of the sinusoid is at the
center of the linear-frequency analysis BW. In fact, the bias is zero.

The majority of results along these lines are in relation to sampled random processes. If one considers only sampling rates that avoid measurable aliasing, as is done in this paper, then it is straightforward to use appropriate transformation techniques to obtain at least some idea of statistics of the continuous-time process parameter from the discrete-time one. One of the most common techniques for nonlinear transformations involves a first order Taylor series expansion. For example, Priestly uses this technique often in [21]. In the case where the observation interval is allowed to approach infinity and the sample rate is fixed, there are also numerous asymptotic results, [20], [21], [22], [23], [24], etc. These results rely on the condition that the estimator is consistent; in other words, that its variance approaches zero as the number of samples approaches infinity. But our parameter estimation problem differs from that of Priestly and others, in that we limit the duration of any realization of the process to a finite time, T. There are many situations where this is the case. For example, one could be constrained to a finite observation time, T, in transmission of a brief voice message over a channel. It could also be that one only has access to data that has been previously recorded over a finite time, T. In the problem of optimal linear estimation with finite impulse response filters, Feuer [32] proved that for a fixed time window length T, as the output is sampled more rapidly, the discrete lattice filter converges to an underlying continuous-time lattice filter. A third case would be where a sliding window of length T is used in signal processing applications involving random processes that are only locally wide sense stationary (WSS). Here, if too large a value of T is used to estimate a parameter that is time varying, estimator performance would be reduced. This case is, in fact, considered in our paper to demonstrate the value of our results. The main difficulty with the restriction of access to only a finite amount of data is that, even though the number of samples, N, approaches infinity as the sampling interval Δ approaches zero, ultimately, estimator performance is limited by T. The variance does not become arbitrarily small,
and so standard limit theorems do not generally apply here. Also, the correlation structure between a fixed number of observations changes. These distinctions pose analysis difficulties that do not arise in the case of a fixed sampling rate.

Central to this work is our observation that estimator performance can be partitioned into three distinct regions. These three distinct regions include a finite sample region in which sample rates ranging from this lowest value permitted to avoid measurable aliasing up to a value equal to four times this rate, a large sample region that extends a decade above this range, and a very large sample region that corresponds to all higher sample rates. We use the sample size terminology here for two reasons. First, for fixed real time observation duration, $T$, relatively speaking, the number of samples will increase as the sampling rate is increased. Second, in estimation theory it is more common to refer to sample size than to sample rate. It will be shown that a significant improvement in estimator performance can be realized by increasing the sampling frequency by a very modest amount in the finite sample region. This payoff becomes less, but is still notable in the large sample region. As one might expect, there is negligible improvement in estimator performance in the very large sample region, since the estimator variance is close to the minimum possible variance controlled by $T$. This result is important, in that it provides a rigorous and quantitative rationale for selection of a sampling rate, in relation to estimation of the BW parameter of a Gauss-Markov process for the situation where the observation time, $T$, is fixed.

Before proceeding to the next section, it is appropriate to briefly explain why we are interested in GM processes. One reason is that they comprise perhaps the most commonly used class of linear processes for characterizing WSS random processes. They have applications in areas ranging from characterizing atmospheric wind profiles [33] to EEG activity [34], to noise in communication systems. A second reason is that, if one considers GM processes having realizations in $d$-dimensional Euclidean space, then they are admitted to serve as the state model in a wide range of state space realizations.
of more complex processes. In this paper we confine our interest to scalar-valued GM processes. The reason for this is our desire to highlight the need for, and value of our results. If we can convince the reader that our problem is worthy of study, in and of itself, and that our results are new and valuable, then one can only hope that researchers, in addition to ourselves, will pursue this more mathematically challenging problem.

The remainder of this paper is organized as follows. In the next section we briefly review the description of a GM process, and some standard results in the literature. In section 3.4 we address the statistics of the lagged-product autocorrelation estimators in careful detail. The performance of the AR(1) parameter is addressed in section 3.5. With the aid of a second order Taylor series expansion, we observe the suggestion of the three distinct regions discussed above. Section 3.6 relates this performance to that of the estimator of the BW parameter of the original GM process. Section 3.7 includes guidelines for selection of the sampling rate. The value of our results is demonstrated in section 3.8, in the context of the prediction problem of a time invariant GM process. The prediction performance of the m-step prediction that uses an assumed deterministic BW parameter $\beta$ as well as that uses an estimated one is investigated. In particular, we show that prediction performance is highly robust with respect to estimation accuracy of $\beta$. This is important, because it allows one to use a surprisingly small observation time $T$ and still achieve nearly optimal performance with perfect knowledge of $\beta$. Finally, we summarize our main theoretical and practical contributions in section 3.9.

3.3 Estimation Procedure

Here, we describe a standard procedure for estimating the BW parameter, $\beta$, associated with a continuous time Gauss-Markov process, $X(t)$, with autocorrelation function $R_X(\tau) = \sigma_X^2 e^{-\beta |\tau|}$. It is a WSS random process that is typically described by the fol-
lowing differential equation

\[ X(t) + \beta X(t) = \epsilon(t) \] (3.1)

This equation may also be viewed as a first order dynamical system, where the “input” is the fictitious white noise process \( \epsilon(t) \) with variance \( \sigma_\epsilon^2 \), and the “output” is \( X(t) \). This is a two-parameter system. One parameter is the variance of the driving noise \( \sigma_\epsilon^2 \), and the other is the BW parameter \( \beta \) with unit rad/s. The term BW arises from the fact that the process power is constant at frequencies below \( \beta \), and rolls off at a rate of 20 dB per decade at frequencies above \( \beta \) rad/s. At the frequency \( \beta \), it has decayed 3 dB from its low frequency value. Hence, beta is more accurately termed the 3 dB process BW. The reader versed in digital signals and systems might observe that, mathematically, this process does not satisfy the Nyquist sampling theorem; a theorem that states a process can be sampled without loss of information only if it has zero energy above some frequency, and only if it is sampled at least twice that frequency. The fact that a GM process does decay with increasing frequency, however, allows one to use sampling rates which, if sufficiently high, result in minimal loss of information associated with aliasing. There is no universal consensus on a lower bound on the sampling frequency, since it depends on the measure of accuracy that is used. In this paper we will only consider sampling frequencies that are at least 20 \( \beta \) rad/s. Thus, any aliasing that does occur will be at frequencies whose energy is at least 23 dB below the low frequency content. It is also assumed that the effect of an antialiasing filter is ignored.

In (3.1), \( X(t) \) is the only process that could be observed. By sampling \( X(t) \) at a uniform sampling interval, \( \Delta \) satisfying the above assumption, we arrive approximately at a discrete-time first order autoregressive (AR(1)) process. If we denote it as \( X_k = X(k\Delta), k = 1, \cdots, N, N = T/\Delta \), then it can be expressed by:

\[ X_k = \alpha X_{k-1} + u_k \] (3.2)
with AR parameter $\alpha = e^{-\beta \Delta}$. The driving noise, or model error $u_k$ is white, with zero mean and variance $\sigma_u^2 = \sigma_X^2 \cdot (1 - \alpha^2)$. It is well known (e.g. Priestley[21], Wen[35]) that for sufficiently small $\Delta$, this variance is approximately $\sigma_u^2 = \sigma_X^2 \cdot 2\beta \Delta$.

Since $\alpha = R_1/R_0$, where $R_r$ is the autocorrelation function associated with $X(t)$ of lags $\tau = r\Delta$, our estimation procedure begins with using the biased lagged-product estimators of these correlations. Then a second-order Taylor expansion is used to estimate $\alpha$, since it is shown that a first-order expansion fails to give reliable estimator bias and variance behavior. Finally, we use a first-order Taylor expansion to estimate $\beta$ from the estimator of $\alpha$.

### 3.4 Autocorrelation Function Estimation

The biased estimator of autocorrelation of sampled process $X_k, R_r$ is:

$$R_r = \frac{1}{N} \sum_{k=1}^{N-|r|} X_k X_{k+r} \quad (3.3)$$

The mean and covariance information associated with the biased, lagged-product estimator of the discrete-time correlation function, $R_r$ are well known (e.g. [21])

$$E(\hat{R}_r) = (1 - \frac{|r|}{N}) R_r \quad (3.4)$$

and

$$\text{cov}\{\hat{R}_r, \hat{R}_{r+v}\} = \frac{1}{N} \sum_{m=-(N-r)+1}^{N-r-v+1} \left\{1 - \frac{\eta(m) + r + v}{N}\right\} \cdot \{R_m R_{m+v} + R_{m+r+v} R_{m-r} + K_4(m, r, v)\} \quad (3.5)$$

refer to $\eta(m)$ in (5.3.22) in [21]. The term $K_4(m, r, v)$ is the fourth cumulant of the distribution of $\{X_t, X_{t+r}, X_{t+m}, X_{t+m+r+v}\}$ (Isserlis [36]), when $X(t)$ is a Gaussian process, as is the case here, all joint distribution are multivariate normal and hence, $K_4(m, r, v) \equiv 0$. 

Both of these expressions are exact. If the process is Gaussian, and if its autocorrelation function decreases sufficiently fast over the observation window, then, for large $N$, the covariance may be approximated by (Bartlett [37]):

$$
\text{cov}\{\hat{R}_r, \hat{R}_{r+v}\} \simeq \frac{1}{N} \sum_{m=-\infty}^{\infty} \{R_m R_{m+v} + R_{m+r+v} R_{m-r}\} \quad (3.6)
$$

Since our interest is in the different sample size case, it is necessary to assess any assumptions related to $N$, in regard to how well (3.5) is approximated by (3.6), as we proceed in our development. We therefore offer Figure 3.1 below, in relation to these two covariance expressions.

![Graph showing the relative error of $\text{cov}\{\hat{R}_r, \hat{R}_{r+v}\}$ for $(r, v) = (0, 0)$, $(0, 1)$, $(1, 0)$ and $\beta T = 20$.](image)

**Figure 3.1** Relative error of $\text{cov}\{\hat{R}_r, \hat{R}_{r+v}\}$ for $(r, v) = (0, 0)$, $(0, 1)$, $(1, 0)$ and $\beta T = 20$

Figure 3.1 is for $\beta T = 20$; a value which corresponds to an observation window equal to 20 process time constants. This value will be used often throughout this work, for the sake of brevity. For this observation window, Figure 3.1 shows that the relative error between the covariance from (3.6) and that from (3.5) is below 3% for the entire range of
\( \beta \Delta \) values, and for all lag combinations noted. Also, significant changes in the relative error gradient occurs near \( \beta \Delta = 0.08 \) and 0.008. In accordance, three sampling rate ranges, namely, finite, large and very large ones could be identified as \( R-1 = [\beta \Delta \geq 0.08] \), \( R-2 = [0.008 \leq \beta \Delta < 0.08] \), and \( R-3 = [\beta \Delta \leq 0.008] \) corresponding to substantial, gradual, almost no gradient change. At \( R-3 \) region \( [\beta \Delta \leq 0.008] \), the relative error is nearly equal to 1.27\% for all three covariance, and stays the same, showing the lower limit of the approximation level in this case. The break point 0.008 corresponds to a sample size \( N = 2500 \), and an analysis BW that is approximately equal to the process -55 dB crossover frequency and the breaking point .08 corresponds to \( N = 250 \), and -35dB crossover frequency.

This analysis provides support for use of the approximate covariance expression at sampling rates typically chosen to avoid aliasing. For above stated sampled GM process, the covariance in (3.6) can be rewritten as an explicit function of the parameter of the process \( \alpha \) as follows:

\[
\text{cov}\{\hat{R}_r, \hat{R}_{r+v}\} \cong \frac{\sigma^4}{N} \cdot \alpha^v \cdot \left[ \frac{1 + \alpha^2 + \alpha^{2r} + \alpha^{2+2r}}{1 - \alpha^2} + v + (2r + v)\alpha^{2r} \right]
\]  

(3.7)

This explicit expression (3.7) is a new results and a key expression to derive later results. Then the predicted variance of \( \hat{R}_0, \hat{R}_1 \) is as follows:

\[
\begin{align*}
\text{Var}(\hat{R}_0) & \cong \frac{\sigma^4}{N} \cdot \frac{2(1 + \alpha^2)}{(1 - \alpha^2)} \\
\text{Var}(\hat{R}_1) & \cong \frac{\sigma^4}{N} \cdot \frac{(1 + 4\alpha^2 - \alpha^4)}{(1 - \alpha^2)} \\
\text{Cov}(\hat{R}_0, \hat{R}_1) & \cong \frac{\sigma^4}{N} \cdot \frac{4\alpha}{(1 - \alpha^2)}
\end{align*}
\]  

(3.8)

Only the first expression in (3.8) is given in previous works [21]. The other two expressions are new. Notice as \( \Delta \to 0 \), the above all three will approach the same expression

\[
\text{Var}(\hat{R}_0) \cong \text{Var}(\hat{R}_1) \cong \text{Cov}(\hat{R}_0, \hat{R}_1) \to \frac{2\sigma^4}{\beta \cdot T}
\]  

(3.9)
This is not unexpected since as $\Delta \to 0$, the random variable $X(t\Delta)$ and $X((t+1)\Delta)$ converges to $X(t\Delta)$ in the mean square sense, such that $\hat{R}_1$ approaches $\hat{R}_0$ and the three terms in (3.9) all converge to $Var(\hat{R}_0)$. It also states that the variance of $\hat{R}_0$ and $\hat{R}_1$ will not go to zero for a fixed window length $T$, instead, the limiting value is proportional to $1/T$.

The covariance behavior for these estimators described in (3.8) is illustrated in Figure 3.2. In the paper, we have analytical expressions which define the behavior for the very large sampling rate region R-3, $[0 < \beta \Delta < 0.008]$, and for the large one R-2, $[0.008 < \beta \Delta < 0.08]$. Our expressions quantify the estimator performance in these two regions, and consequently, define the region associated with what one would describe as finite sampling rate behavior.

![Figure 3.2](image)

Figure 3.2 The predicted variance from (3.8) of autocorrelation function of $X(t)$ sampling at different time interval $\Delta$.

As $\hat{R}_0$ and $\hat{R}_1$ are sums of products of correlated Gaussian random variables, the result in (3.8) and (3.9) may be compared with the corresponding result for the vari-
ance of the sample variance $s^2$ when we have $N'$ independent normal observations; as $Var(s^2) = 2\sigma^2/N'$, then for any given non zero $\Delta$, the equivalent $N'$ will be:

$$N'(\hat{R}_0) \simeq \frac{N(1 - \alpha^2)}{(1 + \alpha^2)}$$

$$N'(\hat{R}_1) \simeq \frac{2N(1 - \alpha^2)}{(1 + 4\alpha^2 - \alpha^4)}$$

and the limiting value of $N'$ for both $\hat{R}_0$ and $\hat{R}_1$ will be

$$\lim_{\Delta \to 0} N' = \beta T$$

In comparison with the distribution of the sample variance $s^2$ of $N'$ independent normal observations from the population of $N(0, \sigma^2)$, we have the following conjecture:

**Conjecture 1:** At a given sampling interval $\Delta$, the marginal distribution of $\hat{R}_0$ will have chi square distribution with $N'(\hat{R}_0)$ degree of freedom and scale parameter $\frac{\sigma^2}{N'(\hat{R}_0)}$, i.e.

$$\hat{R}_0 \sim^d \frac{\sigma^2}{N'(\hat{R}_0)} \chi^2_{N'(\hat{R}_0)}$$

or in other words, it could also be described as gamma distribution

$$\hat{R}_0 \sim^d \text{gamma} \left( \frac{N'(\hat{R}_0)}{2}, \frac{2\sigma^2}{N'(\hat{R}_0)} \right)$$

As $\Delta \to 0$, $\hat{R}_0$ would have a limiting distribution

$$\hat{R}_0 \longrightarrow^d \text{gamma} \left( \frac{\beta T}{2}, \frac{2\sigma^2}{\beta T} \right)$$

where $\sim^d$ means to have the distribution and $\longrightarrow^d$ means to converges in distribution.

In similar way, the marginal distribution of $\hat{R}_1 - E(\hat{R}_1)$ can be stated as chi-square or gamma distribution by replacing the $N'(\hat{R}_0)$ with $N'(\hat{R}_0)$ in Conjecture 1.

An important point here is that when the time window $T$ is fixed, as $\Delta$ approaches 0, the equivalent $N'$ for both the two estimated variance will not go to infinity, instead,
it is a finite number $\beta T$, then the limiting distribution of $\hat{R}_0$ will continue to be a gamma distribution with finite degree of freedom $\beta T$. This is against what might be expected, as it is usually believed that as the process is sample more rapidly, the sample size $N$ approaches infinity, then a Central limit theorem might hold and the estimated variance $\hat{R}_0$ and $\hat{R}_1$ would be approximately multivariate normal. A more rigorous proof of Conjecture 1 is not available yet. However the following sample method also supports the non-normal but gamma marginal distribution of $\hat{R}_0$ and $\hat{R}_1$.

Choose $\beta = 10$, $T = 20/\beta = 2$, the continuous-time Gauss Markov process is sampled at different rates, with $f_s = f_{\text{min}} \cdot n$, $n = 1, 2, \cdots, 32$ and $f_{\text{min}} = 10\beta/\pi$. At each rate, $M = 10000$ samples are obtained. Then samples of $\hat{R}_0$ and $\hat{R}_1$ and their sample marginal distribution could be obtained at each sampling rate with sample size 10000.

An Maximum Likelihood Estimation (MALE) gamma distribution and normal distribution is fitted to the sample data. The likelihood function is also estimated for the two fitted distribution.

![Image of distribution fitting](image)

**Figure 3.3** The sample and gamma and normal fitted distribution of $\hat{R}_1$ at $\beta \Delta = 0.0098$, $n = 32$
Figure 3.4 The sample and gamma and normal fitted distribution of $\hat{R}_0$ at $\beta\Delta = 0.0098$, $n = 32$

Figure 3.3 and figure 3.4 show the sampled and fitted gamma and normal probability distribution function(pdf) when the sample size $N = 2000$ or when the Nyquist rate is at -50dB crossover frequency. It is clear from the two figures that gamma distribution fits the sampled data much better than the normal distribution. Same conclusion could be drawn by comparing the likelihood function of the two pdf, in figure 3.3, $L_{\text{gamma}} = -2081.8$ is bigger than $L_{\text{normal}} = -2698.7$. With $\beta\Delta = 0.0098$, $\beta T = 20$ and $\sigma_z^2 = 1$, both the equivalent degree of freedom $N'(\hat{R}_0)$ and $N'(\hat{R}_0)$ equals $\beta T = 20$, substitute them in (3.14), the gamma distribution will have parameters $(10, 0.1)$, which is very close to the fitted gamma parameters $(10.45, 0.095)$ and $(10.67, 0.094)$ shown in figure 3.3 and 3.4. Another thing to emphasize here is that $\beta\Delta = 0.0098$ is at the sample size region between R-2 and R-3 where the limiting equivalent degree of freedom and distribution of $\hat{R}_0$ is already achieved.
3.5 Statistics of $\alpha$ Estimator, $\hat{\alpha}$

It is well known that this estimator converges to the value one, in the mean square sense, as $\Delta \to 0$. Our use of a second order Taylor expansion yields analytical expressions for the bias and variance of this estimator that are accurate for both very large $N$ and large $N$ regions. We discuss the reasons that a first order expansion failed.

