



8-2000

Symbolization-based analysis of engineering time series

Charles Edward Andrew Finney

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation

Finney, Charles Edward Andrew, "Symbolization-based analysis of engineering time series." PhD diss., University of Tennessee, 2000.
https://trace.tennessee.edu/utk_graddiss/8276

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Charles Edward Andrew Finney entitled "Symbolization-based analysis of engineering time series." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mechanical Engineering.

Ke Nguyen, Major Professor

We have read this dissertation and recommend its acceptance:

Duane D. Burns, C. Stuart Daw, Jeffery W. Hodgson, J. Rodger Parsons


Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

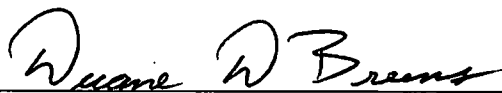
(Original signatures are on file with official student records.)

I am submitting herewith a dissertation by CHARLES EDWARD ANDREW FINNEY entitled *Symbolization-based analysis of engineering time series*. I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mechanical Engineering.

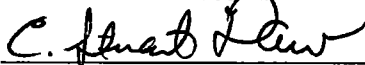


KE NGUYEN, Major Professor

We have read this dissertation and recommend its acceptance.



DUANE D. BRUNS



C. STUART DAW

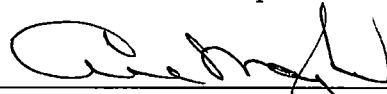


JEFFREY W. HODGSON



J. ROGER PARSONS

Accepted for the Council:



Dean of the Graduate School

Symbolization-based analysis of engineering time series

A Dissertation
Presented for the
Doctor of Philosophy
Degree

The University of Tennessee, Knoxville

CHARLES EDWARD ANDREW FINNEY

August 2000

Acknowledgements

The author thanks Professors DUANE D. BRUNS, JEFFREY W. HODGSON and J. ROGER PARSONS of the University of Tennessee for guidance as members of the doctoral committee. The author expresses appreciation to F.T. CONNOLLY of the Ford Motor Company, J.B. GREEN, JR. of the Oak Ridge National Laboratory, and K.D. EDWARDS and M. VASUDEVAN of the University of Tennessee for supplying experimental data. 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 appreciation to his colleagues K. DEAN EDWARDS, TIANG-YONG TEH, JONATHAN R. VINCENT, SANDEEP RAJPUT, SACHIN SARNOBAT, and JOHNEY B. GREEN, JR., mostly for their patience, as well as to his friends (ARB, DLG, JMH, AAMH, TEK, BWM, WS, RMW to name a few) who expressed interest and concern (and offered occasional goading) during his tenure in the Ph.D. program (likewise to his furry friends, Miss BOOTSIE and Mister BOBBIT, without whose tutelage and divertissement he would not have passed his comprehensive examination). Most importantly, he thanks his parents, RAYMOND and LINDA, and his FAMILY, for their love and support during his protracted graduate studies.

Abstract

Data symbolization, derived from the study of symbolic dynamics, involves discretization of measurement data to aid in observing and characterizing temporal patterns. In this study, symbolization-based methods are developed for analysis of time series from experimental engineering systems to test hypotheses concerning stationarity, temporal reversibility, and synchronization. Stationarity is examined in the context of process control and dynamical state matching; temporal reversibility, in the context of model discrimination and selection of control schemes (linear versus nonlinear); and synchronization, in the context of modes of interactions between system components. Statistical significance is estimated using the method of surrogate data with Monte Carlo probabilities.

Preface

My attitude while writing this dissertation was that the document itself was *pro forma*, and thus when selecting content and scope, I tried to err on the side of succinctness, reasoning that “important” results should be reserved for papers of wider distribution. Here I briefly highlight what is not included but probably should be.

Implicit throughout the work is that the methods described herein rely on the assistance of a digital computer; nowhere are details made explicit. For the most part, the computational coding details are irrelevant. I chose to implement my algorithms in MATLAB, but they could have been implemented in other languages. The thread common to all the techniques is the concept of numerical nonparametric testing, the use of which is enboldened by “inspirational” essays such as one by Efron:¹ “A theory which enables a scientist to understand his data with the help of a high speed computer may now be as useful as a theory which only requires a table of the exponential function, particularly if the latter theory does not exist. Computer assisted theory is no less ‘mathematical’ than the theory of the past, it is just less constrained by the limitations of the human brain.”

Many researchers claim that data symbolization is “robust to noise”; such claims are accepted without rigorous proof, partly because in the tradition of symbolic dynamics and data symbolization, it is common knowledge, if not the very least intuitive, that symbolization allows discrimination of patterns in noisy data. I chose to omit strong claims regarding robustness to noise (both terms of which require careful definition) and resulting trials using my methods because I viewed them as ancillary to the overall thesis. Although I have tested sensitivity to noise, I do not believe the results of those tests necessary for this work; anyone implementing these methods can easily repeat the exercise using different criteria.

With a great many techniques in nonlinear time-series analysis, the nature of the results depends on the selection of parameters associated with the statistical metric and the associated null hypothesis. The question of selecting symbolization

¹Efron B. (1979). Computers and the theory of statistics: thinking the unthinkable, *SIAM Review* 21(4), 460–480.

parameters has been addressed elsewhere, so detailed discussion is not necessary in this study. To some extent, choice of symbolization parameters depends on what is desired to be observed; again, anyone implementing these methods should in the course of learning their utility also explore the effects of symbolization parameters. This dissertation is not comprehensive in that regard.

What is included in this work is a framework for using symbolic data transformations with appropriate metrics to test certain hypotheses. These methods, and resulting derivatives, are expected to find utility in analyzing time series from complex engineering processes and for analysis of time series from measurement systems of severely limited digital precision.

Contents

Acknowledgements	ii
Abstract	iii
Preface	iv
List of tables	viii
List of figures	ix
1 Introduction	1
2 Background and objectives	3
2.1 Précis	3
2.2 Introduction and background	4
2.2.1 Symbolization	4
2.2.2 Surrogates and surrogate generation	8
2.2.3 Dynamical stationarity	10
2.2.4 Temporal reversibility	12
2.2.5 Synchronization	17
2.3 Scope of the present work	19
2.3.1 Methodology	20
2.3.2 Application of methods	21
3 Methodology	25
3.1 Symbolization	25
3.1.1 Basic definitions	25
3.1.2 Data symbolization	28
3.1.3 The Monte Carlo probability	38
3.2 Test for symbolic stationarity	39
3.3 Test for temporal reversibility	41
3.4 Test for symbolic synchronization	45
4 Results and discussion	51
4.1 Stationarity	51
4.1.1 Data	51

4.1.1.1	Experimental apparatus and procedure	51
4.1.1.2	Dynamical behavior of the fluidized-bed data . . .	54
4.1.2	Results of the stationarity test	56
4.2	Temporal irreversibility — T_{irr}	62
4.2.1	Data	62
4.2.1.1	Experimental apparatus and procedure	62
4.2.1.2	Dynamical behavior of the engine data	63
4.2.2	Results of the symbolic temporal-irreversibility test	67
4.3	Temporal irreversibility — S_{Δ}	69
4.3.1	Data	69
4.3.1.1	Experimental apparatus and procedure	69
4.3.1.2	Dynamical behavior of the pulse combustor	72
4.3.2	Results of the symbolic difference statistic test	78
4.4	Synchronization	81
4.4.1	Data	81
4.4.1.1	Experimental apparatus and procedure	81
4.4.1.2	Dynamical behavior of the engine data	82
4.4.2	Two-cylinder synchrograms	87
4.4.3	Three-cylinder synchrograms	101
5	Conclusions and recommendations	107
	References	110
	Vita	117

List of tables

- 1 Stationarity test results for the fluidized-bed data for reference set 4, replicates a and b, and test set C, replicates a and b. 60

List of figures

1	Illustration of temporal reversibility and irreversibility.	13
2	Illustration of the process of symbolization.	30
3	Symbolization and interval keys and resulting symbol sequences. . . .	37
4	Construction of a multi-series code series used in the symbol synchro- gram.	47
5	Schematic of the fluidized-bed system.	52
6	Relationship of the relative conditions for the reference and test sets for the fluidized-bed data.	53
7	Short time-series segments for the reference group from the fluidized bed over a range of fluidization conditions.	55
8	Short time-series segments for the test group from the fluidized bed over a range of fluidization conditions.	57
9	Comparison of the test group against the reference group for the fluidized- bed data using the T_{stat} statistic.	58
10	Short time-series segments from the Ford engine at a range of equiva- lence ratios. Segments are plotted on the same scale.	64
11	Short time-series segments from the Ford engine at a range of equiva- lence ratios. Segments are standardized to aid in visualization of the combustion oscillations.	66
12	Results of the symbolic temporal-reversibility test for the Ford engine data over a range of equivalence ratios for 200 surrogate trials. Sym- bolization parameters were $K_n = \{2\ 2\ 2\ 2\ 2\}$, $K_n = \{4\ 4\ 4\}$, and $K_n = \{8\ 8\}$	68
13	Schematic of the thermal pulse combustor.	70
14	Short pressure time-series segments from the thermal pulse combustor at a residence time of 50 msec. Segments are listed in decreasing equiva- lence ratio, from nearly stoichiometric to very fuel lean, approaching flameout, and are plotted on the same scale.	73
15	Segments of peak-to-trough cycle-magnitude time series from the ther- mal pulse combustor at a residence time of 50 msec. Segments are displayed in decreasing equivalence ratio, from nearly stoichiometric to very fuel-lean, approaching flameout, and are plotted on the same scale.	76

