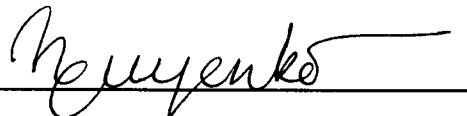


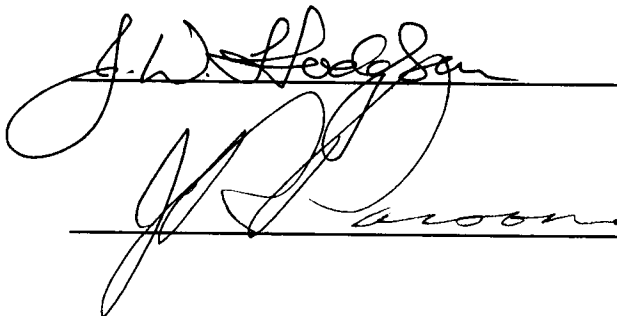
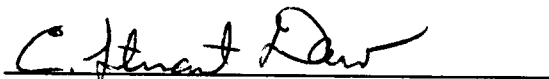
To the Graduate Council:

I am submitting herewith a thesis written by Charles Edward Andrew Finney, entitled "Identification and characterization of determinism in spark-ignition internal combustion engines". I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Mechanical Engineering.



Dr. Ke Nguyen, Major Professor

We have read this thesis and recommend  
its acceptance:



Accepted for the Council:



Associate Vice Chancellor and Dean of  
the Graduate School

## Statement of Permission to Use

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Tennessee, Knoxville, I agree that the Library shall make it available to borrowers under the rules of the Library. Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of the source is made.

Permission for extensive quotation from or reproduction of this thesis may be granted by my major professor, or in his absence, by the Head of Inter-library Services when, in the opinion of either, the proposed use of material is for scholarly purposes. Any copying or use of the material in this thesis for financial gain shall not be allowed without my written permission.

Signature

Charles E. Finney

Date

1995 July 21

# Identification and characterization of determinism in spark-ignition internal combustion engines

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Charles Edward Andrew Finney

August 1995

Copyright © 1995 by Charles Edward Andrew Finney.

All rights reserved.

## Acknowledgments

The author gratefully acknowledges Jeffrey S. Armfield of the Oak Ridge National Laboratory for development of the experimental apparatus and for data acquisition, John F. Thomas of the Oak Ridge National Laboratory for many interesting discussions and suggestions, Prof. Jeffrey W. Hodgson of the University of Tennessee for advice and experimental assistance, and Prof. J. Roger Parsons of the University of Tennessee for guidance as a member of the thesis committee. The author thanks R.W. Anderson of the Ford Motor Company for supplying experimental data sets. The author offers an especial thanks to his mentors and advisors, Prof. Ke Nguyen of the University of Tennessee and Dr. C. Stuart Daw of the Oak Ridge National Laboratory, for their continued guidance and support. Further, he expresses his appreciation to his colleagues M. Vasudevan, K. Dean Edwards and Nick A. van Goor for their patience and camaraderie. Most importantly, he thanks his parents for their love and support during his graduate studies in mechanical engineering.

# Abstract

Methods of chaotic time series analysis are applied to typical engine data in order to identify and characterize deterministic effects. Pressure and engine-speed time series from three different spark-ignition internal combustion engines are examined as example data sets.

The mutual information function is evaluated comparatively with the autocorrelation function to measure temporal coupling in peak-pressure sequences and engine-speed time series. The utility of windowing or Poincaré sectioning the data in order to highlight portions of the engine cycle is developed and discussed. Additionally, estimates of Kolmogorov entropy and correlation dimension and the method of surrogate data suggest a discernible degree of nonlinear deterministic effects in the examined data.

# Contents

|          |                                                          |           |
|----------|----------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                      | <b>1</b>  |
| <b>2</b> | <b>Background</b>                                        | <b>5</b>  |
| 2.1      | Overview . . . . .                                       | 5         |
| 2.2      | Development of engine data-acquisition systems . . . . . | 6         |
| 2.3      | Statistical analysis of engine data . . . . .            | 8         |
| 2.4      | Dynamic analysis of cyclic variation . . . . .           | 11        |
| 2.4.1    | Analytical studies . . . . .                             | 11        |
| 2.4.2    | Experimental studies . . . . .                           | 14        |
| <b>3</b> | <b>Experimental apparatus and data</b>                   | <b>17</b> |
| 3.1      | Overview . . . . .                                       | 17        |
| 3.2      | Test facility at ORNL . . . . .                          | 18        |
| 3.3      | Test facility at UT . . . . .                            | 25        |
| 3.4      | Data from Ford Motor Company . . . . .                   | 26        |
| <b>4</b> | <b>Data analysis and results</b>                         | <b>30</b> |
| 4.1      | Overview . . . . .                                       | 30        |
| 4.2      | Statistical tests . . . . .                              | 31        |
| 4.3      | Measures of temporal coupling . . . . .                  | 36        |
| 4.3.1    | Autocorrelation function . . . . .                       | 36        |
| 4.3.2    | Mutual information function . . . . .                    | 37        |
| 4.3.3    | Application to pressure sets . . . . .                   | 39        |
| 4.3.4    | Application to speed sets . . . . .                      | 46        |
| 4.4      | Observations of higher dimensionality . . . . .          | 53        |
| 4.5      | Estimates of Kolmogorov entropy . . . . .                | 55        |
| 4.5.1    | Definition of the estimator . . . . .                    | 55        |
| 4.5.2    | Data windowing . . . . .                                 | 59        |
| 4.5.3    | Normalization of entropy . . . . .                       | 60        |
| 4.5.4    | Application to pressure sets . . . . .                   | 61        |

## CONTENTS

v

|          |                                              |           |
|----------|----------------------------------------------|-----------|
| 4.5.5    | Application to speed sets . . . . .          | 65        |
| 4.6      | Estimates of correlation dimension . . . . . | 69        |
| 4.6.1    | Definitions . . . . .                        | 69        |
| 4.6.2    | Application to pressure sets . . . . .       | 70        |
| 4.6.3    | Application to speed sets . . . . .          | 76        |
| 4.7      | Tests for nonlinear structure . . . . .      | 76        |
| <b>5</b> | <b>Conclusions</b>                           | <b>84</b> |
|          | <b>List of References</b>                    | <b>86</b> |
|          | <b>Appendix</b>                              | <b>91</b> |
|          | <b>Vita</b>                                  | <b>96</b> |

## List of Figures

|      |                                                                                               |    |
|------|-----------------------------------------------------------------------------------------------|----|
| 2.1  | An indicator diagram from Clerk. . . . .                                                      | 7  |
| 2.2  | A temperature return map from Daily. . . . .                                                  | 13 |
| 3.1  | Twenty engine cycles from STID0629 and STID0630. . . . .                                      | 22 |
| 3.2  | Twenty engine cycles from LNID0629 and LNID0630. . . . .                                      | 23 |
| 3.3  | Time series of STID-FCE, STID-FCD and LNID-FCE. . . . .                                       | 24 |
| 3.4  | Time series segments from FI6V1, FI6V2 and FI6V4. . . . .                                     | 27 |
| 3.5  | Twenty engine cycles from FV6SA27 and FV6SA37. . . . .                                        | 29 |
| 4.1  | Frequency distributions of peak pressures for FV6SA27 and<br>FV6SA37. . . . .                 | 33 |
| 4.2  | Frequency distributions of peak pressures for STID0629 and<br>STID0630. . . . .               | 34 |
| 4.3  | Frequency distributions of peak pressures for LNID0629 and<br>LNID0630. . . . .               | 35 |
| 4.4  | Autocorrelation and mutual information functions of peak pres-<br>sures for FV6SA27. . . . .  | 40 |
| 4.5  | Autocorrelation and mutual information functions of peak pres-<br>sures for FV6SA37. . . . .  | 41 |
| 4.6  | Autocorrelation and mutual information functions of peak pres-<br>sures for STID0629. . . . . | 42 |
| 4.7  | Autocorrelation and mutual information functions of peak pres-<br>sures for STID0630. . . . . | 43 |
| 4.8  | Autocorrelation and mutual information functions of peak pres-<br>sures for LNID0629. . . . . | 44 |
| 4.9  | Autocorrelation and mutual information functions of peak pres-<br>sures for LNID0630. . . . . | 45 |
| 4.10 | Autocorrelation and mutual information functions of engine speed<br>for FI6V1. . . . .        | 47 |
| 4.11 | Autocorrelation and mutual information functions of engine speed<br>for FI6V2. . . . .        | 48 |

|                                                                                                                   |    |
|-------------------------------------------------------------------------------------------------------------------|----|
| 4.12 Autocorrelation and mutual information functions of engine speed for FI6V4. . . . .                          | 49 |
| 4.13 Autocorrelation and mutual information functions of engine speed for STID-FCE. . . . .                       | 50 |
| 4.14 Autocorrelation and mutual information functions of engine speed for STID-FCD. . . . .                       | 51 |
| 4.15 Autocorrelation and mutual information functions of engine speed for LNID-FCE. . . . .                       | 52 |
| 4.16 Phase-space trajectory reconstructed from STID0629. . . . .                                                  | 56 |
| 4.17 Normalized entropy for FV6SA27 and FV6SA37 as a function of normalized and absolute cutoff length. . . . .   | 62 |
| 4.18 Normalized entropy for STID0629 and LNID0629 as a function of normalized and absolute cutoff length. . . . . | 63 |
| 4.19 Normalized entropy for STID0630 and LNID0630 as a function of normalized and absolute cutoff length. . . . . | 64 |
| 4.20 Normalized entropy for LNID-FCE as a function of segment length. . . . .                                     | 66 |
| 4.21 Normalized entropy for FI6V1, FI6V2 and FI6V4 as a function of cutoff length. . . . .                        | 68 |
| 4.22 Dimension for FV6SA27 and FV6SA37 as a function of normalized and absolute cutoff length. . . . .            | 71 |
| 4.23 Dimension for STID0629 and LNID0629 as a function of normalized and absolute cutoff length. . . . .          | 72 |
| 4.24 Dimension for STID0630 and LNID0630 as a function of normalized and absolute cutoff length. . . . .          | 73 |
| 4.25 Dimension for FV6SA27 as a function of segment length. . . . .                                               | 75 |
| 4.26 Dimension for FI6V1, FI6V2 and FI6V4 as a function of normalized and absolute cutoff length. . . . .         | 77 |
| 4.27 Plot of LNID-FCE and a typical surrogate data set. . . . .                                                   | 81 |
| A.1 Example of data windowing on a fictitious time series. . . . .                                                | 93 |
| A.2 Example calculation of the decay function. . . . .                                                            | 94 |

# Chapter 1

## Introduction

Variability in cycle-to-cycle performance of spark-ignition internal combustion engines, especially under idling and fuel-lean conditions, is well known and documented. This cyclic variability coincides with erratic combustion performance and is observable in many aspects of engine operation, because the combustion process directly affects cylinder pressure and greatly influences exhaust composition and instantaneous piston speed and torque. Authors such as Heywood [1] provide comprehensive discussions of these phenomena, and Young [2] and Ozdor *et al.* [3] have completed literature surveys on the topic.

Cyclic variability typically causes differences in the detailed pressure histories of successive cycles. Whereas it is typical for the magnitude and timing of peak pressure within each cycle to be correlated, as quicker burning cycles tend to have larger peak pressures, the correspondence of any pressure-derived

variable, such as IMEP or peak pressure, from cycle to cycle generally has been considered stochastic (e.g., see Barton *et al.* [4]) until recently.

Instantaneous torque and speed variations result in unsteady power output to the crankshaft. In multicylinder engines, the cylinder firing sequence can mitigate effects of individual cylinder variations, but substantial average variability can drastically affect driveability in the case of motor vehicles.

The exhaust composition also depends on the combustion process. Small-scale cycle-to-cycle variations in exhaust composition can be tolerated by emissions-control systems, but exceptional combustion events, such as misfires, have deleterious effects on emissions quality. In the case of misfires, unburned hydrocarbons are expelled to atmosphere, resulting in lower fuel economy and increased pollution.

The extent of variability generally depends on the combustion stoichiometry — as the fuel-oxidizer mixture becomes increasingly fuel-lean, the degree of cyclic variability increases. Whereas some cycles burn completely, others experience partial burning, in which the exhaust valve opens before combustion is complete, or misfires, in which combustion fails to occur. As the mixture becomes increasingly fuel-lean, partial burning and then misfires become more frequent and contribute to extreme cyclic variability.

Of particular engineering concern is whether cyclic variability exhibits prior-cycle effects, constituting a feed-forward propagation of information which

might enable prediction and avoidance of extreme combustion events such as misfires. Before the advent of deterministic chaos theory, many superficially random processes were considered uncontrollable. It is now recognized that nonlinear determinism can sometimes masquerade as randomness, and for such cases techniques for deterministic chaos control can be used to improve control significantly.

The benefits of reducing variability are obvious and substantial from industrial and environmental perspectives, and with very recent advances in cycle-resolved control capabilities (e.g., see Stevens *et al.* [5] and Atkinson *et al.* [6]) and on-board data acquisition and processing, the need to determine the nature and mechanisms of prior-cycle effects is of practical import.

In contrast with the above, some authors suggest that a degree of variability is tolerable if not desirable. Stone [7] suggested that currently implemented knock-control schemes would fail if there were no cyclic variability, for all cycles would either knock or experience normal combustion. Additionally, Heywood [1] remarked that three-way catalytic converters achieve better conversion efficiency as a result of oscillations in the equivalence ratio.

Previous studies of the effects of combustion variability predominantly have focused on the variation of pressure or flame-kernel related parameters, such as kernel growth rate or percentage burn times, over a few successive cycles. Other studies have examined engine rotational speed fluctuations over tens to hundreds of cycles and have attributed these oscillations to lag effects such as manifold

charging delay or fuel metering and transport lag excited in a closed-loop system by combustion variability and perpetuated by the idle-speed controller [8, 9, 10]. These later works have not examined whether speed fluctuations, resulting from variability in the interaction of the engine and the drivetrain, can exist in the absence of the controller.

The primary objective of the present study is to demonstrate how the application of time serial analytical methods based on deterministic chaos theory can be used to characterize experimental engine data and to determine if prior-cycle effects exist.

A secondary objective is to compare chaotic time serial methods with some of the more standard analytical techniques.

Chapter 2 presents a brief literature survey of the methods used to study cyclic variability and of studies which suggest that application of time serial analytical techniques based on deterministic chaos theory would be useful if not more desirable than conventional methods. An overview of the experimental apparatus and procedures employed to collect example data sets to test the analytical techniques is given in Chapter 3. Chapter 4 explains the development of the analytical methodology and discusses selected results. Chapter 5 offers conclusions of this study as well as its impact and implications for future study.

## Chapter 2

# Background

### 2.1 Overview

This chapter reviews previous studies on the characterization of cyclic variability. The first section highlights the development of data-acquisition systems which have facilitated sophisticated analysis of cyclic variability. The second section reviews the application of conventional statistical analytical techniques, particularly those designed to detect deterministic prior-cycle effects. The third section presents results of more recent studies involving the methodologies of time series analysis and information theory.

The mechanisms of cyclic variability are beyond the scope of this thesis and will not be reviewed in the present chapter. Although other variables, such as optical flame-front observations, instantaneous speed fluctuations, or acoustic

signatures, may be used to monitor cyclic variability, the following reviews will be limited primarily to those utilizing cylinder-pressure measurements.

## 2.2 Development of engine data-acquisition systems

Cycle-to-cycle irregularity in combustion was noticed very early in the development of spark-ignition internal combustion engines. By the end of the nineteenth century, experimenters had developed indicator diagrams, one form of which used an arc lamp mechanically linked with the engine to trace a pattern on photographically sensitive paper to produce an effective pressure-volume diagram of the engine cycle. When these indicator diagrams were produced over successive cycles, the irregularity in the superimposed traces, particularly in the combustion and expansion strokes, was noticeable.

Clerk's 1893 work [11] displayed several indicator diagrams, which at the time were used to explain the differences between ideal and actual engine performances and to characterize the effects of parametric differences such as compression ratio or throttle position. One of Clerk's diagrams is reproduced in Figure 2.1. Contemporary explanations for the cyclic variability were that fuel was not totally atomized or that exhaust residual led to charge stratification, reasons echoed in 1909 by Jones [12] and in 1911 by Homans [13].

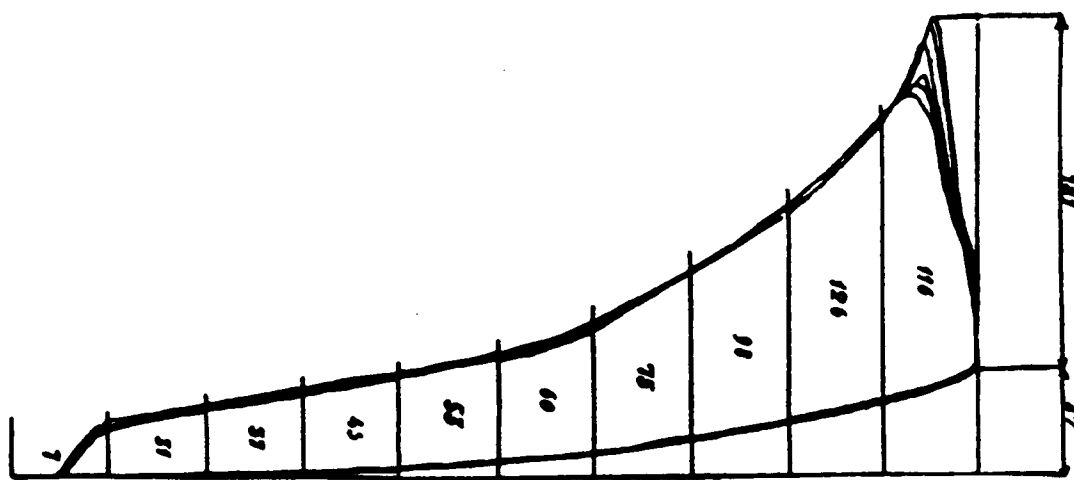


Figure 2.1: An indicator diagram from Clerk. Source: Clerk D. (1893). *The gas engine*, fourth edition. John Wiley and Sons.

With the development of the oscilloscope, experimenters were able to capture and examine several cycles of pressure traces easily in real time, and instrumentation signals could be recorded onto magnetic tape for later processing. Curry's 1963 work [14] was one of the first studies to link instrumentation with a digital electronic computer. His work was heralded by his contemporaries for enabling him to acquire multivariate measurements from more than 2000 cycles at a time, compared with 20 at a time on an oscilloscope screen or photograph, thus gaining a statistical advantage.

With these advances in data-acquisition systems, researchers were able to progress from very simple statistical characterization based on reduced variables, such as magnitude or timing of peak pressure, to more sophisticated analysis based on the complete time histories of many variables. The proliferation of inexpensive and powerful microcomputer-based data-acquisition systems has facilitated analysis and possible control of combustion irregularities.

## 2.3 Statistical analysis of engine data

Some of the earliest available pressure-related variables were reduced variables, those which characterized the cycle with a single or few numbers. Examples are magnitude of peak pressure, crank angle or position of peak pressure, maximum rate of pressure rise, and IMEP. For measurements conducted over

multiple cycles, statistics such as mean and variance could be calculated on the individual variables and cross-correlation coefficients could be calculated on pairs of variables.

Heywood [1] presented a commonly used and widely accepted statistic, the *coefficient of variance* of IMEP, defined as

$$COV = \frac{\sigma_{IMEP}}{\overline{IMEP}}$$

where  $\sigma_{IMEP}$  is the standard deviation of IMEP and  $\overline{IMEP}$  is the arithmetic mean of the IMEP. Heywood reported that when the *COV* exceeds ten percent, vehicular driveability is reduced.

Patterson [15] reported that Vichnievsky [16] had studied peak-pressure variability statistically to see the effects of different fuels and operating conditions and that Soltau [17] had defined a new variable to describe variability. Soltau defined variability as the average difference in peak pressure between each cycle and the mean cycle and reported his results as a fraction of the mean value of peak pressure or IMEP. Some of Soltau's conclusions were based on as few as ten-cycle samples.

Patterson [15] used histograms to approximate the probability density functions of variables such as magnitudes of peak pressures and maximum rates of pressure rise. His work relied on having several thousand measurements in order to obtain "smooth" histograms, and he rejected from analysis cycles outlying a

$\pm 2\sigma$  band about the mean. Patterson also examined the effects of sample size on statistical consistency, noting that whereas 2000-cycle samples were more than sufficient, 20-cycle samples, which represented the maximum number of traces which could be accurately identified on oscilloscope photographs, were inadequate for statistical characterization.

Barton *et al.* [4] statistically analyzed four pressure-related variables: magnitude and timing of peak pressure and magnitude and timing of the maximum pressure-rise rate. Their apparatus consisted of a single-cylinder engine and a computer-controlled data-acquisition system which allowed them to collect data from hundreds of successive cycles. Based on statistical measures of skewness, a characterization of symmetry about the mean in the data distribution, and kurtosis, a metric of peakedness or shape of the distribution, they concluded that peak pressures were normally distributed. Moreover, as the autocorrelation coefficients of peak pressure were all insignificant, they surmised that peak pressures exhibited no cycle-to-cycle determinism and represented a gaussian random process.

Matekunas [18] examined IMEP in addition to the same four pressure-related variables as Barton *et al.* for correlations with burn rates. He suggested that IMEP might be a better indicator than the peak-pressure variables because it represented a more direct measure of the response of the engine to combustion instabilities.

The traditional statistical methods described previously assume that the data are normally distributed, independent and random events. Because of that assumption, the use of traditional statistical methods precludes the possibility of serially correlated data and therefore are inadequate to characterize temporal correlations in data, which would be present if deterministic effects exist. Moreover, with the exception of the autocorrelation function, the standard methods characterize static relations in the data, whereas time series analytical techniques generally highlight dynamical relations in the data.

## 2.4 Dynamic analysis of cyclic variation

Almost all of the studies reviewed in the previous section did not use time-based measures but relied on a single statistic, such as the coefficient of variability or variance of peak pressures, to describe a data set of many samples. This section reviews studies that depended on temporal correlations for analysis of cyclic variability or for detection of prior-cycle effects.

### 2.4.1 Analytical studies

Kantor [19] presented a discrete model of dynamical combustion instability in spark-ignition engines in which the mean exhaust temperature oscillates from cycle to cycle because of the coupling of the residual gas fraction and

combustion rate. The Kantor model assumed a constant inter-cyclic residual gas fraction and modeled the combustion rate with a laminar flame and Arrhenius reaction rate.

Daily [20, 21] explored the Kantor model and suggested parametric values, such as activation energy, which could yield chaotic behavior. Daily's major results included a complex iterate map of exhaust temperature which suggested chaotic behavior. Figure 2.2 is a reproduction of one of Daily's temperature return maps, which illustrates the chaotic nature of the Kantor model. The structure of the return map is off of the identity line, which represents periodic or constant inter-cyclic temperature, yet is not space-filling and amorphous, which would be the case in a random inter-cyclic process.