Since $\alpha$ is also the autocorrelation coefficient of the sampled process $X_k$ of first lag, a natural estimator of $\hat{\alpha}$ is:

$$\hat{\alpha} = \frac{\hat{R}_1}{\hat{R}_0} \quad (3.16)$$

A standard method of approximating the mean and variance of (3.16) is known in statistics as the $\delta$-method; which is simply a first order Taylor series expansion of the nonlinear function of the correlation lag estimators about their means.

$$\hat{\alpha} = \frac{\hat{E}(\hat{R}_1)}{\hat{E}(\hat{R}_0)} - \frac{\hat{E}(\hat{R}_1)}{(\hat{E}(\hat{R}_0))^2}(\hat{R}_0 - \hat{E}(\hat{R}_0)) + \frac{1}{\hat{E}(\hat{R}_0)}(\hat{R}_1 - \hat{E}(\hat{R}_1)) \quad (3.17)$$

This method gives the following estimator bias and variance approximations:

$$\text{Bias}(\hat{\alpha})_{1st} \sim -\frac{\alpha}{N} \quad (3.18)$$

$$\text{Var}(\hat{\alpha})_{1st} \sim \frac{1 - \alpha^2}{N} \quad (3.19)$$

The bias and variance of $\hat{\alpha}$ in (3.18) and (3.19) are well known results as (5.3.34) and (5.2.33) by Priestley [21] and by Barlett [38], P.241. However the above first order expansion assumes that the deviation of the random variable $\hat{R}_1$ and $\hat{R}_0$ from its mean is small. In fact, (3.9) indicates that the variance of the covariance estimation will not converge to 0 as $\Delta \to 0$, which means the second order term will persist no matter which $\Delta$ is used. It also assumes by linearization that the expected value of a nonlinear function of random variables now equals the function of the expected values of the random variables. Expressed in formula, it says that
If, x, and y are r.v.s.

\[ E(g(x, y)) = g(E(x), E(y)) \]  

(3.20)

Though the above generally is not correct for nonlinear functions, by linearization, it holds. In order to evaluate the correctness of the above assumptions. A second order Taylor expansion of (3.16) is made in the vicinity of \( E(\hat{R}_1) \) and \( E(\hat{R}_0) \)

\[
\hat{\alpha} = \frac{E(\hat{R}_1)}{E(\hat{R}_0)} - \frac{E(\hat{R}_1)}{(E(\hat{R}_0))^2}(\hat{R}_0 - E(\hat{R}_0)) + \frac{1}{E(\hat{R}_0)}(\hat{R}_1 - E(\hat{R}_1)) + \left[ \frac{E(\hat{R}_1)}{(E(\hat{R}_0))^2}(\hat{R}_0 - E(\hat{R}_0))^2 - \frac{1}{E(\hat{R}_0)}(\hat{R}_0 - E(\hat{R}_0))(\hat{R}_1 - E(\hat{R}_1)) \right] 
\]

(3.21)

Then the expected value of \( \hat{\alpha} \) will be

\[
E(\hat{\alpha}) = \frac{E(\hat{R}_1)}{E(\hat{R}_0)} + \left[ \frac{E(\hat{R}_1)}{(E(\hat{R}_0))^2} Var(\hat{R}_0) - \frac{1}{(E(\hat{R}_0))^2} \cdot Cov(\hat{R}_0, \hat{R}_1) \right] 
\]

(3.22)

Substitute (3.4) and (3.8) into (3.22), we then will have

\[
Bias(\hat{\alpha})_{2nd} \sim -\frac{3\alpha}{N} 
\]

(3.23)

Comparing (3.18) and (3.23), both of them will approaches 0, as the sampling interval \( \Delta \to 0 \), they are offset by a factor of 3. Figure 3.5 supports the second order expansion \( \hat{\alpha} \) bias results.

In particular, at low sampling rate in the finite and large sample size \( \beta \Delta > 0.008 \), (3.23) give much better prediction on the bias of \( \hat{\alpha} \) than (3.18). This is a new result since it is usually expected that 1st order expansion gives reasonably good bias estimation.

We then derive the variance of \( \hat{\alpha} \) based on second order expansion. \( \hat{\alpha} \) is a function of two random variables \( \hat{R}_1 \) and \( \hat{R}_0 \). We need to have joint statistics of them. Though in previous section, it is found that the marginal distribution of the estimated correlation would be gamma distribution, here \( \hat{R}_1 \) and \( \hat{R}_0 \) could be approximately jointly multivariate normal with mean and variance expressed in (3.4) and (3.8). Substitute the above results in the second order expansion of \( \hat{\alpha} \) in (3.21), we then have,

\[
Var(\hat{\alpha})_{2nd} \sim \frac{2}{\beta T} \cdot (1 + \frac{4}{\beta T}) \cdot (1 - \alpha + \frac{\alpha}{N})^2 
\]

(3.24)
Figure 3.5 The sample and predicted bias $\hat{\alpha}$ from 1st and 2nd expansion at different sampling rates.

Comparing the variance of $\hat{\alpha}$ based on 1st and 2nd order expansion, Figure 3.6 and Figure 3.7 shows that, $\text{Var}(\hat{\alpha})_{2nd}$ tend to overestimate the variance, while $\text{Var}(\hat{\alpha})_{1st}$ tend to underestimate it. However the absolute relative difference from sample $\text{Var}(\hat{\alpha})_{2nd}$ is less than half of that of $\text{Var}(\hat{\alpha})_{1st}$ except at the finite range of sample size. As $\Delta \to 0$, $\alpha = e^{-\beta \Delta} \sim 1 - \beta \Delta \to 1$. Both (3.24) and (3.19) converges to 0 at order of $O(\beta^2 \Delta^2)$ as $\Delta \to 0$ but with different factors.

\[
\lim_{\Delta \to 0} \text{Var}(\hat{\alpha})_{1st} \cong \lim_{\Delta \to 0} \frac{2}{\beta T} (\beta \Delta)^2
\]

(3.25)

\[
\lim_{\Delta \to 0} \text{Var}(\hat{\alpha})_{2nd} \cong \lim_{\Delta \to 0} \frac{2}{\beta T} (1 + \frac{4}{\beta T})(1 + \frac{(1 - \beta \Delta)}{\beta T})^2 (\beta \Delta)^2
\]

(3.26)

Since both the bias and variance of $\hat{\alpha}$ converges to 0 as $\Delta \to 0$. It follows that $\hat{\alpha} - \alpha$ converges to 0 in the mean square sense and with probability one. By Central Limit Theorem (C.L.T.), as $\Delta \to 0$, $\hat{\alpha}$ will be normally distributed. We then discuss the distribution of $\hat{\alpha}$, in particular in terms of the three regions in figure 3.2.
Figure 3.6  The sample and predicted variance of $\hat{\alpha}$ from 1\textsuperscript{st} and 2\textsuperscript{nd} expansion of $\hat{\alpha}$ at different sampling rates with data length $T = 30/\beta$.

To save space, the sample and MLE fitted probability density function (pdf) at the large sample size $\beta \Delta = 0.0098$ is shown in Figure 3.8 from 20000 simulations at different sampling rates. It shows that the beta distribution fits the sample distribution of $\hat{\alpha}$ better than the Normal distribution since the log likelihood function is bigger for beta distribution than sample distribution. One reason that beta pdf is considered is that it is on (0, 1) corresponding to the region of $\hat{\alpha}$. In addition, the beta distribution captures the asymmetry and the right tail. Even for the sample size as large as 2037 i.e. when the analysis BW is at -53dB crossover frequency in the large sample size region, the sample distribution is more likely to be beta than normal. The absolute difference in the cumulative probability is 0.2144 for Gaussian distribution and is only .0712 for beta distribution, providing another reason for preference to beta distribution.
3.6 Statistics of $\beta$ Estimator, $\hat{\beta}$

It is also important to estimate the continuous-time parameter $\beta$ of the underlying continuous-time process. As discussed in previous sections, sampled version of process at very high sampling rate converges to a process that is independent of the continuous-time parameter $\beta$. $\beta$ as a band width parameter gives us direct information about the continuous process and would allow us to do more inference about the process than the discrete-time parameter $\alpha$. In second order continuous-time processes, the continuous-time parameter is related to the fundamental frequency and damping ratio of the dynamic systems. So they are more inherent to the system or process than the discrete time parameters. However, there is little work on the statistics of the continuous-time parameter $\beta$ estimation in relation to sampling rates. We will then discuss the statistics of estimation of $\beta$, $\hat{\beta}$. 

Figure 3.7 The absolute relative difference of predicted variance of $\hat{\alpha}$ from 1st and 2nd expansion from the sample variance at different sampling rates with data length $T = 30/\beta$. 
A natural estimator of $\beta$ is simply

$$\hat{\beta} = -\frac{\ln \hat{\alpha}}{\Delta} \quad (3.27)$$

Applying the $\delta$-method to this nonlinear function of $\hat{\alpha}$ gives

$$\hat{\beta} = \beta - \frac{1}{\Delta} \cdot \frac{1}{\alpha}(\hat{\alpha} - \alpha) \quad (3.28)$$

Again here, a first order expansion of $\hat{\alpha}$ in the vicinity of $\alpha$ is applied, the assumption of which is also the deviation of $\hat{\alpha}$ from $\alpha$ is small. Since $\hat{\alpha} - \alpha$ converges to zero in the mean square sense, at small $\Delta$, the assumption will hold well. Followed from (3.28), the relationship of the bias and variance of $\hat{\beta}$ with that of $\hat{\alpha}$ will be:

$$Bias(\hat{\beta}) = -\frac{Bias(\hat{\alpha})}{\alpha \Delta} \quad (3.29)$$

$$Var(\hat{\beta}) = \frac{Var(\hat{\alpha})}{\alpha^2 \Delta^2} \quad (3.30)$$
If the first order expansion results of mean and variance of $\hat{\alpha}$ is applied, the mean and variance of $\hat{\beta}$ can be expressed as:

$$E(\hat{\beta})_{1st} \sim \beta + \frac{1}{N\Delta} = \beta + \frac{1}{T}$$  \hspace{1cm} (3.31)

$$Var(\hat{\beta})_{1st} \sim \frac{1 - \alpha^2}{T\Delta\alpha^2} \approx \frac{2\beta}{T} + \frac{2\beta^2\Delta}{T}$$  \hspace{1cm} (3.32)

Substituting the mean and variance of $\hat{\alpha}$ in terms of second order expansion, we will then have:

$$E(\hat{\beta})_{2nd} \sim \beta + \frac{3}{N\Delta} = \beta + \frac{3}{T}$$  \hspace{1cm} (3.33)

$$Var(\hat{\beta})_{2nd} \sim \frac{2}{\beta^2 \cdot T} \cdot (1 + \frac{4}{\beta^2 T}) \cdot \frac{(1 - \alpha + \frac{\beta^2}{\Delta^2})^2}{\alpha^2 \Delta^2}$$  \hspace{1cm} (3.34)

$$\sim \frac{2}{\beta^2 \cdot T} \cdot (1 + \frac{4}{\beta^2 T}) \cdot \frac{(1 - \alpha - 1 + \frac{1}{N})/\Delta^2}{\alpha^2 \Delta^2}$$

$$\sim \frac{2\beta}{T} \cdot (1 + \frac{4}{\beta^2 T}) \cdot (1 + \frac{\beta \Delta}{2} + \frac{1}{\beta T})^2$$  \hspace{1cm} (3.35)

Compared (3.31) with (3.33), though the bias of $\hat{\alpha}$ converges to 0 for both 1st and 2nd order expressions, the bias of $\hat{\beta}$ persists as $O(1/T)$ even as $\Delta \to 0$, the two expression are off by a factor of 3. Since the 2nd order expression for the bias of $\hat{\alpha}$ is more accurate, let’s compare the two bias expression to the sample bias of $\hat{\beta}$ shown in figure 3.9. Clearly the bias from 2nd order expression gives much better prediction than from 1st order one. That’s is another reason why a 2nd order Taylor expansion rather than a 1st one is required in order to arrive at a correct bias prediction of $\hat{\beta}$.

Figure 3.10 compares the sample and predicted variance of $\hat{\beta}$ based on 1st and 2nd order expansion in (3.32) and (3.34) and (3.35). From this figure, apparently, predicted variance of $\hat{\beta}$ based on 2nd order expansion in (3.34) and (3.35) matches the sample ones much better than that on 1st order expansion not only at high sampling rate but also at very slow rate. In particular, at $\beta \Delta = 0.157$ and $\beta \Delta = 0.0049$, (3.34) and (3.35) almost the same as the sampled ones. The only difference of (3.35) (with approximation) from (3.34) (without approximation) occurs at the slowest sampling rate when $\beta \Delta = 0.314$. 
Figure 3.9  Sample and predicted $\text{Bias}(\hat{\beta})$, based on 1st and 2nd expansion for $\beta = 10$, $\beta T = 30$.

Figure 3.10  Sample and predicted $\text{Var}(\hat{\beta})$, based on 1st in (3.32) and 2nd expansion in (3.35) with approximation and in (3.34) without approximation, for $\beta = 10$, $\beta T = 30$. 
So (3.35) is a very good approximation of (3.34). Figure 3.11 presents the same figure except for $\beta T = 60$. It shows that the approximated version (3.35) actually predicts the variance of $\hat{\beta}$ better than the unapproximated version in (3.34).

Both figure 3.10 and 3.11 support the three sample size regions previously indicated in figure 3.2, i.e., finite region R-1: $\beta \Delta \in (0.08, 0.314)$ where the variance of $\hat{\beta}$ has a substantial decrease as sampling rate increases; large sample size region R-2: $\beta \Delta \in (0.008, 0.08)$, where the variance of $\hat{\beta}$ only has slightly gradual decrease as sampling rate increases; very large sample size one R-3: $\beta \Delta < 0.008$, where the variance hardly decreases as rate continues to increases.

According to C.L.T, $\hat{\beta}$ will be asymptotically Normally distributed. Figure 3.12 shows the sample, MLE Normal and Gamma fitted distribution of $\hat{\beta}$ when the sample rate is at the boundary of region R-2 and R-3. It also shows that even for the sample
size as large as 2000, $\hat{\beta}$ is more likely to have gamma distribution than to have normal one. Since the MLE fitted Gamma distribution has slightly bigger log likelihood function $-52609$ than that of normal $-53775$. In addition, about .2168 of $\hat{\beta}$ does not falls in the Normal pdf while only 0.0706 of $\hat{\beta}$ does not fall in Gamma pdf. This is obtained by integrating the absolute difference of Gamma pdf and Normal pdf from the sample one. Either pdf could be an approximation of the sample pdf, though the Gamma pdf will fit the samples slightly better.

![Figure 3.12 Sample and Normal $L_{Norm} = 53775$ and Gamma $L_{Gamma} = -52609$ fitted pdf of $\hat{\beta}$ for $\beta = 10$, $\beta T = 20$, $\Delta = 0.00098$, sample size $N = 2037$ from 20000 simulations.](image)

### 3.7 Guidelines in Selecting $\Delta$

For a first order Gauss Markov Process with parameter $\beta$, its -3dB BW is $\beta/2\pi$ Hz and its -23dB BW would be $f_{Ny} = 10\beta/2\pi$. According to sampling theorem, to avoid less than -23dB aliasing, the sampling rate must satisfy $f_s \geq f_{s_{min}} = 2f_{Ny}$. However,
with the present advances of technology, people tend to sample the process at a rate way too higher than the minimum one $f_{\text{min}}$ required by sampling theorem. If one's only concern is to achieve optimal estimation performance, then he should sample as fast as you can. However, the point is, whether such gain of fast sampling is worth the cost. In reality, performance is not the only thing that matters. The computation and storage cost and the practicability of fast sampling also have to be considered. If one were to be concerned with effective performance at reasonable cost, then a proper sampling rate should be chosen to balance the gain and the cost. With the three sampling rate regions defined, guidelines of choosing sampling rate could be provided.

Both figure 3.2, 3.10, and 3.11 supports the three sampling rate region, where the pay off of increasing sampling rate is significantly different. In finite sampling rate region R-1, $\beta \Delta \in (0.08, 0.314)$, where the sampling rate ranges from the minimum rate $f_{\text{min}} = 20\beta/2\pi$ Hz to a value four times as much the $f_{\text{min}}$, or the analysis BW is from -23dB to -35dB cross over frequency, both the variance of estimators of autocorrelation $\hat{R}_0$ and that of continuous-time parameters $\hat{\beta}$ has a significant drop. According to (3.35), at any sampling interval $\Delta$, the drop rate of the variance of $\hat{\beta}$ is $\frac{d}{d} \frac{\text{Var}(\hat{\beta})}{f_{\text{s}}} = \frac{2\beta^2 \Delta^2}{\tau} \cdot (1 + \frac{1}{\beta})(1 + \beta \Delta/2 + 1/\beta T)$, it is approximately inversely proportional to the square of the sampling rate. Thus the higher the sampling rate is, the smaller the drop rate is. For example, if $\beta T = 30$ and $\beta = 10$, then the drop rate in R-1 is approximately between 0.0887 and 0.0052. In this region, a maximum pay off by increasing of sampling rate is obtained.

In the large sampling rate region R-2, $\beta \Delta \in (0.008, 0.08)$, where the sampling rate is from four times to forty times as much $f_{\text{min}}$, or the analysis BW is from -35dB to -55dB cross over frequency, the drop rate is approximately between 0.0052 and 0.0001. We gain somewhat payoff from the increase of sampling rate.

In the vary large sampling rate region R-3, $\beta \Delta \in (0.0, 0.008)$, where sampling rate is higher than forty times as much $f_{\text{min}}$, or the analysis BW is beyond -55dB cross over...
frequency, the drop rate is less than 0.0001. We gain minimal decrease in the variance of $\hat{\beta}$.

Hence, if you want to get most out of your cost while achieving sub-optimal performance, then, the sampling rate should be chosen at the boundary of R-1 and R-2, $\beta \Delta = 0.08$. If nearly optimal performance is required regardless of the cost, then the boundary of R-2 and R-3 should be selected as the sampling interval, $\beta \Delta = 0.008$. If in a mission critical application, optimal performance is the only thing that concerns, the sampling interval should be chosen in R-3 region depend on your precision tolerance.

In the context of parameter estimation, according to the minimum variance of $\hat{\beta}$ rule, we would be able to generate quantitative guidelines in selecting $\Delta$ in relation to a variance precision tolerance value $\epsilon$.

If $\beta$ is estimated from a finite duration of process, then there is a non-zero lower bound of the variance of $\beta$ among all the $\Delta$.

$$\text{Var}(\hat{\beta})_{\text{min}} = \min_{\Delta} \text{Var}(\hat{\beta}) = \lim_{\Delta \to 0} \text{Var}(\hat{\beta})$$

$$\sim \frac{2\beta}{T} \cdot (1 + \frac{4}{\beta T}) \cdot (1 + \frac{1}{\beta T})^2$$  \hspace{1cm} (3.36)

In order that, at a give $\Delta$, the relative deviation of the variance of $\hat{\beta}$ from its lower bound is no bigger than a chosen value $\epsilon$, i.e., $\frac{\text{Var}(\hat{\beta}) - \text{Var}(\hat{\beta})_{\text{min}}}{\text{Var}(\hat{\beta})_{\text{min}}} \leq \epsilon$, where $0 \leq \epsilon \leq 1$, we will have:

$$\frac{\beta \Delta}{1 + \frac{1}{\beta T}} \leq \epsilon$$  \hspace{1cm} (3.37)

Equation (3.37) also means that the analysis BW should be at $-20 \log_{10}\left(\frac{\pi}{\epsilon (1 + \frac{1}{\beta T})}ight) + 3$ dB crossover frequency.

Before proceeding to an example to demonstrate the utility of these guidelines, a few comments on the estimation bias are in order. From (3.33) and figure 3.9, it was noted that bias of $\hat{\beta}$ is much less sensitive to sample rate selection than is variance. Even so, bias in certain applications can be an important concern. Figure 3.9 and other simulations show that in region R-1, the bias has a bigger drop to a value close to $3/T$.
at the boundary between R-1 and R-2. In R-2, the bias slightly drops to $3/T$. We arrive at similar guidelines with the bias rule.

### 3.8 Application in Prediction

In previous sections, we discuss how the choice of sampling interval $\Delta$ influences the parameter estimation of the underlying continuous-time GM random process and provide guidelines in the choice of sampling interval. To show the value of these results, in this section, we apply the results to one of the most important problems in the study of stochastic processes—prediction, which has been of interest in economic systems, for instance, forecasting the future value of a stock price, as well as in different types of physical systems, such as flight tracking and control.

We will consider time $t_0$ ahead prediction of a Gauss Markov process with constant BW parameter $\beta$, given all previous values of the process up to time $t$. We will compare the case of prediction mse that uses known true $\beta$ and the one that uses $\hat{\beta}$, which is estimated from previous process values of window length $T$. A root relative error $rrmse$ is defined to make the comparison. We then derive formulae for the $rrmse$ and understand how the parameters $\beta\Delta$, $\beta t_0$ and $\beta T$ influence the $rrmse$. Contour plot of $rrmse$ are provided and an example of how we use such contour plot to design the prediction filter in an active noise cancellation system is provided.