16	Segments of peak-to-trough cycle-magnitude time series from the thermal pulse combustor at a residence time of 50 msec. Segments are displayed in decreasing equivalence ratio, from nearly stoichiometric very fuel-lean, approaching flameout. Segments are standardized to unit variance to highlight cycle-to-cycle rise and fall behavior.	77
17	Behavior of the difference statistic S_{Δ} for peak-to-trough cycle magnitudes for the pulse-combustor data over a range of equivalence ratios.	79
18	Short time-series segments from the Expedition engine at an equivalence ratio of 0.59. Segments are listed in firing order from top to bottom and are plotted on the same scale.	84
19	Short time-series segments from the Quad-4 engine at equivalence ratio of 0.536. Segments are listed in firing order from top to bottom and are plotted on the same scale.	86
20	Symbol synchrograms for the Expedition data at an equivalence ratio of 0.59.	88
21	Randomization of the symbol series for cylinders 0 and 1 destroys patterns of anticorrelation in the synchrogram.	90
22	Modified Shannon entropy of the symbol synchrograms depicted in Figure 20	92
23	Correlation statistic S of the symbol synchrograms depicted in Figure 20 . The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data. Cylinder groupings (labeled by firing order) are: 0 and 3 (a), 0 and 5 (b), 0 and 1 (c).	94
24	Symbol synchrograms for the Quad-4 data at an equivalence ratio of 0.536.	97
25	Modified Shannon entropy of the symbol synchrograms depicted in Figure 24	98
26	Correlation statistic S of the symbol synchrograms depicted in Figure 24	100
27	Symbol synchrogram depicting correlation patterns between cylinders 1, 2 and 3 (firing order) for the Quad-4 data at an equivalence ratio of 0.536.	104
28	Modified Shannon entropy of the symbol synchrograms for cylinders 1, 2, and 3 (firing order) depicted in Figure 27 . The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data.	104
29	Correlation statistic S of the symbol synchrograms for cylinders 1, 2, and 3 (firing order) depicted in Figure 27 . The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data.	105

Chapter 1

Introduction

Ancient humans viewed the heavens with wonderment and reverence, but eventually with pragmatism. The ancient “engineers”, those who developed technology to further civilization, used the regular behavior of terrestrial rotation and planetary orbits to govern the activities of industry and society — they correlated geophysical events such as rainfall and flooding, important for agriculture; they developed a system for time, dates and for navigation. The regularity of nature’s cycles paced the activities of civilization.

During the progress of natural philosophy, the prevailing world view depended on the concept of determinism. From Aristotle’s celestial-sphere model to the work of Newton, the universe was viewed to behave deterministically, like clockwork. This view pervaded the development of science and technology until the nineteenth century, when dynamicists such as Poincaré modified the view of absolute determinism. On the microscopic scale, the quantum-mechanical viewpoint introduced the philosophically unsettling notion of probabilistic behavior. Lorenz and his successors modified the world view yet again with the development of chaos “theory”, the recognition that some behavior in nonlinear systems is neither perfectly regular nor random, not even stochastic.

In the course of engineering study and practice, the prevailing view had been that

any complex, aperiodic behavior must be random in nature, a view which only in the 1990s began to change as chaos and the more applicable methods of nonlinear time-series analysis began to be adopted. Two impediments have retarded the adoption of a chaos framework into analysis of engineering systems — utility and measurability. The utility of certain chaos methods has been demonstrated in engineering systems whose behavior is either enhanced by chaos or whose behavior may be controlled out of chaotic regimes. The measurability problem in some cases is unique to engineering systems, those in the domain of the “real world”. Nonlinear time-series analysis was primarily developed by mathematicians and physicists using mathematical models or simple experimental systems. Their resulting methods did not necessarily handle the effects of noise gracefully or robustly.

Symbolic dynamics uses a data simplification to characterize complex behavior. Symbolization-based data analysis seeks to trade a degradation of measurement precision for a reduction in the effects of noise on statistics and descriptive metrics; alternatively, it seeks to reconstruct dynamical patterns out of intrinsically low-precision measurements. Data symbolization has not seen significant application to measurements from engineering systems — the present study seeks to remedy that deficiency.

This dissertation is organized as follows. Chapter 2 provides background, statements of problems and the motivation behind the present study. Chapter 3 introduces the methodology developed in this work. Chapter 4 presents application of the methods and discusses the relevance of the results to systems of engineering relevance. Chapter 5 offers conclusions from the present study and recommendations for future study.

Chapter 2

Background and objectives

2.1 Précis

Methods of chaotic or nonlinear time-series analysis were primarily developed using noise-free numerical models or on simple laboratory experiments. These methods are powerful for analyzing and predicting dominant systemic dynamics, but for application to a broad spectrum of engineering systems, they must be able to cope with measurement and/or dynamical noise.

The predominant techniques for coping with noise involve either linear filtering (e.g., impulse-response filters, principal-components analysis) or nonlinear noise-reduction filtering based on predictive models. Noise-reduction techniques can be undesirable because they either use models inappropriate for the dynamics of interest (e.g., linear models) or their effects on the dynamics of interest are unknown because they assume a certain model for the noise. Additionally, they can be ill-suited for real-time analysis because they can be computationally resource-intensive.

Symbolization or quantization techniques initially seem counterintuitive because they effectively discard information, using a lower-precision representation of higher-precision measurements. The objective of symbolization is to employ a coarse-grained description of signal behavior without allowing the high-dimensional or noise effects

to destroy statistical discrimination.

In this study, the development of symbolic time-series analysis for noisy engineering measurement time series is presented. Novel techniques for testing for stationarity, temporal reversibility, and multivariate synchronization are introduced. Specific applicability to combustion and multiphase flow systems of engineering importance and relevance is emphasized. More specifically, this work will highlight the relevance of temporal irreversibility and synchronization to selection of models for and control schemes for cyclic variability in internal combustion engines, the nature of temporally irreversible dynamics in pulse combustion, and the relevance of stationarity for dynamical state matching in fluidization dynamics.

2.2 Introduction and background

2.2.1 Symbolization

Data symbolization is derived from the discipline of *symbolic dynamics*. At the end of the nineteenth century, dynamicists were developing methods to describe the dynamical characteristics of complicated systems. Poincaré developed a manner of stroboscopic sampling and projection onto a planar cross-section of the phase-space trajectory, a representation now referred to as a *Poincaré section*.¹ Another approach was to represent sections of the phase space in a coarse-grained manner using a discrete, symbolic data representation; the transitions of the symbols would then describe the dynamics of the system. This latter approach led to the discipline of *symbolic dynamics* (Morse and Hedlund 1938).

¹Fully termed a *surface of section*.

Symbolic dynamics is a well-studied topic among mathematicians and dynamists, and consequently the main bent of research has focused on modeling dynamical systems. Only recently has the use of a symbolic data representation been applied to statistical analysis of measurement time series. Time-series analysts, particularly those from the nonlinear and chaotic time-series communities, have begun to appreciate symbolization as a useful approach to noisy, complex time series. In the chaotic time-series analysis community, there have been two primary approaches to characterizing the strange attractors which serve as the icons of chaos: metric and topological. In the metric approach, attractors are characterized by the estimation of metrics or statistics, e.g., entropy, dimension, and Lyapunov spectrum. In the topological approach, attractors are characterized by the intertwining of (unstable) periodic orbits and how these orbits transition from one family to another. It is in this latter area that symbolic dynamics and more lately a data-symbolization approach have proved most useful. Only in the past decade has there been significant interest in applying symbolization to the metric approach.

Data symbolization refers to the process of converting higher-precision data to lower-precision data (an example might be rounding off times recorded to the nearest minute or second to the nearest hour).² *Symbolic time-series analysis* (also known as *symbol-sequence analysis*) refers to a set of techniques merging the data structures of symbolic dynamics and the generality of information theory. This new approach to time-series analysis seeks to be more flexible with complex, nonlinear time series than linear time-series analysis as well as be less affected by noise than standard methods

²A form of coarse-graining is making histograms of data, in which case the bins are the symbols and the number of items in each bin is the summary statistic. Histograms describe the frequency of the data symbolically, but temporal relationships between individual data records are destroyed.

of nonlinear or chaotic time-series analysis. As such, symbolic time-series analysis holds promise for application to time-series measurements from engineering systems.

Although there has been significant recent interest in symbolization, its ideas are not new. Early in the age of digital computing, analysts had to cope with low-precision digital data. One of the first applications of the digital computer was in computing the autocorrelation functions of time series for the production of graphical displays called *correlograms*. In the 1940s, researchers in Britain concerned with air warfare were among the first to employ digital computers (Cunningham and Hynd 1946):

Being a digital machine, the relay computer is in one sense perfectly accurate. It is, however, necessary to group the data to the nearest integer in the range ± 63 ; in general, this has a negligible effect on the correlogram, but if the grouping be very drastic, it is possible to introduce corrections analogous to Sheppard's corrections, which are valid when certain plausible assumptions are satisfied . . .

From this statement, it is seen that data symbolization was forced by precision limitations of the machine: there are 127 values in the range of integers from -63 to $+63$ (including 0), so the machine computed with 7-bit precision (because $127 = 2^7 - 1$).

As digital computing for data analysis became more sophisticated, researchers became accustomed to expecting or even requiring higher measurement precision, perhaps 10-bit (1024 values) or 12-bit (4096 values) or up to today's 16-bit (65536 values) precision. Even as the trend for higher-precision measurements progressed, analysts found data symbolization convenient, particularly in cases in which the computer hardware made symbolization desirable. A 1968 monograph (Enochson and

Otnes 1968) reported:

Two different types of time series data arise that are to be processed by a digital computer. Some processes are discrete by their very nature, say for example, the daily closing prices of a given stock. On the other hand, some data arise naturally as a continuous record such as the continuously recorded output of an accelerometer on the skin of a missile structure, which is intended to give a measure of the vibration at that point. The continuous record is digitized for digital analysis by an analog-to-digital conversion procedure. In either case, an important observation to make is that only a finite number of bits in a binary digital computer are required to represent any given individual data point. In actual practice, many analog-to-digital converters present their output at a precision of 8 to 10 bits (including the sign bit). For almost all applications, the recording instruments are no more accurate than one part in 256 to 1024 and hence, that quantization is fine enough. . . .