Foakes and Pollard [22] later claimed that for realistic thermochemical properties, the Kantor model was not chaotic. For instance, much of the chaotic regime of the bifurcation diagram from the Kantor model yielded extremely large burn angles, the time necessary to complete combustion, for values of activation energy not found in known fuels; for values consistent with common fuels, the burn angles reduced to values non-chaotic and consistent with experimental observations. This finding implied that the fuel-chemistry effects of the Kantor model were not the predominant factors in cyclic variability, and Foakes and Pollard recommended that the issue of determinism in cyclic variability be resolved experimentally, citing work by Martin *et al.* [23], which found that certain

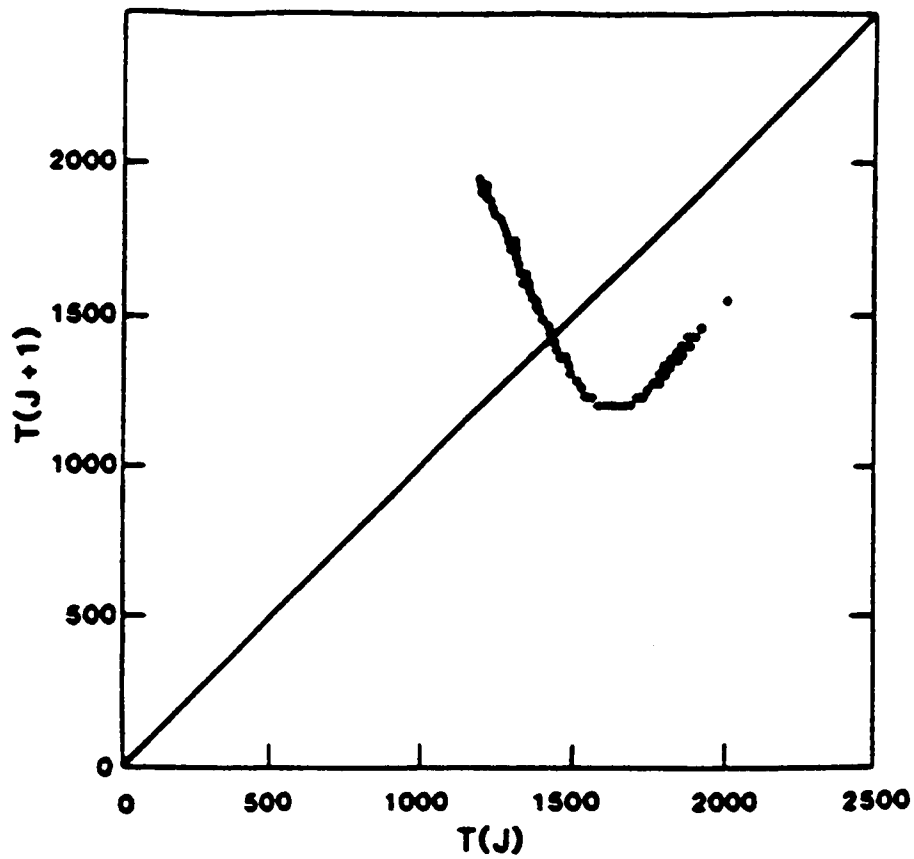


Figure 2.2: A temperature return map from Daily. Source: Daily J.W. (1988).

“Cycle-to-cycle variations: a chaotic process?”, *Combustion Science and Technology* **57**, 149-162.

lean-burn modes in a spark-ignition engine showed prior-cycle effects.

Wagner *et al.* [24] demonstrated numerical control on the Kantor model using an established chaos control method based on that suggested by Ott *et al.* [25]. The control scheme was based on simple proportional feedback applied to a linearized region near the control fixed point on the exhaust-temperature iterate map and made small cycle-to-cycle perturbations to the spark advance to achieve constant-temperature control.

## 2.4.2 Experimental studies

Belmont *et al.* [26] established a “memory” effect in cyclic variability by demonstrating non-zero autocorrelation coefficients in peak pressure of a single-cylinder engine. Their autocorrelation procedure was similar to that of Barton *et al.* [4], who had concluded that cycle-to-cycle variations in peak pressure were random. They used a single-cylinder side-valve engine, anticipating that the slow-burn characteristics of this engine coupled with cyclic variability would highlight the memory effects that they sought to observe. In many engine operating conditions, analysis of 4096 consecutive peak-pressure values yielded statistically significant autocorrelation coefficients. They concluded that the second autocorrelation coefficient, expressing coupling of one cycle and its immediate successor, represented the degree of “memory” from one cycle to the next.

Martin *et al.* [23] examined different burn modes in a single-cylinder

engine. They found, based on inspection of flows on IMEP return maps, that for some fuel-lean engine operating conditions, there were discernible cycle-to-cycle patterns in IMEP, but for stoichiometric conditions, there were no apparent deterministic effects. They suggested that same-cycle and prior-cycle effects are always present in different degrees of dominance in engines, making visibility of determinism depend on the nature and dominance of the same-cycle or prior-cycle as well as stochastic effects.

Most recently, Grünefeld *et al.* [27] reported strong evidence of deterministic prior-cycle effects in IMEP in a four-cylinder port-injected engine and attributed the variability to changes in the equivalence ratio and effects of residual gases on the fresh charge. The equivalence ratio was found to oscillate from cycle to cycle because of an inventory effect of fuel wetting the intake port. Their claims of determinism were based on significant correlation on scatter plots of IMEP with water content and with equivalence ratio of the succeeding cycle.

Chew *et al.* [28] claimed that observations of in-cylinder pressure “conclusively” demonstrated chaos. They suggested that the intake runner affected the charge conditions and presented a model to account for this effect. Their conclusions that cyclic variability is chaotic were founded on a projection of the phase-space trajectory in the peak-pressure and angle-of-peak-pressure plane, and they also showed a Poincaré section in the peak-pressure and maximum pressure-rise rate plane.

Finney *et al.* [29] presented experimental observations demonstrating deterministic prior-cycle effects in a single-cylinder engine. They derived a predictability factor based on a maximum-likelihood estimator of the Kolmogorov entropy and found that short-term deterministic effects exist in cylinder pressure traces. Their measure of determinism was based on comparing the entropy values of the experimental data with values of a random signal with the same probability distribution as the original data, and they demonstrated a method of quantifying the level of determinism based on the observational length scale of importance.

Past studies generally have approached analysis of engine data from a static, one-dimensional perspective, *i.e.*, based on a single quantifier of cycle or data-set performance without regard for temporal causality. Of the studies that have used dynamic analysis, most of the reduced variables have been one-dimensional (*e.g.*, peak pressure or IMEP). The present study will approach analysis of engine data from a dynamic, multi-dimensional perspective and will employ time serial methods which account for temporal sequence and which use several data per engine cycle to describe the dynamical behavior of the engine.

## Chapter 3

# Experimental apparatus and data

### 3.1 Overview

This chapter describes the experimental apparatus and example data sets used in this project. Each section presents the components and data-acquisition capabilities and summarizes the data types and sets from the engine test facilities which produced test-case data for the analytical methods developed in this project.

Three separate engine test facilities produced data which were used in this project. The primary stand was located at the Oak Ridge National Laboratory, and the secondary stand was located at the University of Tennessee. Additionally, R.W. Anderson of the Ford Motor Company supplied several representative data sets [30].

The test engines were chosen on the bases of availability and ability

to generate data types which are typical in engine analysis. Therefore, for the purpose of this study, the important feature of each apparatus is not the quality of the system but rather its ability to produce representative data sets suitable to serve as examples and test cases for the algorithms developed herein.

## 3.2 Test facility at ORNL

The experimental test facility at the Advanced Propulsion Technology Center features several experimental engines, microcomputer-based data-acquisition systems and emissions analyzers. The engine used in this study was a Kohler Magnum 12, which had a single cylinder of 85.9 mm (3.38 in) bore and 82.6 mm (3.25 in) stroke, for a 0.476 L (29.1 cu.in) displacement. The engine was equipped with port electronic fuel injection governed by an Electromotive TEC engine management system using a speed-density strategy, which allowed for independent control of air-fuel ratio and spark timing. The cylinder head was an L-configuration and had a machined portal into which a flush-aligned, water-cooled Kistler 7063A pressure transducer was mounted. A 25-hp eddy current dynamometer was elastically linked to the crankshaft, and the dynamometer brake power could be set electronically. Instrumentation and control leads were connected to analog-to-digital conversion boards attached to two microcomputer systems. An Apple Macintosh running LabView served as the

data-acquisition computer, whereas an Intel-based microcomputer running a program called SUPER (supplied by Electromotive) served as the control computer. The LabView program provided a virtual-instrumentation user interface which allowed computer-controlled data-acquisition and storage of several user-selected variables. The SUPER program interfaced with the engine control module and allowed user inputs to change parameters such as throttle position, spark advance and fuel feed rates. A National Instruments NB-MIO-16L data-acquisition board was used for low-speed sampling, and a National Instruments NB-A2000 was used for high-speed sampling. Both boards had 12-bit digitization resolution.

Analog signals were amplified but not filtered before digitization. Under certain conditions, line-noise contamination and aliasing corruption of the digitized signals were evident, and only data sets with indiscernible contamination were chosen for further analysis.

Four high-speed and three low-speed data sets were selected as example sets. As cyclic variability generally is more prevalent at fuel-lean cylinder mixtures and idle speeds, only data sets from fuel-lean, low-load and stoichiometric conditions were chosen for study. The high-speed sets were time series of cylinder pressure sampled on a time basis. The low-speed sets were multivariate time series sampled on a time basis, and the most significant variables were engine speed measured with a magnetic crankshaft encoder, air-fuel ratio, concentrations of  $CO$ ,  $CO_2$ ,  $O_2$ ,  $NO_x$  and UHC, and manifold pressure and temperature.

All runs were conducted using indolene fuel. All experimentation was performed under the guidance of Jeffrey S. Armfield, a research staff member at ORNL.

Table 3.1 summarizes the operating conditions and characteristics of the selected data sets, and figures 3.1-3.3 show typical segments of the time series. The pressure data are centered about a mean value of zero, which is a result of the high-pass filtering effects characteristic of piezoelectric transducers. As a result, there are large negative values of pressure because the zero is referenced to the data-set mean rather than atmospheric pressure. In the pressure time series, variations from cycle to cycle, particularly in the peak and trough magnitudes, are noticeable. In the speed time series, oscillations persist when the fuel controller is disabled (compare STID-FCE and STID-FCD). Additionally, intermittent large-amplitude and long-term excursions from the nominal engine speed are common, as in LNID-FCE.

Typically, the idle-speed fuel controller independently changed spark advance and the fuel-injector pulse width in response to engine speed and manifold pressure. In the data set in which the fuel controller was disabled (STID-FCD), the spark advance was fixed to 27 degrees before TDC, and the fuel-injector solenoid pulse width was set to a constant value. As the amount of fuel per cycle was fixed, the air-fuel ratio was governed only by the amount of air drawn in each cycle, and the stoichiometry would be expected to change with engine speed.

Table 3.1: Description of data sets from Kohler engine at ORNL.

| Label               | Records | Sampling rate [Hz] | Engine cycles | Points/cycle | Description                                   |
|---------------------|---------|--------------------|---------------|--------------|-----------------------------------------------|
| High-speed sampling |         |                    |               |              |                                               |
| STID0629            | 360000  | 9000               | 498           | 722          | Stoichiometric idle, June 29                  |
| LNID0629            | 360000  | 9000               | 481           | 747          | Lean idle, June 29                            |
| STID0630            | 360000  | 9000               | 584           | 692          | Stoichiometric idle, June 30                  |
| LNID0630            | 360000  | 9000               | 517           | 696          | Lean idle, June 30                            |
| Low-speed sampling  |         |                    |               |              |                                               |
| STID-FCE            | 6000    | 10                 | —             | —            | Stoichiometric idle, fuel controller enabled  |
| STID-FCD            | 6000    | 10                 | —             | —            | Stoichiometric idle, fuel controller disabled |
| LNID-FCD            | 27748   | 12                 | —             | —            | Lean idle, fuel controller enabled            |

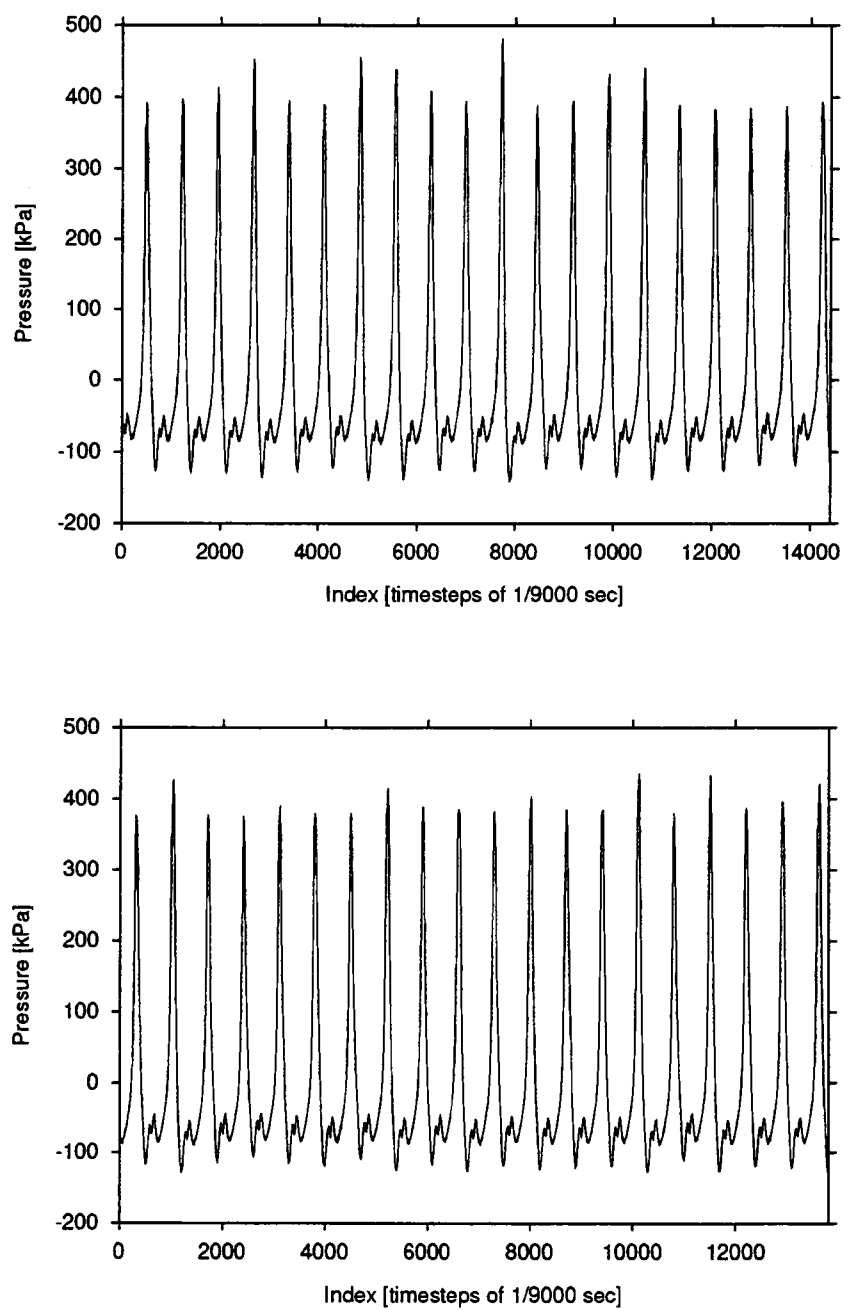


Figure 3.1: Twenty engine cycles from STID0629 (top) and STID0630 (bottom).

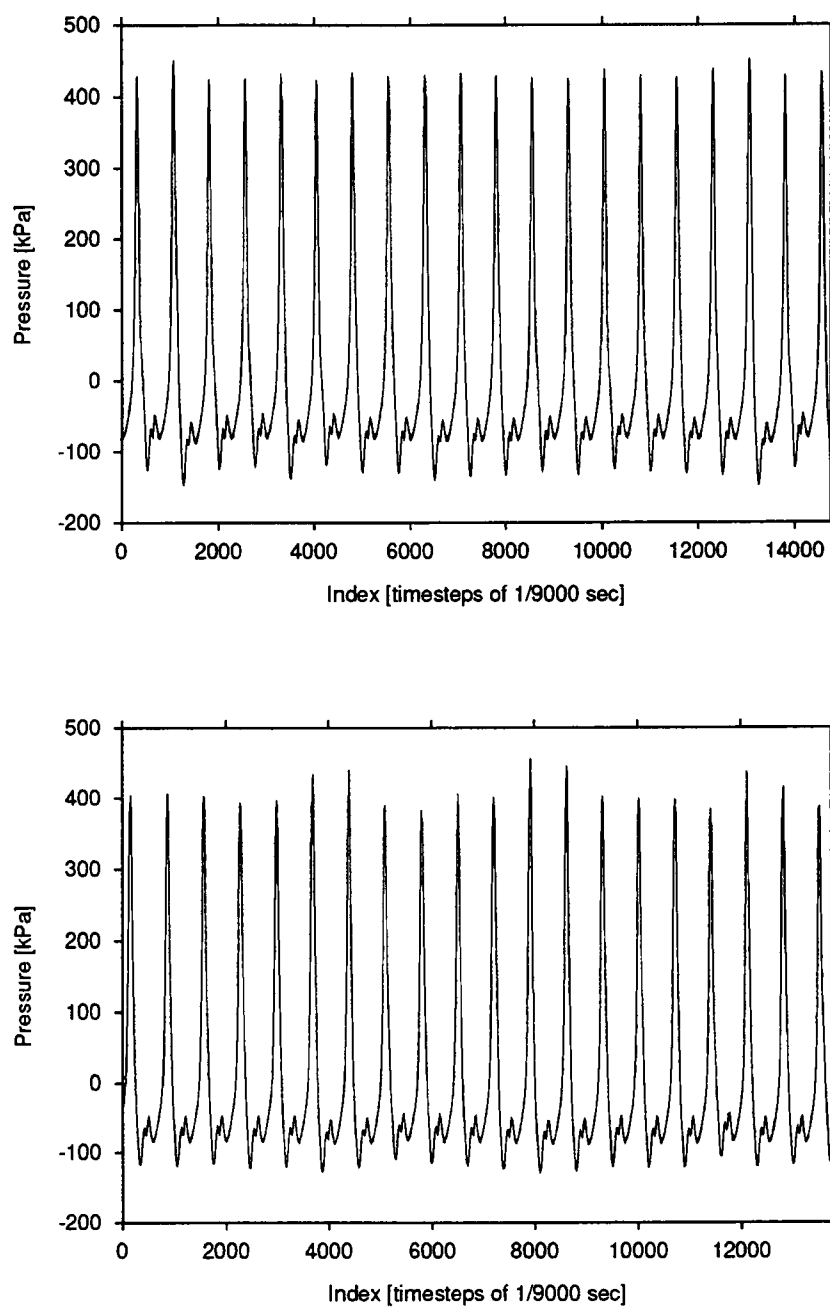


Figure 3.2: Twenty engine cycles from LNID0629 (top) and LNID0630 (bottom).

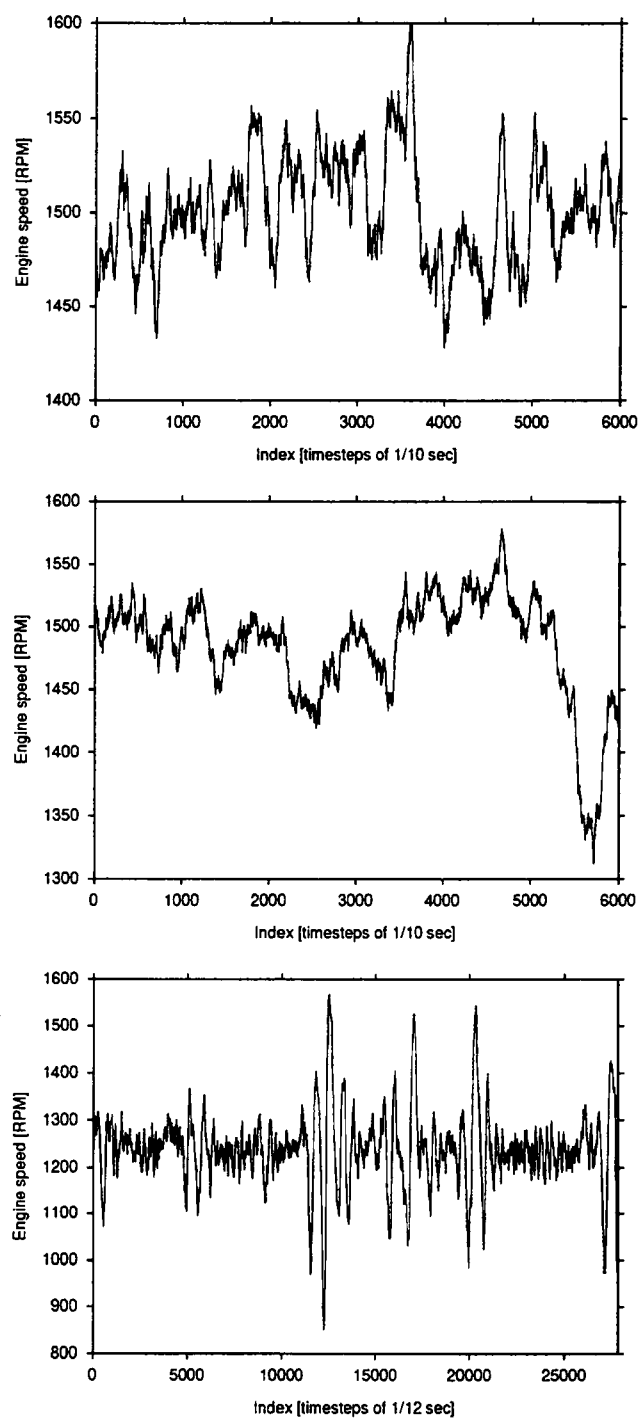


Figure 3.3: Time series of STID-FCE (top), STID-FCD (middle) and LNID-FCE (bottom).

### 3.3 Test facility at UT

The engine test stand consisted of a production engine and a data-acquisition system. The engine was a 1972 Ford 200 CID, which had six in-line cylinders with a total of 3.3 L (200 cu.in) displacement. In place of the carburetor was a cylindrical fuel-oxidizer mixing chamber, into which gaseous propane from a pressurized cylinder was introduced at a regulated, constant rate and ambient air was drawn by the action of the engine. Cylindrical extensions to the mixing chamber allowed its volume to be increased by factors of 2, 3 and 4. Below the mixing chamber was the throttle, and a set screw maintained throttle position. A Weston Tachometer Generator was linked to the crankshaft, and the output of the tachometer was 12 VDC at a maximum speed of 2000 RPM. The analog signal from the tachometer was band-pass filtered from 0.1 to 30 Hz with a Rockland 852 filter. The signal was high-pass filtered to remove the DC offset to achieve maximum digitization resolution of engine-speed variations, and the signal was low-pass filtered to reduce contamination by the ambient 60 Hz line field, which was rather strong and noticeable without filtering. A Nicolet 440 digitizing oscilloscope recorded the filtered signal with 12-bit analog-to-digital resolution.

Three data sets of engine speed sampled on a time basis were chosen as example sets. The fuel was metered with a rotameter for reproducibility. All data sets were recorded at 100 samples per second for ten minutes, for a total

of 60000 points per data set. The throttle was left unchanged at a nearly closed position throughout the runs. Table 3.2 summarizes the operating conditions of the selected data sets, and Figure 3.4 displays typical segments of the time series. In these figures, as the mixing-chamber volume is increased, the magnitude and time scales of engine-speed fluctuations increase noticeably.