#### 3.8.1 Prediction of a Time Invariant GM Process for Various Prediction steps

The prediction of a GM process of form (3.1) with constant BW parameter $\beta(t) = \beta$ is considered. The time constant of the GM process is $\tau = 1/\beta$, and sampling interval is $\Delta$. Let’s assume the prediction time is $t_0$ and $t_0 = m\Delta$. The number of the time constant of $t_0$ is $\beta t_0$ and the number of sampling interval of $t_0$ is $m$. Given $X(s), s \leq t, t = k\Delta$
with variance $\sigma_X^2$, the true $X(t + t_0)$, the estimate $\hat{X}(t + t_0)|_{\beta}$ with known $\beta$ and the estimate $\hat{X}(t + t_0)|_{\hat{\beta}}$ with assumed $\hat{\beta} = \beta + \delta \beta$ will respectively be:

$$X(t + t_0) = x = e^{-\beta t_0} X(t) + v(t)$$

$$\hat{X}(t + t_0)|_{\beta} = \hat{x}(\beta) = e^{-\beta t_0} X(t)$$

$$\hat{X}(t + t_0)|_{\hat{\beta}} = \hat{x}(\hat{\beta}) = e^{-(\beta + \delta \beta) t_0} X(t)$$

(3.38)

where $v(t)$ is the convolution of the white noise with the system impulse response in $(t, t + t_0)$. $v(t)$ will be a moving average process with variance

$$Var(v(t)) = \sigma_X^2 \cdot (1 - e^{-2\beta t_0})$$

(3.39)

When $m = 1$, (3.39) is obvious since $v(t)$ is just the one step prediction noise $u_{k+1}$ with variance $\sigma_X^2 (1 - e^{-2\beta \Delta})$. When $m > 1$ and $m$ is integer, according to (3.2), $v(t)$ will have the form:

$$v(t) = e^{-\beta \Delta (m-1) u_{k+1} + e^{-\beta \Delta (m-2)} u_{k+2} + \cdots + u_{k+m}}$$

$$= \sum_{j=1}^{m} e^{-\beta \Delta (m-j)} u_{k+j}$$

(3.40)

In (3.40), all the $u_{k+j}$, $j = 1, \cdots, m$ are independent white sequences with variance $\sigma_X^2 \cdot (1 - e^{-2\beta \Delta})$. Then the variance of $v(t)$ for m-step prediction is

$$Var(v(t)) = Var(u_k) \cdot \sum_{j=1}^{m} e^{-2\beta \Delta (m-j)}$$

$$= \sigma_X^2 \cdot (1 - e^{-2\beta \Delta}) \cdot \left( \frac{1 - e^{-2m\beta \Delta}}{1 - e^{-2\beta \Delta}} \right) = \sigma_X^2 \cdot (1 - e^{-2m\beta \Delta})$$

(3.41)

Therefore, (3.39) also holds for $m > 1$. Then the m-step prediction root mean square error (rmse) of $\hat{X}(t + t_0)|_{\beta}$ and $\hat{X}(t + t_0)|_{\hat{\beta}}$ would respectively be:

$$rmse(\beta) = \sigma_X \cdot (1 - e^{-2\beta t_0})^{0.5} = \sigma_X (1 - e^{-2\beta m \Delta})^{0.5} = \sigma_X (1 - \alpha^m)^{0.5}$$

$$rmse(\hat{\beta}) = \sigma_X \cdot (1 - 2e^{-(\hat{\beta} + \beta) t_0} + e^{-2\hat{\beta} t_0})^{0.5}$$

(3.42)

(3.43)
As expected, the prediction performance depends only on the real prediction interval, \( t_0 \), which can be achieved for a variety of \((m, \Delta)\) combinations. Define the prediction performance index

\[
I(\beta \Delta, m) = 1 - (1 - e^{-2\beta \Delta m})^{0.5} = 1 - (1 - e^{-2\beta_0})^{0.5}
\]  

(3.44)

Then for \( \beta \Delta = 0.30 \), which is in the high variance region for \( \hat{\beta} \) in figure 3.11, and corresponds to the largest sampling interval \( \Delta \) to avoid measurable aliasing, we find that this index equals 0.33 for one-step ahead \( (m=1) \) prediction. By most standards, this level of performance would not be considered impressive. For this reason, we will consider values of \( \beta \Delta = 0.1 \), and 0.01 for our investigation in this section. These values correspond to the regions in figure 3.11, wherein, respectively, the variance is not so high, and where it has nearly achieved its lower bound. If we demand a performance index of at least 0.5, then the maximum number of prediction steps allowed for consideration includes: \( m = 1 \) for \( \beta \Delta = 0.1 \), and \( m = 14 \) for \( \beta \Delta = 0.01 \). For convenience, these parameter ranges to be used throughout our investigation are given in table 3.1.

Table 3.1 Parameter values to be used in subsequent investigations

<table>
<thead>
<tr>
<th>( \beta \Delta )</th>
<th>( m_{\text{max}} )</th>
<th>( I(\beta \Delta, m_{\text{max}}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1</td>
<td>0.57</td>
</tr>
<tr>
<td>0.01</td>
<td>14</td>
<td>0.51</td>
</tr>
</tbody>
</table>

It is also necessary to identify a range for the estimation window size, \( T \), so that the prediction performance using \( \hat{\beta} \) is measurably different from the optimal performance using \( \beta \). For, if \( T \) is too large, then there will be no notable difference. And one goal of this investigation is to identify a minimum acceptable estimation window length, so that in the case to be considered where \( \hat{\beta} \) is slowly time-varying, we can slide this window in time, and not introduce unnecessary smoothing.

To this end, suppose for the moment that \( \hat{\beta} \), our assumed value for \( \beta \), differs from \( \beta \) by a none random amount \( \delta \beta \). Let \( \text{mse}(\hat{\beta}) \) and \( \text{mse}(\beta) \) denote the prediction mean
squared error (mse) associated with $\beta$ and $\beta$. It could be proved that

$$mse(\hat{\beta}) - mse(\beta) = E(\hat{x}(\hat{\beta}) - \hat{x}(\beta))^2 = (e^{-\beta_{\theta}} - e^{-\beta_{\theta}})^2 \sigma_X^2$$ \hspace{1cm} (3.45)

(3.45) shows that the difference of the prediction mse related to $\hat{\beta}$ and to $\beta$ is also the mean squared error of the prediction from $\hat{\beta}$ and that from $\beta$. To gain a percentage of this mse difference, we can normalize it in two ways. One way is to divide this difference with the prediction $mse(\beta)$. Thus we will arrive at the deterministic relative error of the prediction $mse(\hat{\beta})$ from $mse(\beta)$, which we will express as $drmse(\beta, \delta \beta)$. The other way is to divide it with $E(\hat{x}(\beta))^2$. This ratio expressed as $drmse(\beta, \delta \beta)$ means the deterministic relative error of the $\hat{x}(\hat{\beta})$ from $\hat{x}(\beta)$. The two denominators are only functions of $\beta_{\theta}$, so the two relative errors are just off by a constant factor that is independent of $\delta \beta$. The reason for the term deterministic is that no statistical analysis was performed. The error, $\delta \beta$, was taken as a deterministic error, not a random one.

We may be more interested in the relative error of prediction in the latter case. Let us look at the two relative errors respectively. In fact, the relationship between them are:

$$drmse(\beta, \delta \beta) = \frac{mse(\hat{\beta}) - mse(\beta)}{mse(\beta)} = \frac{(1 - e^{-\delta \beta_{\theta}})^2}{(e^{2\beta_{\theta}} - 1)}$$ \hspace{1cm} (3.46)

$$drmse(\beta, \delta \beta) = \frac{E(\hat{x}(\hat{\beta}) - \hat{x}(\beta))^2}{E(\hat{x}(\beta))^2} = (1 - e^{-\delta \beta_{\theta}})^2$$ \hspace{1cm} (3.47)

$$drmse(\beta, \delta \beta) = drmse(\beta, \delta \beta) \cdot \frac{E(\hat{x}(\beta))^2}{mse(\beta)} = \frac{drmse(\beta, \delta \beta)}{e^{2\beta_{\theta}} - 1}$$ \hspace{1cm} (3.48)

Since both (3.46) and (3.47) contains a squared term, we will take a square root of $drmse(\beta, \delta \beta)$ and $drmse(\beta, \delta \beta)$ such that root relative errors in percentage are obtained:

$$drrmse(\beta, \delta \beta) = \left( \frac{mse(\hat{\beta}) - mse(\beta)}{mse(\beta)} \right)^{0.5} \frac{|1 - e^{-\delta \beta_{\theta}}|}{(e^{2\beta_{\theta}} - 1)^{0.5}} = \frac{|1 - e^{-\delta \beta m_{\Delta}}|}{(e^{2\beta m_{\Delta}} - 1)^{0.5}}$$ \hspace{1cm} (3.49)

$$drrmse(\beta, \delta \beta) = \left( \frac{E(\hat{x}(\hat{\beta}) - \hat{x}(\beta))^2}{E(\hat{x}(\beta))^2} \right)^{0.5} = |1 - e^{-\delta \beta_{\theta}}| = |1 - e^{-\delta \beta m_{\Delta}}|$$ \hspace{1cm} (3.50)
Notice that (3.49) and (3.50) depends on $\beta m \Delta$, and $\delta \beta / \beta$. In table 3.1, we show that to have a performance indexed at least bigger than 0.5, $\beta t_0$ to be considered will be 0.1 and 0.14. For this range of $\beta t_0$ and $-0.5 < \delta \beta / \beta < 1$, we could expand the form of $e^y$ with $1 + y$ in (3.49) and (3.50). Then the deterministic root relative error will approximately be,

$$drrmse_0(\beta, \delta \beta) = \frac{|1 - e^{-\delta \beta m \Delta}|}{(e^{\delta \beta m \Delta} - 1)^{0.5}} \approx \frac{|\delta \beta|}{\beta} \sqrt{\frac{\beta m \Delta}{2}}$$

(3.51)

$$drrmse_1(\beta, \delta \beta) = |1 - e^{-\delta \beta m \Delta}| \approx \frac{|\delta \beta|}{\beta} \beta m \Delta$$

(3.52)

(3.51) and (3.52) shows that the two $drrmse$ are functions of $\delta \beta / \beta$ and $\beta m \Delta$. In particular, for $\beta m \Delta < 0.14$ and $\delta \beta / \beta \in (-0.5, 1)$, both are approximately linear functions of $|\delta \beta / \beta|$ with a slope of $\sqrt{\frac{\beta m \Delta}{2}}$ and $\beta m \Delta$ respectively. As long as $\beta m \Delta$ is fixed, no matter which combination of $m$ and $\beta \Delta$ is chosen, the $drrmse$ will be the same. Figure 3.13 and 3.14, though obtained from (3.49) and (3.50), illustrate this approximate functional relationship of $drrmse_0$ and $drrmse_1$ to $\beta m \Delta$ and $\delta \beta / \beta$ very well. Note that the actual slopes are very close to that predicted in (3.51) and (3.52). Also note that, as expected, the $drrmse_0$ and $drrmse_1$ for $m = 1$, $\beta \Delta = 0.1$ is equal to that for $m = 10$, $\beta \Delta = 0.01$. But more importantly, the prediction $drrmse_1$ is less than 14% when $\hat{\beta}$ is as much as twice the true value $\beta$. Thus, while there are certainly many applications of GM processes wherein accurate estimation of the GM parameter is crucial, the prediction application does not appear to be one of them. Exceptions might include high accuracy navigation and target settings, wherein inaccuracies on the order of 10% cannot be tolerated.
Figure 3.13  Plots of the $drrmse_0$, (3.49) as a function the number of prediction steps $m$, and of $\delta \beta / \beta$ for $\beta \Delta = 0.1$ with $m = 1$, and for $\beta \Delta = 0.01$, with $m = 1, 5, 10, \text{and} 14$.

Figure 3.14  Plots of the $drrmse_1$ in (3.50) as a function the number of prediction steps $m$, and of $\delta \beta / \beta$ for $\beta \Delta = 0.1$ with $m = 1$, and for $\beta \Delta = 0.01$, with $m = 1, 5, 10, \text{and} 14$. 
If we interpret $\delta \beta$ as the $2 - \sigma$ uncertainty associated with the estimator $\hat{\beta}$, then for the range of interest shown in figure 3.13 along with the associated variance expression (3.35), we have

$$
\frac{\delta \beta}{\beta} = \frac{2\sigma_{\beta}}{\beta} \approx 2\sqrt{\frac{2}{\beta T} \left(1 + \frac{4}{\beta T}\right) \left(1 + \frac{\beta \Delta}{2} + \frac{1}{\beta T}\right)^2}
$$

(3.53)

If we set (3.53) equal to the largest considered value, $\delta \beta/\beta = 1$, and solve for $\beta T$, we find that $\beta T = 25$. This solution is nearly independent of which of the values, 0.1 or 0.01, associated with $\beta \Delta$ in table 3.1, is used in (3.53). The reason is that both of these values are negligible relative to 1.

For a sampling interval $\Delta = 0.1$ seconds (s), $T = 25s$, and $\beta = 1 rad/s$, figure 3.15 shows a segment of the relative error between $\beta$ and $\hat{\beta}(t)$, as well as the associated root relative error between the mse 1-step predictor that uses $\beta$ and the one that uses $\hat{\beta}(t)$. Figure 3.16 is the same as figure 3.15 except for $\Delta = 0.01$, $m = 10$.

![Figure 3.15](image)

Figure 3.15 The relative error (%) for $(\beta, \hat{\beta})$ and for $(\hat{x}_m(\beta), \hat{x}_m(\hat{\beta}))$ for a segment of a sampled GM process with $\sigma^2_x = 1$, $\beta = 1$, $\Delta = 0.1$, $m = 1$ and sliding window length $T = 25$ seconds.
The relative errors in figure 3.14 are consistent with the above discussion. Even though the error between $\beta$ and $\hat{\beta}(t)$ ranges between -50% and +100%, this results in sub-optimal prediction on the order of only 5%. Furthermore, this holds for both $(\Delta = 0.1, \ m = 1)$ and $(\Delta = 0.01, \ m = 10)$. Thus, it would appear that the deterministic error analysis that resulted in (3.51) and (3.52) can provide valuable insight in the case where the actual error is a r.v., if one identifies the error, $\delta \beta$ with a range of variation of this r.v.

To complete this example, we will develop a formula similar to (3.51) and (3.52), but where $\hat{\beta}$ is a r.v. and is the actual estimated $\beta$ and the error $\delta \beta$ is, indeed, the actual error $\hat{\beta} - \beta$. This is a simple development, due to the fact that $\alpha$ and $\hat{\alpha}$ are exponential functions of $\beta$ and $\hat{\beta}$. For this reason, our development is a straightforward application of moment generating functions. We assume $\hat{\beta}$ is estimated from another sample realization of $X(t)$ of window size $T$, so that $\hat{\beta}$ and $X(t)$ is independent. And $\hat{\beta}$ is a normal r.v. with mean $\mu_\beta$ and $\sigma^2_\beta$. The mean and variance of $\hat{\beta}$ can be calculated
from (3.33) and (3.35). Then (3.45) will now become:

\[ mse(\hat{\beta}) - mse(\beta) = E(\hat{x}(\hat{\beta}) - \hat{x}(\beta))^2 = \sigma_X^2 \cdot E(e^{-\beta_0} - e^{-\beta_0})^2 \] (3.54)

We will define similar relative error as \textit{drrmse0} and \textit{drrmse1}, called \textit{srrmse0} and \textit{srrmse1}, which means statistical root relative error.

\[ srrmse0(\beta, \delta\beta) = \left( \frac{mse(\hat{\beta}) - mse(\beta)}{mse(\beta)} \right)^{0.5} = \left( \frac{E(e^{-\delta\beta m\Delta} - 1)^2}{e^{2\beta m\Delta} - 1} \right)^{0.5} \] (3.55)

\[ srrmse1(\beta, \delta\beta) = \left( \frac{E(\hat{x}(\hat{\beta}) - \hat{x}(\beta))^2}{E(\hat{x}(\beta))^2} \right)^{0.5} = \left( E(e^{-\delta\beta m\Delta} - 1)^2 \right)^{0.5} \] (3.56)

According to (3.33) and (3.35), we also have

\[ \delta\beta = \hat{\beta} - \beta \sim^d N\left(\frac{3}{T}, \sigma^2_{\beta}\right) \]

\[ \delta\beta m\Delta \sim^d N\left(\frac{3}{T} m\Delta, \sigma^2_{\beta} m^2 \Delta^2\right) \] (3.57)

where, \( \sim^d \) means to have the probability distribution. If \( Y \) is a normal r.v. with mean \( \mu \) and variance \( \sigma^2 \), then \( e^Y \) will be a lognormal r.v. with mean \( e^{\mu + \frac{1}{2} \sigma^2} \) and variance \( e^{2(\mu + \sigma^2)} - e^{2\mu + \sigma^2} \). Using this lognormal distribution, (3.55) and (3.56) will be:

\[ srrmse0(\beta, \delta\beta) = \left( \frac{e^{-\frac{3m\Delta}{T} + 2m^2 \Delta^2 \sigma^2_{\beta} - 2e^{-\frac{3m\Delta}{T} + m^2 \Delta^2 \sigma^2_{\beta}} + 1}}{e^{2\beta m\Delta} - 1} \right)^{0.5} \] (3.58)

\[ srrmse1(\beta, \delta\beta) = \left( \frac{e^{-\frac{6m\Delta}{T} + 2m^2 \Delta^2 \sigma^2_{\beta} - 2e^{-\frac{6m\Delta}{T} + m^2 \Delta^2 \sigma^2_{\beta}} + 1}}{e^{2\beta m\Delta} - 1} \right)^{0.5} \] (3.59)

In (3.58) and (3.59), \( \frac{m\Delta}{T} \) can be written as \( \frac{m\beta\Delta}{\beta T} \), which is a function of \( m\beta\Delta \) and \( \beta T \). \( m^2 \Delta^2 \sigma^2_{\beta} \) can also be written as \( (m\beta\Delta)^2 \left( \frac{\sigma^2_{\beta}}{m^2} \right) \). And \( \frac{\sigma^2_{\beta}}{m^2} = \frac{2}{3T}(1 + \frac{4}{3T})(1 + \frac{\Delta}{2T} + \frac{1}{2T})^2 \). Therefore, \textit{srrmse0} in (3.58) and \textit{srrmse1} in (3.59) are functions of \( \beta\Delta, m\beta\Delta \) and \( \beta T \). These three variables, are the only independent variables. We will explore \textit{srrmse0} and \textit{srrmse1} as functions of these variables. For the moment, only figures of \textit{srrmse0} is investigated since there is only a factor between the two.
Figure 3.17 shows the plot of $srrmse0$ as a function of $\beta T$, at different $\beta \Delta$ values shown in the legend, for fixed $m\beta \Delta = 0.1$. It shows for the same $\beta \Delta$, as $\beta T$ increases, $srrmse0$ decreases monotonically. The decrease rate continuously drops. When $\beta T$ reaches 100, $srrmse0$ is 0.03, very close to the smallest one. Figure 3.17 also shows that for the same $\beta T$ and $m\beta \Delta$, $srrmse0$ is very close for different $\beta \Delta$. This means $\beta \Delta$ has much smaller influence on $srrmse0$ than $\beta T$ does. To understand how $\beta \Delta$ influences $srrmse0$, we draw figure 3.18 and 3.19 which show $srrmse0$ as a function of $\beta \Delta$, at different $m\beta \Delta$ shown in the legend and for $\beta T = 100$ and 20. These two figures show that $srrmse0$ is almost a linear function of $\beta \Delta$. As $\beta \Delta$ decreases, $srrmse0$ also decreases but at a very modest rate. For $\beta T = 100$, this rate is nearly the same for the five different $m\beta \Delta$ values. For example, when $\beta \Delta$ drops from 0.2 to 0.01, $srrmse0$ only decreases by 0.004. Compared with the influence of $m\beta \Delta$ on $srrmse0$, the effect of $\beta \Delta$ can almost be ignored. Another phenomenon is that $srrmse0$ is the same for
$m\beta \Delta = 0.6 \text{ and } 1.0.$

Figure 3.18 Plots of the $srrmse_0$ versus $\beta \Delta$ at different $m\beta \Delta$ values for $\beta T = 100$.

Figure 3.19 Plots of the $srrmse_0$ versus $\beta \Delta$ at different $m\beta \Delta$ values for $\beta T = 20$. 
We then show $srrmse_0$ as a function of $m\beta\Delta$ at different $\beta T$ in figure 3.20. Figure 3.20 shows that when $m\beta\Delta$ increases from 0 to 2, $srrmse_0$ first increases to a maximum value and then decreases. To summary, the four figures of $srrmse_0$ shows that there appears to be a simpler functional relationship between $srrmse_0$ and $\beta T$, $m\beta\Delta$, and $\beta\Delta$ than that in (3.58).