As it turns out, for many time series of interest, much coarser quantization is often acceptable. In the extreme case, one can quantize to a single bit. That is, if the value of a signal is larger than or equal to zero, set the bit equal to zero; if the value of the signal is less than zero, set the bit equal to one. In simpler terms, the sign bit of the data point is the only information retained.

The above-referenced monograph detailed the methods for either 1- or 4-bit quantization for computing autocorrelation functions. The advantage of the symbolization schemes was that for a small number of symbols (different data values), the results from every possible multiplication between two data records could be stored in a look-up table and accessed very efficiently with an intrinsic machine command. As autocorrelation calculations depend on many multiplications, reducing the calculations to a look-up process was more efficient. This is an example of a high-precision machine attenuating data precision to reduce computational time.

Presently, computers are capable of performing many high-precision numerical operations extremely quickly, and memory and storage are becoming plentiful and inexpensive, so why is there interest recently in symbolization as a time-series analysis tool? Symbolic time-series analysis has been shown to be suited to handle data in the presence of dynamical or measurement noise, where having greater numerical precision can be more harmful than helpful. In some applications, being able to manipulate symbolic or low-precision data, particularly noisy data, might be advantageous for memory or processing-speed concerns, so data symbolization is an important and apposite engineering research topic.

Recently, symbolization has found applications over a wide range of disciplines such as astronomy (Schwarz, Benz, Kurths, and Witt 1993), psychology (Kurths, Schwarz, Witt, Krampe, and Abel 1996), and medicine (Kurths, Voss, Saperin, Witt, Kleiner, and Wessel 1995; Saperin, Gowin, Kurths, and Felsenberg 1998). In systems of engineering relevance, symbolization has been applied to internal combustion engines (Finney, Green, and Daw 1998; Daw, Kennel, Finney, and Connolly 1998; Green, Daw, Armfield, Finney, Wagner, Drallmeier, Kennel, and Durbetaki 1999), thermal pulse combustors (Edwards, Finney, Nguyen, and Daw 1998), and fluidization (Halow and Daw 1994; vander Stappen 1996; Finney, Nguyen, Daw, and Halow 1998) and multiphase flow systems (van der Welle 1985; Angeli and Hewitt 1996; Daw, Finney, Nguyen, and Halow 1998).

2.2.2 Surrogates and surrogate generation

Calculation of the symbol statistics from different data-series segments allows comparison of the data series based on difference measures of the symbol statistics. In absence

of *a priori* estimates of significance, statistical significance of difference measures may be accomplished using a method of surrogate data. In this method, surrogate data sets are created from the original data to test a specified null hypothesis. When this process is repeated many times, the mean and variance of the surrogate-data statistical measures may be calculated to estimate significance of the original data measure. For instance, under the null hypothesis of no temporal correlation within a time series, surrogate data sets would be created by shuffling the order of the time series, and then a correlation metric would be calculated for each surrogate set. Given a battery of surrogates, the mean and variance of the surrogate correlation metrics could be used to test whether the correlation metric of the original time series shows correlation with statistical significance.

The method of surrogate data as described above was used by Theiler, Eubank, Longtin, Galdrikian, and Farmer (1992). The process of creating surrogate data and of random data permutations is a combination of bootstrapping and Monte Carlo techniques and is common in nonlinear time-series practice (Rapp, Albano, Zimmerman, and Jiménez-Montaño 1994; Voss and Kurths 1998; Timmer, Schwarz, Voss, Wardinski, Belloni, Hasinger, van der Klis, and Kurths 2000). Although Fisher (1935) advocated using randomization for testing statistical validity, the computer-age methodology of Monte Carlo techniques perhaps owes its origins to Barnard (1963)³. One disadvantage of the method of surrogate data is that it can be computationally intensive, so a test with *a priori* significance, if available, is preferred, provided that the assumptions underlying the statistical model be valid.

³*op. cit.*: "... provided one has access to a reasonable amount of time on a reasonably powerful computer, an exact test of significance is something one never need be without." In his comments, Barnard provided a procedure and test for significance.

Hope (1968) provides a succinct explanation of Monte Carlo significance:

Monte Carlo significance test procedures consist of the comparison of the observed data with random samples generated in accordance with the hypothesis being tested. A test criterion is chosen to facilitate this comparison. The outcome of the test is determined by the rank of the test criterion of the observed data relative to the test criteria of the random samples forming the reference set.

Section 3.1.3 presents mathematical details of testing with surrogate data.

2.2.3 Dynamical stationarity

Dynamical stationarity refers to whether a signal remains statistically consistent over time. As applied to time-series analysis, stationarity may be related to ergodicity. Interesting signals (*i.e.*, with finite entropy) contain many types of oscillations on many different time and length scales. A simple test for stationarity might involve statistical comparison of a short time-series segment with another observed some time afterward (*e.g.*, comparing the data mean for the front and back halves of a time series).

There are many different manifestations of nonstationarity in signals. For instance, there are simple drifts in the local mean (*e.g.*, a steady trend, as in the Dow Jones Industrial Average) or seasonal changes (*e.g.*, the mean outdoor temperature trend over a year). Of interest in nonlinear time-series analysis is whether nonstationarity in the signal fluctuations arises from measurement-system changes (*i.e.*, drifts in signal mean or attenuation of signal amplitude, both of which might occur because of instrument dynamics) or are real changes in the nature of the dynamics. Thus, the

question of dynamical stationarity is one of whether the bounds of the phase-space attractor, as best one can define and measure, shifts over time.

Stationarity is an important assumption in most methods of nonlinear time-series analysis. Until recently, stationarity was more assumed (or wished for) than verified because of the lack of practical tests.

Schreiber (1997) proposed a stationarity test using nonlinear cross predictions, in which prediction errors from temporally separated segments of the time series, one used as the database and the other as the test segment, are evaluated. This approach is rather novel in that it compares portions of the time series directly rather than statistical parameters derived from the time series. Additionally, this method is useful if the nonstationarities arise from trajectory shape shifts in which dynamical invariant measures remain unchanged.

Kennel (1997) presented a test based on the temporal distributions of near neighbors. The premise of this test is that under stationarity, similar trajectory segments (near neighbors) should be equally likely to occur throughout the temporal record of the time series. The Kennel test is much more than an extension of the general class of tests outlined by Schreiber — it tracks temporal rather than spatial relations. This method was extended by Yu, Lu, and Harrison (1999).

Witt, Kurths, and Pikovsky (1998) offered a series of tests based on temporal independence of probability density functions (data histograms) and power spectra. The implementation of this test relies on dividing the temporal record of the time series into segments, a process which is standard practice with nonlinear statistical measures. An undesirable feature of the method of Witt, Kurths, and Pikovsky (1998) is its reliance on a two-step data prefiltering, one a local linear detrender and

the second a Butterworth filter.

Missing in the literature and in practice is a test for dynamical stationarity based on data symbolization. Any such symbolic stationarity test should complement other symbolic tests, such as for temporal irreversibility, using the same symbolization parameters, so that the two tests would have the same basis and not be disjointed in their data representations. This dissertation will present a symbolic approach for evaluating stationarity.

2.2.4 Temporal reversibility

A time series is termed *temporally reversible* if it appears similar whether viewed in forward or reverse time. Tong (1990) gives an accepted formal definition:

Definition 4.6: A stationary time series $\{X_t\}$ is *time reversible* if for every positive integer n , and every $t_1, t_2, \dots, t_n \in \mathbf{Z}$, the vectors $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ and $(X_{-t_1}, X_{-t_2}, \dots, X_{-t_n})$ have the same joint distributions.

A stationary time series which is not time reversible is said to be *time irreversible*.

In other words, as phrased by Diks, van Houwelingen, Takens, and DeGoede (1995), “A time series is said to be reversible if its probabilistic properties are invariant with respect to time reversal. Otherwise it is said to be irreversible.” An important qualifier in this formal definition is the property of dynamical *stationarity*.

An illustration of reversibility and irreversibility in time series is shown in **Figure 1**. The first signal **(a)** is a temporally symmetric sawtooth wave; the trough-to-peak rise time equals that of the peak-to-trough fall time. Thus, the signal looks the same whether viewed in forward or reverse time. The second signal **(b)** is a temporally

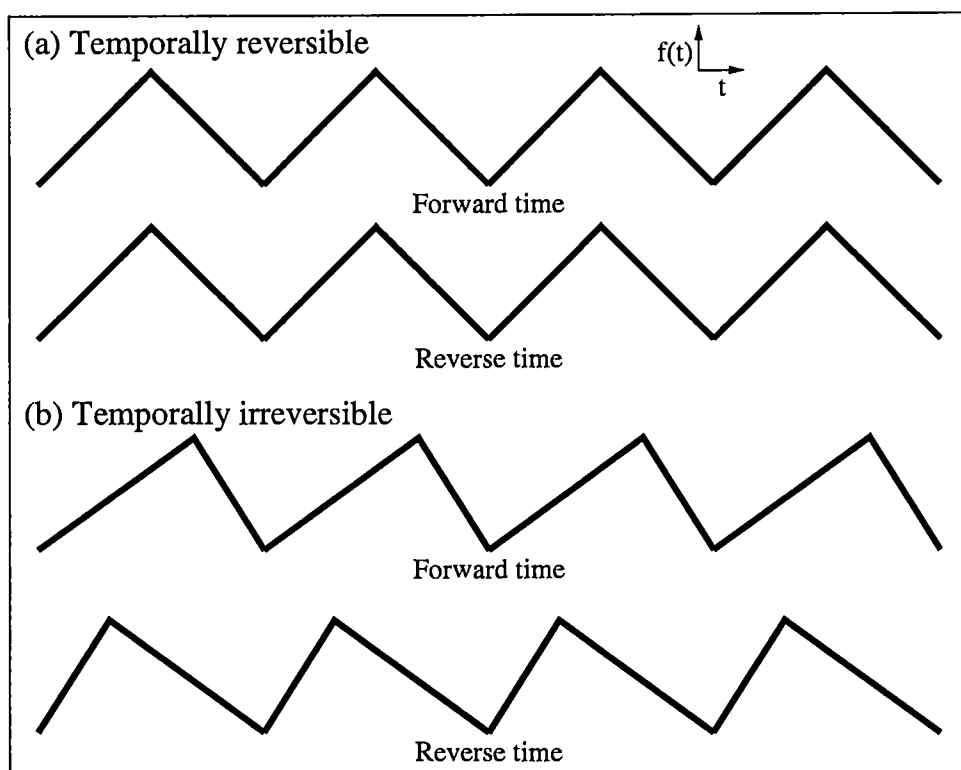


Figure 1: Illustration of temporal reversibility (a) and irreversibility (b).

asymmetric sawtooth wave; the trough-to-peak rise time is twice that of the peak-to-trough fall time. Thus, the signal in forward time looks much different than that in reverse time, where the trough-to-peak rise time is one half that of the peak-to-trough fall time.