Table 3.2: Description of data sets from Ford engine at UT.

| Label | Mixing-chamber<br>volume [L] |
|-------|------------------------------|
| FI6V1 | 0.75                         |
| FI6V2 | 1.5                          |
| FI6V4 | 3.0                          |

### 3.4 Data from Ford Motor Company

R.W. Anderson of the Powertrain Research Laboratory of Ford Motor Company supplied two data sets to serve as examples of time series sampled on a crank-angle basis rather than a time basis. The engine was a production Ford V6, and one of the cylinders was equipped with a flush-mounted piezoelectric pressure transducer. Pressure data were collected in a 540 crank-angle-degree window, and digitized data on the A/D conversion board were transferred to microcomputer during the remaining 180 crank angle degrees (CAD), which corresponded to the

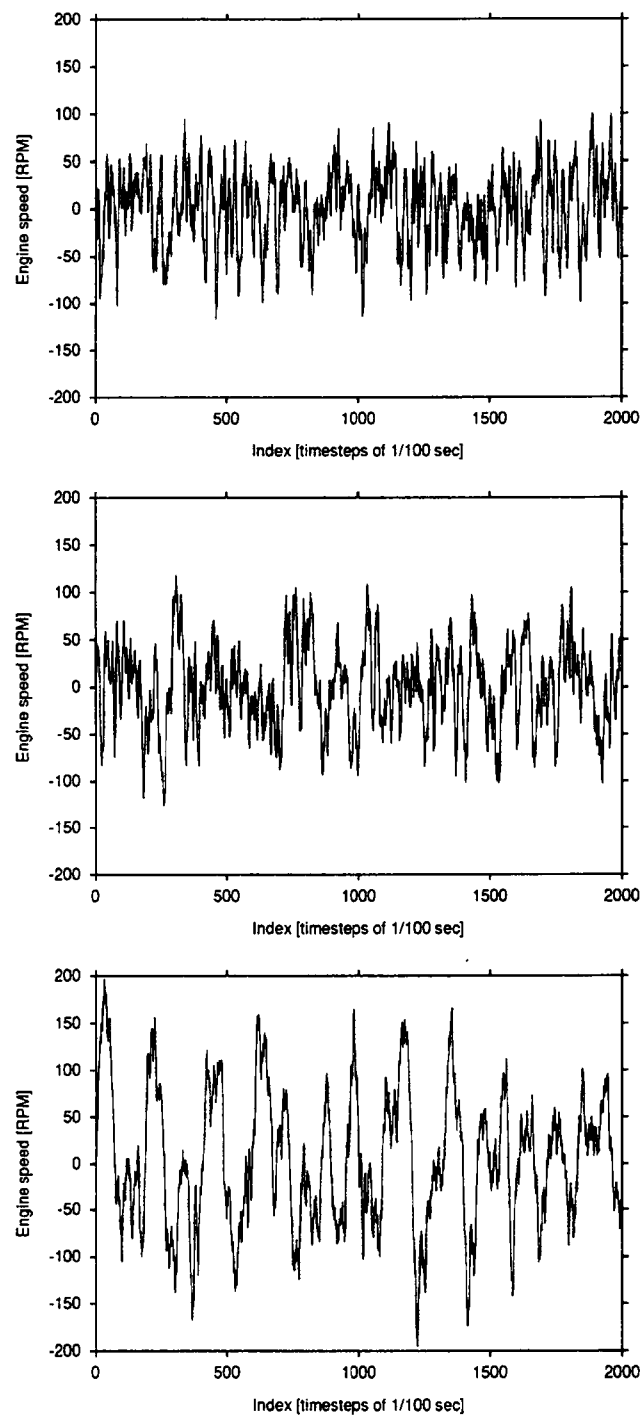


Figure 3.4: Time series segments from FI6V1 (top), FI6V2 (middle) and FI6V4 (bottom).

intake stroke.

Two data sets were selected as examples. All experimental conditions were similar, with the exception of spark advance, which was 27 CAD for data set FV6SA27 and 37 CAD for data set FV6SA37. Engine speed was approximately 1500 RPM with 2.62 bar load and stoichiometric air-fuel mixtures. Data were collected for 256 engine cycles. Figure 3.5 displays typical segments of the pressure time series. In these plots, the missing data points from the intake stroke have been linearized in order to create continuous time series. The cyclic variability in these segments is pronounced.

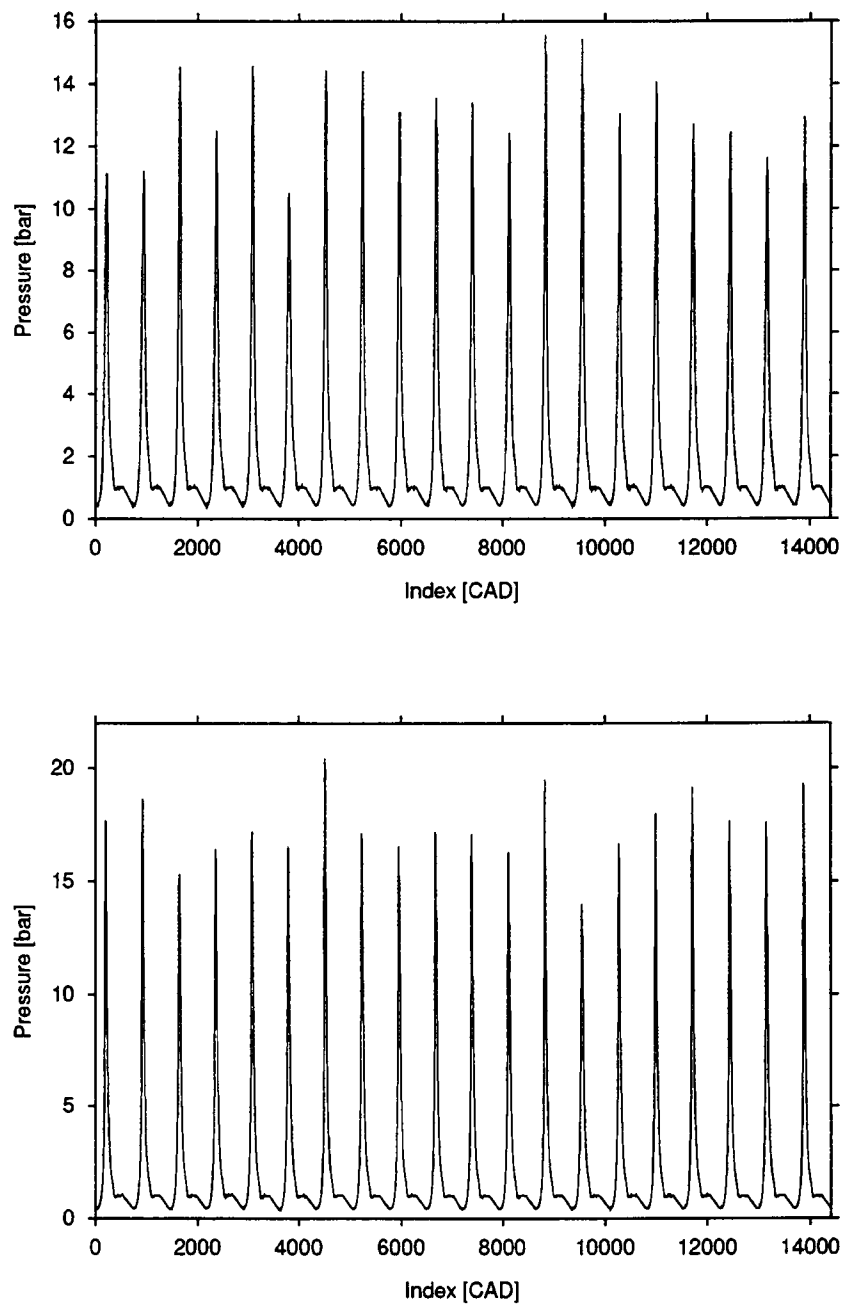


Figure 3.5: Twenty engine cycles from FV6SA27 (top) and FV6SA37 (bottom).

## Chapter 4

# Data analysis and results

### 4.1 Overview

This chapter describes the methods used to analyze the experimental data. All analysis was performed on digital electronic computers using FORTRAN programs implementing algorithms developed for engine time series or adapted from the chaotic time series analysis program library of C.S. Daw of the Oak Ridge National Laboratory [31].

In this chapter, the term *pressure sets* will refer to the six data sets with pressure as the dependent variable: FV6SA27, FV6SA37, STID0629, STID0630, LNID0629 and LNID0630. Likewise, the term *speed sets* will refer to the six sets with crankshaft rotational speed as the dependent variable: FI6V1, FI6V2, FI6V4, STID-FCE, STID-FCD and LNID-FCE.

## 4.2 Statistical tests

Peak pressures were determined for the six pressure sets, and the sets of peak pressures were analyzed using techniques defined in the literature to serve as a basis for comparison with previous studies. Two discriminating statistics were calculated: the third moment, or *skewness*, and the fourth moment, or *kurtosis*. The coefficient of skewness is defined as

$$\text{skewness} = \frac{1}{(N-1)\sigma^3} \sum_{1 \leq i \leq N} (x_i - \bar{x})^3$$

and the coefficient of kurtosis is defined as

$$\text{kurtosis} = \frac{1}{(N-3)\sigma^4} \sum_{1 \leq i \leq N} (x_i - \bar{x})^4$$

where  $\sigma$  is the standard deviation in  $x$ ,  $\bar{x}$  is the set mean, and  $N$  is the number of elements in the set. The coefficient of skewness quantifies the degree of symmetry about the mean of the data distribution, and the coefficient of kurtosis describes the shape of the distribution. For a normal distribution, skewness is zero, and kurtosis is three. A negative skew denotes a frequency bias below the mean, whereas a positive skew denotes a frequency bias above the mean. Calculating the skewness and kurtosis of a data set is important in order to test the validity of the assumption made by many statistical tests that the data are random and normally distributed.

Physically, normally distributed data often belie a superposition of many independent random sources, whereas a non-gaussian distribution suggests a few

dominant independent sources. Therefore, the nature of data distribution is important to establish the dimensionality or complexity of the system, which would have implications on analytical and control applications.

Table 4.1 summarizes the statistical characteristics of the peak-pressure sets. The sets FV6SA27 and FV6SA37 are almost normally distributed, similar to the findings of Barton *et al.* [4]. The Kohler sets are distinctly non-gaussian, as evidenced by their large values of skewness and kurtosis. Figures 4.1-4.3 display the frequency distributions for the peak pressures of the pressure sets. The differences in the distributions for the Ford and Kohler sets might be attributed to different engine configurations (number of cylinders), operating conditions or cylinder geometries. It should be emphasized that these sets have fewer elements than might be desirable for a confident statistical characterization.

Table 4.1: Summary of statistics for peak-pressure sets.

| Label    | Records | Skewness | Kurtosis |
|----------|---------|----------|----------|
| FV6SA27  | 249     | 0.03     | 2.61     |
| FV6SA37  | 249     | -0.21    | 2.91     |
| STID0629 | 498     | 1.92     | 7.11     |
| STID0630 | 584     | 1.49     | 5.79     |
| LNID0629 | 481     | 2.36     | 10.3     |
| LNID0630 | 517     | 1.78     | 6.76     |

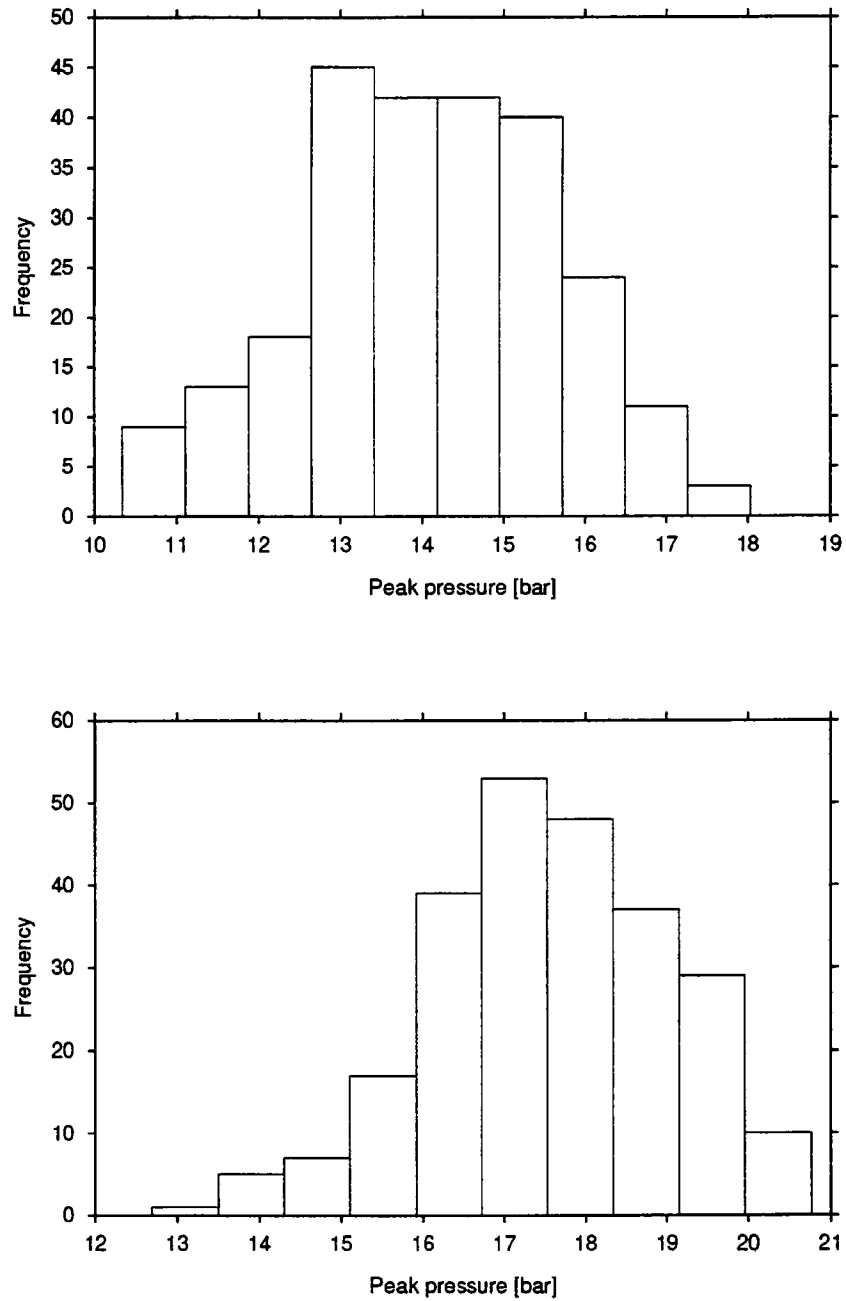


Figure 4.1: Frequency distributions of peak pressures for FV6SA27 (top) and FV6SA37 (bottom).

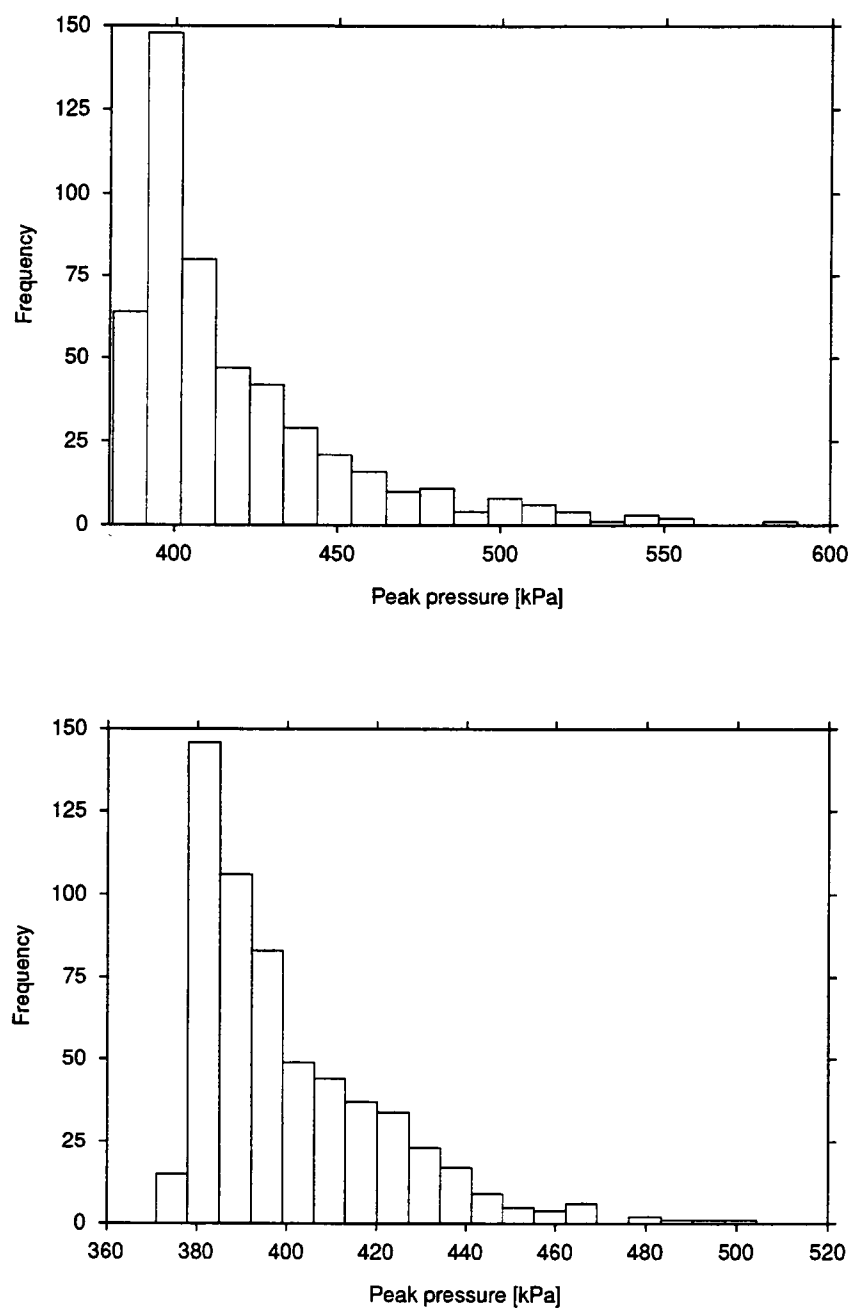


Figure 4.2: Frequency distributions of peak pressures for STID0629 (top) and STID0630 (bottom).

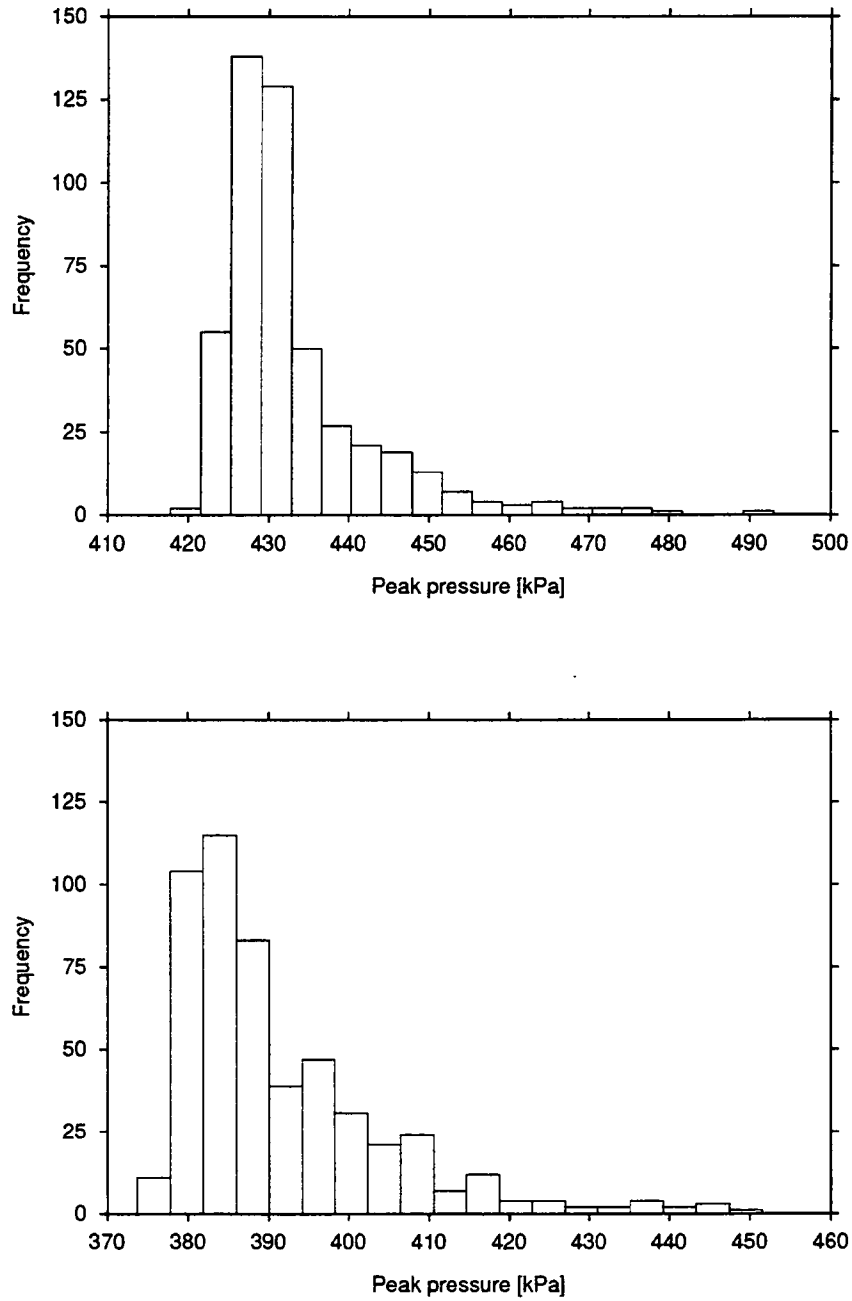


Figure 4.3: Frequency distributions of peak pressures for LNID0629 (top) and LNID0630 (bottom).

## 4.3 Measures of temporal coupling

### 4.3.1 Autocorrelation function

Evaluation of the autocorrelation of a temporal sequence is used to establish causality or serial coupling between elements in a time series. The “most satisfactory estimate” [32] of the autocorrelation function is defined as

$$r_k = \frac{c_k}{c_0}$$

for

$$c_k = \frac{1}{N} \sum_{1 \leq i \leq N-k} (x_i - \bar{x})(x_{i+k} - \bar{x})$$

where  $k$  is the lag or number of events in time,  $c_k$  is the cross-correlation at lag  $k$ ,  $c_0$  is the variance in  $x$ ,  $\bar{x}$  is the set mean, and  $N$  is the number of elements in the set.

By definition, the first autocorrelation coefficient is unity, as the cross-correlation at lag zero is the data variance. For higher-order coefficients, non-zero values indicate the degree of coupling in time. For a zero-mean, random process, all higher-ordered coefficients tend to zero. The first zero crossing, or minimum in the absence of a zero crossing, of the autocorrelation function is taken as the *decorrelation time*, or the number of timesteps on average between totally decoupled events. Peaks beyond the initial dip in the autocorrelation function denote periodicities or strong coupling and are termed here as *correlation*

times. As many control strategies rely on predictions, the dominant timescales as indicated by the autocorrelation function are important indicators of the expected time window for accurate prediction.

### 4.3.2 Mutual information function

Whereas the autocorrelation function is a linear measure, the mutual information function is a nonlinear function of temporal coupling. Mutual information is a development of information theory and is a measure of the average amount of predictability about a future measurement given knowledge of a present measurement. The basis of the mutual information function was first proposed by Shannon [33]. Because mutual information assumes no functional relationship between variables, it is useful for analysis of nonlinear time series.

The entropy or uncertainty  $H$  of a system  $S$  of measurements  $s_i$  is the average amount of information gained about the system by taking the measurement and is defined as

$$H(S) = - \sum_i P_s(s_i) \log P_s(s_i)$$

where  $P_s(s)$  is the probability density at  $s$ . If the logarithm is in base  $e$ , then the unit of  $H$  is in *nats*; if in base two, which is common in information theory, then in *bits*.

Based on the concept of entropy, the *mutual information* of a measure-

ment  $S$  and a measurement  $Q$  separated in time is

$$I(Q, S) = H(Q) + H(S) - H(S, Q)$$

where  $H(Q)$  and  $H(S)$  are the independent uncertainties and  $H(S, Q)$  is the joint uncertainty, defined as

$$H(S, Q) = - \sum_{i,j} P_{sq}(s_i, q_j) \log P_{sq}(s_i, q_j)$$

Practically, the probability functions  $P$  are calculated by binning the data and constructing frequency distributions.

For a random or non-deterministic sequence of measurements, the mutual information function decays to a baseline immediately, i.e., at a lag of one timestep. This behavior may be interpreted as a result that for this random sequence there is no temporal coupling between measurements and that one predicts equally poorly one timestep apart as infinity timesteps apart. In other words, knowledge of present events yields no advantage in guessing future events. Periodic signals have relatively flat mutual information functions, except at lags, such as one quarter of the function period for a sinusoidal wave, in which there is great joint uncertainty in the variable. The mutual information function for a chaotic signal will experience a general decaying trend at high lags with occasional peaks at dominant timescales.

Fraser and Swinney [34] were among the first authors to apply the mutual information function to chaotic time series analysis, and the method has since

achieved widespread usage. Like the autocorrelation function, the mutual information function is useful for control purposes, as it measures the average amount of predictability in a signal.