![Figure 3.20 Plots of the $srrmse_0$ versus $m\beta\Delta$ at different $\beta T$ values for $\beta\Delta = 0.04$.](image)

For $\beta T > 20$, and $m\beta\Delta < 1$, $\frac{m\Delta}{T} < 0.05$ and $(m\beta\Delta)\frac{\sigma^{2}_{\beta}}{(\frac{\beta\Delta}{\beta T})^{2}} < 0.3$, thus the term in the exponential function in (3.58) and (3.59) is very close to 0. In the same way as we arrive at an approximation of the $drrmse_0$, we could also expand the exponential term $e^{Y}$ into $1 + Y$. If we also substitute the expression of $\sigma^{2}_{\beta}$ into (3.58) and (3.59), these two equations will become:

\[
srrmse_0(\beta, \delta, \beta) \sim \sqrt{\frac{\beta m\Delta}{2}} \cdot \frac{\sigma_{\beta}}{\beta} \approx \sqrt{\frac{\beta m\Delta}{2}} \cdot (1 + \frac{\beta\Delta}{2} + \frac{1}{\beta T}) \left[2 + \frac{2}{\beta T} \left(1 + \frac{4}{\beta T}\right)\right] \quad (3.60)
\]

\[
srrmse_1(\beta, \delta, \beta) \sim (\beta m\Delta) \cdot \frac{\sigma_{\beta}}{\beta} \approx (\beta m\Delta) \cdot (1 + \frac{\beta\Delta}{2} + \frac{1}{\beta T}) \left[2 + \frac{2}{\beta T} \left(1 + \frac{4}{\beta T}\right)\right] \quad (3.61)
\]
The approximation in (3.60) and (3.61) is a simpler function. Compared with the deterministic root relative error in (3.51) and (3.52), the statistical root relative error in (3.60) and (3.61) only replaces $\frac{\Delta}{\beta}$ with the term $\frac{\Delta}{\beta}$. From (3.60), we see that $srrmse_0$ is proportional to $\beta\Delta$ and is a strictly decreasing function of $\beta T$. $srrmse_0$ is also proportional to $\sqrt{m/\beta\Delta}$. These functional relationships are well reflected in figure 3.17, 3.18, 3.19, and 3.20. In particular, the slope of $srrmse_0$ with respect to $\beta\Delta$ in figure 3.18 and 3.19 can be predicted from (3.60). It could also be shown that the approximation in (3.60) and (3.61) is more close to (3.58) and (3.59) for smaller $m\beta\Delta$ and bigger $\beta T$.

The only difference of $srrmse_1$ from $srrmse_0$ is that it is proportional to $m\beta\Delta$ rather than $\sqrt{m\beta\Delta}$. This linear relationship of $srrmse_1$ and $m\beta\Delta$ is illustrated in figure 3.21, which shows that for $\beta T \geq 50$, the curve of $srrmse_1$ versus $m\beta\Delta$ is almost a straight line.

![Figure 3.21 Plots of the srrmse1 versus m\beta\Delta at different \beta T values for \beta\Delta = 0.04.](image)

Overall, $\beta T$ and $m\beta\Delta$ has much bigger influence on the root relative error. The
influence $\beta \Delta$ can be ignored for $\beta \Delta \leq 0.2$. This also means, in prediction, smaller sampling interval does not provide us as much advantage as bigger window size. And for requested root relative error, the above analysis can provide how to choose the estimation window size $T$ and sampling interval $\Delta$ given a rough idea of the BW parameter $\beta$ of the GM process.

For designing purpose, figure 3.22 shows the contour plot of $srrmse0$ versus $\beta T$ and $m\beta \Delta$ for $\beta T \in (20, 200)$ and $m\beta \Delta \in (0.01, 0.1)$. The contour plot for $m\beta \Delta > 0.1$ is not shown since the predictability performance index $I < 0.57$ for this range. This plot exhibits how different $\beta T$ and $m\beta \Delta$ values can result in the same $srrmse0$. The contour plot is independent of $\beta \Delta$ values because the influence of $\beta \Delta$ on $srrmse0$ can be ignored. Figure 3.23 is a zoomed in figure 3.22 with $\beta T \in (20, 40)$.

![Contour plot of srrmse0 versus $\beta T$ and $m\beta \Delta$, $\beta T \in (20, 200)$.](image)

(The numbers on the contour curve is the $srrmse0$ values.)

According to figure 3.22, as $\beta T$ increases, $srrmse0$ decreases for the same $m\beta \Delta$; in
the mean time, the optimal prediction mse that uses $\beta$ does not change since $mse(\beta) = 1 - e^{-2m\beta\Delta}$. As a result, the prediction mse that uses $\hat{\beta}$ will decrease, which is good. We also notice, that bigger $\beta T$ and bigger $m\beta\Delta$ may have equal $srrmse0$. While bigger $m\beta\Delta$ leads to bigger $mse(\beta)$. In a case when we wish to control our prediction mse, with the above information, we can decide how far $m\beta\Delta$ prediction we can make, and also decide how big window length $\beta T$ of the process is needed to estimate $\beta$.

Let’s give an example on how we can use figure 3.22 or 3.23 to design m-step prediction. In designing an active noise cancellation system, we have a sound field and desire to cancel the noise at location $P_0$ by injecting a sound wave to the sound field such that the injected noise will cancel the original sound at location $P_0$. Then the original sound at location $P_0$ will be cancelled. Figure 3.24 draws such an active noise cancellation system. In figure 3.24, the plant is the original sound field; the controller is a
prediction filter and the actuator is the speaker. Active noise cancellation works best for sound fields that are spatially simple and for sound waves in low-frequency ranges. Classical examples include a closed vehicle cabin. Here we are only interested in the noise cancellation at location \( P_0 \), assume \( P_0 \) is where a person sits in the cabin.

At location \( P_0 \) and time \( t \), the noise process is measured with a microphone as \( X(t) \), then the measurement of \( X(t) \) is fed into a controller. After some signal processing, the output of the controller \( Y(t) \) is used to drive the speaker to generate a sound \( Y(t) \). The distance between the speaker and location \( P_0 \) is \( L \). There may be some restriction that results in \( L > L_0 \). If we let this \( L_0 = 1m \). Assume the noise process from the speaker is \( Y(t) \). By the time \( Y(t) \) arrives location \( P_0 \), the original noise at \( P_0 \) is now \( X(t + t_0) \). If \( Y(t) \) has the same magnitude and negative phase as that of \( x(t + t_0) \), the noise at \( P_0 \) will be cancelled. Therefore, the filter has to predict \( X(t + t_0) \) from \( x(s), s < t \).

If we assume the time between the instant of the measurement signal \( x(s), s \leq t \) and the output of speaker is negligible compared to the time of the sound wave propagating from the speaker to location \( P_0 \) and the sound speed in air is \( c \), then \( t_0 = L_0/c \) is a restriction parameter. This is a typical prediction problem. Let’s assume \( L_0 = 1m \) and

---

**Figure 3.24** An active noise cancellation system
$c = 340m/s$, then $t_0 = 1/340s$. We will assume the noise process is a band limited 1st order GM process with constant BW parameter $\beta$. Then the optimal prediction mse that uses $\beta$ will be $mse(\beta) = 1 - e^{-2\beta t_0}$. If we require the optimal root mean square error is at most 10%, which means we have at least 90% predictability, then $mse(\beta) \leq 0.01$. This will lead to $\beta t_0 \leq 0.053$, which means we can at most predict $X(t)$ up to $\beta t_0 = 0.053$. Given this, $\beta \leq 17rad/s$, which means we can predict a GM process with maximum $\beta = 17rad/s$. Any GM process whose $\beta$ is bigger than $17rad/s$ will not be able to be dealt with in this situation.

A second step would be to use the contour plot of $srrmse0$ to decide the window length $T$ of the previous process values that need to be used to estimate $\beta$. We will also require that the root relative error of mse that uses $\hat{\beta}$ from that uses $\beta$ is 0.052, i.e., $srrmse0 = 0.05$. As $\beta t_0 = \beta m \Delta = 0.053rad$, from figure 3.23, the $\beta T$ required for $srrmse0 = 0.0516rad$ and 0.0485rad, is 26 and 29. In order that $srrmse0 \leq 0.05$, $\beta T \geq 27.5rad$. As $\frac{1}{\beta} \geq \frac{1}{17}s/rad$, then $T \geq \frac{\beta T}{\frac{1}{17}} = 1.62s$. With this information, a minimum $T$ could help to reduce the filter computation time. A maximum sampling interval would then be $t_0$ if $\beta t_0$ satisfies the antialiasing requirement. In that, we will maintain the prediction performance while paying minimal computation and storage cost.

Therefore, the prediction filter has to used at least previous 1.62s of process data to make $\frac{1}{340}s$ ahead prediction with at least 90% predictability. As $m \geq 1$, $\beta \Delta \leq 0.053rad$. The maximum sampling interval $\Delta$ will be $\frac{0.053}{\beta}s/cycle$. However we could also choose $m = 2$, but we have to sample the process faster with $\Delta \leq \frac{0.0265}{\beta}s/cycle$. As long as $\beta \leq 17rad/s$, with the above setting, the 90% predictability will be obtained. In fact, when $\beta < 17rad/s$, using the above design parameters, we will have a better than 90% predictability. This is because smaller $\beta m \Delta$ will lead to smaller $mse(\beta)$, and with the same ratio of $\beta m \Delta$, and $\beta T$, the $srrmse0$ will be even smaller.

The above discussion on how the root relative error of prediction mse is influenced
by \( m\beta \Delta \) and \( \beta T \) as well as \( \beta \Delta \) is based on the mean and variance expressions of \( \hat{\beta} \).

We offer figure 3.25 which shows sample prediction \( mse(\hat{\beta}) \) with estimation window length \( T = 200/\beta \) of a GM r.p. sampled at five different rates to show that a higher sampling rate offers almost no advantage on the prediction mse over a smaller rate. The relative difference among the mse at different rates is less than 1%. This gives another illustration that higher sampling rate does not decrease prediction mse.

Figure 3.25 Sample prediction \( mse(\hat{\beta}) \) versus \( \beta t_0 \) for a GM process with \( \beta = 10, \beta T = 200 \), from 500 simulation at 5 different sampling rates with the shown \( \beta \Delta \) value in the legend, using all previous value to estimate \( \hat{\beta} \).

3.9 Conclusion

The influence of sampling interval and observation time on the parameter estimation of a continuous-time GM process sampled at different rates was considered. The problem was approached by assuming the observation time \( T \) fixed while allowing the sampling interval \( \Delta \) to approach 0, such that the sample size \( N = T/\Delta \) approaches in-
finity, which corresponds to applications of many areas of science and engineering. Half analytical expressions of bias and variance of both the discrete-time and continuous-time parameter estimator based on the 1st order and 2nd order expansion of $\hat{\alpha}$ in the vicinity of the estimated biased autocorrelation $\hat{R}_0$ and $\hat{R}_1$ were provided and compared with simulation results. It was shown that a second order expression provides better results in both the bias and variance of parameter estimation, in particular, the continuous-time parameter estimation. From the above results, three sample size region—R-1 finite, R-2 large and R-3 very large ones were identified according to the gradient change of variance of the estimator of parameters and autocorrelations. A conjecture about the distribution of autocorrelation estimator was provided. The distribution of $\hat{\alpha}$ and $\hat{\beta}$ were also investigated. It was shown that the sampling rate should not be higher than that in the middle of R-2 large sample size region, i.e., $\beta \Delta \geq 0.04$, to obtain a near optimal performance of parameter estimation. It gave statistical point of view why a 40kHz crossover frequency Nyquist rate is usually applied in engineering. Guidelines in the choice of sampling interval were provided numerically. The application section applied the analytical results in prediction problem of a GM process with constant BW parameter to achieve the required prediction performance with minimal computation and storage cost. A deterministic and statistical analysis on the root relative errors of prediction mse of a GM process with constant BW parameter is performed. The explicit functional relationship between these variables and $m\beta \Delta$, $\beta T$ and $\beta \Delta$ were provided. It was showed that in prediction, the influence of sampling interval $\Delta$ on the prediction performance can be ignored compared with that of $T$ and $t_0$ as long as antialiasing is satisfied. The above information may be used to design an m-step prediction of a GM process with different restrictions, for example to design the prediction controller of an active noise cancellation system. Overall, this chapter provided statistical point of view that slow sampling rate should be applied in terms of parameter estimation and prediction of a continuous-time stochastic process with fixed and finite observation time.
3.10 Appendix

3.10.1 Derivation of Bias(\(\hat{\alpha}\)) Based on a 2\textsuperscript{nd} Order Taylor Expansion

To save the reader time and effort, the derivation of the bias and variance of \(\hat{\alpha}\) based on a 2\textsuperscript{nd} order Taylor expansion is provided here. The derivation of the bias of \(\hat{\alpha}\) does not require the joint pdf of \(\hat{R}_0\) and \(\hat{R}_1\). Instead, only the mean and covariance of \(\hat{R}_0\) and \(\hat{R}_1\), which is expressed in (3.4) and (3.8), is needed. Recall (3.21).

\[
E(\hat{\alpha}) = \frac{E(\hat{R}_1)}{E(\hat{R}_0)} + \left[ \frac{E(\hat{R}_1)}{(E(\hat{R}_0))^3} Var(\hat{R}_0) - \frac{1}{(E(\hat{R}_0))^2} \cdot Cov(\hat{R}_0, \hat{R}_1) \right]
\]

(3.62)

\[
E(\hat{\alpha}) = \frac{E(\hat{R}_1)}{E(\hat{R}_0)} + \left[ \frac{E(\hat{R}_1)}{(E(\hat{R}_0))^3} Var(\hat{R}_0) - \frac{1}{(E(\hat{R}_0))^2} \cdot Cov(\hat{R}_0, \hat{R}_1) \right]
\]

\[
= (1 - \frac{1}{N})\alpha + (1 - \frac{1}{N^2}) \cdot \frac{2\alpha(1 + \alpha^2)}{N(1 - \alpha^2)} - \frac{4\alpha}{N(1 - \alpha^2)}
\]

\[
= (1 - \frac{1}{N})\alpha + \frac{2(\alpha^3 - \alpha)}{N(1 - \alpha^2)} - \frac{2\alpha(1 + \alpha^2)}{N^2(1 - \alpha^2)}
\]

\[
= \alpha - \frac{3\alpha}{N} - \frac{2\alpha(1 + \alpha^2)}{N^2(1 - \alpha^2)}
\]

\[
\sim \alpha - \frac{3\alpha}{N}
\]

Thus the bias of \(\hat{\alpha}\) in (3.23) could be obtained.
3.10.2 Derivation of the Variance of $\hat{\alpha}$ Based on a 2nd Order Taylor Expansion

From (3.21), the variance of $\hat{\alpha}$ could be expressed as:

$$
\text{Var}(\hat{\alpha}) = \left(\frac{E(\hat{R}_1)}{E(\hat{R}_0)}\right)^2 \text{Var}(\hat{R}_0) + \frac{1}{R_0^2} \text{Var}(\hat{R}_1) \\
- \frac{2E(\hat{R}_1)}{R_0^3} \text{Cov}\{(\hat{R}_0 - E(\hat{R}_0)), (\hat{R}_1 - E(\hat{R}_1))\} \\
+ \left(\frac{E(\hat{R}_1)}{R_0^3}\right)^2 \cdot \text{Var}((\hat{R}_0 - E(\hat{R}_0))^2) \\
+ \left(\frac{1}{R_0^3}\right)^2 \text{Var}((\hat{R}_0 - E(\hat{R}_0)) \cdot (\hat{R}_1 - E(\hat{R}_1))) \\
- \frac{2E(\hat{R}_1)}{R_0^3} \cdot \text{Cov}\{(\hat{R}_0 - E(\hat{R}_0), (\hat{R}_0 - E(\hat{R}_0))^2\} \\
+ \frac{2E(\hat{R}_1)}{R_0^3} \cdot \text{Cov}\{(\hat{R}_1 - E(\hat{R}_1), (\hat{R}_0 - E(\hat{R}_0))(\hat{R}_1 - E(\hat{R}_1))\} \\
+ \frac{2E(\hat{R}_1)}{R_0^3} \cdot \text{Cov}\{(\hat{R}_1 - E(\hat{R}_1), (\hat{R}_0 - E(\hat{R}_0))^2\} \\
- \frac{2}{R_0^3} \cdot \text{Cov}\{(\hat{R}_1 - E(\hat{R}_1), (\hat{R}_0 - E(\hat{R}_0))(\hat{R}_1 - E(\hat{R}_1))\} \\
+ \frac{2E(\hat{R}_1)}{R_0^3} \cdot \text{Cov}\{(\hat{R}_0 - E(\hat{R}_0))^2, (\hat{R}_0 - E(\hat{R}_0))(\hat{R}_1 - E(\hat{R}_1))\}
$$

This is a fairly complicated equation containing ten terms on the right hand side. We only have the mean and covariance of $\hat{R}_0$ and $\hat{R}_1$. Without the joint pdf information of $\hat{R}_0$ and $\hat{R}_1$, it is not possible to get the higher order joint moments. To solve the problem, we need some assumption of the joint pdf. Though we have the conjecture stating that the marginal pdf of $\hat{R}_0$ and $\hat{R}_1$ are gamma distribution. Here we assume that $\hat{R}_0 - E(\hat{R}_0)$ and $\hat{R}_1 - E(\hat{R}_1)$ have jointly multivariate normal distribution with mean 0 and variance stated in (3.8) and (3.9). If we express the joint pdf of two random variable
\( U = \hat{R}_0 - E(\hat{R}_0) \) and \( V = \hat{R}_1 - E(\hat{R}_1) \) as:

\[
\begin{pmatrix}
U = \hat{R}_0 - E(\hat{R}_0) \\
V = \hat{R}_1 - E(\hat{R}_1)
\end{pmatrix}
\sim d
\begin{pmatrix}
0 \\
0
\end{pmatrix}
\begin{pmatrix}
\sigma_0^2 & \rho \sigma_0 \sigma_1 \\
\rho \sigma_0 \sigma_1 & \sigma_1^2
\end{pmatrix}
\]

Then we have:

\[
\text{Cov}\{U, V\} = \rho \sigma_0 \sigma_1
\]
\[
\text{Cov}\{U, U^2\} = \text{Cov}\{U, V^2\} = \text{Cov}\{V, U^2\} = \text{Cov}\{V, V^2\} = 0
\]
\[
\text{Cov}\{U^2, UV\} = 2\rho \sigma_0 \sigma_1 \sigma_0^2
\]
\[
\text{Var}(U^2) = 2\sigma_0^4
\]
\[
\text{Var}(V^2) = 2\sigma_1^4
\]
\[
\text{Var}(UV) = \sigma_0^2 \sigma_1^2 (1 + \rho^2)
\]

Even with the above assumption, the form of the variance of \( \hat{\alpha} \) would still be very complicated. From figure 3.2, the three term in the covariance matrix of \( \hat{R}_0 \) and \( \hat{R}_1 \) are already very close to there limiting values \( \frac{2\sigma_0^4}{\beta T} \) at the boundary of sampling rate region R-1 and R-2. We then further assume:

\[
\sigma_0^2 = \rho \sigma_0 \sigma_1 = \sigma_1^2 = \frac{2\sigma_0^4}{\beta T} = \frac{2R_0^2}{\beta T}
\]

With the above two approximation, we then have:

\[
\text{Var}(\hat{\alpha}) \sim \frac{\sigma_0^2}{R_0^2} \left[ \frac{(1 - \alpha/N)^2 \alpha^2}{R_0^2} + \frac{1}{R_0^2} - \frac{2(1 - \alpha/N)\alpha}{R_0^2} \right]
\]
\[
+ \frac{2\sigma_0^2}{R_0^2} \left[ \frac{(1 - \alpha/N)^2 R_0^2 \alpha^2}{R_0^2} + \frac{1}{R_0^2} - \frac{2(1 - \alpha/N)R_0 \alpha}{R_0^2} \right]
\]
\[
\sim \frac{\sigma_0^2}{R_0^2} \left( 1 + \frac{2\sigma_0^2}{R_0^2} \right) \left[ (1 - \frac{\alpha}{N})^2 \alpha^2 + 1 - 2(1 - \frac{\alpha}{N}) \alpha \right]
\]
\[
\sim \frac{2}{\beta T} \left( 1 + \frac{4}{\beta T} \right) \left( 1 - \alpha + \frac{\alpha}{N} \right)^2
\]

Even though two approximation are made in third derivation, the resulted variance of both \( \hat{\alpha} \) and \( \hat{\beta} \) matches the sample results even at the middle of finite sampling rate region R-1 (refer to figure 3.6, 3.10 and 3.11).
CHAPTER 4 APPLICATION OF PREVIOUS RESULTS ON SOUND PRESSURE DATA ANALYSIS

In previous chapters, theoretical results of two problems were presented. One was to detect tones using variability related to families of spectral estimators of mixed random process. The other was on the statistical influence of sampling interval and observation time on the parameter estimation of a continuous-time Gauss Markov process. The goal of this chapter is to demonstrate how these results can be used in relation to:

- characterizing the stochastic structure of the noise from an engine cooling system;
- constructing hypothesis tests for deciding whether a change in the engine shape has a significant effect on components of this noise.