Temporal irreversibility is an important characteristic of time series because the presence of irreversibility eliminates linear Gaussian random processes (LGRP), or static nonlinear transformations thereof, as potential models for the generating process; conversely, irreversibility implies nonlinearity or (less likely) non-Gaussian linearity. The implicit temporal reversibility of LGRP was first formalized by Weiss (1975) and has important implications in modeling and analysis. As Lawrance (1991) notes:

The view taken here is that directionality [temporal irreversibility] is an aspect of time series analysis which deserves wider recognition; for instance, it does not make sense to forecast with a time series model which is reversible, when past data are definitely irreversible. In simulating inputs to a system based on directional historical data, directional simulated data should be used. Such obvious requirements are not met by the use of Gaussian ARMA [autoregressive moving average] models.

This applicability to modeling extends to analysis — most linear time-series analysis is inadequate to capture features such as temporal irreversibility which occur in nonlinear time series. Within the past decade, temporal irreversibility has increasingly gained favor in nonlinear time-series applications ranging from economics to medicine to ecology to astrophysics to engineering systems (Diks, van Houwelingen, Takens, and DeGoede 1995; van der Heyden, Diks, Pijn, and Velis 1996; Paluš 1996; Ramsey and Rothman 1996; Stone, Landan, and May 1996; Hoekstra, Diks, Allesie, and DeGoede 1997; Stam, Pijn, and Pritchard 1998; Green, Daw, Armfield, Finney, Wagner,

Drallmeier, Kennel, and Durbetaki 1999; Timmer, Schwarz, Voss, Wardinski, Belloni, Hasinger, van der Klis, and Kurths 2000). However, statisticians working with time series had been aware for quite some time that temporal irreversibility implies non-linearity, as evidenced in the discussion of a meeting of the Royal Statistical Society (Campbell and Walker 1977) :

Mr R. F. Galbraith's point, also made by Professor H. Akaike, about the intervals from a minimum of the series to the following maximum being consistently about 2 years shorter than those from a maximum to the following minimum is most important. We certainly ought not to have overlooked this, and should have noted that our model, at least if Gaussian, could not possibly explain the asymmetry because of its time reversibility. We agree with Professor Cox that the time irreversibility thus indicated is more likely to arise from a non-linear process rather than a linear non-Gaussian process.

On a historical note, the earliest known mention of a "directional effect" appears to have been by Daniels (1946) in the context of analyzing time series from textile processes (Lawrance 1991).⁴

There have been several types of tests for temporal reversibility or irreversibility in time series. Cox (1981) suggested a test based on a lagged third cumulant estimate:

$$\nu_{s,h} = \frac{\langle (Y_s - Y_{s-h})^3 \rangle}{[\frac{1}{2}\langle (Y_s - Y_{s-h})^2 \rangle]^{\frac{3}{2}}}, \quad (1)$$

⁴*op. cit.*, p.87: "I should like to ask Mr. Foster whether he has found any evidence in cotton series of what might be called a directional effect. From the theory of drafting outlined in his paper, one might expect the thickness to increase relatively slowly up to a point, and then to diminish rapidly as the tuft is pulled through the front rollers, producing a kind of saw-toothed appearance in the thickness curve. That such an effect may exist in wool is suggested by a tendency to skewness in the form of the frequency distribution of first differences, though the evidence is admittedly inconclusive in the absence of a suitable test of significance. It is perhaps worth observing that a directional effect of this nature is produced in an autoregressive series when the distribution of the residuals is skew."

where $\nu_{s,h}$ is the cumulant based on temporal index s and delay h , $\langle \cdot \rangle$ denotes expected value, Y_s is a measurement at temporal index s , and Y_{s-h} is a measurement h in time preceding Y_s .

Similarly, Pomeau (1982) suggested:

$$\psi''(\tau) = \langle x^3(t)x(t+\tau) - x(t)x^3(t+\tau) \rangle, \quad (2)$$

where ψ is the temporal-symmetry index, τ is the delay, $x(t)$ is a measurement at temporal index t , and $x(t+\tau)$ is a measurement τ in time succeeding $x(t)$.

Following the terminology of Lawrance (1991), it should be noted that both the Cox and Pomeau statistics, as well as an alternate by Timmer et al. (1993), quantify *lagged reversibility* and rely on simple relationships of pairs of measurements. An advantage of this type of statistic is its simplicity, but a disadvantage is the low dimensionality of the input data (groups of two time-ordered data).

Stone, Landan, and May (1996) offered a test based on a nonlinear predictive model using nearest-neighbor prediction. By comparing the effectiveness of the predictive regression for forward- and reverse-time versions of the time series, a measure of irreversibility is obtained. This test would be expected to be more robust than the tests mentioned above because of the higher dimensionality of the input data (the embedding process typically groups 2–10 time-ordered data). The scheme presented by Stone et al. does not employ any prefiltering before the phase-space embedding, so this test might be expected to be hampered by the effects of noise.

An alternative phase-space test for irreversibility was proposed by Diks, van

Houwelingen, Takens, and DeGoede (1995) based on an attractor shape metric presented in Diks, van Zwet, Takens, and DeGoede (1996). The test compares the metrics between phase-space trajectories embedded from forward- and reverse-time versions of the time series. This test has been used in a variety of applications (van der Heyden, Diks, Pijn, and Velis 1996; Hoekstra, Diks, Alessie, and DeGoede 1997). For noisy time series, a filtering noise-reduction procedure would have to be employed at higher noise levels.

Missing in the literature and in practice is a test for temporal reversibility which utilizes symbolization and which does not rely on Fourier-transform surrogates. This dissertation will provide such a symbolization-based test.

2.2.5 Synchronization

Synchronization in dynamical systems refers to the phenomenon of multiple systems exhibiting similar behavior simultaneously, to within an appropriately defined level of discrimination of “similarity” and “simultaneity”. For physical relevance, the distinct systems should be coupled in an identifiable manner (e.g., gravity, mechanical linkages, mass exchange) so that the exchange of information between the systems, and not chance, may be identified as the source or driver of synchronization.

The classical, canonical example of synchronization was described by Huygens (1673) in his treatise on the pendulum clock. Huygens described how the pendula of two pendulum clocks, mounted on a common beam, tended to synchronize in anticorrelated motion due to transmission of vibrations along the beam⁵; without the means

⁵*op. cit.*, p. 30: “It is quite worth noting that when we suspended two clocks so constructed from two hooks imbedded in the same wooden beam, the motions of each pendulum in opposite

of conveyance of information, the two clocks would not be expected to synchronize by chance, especially after application of disruptions intended to desynchronize.

Rayleigh (1896) described synchronization of two organ pipes placed in sufficient proximity. When pipes of similar but distinct frequencies are sounded separately, frequency beating occurs when the pipe openings are distant, but when the openings are near, a single, mutual frequency sounds. With acoustical and pressure-based phenomena, this synchronization may be seen in other systems. For instance, two candles will flicker independently when placed apart, but when placed together, their flickering patterns are nearly identical and synchronous, because each flame's combustion, with surrounding convective currents, affects the other's.

Synchronization is understood conceptually in many scientific disciplines, but there is no established, universal definition. Recognizing this deficiency, Schäfer, Rosenblum, Abel, and Kurths (1999) noted:

Synchronization is a universal phenomenon that occurs due to the coupling of two or more nonlinear oscillators. A number of quite different effects are referred to as synchronization. Understood in a wide sense as the mutual time conformity of two or more processes, this phenomenon lacks a unique definition and requires a more precise description in particular cases. For example, in the context of the interaction of chaotic oscillators one distinguishes between complete, generalized, phase and lag

swings were so much in agreement that they never receded the least bit from each other and the sound of each was always heard simultaneously. Further, if this agreement was disturbed by some interference, it reestablished itself in a short time. For a long time I was amazed at this unexpected result, but after a careful examination finally found that the cause of this is due to the motion of the beam, even though this is hardly perceptible. The cause is that the oscillations of the pendula, in proportion to their weight, communicate some motion to the clocks. This motion, impressed onto the beam, necessarily has the effect of making the pendula come to a state of exactly contrary swings if it happened that they moved otherwise at first, and from this finally the motion of the beam completely ceases. But this cause is not sufficiently powerful unless the opposite motions of the clocks are exactly equal and uniform."

synchronization, with all these states being defined in different ways.

Schäfer et al. explained that because of the effects of noise, phase- and frequency-locking synchronization can experience *phase slips*, interruptions in the synchronicity of the oscillators, so that synchronization is best measured statistically.⁶ They offered a graphical representation, called a *synchrogram*, of bivariate, multimodal synchronization based on stroboscopically measured phase differences. The synchrogram, so defined, has the advantage over traditional phase-difference metrics in that it can represent multimode (i.e., $n : m$ rather than $n : 1$) synchronous epochs, but it has the disadvantage of being limited to a bivariate relationship.