### 4.3.3 Application to pressure sets

The sequences of peak pressures from the pressure sets were analyzed with autocorrelation and mutual information techniques. For each peak-pressure set, ten surrogate sets in which the order of the elements was shuffled randomly were created to test a null hypothesis that the autocorrelation and mutual information coefficients were insignificant and not the result of the temporal order within the peak-pressure set. Because the data were expected to have weak-to-negligible coupling, a short window of ten engine cycles was used.

Figures 4.4-4.9 display plots of the calculated functions for the peak-pressure sequences and their randomly shuffled surrogates. For visualization, the original-set function points are connected with a cubic spline function, and the mutual information functions are plotted on a base-two logarithmic scale. These plots indicate that most of the peak-pressure sequences exhibit weak or no temporal coupling. Particularly noteworthy are the cases in which the original-set functional values are significantly higher than those for the ten randomly shuffled surrogate sets, as these instances would indicate temporal correlation. The results may be interpreted as either a revelation that there was indeed no

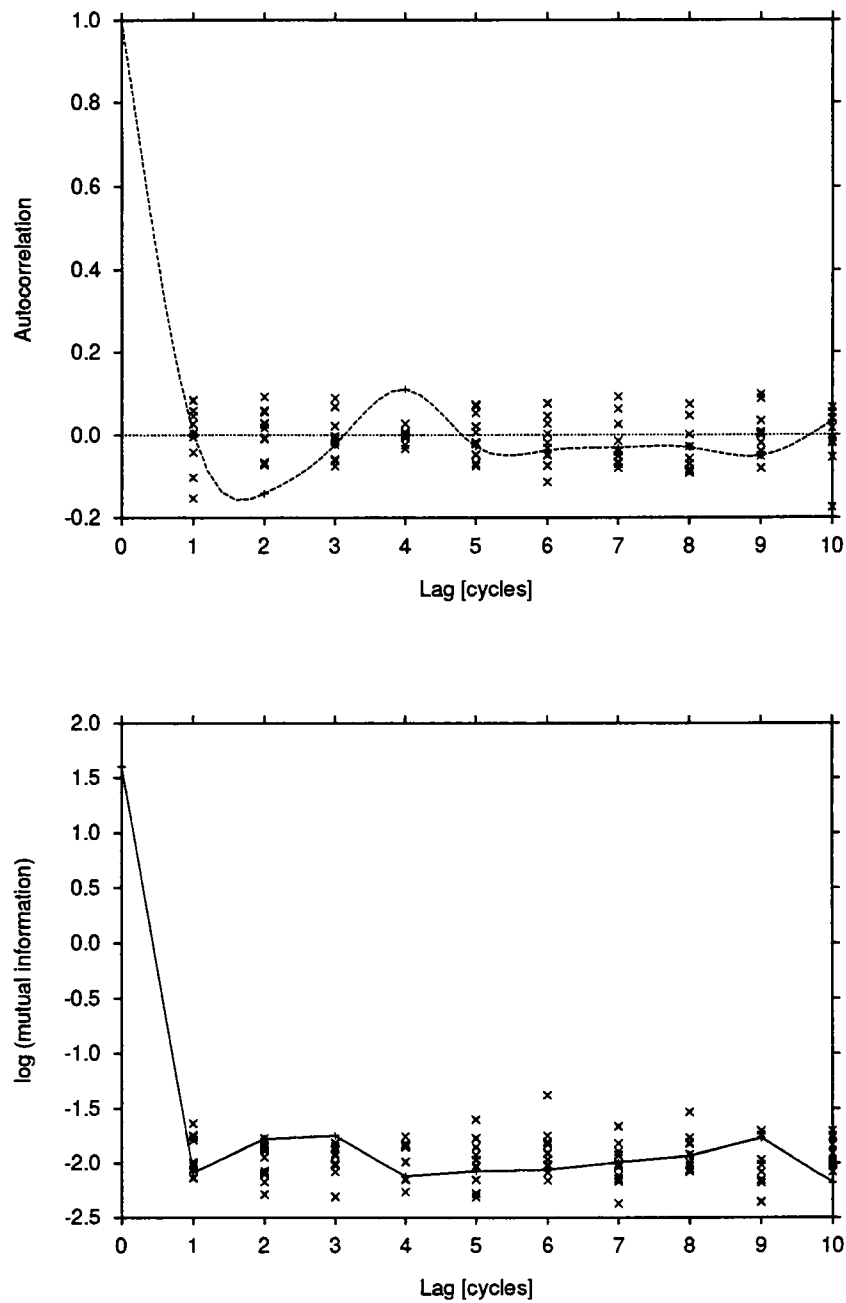


Figure 4.4: Autocorrelation and mutual information functions of peak pressures for FV6SA27. Lines connect values for the original data set, and the unconnected points are for the ten randomly shuffled surrogate data sets.

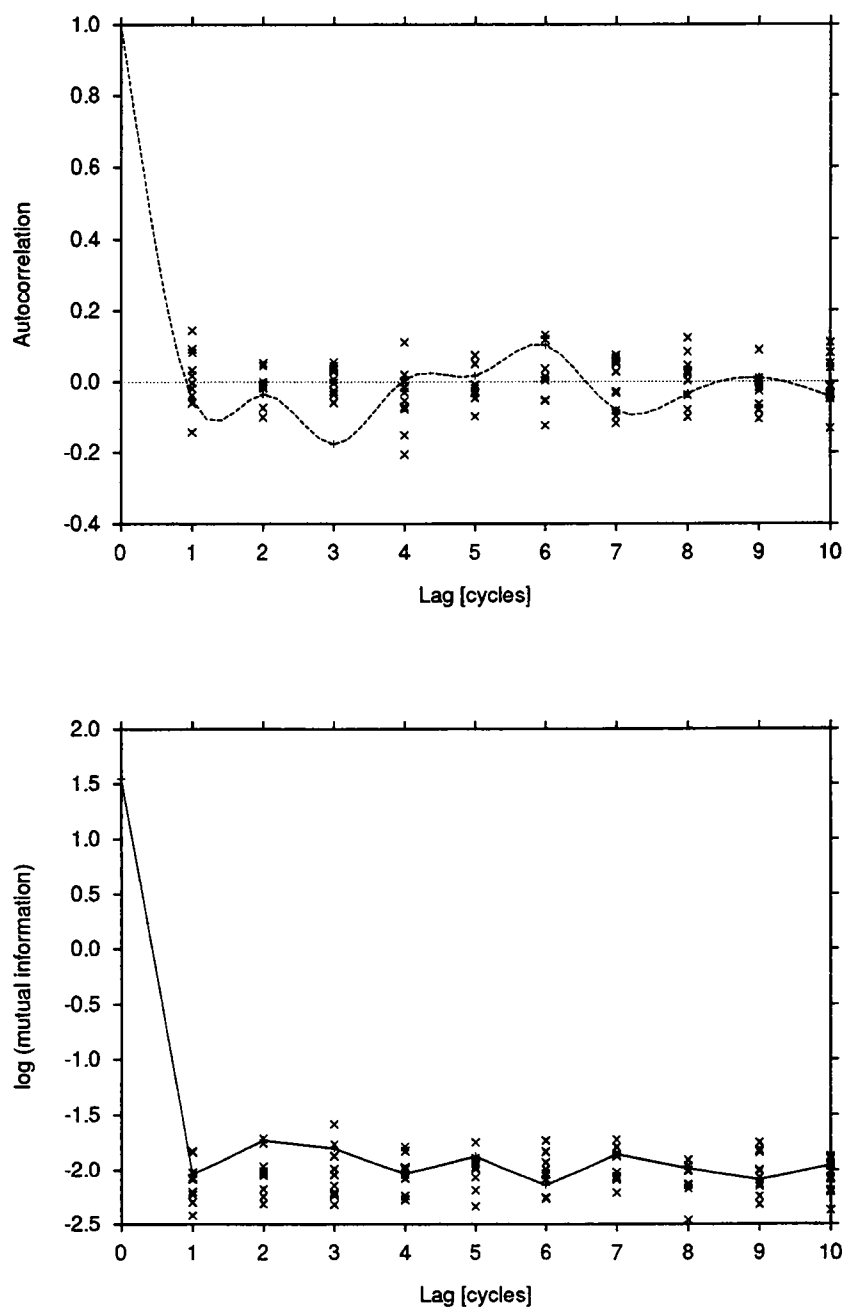


Figure 4.5: Autocorrelation and mutual information functions of peak pressures for FV6SA37. Lines connect values for the original data set, and the unconnected points are for the ten randomly shuffled surrogate data sets.

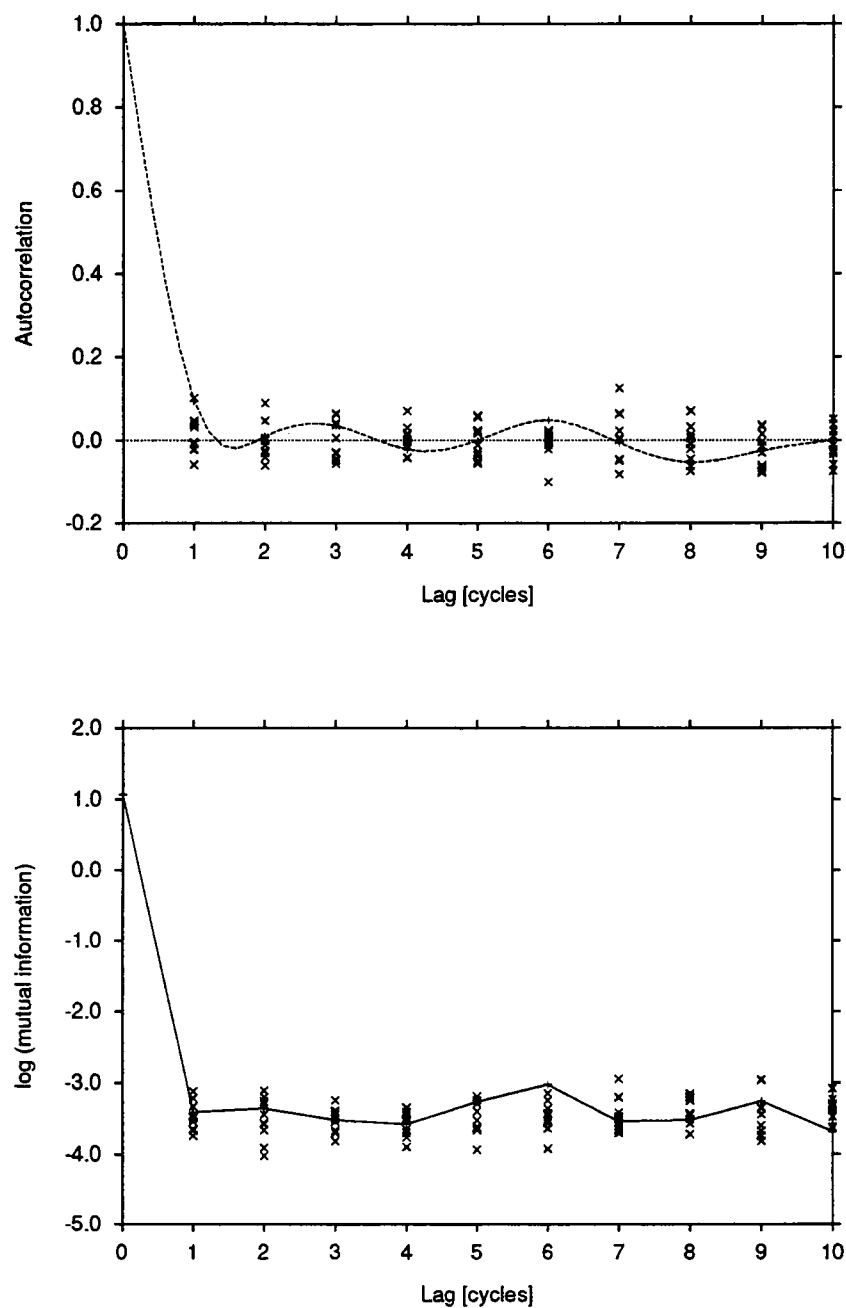


Figure 4.6: Autocorrelation and mutual information functions of peak pressures for STID0629. Lines connect values for the original data set, and the unconnected points are for the ten randomly shuffled surrogate data sets.

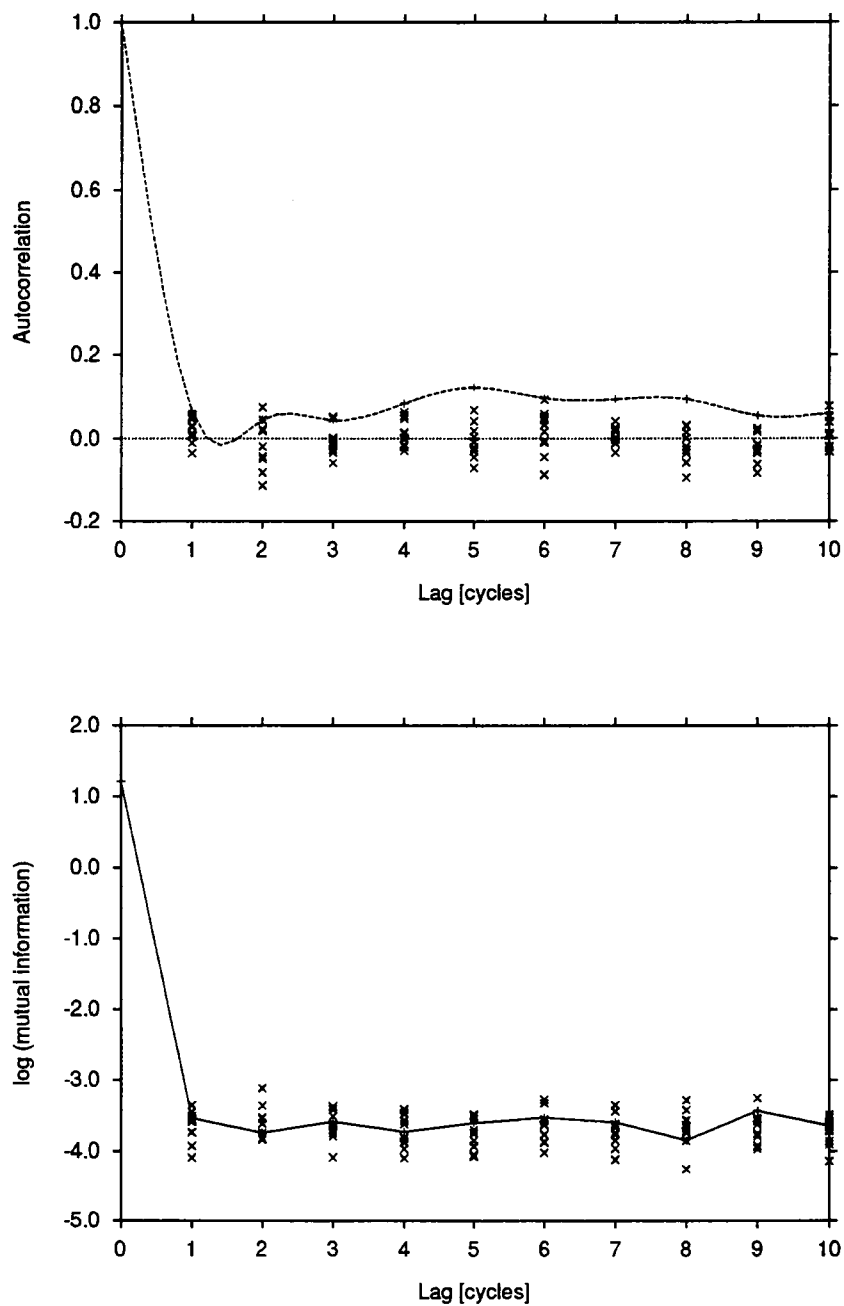


Figure 4.7: Autocorrelation and mutual information functions of peak pressures for STID0630. Lines connect values for the original data set, and the unconnected points are for the ten randomly shuffled surrogate data sets.

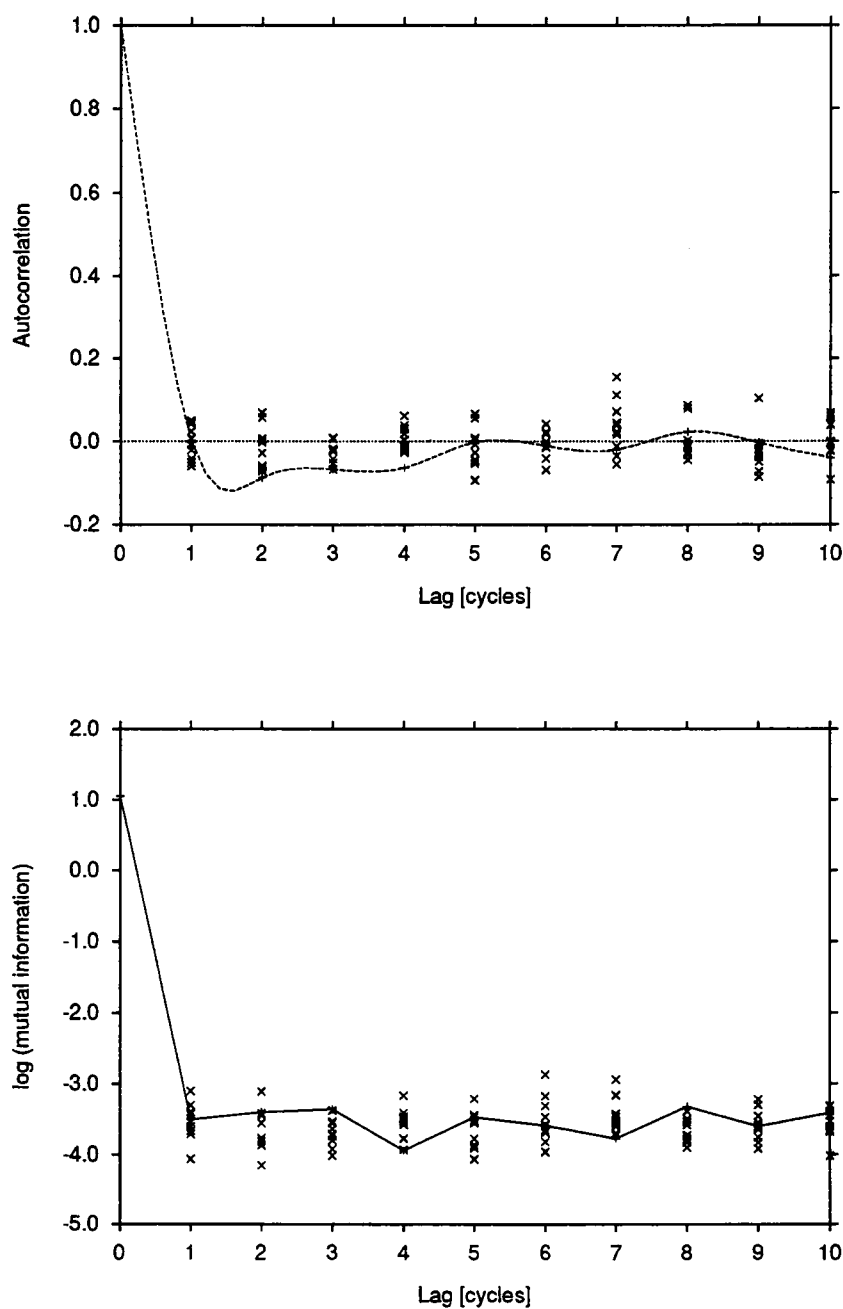


Figure 4.8: Autocorrelation and mutual information functions of peak pressures for LNID0629. Lines connect values for the original data set, and the unconnected points are for the ten randomly shuffled surrogate data sets.

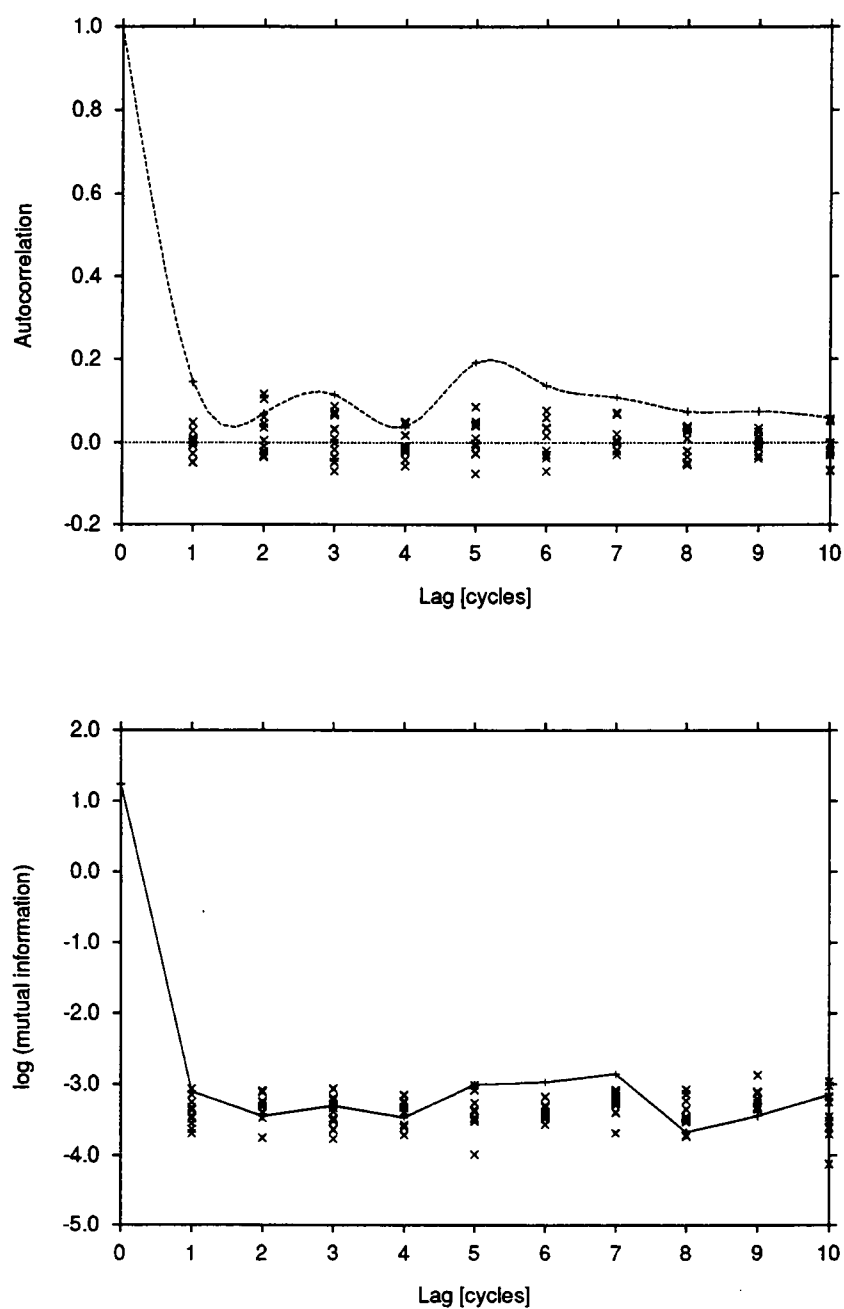


Figure 4.9: Autocorrelation and mutual information functions of peak pressures for LNID0630. Lines connect values for the original data set, and the unconnected points are for the ten randomly shuffled surrogate data sets.

serial correlation in the data or that there were too few data points to resolve properly any intercylic patterns with the two applied statistical functions.

The autocorrelation functions for FV6SA27 in Figure 4.4 and FV6SA37 in Figure 4.5 differ in dominant time scales (two and four cycles for FV6SA27 and three and six cycles for FV6SA37). Such a time-scale shift might be attributed to the changes in spark advance which might have altered memory effects.

#### 4.3.4 Application to speed sets

The speed sets were examined with autocorrelation and mutual information techniques. Inspection of the time series suggested that relatively long lags would be needed in order to observe the dominant time scales in the data.