4.1 The Objective of the Measurement

The purpose of this experiment\(^1\) is to analyze the noise generated from a tractor engine cooling system and to identify sources of important frequency component of the noise. In an engine cooling system, the inlet air flows through a radiator and a fan with a shroud, and then the discharge of the fan passes the engine block. The front view of the radiator is shown in figure 4.1.

Currently, tractor engines can be made very quiet. As a result, the engine cooling system becomes the major noise source in a tractor. When a fan is operated together

\(^1\)The experiment was designed by Professor Adin Mann and his graduate student, Chen Yu, for a tractor engine cooling system noise control project under the support of John Deere.
with the radiator and engine block, the noise spectrum becomes very complex and is not completely understood, although there is mature theory [39] to predict the noise generated from a rotating fan in free field. This noise of a free rotating fan is primarily caused by fluctuating forces exerted by rotating blades on the surrounding medium and is generally harmonics of blade passing frequency (BPF). There is no theory to predict the noise from an engine cooling system. What we do know is that when the air passes through the radiator and the engine, it may produce broadband noise, and with its passage through the fan in between the radiator and the engine, tones might be added to the broadband noise or the broad band noise might become modulated. When the inlet airflow hits the fan blade, it will change its flow direction. Part of the airflow goes up and hits the upper edge of the fan shroud. Part of it goes down and hits the lower shroud edge and gets bounced back. Because of the shroud asymmetry and the
fan blades rotation, unsteady fluctuating forces could be generated. The above facts all adds to the complexity of the process.

In their experiment, one problem drawing their attention is how the physical change will affect the sound power level (SPL). Experimental methods used to estimate the overall SPL, are insufficient to characterize the noise structure, since the noise contains both narrow band and broad band spectrum. The latter has a masking effect, as it carries the majority of the acoustic energy. In [40], a semi-empirical computer model is provided to predict the noise spectrum from five different engine cooling fan assemblies running under various working conditions. Unfortunately, the method focuses more on computation, with no regard as to the possible mixed spectral nature of the process.

The factors that may influence the noise from the cooling system include the fan type (the number of blades in the fan for instance), fan speed, geometry of the shroud, geometry of the engine, the distance and relative location of the fan and the engine. However it is still not well understood how the noise is generated with the influence of all those factors. This project by itself is a very big and complex one. It is not the purpose of the author to complete the noise control project. Instead, the focus of this chapter is to characterize the stochastic structure of the noise process of an engine cooling system and to correlate the physical change of the engine block with the noise structure using the results from previous chapter.

4.2 Description of the Experiment and Measurement Data Set

The sound pressure is measured as the noise process for analysis. The experiment was conducted in the acoustics lab in Howe Hall, which is an anechoic chamber such that a free field with minimum ambient noise is obtained.

Figure 4.2 shows the right view of the experimental stand installed in this chamber. The radiator, the fan and a contoured shroud, and mock engine across the stand are
indicated in figure 4.2. The fan is 33cm behind the radiator. A plastic shield is installed surrounding the cooling system to protect the experimenter. For the first test, the rectangular prism-shaped mock engine was not installed.

Figure 4.2  The right side view of experimental stand including the radiator, fan and shroud, mock engine, shaft and shield

The sound pressure is measured by a free field, 1/2in microphone located 117cm away from the front surface of the radiator. It is pointed to the radiator so that a perpendicular incidence is obtained.

Strictly, calibration is an end-to-end check of an instrument over its entire useful frequency range and is generally a time-consuming procedure involving specialized equipment[41]. The sound level meter and microphone calibration is checked and adjusted regularly in this laboratory. Before the measurement was made, a loose calibration of the microphone at one frequency is made. An externally generated electrical signal of known amplitude and frequency, 1000Hz, is injected into the microphone amplifiers circuit, slight compensation can be made such that the output at 1000Hz is 64dB. That
compensation is called a calibration factor. It is 2.76e+6 in this experiment. The input gain is 30dB and output gain is 0dB.

The microphone is connected to the Labview data acquisition system. The data acquisition system includes a Dual-channel Low pass filter (Model SR640 from Stanford Research Inc.) The maximum sampling rate of the data acquisition system is 30KHz. Table 4.1 described each data set and the condition under which the measurement is made.

Table 4.1 Description of measurement data sets and conditions of the measurement

<table>
<thead>
<tr>
<th>No</th>
<th>Data name</th>
<th>Mock engine</th>
<th>Sample rate $f_s$ (Hz)</th>
<th>Engine speed (rpm)</th>
<th>Fan speed (rpm)</th>
<th>length of observation time (second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>nomock16000.dat</td>
<td>no</td>
<td>16000</td>
<td>1020</td>
<td>1930</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>mock16000.dat</td>
<td>yes</td>
<td>16000</td>
<td>1020</td>
<td>1930</td>
<td>2</td>
</tr>
</tbody>
</table>

The file is conformed with Matlab format. It can be loaded in Matlab using the command “load filename.dat” directly. Each data set contains 4 columns, the physical meaning of the data in each of the columns is listed in table 4.2.

Table 4.2 Description of each column in the measurement data sets

<table>
<thead>
<tr>
<th>Column No.</th>
<th>Physical meaning of the column</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>frequency points in Hz</td>
</tr>
<tr>
<td>2</td>
<td>linear frequency response in dB</td>
</tr>
<tr>
<td>3</td>
<td>time points in ms (millisecond)</td>
</tr>
<tr>
<td>4</td>
<td>time signal</td>
</tr>
</tbody>
</table>
4.3 Noise Generation Mechanism of the Engine Cooling System

All too often, statistical analysis of data is performed by either an engineer who understands the physics of the problem under study, but has limited understanding of statistics, or by a statistician who has a very limited understanding of the physics. It is our belief that in order to contribute to a better understanding of the problem, statistical analysis should be conducted only after gaining some understanding of the physics of the problem. The purpose of this section is, therefore, to summarize our effort to gain this understanding, in relation to the problem at hand. This will include a review of the major components and noise generation mechanisms, in relation to Figure 4.2. The major components include the fan and shroud, the radiator, and the presence of the engine block. The mechanisms include air flow through the radiator, fan noise, and air flow past the engine.

4.3.1 Air Flow Across the Cooling Radiator

The radiator is composed of a collection of copper tubes and plates, or fins. In this subsection we restrict our interest to noise associated with a single tube or fin. Figure 4.3 illustrates the fluid dynamics related to the presence of these elements.

When airflow passed through the radiator, as the air is viscous, a boundary layer will build up on the surface of both the tubes and fins along the streamline. As the boundary layer meets the edge of the tubes and fins, it has nowhere to attach but to break off, causing a fluctuation vortex with characteristic frequency. The tube could be considered as a cylinder.

When airflow passes an object, both the air and the object are affected. The response of the air depends on the Reynolds number \((Re)\), a non-dimensional parameter defined as \(Re = \frac{LU}{\nu}\), where \(L\) is the diameter of the cylinder, \(U\) is the airflow speed and \(\nu\) is the
kinematic viscosity of the air. At room temperature, $\nu$ is about about $1.6 \times 10^{-4}/s$. The airflow speed is around $20 ft/s$ according to measurement, and the tube diameter is about 2 inch, then the $Re$ is about $2 \times 10^4$.

Reference [42] made a flow visualization of the formation of a vortex street of flow around a cylinder at $Re$ between 40 and 200. Their flow visualization shows that for high $Re$, $Re \geq 60$, the vortices are shaped and shed alternatively away from opposite sides of the cylinder at regular intervals, forming a von Kalman vortex street[43]. Turbulence is developed, as $Re$ reaches 400. A figure of detailed kalman vortex cited from [44] is shown in figure 4.4 for the convenience of the reader.

The relationship between vortex shedding frequency and airspeed was discovered over 100 years ago. In 1878, V. Strouhal noticed an interesting phenomenon. Tones generated by a wire in the passing wind were proportional to the wind’s speed divided by the wire’s thickness. For high $Re$ flows (where viscous forces are negligible), the vortex shedding frequency $f_v$ depends only on $L$ and $U$, through a dimensionless quantity, the Strouhal number $St = \frac{f_v L}{U}$, where $L$ and $U$ mean the same as in $Re$. The Strouhal number $(St)$ is a constant and equal to 0.2 when the $Re$ ranges from $10^3$ to $10^4$(refer to figure 4.5).
Figure 4.4 A detailed kalman vortex formed from airflow passing an cylinder when $Re > 300$, from [44]

Figure 4.5 Strouhal number for vortex shedding frequency behind a circular cylinder. (Refer to [45])

Figure 4.5 suggests the $St$ of the radiator with $Re$ being $2 \times 10^4$ should be 0.2. The
tube diameter is 2 inches and the free airflow speed is about 22.9 ft/s. Then the vortex shedding frequency \( f_v \) of the radiator would be \( f_v = \frac{2\pi U}{L} \), where \( U = \frac{0.2 \times 22.9}{1/6} = 27.5 \text{ Hz} \).

The square-shaped fins also have a dimension of 2 inches, yielding the same vortex shedding frequency. The turbulence of air passing the radiator with the shedding frequency 27.5 Hz might be a possible noise tone source.

A similar phenomenon happens when the airflow passes objects of other shapes such as rectangular cylinder, causing air fluctuation with characteristic frequency determined by the dimension of the objects and the free stream speed. Therefore when the airflow passes through the radiator in the upstream of the assembly, the airflow become turbulent before it enters the fan. As in a typical heat exchanger, the fin plate spacing is much smaller than the tube diameter. The small fin plates serve to stabilize the flow by viscous interaction, effectively increasing the viscosity of the fluid and reducing the Reynolds number into the stable regime [46].

### 4.3.2 Fan Noise

There is an enormous literature relating to fan noise. This literature spans a period of over 50 years, and ranges from highly theoretical to very experimental, in nature. Our intent here, is to gain a basic appreciation of this topic, in relation to the configuration under study. Noise of a fan is a function of fan speed, turbulent airflow, cavitation, airflow noise, and bearing noise [47]. Increasing the fan speed contributes substantially to increasing both the flow rate and noise levels.

This configuration includes an axisymmetric fan with a hub diameter, \( d = 6\frac{3}{4} \text{ in} \), and with six blades. The characteristic dimensions of any blade are its length, \( l = 6\frac{1}{16} \text{ in} \), width, \( w = 5\frac{7}{8} \text{ in} \) and pitch, \( p = 29.2^\circ \). The blade tip diameter of the fan is 21.5 in.

Another important component is the shroud that surrounds the fan. The shroud has two purposes. One is to shield the fan, and the second is to improve performance by forcing more air to flow in the desired direction. The important parameters related to the
shroud are its width, \( ws = 22.5\text{in} \) and height \( h = 27\text{in} \) and the distance between it and the blade tip, tip clearance, \( dv = 3\text{in} \) along the vertical direction and \( dh = 0.5\text{in} \) along the horizontal direction.\(^2\) A tight tip clearance is desirable since it will improve flow at a given fan speed. Its byproduct is some noise reduction due to reduced turbulence near the blade tips. A reduction from a tip clearance of 3\% to 1\% of the fan diameter can be expected to yield about a 10\% improvement in flow. The additional flow can then allow a slower fan speed [48], which will in turn reduce the noise level. The fan shroud is often made asymmetric because of the compactness of the engine compartment. It is this shroud asymmetry that causes a significant amount of unsteady fluctuating forces, which is far more effective in generating sound than the steady fluctuating force. In particular, the former has the potential to increase the spectrum level at the harmonics of the BPF.

The fundamentals of fan noise generation involve the passage of air through the fan. As air is pulled through the fan, its flow field is changed from a free, steady and less turbulent field to a highly turbulent, quasi-periodic unsteady field. Therefore, the noise generated will include a more broadband component associated with this turbulence, as well as narrow band components associated with the spatially periodic nature of the blades. There are a variety of fan noise models that could be used to predicted the structure of the noise associated with our configuration [39], [40], [49], [50], [51] and [52]. The purpose of this subsection is not to offer a specific prediction model. Such an effort is beyond the scope of the limited current investigation. The investigation in this chapter will rely more on the data than on theory. The point of this subsection was to convey our appreciation of the mechanisms related to fan noise generation, in relation to the data that will be analyzed in the next section, and to point out that a more in-depth investigation could be pursued using well-established analytical tools.

\(^2\)Because it is a contoured and asymmetry shroud, it is difficult to measure its dimension. The above measurement is somewhat rough.
4.3.3 Influence of the Engine Block

A major goal of the research on this problem is to identify how the engine block influences the noise field. If the current block shape is found to have an amplifying effect on the noise, then it may be possible to reduce this effect via certain shape modifications. In subsection 4.3.1, it is noted that there are well-established theoretical results for air flow past simple shapes. However, engine blocks come in many shapes, and they are all complex. Because no readily available analytical models to predict the influence of the block exist, the approach used in this research effort is to evaluate the influence of various design changes using mock engine blocks.

And it is in relation to this evaluation that the tools we have developed in previous chapters may have the greatest value. The reason is that the evaluation must be based on experimental data associated with random processes that are a mixture of broadband and very narrow band components. So, because the evaluation is based on changes associated with frequency information related to estimated power spectral densities (p.s.d.s), the Chapter 2 results will be especially useful. In addition, in estimating changes in octave and 1/3 octave bands, We will show that the results in Chapter 3 can also be useful. In particular, if a given band is associated with a purely broadband spectrum, then it may be possible to characterize it via a Markov process. In this case, the influence of an engine block design change could be captured in the process model parameters. And statistically, test for a change in a parameter is more robust that test for a change based on the p.s.d..

The existence of the engine block located in the downstream makes the noise spectrum further more complicated. The compactness of the engine compartment not only causes the recirculation of the discharged airflow to the intake of the fan, generating the so-called chopping sound, but also raises the static pressure drop across the cooling system, which has a direct impact on the resulting flow rate, hence on the cooling
performance, but has little influence on the overall spectra.

All the above factors contribute to the complexity of the noise generated from such an engine cooling system.

4.4 Preliminary Analysis of the Sound Pressure Measurement

In this section, we proceed to analyze the sound data without concern for whether it has a mixed spectral structure. Many practitioners and researchers are not well versed in the theory and statistics of mixed random processes. Hence, one could say that in this section, the data will be studied through the eyes of a typical practitioner.

It has to be stressed that only the sound pressure at one location is measured because of the experimental limitation. Therefore the conclusion is only applicable to understand the sound pressure process at that specific location. Figure 4.6 plots part of the time signal denoted as $x(t)$ with unit voltage in mock16000.dat.

![Figure 4.6 The time signal of the sound pressure measurement $x(t)$ in mock16000.dat](image-url)
The time signal apparently exhibits some periodic phenomenon and no obvious time varying trend. In addition, the measurement is not synchronized with the shaft speed. Therefore, the sound pressure process is assumed to be WSS in the following analysis, as is a common assumption in the engineering world.

The hardware sample interval precision is 0.001ms. Since the required sampling interval for a designed rate $f_s = 16000\,Hz$ is 0.0625ms, the actual sampling interval is $\Delta = 0.062\,ms/cycle$ and sampling rate is $f_s = 16129\,Hz$.

As the sample size of $x(t)$ in mock16000.dat is $N = 32000$, the true observation time of mock16000.dat is $T = N\Delta = 1983.9\,ms = 1.984s$. For this time window, a p.s.d. estimate obtained from a non-averaged periodogram would have the highest frequency resolution, $\Delta f = 1/T = 0.504\,Hz$. But recall from Chapter 2, that this p.s.d. estimator also has the greatest variability. Hence, it is very seldom used. Instead, the data is partitioned into subintervals, and the periodograms of those intervals are averaged. This is the standard approach, and it is the one utilized in practically all p.s.d. software packages, including Matlab. It should also be noted that such packages usually utilize a Hanning window to taper the data in each subinterval, so that spectral leakage associated with discontinuities of potential tonal components is reduced by a factor of approximately 2. It should also be mentioned that the use of such a window usually includes a correction factor to account for the effect of the taper operation, which is to reduce the total power associated with the unwindowed data. But this factor, which is used in the Matlab p.s.d. package, assumes the process has not only a purely continuous spectrum, but a relatively flat one. If the spectrum were a pure line spectrum, then a different correction factor should be used. But in our situation, which is very common in problems involving rotating machinery, the spectrum is mixed. Unfortunately, there is not a p.s.d. analysis package on the market that is sufficiently intelligent to identify a mixed spectrum, and then to apply the proper correction factors to its point and continuous components. We will not pursue this in this investigation. But the reader
should be aware of it. It is mentioned here because much of this dissertation involves processes having a mixed spectrum. So, if one were to investigate such an intelligent p.s.d. estimation approach, our results could be valuable.

Here, We choose to partition the data into 7 non-overlapping subintervals of size $4096^3$. Thus, the unwindowed averaged periodogram would have a spectral resolution of $\Delta f = 1/(4096\Delta) = 3.94\text{Hz}$. But since in this section, the data is analyzed through the eyes of a typical practitioner, a Hanning window of size 4096 is utilized. Hence, the effective spectral resolution becomes $\Delta f = 7.88\text{Hz}$. Such an estimated PSD(4096) and its 95% estimated Confidence Interval(C.I.) is plotted in figure 4.7.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.7.png}
\caption{The estimated PSD(4096) and its 95% C.I. of the measure sound pressure process $x(t)$ in mock16000.dat}
\end{figure}

The C.I. of the mean of the estimated PSD(n) in *Matlab* is estimated as follows [9], [17], [53].

- First, the signal $\{X(i), i = 1, \cdots, N\}$ is divided into $k$ non-overlapping sections

\footnote{Since the sample size $N > 7 \times 4096$, the rest of data is thrown away.}
of data of size \( n \), \( \{x_h(m) = X((h - 1)n + m), m = 1, \ldots, n, h = 1, \ldots, k\} \). If \( kn < N < (k + 1)n \), the extra data is thrown away.

- Second, we denote \( I_h(\omega_p) = |X_h(\omega_p)|^2 = \frac{1}{\sum_{m=1}^{n} v(m)^2} \left| \sum_{m=1}^{n} x_h(m)v(m)e^{-j\omega_pm} \right|^2 \) as the squared FFT of each windowed section, where \( v(m) \) is the window function and \( \omega_p = \frac{2\pi p}{n}, p = 0, \ldots, \frac{n}{2} \) (\( n \) even) is the angular frequency. If we assume that \( \{x_h(m)\}_{m=1}^{n} \) is a sample realization of a regular Gaussian WSS random process with zero mean and purely continuous spectrum \( f(\omega) \), \( I_h(\omega_p)/\frac{f(\omega_p)}{2} \) has a \( \chi^2 \) distribution with degree of freedom 2.

- Third, the averaged \( I_h(\omega_p) \) is the estimated PSD(n), which is expressed as \( \hat{S}(\omega_p) = \frac{1}{k} \sum_{h=1}^{k} I_h(\omega_p) \). If we also assume that the sections of \( \{x_h(m), h = 1, \ldots, k\} \) are uncorrelated, then \( \hat{S}(\omega_p)/\frac{f(\omega_p)}{2k} \) will be a \( \chi^2 \) distribution with degree of freedom 2k.

- Last, the \( (1 - \alpha)100\% \) C.I. of the \( f(\omega_p) \) will be \( \left[ \frac{\hat{S}(\omega_p)}{\chi^2_{2k,1-\alpha/2}}, \frac{2k\hat{S}(\omega_p)}{\chi^2_{2k,\alpha/2}} \right] \), where \( \chi^2_{2k,\alpha/2} \) denotes the value, such that the probability of a \( \chi^2_{2k} \) r.v. smaller than this value is \( \alpha/2 \), i.e. \( P(\chi^2_{2k} < \chi^2_{2k,\alpha/2}) = \alpha/2 \).