Missing in the literature and in practice is a representation of multivariate synchronization. Of particular engineering relevance is examination of the degrees of and conditions leading to synchronicity in multicylinder internal combustion engines in fuel-lean operation. Symbolization has successfully been applied to engine data to provide fundamental insight into the nature of unstable combustion variations, so the use of a multivariate symbolic representation of synchronization seems an appropriate area for research. This dissertation will present such a representation.

2.3 Scope of the present work

This section introduces the development of the methodology for the proposed study and the application of these methods to observations from physical systems and models of engineering relevance.

⁶As noted by Tass et al. (1998), "Strong noise can cause *phase slips*, i.e., rapid unit jumps of the relative phase. In this case the question 'synchronous or not synchronous' cannot be answered unambiguously, but can be treated only in a statistical sense."

2.3.1 Methodology

In the course of the present research, three techniques will be developed to analyze noisy signals from engineering measurements using symbolic time-series analysis. The first is a test for dynamical stationarity. The second is a test for temporal reversibility. The third is a representation of and test for synchronization of multiple measurement signals.

As noted in **Section 2.2.3**, dynamical stationarity is a requirement of most tests developed in nonlinear time-series analysis. Until recently, stationarity has more often been assumed than verified because of the paucity of research into defining stationarity more appropriately in the context of noisy, nonlinear time series. Of the published tests for stationarity developed by nonlinear time-series analysts (Schreiber 1997; Kennel 1997; Witt, Kurths, and Pikovsky 1998; Yu, Lu, and Harrison 1999), none employ symbol-sequence analysis.⁷ A symbolic test is therefore warranted to define stationarity in a context more meaningful for other methods of symbolic time-series analysis. For example, confirmation of stationarity is a requirement before a signal may be tested for temporal reversibility, as nonstationary signals are trivially irreversible.

As discussed in **Section 2.2.4**, temporal irreversibility is a symptom of nonlinearity, and recent appreciation of this dynamical characteristic has led to increased research into irreversibility. Most recent tests for irreversibility rely on regression models or raw state-space methods, and to handle the effects of noise, they rely on

⁷At the time of writing, Kennel and Mees (2000) introduced a symbolic stationarity test. This test evaluates statistical significance differently than in this dissertation but is algorithmically more complicated.

some sort of filtering noise reduction. Most standard tests for nonlinearity rely on Fourier-transform surrogates and thus inherit the limitations and disadvantages associated with this method (Rapp, Albano, Zimmerman, and Jiménez-Montaña 1994; Stam, Pijn, and Pritchard 1998; Kugiumtzis 1999). A test for temporal reversibility based on symbol-sequence analysis will be presented, and this test will provide statistical significance.

Synchronization of signals from coupled dynamical systems is a topic of increased interest in the nonlinear-dynamics research community. A novel symbolic method for representing synchronization from multiple simultaneously measured signals will be presented. This algorithm will allow complete characterization of the system state using more than two signals, a distinct advantage over current methodologies which demonstrate synchronization in pairs of signals only.

2.3.2 Application of methods

The utility of the proposed methods will be demonstrated on measurement signals from several systems of engineering relevance. The first type of signal will be heat-release and work estimates from several different spark-ignited internal combustion engines. The second type of signal will be pressure measurements from a thermal pulse combustor. The third type of signal will be differential pressure measurements from a fluidized bed.

The dynamical behavior of internal combustion engines is very interesting from a nonlinear-dynamics perspective and extremely relevant from an engineering perspective. The cycle-to-cycle patterns of heat-release, work and pressure-derived quantities

have been demonstrated to be nonlinear in nature, particularly at lean fueling, and exhibit characteristics ranging from stochastic to deterministically chaotic. This latter characteristic has important implications for control. The physics of the combustion process dictates that the measurement signals be inherently noisy, because of fluid-dynamical and operational effects, and coupled with additive-noise effects, analysis of the high-dimensional signals can be problematic. Here, symbolization is a useful and appealing basis for signal analysis, and it fits in naturally with limited computational capabilities of existing hardware in production engine systems. In this study, data from several types of internal combustion engines will be examined temporal reversibility and synchronization. Temporal irreversibility places limits on the types of models of cyclic variability which one may employ and, as a measure of nonlinearity, anticipated control strategies. Synchronization among various engine cylinders is very interesting in that it has not received much attention in the literature and in that exploration of the nature of synchronization can contribute greatly to the understanding of commonly known performance limitations such as the acoustic modes of manifold dynamics (e.g., intake-manifold tuning, a performance-enhancing tactic in automobile racing).

Acoustical interactions are important for combustion systems, in the manifold dynamics of internal combustion engines and in the interactions of multiple flames in boiler systems. In the limiting case is the pulse combustor, a continuous-flow system in which the combustion occurs in phase with acoustic oscillations of the exhaust tailpipe. The dynamics of combustion-chamber pressure measurement time series has been shown to range from a stable, approximate periodicity near stoichiometric fueling to highly unstable oscillations near the fuel-lean flammability limit. Because

the pressure variations primarily occur in phase with the tailpipe acoustical standing wave, the pressure measurement time series are dominated by strong periodicities, and instabilities associated with irregular combustion appear as modulation to the amplitude of the periodic oscillations. Recent methods for testing for nonlinearity based on the method of surrogate data using Fourier-transform surrogates do not function properly in cases of strong predominant periodicities (Stam, Pijn, and Pritchard 1998), so a test for temporal irreversibility should prove more definitive for pulse-combustor data. Knowledge of the behavioral nature of pulsed combustion near the fuel-lean limit is important, as nonlinear and linear control responses differ. Additionally, the strong periodicities in the continuous measurement signals present unique challenges heretofore unmet by established symbolization analysis methods. The proposed study will demonstrate novel symbolization schemes well-suited for pulse-combustion time series and will test combustor behavior near the fuel-lean flammability limit for significant temporal irreversibility. The data have been obtained from a laboratory-scale thermal pulse combustor.

A very relevant question in the modeling of fluidization dynamics is how long the simulation should be run to capture converged dynamics. For instance, analysis of start-up transients is not necessarily indicative of the converged, post-transient dynamical state of the fluidization simulation (*i.e.*, the actual dynamics of interest), so identification of the convergence time is quite important. Equally relevant is whether experimental fluidization systems exhibit long-term dynamical drifts even after startup transients have passed. This question of ergodicity is important for dynamical state matching for process control (*e.g.*, how similar a test time series is to previously observed reference time series) and for determining the minimal acceptable

observational time span, as well as numerical simulation duration, at a given operating condition. In this study, observations from an experimental fluidization system will be examined to demonstrate methods of stationarity testing for autostationarity and for dynamical state matching.

Chapter 3

Methodology

3.1 Symbolization

Data symbolization is the first step in symbolic time-series analysis. This section presents terminology and basic algorithms for symbolization.

3.1.1 Basic definitions

A *datum* is an item of information, usually numeric; its plural is *data*. A datum might be the outdoor ambient temperature at a given instant or the high temperature the previous day — such information might be quantitative (*e.g.*, 291.32 K or <300 K) or qualitative (*e.g.* “warm” or “warmer than yesterday”). Although often used interchangeably, the terms *signal* and *time series* are differentiated. A *signal* is a continuously variable analog information source (although in electronics, digital information sources are also signals). A *time series* is a time-ordered sequence or collection of data, usually the result of digitization of a signal, in which case it is a discrete, finite-numerical-precision representation of the signal plus measurement and conversion uncertainty and channel noise.

Time series are of two types, discrete and continuous, and the differences subjectively depend on the timescales of measurement. A *discrete time series* is recorded

on an event basis to produce a datum representative of a natural period of oscillation in a signal. A *continuous time series* is recorded at equally spaced intervals or on fast timescales relative to the natural period of oscillation in a signal. From a modeling standpoint, discrete time series are best represented by maps, continuous time series by differential equations. Examples of discrete time series might be the maximal daily outdoor temperature recorded over many days or the duration of a geyser spray each eruption. In the former case, days are cycles of equal duration. In the latter case, the geyser erupts erratically at unequal intervals between bursts. Examples of continuous time series might be the instantaneous outdoor temperature measured each hour over many days or the pressure in an internal combustion engine cylinder measured once per crank-angle degree. In the former case, hours are equal intervals, and assuming that on average temperature oscillates between a minimum and maximum within a day, there would be 24 measurements per natural period of oscillation of the temperature. In the latter case, a typical four-stroke engine has 720 crank-angle degrees per engine cycle, so there would be 720 measurements per natural period of oscillation; because the piston accelerates and decelerates throughout the engine cycle, the time intervals between crank-angle degrees are unequal, but the basis for measurement is equal.

The terms *stream*, *series* and *sequence* are also differentiated. A *stream* is a continuous data flow. A stream usually refers to an abstract source of available information (for instance, the output of a measurement transducer which is always working while the system is running regardless of whether one observes its output). A *series* is a finite-length, sampled subset of a stream. Usually, the beginning and ending instants of a series are chosen. A *sequence* is a short, ordered subset of a

series, usually grouped together for numerical analysis.

Individual items in a series are *records*. The location of a record within a series is the *record index*, usually counted from 1 to the number of records in the series. Individual items in a sequence are *elements*. The location of an element within a sequence is the *step*. (Individual items in a stream are simply *data*, and their location is measured not by an index but by an external, usually absolute, time source. Streams here are defined as abstract entities, so their internal structure and terminology are not of much interest here.)

A time series and a data series are usually the same. In series, most data occur in temporal order, but this is not always the case. The data might have been sampled from randomly selected locations in a time series to create a data series — the temporal order would be destroyed. The order of a time series might be shuffled for different types of numerical analysis — this would create a data series, even though it might be referred to as a “shuffled time series”, so that the origin of the data would be known.