Figures 4.10-4.15 display plots of the calculated functions. For the sets FI6V1, FI6V2 and FI6V4, the temporal coupling decays quickly. As the mixing-chamber volume is increased from set FI6V1 to FI6V4, the strength of temporal coupling increases, as seen in the increasing magnitudes of the peaks in the plots. The dependence of engine speed on the intake runner volume is reminiscent of the findings of Chew *et al.* [28], who presented a model for combustion variability with a dependence on the runner volume. This model was an adaptation of the standard Helmholtz resonator model of the cylinder volume and hydrodynamic resonance of air oscillations in the intake runner. The Chew model incorporated variable compressibility and viscous damping effects which comprised a nonlinear

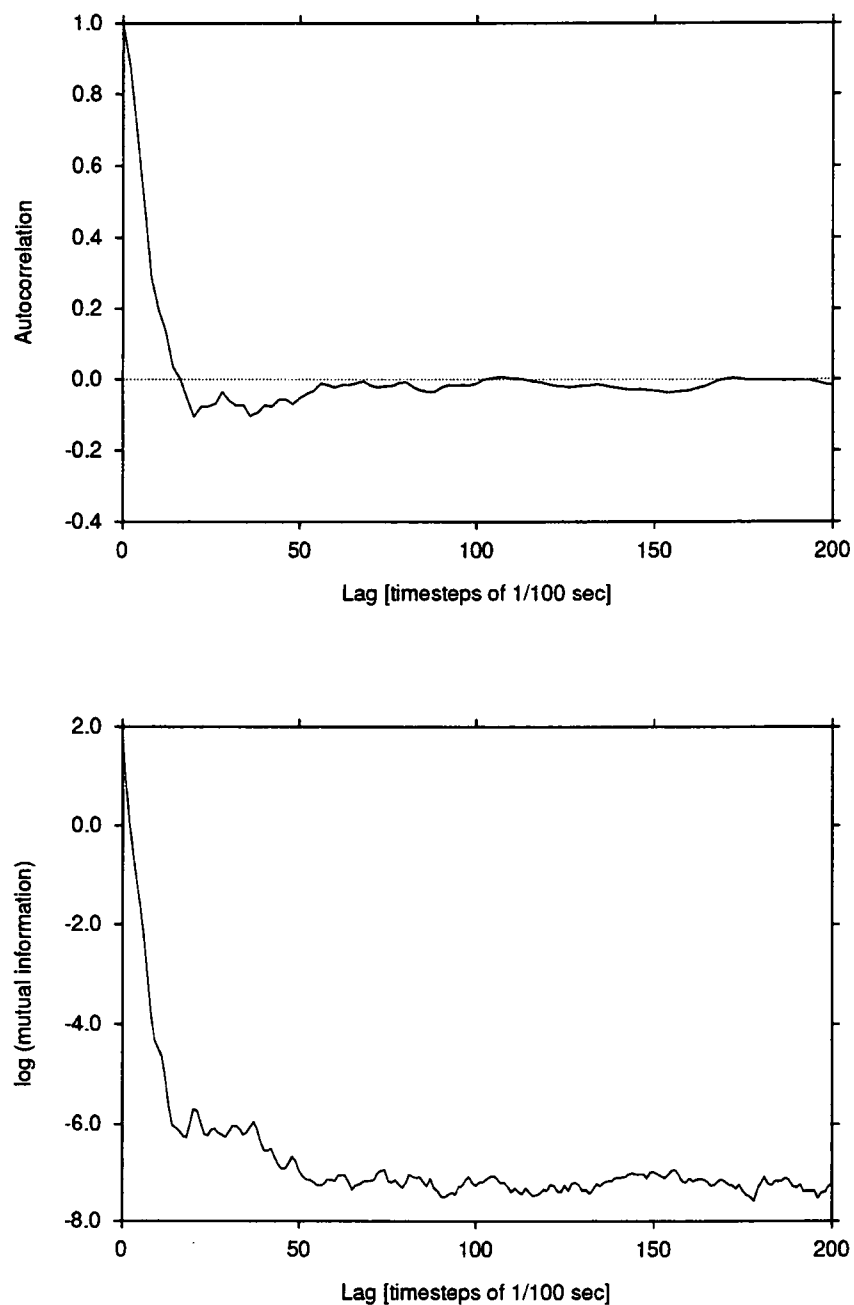


Figure 4.10: Autocorrelation and mutual information functions of engine speed for FI6V1.

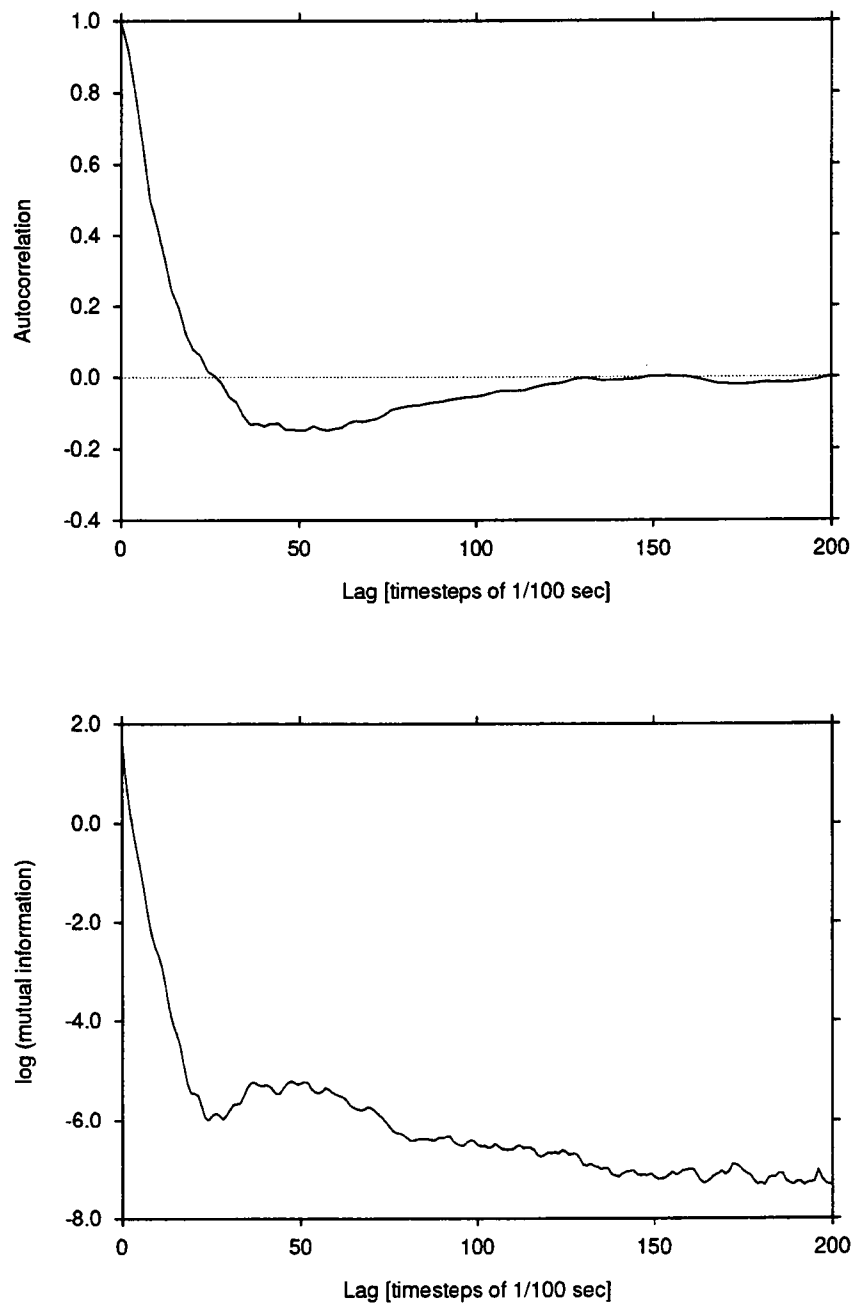


Figure 4.11: Autocorrelation and mutual information functions of engine speed for FI6V2.

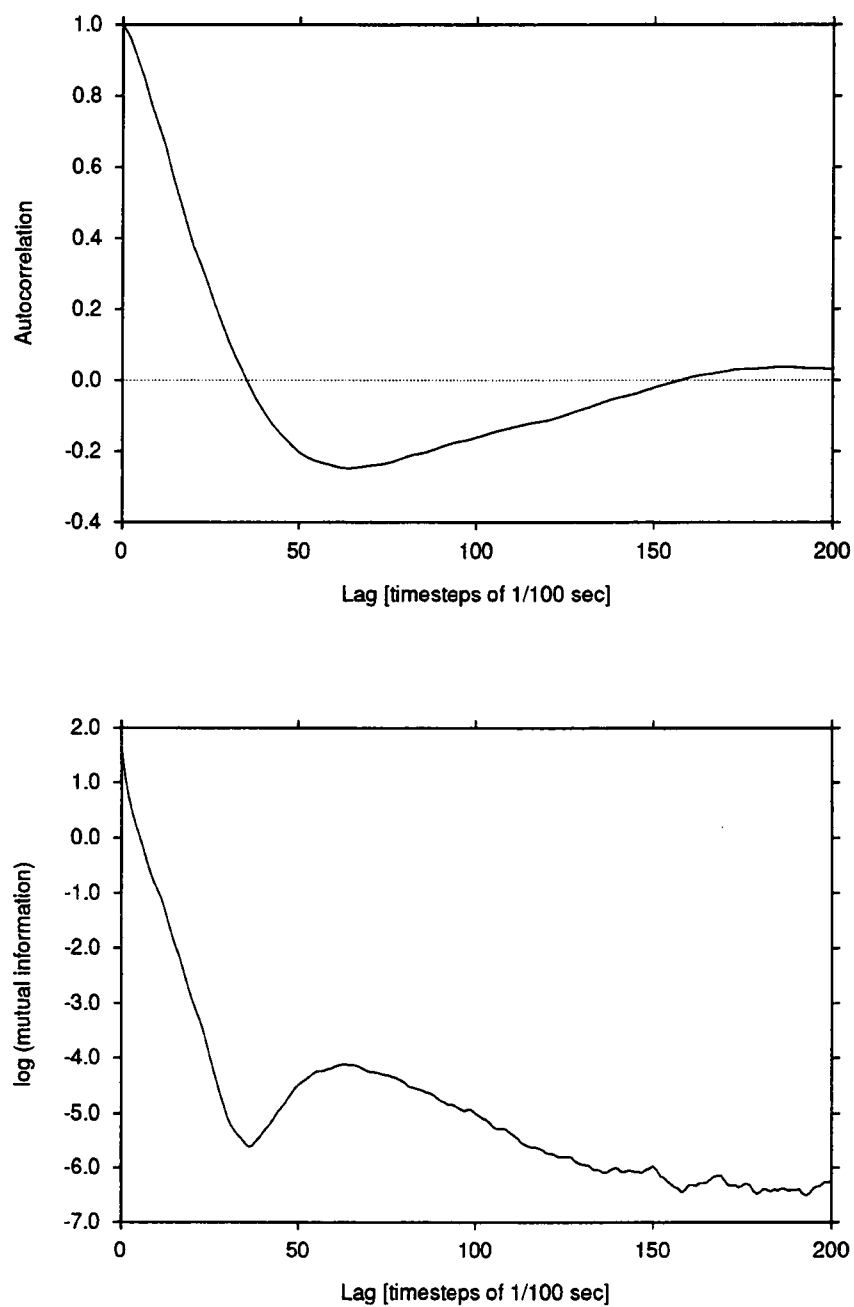


Figure 4.12: Autocorrelation and mutual information functions of engine speed for FI6V4.

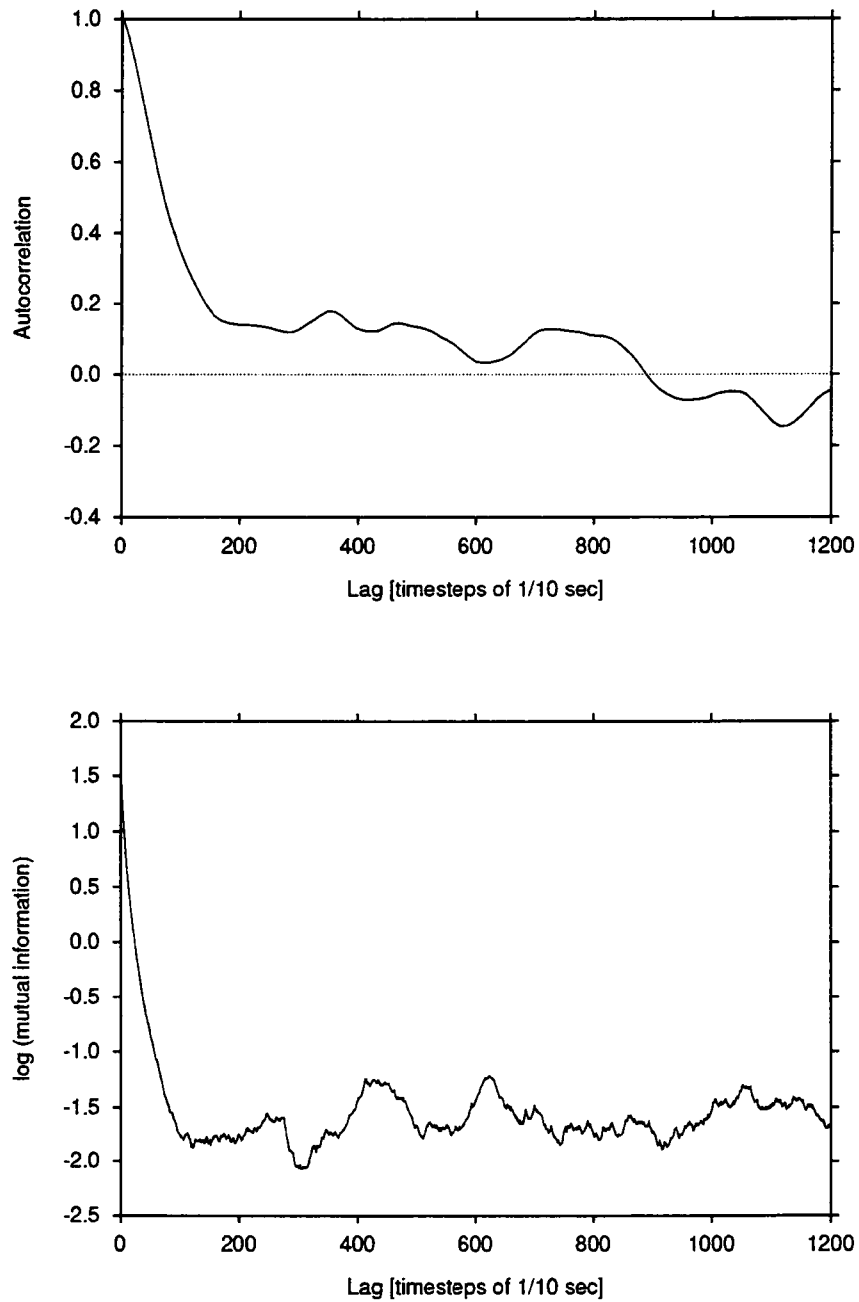


Figure 4.13: Autocorrelation and mutual information functions of engine speed for STID-FCE.

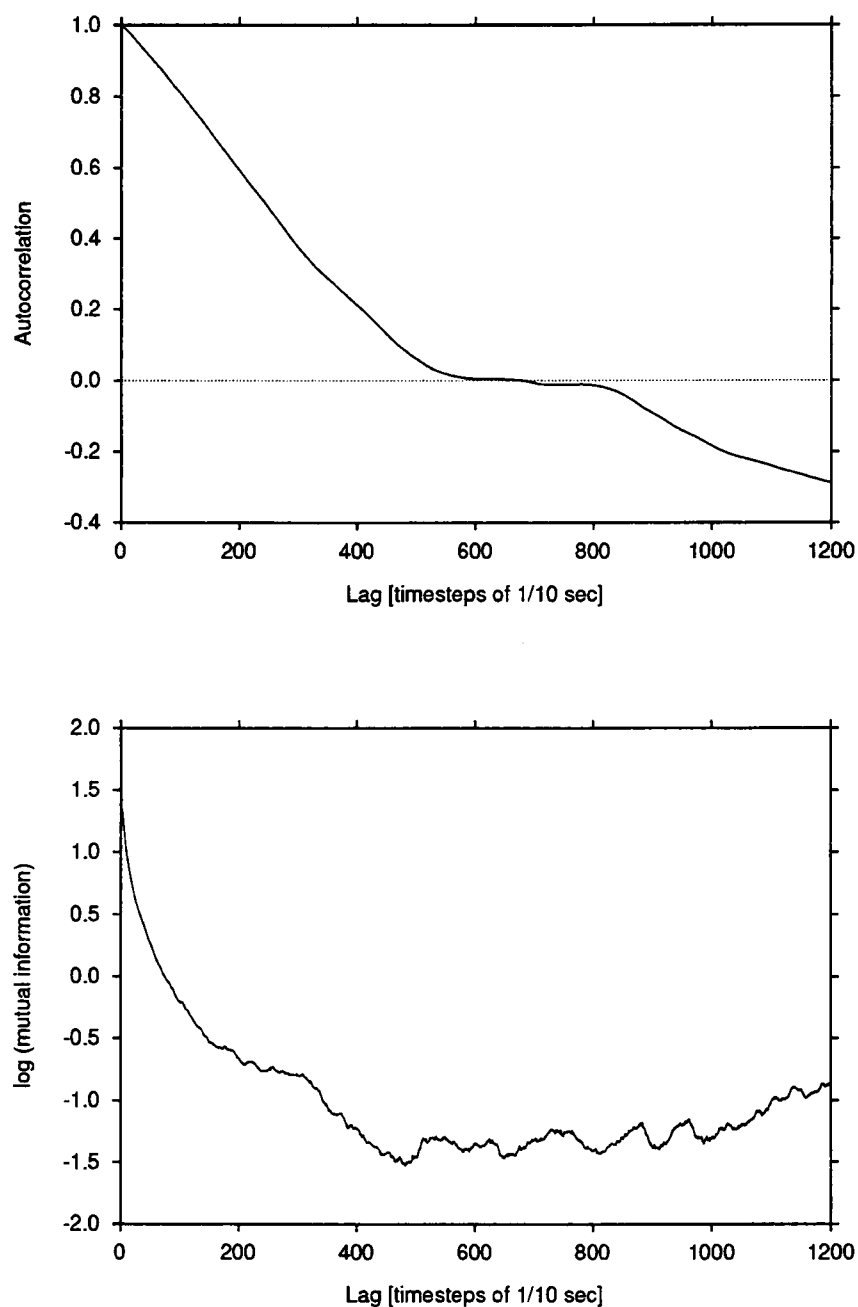


Figure 4.14: Autocorrelation and mutual information functions of engine speed for STID-FCD.

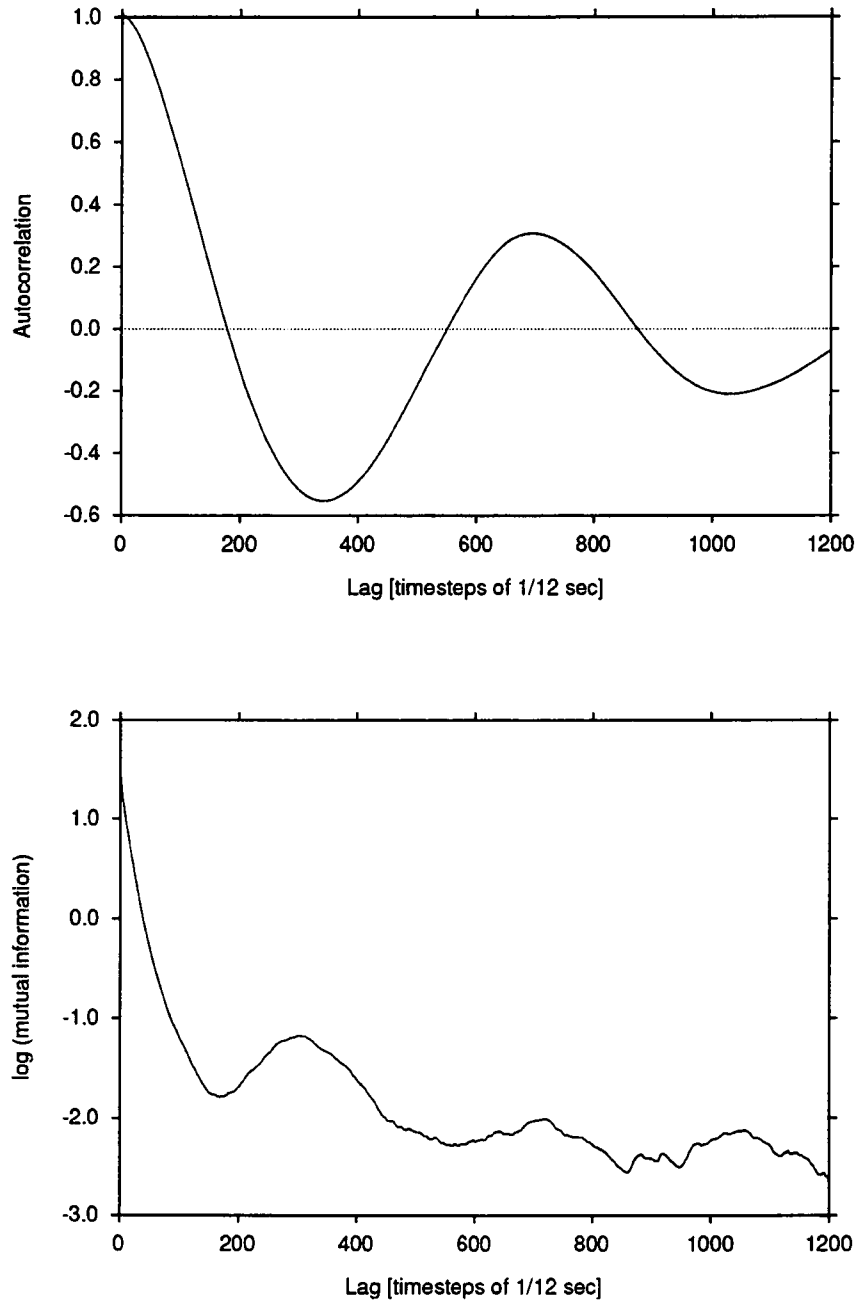


Figure 4.15: Autocorrelation and mutual information functions of engine speed for LNID-FCE.

coupled system which could lead to chaotic behavior.

For the Kohler data, all sets exhibit extremely long decorrelation times, suggesting that a predictive model could be effective over very long time scales.

The autocorrelation and mutual information functions agree in the dominant time scales for the Ford data, as seen in decorrelation times of 30 to 40 timesteps and correlation times of 50 to 60 timesteps. For the Kohler set LNID-FCE (Figure 4.15), the autocorrelation and mutual information time scales are shifted, as the mutual information records significant coupling in shorter periods than the autocorrelation. This disagreement arises from a nonlinear component in the data which the mutual information function observes with a different degree of sensitivity.

## 4.4 Observations of higher dimensionality

An important aspect of deterministic chaos theory has been the development of embedding theory. Although multivariate embedding of data into vector spaces has been employed in classical dynamics for many decades, Takens [35] and Packard *et al.* [36] were the first to demonstrate the utility of time-delay embedding in time series analysis. According to the embedding theorem of Takens, a scalar time series may be embedded into a higher-dimensional vector space, and the resulting set of vectors forms a phase-space trajectory which is geometrically

similar to the state space of the system from which the time series was measured.

In the *method of delays*, a scalar time series

$$x = \{x_1, x_2, x_3, \dots, x_N\}$$

is transformed into a set of vectors of the form

$$\xi_i = \{x_i, x_{i+\tau_E}, \dots, x_{(m-1)\tau_E}\}$$

where  $i$  is the vector index,  $\tau_E$  is the embedding time lag, and  $m$  is the embedding dimension.