The PSD(4096) estimate in figure 4.7 exhibits a mixture of broadband continuous spectrum with very strong narrow peaks centered at the integral times of the fundamental frequency 192.9Hz, which is the blade passing frequency (BPF). The ratio between the fan speed to the engine speed is tested to be about 1.89. The fan has 6 blades, assuming constant engine speed, the BPF \( f_b \) would be:

\[
f_b = \frac{\text{enginespeed}(\text{rpm}) \times 1.89 \times 6}{60} = \frac{\text{fanspeed}(\text{rpm}) \times 6}{60}
\]

(4.1)

Since the fan speed of mock16000.dat is 1930 rpm, the BPF \( f_b \) would be approximately 193 Hz. The fan has 6 blades, then the shaft frequency is approximately 32 Hz. From the discussion in section 4.3.1, we saw that the shedding frequency associated with a radiator tube is about 27.5 Hz, which is close to the shaft frequency. There is also a
strong peak at this shedding frequency 27.5Hz. To show the p.s.d. around this frequency more clearly, part of figure 4.7 with a frequency range of 0–300Hz is shown in figure 4.8.

![Figure 4.8](image)

Figure 4.8 The 0–300 Hz estimated PSD(4096) and its 95% C.I. of the measured sound pressure process $x(t)$ in mock16000.dat

However, there are a number of uncertain issues associated with figure 4.7. The EPF might not necessarily fall on the bin frequency. The engine speed may also be wandering around a nominal speed. In addition, harmonics do not wander independent of one another. If the shaft speed varies an amount of $\Delta f$Hz, then the BPF will vary by $6\Delta f$Hz and the $m^{th}$ harmonic of the BPF varies by $6m\Delta f$Hz.

Through the eyes of a regular researcher who is not well trained in statistics or dsp, those narrow peaks appear to be tones and the height of the peaks seems to be significant. From our discussion in Chapter 2, the narrow peaks are not necessarily tones. If they are, then their magnitudes are strongly dependent on $T$. Recall, the information in a p.s.d. is the area, not the magnitude. They could also be narrow resonance peak and it
is the area under the peak that indicates the power on the narrow frequency band. Even
the C.I. obtained from Matlab is based on the assumption of uncorrelated subsections of
regular Gaussian WSS r.p. with purely continuous spectrum, which requires a transient
autocorrelation of the process if the window length is large. The autocorrelation function
of a process containing line spectrum will have a periodic component which will never
decay to 0 with increasing number of lags. In the case of mixed spectrum, the actual C.I.
of the mean of the p.s.d. estimator at the tone's frequency can be properly estimated
with the formula (2.35) instead of (2.34).

4.5 Octave Band Analysis of Sound Pressure Process with and
without Mock Engine

In this section, a standard investigation of the general structure of the noise is con­
tinued using A-weighted octave and 1/3 octave analysis. The reason is that this type of
analysis is commonly used in attempting to identify the regions that are most important
in relation to noise, in the audible sense.

The way the octave band level is viewed in this section is what a typical engineer
will use. Results will be drawn from such non-statistical view of point. In next section,
a statistical view of point will be used to analyze the process. Then the comparison of
results from statistical and non-statistical point of view will be made, which will show
how a statistical analysis of the measured random process (r.p.) will value.

4.5.1 Using Bandpass Filter to Obtain the Octave Band Level

There are two ways to estimate the octave band level. Recall the Wiener-Khinchine
relation.

\[
R_x(0) = E[X^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X(j\omega) \, d\omega
\]  

(4.2)
where $R_x(0)$ is the autocorrelation of the r.p. $X(t)$ and $S_X(j\omega)$ is its p.s.d. at angular frequency $\omega$.

One method is to use an octave band pass filter to filter the process into each octave band, and then to estimate the autocorrelation $R_x(0)$ of each filtered process. This method estimates the level in the time domain on the LHS of (4.2). The sound pressure process is measured in voltage. The voltage signal is an analog of the true sound pressure signal. There is a linear relationship between them. Hence, the voltage signal is directly used. The octave band level of the voltage signal in decibel is 10 times of the logarithm to the base 10 of the estimated autocorrelation of the filtered voltage signal. Then a 60dB shift is added to express the level of the original sound pressure signal.

A second method is to directly integrate the estimated p.s.d. of the r.p. in each octave band. This case involves the choice of window size $n$. However, the p.s.d. of a mixed r.p. at the tone's frequency is proportional to $n$, for large $n$. To avoid this potential problem, the first method is used.

A standard Butterworth bandpass filter is used to design each octave band pass filter. However at low frequency band where the bandwidth (BW) is very narrow relative to the Nyquist rate, the resulting bandpass filter will be away from the designed one.

When a filter is used, the FRF that relates a input sinusoid to the filter output sinusoid is only valid in steady state. But if the length of the filter impulse response is very long relative to that of the data, the steady state will never be achieved. When the normalized passband width is very close to 0, the length of the impulse response of the filter will be very large, since it is inversely proportional to the normalized bandwidth. The normalized BW is the actual BW divided by half the sample rate. The filtered process will be dominated by the impulse response of filter. This will happen to low frequency octave bands since their BW is very narrow relative to sample rate. To avoid this problem, the original process is down sampled by proper factors such that the normalized passband width is not small.
4.5.2 Observation from Octave Band and 1/3 Octave Band Analysis of the Measurement with Mock Engine

The estimated A-frequency weighed octave and 1/3 octave band sound pressure level of the signal $x(t)$ in mock16000.dat is shown in figure 4.9 and 4.10 respectively.

![Graph showing A-weighted octave band sound pressure level of mock16000.dat, with the 60dB shift from voltage level to sound pressure level.](image)

Figure 4.9 The A-weighted octave band sound pressure level of mock16000.dat, with the 60dB shift from voltage level to sound pressure level.

A-frequency weighting (refer to [54]) is applied in accordance of human ear response to sound pressure level. It gives much lower weights at low frequencies where human ears are quite 'deaf'.

From a non-statistical view, figure 4.9 suggests that the majority of the A-weighted sound pressure power is contained in 1000Hz, 500Hz, and 250Hz bands. These three bands contains about 36.4%, 30.5%, and 15.8% of the total A-weighted power of the sound process in mock16000.dat.

Figure 4.10 provides more detailed information related to the three dominant bands.
Figure 4.10 The A-weighted 1/3 octave band sound pressure level of mock16000.dat, with the 60dB shift from voltage level to sound pressure level.

The 1/3 octave bands containing the majority of the A-weighted power are centered on 200Hz, 400Hz, 630Hz, 800Hz and 1000Hz. These five bands contain approximately 8.9%, 11.2%, 14.6%, 14.6% and 15.2% of the total A-weighted power.

Table 4.3 summarizes the most important octave and 1/3 octave bands of the sound pressure process in mock16000.dat. The bands are given in the order of power contribution. The power contribution of each band in percentage of the total power is also shown in the 4th column in table 4.3. The potential harmonic frequencies (HFs) of the BPF contained in each band are also included in the table. The fact that the most important bands contain all the first five HFs of BPF suggests that the high power of those bands might be due to the power of possible harmonics of BPF. To answer this question, a comparison of the octave band level and the power of possible harmonics is made in the next subsection.
Table 4.3  The most important octave and 1/3 octave bands as well as the contained HFs of the BPF of the process in mock16000.dat

<table>
<thead>
<tr>
<th>Octave band order of power contribution</th>
<th>Center frequency (Hz)</th>
<th>Upper and lower limit (Hz)</th>
<th>Power contribution (%) to the total power</th>
<th>Contained possible HFs of BPF (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1000</td>
<td>(707,1414)</td>
<td>36.4%</td>
<td>771.7, 964.7, 1157.3, 1351</td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>(353,707)</td>
<td>30.5%</td>
<td>385.6, 578.6</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>(176, 353)</td>
<td>15.75%</td>
<td>193</td>
</tr>
<tr>
<td>1/3 octave band order of power contrib-</td>
<td>Center frequency</td>
<td>Upper and lower limit</td>
<td>Power contribution (%) to the total power</td>
<td>Contained possible HF of BPF</td>
</tr>
<tr>
<td>ution</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1000</td>
<td>(880, 1130)</td>
<td>15.2%</td>
<td>964.7, 1157.3</td>
</tr>
<tr>
<td>2</td>
<td>630</td>
<td>(565, 707)</td>
<td>14.6%</td>
<td>578.6</td>
</tr>
<tr>
<td>2</td>
<td>800</td>
<td>(707, 880)</td>
<td>14.6%</td>
<td>771.7</td>
</tr>
<tr>
<td>4</td>
<td>400</td>
<td>(353, 440)</td>
<td>11.2%</td>
<td>385.6</td>
</tr>
<tr>
<td>5</td>
<td>200</td>
<td>(176, 225)</td>
<td>8.9%</td>
<td>193</td>
</tr>
</tbody>
</table>

4.5.3 Comparison of the Octave and 1/3 Octave Band Level with the Power at the Harmonic Frequencies of BPF

Figure 4.11 shows the comparison of estimated unweighted power of each octave and 1/3 octave band with that of the potential harmonics of BPF contained in that band of the voltage signal.

The y-axis in Figure 4.11 is shown in linear scale instead of dB for the visualization of percentage. Since the important bands are already identified, unweighted power of each band is given for the convenience of comparison. Figure 4.11 shows, approximately, how much of the energy in the octave band is contributed by the potential harmonics.

The power of the first ten potential harmonics shown in vertical lines with stars, is estimated, by summing the area under the estimated PSD(4096) of a narrow frequency band centered on the bin frequency closest to the HFs of BPF. The band width is $2\Delta f$Hz.
Figure 4.11 Comparison of estimated unweighted power of each octave band and 1/3 octave band with that of the potential harmonics of BPF shown in original magnitude

It has to be mentioned that this estimation very rough.

The percentages of the power of potential harmonics to that of the corresponding octave bands and 1/3 octave bands could be estimated from figure 4.11 and are listed in table 4.4.

Table 4.4 The percentage of the power of tone to corresponding octave and 1/3 octave level of $x(t)$ in mock16000.dat

<table>
<thead>
<tr>
<th>No.</th>
<th>Harmonic frequency (Hz)</th>
<th>Center frequency of octave band (Hz)</th>
<th>Percentage of octave band level (%)</th>
<th>Center frequency of 1/3 octave band (Hz)</th>
<th>Percentage of 1/3 octave band level (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>193</td>
<td>250</td>
<td>59.5</td>
<td>200</td>
<td>72.4</td>
</tr>
<tr>
<td>2</td>
<td>385.6</td>
<td>500</td>
<td>41.2</td>
<td>400</td>
<td>78.9</td>
</tr>
<tr>
<td>3</td>
<td>578.6</td>
<td>500</td>
<td>14.7</td>
<td>630</td>
<td>42.1</td>
</tr>
<tr>
<td>4</td>
<td>771.7</td>
<td>1000</td>
<td>10.8</td>
<td>800</td>
<td>23.8</td>
</tr>
<tr>
<td>5</td>
<td>964.7</td>
<td>1000</td>
<td>9.23</td>
<td>1000</td>
<td>23.7</td>
</tr>
<tr>
<td>6</td>
<td>1157.3</td>
<td>1000</td>
<td>1.42</td>
<td>1250</td>
<td>8.86</td>
</tr>
</tbody>
</table>
The potential 8\textsuperscript{th}, 9\textsuperscript{th} and 10\textsuperscript{th} harmonics are in the octave band on 2000Hz and contain 8.4\% of overall band power. We offer the following explanation of the power contribution of the tone in percentage in table 4.4. The r.p. $x(t)$ could be expressed as a sum of its continuous spectrum component $x_c(t)$ and discrete one $x_d(t)$, i.e. $x(t) = x_c(t) + x_d(t)$. The power of the potential 1\textsuperscript{st} harmonic of BPF contains approximately 59.5\% of that of the octave band centered on 250Hz of $x(t)$. This means if the tone could be removed, the A-weighted level (dB(A)) of the 250Hz octave band will drop by approximately -3.9dB to 40.9dB. This would make this band less important since the level of the band centered on 4000HZ is about 40.9dB.

Similar calculation were made pertaining to the percentage of the power of other potential harmonics in other bands. The dB(A) of the octave bands centered on 500Hz, 1000Hz, 2000Hz of the continuous spectrum component $x_c(t)$ would drop by approximately -3.54dB, -1.11dB and -0.36dB to 44.1dB, 47.3dB, 43.4dB respectively. However, they would still be the most important three octave bands.

The first five potential harmonics add 5.6dB, 6.8dB, 2.4dB, 1.2dB and 1.2dB to the dB(A) of 1/3 octave band of $x_c(t)$, respectively. From figure 4.10, the dB(A) of 1/3 octave band centered at 200Hz, 400Hz, 630Hz, 800Hz and 1000Hz of $x_c(t)$ would then be 36.4dB, 36.3dB, 41.7dB, 42.7dB and 42.8dB. Consequently the most important 1/3 octave bands will be those centered on 630Hz, 800Hz and 1000Hz.

Finally, we remark that this analysis presumed that the estimated octave and 1/3 octave band levels were accurate. But to proceed with design changes, a decision made without a measure of accuracy could be a costly mistake.

4.5.4 Comparison of Octave Band and 1/3 Octave Band Analysis of the Measurement with and without Mock Engine

There are two operating conditions in the experiment, namely, with and without mock engine block installed. The above analysis was with mock engine block. Figure
4.12 and 4.13 compare the A-weighted octave and 1/3 octave band level of the two conditions.

Figure 4.12 Comparison of A-weighted octave band level of $x(t)$ in mock16000.dat and nomock16000.dat

Table 4.5 summarizes the comparison results of the sound pressure level with and without mock engine from figure 4.12 and 4.13. It shows which octave or 1/3 octave bands have equal, or higher levels under one condition than that under the other. For instance, the second row in table 4.5 means the dB(A) of octave bands centered on 31.5, 63, 500Hz with and without mock engine are equal.
Figure 4.13 Comparison of A-weighted 1/3 octave band level of $x(t)$ in mock16000.dat and nomock16000.dat

Table 4.5 Comparison of the A-weighted octave and 1/3 octave band and overall sound pressure level with and without mock engine

<table>
<thead>
<tr>
<th>Comparison results of octave band</th>
<th>Center frequency in Hz (dB difference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal</td>
<td>31.5, 63, 500</td>
</tr>
<tr>
<td>Mock higher</td>
<td>125 (2.5dB)</td>
</tr>
<tr>
<td>Nomock higher</td>
<td>250, 1000, 2000, 4000 (all 2.5dB)</td>
</tr>
<tr>
<td>Comparison results of 1/3 octave band</td>
<td>Center frequency in Hz (dB difference)</td>
</tr>
<tr>
<td>Equal</td>
<td>50, 63, 80, 250, 400</td>
</tr>
<tr>
<td>Mock higher</td>
<td>125 (4.5dB), 630 (2.5dB), 800 (1dB)</td>
</tr>
<tr>
<td>Nomock higher</td>
<td>200, 315, 500, 1000 (all 2.5dB)</td>
</tr>
<tr>
<td>Comparison result of overall level</td>
<td>dB difference</td>
</tr>
<tr>
<td>Nomock higher</td>
<td>1.7dB</td>
</tr>
</tbody>
</table>
The above is from a non-statistical view, a typical engineer would conclude that mock engine decreases the overall sound pressure level at the measured spot by 1.7dB.

However one has to note that all the above octave band levels are just estimates, which have mean and variance. Without looking at the C.I. of an estimator, one should not conclude whether the difference of two estimates are significant or not. Therefore the above comparison results might be misleading since the C.I. of the estimate is not in consideration.

To gain the confidence interval information of the octave band level, the structure of the process has to be firstly understood, i.e. whether the process is regular stationary process with only continuous spectrum or mixed r.p. with both continuous and discrete ones has to be firstly decided. The variance of the autocorrelation estimator of a mixed random process could be calculated according to (2.24) and Results 3 in Chapter 2. That variance expression is a frequency domain one that involves the decomposition of the continuous spectrum from discrete one. Without the knowledge of the separated spectrum, it is impossible to get valid estimation of the variance of the autocorrelation information. The detailed spectrum structure of the process in those important octave or 1/3 octave bands is then investigated.

In the next section, we offer the results of tone detection at each HFs of the BPF in order to perform statistical analysis on the octave band level.

4.6 Potential Tones and Estimated Tone Power of the Sound Pressure Process with and without Mock Engine

The tone detection is a very important procedure to perform prior to statistical analysis of each octave band. The first 10 HFs of the BPF appear to have strong harmonics. Tone detection is performed on these using the convergence property of families of p.s.d. estimators discussed in Chapter 2. Before the detailed detection processes is presented,
a summary on whether there is a tone and the tone power estimate at each of the first 10 HFs of BPF is listed in table 4.6. The upper limit of the tone power in dB from the MV spectra has the meaning of $10 \log_{10} \frac{A^2}{2}$. For one-sided spectrum, the power of the tone is $\frac{A^2}{2}$ in figure 4.11. $A$ is the amplitude of the tone.

Table 4.6 Comparison of the estimated power at HFs of BPF of the sound pressure process with and without mock engine

<table>
<thead>
<tr>
<th>Harmonic number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (Hz)</td>
<td>192.7</td>
<td>385.6</td>
<td>578.6</td>
<td>771.7</td>
<td>963.9</td>
<td>1157</td>
<td>1371</td>
<td>1543</td>
<td>1736</td>
<td>1930</td>
</tr>
<tr>
<td>Mock, tone?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>M</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Mock, power (dB)</td>
<td>-11.96</td>
<td>-15.5</td>
<td>-19</td>
<td>-24.5</td>
<td>-24.3</td>
<td>N</td>
<td>-31.9</td>
<td>-34.5</td>
<td>-33.7</td>
<td>-40.7</td>
</tr>
<tr>
<td>No mock, tone?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>No mock, power (dB)</td>
<td>-7.94</td>
<td>-16.0</td>
<td>-22.8</td>
<td>-26.9</td>
<td>-14.8</td>
<td>-25.8</td>
<td>-34</td>
<td>-33.2</td>
<td>-29.2</td>
<td>N</td>
</tr>
</tbody>
</table>

In the table, "Y" means there is tone; "N" means no tone; "M" means possible tone; "Power" actually is the upper limit of the power. The tonal power without mock engine at the 1st and 5th HFs of BPF is substantially higher than that with mock engine. The difference is 4dB and 10dB respectively. Only the tones at 578.6Hz and 1371Hz have higher power with mock engine than that without it. There is no tone on the 6th HF, 1156.6Hz of the voltage signal in mock16000.dat, but the tone on 1156.6Hz without mock engine exists and has -25.8dB upper limit of power. In either case, the contribution of this component is negligible. The 4th harmonic with mock engine may or may not be a tone. The AR and MV spectra of the heterodyned process of mock16000.dat with analysis BW of (0, 800)Hz and those of (707, 880)Hz show inconsistency of the convergence property at the 4th HFs of BPF, 771.7Hz. A possible reason could be that, the tone's
frequency is actually wandering between several $\Delta f_s$. By downsampling, the tone frequency variability in the heterodyned process of $(707, 880)\text{Hz}$ is also decreased, resulting a strong convergence behavior to $-24.5\text{dB}$. Hence the $4^{\text{th}}$ harmonic is an interesting one, and need more information to investigate.

The detailed tone detection procedure is included in appendix for interested reader such that the reader will not be distracted from the statistical analysis of the octave band levels of the two conditions. It has to be noted, that the tone power in table 4.6 is estimated from MV spectra of each heterodyned process of the $1/3$ octave band. In the appendix, the AR and MV spectrum of heterodyned process with broader frequency range such as $(0, 800)\text{Hz}$ and $(800, 1600)\text{Hz}$ is shown for illustration, which may provide different upper limit for the tone power.

In figure 4.11, the one-sided tonal power is estimated from the integral of the p.s.d. estimator, while table 4.6 provides two-sided tonal power from convergence of MV spectra. For comparison, the one-sided tonal power ($\frac{A^2}{2}$ in original magnitude) of the noise with mock engine obtained from the two methods are listed in table 4.7.

Table 4.7  Comparison of the estimated one-sided tonal power at HFs of BPF of the noise with mock engine from MV spectra and that from p.s.d. estimator integral.

<table>
<thead>
<tr>
<th>Harmonic number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (Hz)</td>
<td>192.7</td>
<td>385.6</td>
<td>578.6</td>
<td>771.7</td>
<td>963.9</td>
<td>1157</td>
<td>1371</td>
<td>1543</td>
<td>1736</td>
</tr>
<tr>
<td>Power (MV)</td>
<td>0.127</td>
<td>0.0564</td>
<td>0.0252</td>
<td>0.0071</td>
<td>0.0074</td>
<td>N</td>
<td>0.0013</td>
<td>7e-4</td>
<td>9e-4</td>
</tr>
<tr>
<td>Power (p.s.d.)</td>
<td>0.144</td>
<td>0.0483</td>
<td>0.0173</td>
<td>0.0075</td>
<td>0.0065</td>
<td>0.0008</td>
<td>0.0011</td>
<td>7e-4</td>
<td>8e-4</td>
</tr>
</tbody>
</table>

Table 4.7 shows that the tonal power estimated from MV spectra is slightly higher than that from the p.s.d. estimator integral except at the $1^{\text{st}}$ and $4^{\text{th}}$ HFs of the BPF. This is because the MV method give the upper limit of the tonal power, while spectral
leakages of a tone goes beyond 2 bin frequencies in a p.s.d. estimator. However, since the tonal power is high at the BPF, the p.s.d. estimator variance is high which result in a high estimated tonal power. The 4th harmonic may or may not be a tone.