As implied above, most real time series involve *uncertainty* and *noise*. *Uncertainty* describes the quality of a measurement in terms of *accuracy* and *precision*. For instance, the ambient outdoor temperature might actually be 291.32 K, but the transducer might indicate a temperature of 290.6 K. The transducer’s reading is slightly inaccurate because of its design. Furthermore, the precision (least count) of its display or digital output might not be sufficient to distinguish between 290.58 and 290.62 K — both would be reported as 290.6 K. Both inaccuracy and imprecision are factors which compromise the quality of physical measurements.

Noise generally describes dynamics other than those of interest, for lack of a vaguer definition. In the thermocouple example, on the smallest length scales, there is thermal noise, the stochastic oscillations and vibrations of atoms and molecules which vary with the temperature and energy of the system. This thermal noise can be significant enough to cause the actual analog output (voltage) of the thermocouple to vary erratically over short time scales, even when supposedly maintained at constant temperature. Additionally, in the connectors between the transducer and the recording device, there is thermal noise and electromagnetic interference (*e.g.*, from nearby alternating currents in power lines). These factors all add fluctuations to the signal totally unrelated to the signal of interest (the actual temperature), even *before* the measurement is recorded. Additionally, all of the recording equipment suffer from the effects of environmental noise contamination, bit noise and/or digital imprecision.

All real measurement series should be affected by uncertainty and noise — it is up to the researcher to understand how significant both factors are. In addition to the measurement noise described above, most physical systems contain *dynamical noise* which sometimes affects their behavior and/or shows up in measurements. Dynamical noise generally arises from high-dimensional or complex physical processes in the system, and these processes can interact with the dynamics of interest, usually resulting in more complex behavior.

3.1.2 Data symbolization

Data *symbolization* (*quantization*) involves the conversion of data of many possible values into symbols of only a few distinct values. The actual quantization is applied to individual data records, but symbolization may be represented as a procedure

producing a symbol stream from a signal (if symbolization occurs at the time of recording) or a symbol series from a data series. This coarse-graining has the practical effect of producing low-resolution data from high-resolution data; only large-scale details are preserved, and in some cases the effects of dynamic and measurement noise on statistical algorithms are reduced.

The basic idea in data symbolization is to quantize data into a few possible values. Depending on the value of a datum, it is assigned one of n symbolic values (e.g., 0 or 1 for $n = 2$; 0, 1 or 2 for $n = 3$). The value n is termed the *symbol-set size*, which denotes the number of symbols available to symbolize the data. As n becomes increasingly large, more detail of the original data is included, with the tradeoff that more noise is also included. In the limit, when n equals the number of distinct values in the time series, the symbol series and the data series contain the same degree of precision, with the data representation (using symbols instead of numbers) being the only difference.

Figure 2 illustrates the process of data symbolization (a) and a representation of the symbolized dynamics called a *symbol-sequence histogram* (b). Data are symbolized based on their value relative to boundaries that partition the data range to produce a symbol series. Sequences are formed by defining a finite-length template to group consecutive symbols, and this template is shifted along the symbol series. Each sequence is decimally encoded to form a code series. A symbol-sequence histogram is a tally of the absolute counts or relative frequencies of each sequence code. The symbol-sequence histogram is a tabulation of all the dynamical patterns in a time series. This tabulation may be expected to change based on the symbolization parameters, viz., the symbol-set size, the sequence length, and the intersymbol

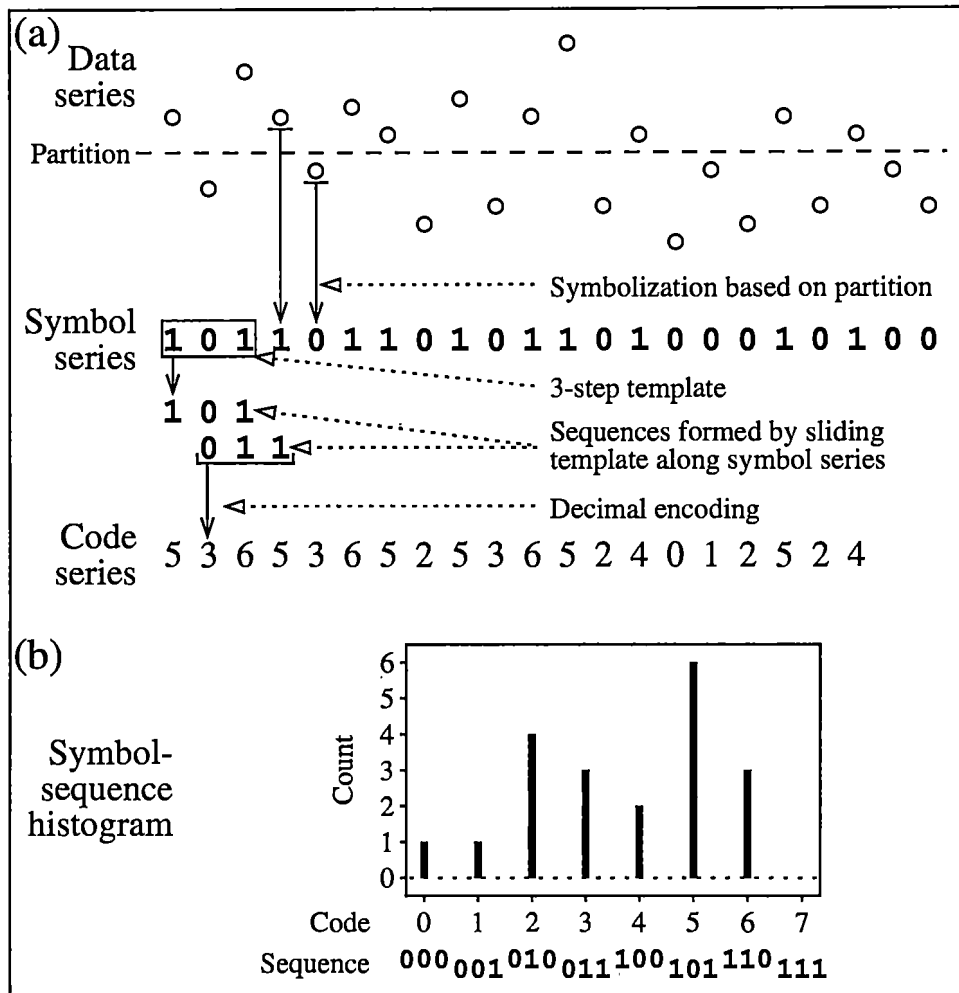


Figure 2: Illustration of the process of symbolization.

interval.

The symbol-set size refers to the number of symbols used to partition the data. In the schematic (**Figure 2**), there is a single partition, so only two symbols (0 and 1) form the symbol set. As this symbol-set size increases, more details about the variations in the data are included, with the tradeoff that the effects of noise become more pronounced. Practical symbol-set sizes range from 2–10 (Daw, Kennel, Finney, and Connolly 1998). The choice of how to partition the data space to capture the temporal patterns faithfully, with some degree of sensibility, may be arbitrary. In formal symbol dynamics, there is a theoretical selection called a *generating partition*, in which a unique, perfect representation of the dynamics is obtained, for a noise-free, infinite sequence of data. In symbol-sequence analysis of real (noisy) data, this generating partition does not exist, even in theory, because of the effects of noise (Crutchfield and Packard 1983). Although some authors attempt to optimize partition locations (Lehrman, Rechester, and White 1997), others note that reasonable partition choices yield satisfactory results (Kurths, Schwarz, Witt, Krampe, and Abel 1996; Tang and Tracy 1998). A convenient and intuitive choice is to select partitions which make each symbol in the symbol series equally probable; this is called *equiquantal* (Paluš 1996) or *equiprobable* partitioning (Finney, Green, and Daw 1998).

Sequences are short, sequential groupings of symbols and depend on two parameters: sequence length and symbolization interval. The *sequence length* is the number of symbols in the sequence, and the *symbolization interval* is the number of symbols in the original symbol series between each element in the sequence. For discrete time series, a natural choice for symbolization interval is one iterate, so that consecutive symbols are groups in the sequence. For continuous time series, the effects

of oversampling result in many consecutive instances of the same interval, so that a decorrelation interval is necessary to capture “interesting” dynamical features of the signal. There is no immediately obvious choice of sequence length. The process of selecting symbolization parameters can be complicated and depends on the nature of the data and on which dynamical features are to be highlighted. Finney, Green, and Daw (1998) and Daw, Finney, Nguyen, and Halow (1998) review criteria for selecting symbolization parameters.

The sequence length dictates how much temporal information is included in the sequence; in this respect, it is analogous to the embedding dimension in phase-space reconstruction. Practically, sequence length may be selected based on the autocorrelation of the data or by convenience given a chosen symbol-set size. The number of possible sequences (bins in the symbol-sequence histogram) follows the relation

$$N_{\text{seq}} = n^L, \quad (3)$$

where n is the symbol-set size and L is the sequence length.

Symbol-sequence histograms summarize the dynamic information of present, so metrics of the histograms can be used to characterize the generating time series. One metric used to describe the degree of organization is a modified form of the Shannon entropy:

$$H_{SM} = -\frac{1}{\log N_{\text{obs}}} \sum_i p_i \log p_i, \quad (4)$$

where p is the symbol-sequence histogram (with relative frequencies) and i is indexed

over all N_{obs} bins with non-zero frequency of occurrence. Using an equiprobable partitioning scheme and normalizing the entropy by N_{obs} results in values of H_{SM} ranging from 0 to 1. For nearly “random” data, the modified Shannon entropy tends to unity, and lower values of entropy imply more organization in the data, to within finite-sample fluctuations.