In his analysis of complex dynamical systems, Poincaré developed a method of strobing the phase-space trajectory at regular time intervals or at instances in which the trajectory crosses a defined surface in the phase space (e.g., see [37]). These Poincaré sections allow a simplified but multi-dimensional view of data and can be defined in physically meaningful manners. For instance, in engine pressure traces, only portions of the time series in physically significant portions of the engine cycle, such as the combustion and power strokes, may be examined, and the remaining portions of the pressure traces may be ignored. For data-acquisition systems such as the Ford V6 system described in Chapter 3, portions of the engine cycle are unobservable due to hardware limitations. In subsequent sections of this chapter, selectively highlighting portions of the engine data will be referred to as *data windowing*, but these data windows are simply applications of the Poincaré-sectioning techniques arising from classical dynam-

ical theory. Indeed, the standard practice of characterizing the engine cycle with magnitude or timing of peak pressure represents a one-dimensional Poincaré section, whereas the data-windowing techniques employed in this thesis represent multi-dimensional Poincaré sections.

Figure 4.16 is a three-dimensional representation of a phase-space trajectory reconstructed from STID0629. In the plot, the compression stroke is contained within a narrow range with little divergence from cycle to cycle, and pressure traces diverge soon after spark. Following blowdown, the trajectories during the exhaust and intake strokes are limited to the small loops seen in the plot. The plane depicts the surface through which Poincaré sections are made, and the bold lines are three-dimensional projections of 30-dimensional vectors which are formed in the data windowing process. In this case, data windows are defined when the pressure trace crosses zero and contain 30 data points thereafter.

## 4.5 Estimates of Kolmogorov entropy

### 4.5.1 Definition of the estimator

Kolmogorov entropy is another concept derived from information theory and may be regarded as a measure of predictability in a signal or, conversely, as a measure of information loss in a signal. Proposed by Kolmogorov in 1958 [38], the Kolmogorov entropy has become a commonly used metric in chaotic time

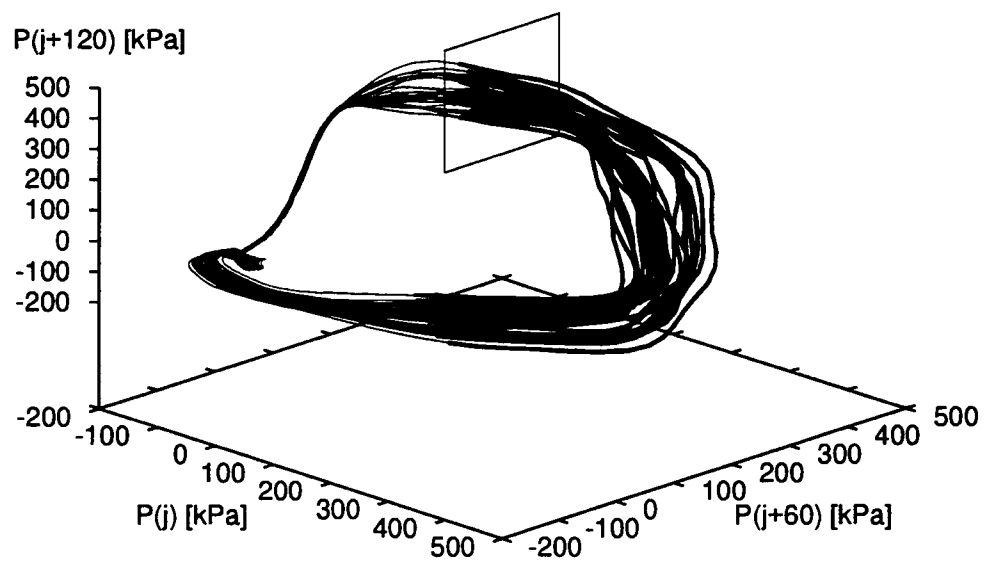


Figure 4.16: Phase-space trajectory reconstructed from STID0629. The plane is the Poincaré sectioning plane, and the bold lines represent data windows.

series analysis (e.g., see Grassberger and Procaccia [39], Schouten *et al.* [40], van der Stappen *et al.* [41]). The Kolmogorov entropy is zero for a periodic or cyclic signal, infinite for a random or stochastic signal, and finite for a deterministically chaotic signal.

Schouten *et al.* [40] proposed a maximum-likelihood estimator of the Kolmogorov entropy, and Finney *et al.* [29] applied a variation of this maximum-likelihood estimator for analysis of engine cylinder pressure traces. The method of Schouten *et al.* is based on the calculation of the average time for initially close trajectory segments to diverge beyond a specified tolerance, or *cutoff length*. Candidate pairs of points are randomly sampled from the time series, and the trajectory segments are tested pointwise using the maximum norm to detect divergence beyond the cutoff length. Candidate pairs which do not diverge beyond the test *segment length* are classified as *qualifying pairs*, and each qualifying pair is tracked until divergence beyond the cutoff length occurs. The *maximum-likelihood estimate* of Kolmogorov entropy is

$$K_{\text{ML}} = -\frac{1}{\tau_s} \log \left[ 1 - \frac{1}{\bar{b}} \right]$$

for

$$\bar{b} = \frac{1}{M} \sum_{1 \leq i \leq M} b_i$$

where  $\tau_s$  is the timestep size,  $b$  is the divergence time in timesteps, and  $M$  is the number of qualifying pairs.

The method employed in the present study is that of Finney *et al.* [29]. In this method, a population of randomly sampled qualifying pairs is identified. At each timestep exceeding the test segment length, the entire population of valid qualifying pairs is tested for divergence beyond the cutoff length, and the number of valid qualifying pairs is decremented accordingly. The pairs are tracked until only a small fraction of the original qualifying-pair population has not diverged. According to Schouten *et al.*, the fraction of undiverged trajectory pairs from the initial sample population decays according to the relation

$$F_\tau \propto \exp(-\kappa\tau_s)$$

where  $\tau_s$  is the timestep size and  $\kappa$  is the decay constant. Solving for this constant yields a maximum-likelihood estimate of the Kolmogorov entropy

$$\hat{K} = -\log F$$

When the logarithm is taken to base two, the unit of  $\hat{K}$  is in *bits*.

Typically, the segment length is specified according to a dominant time scale of the data, conveniently defined as the average time between upward crossings of the data mean, and cutoff length is specified with respect to a characteristic length scale, such as the average absolute deviation about the mean.

### 4.5.2 Data windowing

In a cyclic signal such as engine cylinder pressure, portions of the trajectory are constrained and are less interesting than more variable and physically significant portions. For example, in the analysis of combustion variability, the primary effects are visible in the combustion and power strokes, whereas in the exhaust and intake strokes, the pressure is largely determined by the motion of the piston in the cylinder, resulting in a significantly a more regular signal during these strokes. As estimates of entropy rely on a defined cutoff length as a discriminating metric, it is desirable to examine and define length scales based only on the relevant portions of the cycle.

Finney *et al.* [29] adapted their maximum-likelihood estimator to examine windowed data. The method of windowing as described below actually defines a section of a Poincaré surface in the trajectory. Data windows of constant lengths are defined, and windows are tested pointwise using the maximum norm for divergence beyond the defined cutoff length. Divergence at any point in the windows disqualifies a pair of windows. For an initial population of qualifying window pairs, the fraction of the population not diverged decays as

$$F_W \propto \exp(-\kappa W)$$

where  $W$  denotes a data window. Solving for  $\kappa$  yields the windowed maximum-

likelihood estimate of Kolmogorov entropy

$$\hat{K}_W = -\log F$$

When the logarithm is taken to base two, the unit of  $\hat{K}_W$  is in *bits*.

The Appendix contains a detailed example of the windowed entropy calculation.

### 4.5.3 Normalization of entropy

The method of Schouten *et al.* lacks a practical definition of an infinite value for entropy. They presented a maximum entropy based solely on the sample size of qualifying pairs, but this definition, particularly in the method used in the present study, does not account for the probability distribution of the data.

Daw and Finney [42] have developed a method of normalizing the entropy value against an *ultimate entropy*, which is the entropy of a random signal with the same probability distribution as the original signal. The ultimate entropy accounts for the probability that a certain fraction of point pairs will not diverge purely by chance, *i.e.*, as governed by the probability distribution of the data. This statistic is calculated practically by examining the initial random sample of point pairs at one tracking time unit beyond the test segment length according to

$$C'_{\ell_c} = \frac{1}{M} \sum_{i \neq j} \Theta(\ell_c - \|x_i - x_j\|)$$

where  $\Theta$  is the Heaviside function,  $\ell_c$  is the cutoff length,  $x$  is the data vector, and  $M$  is the number of pairs in the initial random sample. As defined here,  $C$  is a one-dimensional correlation integral. The ultimate entropy then is

$$\hat{K}_{\text{ult}} = -\log C'_{\ell_c}$$

The ultimate entropy provides a practical limit of infinite entropy and can serve as a basis to measure determinism, as defined with the *normalized entropy* or *entropic determinism factor*

$$\langle \hat{K}_\ell \rangle = \frac{\hat{K}}{\hat{K}_{\text{ult}}}$$

The dependence of entropy on cutoff length and segment length will be examined below.

#### 4.5.4 Application to pressure sets

The pressure sets were analyzed with the windowed entropy estimator, and the results are plotted in figures 4.17-4.19. These plots demonstrate the dependence of length scale on the level of determinism, as measured with normalized entropy. The differences in the curves depend on which length scale, relative or absolute, is used. The value of the entropy-length spectrum is that it defines the average degree of predictability which might be achieved to within the specified cutoff length, and this information is important for control purposes.

For the Ford data sets (Figure 4.17), the normalized entropy of FV6SA37

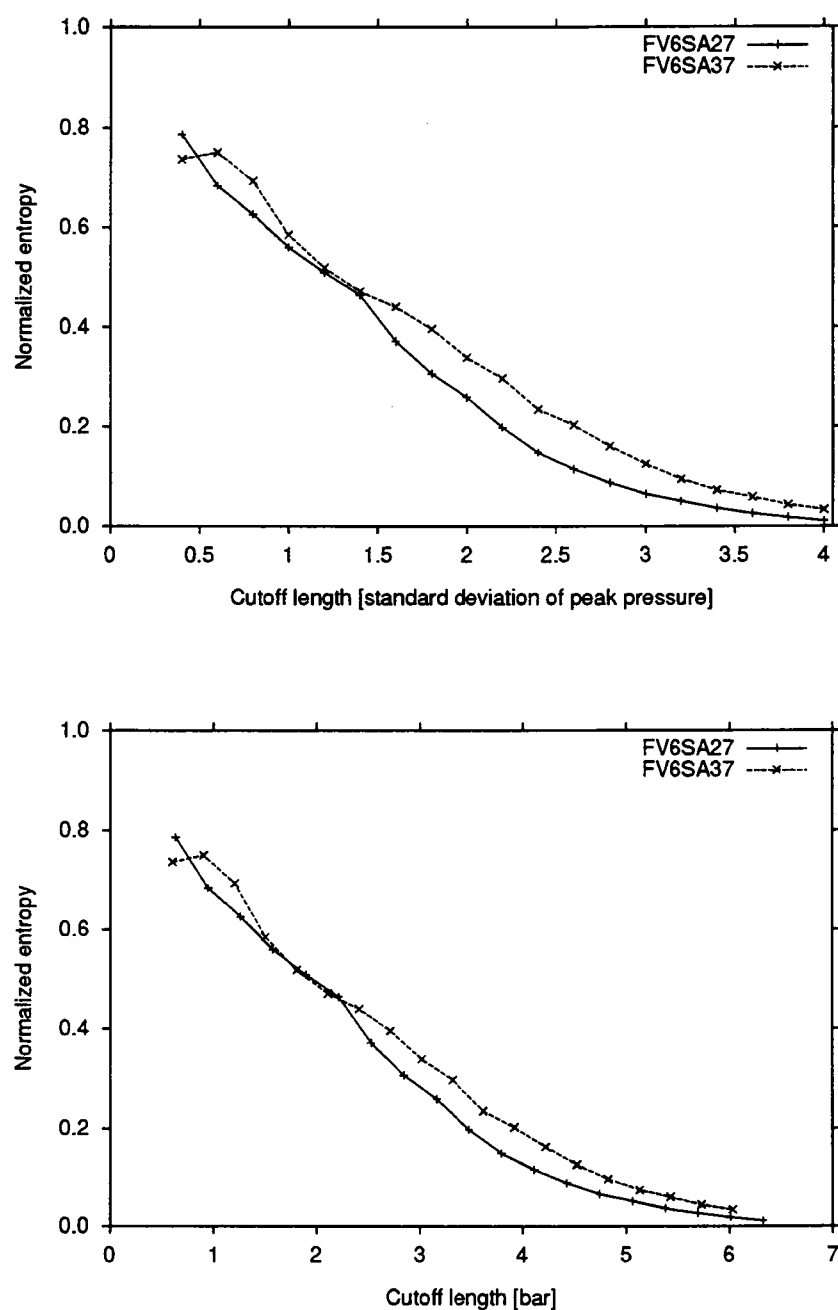


Figure 4.17: Normalized entropy for FV6SA27 and FV6SA37 as a function of normalized (top) and absolute (bottom) cutoff length.

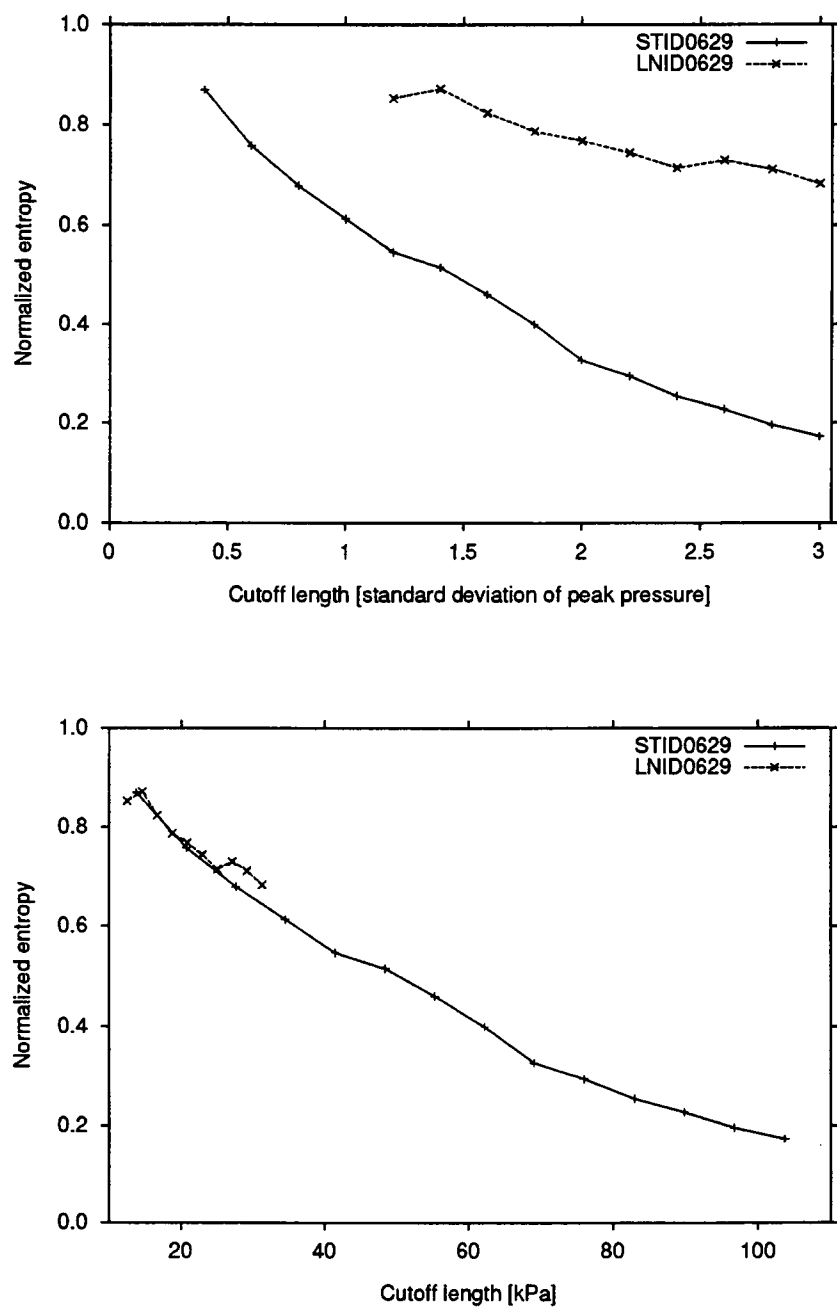


Figure 4.18: Normalized entropy for STID0629 and LNID0629 as a function of normalized (top) and absolute (bottom) cutoff length.

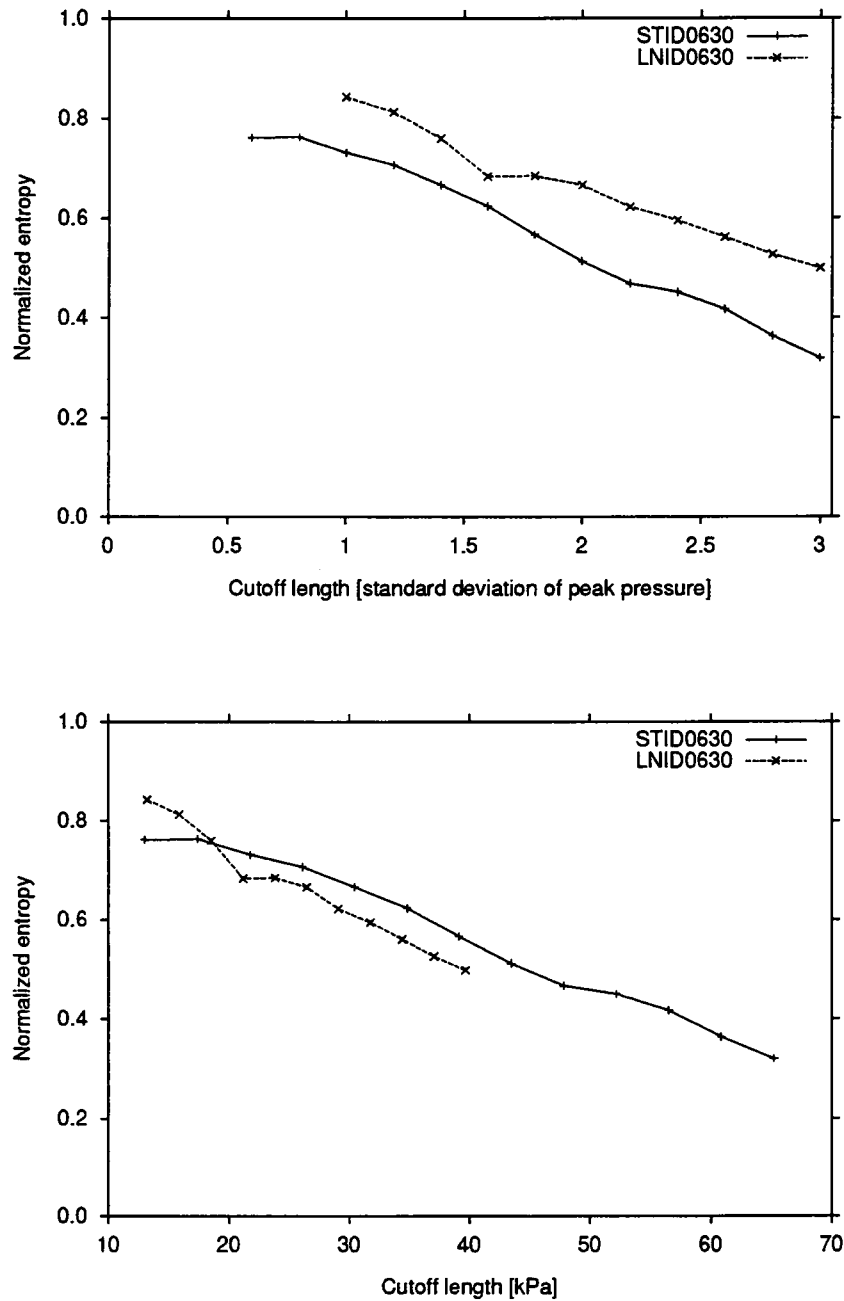


Figure 4.19: Normalized entropy for STID0630 and LNID0630 as a function of normalized (top) and absolute (bottom) cutoff length.

is consistently higher than or equal to that for FV6SA27 over all length scales. This relation suggests that FV6SA37 exhibits lower levels of determinism or predictability than FV6SA27, differences which might be attributed to the change in spark advance and the subsequent effects on engine performance.

For the Kohler data sets (figures 4.18-4.19), the lean conditions, viewed on a relative or normalized length-scale basis, are consistently less predictable than the stoichiometric conditions. Viewed on an absolute length-scale basis, the two conditions exhibit similar levels of predictability, but such comparison is misleading because it does not account for individual data-set variability. The differences between the stoichiometric and lean conditions are significantly more pronounced in the June 29 data (Figure 4.18) than in the June 30 data (Figure 4.19), suggesting that although the conditions from day to day were nominally the same, engine behavior was different.

#### 4.5.5 Application to speed sets

The engine-speed data sets were analyzed with the entropy estimator, and sensitivity to segment and cutoff lengths was examined.

Figure 4.20 shows normalized entropy as a function of segment length for LNID-FCE for a cutoff length of one average absolute deviation about the data-set mean. After three timesteps, the normalized entropy has converged to a plateau, implying that three timesteps is the minimum embedding dimension necessary to

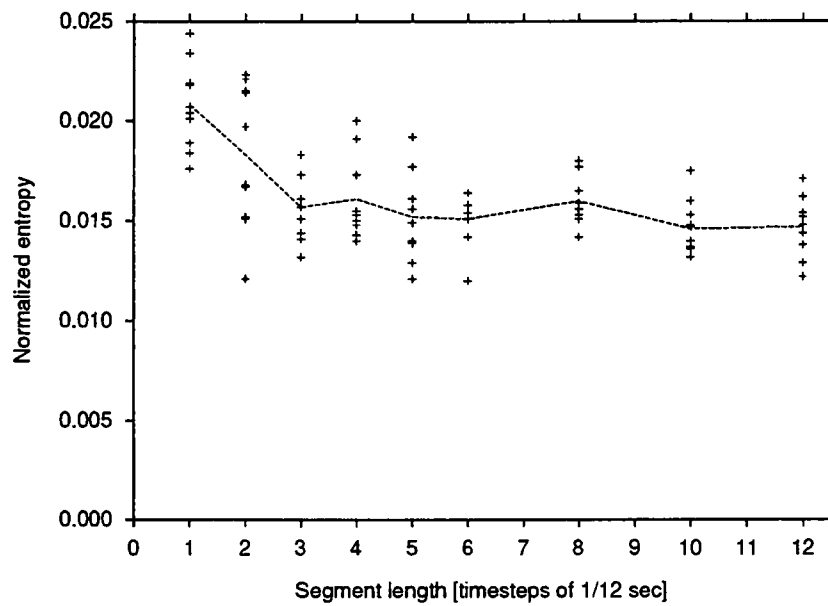


Figure 4.20: Normalized entropy for LNID-FCE as a function of segment length.

The line connects the mean value for ten entropy estimates at each segment length.

obtain the best level of predictability at the observational length scale. For much more stringent tolerances, corresponding to lower cutoff lengths, the minimum segment length required to achieve convergence would be expected to increase; *i.e.*, more information would be required to describe the complexity of the signal.

Figure 4.21 shows normalized entropy as a function of cutoff length for the Ford speed data sets for a segment length of one average cycle time, defined by successive upcrossings of the data-set mean. As is evident in the trends of the curves, increasing the mixing-chamber volume appears to increase the level of determinism or predictability, as FI6V4 has significantly lower values of normalized entropy than FI6V1. All three data sets exhibit similar trends: an initial decrease in entropy as cutoff length is increased and noise effects become less significant, a slight plateau or trough at a narrow range in cutoff lengths, and a rise in entropy as cutoff length is increased toward the limits of the signal. The trough values represent a preferred observational length scale in which to view the signal dynamics, and these minima occur at 1.0 AAD for FI6V1, 1.6 AAD for FI6V2 and 2.0 AAD for FI6V4.