4.7 Statistical Analysis of Octave Band and 1/3 Octave Band Level

The C.I. of each 1/3 octave band level is estimated by firstly heterodyning the data into each octave band. Heterodyning is a procedure including first bandpass filtering the process, then multiplying it with a cosine function whose frequency \( f_0 \) is the lower limit of the band, such that its frequency is shifted by \( -f_0 \), finally low pass filtering and down sampling the frequency shifted process so that the new frequency range is in each band. The heterodyned process is denoted as \( x_h(t) \).

If a tone’s frequency in the original process is very close to the upper or lower limit of a 1/3 octave band, then aliasing of that tone may occur in its neighboring band. Special care should be given to this case. The 1/3 octave band of (1760, 2250)Hz is such a case. A harmonic frequency of 1737.2 Hz is very close to the lower limit of this band. Analysis of the influence of such a tone requires consideration of both bands. In this work, a simple solution to this is to shift the boundary of the 1/3 octave band, such that the tone’s frequency is farther away from the bands. For antialiasing purpose, the band of (1760, 2250)Hz is thus shifted to (1800, 2250)Hz.

Once \( x_h(t) \) of each band is obtained, usually, it is just a single tone plus broad band noise or simply broad band noise. With such a simple spectrum structure, the power of the tone can be estimated and both the mean and the variance of the autocorrelation estimator of \( x_h(t) \) are estimated using (2.24) in Chapter 2. In (2.24), the tone power \( A^2 \) is from table 4.6 and the true noise spectrum is replaced with the estimated ones. The C.I. of the p.s.d. estimator is then estimated using (2.32) at non-tone's frequency
and (2.35) at tone’s frequency. After heterodyning, any tone whose frequency is away
from the center of the analysis BW is now much closer to the center of the new analysis
BW. The tonal nature is easier to be detected. However, in the case of a wandering tone
frequency, heterodyning may also decrease the tone frequency variability.

Figure 4.14 and 4.15 make a comparison on the standard deviation and 95% C.I.
of the p.s.d. estimator using the method in *Matlab* and my method (2.35) respectively
on the heterodyned sound pressure process in octave band of (176, 353)Hz with mock
engine. Only the frequency range of (176, 225) Hz is shown for clear view. This process
in (176, 353)Hz is a locally white noise plus a sinusoid at 193Hz.

From figure 4.14 and 4.15, we see that at frequencies away from the tone’s frequency,
the standard deviation and C.I. from the two methods are the same. While at tone’s
frequency, there is a notable difference between results from the two methods. The
standard deviation from my method, which considers the mixed spectral nature, is more
than 8dB smaller than that from *Matlab*, which assumes purely continuous spectrum.
This difference is also reflected in the C.I. in figure 4.14, which shows that both the
upper and lower limits of the 95% C.I. from my method are 8dB smaller than those
from *Matlab*. This difference will be even bigger if a larger window size is used. In fact,
*Matlab* overestimates the standard deviation of the p.s.d. estimator asymptotically by a
factor proportional to the window size.
Figure 4.14 Comparison of the standard deviation of p.s.d. estimator from Matlab and that from our method, (2.35), of the heterodyned process in mock16000.dat, with a frequency range (176, 225) Hz, the window size is $n = 256$, the sampling rate is 366Hz.

Figure 4.15 Comparison of the 95% C.I. of p.s.d. estimator from Matlab and that from our method, (2.35), of the heterodyned process in mock16000.dat, with a frequency range (176, 225) Hz.
Therefore when a mixed r.p. is involved, the C.I. of the p.s.d. estimator from Matlab is misleading because it is based on inappropriate assumption. Such misleading information may result in a mistake in military specifications for helicopter drive train or in evaluating the effect of a design change on machinery sound and vibration where the C.I. of the p.s.d. estimator played an important role while the r.p.es involved contain both continuous and discrete spectrum components. For example, the method on Matlab may fail to identify an actual significant effect of a design change on the noise because of the overestimation on the C.I..

Figure 4.16 shows the 2 - σ C.I. of A-weighted 1/3 octave band level of the sound pressure process in mock16000.dat and nomock16000.dat.
From the above C.I. of the 1/3 octave band level, a more statistically sound inference could be made. Table 4.8 summarizes a 1/3 octave level and overall level comparison results based on the $2 - \sigma$ C.I.

Table 4.8 Comparison of 1/3 octave and overall sound pressure level with and without mock engine from the C.I. information

<table>
<thead>
<tr>
<th>Comparison results of 1/3 octave band</th>
<th>Center frequency in Hz (C.I. separation in dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal (C.I. Overlap &gt; 0)</td>
<td>50, 63, 80, 125, 250, 400, 500, 2500</td>
</tr>
<tr>
<td>Mock higher</td>
<td>630(1dB), 800(0.05dB)</td>
</tr>
<tr>
<td>Nomock higher</td>
<td>200(1.5dB), 315(0dB), 1000(3dB), 1250(1.5dB), 1600(1.3dB)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparison result of A-weighted overall level</th>
<th>C.I. difference in dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nomock higher</td>
<td>1.1dB</td>
</tr>
</tbody>
</table>

That the $2 - \sigma$ C.I.s of the two conditions do not overlap means that the mean of 1/3 octave band level is significantly different at a 0.05 significance level. If they overlap, then for a significance level $\alpha = 0.05$, they are not significantly different.

With the C.I. interval, the level of the two conditions of the 1/3 octave band centered on 125Hz that was considered different from non-statistical view, is indeed not significantly different at a 0.05 significance level. The $2 - \sigma$ C.I.s of the overall A-weighted level are (52.62, 52.93)dB (Mock) and (54.00, 54.48)dB (No mock). The two overall levels are significantly different at 0.05 significance level, since the $2 - \sigma$ C.I.s do not overlap.

The 1/3 octave band centered on 800Hz is one of the most important bands. The difference of the 1/3 octave level is 1.24dB from figure 4.12, but the difference of the $2 - \sigma$ C.I. of that band level of mock and no mock condition, is 0.05dB, which is only slightly bigger than 0. This means the two 1/3 octave band level is statistically different at 0.05 significance level. However, if the difference between the C.I. of the two conditions is modestly bigger than or smaller than 0, it has to be treated carefully. A slightly bigger than 0 difference, means it is significant at $\alpha = 0.05$ level, but might not be significant.
for a significance level of $\alpha < 0.05$. Therefore, a test of hypothesis could be performed and the P-value could be calculated to get the significance level.

If we assume the mean of the 1/3 octave band level centered on 800Hz is $L_m$ for mock and $L_n$ for no mock condition, then $\hat{L}_m = 0.032$ and $\hat{L}_n = 0.024$. The estimated standard error of $\hat{L}_m$ and $\hat{L}_n$ is $Std_m = 0.0018$ and $Std_n = 0.0023$. As $\hat{L}_m$ and $\hat{L}_n$ are estimated from large sample of size 500 and 250 respectively, they could be assumed to be normal. A null hypothesis could be tested:

$$H_0 : L_m = L_n$$
$$H_a : L_m > L_n$$

We chose a one-sided alternative hypothesis because estimation showed that $\hat{L}_m$ is bigger than $\hat{L}_n$. Then if $H_0$ is true, the $Z$ statistic is $Z_0 = \frac{\hat{L}_m - \hat{L}_n}{\sqrt{Std_m^2 + Std_n^2}} = 2.72$. If $H_0$ is true, then $Z$ is a standard normal r.v. Then the P-value will be $P(Z > Z_0) = 0.0033$. Then the null hypothesis is rejected at a level of 0.004.

4.7.1 Analysis of 1/3 Octave Bands Having No Tones

A number of 1/3 octave bands, such as (225, 283)Hz and (283, 353)Hz, contain no tones. In that case, the process could be assumed to be GM process. The mean and variance of the GM process parameter can be estimated using the results from Chapter 3 and used to test for a possible change of the process due to physical change which results in a change in the BW parameter. A hypothesis test can be constructed on the parameters of the two conditions. The null hypothesis would be that the GM parameters with and without mock is the same. In contrast, the alternative two-sided hypothesis would be that they are not equal. A one-sided alternative hypothesis could also be chosen if whether one is bigger than the other is of interest. Such a parametric test is better than a non-parametric one since the estimation of a single scalar parameter is far more robust for a given amount of data than non-parametric estimation.
We will apply this test for the bands \( (440, 565) \text{Hz} \) and \( (283, 353) \text{Hz} \). \( x_h(t) \) in these two bands has purely continuous spectrum. We assume they are discrete-time AR(1) processes of the form (3.2) since the p.s.d. estimate of the \( x_h(t) \) strongly suggest so. Express the AR(1) parameter as \( \alpha_n \) (no mock) and \( \alpha_m \) (mock). Let \( \hat{\alpha}_n \) (no mock) and \( \hat{\alpha}_m \) (no mock) to be the estimated ones. As the sample rate is around 150 Hz, it is not a fast sampling case relative to the dynamics of the underlying continuous-time process. \( \alpha \) is closer to 0 rather than to 1 and the estimated \( \beta T \) are very big for all cases. So a standard bias and variance formula of \( \hat{\alpha} \) is used. The bias of \( \frac{3\alpha^2}{N} \) in (3.23) can be ignored for big \( N \) and small \( \alpha \). Then \( \hat{\alpha} \) will be approximately normal with mean \( \alpha \) and variance \( \left(1 - \frac{\alpha^2}{N T^2}\right), \) where \( N \) is the sample size of \( x_h(t) \). The mean and variance are estimated by replacing \( \alpha \) with \( \hat{\alpha} \). Results show that \( \hat{\alpha}_n \) is bigger than \( \hat{\alpha}_m \) for both bands. In order to see how significant this difference is, a one sided hypothesis is performed for the \( x_h(t) \) in the two bands:

\[
H_0 : \alpha_n = \alpha_m \quad \quad \quad H_a : \alpha_n > \alpha_m
\]

Table 4.9 shows the estimated AR(1) parameter \( \alpha \) and its standard deviation of \( x_h(t) \) in the two bands. Table 4.10 shows z-statistic and the P-value as well as the hypothesis test conclusion. From the two tables, it can be concluded that the AR(1) parameters \( \alpha \) of the process in the band \( (440, 565) \text{Hz} \) with and without mock are not significantly different, but the ones in the band \( (283, 353) \text{Hz} \) are significantly different at a significance level 0.05 under the two conditions.
Table 4.9  Estimated parameters and their standard errors of $x_h(t)$ in the bands (440,565) Hz and (283,353) Hz with and without mock engine.

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\hat{\alpha}$</th>
<th>Sample size $N$</th>
<th>Sample rate (Hz)</th>
<th>$\beta T$</th>
<th>$\beta \Delta$</th>
<th>Standard error($\hat{\alpha}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mock</td>
<td>0.0795</td>
<td>500</td>
<td>252</td>
<td>1266</td>
<td>2.53</td>
<td>0.0447</td>
</tr>
<tr>
<td>No mock</td>
<td>0.1469</td>
<td>250</td>
<td>252</td>
<td>480</td>
<td>1.98</td>
<td>0.0632</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\hat{\alpha}$</th>
<th>Sample size $N$</th>
<th>Sample rate (Hz)</th>
<th>$\beta T$</th>
<th>$\beta \Delta$</th>
<th>Standard error($\hat{\alpha}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mock</td>
<td>0.1445</td>
<td>286</td>
<td>144</td>
<td>553</td>
<td>1.935</td>
<td>0.0585</td>
</tr>
<tr>
<td>No mock</td>
<td>0.3155</td>
<td>143</td>
<td>144</td>
<td>165</td>
<td>1.154</td>
<td>0.0794</td>
</tr>
</tbody>
</table>

Table 4.10  Test of hypothesis of $x_h(t)$ in the bands (440,565) Hz and (283,353) Hz under the two conditions.

<table>
<thead>
<tr>
<th>Bands of $x_h(t)$ Hz</th>
<th>$\hat{\alpha}_n - \hat{\alpha}_m$</th>
<th>Standard error of $\hat{\alpha}_n - \hat{\alpha}_m$</th>
<th>$z = \frac{\hat{\alpha}_n - \hat{\alpha}_m}{\text{Std}(\hat{\alpha}_n - \hat{\alpha}_m)}$</th>
<th>$P_{value} = P(Z &gt; z)$</th>
<th>Conclusion at significance level 0.05</th>
</tr>
</thead>
<tbody>
<tr>
<td>(440,565)</td>
<td>0.0674</td>
<td>0.0775</td>
<td>0.8694</td>
<td>0.1922</td>
<td>$H_0$ not rejected</td>
</tr>
<tr>
<td>(283,353)</td>
<td>0.1710</td>
<td>0.0986</td>
<td>1.724</td>
<td>0.0425</td>
<td>$H_0$ rejected</td>
</tr>
</tbody>
</table>
4.8 Summary and Conclusion

The sound pressure process of an engine cooling system with and without mock engine were studied in this chapter. Our analysis procedure included

- Studying the general noise generation mechanism of a tractor engine cooling system;

- Performing a preliminary p.s.d. and C.I. analysis of the data using Matlab;

- A standard octave and 1/3 octave band analysis;

- Using families of p.s.d. estimators including AR and PER to reveal the spectrum structure and to detect tones at HF's of the BPF. The tonal power is estimated from the convergency property of a family of MV spectra. This could not be done with a single p.s.d. estimator.

- A statistical analysis on the 1/3 octave band level;

- Hypothesis test on the 1/3 octave band level of the two conditions in the bands where there is tone and parametric hypothesis test of the two conditions in the bands containing no tone.

One of this chapter's goal was a detailed characterization of the stochastic structure of the sound pressure process using results from previous chapters. The results for this purpose included:

- The sound pressure process is a mixed random process containing broadband spectrum as well as tones at most harmonics frequencies of the BPF. The tonal powers of the first 5 HF's of the BPF contain most of the line spectrum energy.

- The most important octave bands are those centered on 1000, 500 and 250 Hz, which contains 36.4%, 30.5% and 15.75% of total energy respectively;
The most important 1/3 octave bands are those centered on 1000, 630, 800, 400 and 200 Hz, which contains 15.2%, 14.6%, 14.6%, 11.2% and 8.9% of total energy respectively;

The first three harmonics of the sound pressure with mock engine installed contain 59.5%, 41.2% and 14.7% energy of the corresponding 1/3 octave bands on 200, 400, 630 Hz respectively.

A second goal of this chapter was to demonstrate the usefulness of our results from previous chapters including:

- C.I. estimation and comparison of the p.s.d. of a mixed random process from our formula in (2.35) and that from Matlab;
- C.I. estimation of 1/3 octave band levels of a mixed random process using (2.24);
- Hypothesis test and its implication for the mock engine influence of the bands with tone (Chapter 2) and the bands without tone (Chapter 3).

The results of this goal contained:

- At 193Hz, the 1st BPF where there is a tone, the estimated standard deviation and the lower and upper limit of 95% confidence interval of p.s.d. from Matlab are 8dB higher than those using our formula (2.35). This difference will be even more substantial if a bigger window size is used. The erroneous C.I. of p.s.d. in Matlab stems from the inappropriate assumption of a process without tone while it is actually there. And using C.I. from softwares such as Matlab will result in failure to find an actual significant effect of a design change on the noise component, which will in turn affect the design decision and is very costly. At frequencies where there is no tone, the Matlab C.I. is correct.
An example hypothesis test performed on the dB(A) of 1/3 octave band centered on 800Hz under the two conditions showed that the dB(A) with mock engine is higher than that without mock engine at a significance level 0.05. The calculated P-value is 0.0033.

In the bands where there is no tone, the heterodyned process was modeled as an AR(1) GM r.p.. The test of hypothesis on the AR(1) parameters showed that in the band of 500Hz, there is no significant effect, while in the band of 315Hz, there is a significant effect associated with the mock engine. This parametric test is more robust and reliable than a non-parametric one, such as p.s.d., since it only involves one parameter estimation while a p.s.d. estimator involves estimation of the spectrum on a large number of frequencies.

In a word, the key contribution of this chapter is that by comparing results from standard analysis and those from our proposed statistical analysis on the sound pressure process of a tractor cooling system, we show that the analysis of r.p. from mechanical systems should not be done without considering the mixed spectral nature of the process. If there are tones, then the Matlab C.I. of p.s.d. estimator at the tone’s frequency is faulty and may lead to costly wrong decision. The more appropriate and reliable procedure is to first understand the physics of the phenomenon, then determine whether there are tones at specific frequencies indicated by the physics and then perform spectral analysis and associated statistical tools such as C.I. and parametric hypothesis test corresponding to the spectral nature of the process. No software package on the current market is able to perform such an advanced analysis. The practitioners have to be aware of this and apply our method more carefully instead of only relying on the software package to conduct the analysis.

Further work could be done on tone tracking and depressing using extended Kalman filter. However since tone tracking is not part of the thesis, it is not included in this
4.9 Appendix: Detailed tone detection procedure

4.9.1 Using A Family of Averaged Periodograms

Since from the octave band analysis, it is shown that the octave band between 200Hz and 1400Hz is the dominating bands. Just for a view of how a family of averaged periodograms could be used to locate a possible tone.

The processes of mock16000.dat is down sampled to the frequency range of (0,2016)Hz by a factor of 4. The new data size is 8000. It is denoted by $X_4$. Section 4.4 already introduced averaged periodogram as a commonly used none parametric p.s.d. estimator. In this section, the averaged periodogram of window size $n$ will be denoted as $\text{APER}(n)$ to distinguish it from other p.s.d. estimators.

![Diagram](image.png)

Figure 4.17 Estimated $\text{APER}(512)$ and $\text{APER}(1024)$ of the decimated process $X_4$, in mock16000.dat, with a frequency range of (0, 2016) Hz.
APER(512) and APER(1024) is then shown in figure 4.17. \( n \) is window size or the correlation lag used. The magnitude of APER(n) at a tone’s frequency, is meaningless, since it is asymptotically \( O(n) \). At frequency away from tones, the magnitude of APER(n) is independent of window size, or \( O(1) \). Therefore, a family of periodogram with doubling window size could be used to detect tones. If a tone exists, then at that tone’s frequency, the PSD should have approximately 3dB increase per order doubling; at other frequencies, the estimated APER(n) should stay at the same level but with different variance. As the total length of data \( N \) is fixed, if the window size \( n \) is chosen to be bigger, the frequency resolution will be higher but less number of PER(n) could be used for averaging. As a result, the variance of APER(n) will be bigger. There is a trade off between frequency resolution and estimator variance. Equation (2.32) for non tone frequency and (2.34) (tone frequency) in chapter 2 provide the variance expression of APER(n).

The dB increase of APER(1024) from APER(512) at multiples of the fundamental frequency 193 Hz is shown in table 4.11. The dB increase might not be exactly 3dB since the tone frequency might lie on a bin frequency. We then turn to Autoregressive (AR) and Minimum variance (MV) spectrum.

<table>
<thead>
<tr>
<th>frequency(Hz)</th>
<th>193</th>
<th>385.6</th>
<th>578.6</th>
<th>771.7</th>
<th>964.7</th>
<th>1157</th>
<th>1350</th>
</tr>
</thead>
<tbody>
<tr>
<td>dB</td>
<td>4.62</td>
<td>3.415</td>
<td>2.952</td>
<td>2.47</td>
<td>3.264</td>
<td>2.15</td>
<td>2.538</td>
</tr>
</tbody>
</table>
4.9.2 Using A Family of Autoregressive (AR) and Minimum Variance Spectrum

Autoregressive (AR) spectrum is a parametric p.s.d. estimator of a WSS r.p. It has the advantage of higher frequency resolution using smaller number of lags, which could also be called order. In the case of a mixed r.p. containing both continuous and line spectrum, the family of AR spectra will exhibit different properties at the tone and non-one frequency. At frequencies away from tones, it will stay the same level regardless of order. While, at tone’s frequency, the family of AR spectra will be asymptotically $O(n^2)$ or increase by approximately 6dB per order doubling.