Algorithmically, the process of creating symbol sequences from a symbol series is the same as embedding vectors from a time series using time-delay embedding. However, there does not appear to be any obvious analogue of the geometric construction underlying time-delay embedding. The time-delay reconstruction theorems (Takens 1980; Sauer, Yorke, and Casdagli 1991) are valid only with smooth transformations of the original state space, which do not exist with symbolized time series, which are discrete and not smooth.

The *symbol tree* is a graphical representation of the symbol statistics as a function of the symbol-sequence length; this representation is derived from the study of symbolic dynamics. Here is an example symbol tree:

$$\begin{array}{cccc}
 & p_0 & & p_1 \\
 p_{00} & & p_{01} & \\
 & p_{000} & p_{001} & p_{010} & p_{011} & p_{100} & p_{101} & p_{110} & p_{111}
 \end{array}$$

In the bottom level of the tree, p_{xyz} is the probability that the symbol sequence $x y z$ occurs. In this example, the symbol-sequence length, denoted L , is 3, representing the probability of three successive symbols to occur. Some researchers refer to L as the tree level because it represents the depth in the symbol tree at which the symbol statistics are evaluated (Tang, Tracy, Boozer, deBrauw, and Brown 1995; Tang and Tracy 1998). The symbol statistics vary as a function of the symbol-sequence length.

A convenient numerical representation of each symbol sequence is achieved by converting the base- n sequence into a decimal (base-10) equivalent number, termed the *sequence code*. For instance, for the binary sequence 0 0 0 the sequence code is 0, for 0 0 1 is 1, for 0 1 0 is 2, for 0 1 1 is 3 and so on. This encoding process is very similar to the approach used by Lehrman, Rechester, and White (1997). Encoding each symbol sequence also allows for the symbol statistics to be displayed as a histogram, with the sequence code serving as the bin number; this representation of the symbol statistics is termed a *symbol-sequence histogram*. With equiprobable partitioning, the relative frequencies for truly random data sequences are equal, and all histogram bins are equally probable, within statistical fluctuations of finite-size sampling populations. Thus, any significant deviation from equiprobability is indicative of temporal correlation and deterministic structure.

All discussion hitherto has focused on the choice of a fixed symbol-set size to calculate symbol statistics for a given symbol-sequence length and symbolization interval. In some cases, it is not desirable to have the same n at each step in the sequence. For instance, in the cases of fitting models to experimental data, using a different number of symbols in each step in the symbol sequence can provide better discrimination (Daw, Kennel, Finney, and Connolly 1998). Differing symbol-set sizes can provide better fits because the one-dimensional marginal distributions of trial simulations are constrained to match the data more closely when symbol partitions are not commensurate. Instead of using the parameters n and L to describe symbol-tree geometry, a more generic *symbolization key*, denoted K_n , may be used. This key is a list of symbol-set sizes at each step in the symbol sequence. For example, a symbolization key of {3 2} denotes a symbol-set size of 3 for the first step of the

symbol sequence and a symbol-set size of 2 for the second step. Graphically, the symbol tree would be:

$$\begin{array}{cccccc}
 & p_0 & & p_1 & & p_2 \\
 p_{00} & p_{01} & p_{10} & p_{11} & p_{20} & p_{21}
 \end{array}$$

Mathematically, the symbolization transformation $Q(n; \cdot) : \mathbf{x} \mapsto \mathbf{s}$ is defined such that the data series \mathbf{x} is discretized using n symbols based on a partitioning function P which discretizes the range of \mathbf{x} . Here, the partitioning function is defined *a posteriori* such that the observed probabilities of each symbol in \mathbf{s} are the same (i.e., using equiprobable partitioning).

The encoding transformation $E(L, I; \cdot) : \mathbf{s} \mapsto \mathbf{c}$ is defined such that the symbol series \mathbf{s} is “embedded” into sequences of length L with intersymbol interval I and then decimally encoded to a scalar code series \mathbf{c} . The code series \mathbf{c} , then, contains the symbol statistics of time series \mathbf{x} given symbolization parameters n , L , and I . The symbolization described here is homogeneous; i.e., fixed symbol-set sizes n and sequence lengths L are employed.

The above steps are combined into a single transformation

$$S(n, L, I; \cdot) : \mathbf{x} \mapsto \mathbf{c}. \quad (5)$$

More generally, to account for nonhomogeneous symbolization, S may be defined as

$$S(K_n, K_i; \cdot) : \mathbf{x} \mapsto \mathbf{c}, \quad (6)$$

where K_n is the symbolization key (the number of symbols at each step in the sequence) and K_i is the interval key (the number of timesteps or symbols between each

step in the sequence). Note that $\dim K_i = \dim K_n - 1$.

The relationship between the symbolization and interval keys may be seen in **Figure 3**. In the figure, the smooth curve is the analog signal, and the open symbols represent measurement values. To the right of the signal are the partitions used to quantize the data range with symbol-set sizes of 2,3 and 4. Below the signal are three sequences. Sequence 0 1 1 1 is formed by homogeneously symbolizing the data series with a binary symbol set at each of the four steps in the sequence, so the symbolization key is {2 2 2 2}; consecutive measurements are used at each step in the sequence, so the interval key is {1 1 1}. Sequence 0 2 3 is formed by homogeneously symbolizing the data series with a quaternary symbol set at each of the three steps in the sequence, so the symbolization key is {4 4 4}; every other measurement is used at each step in the sequence, so the interval key is {2 2}. Sequence 0 1 1 is an example of nonhomogeneous symbolization. The symbolization key is {2 4 3}, meaning that a symbol-set size of 2 is used for the first step, of 4 for the second step, and of 3 for the third step in the sequence. The interval key is {1 3}, meaning that consecutive measurements are used between the first and second steps in the sequence, and three measurements separate the second and third steps in the sequence. Denoting this nonhomogeneous symbolization is awkward, but the nomenclature is useful for describing multichannel synchronization.

A *symbol-sequence histogram* is the frequency histogram of \mathbf{c} , denoted

$$\mathbf{p} = (p_1, \dots, p_{N_{\text{seq}}}), \quad (7)$$

where p_i is the relative frequency of sequence i and N_{seq} is the number of possible sequences

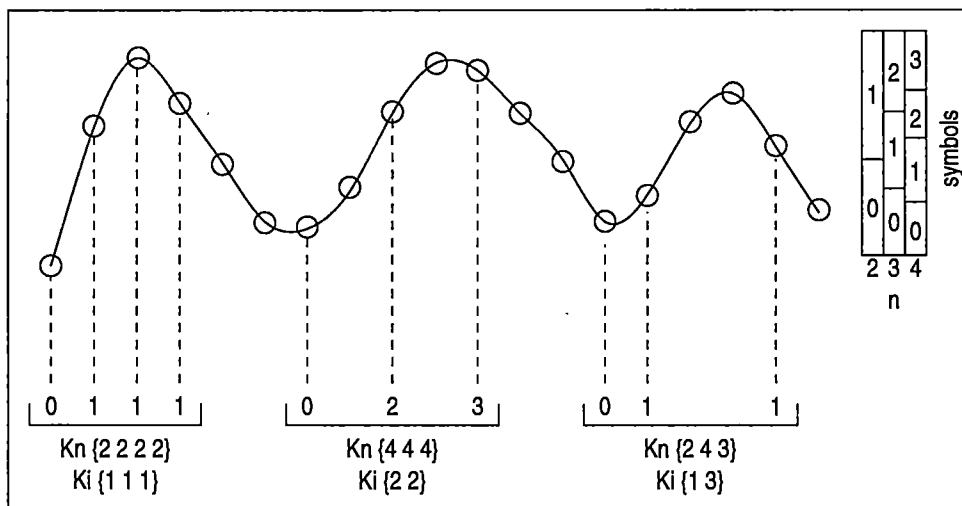


Figure 3: Symbolization and interval keys and resulting symbol sequences.

(n^L for homogeneous symbolization and $\prod K_n$ generally).

3.1.3 The Monte Carlo probability

Using the method of surrogate data, a measure (usually a simple statistic) is compared for the original data and a large sample of surrogate realizations. There are two ways to test a chosen measure. First, if known that the distribution of the measure is Gaussian, then the significance may be computed using

$$S = \frac{|M_{\text{orig}} - \langle M_{\text{surr}} \rangle|}{\sigma_{\text{surr}}}, \quad (8)$$

where M_{orig} is the measure value for the original time series, $\langle M_{\text{surr}} \rangle$ is the expected (mean) value of all the surrogate time-series measures, and σ_{surr} is the standard deviation of the surrogate time-series measures. After Rapp, Albano, Zimmerman, and Jiménez-Montaño (1994), the corresponding probability is then

$$P_G = \Pr[M \leq M_{\text{orig}}] = \frac{1}{2}[1 + \text{erf}(S/\sqrt{2})]. \quad (9)$$

Thus, $S > 2.6$ rejects the null hypothesis that the original and surrogate time series are the same with significance level $\alpha = 0.01$, or with 99 percent confidence.

Alternately, the *Monte Carlo probability* (Rapp, Albano, Zimmerman, and Jiménez-Montaño 1994) may be evaluated if it is not known whether M follows a Gaussian distribution:

$$P_M = \frac{\#\{i : M_i \leq M_{\text{orig}}\}}{N_{\text{surr}} + 1}, \quad (10)$$

where N_{surr} is the number of surrogate trials. Under a null hypothesis $H : M_{\text{orig}} \leq M_{\text{surr}}$, the null is rejected if $M_{\text{orig}} < M_{\text{surr}}$, for all surrogates, with p -value $p = 1/(N_{\text{surr}} + 1)$. Significance probabilities also may be calculated exactly using less-strict criteria.

This procedure may be modified for the inverse null hypothesis, $H : M_{\text{orig}} \geq M_{\text{surr}}$, and adapted for two-tail tests.