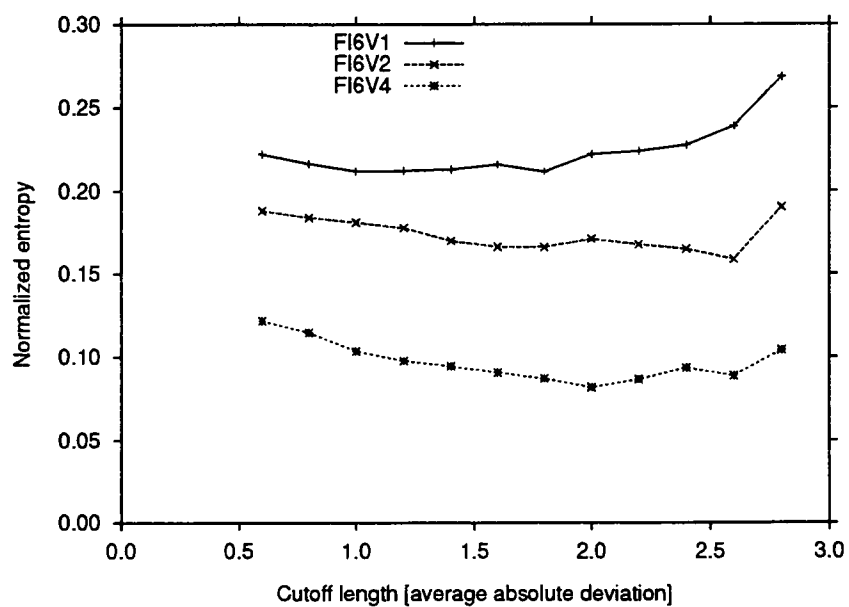


Figure 4.21: Normalized entropy for FI6V1, FI6V2 and FI6V4 as a function of cutoff length.

## 4.6 Estimates of correlation dimension

### 4.6.1 Definitions

The correlation dimension is a measure of signal complexity and may be likened to the degrees of freedom or number of significant components present in the signal. The correlation integral is defined as [43]

$$C_\ell \approx \frac{1}{N^2} \sum_{i \neq j} \Theta(\ell - \|x_i - x_j\|)$$

where  $\Theta$  is the Heaviside function,  $x$  are multi-dimensional vectors,  $\ell$  is a tolerance or cutoff length, and  $N$  is the number of points in the data set. As  $\ell \rightarrow 0$  and  $N \rightarrow \infty$ , the correlation integral scales as

$$C_\ell \approx \ell^D$$

where  $D$  is the *correlation dimension*.

Schouten *et al.* [44] presented a maximum-likelihood estimate of the correlation dimension. Their method approximates the correlation integral by examining a population of randomly sampled point pairs and counting the number of pairs separated by a distance less than  $\ell$  for various  $\ell \leq \ell_0$ , where  $\ell_0$  is the *upper cutoff length*. By choosing a small upper cutoff length, the scaling relation of  $\ell^D$  should be maintained. The correlation dimension is then the slope of  $\log C_\ell$  on a full logarithmic scale.

The *maximum-likelihood estimate* of the correlation dimension is given as

$$D_{\text{ML}} = \left[ -\frac{1}{M} \sum_{1 \leq i \leq M} \log r_i \right]^{-1}$$

where  $r_i$  is the maximum separation normalized to  $\ell_0$  for a population of  $M$  qualifying pairs. This estimate of correlation dimension is much less computationally intensive than the estimate of Grassberger and Procaccia but for noisy signals tends to overestimate the actual correlation dimension of the same noiseless signal.

For the present study, a windowed version of the  $D_{\text{ML}}$  was developed for analysis of engine pressure traces. The data windows are defined as described in the previous section on entropy estimates. Windowing the data reduces the  $D_{\text{ML}}$  significantly, and this reduction should be accounted for when interpreting results or comparing with non-windowed data.

#### 4.6.2 Application to pressure sets

The pressure sets were analyzed with the windowed dimension estimator, and the results are plotted in figures 4.22-4.24. The plots exhibit strong dependence on length scale. There are two significant length-scale regions: at lower cutoff lengths (less than one standard deviation of peak pressures), the increasing dimensionality as cutoff length is decreased is a consequence of dynamic noise effects; at higher cutoff lengths, there is a gradual decrease in signal complexity

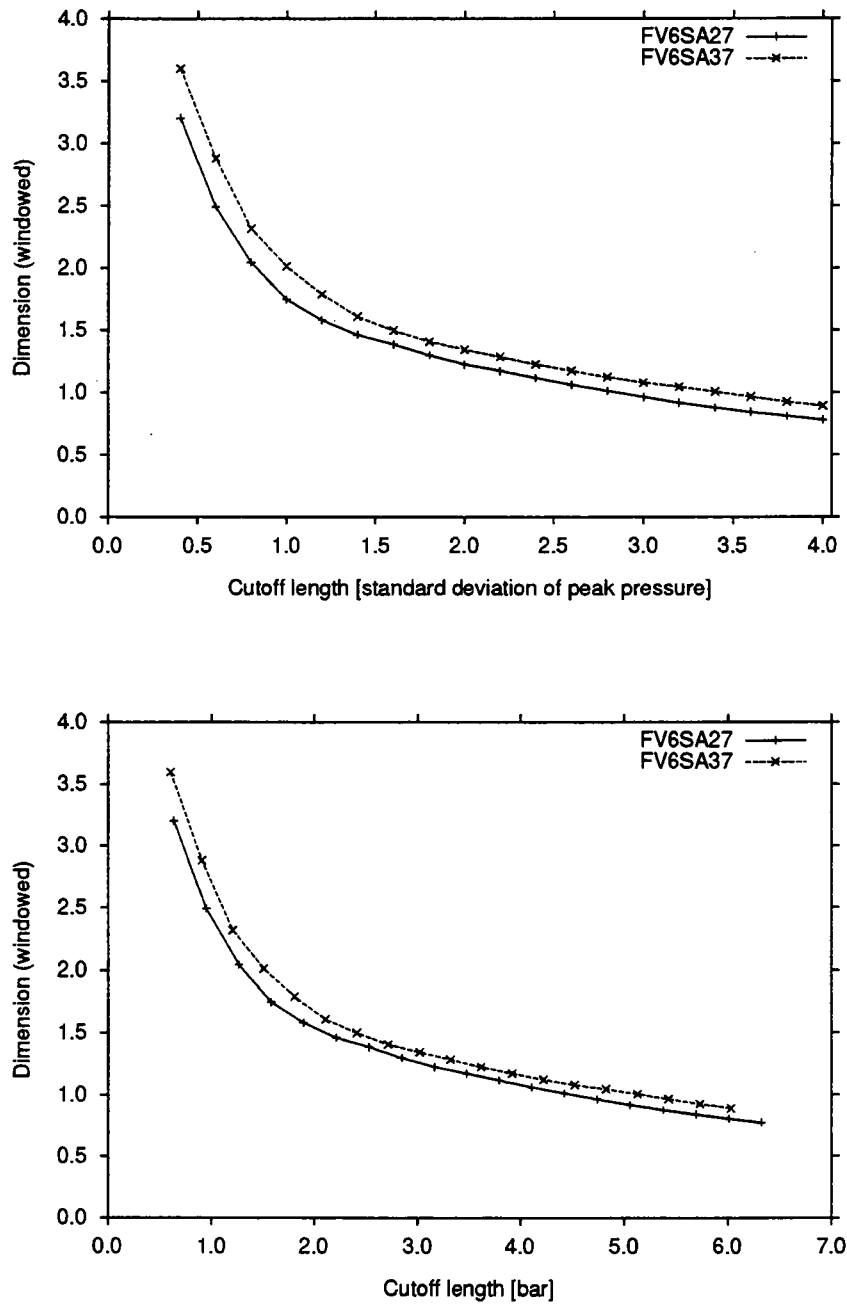


Figure 4.22: Dimension for FV6SA27 and FV6SA37 as a function of normalized (top) and absolute (bottom) cutoff length.

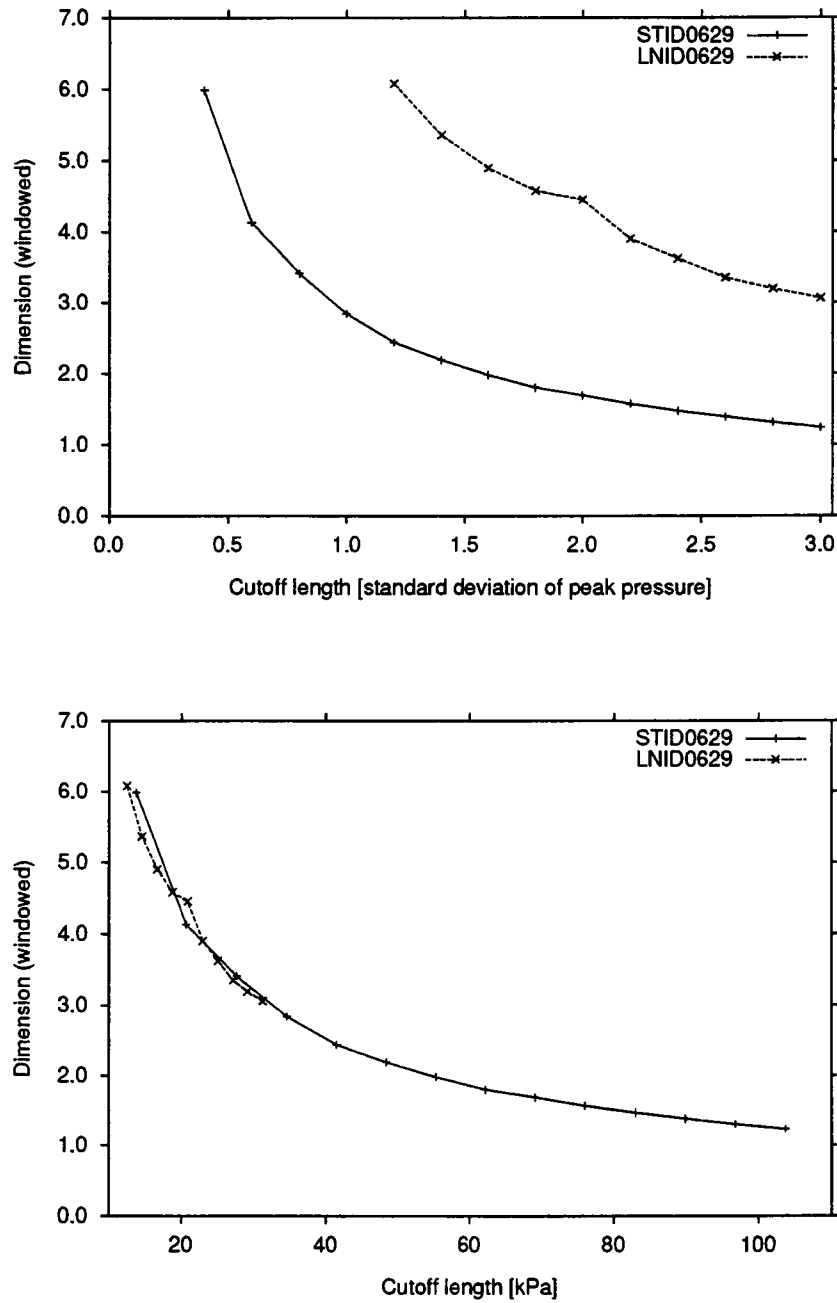


Figure 4.23: Dimension for STID0629 and LNID0629 as a function of normalized (top) and absolute (bottom) cutoff length.

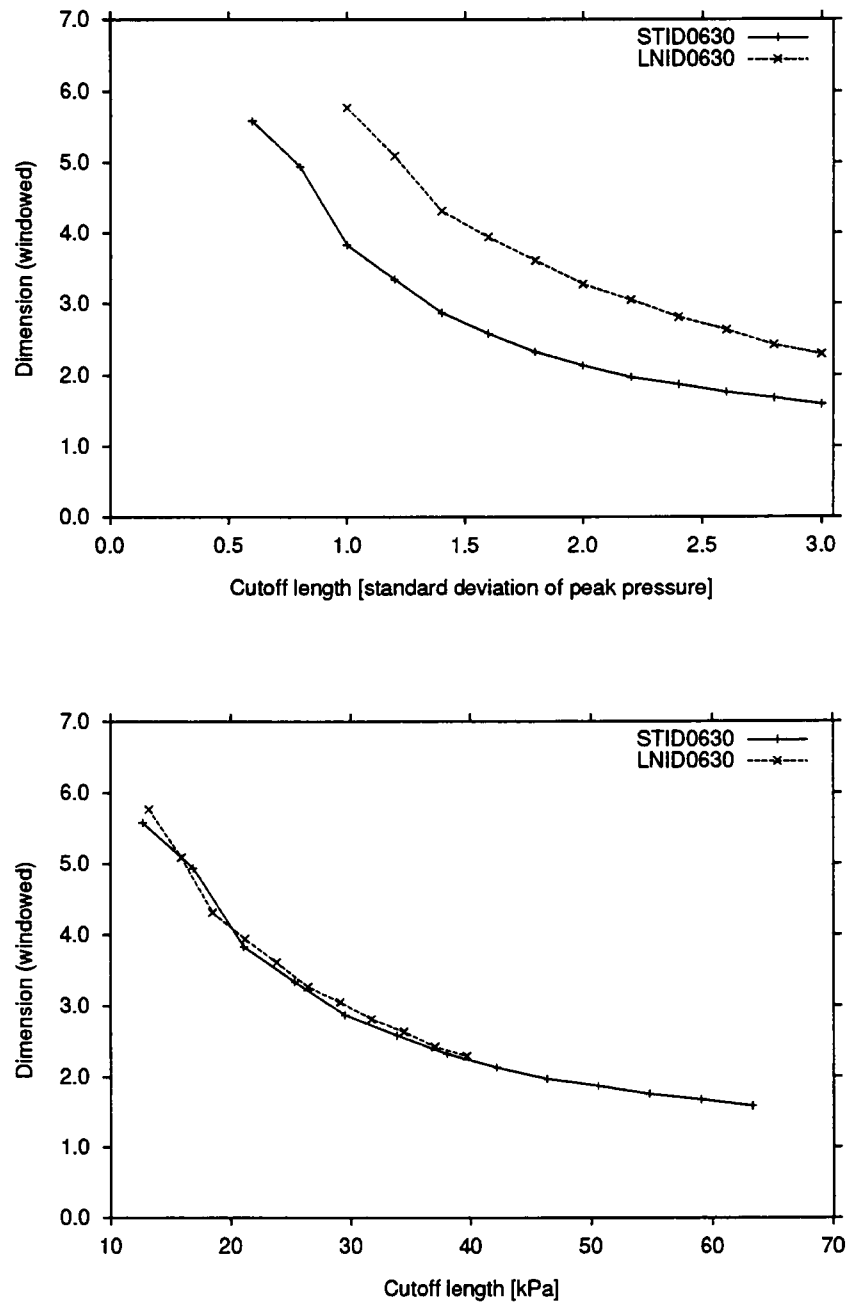


Figure 4.24: Dimension for STID0630 and LNID0630 as a function of normalized (top) and absolute (bottom) cutoff length.

as the observational scale increases.

For the Ford data (Figure 4.22), the signal from FV6SA37 is consistently more complex and exhibits higher dimensionality than that from FV6SA27. This result supports the relation between the two sets as seen with normalized entropy (Figure 4.17). For the Kohler data, the differences between the stoichiometric and lean conditions are significantly more pronounced in the June 29 data (Figure 4.23) than in the June 30 data (Figure 4.24), suggesting that although the conditions from day to day were nominally the same, engine behavior was different. These differences were also seen in the normalized entropy (figures 4.18-4.19), suggesting a complementary relationship of predictability and signal complexity.

In addition to exhibiting a strong dependence of cutoff length, dimensionality depends on the segment or window length, both of which define the embedding dimension of the trajectory segment of observation. Figure 4.25 shows dimension versus cutoff length for several data windows centered on TDC of the combustion stroke using an upper cutoff length of one average absolute deviation about the mean in the windowed data subset. The dimension of the trace converges at embedding window of 40 CAD, and in all converged embeddings a dimension of 4.2 at  $\log \ell_0 = -0.8$  is evident. This convergence suggests that a window length of 40 CAD is sufficient to observe the inherent low-dimensional behavior of the pressure-trace variations.

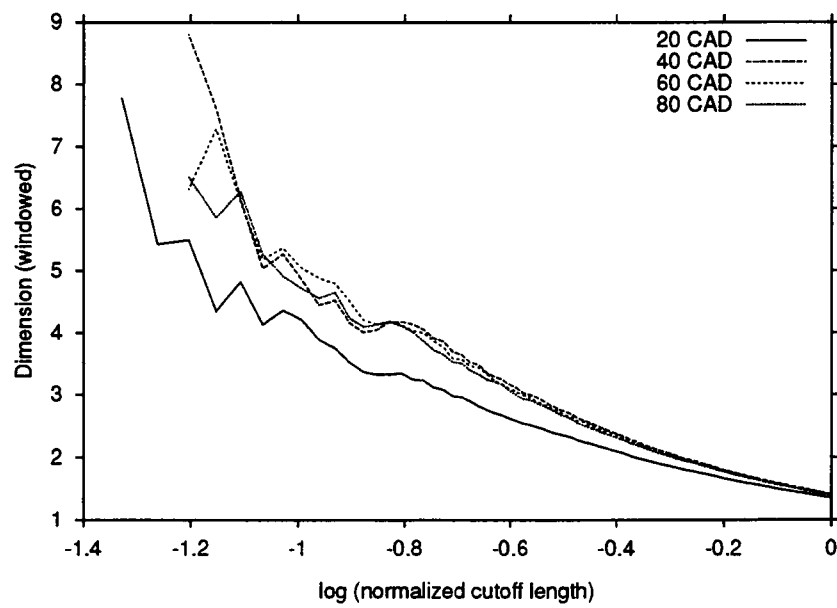


Figure 4.25: Dimension for FV6SA27 as a function of segment length.

### 4.6.3 Application to speed sets

The Ford speed data sets were analyzed with the maximum-likelihood estimator of correlation dimension, and the results are plotted in Figure 4.26. The curves for FI6V1 and FI6V2 are distinct from and higher in value than that of FI6V4, implying that the engine-speed signal becomes less complex as the mixing-chamber volume is increased. If viewed on an absolute length scale, then the curve for FI6V4 exhibits the highest dimensionality. Such a viewpoint is misleading because it does not account for the fact that FI6V4 has larger-magnitude oscillations and that this scale shift *per se* is not the primary indicator of increased complexity.

## 4.7 Tests for nonlinear structure

A problem with analysis of signals from complex systems is characterization of the nature of the complexity. For instance, many statistical tests do not indicate whether signals arise from chaos, nonchaotic nonlinear determinism, linearly correlated noise, or a combination of these sources. In the previous sections of this chapter, surrogate data sets for peak-pressure sequences were created by randomly shuffling the order of the sets in order to test a null hypothesis that the data were random events with no temporal correlation. This null hypothesis was tested by comparing the surrogate sets with the original sets using selec-

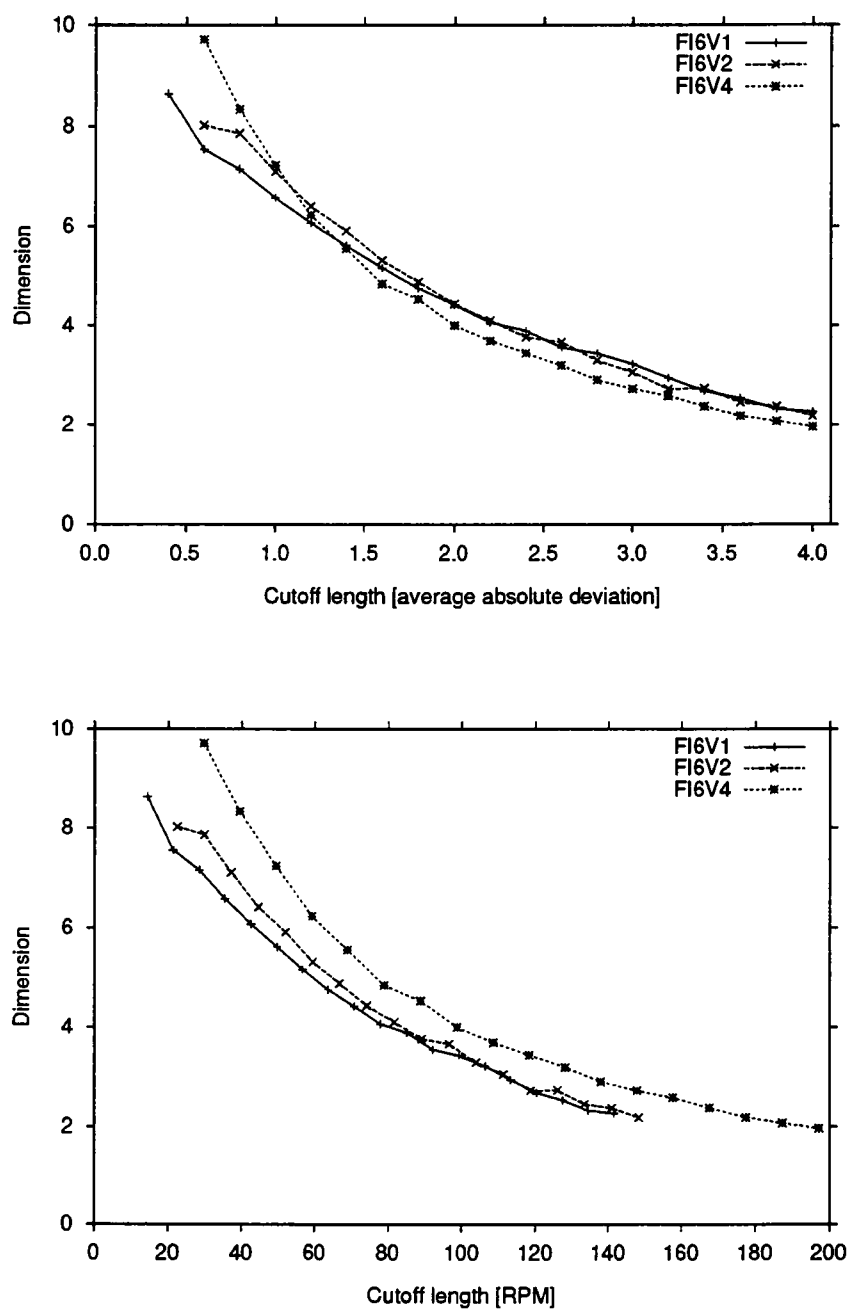


Figure 4.26: Dimension for FI6V1, FI6V2 and FI6V4 as a function of normalized (top) and absolute (bottom) cutoff length.

ted discriminating statistics designed to measure temporal coupling, namely the autocorrelation and mutual information functions.

Whereas there exists no direct test of whether data are chaotic in nature, a commonly employed technique, the so-called *method of surrogate data*, can provide a statistical means to distinguish nonlinear, either chaotic or nonchaotic, and random effects in data. This section describes implementation of the method of Fourier phase-randomization as described by Theiler *et al.* [45, 46]. Other authors [47] have suggested alternatives to the Fourier-transform methods presented by Theiler, whereas others have highlighted the limitations of phase-randomization surrogates [48].

In the Fourier phase-randomization method, time-domain data first are transformed into the frequency domain with a forward Fourier transform, and the phases of the frequency-domain data are scrambled randomly before the data are transformed back into the time domain with a reverse Fourier transform. The method used in this thesis employs non-windowing Fourier transforms and rescales the surrogate data sets, which have a gaussian distribution after the reverse Fourier transform, to have the same distribution as the original data set by using rank substitution with the original set. The resulting surrogate set has the same distribution and autocorrelation as the original set and is suitable to test the null hypothesis that structure in the original data arises from a linear stochastic process or may be described as linearly correlated noise. In other words, if the

original set had been random, then scrambling its temporal sequence would have no effect on appropriate statistical descriptions of the data.

Theiler *et al.* suggest several discriminating statistics, among which are correlation dimension, forecasting error, Lyapunov exponents and comparisons of correlation integrals, of the original and surrogate data sets. Any significant difference between the original and surrogate sets would serve to reject the null hypothesis and to suggest that the data contain nonlinear structure, which for a time series implies determinism. Whereas usage of several discriminating statistics is prudent practice in order to help compensate for the sensitivity of any one test to effects of noise or paucity of data, only one test, comparison of correlation integrals using the Kolmogorov-Smirnov test, will be used in this thesis as an example of how the method of surrogate data could be implemented on engine data.