The Capon MV spectrum is another spectrum which could be used to character the spectrum of a mixed r.p. It is not a p.s.d., but is a power spectrum. The family of MV spectra asymptotically converges monotonically downward to the line spectrum of the tone, as the order $n \to \infty$.

The Capon MV spectrum was originally introduced in 1969 by Capon [55] as a solution to a constrained minimization problem. The convergency properties of the $n^{th}$ order Capon spectrum was developed recently in [27]. The asymptotic rate of convergency of Capon MV spectra for the case of sinusoid plus white noise was achieved by Sherman and Lou [7]. They found the MV(n) spectra will drop by 3dB per order doubling at all frequency except at the tones. At the tone’s frequency, it converges monotonically to the tone’s power, $\frac{\sigma^2}{4}$. So, MV spectrum convergency value at the lowest order provides an upper bound of the tone’s power $\frac{\sigma^2}{4}$.

Figure 4.18 and 4.19 shows a family of AR and MV spectrum of order 40, 80, 160, between the range 0-1500 Hz from $X_4$ of the process mock16000.dat measured with mock engine. The frequency resolution for these two figures are about 0.50Hz. The small arrows in figure 4.18 indicate the position of those peaks of order $p = 40$ and 80.
Figure 4.18  AR spectrum of decimated mock16000.dat, $X_4$, at order $p = 40, 80$ and $160$.

Figure 4.19  MV spectrum of decimated mock16000.dat, $X_4$, at order $p = 40, 80, 160$ and $320$. 
The increasing AR spectrum peaks occur at the same frequency location of the 1st, 2nd and 3rd HFs of the BPF. At around the 4th, 5th, and 7th HFs of BPF, the AR increasing peaks locations actually oscillating in a narrow frequency band for different order. The AR spectrum corresponds to the 6th HF of BPF does not even has a single increasing peak suggesting there is no tones at all. The AR spectrum at frequencies between those HFs of BPF stays at the same level which could be trusted that there is no tones in between.

Confidence interval is not available for Capon MV spectrum, however, the convergence property of the MV spectrum family could also help to detect tones. The MV(n) spectra in figure 4.19 shows strong convergence on the 3rd, somewhat convergence on the 1st and 2nd HFs, almost non-convergence on the 4th, 6th and 7th HF. Conclusion could not be drawn since the process has still a relatively broad frequency range and there are quite a few possible tones. The MV spectrum with a narrower frequency range is necessary. Finally, the MV spectrum of heterodyned process containing only a single tone should provide the best estimation.

Figure 4.20 shows the C.I. of the AR(p) spectra family at the first 4 HFs of order p= 40, 80, 160. The variance of AR(n) spectrum is estimated from equation (2.37) in Chapter 2. If there is a tone, then the C.I. should be separated at tone's frequency by approximately 6dB. The C.I. at the first three HFs of BPF has about 2dB to 6dB increase. The C.I. at the 4th HF of BPF actually overlap for order 40 and 80 and the peak occurs at different frequency location. This suggests the 4th HF of BPF is either not a pure tone, or has a too small power to be detected.

With the family of AR and MV spectra, the 4th and 6th harmonics might be either missing or have too small power. This could not be found simply from a PSD in figure 4.17. The reason of the possible missing 4th and 6th harmonic could either be some directivity pattern such that at the measurement location, it is very small or be some physical changes causing the cancellation of that harmonic. There is not enough infor-
Figure 4.20 Zoomed in figure of Confidence Interval of AR spectrum of decimated mock16000.dat, $X_t$, at order $p = 40, 80, 160$ at the potential HFs of BPF.

It is very likely that there are pure tones on the 1\textsuperscript{st}, 2\textsuperscript{nd}, and 3\textsuperscript{rd} HFs of BPF. To have higher frequency resolution, the process is further heterodyned into two processes $X_{s0}$ and $X_{s1}$ with the frequency range of (0,1000)Hz and (800, 1600)Hz respectively. The family of AR and MV spectra of order 40, 80 and 160 of $X_{s0}$ is shown in figure 4.21 and 4.22. The frequency resolution of these two figures is 0.25Hz. The MV spectra shows strong convergency to -11.96dB, -17.0dB and -20.4dB at 192.95Hz, 385.9Hz and 578.6Hz respectively. The 4\textsuperscript{th} harmonic still appears to be missing.
Figure 4.21  AR spectrum of decimated mock16000.dat, $X_{80}$ with a frequency range of $(0, 1000)$Hz, at order $p = 40, 80, 160$.

Figure 4.22  MV spectrum of decimated mock16000.dat, $X_{80}$ with a frequency range of $(0, 1000)$Hz, at order $p = 40, 80, 160$. 
Figure 4.23 and 4.24 shows the family of AR and MV spectra for the heterodyned process $X_{s_1}$.

These two figures show that the tone at the 5th HF of BPF, 964.4Hz, converges to -27.1dB. The 6th Harmonic of BPF is missing apparently since the MV spectra do not converge to any line spectrum at around 1156Hz. There is no clear convergency in the MV spectra at the 7th and 8th HFs of BPF either. The AR(160) has such high variability that the increasing peaks at the possible tone is not obvious.

To see more clearly on the 4th harmonic, the process is heterodyned to the frequency range (707, 780)Hz. The AR and MV spectrum of this heterodyned process denoted by $X_{16}$ is shown in figure 4.25 and 4.26. This shows that the 4th harmonic seems to converges to -26.1dB, but the peak of AR(128) is even smaller than that of AR(64). So
Figure 4.24 MV spectrum of hetordyned mock16000.dat, $X_{81}$ with a frequency range of (800, 1600)Hz, at order $p = 40, 80, 160$.

it could be said there is possible tone on the 4th HF. The sample size of $X_{16}$ is 728. The MV(64) is 24.5dB. As order 128 is relatively big to the sample size 728. The MV(128) is considered to have higher variance.
Figure 4.25  AR spectrum of hetordyned mock16000.dat, $X_{16}$ with a frequency range of (707, 880)Hz, at order $p = 16, 32, 64, 128$

Figure 4.26  MV spectrum of hetordyned mock16000.dat, $X_{16}$ with a frequency range of (707, 880)Hz, at order $p = 16, 32, 64, 128$
4.9.3 Detect Tones and Estimate Amplitudes in Nomock16000.dat

Similar detection procedure is applied to nomock16000.dat. The APER(512) and APER(1024) of the decimated process, $X_4$ of nomock16000.dat is shown in figure 4.27. The AR spectrum family of order 40, 80 and 160 is shown in figure 4.28. Figure 4.29 shows the MV spectrum family of order 40, 80, 160 and 320.

![Figure 4.27](image)

**Figure 4.27** APER(512) and APER(1024) of decimated nomock16000.dat, $X_4$ with a frequency range of (0, 2016)Hz

The AR and MV spectra between the range of 0 and 1500Hz do not provide enough frequency resolution. The process in nomock16000.dat is heterodyned to two processes $X_{80}$ and $X_{81}$ with frequency range of (0, 1000)Hz and (800, 1600)Hz.

Figure 4.30 (0-800Hz) and 4.32 (800-1600Hz) show the AR spectra of order 40, 80, and 160 of the heterodyned process of nomock16000.dat. Figure 4.31 (0-800Hz) and 4.33 (0-800Hz) show the corresponding MV spectra of the same order. The reason that the highest order only goes to 160 is because the new data sample size is only 4000.
The comparison results from the family of APER, AR and MV p.s.d. estimator of the sound pressure process with and without mock engine is already shown in table 4.6. There is no need to discuss each figures in detail.

Overall, by using the families of APER, AR and MV spectra, it could be found that the presence of mock engine may increase the power spectrum at the 3rd HFs of BPF, 578.6Hz, while decrease the power spectrum at the 1st, 2nd and 5th HFs of BPF, in particular, the power of the 5th harmonics drop by 10dB because of mock engine. The mock engine also appears to smear the 4th and 6th harmonics at 771.7Hz and 1156.6Hz.
Figure 4.29 MV spectrum of decimated nomock16000.dat, $X_4$ with a frequency range of $(0, 2016)$ Hz, at order $p = 40, 80, 160, 320$

Figure 4.30 AR spectrum of decimated nomock16000.dat, $X_{80}$ with a frequency range of $(0, 1000)$ Hz, at order $p = 40, 80, 160$
Figure 4.31  MV spectrum of decimated nomock16000.dat, $X_{80}$ with a frequency range of (0,1000)Hz, at order $p = 40, 80, 160$

Figure 4.32  AR spectrum of heterodyned nomock16000.dat, $X_{81}$ with a frequency range of (800,1600)Hz, at order $p = 40, 80, 160$
Figure 4.33 MV spectrum of hetordyned nomock16000.dat, $X_{81}$ with a frequency range of (800, 1600)Hz, at order $p = 40, 80, 160$
5.1 Summary and Conclusion

The purpose of this dissertation was to contribute to the knowledge base of statistical signal processing, in relation to mechanical systems. In general, statistical signal processing has its "home" in electrical engineering. It combines statistics, random process theory, systems theory and digital signal processing; all of which are addressed in both undergraduate and graduate level electrical engineering curricula. Consequently, even though courses and research are conducted in other areas of science and engineering, the majority of the basic research in this area is conducted by electrical engineers. The interest of this area on the part of mechanical engineers has, and continues to be predominantly application in nature. Hence, these researchers usually take theory that has been developed by electrical engineers and apply it to mechanical systems. The popularity of this application research is evidenced in the journal Mechanical Systems and Signal Processing, which is devoted exclusively to mechanical systems.

In spite of the popularity of the area of statistical signal processing in relation to mechanical systems, because the tools are most often derived from research conducted by electrical engineers, many of them are not developed with mechanical systems in mind. But random processes associated with mechanical systems are often quite different, and often more complex than processes associated with application areas of electrical engineering. They are, nonetheless, often applied to mechanical systems; and with little or no regard as to the limitations of their applicability. This is not intended as a criticism,
for the fact is, it is no easy task for a researcher trained in mechanical engineering, or any single discipline (including disciplines within electrical engineering) to develop a firm enough grasp of these various areas in order to integrate them in addressing a real world application.

My Ph.D. studies have included development of a firm basis in these areas, as evidenced by my course work and publications to date. So it was natural, and indeed, it was my intent to pursue a dissertation that would take advantage of this basis, and serve to contribute to the knowledge base, in relation to mechanical systems. Rather than focusing on a single application, I chose to investigate a number of more generic problems; problems that have common and direct application in relation to mechanical systems.

One of the problems, addressed in Chapter 2, was that of characterizing the statistics of popular spectral estimators, in relation to mixed random processes. Because most mechanical systems exhibit behavior that is periodic in nature (e.g. motors, pumps, turbo-machinery, lathes, etc.), processes related to them, such as flow, pressure, vibration, and sound will invariably have a mixed spectrum; that is, spectral energy that is a mixture of a continuous spectral density and a discrete line structure. In the case of sound radiation from a compressor, the continuous spectrum might be associated with flow turbulence or rubbing between two components, since these phenomena are more random in nature. The line structure might be related to the lower harmonics associated with periodic impacts between the exhaust port valve and its seat. In fact, if one also considers the time-varying load exerted on components, and impulsive expulsions of gas when the valve opens, it is safe to say that such complex types of random processes are rarely found in electrical engineering applications.

Statistics of spectral estimators for regular random processes are well known. In fact, Matlab’s *psd.m* function for estimation of the confidence interval associated with a power spectral density (p.s.d.) estimate is based on the assumption of a regular process.
These confidence intervals play a central role in areas such as military specifications for helicopter drive train vibration, and in evaluating the effect of design changes on machinery sound and vibration.

The primary contribution of this chapter relative to present literature is a detailed statistical analysis of the statistics of certain families of spectral estimators in particular to the case of a mixed random process. Our key results included:

- The arithmetical mean and variability of families of theoretical spectra were derived. The limiting behavior of these families of theoretical spectra was also derived for families of three commonly used spectral estimators as convergence rate. This could be valuable for the case where the amount of data far exceeds than the lag of autocorrelation used.

- The marginal statistics of the three families of spectral estimators were investigated in a more rigorous manner for the case of a mixed random process.

- More importantly, we introduced the new idea of using the average of a family of spectral estimators as a method to use the family. The statistical properties of averages of families of p.s.d. estimators were obtained from simulation because of the more mathematical difficulty in deriving the joint distribution of p.s.d. estimators of different orders. It was shown that, by averaging a non-convergence FT(n) spectrum estimators, a convergence PER(n) estimator is obtained. The averaged spectral estimators also have notably reduced statistical variability than any of the corresponding single one.

- One of our key results in Chapter 2 allowed us to estimate confidence intervals that take into account the mixed nature of the sound radiation associated with the engine cooling fan experiment in Chapter 4. It was shown that there is a marked difference between the estimates of the p.s.d. confidence intervals using
Matlab's `psd.m` function versus our results of Chapter 2. The difference is even more considerable for bigger window size. This will allow one to supersede Matlab's method to get improved estimates.

The second problem, addressed in Chapter 3, concerned the estimation of the bandwidth parameter, $\beta$, for a Gauss-Markov (GM) process. A GM process model is a more realistic model for all types of broadband processes than the white noise model is. In relation to mechanical systems, common examples of processes that are assumed to be white include sound and vibration associated with sliding surfaces (such as a piston and cylinder), with highly turbulent flow, or with sound radiation from a machine in an octave band that contains no strong narrow band or harmonic content. Even though the white noise assumption is common, and is mathematically tractable, it is not entirely realistic. There are many phenomenological and mathematical reasons for this. For example, mathematically, white noise makes the presumption independence between the random variables $X(t)$ and $X(t + \Delta)$, where $\Delta$ is the sampling interval. But very often application settings of the selection of $\Delta$ is ignored in making this assumption. Furthermore, there are many processes associated with mechanical systems that are inherently not white. One example is "pink" noise associated with wind and other forms of fluid turbulence.

Under the more realistic assumption of a GM process, one can not only better characterize the phenomena under study, but also characterize it in a very parsimonious manner; that is, via the single parameter, $\beta$. Because $\beta$ alone captures the spectral shape of the process, it is far more robust than a non-parametric p.s.d. estimator. The reason is that, for a given amount of data, the p.s.d. entails estimation of energy at a large collection of frequencies. Unless this collection is very small relative to the amount of data used, the p.s.d. estimator will have high variability. On the other hand, since $\beta$ is a single parameter, even a modest amount of data can yield a very reliable estimate.
of it, hence of the corresponding GM process p.s.d..

Even though there is a wealth of literature concerning estimation of $\beta$, almost all of it concerns the setting wherein the sample interval is fixed and where the amount of data is allowed to approach infinity. In the case of mechanical systems, wherein one has essentially unlimited access to a given process, these results are directly applicable. But in many critical applications, such as monitoring the condition of a nuclear reactor cooling pump, the goal is to identify when vibration associated with normal operation is evolving into a different process associated with faulty operation. In this instance, even though one can measure vibration for days or weeks at a time, the process being measured is not one, but many. Hence, it is natural and expedient to address the situation wherein one has access to a GM process for only a fixed amount of time, say, $T$.

In Chapter 3 we addressed the problem of estimating $\beta$ in this finite observation time setting. Because of a finite observation time $T$, the variance of the estimated autocorrelation will not approach 0 even though as the process is sampled more rapidly, the sample size will approach infinity. This leads to the use of a second order expansion. Our key results included:

- Based on a second order Taylor expansion, we first developed expressions for the mean and variance of the least squared estimator of $\alpha$ and $\beta$ that are significantly more accurate than those commonly found in the literature (e.g. Priestly [21]) derived from a first order expansion. Though a second order expansion is not a novel method, by using it, we are able to correct results in the literature which assumes a first-order expansion is accurate enough.

- Three sample size regions—R-1 finite: $\beta\Delta \in (0.08, 0.314)$, R-2 large: $\beta\Delta \in (0.008, 0.08)$ and R-3 very large ones: $\beta\Delta < 0.08$, were located according to the gradient change of variance of the estimator of parameters and autocorrelations.
• A conjecture about the distribution of autocorrelation estimator was provided. The distribution of $\hat{\alpha}$ and $\hat{\beta}$ was also investigated.

• Guidelines in the choice of sampling interval were provided numerically according to the parameter estimation mse requirement.

• We then demonstrated the value of these expressions in relation to the prediction problem, with a simple application to active noise cancellation. One of our key contributions included a "steam table" type of graph that allows the researcher to balance prediction performance and calculation and storage cost.

In Chapter 4, we demonstrated the value of our results from Chapter 2 and 3 in a sound pressure analysis of a tractor engine cooling system. Our key results included:

• The sound pressure process is a mixed random process contain broadband spectrum as well as tones at most harmonics frequencies of the BPF. The tonal powers of the first 5 HF's of the BPF contain most of the line spectrum energy.

• The most important octave bands are those centered on 1000, 500 and 250 Hz, which contains 36.4%, 30.5% and 15.75% of total energy respectively; The most important 1/3 octave bands are those centered on 1000, 630, 800, 400 and 200 Hz, which contains 15.2%, 14.6%, 14.6%, 11.2% and 8.9% of total energy respectively;

• The first three harmonics of the sound pressure with mock engine installed contain 59.5%, 41.2% and 14.7% energy of the corresponding 1/3 octave bands on 200, 400, 630 Hz respectively.

• We showed the Matlab C.I. of p.s.d. estimator at the tone's frequency, for example, the BPF at 193 Hz, is incorrect and may lead to costly wrong design decision.

• We showed how to use families of spectral estimators in Chapter 2 can be used to detect tones. This is meant to offer advice to practitioners who use software
packages on the current market without knowing that the mixed spectral nature is ignored by the software packages. They have to detect the mixed spectral nature using our proposed method from Chapter 2 before they arrive at any decisions.

• An example hypothesis test performed on the dB(A) of 1/3 octave band centered on 800Hz under the two conditions showed that the dB(A) with mock engine is higher than that without mock engine at a significance level 0.05. The calculated P-value is 0.0033.

• Another key result applied from Chapter 3 is to model the process in the bands where there is no tone as a GM one and to perform a hypothesis test on the parameter for deciding whether a given mock engine design influences spectral of the sound pressure in that bands.

In conclusion, we believe that by addressing these two generic problems in a rigorous manner that integrates elements of a range of disciplines, we have been able to obtain results that have demonstrated value in relation to mechanical systems. This belief is supported by the ASME publication [25] associated with our Chapter 2 results, and, in part, by the conference publications [35] associated with our Chapter 3 results. With respect to the latter, we are in the process of submitting a manuscript to the Journal of Mechanical Systems and Signal Processing. As a final note, I believe that this dissertation has provided me with the basis to make further contributions to the field in whatever capacity I may find myself in throughout the development of my professional career; be it as an engineer in industry or a professor in universities.

5.2 Recommendation for Future Works

In using a family of p.s.d. estimators to characterize spectral information of a mixed r.p., the average of the family of p.s.d. estimators is used. However, in studying the
statistical variance of the the average of the family of p.s.d. estimators, we have to turn to simulation for lack of the joint distribution of the family of spectral estimators. To find the joint pdf of families of p.s.d. estimators is a more mathematically challenging problem. Since the spectral estimator of each order is from the same autocorrelation estimator, they are highly correlated. Currently only the asymptotic marginal distribution of the spectral estimator of each order is known which limits the usage of families of p.s.d. estimators. If the joint distribution is pursued in the future, then many functions of such families of p.s.d. estimators could be applied to better characterize the spectral information of a mixed r.p..

In fact, the spectral estimator of each order could be expanded as a linear function of the estimated autocorrelation function. Then a family of such p.s.d. estimators could be expressed as the product of a known matrix and the estimated autocorrelation. Given the statistics of the estimated lagged product autocorrelation, by δ method and with proper assumption, the joint statistics of such a family of p.s.d. estimators could be obtained. This method is specially useful in deriving the joint statistics of the family of PER(n) p.s.d. estimators. However, for the family of AR(n) spectrum, as the AR parameters has to be firstly estimated, then the families of AR(n) spectral estimators can be expressed as functions of AR parameters. It will be a more complicated problem to derive the joint statistics of a family of AR(n) spectral estimators.

In the statistical problem in the choice of a sampling rate in relation to parameter estimation, the process studied is a first order scalar GM process. Thus the results is suitable for this type of process, though we believe it could be extended to higher order processes since they can be decomposed into first order systems. Given the many type of r.p.es in mechanical systems, similar analysis could be applied to a second order system, whose parameter is a function of the fundamental frequency and damping ratio of the underlying system.

In summary, this dissertation provided a statistical foundation in generic problems
commonly observed in mechanical signal processing. Because of limitation of time and energy, works indicated above can be performed in the future.
BIBLIOGRAPHY