The three speed data sets from the Kohler were chosen as example time series. The original time series were smoothed with a 20-point moving average filter before surrogates were generated in order to reduce resolution of high-frequency events which were considered to be insignificant compared with the longer time-scale events which were viewed to dominate the engine-speed fluctuations. From the first 4096 points of the sets STID-FCE and STID-FCD, ten surrogate sets of 4096 points were generated for each original set, whereas ten surrogate sets of 16384 points were generated from the first 16384 points of the

set LNID-FCE.

Figure 4.27 displays the filtered original time series from LNID-FCE and a typical surrogate set. The surrogate set visibly retains many of the features of the original set, including similar magnitudes and time scales of speed oscillations.

The discriminating statistic used to compare the original and surrogate sets was chosen as the Kolmogorov-Smirnov statistic, such as described in [49], of the correlation integrals of any two test sets. The correlation integral for each data set was calculated by choosing 1000 randomly sampled point pairs and tabulating the maximum norm of each point pair. For the sets STID-FCE and STID-FCD and their corresponding surrogates, each point was embedded in 200 dimensions with a one timestep lag, whereas for the set LNID-FCE and its surrogates, each point was embedded in 240 dimensions with a one-timestep lag. In the terminology of the preceding sections of this chapter, the embedding dimension is the segment length.

Table 4.2 summarizes the results of multiple tests of the discriminating statistic for the three original sets and their corresponding surrogates. The tabulated values are the number of times out of ten trial calculations of the Kolmogorov-Smirnov statistic between the correlation integrals that the null hypothesis that the two correlation integrals were the same was rejected at a 95 percent confidence level. As expected, all of the surrogate data sets were clearly different from the original data sets, implying that the original sets contained structure

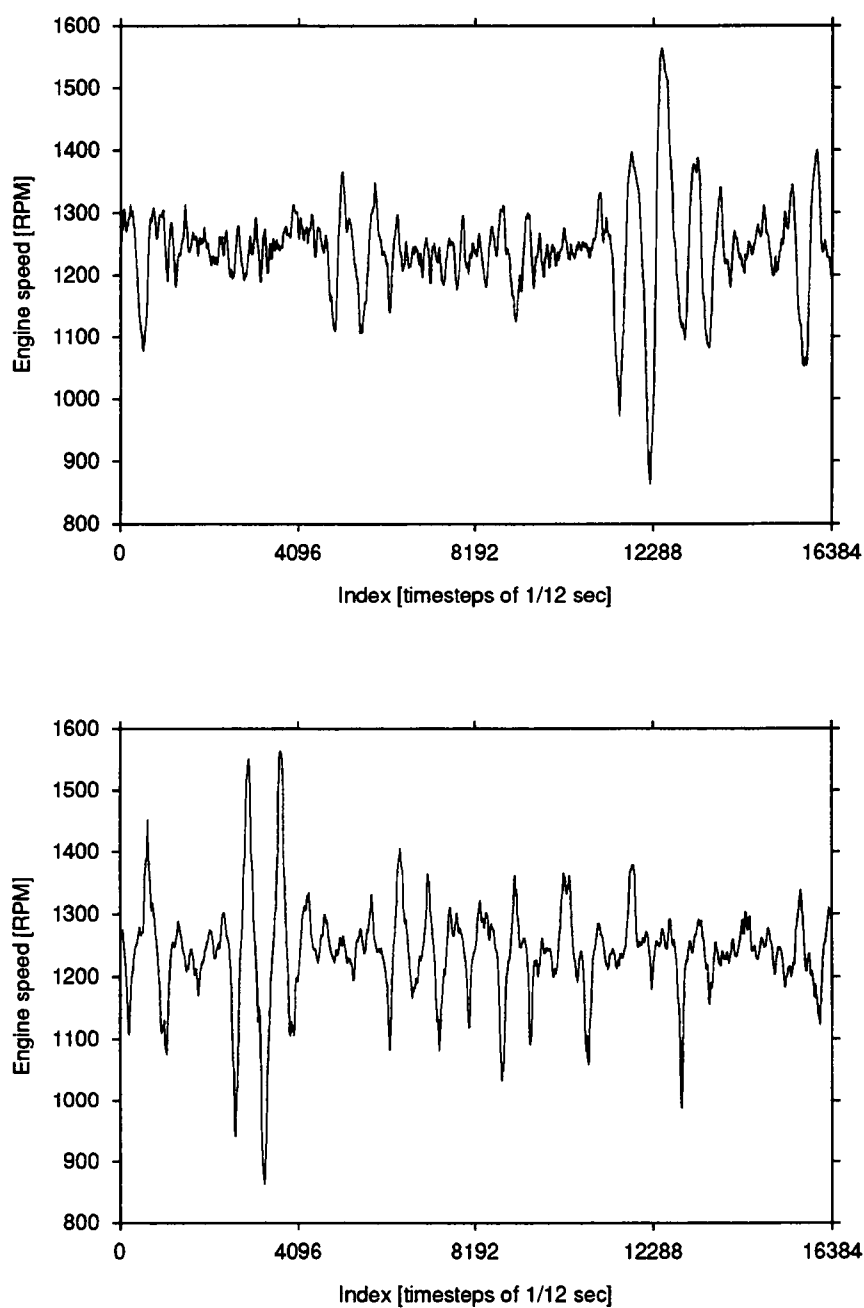


Figure 4.27: Plot of LNID-FCE (top) and a typical surrogate data set (bottom).

Table 4.2: Summary of comparisons between original and surrogate speed data sets.

| Test set    | STID-FCE         |              | STID-FCD         |              | LNID-FCE         |              |
|-------------|------------------|--------------|------------------|--------------|------------------|--------------|
|             | Against original | Against self | Against original | Against self | Against original | Against self |
| Original    | —                | 0            | —                | 1            | —                | 0            |
| Surrogate 0 | 10               | 0            | 10               | 0            | 10               | 0            |
| Surrogate 1 | 10               | 1            | 10               | 1            | 10               | 0            |
| Surrogate 2 | 10               | 0            | 10               | 0            | 10               | 0            |
| Surrogate 3 | 10               | 0            | 10               | 2            | 10               | 0            |
| Surrogate 4 | 10               | 0            | 10               | 2            | 10               | 0            |
| Surrogate 5 | 10               | 0            | 10               | 0            | 10               | 1            |
| Surrogate 6 | 10               | 0            | 10               | 3            | 10               | 0            |
| Surrogate 7 | 10               | 0            | 10               | 0            | 10               | 0            |
| Surrogate 8 | 10               | 0            | 10               | 0            | 10               | 0            |
| Surrogate 9 | 10               | 0            | 10               | 0            | 10               | 3            |

that was destroyed in the scrambling process. Some sets when compared against themselves were significantly different, and these differences might be attributed to shortness of the data sets or to sampling problems (serial correlation in the "random" samples as a result of the very long time scales present in the data).

Because the original and surrogate sets are significantly different according to the discriminating statistic, the primary null hypothesis that nature of the speed data sets may be explained as linearly correlated noise must be rejected, which implies that a degree of nonlinear determinism exists in the data. It should be emphasized that a rigorous test of the primary null hypothesis would involve evaluation of several discriminating statistics.

## Chapter 5

# Conclusions

Methods of time series analysis have been applied to selected representative types of engine data. These methods use nonlinear measures of causality or predictability, in contrast with more traditional linear measures. Furthermore, the estimates of entropy and dimension are based on multi-dimensional data representations which contain more information than conventional single-dimensional representations such as peak-pressure-, pressure-rise- or IMEP-related parameters.

The pressure data examined in this thesis exhibit weak but discernible levels of deterministic prior-cycle effects. Whereas temporal coupling is barely observable in the autocorrelation and mutual information functions of peak pressures, the estimator of Kolmogorov entropy, which utilizes multi-dimensional information, is able to quantify a good degree of serial correlation. Moreover, the

dimensionality of the signal is sufficiently low to suggest the feasibility of cycle-resolved control.

The speed data show strong temporal correlation over very long time scales incorporating hundreds of engine cycles (e.g., over 60 sec for the Kohler data). As the speed oscillations exist without the presence of a closed-loop idle-speed controller (STID-FCD, FI6V1, FI6V2 and FI6V4), a coupling between short-time-scale combustion variations and long-time-scale mechanical effects, such as postulated by previous studies [8, 9, 10], is suggested as an important and appropriate subject for future study.

The apparent insensitivity of the mutual information function for the peak-pressure sets most likely is a result of the sparseness of the sets preventing a convergence of statistical values. A secondary cause might be the method in which the probability distributions were defined — an optimized algorithm might better be able to handle small data sets.

The methods employed in this thesis have identified and characterized deterministic effects in a variety of engine configurations and measurement variables. This confirmation of pronounced variability and instability in the lean-burn limit is of practical importance for industrial applications, and a suitable topic of future studies would be a more systematic parametric survey using production engines as well as the exploration of predictive control models which might be suitable for reduction of engine combustion and performance irregularities.

## List of References

## List of References

- [1] Heywood J.B. (1988). *Internal combustion engine fundamentals*. McGraw-Hill.
- [2] Young M.B. (1981). "Cyclic dispersion in the homogeneous-charge spark-ignition engine — a literature survey", SAE Paper 810020.
- [3] Ozdor N., Dulger M., Sher E. (1994). "Cyclic variability in spark ignition engines: a literature survey", SAE Paper 940987.
- [4] Barton R.K., Kenemuth D.K., Lestz S.S., Meyer W.E. (1970). "Cycle-by-cycle variations of a spark ignition engine — a statistical analysis", SAE Paper 700488.
- [5] Stevens S.P., Shayler P.J., Ma T. (1992). "A basis for the predictive control of cyclic dispersion in a spark ignition engine", in *Combustion in Engines: Technology, Applications and the Environment* (London), 1992 December 1-3, 175-182.
- [6] Atkinson R.J., Tennant C.J., Traver M.L., Atkinson C.M., Clark N.N. (1994). "A controller for a spark ignition engine with bi-fuel capability", SAE Paper 942004.
- [7] Stone R. (1992). *Introduction to internal combustion engines*, 2nd ed. Society of Automotive Engineers.
- [8] Washino S., Nishiyama R., Ohkubo S. (1986). "A fundamental study for the control of periodic oscillation of SI engine revolutions", SAE Paper 860411.
- [9] Nishimura Y., Ishii K. (1986). "Engine idle stability analysis and control", SAE Paper 860415.

- [10] Ando H., Motomichi M. (1987). "Contribution of fuel transport lag and statistical perturbation in combustion to oscillation of SI engine speed at idle", SAE Paper 870545.
- [11] Clerk D. (1893). *The gas engine*, 4th ed. John Wiley and Sons.
- [12] Jones F.R. (1909). *The gas engine*, 1st ed. John Wiley and Sons.
- [13] Homans J.E. (1911). *Self-propelled vehicles*. Theodore Audel and Company.
- [14] Curry S. (1963). "A three-dimensional study of flame propagation in a spark ignition engine", SAE Transactions **71**, 628-650 (with discussion).
- [15] Patterson D.J. (1966). "Cylinder pressure variations, a fundamental combustion problem", SAE Paper 660129.
- [16] Vichnievsky R. (1955). "Combustion in petrol engines", *Proceedings, Joint Conference on Combustion*, 288.
- [17] Soltau J.P. (1960-1). "Cylinder pressure variations in petrol engines", *Proceedings of the Institution of Mechanical Engineers* **2**, 99.
- [18] Matekunas F.A. (1983). "Modes and measures of cyclic combustion variability", SAE Paper 830337.
- [19] Kantor J.C. (1984). "A dynamical instability of spark-ignited engines", *Science* **224**, 1233-1235.
- [20] Daily J.W. (1987). "Cycle-to-cycle variations: a chaotic process?", SAE Paper 870165.
- [21] Daily J.W. (1988). "Cycle-to-cycle variations: a chaotic process?", *Combustion Science and Technology* **57**, 149-162.
- [22] Foakes A.P., Pollard D.G. (1993). "Investigation of a chaotic mechanism for cycle-to-cycle variations", *Combustion Science and Technology* **90**, 281-287.
- [23] Martin J.K., Plee S.L., Remboski D.J. Jr. (1988). "Burn modes and prior-cycle effects on cyclic variations in lean-burn spark-ignition engine combustion", SAE Paper 880201.
- [24] Wagner R.M., Daw C.S., Thomas J.F. (1993). "Controlling chaos in spark-ignition engines", *Proceedings, Central and Eastern States Joint Technical Meeting of the Combustion Institute* (New Orleans), 1993 March 15-17.

- [25] Ott E., Grebogi C., Yorke J.A. (1990). "Controlling chaos", *Physical Review Letters* **64**:11, 1196-1199.
- [26] Belmont M.R., Hancock M.S., Buckingham D.J. (1986). "Statistical aspects of cyclic variability", SAE Paper 860324.
- [27] Grünefeld G., Beushausen V., Andresen P., Hentschel W. (1994). "A major origin of cyclic energy conversion variations in SI engines: cycle-by-cycle variations of the equivalence ratio and residual gas of the initial charge", SAE Paper 941880.
- [28] Chew L., Hoekstra R., Nayfeh J.F., Navedo J. (1994). "Chaos analysis of in-cylinder pressure measurements", SAE Paper 942486.
- [29] Finney C.E.A., Nguyen K., Daw C.S. (1994). "Entropic characterization of cyclic variability in internal combustion engines using chaotic time series analysis", *Proceedings of the 32nd Japanese Combustion Symposium* (Sendai), 1994 November 21-23, 400-402.
- [30] Anderson R.W., Moore M.J. Personal communication of 1993 July 2.
- [31] Daw C.S. Personal communications from 1993 June to 1995 June.
- [32] Box G.E.P., Jenkins G.M (1970). *Time series analysis: forecasting and control*. Holden-Day.
- [33] Shannon C.E., Weaver W. (1949). *The mathematical theory of communication*. University of Illinois Press.
- [34] Fraser A.M., Swinney H.L. (1986). "Independent coordinates for strange attractors from mutual information", *Physical Review A* **33**:2, 1134-1140.
- [35] Takens F. (1981). "Detecting strange attractors in turbulence", in *Lecture Notes in Mathematics* **898**, 366-381.
- [36] Packard N.H., Crutchfield J.P., Farmer J.D., Shaw R.S. (1979). "Geometry from a time series", *Physical Review Letters* **45**:9, 712-716.
- [37] Poincaré H. (1892-1899). *Les méthodes nouvelles de la mécanique céleste*, 3 vol. Guthier-Villars et fils. (in French)
- [38] Kolmogorov A.N. (1958). "A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces", *Doklady Akademiy Nauk USSR* **119**:5, 861-864. (in Russian)

- [39] Grassberger P., Procaccia I. (1983). "Estimation of the Kolmogorov entropy from a chaotic signal", *Physical Review A* **28**:4, 2591-2593.
- [40] Schouten J.C., Takens F., Bleek C.M. van den (1994). "Maximum-likelihood estimation of the entropy of an attractor", *Physical Review E* **49**:1, 126-129.
- [41] Stappen M.L.M. van der, Schouten J.C., Bleek C.M. van den (1995). "Chaotic hydrodynamics and scale-up of gas-solids fluidized beds: using the Kolmogorov entropy for quantification", in *Fluidization VIII: International Symposium of the Engineering Foundation* (Tours).
- [42] Daw C.S., Finney C.E.A. (1994). To be published.
- [43] Grassberger P., Procaccia I. (1983). "Characterization of strange attractors", *Physical Review Letters* **50**:5, 346-349.
- [44] Schouten J.C., Takens F., Bleek C.M. van den (1994). "Estimation of the dimension of a noisy attractor", *Physical Review E* **50**:3, 1851-1861.
- [45] Theiler J., Galdrikian B., Longtin A., Eubank S., Farmer J.D. (1992). "Using surrogate data to detect nonlinearity in time series", in *Non-linear modeling and forecasting*, Volume XII of the Sante Fe Institute Studies in the Sciences of Complexity, 163-188.
- [46] Theiler J., Eubank S., Longtin A., Galdrikian B., Farmer J.D. (1992). "Testing for nonlinearity in time series: the method of surrogate data", *Physica D* **58**, 77-94.
- [47] Takens F. (1993). "Detecting nonlinearities in stationary time series", *International Journal of Bifurcation and Chaos* **3**:2, 241-256.
- [48] Rapp P.E., Albano A.M., Zimmerman I.D., Jiménez-Montaña M.A. (1994). "Phase-randomized surrogates can produce spurious identifications of non-random structure", *Physics Letters A* **92**, 27-33.
- [49] Press W.H. et al. (1992). *Numerical recipes in FORTRAN: the art of scientific computing*, 2nd ed. Cambridge University Press.

## Appendix

# Appendix

## Example entropy calculation

This appendix illustrates the steps necessary to calculate the windowed entropy as defined in Chapter 4.

The first step is the definition of the relevant data windows. Figure A.1 shows three cycles from a fictitious time series. Data windows begin when the signal crosses the threshold (dashed) line, and the bold curves represent the windowed data segments. Each data window contains the same number of points. Cycle 1 represents a fast cycle, as its data window ends below the threshold line. Conversely, Cycle 3 represents a slow cycle, as its data window ends above the threshold line. The dotted portions of the time series are ignored.

The second step is the definition of relevant length scales. This step will be illustrated below.

The third step is selection of a sample set a test windows. Figure A.2

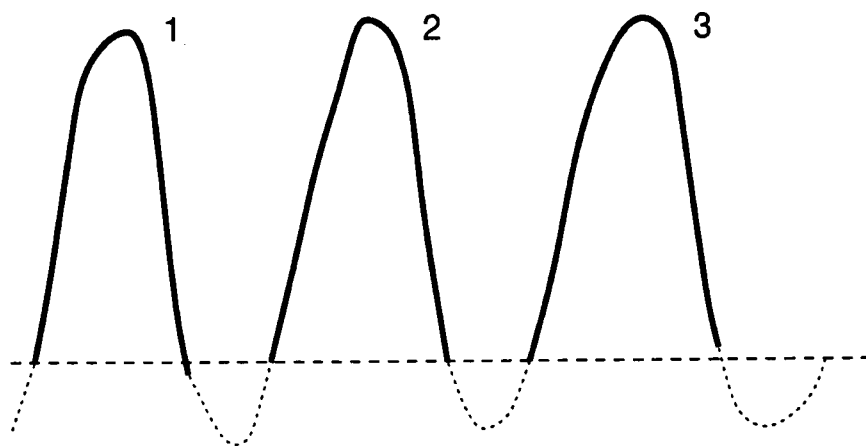


Figure A.1: Example of data windowing on a fictitious time series. The bold curves indicate portions of the time series included in data windows, and the dashed line indicates a threshold for defining the windows.

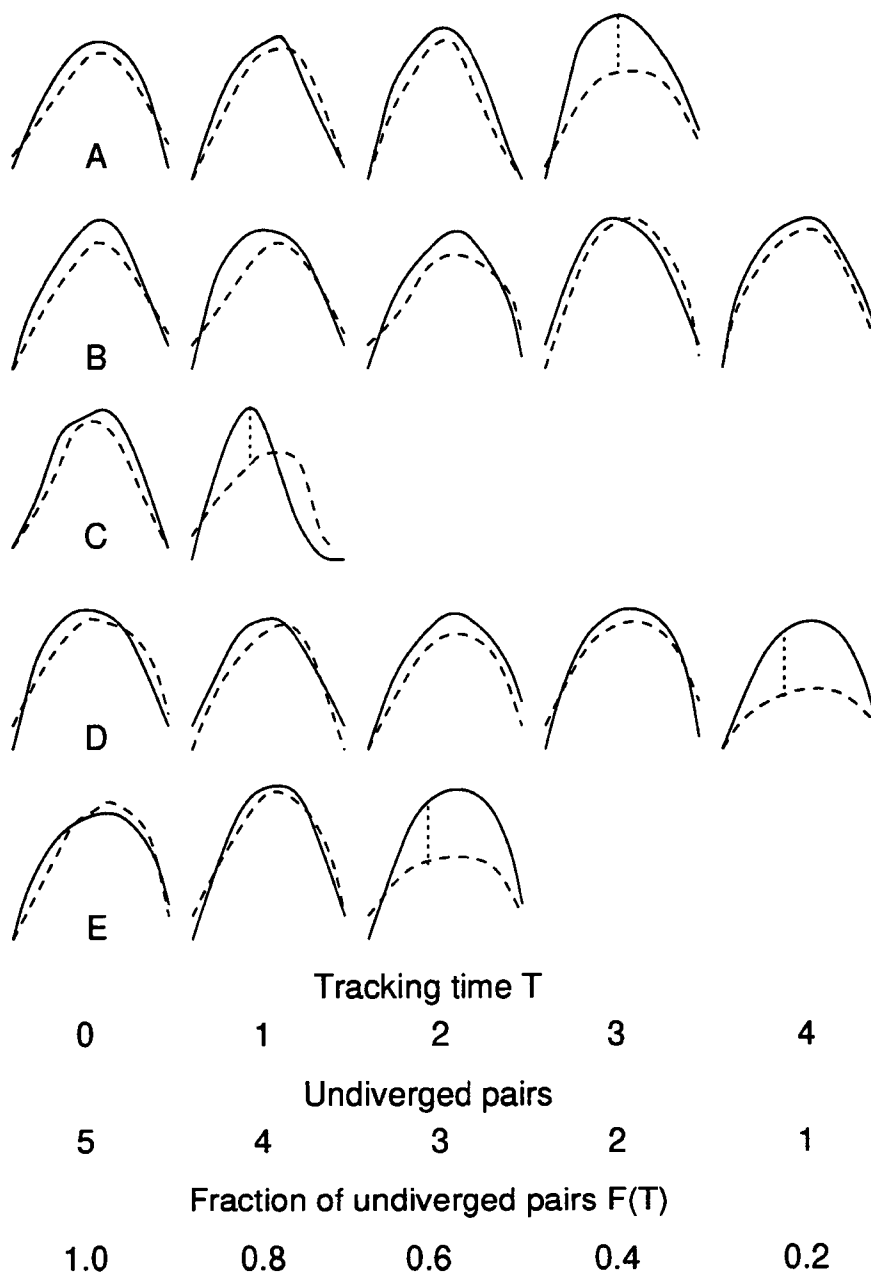


Figure A.2: Example calculation of the decay function. Five sets of undiverged pairs A-E at tracking time  $T=0$  are tested for divergence at window tracking times  $T=1$  to  $T=4$ , and the fraction of pairs still undiverged at each time increment is calculated.

displays five sample test window pairs, designated literally 'A' to 'E'. In the initial test time period, enumerated  $T=0$ , the five pairs, when compared pointwise, are separated by less than a desired tolerance (length scale). When each pair is tracked forward in time (cycles), some pairs diverge beyond the desired length-scale tolerance, as denoted by the dotted line between the windows for pair C in tracking time  $T=1$ . Each window pair in the initial sample population is tracked four time increments (cycles) ahead in time, and the number of undiverged pairs is noted at each time index. Based on a initial undiverged population of five window pairs, the fraction of undiverged pairs referenced with the initial population may be calculated and is designated  $F(T)$  in the figure. The function  $F$  is termed the *decay function*.

The estimate of the entropy is the negative slope of this decay function on a semi-logarithmic scale. In the present example, the gross slope is

$$\frac{\log(0.2) - \log(1.0)}{4 - 0} = -0.58$$

for a base two logarithm. Therefore, the entropy is 0.58 bits per data window. For cases in which the decay function is not as smooth as in the present example, the slope would be determined with a linear least-squares fit.

## Vita

Charles Edward Andrew Finney was born 1969 January 8 in Knoxville, Tennessee. After graduating from Maryville High School in 1987, he attended the University of Virginia and received a B.A. in Astronomy-Physics in 1991. He next studied mechanical engineering at the University of Tennessee, obtaining a B.S.M.E. in 1993 and M.S. in 1995 and qualifying as an Engineering-in-Training in 1992.