3.2 Test for symbolic stationarity

Two time series are defined to be *symbolically stationary* if their symbol statistics are not significantly different. Because “significantly different” depends on how the difference is measured, this definition is not rigorous in itself. Stationarity, then, depends on the context of usage.

Evaluation of stationarity may be used to determine whether a system has achieved ergodicity (e.g., passed startup or changepoint transients) or whether a reference state is dynamically similar to a reference or library state. As mentioned in **Section 2.2.3**, nonlinear time-series analysis practice lacks an accepted, standard test, partly because stationarity is not a well-defined concept. In the context of symbolic analysis, verifying stationarity on the basis of a symbolic test is a reasonable and consistent approach.

A simple difference metric between two symbol-sequence histograms is defined using a norm, such as Euclidean, maximum or superior, or chi-square. The Euclidean

norm is employed in this study, and for two time series A and B is defined as

$$T_{\text{stat}} = [\langle (\mathbf{p}_i^A - \mathbf{p}_i^B)^2 \rangle]^{\frac{1}{2}}, \quad (11)$$

where \mathbf{p} denotes the symbol-sequence histogram and $\langle \cdot \rangle$ denotes the expected value evaluated over all sequences in the histogram.

Ascribing statistical significance to T_{stat} is problematic because sequences in each symbol-sequence histogram are temporally dependent. For instance, the binary sequences 0 1 0 and 1 0 \square are correlated, as the latter sequence is a temporal shift of the former (\square could be either 0 or 1). Direct sequence-by-sequence comparison of the symbol-sequence histograms is possible but evaluating significance introduces the problem of multiple testing (Miller 1966; Timmer, Schwarz, Voss, Wardinski, Belloni, Hasinger, van der Klis, and Kurths 2000) along with presumption of the distribution of the test statistic.

An intrinsic test may be fashioned using Monte Carlo probabilities. Under the null of stationarity, the expected value of T_{stat} is zero. The range of variability may be evaluated by calculating T_{stat} for surrogate symbol series or histograms, with the final probability calculated according to Equation 10, where T_{stat} serves as the measure M . The use of the Monte Carlo process allows direct nonparametric (*i.e.*, not assuming Gaussianity) estimation of significance and obviates the need to account for correlation in symbol sequences but is computationally more intensive than direct evaluation with an *a priori* estimation of significance.

The test procedure is as follows. Two time-series segments \mathbf{x}_A and \mathbf{x}_B are selected; these segments could be the front and back halves of a long time series (to test for

dynamical stationarity), two time series measured from the same process at separated times (to test for process stationarity), or one a test series and the other a library series (to test for achieving process dynamical similarity). Each time series is then transformed according to Equation 6, and the symbol-sequence histograms \mathbf{p}_A and \mathbf{p}_B are tallied from \mathbf{c}_A and \mathbf{c}_B . Then, $T_{\text{stat}}^{\text{orig}}$ is calculated, according to Equation 11. For each of N_{surr} (typically 50–200) surrogates, the symbol series \mathbf{s}_A and \mathbf{s}_B are shuffled (i.e., the order is randomized, much like shuffling a deck of cards) using the randomization operator

$$R(\cdot) : \mathbf{x} \mapsto \tilde{\mathbf{x}}, \quad (12)$$

where $\tilde{\mathbf{x}}$ is the randomly shuffled version of \mathbf{x} . Thus, for each randomized pair of surrogate symbol series $\tilde{\mathbf{s}}_A$ and $\tilde{\mathbf{s}}_B$, a value of $T_{\text{stat}}^{\text{surr}}$ is calculated. When all of the values of $T_{\text{stat}}^{\text{surr}}$ have been recorded over all randomized surrogates, then the Monte Carlo probability may be computed according to Equation 10 and the null hypothesis of stationarity accepted or rejected based on P_M .

3.3 Test for temporal reversibility

A time series is defined to be *symbolically temporally reversible* if its symbol statistics are not significantly different under a temporal reversal transformation. As with symbolic stationarity, the term “significantly different” depends on how the difference is measured, so this definition is not rigorous. As discussed in Section 2.2.4, rejection of a null hypothesis of temporal reversibility rules out Gaussian models, or static transformations of a Gaussian, as a source for the time series. ⁶

For an observed time series, temporal reversibility may be tested by comparing the original time series with a temporally reversed version of itself. The reversal may be achieved using the reversal operator

$$\mathbb{T}(\cdot) : \mathbf{x} \mapsto \mathbf{x}, \quad (13)$$

where \mathbf{x} is the reversed version of \mathbf{x} .¹

In this study, the test used for measuring temporal reversibility is algorithmically similar to that used for stationarity. The metric is the Euclidean norm between the symbol-sequence histogram of the time series and its temporal inverse:

$$T_{\text{rev}} = \langle (\mathbf{p}_i - \mathbf{p}_i)^2 \rangle^{\frac{1}{2}}, \quad (14)$$

where \mathbf{p} denotes the symbol-sequence histogram and $\langle \cdot \rangle$ denotes the expected value evaluated over all sequences in the histogram.

For hypothesis testing, temporally reversible surrogates of a code series may be generated using the symmetrization operator

$$\mathbb{W}(\cdot) : \mathbf{c} \mapsto \mathbf{c}. \quad (15)$$

The temporal symmetrization is achieved by flipping each temporally asymmetric sequence in \mathbf{c} between its time-forward instance or its time-reverse instance with probability $\frac{1}{2}$. The result is a temporally reversible code series, within fluctuations of finite-sample statistics. As a concrete example for $n = 2$ and $L = 3$, possible

¹Explicitly, if $\mathbf{x} = (x_1, x_2, \dots, x_{N-1}, x_N)$, then $\mathbf{x} = (x_N, x_{N-1}, \dots, x_2, x_1)$.

sequences are: 0 0 0, 0 0 1, 0 1 0, 0 1 1, 1 0 0, 1 0 1, 1 1 0, 1 1 1. Sequences 0 0 0, 0 1 0, 1 0 1 and 1 1 1 are temporally symmetric, as their forward and reverse instances are the same (e.g., 0 0 0 forward is the same as 0 0 0 backward). The other sequences are temporally asymmetric.

Alternately, the metric in Equation 14 may be rewritten as

$$T_{\text{irr}} = \langle (\mathbf{p}_{\vec{i}} - \mathbf{p}_{\leftarrow i})^2 \rangle^{\frac{1}{2}}, \quad (16)$$

where $\mathbf{p}_{\vec{i}}$ represents the collection of time-forward instances of temporally asymmetric sequences and $\mathbf{p}_{\leftarrow i}$ their corresponding time-reverse instances. For instance, if $\mathbf{p}_{\vec{i}}$ represent the collection of sequences 0 0 1 and 0 1 1, then $\mathbf{p}_{\leftarrow i}$ would represent the collection of sequence 1 0 0 and 1 1 0, the corresponding time-reverse sequences; redundant sequence pairs are ignored (e.g., the difference between 0 0 1 and 1 0 0 and between 1 0 0 and 0 0 1 are the same, so only one difference is recorded in Equation 16).

The test procedure is as follows. Given time series \mathbf{x} , the time series is transformed according to Equation 6, and T_{irr} is calculated from Equation 16. For each of N_{surr} (typically 50–200) surrogates, a temporally reversible surrogate code series $\mathbf{c}_{\leftrightarrow}$ is created, and from this a value of $T_{\text{irr}}^{\text{surr}}$ is calculated. When all of the values of $T_{\text{irr}}^{\text{surr}}$ have been recorded over all randomized surrogates, then the Monte Carlo probability may be computed according to Equation 10 and the null hypothesis of temporal reversibility accepted or rejected based on P_M .

This test relies on the method of surrogate data. Another test, targeted false flipped symbols (TFFS), has recently been developed and provides an *a priori* estimate of statistical significance of a symbolization-based test without the need for generating

surrogates. Daw, Finney, and Kennel (2000) provide a discussion to the advantages of TFFS over the method employed in this study.

The above test for temporal reversibility directly compares symbol statistics of a time series by measuring the sensitivity of the time series of simple temporal reversal. Another test may be fashioned under the premise that rises and falls in data are, statistically, of equal duration, which would be expected of temporally reversible data.² Naïvely, a statistic to measure such behavior would be calculated by tallying the number of positive versus negative first differences in the data.³ The utility of the statistic should not be limited to first differences, however, as longer timescales may be more sensitive to rises and falls in data. Therefore, a lagged difference may be used more generically instead of the first difference, which has unit lag.

The difference symbolization statistic is defined as

$$S_{\Delta}(\tau) = \langle \text{sign}[x_{i+\tau} - x_i] \rangle, \quad (17)$$

where τ is the lag, $\text{sign}[\cdot]$ is the algebraic sign indicator $(-1, 0, +1)$, x_i is record x at index i , $x_{i+\tau}$ is record x at index $i + \tau$, and $\langle \cdot \rangle$ is the expected value. When $\tau = 1$, $x_{i+1} - x_i$ is the familiar first difference of the data series. For temporally reversible data, $E[S_{\Delta}] = 0$, where $E[\cdot]$ is the probabilistic expectation. The significance of S_{Δ} may be determined using the method of surrogate data and the Monte Carlo probability, based on a two-tailed test, as S_{Δ} may be positive, negative, or zero. In this case, surrogate data sets are created by randomly shuffling the original time

²This observation is ubiquitous in the literature; *vid.* Section 2.2.4.

³Difference symbolization is termed by Kurths et al. (1996) as a *dynamic transformation*.

series before applying the difference symbolization. A shuffle of a reversible time series should have minimal impact on S_Δ , whereas a shuffle of an irreversible time series should force S_Δ to approximately zero.

The test procedure is as follows. Given time series \mathbf{x} , the symbolic lagged-difference statistic S_Δ^{orig} is calculated, according to Equation 17. For each of N_{surr} (typically 50–200) surrogates, a randomized time series $\tilde{\mathbf{x}}$ is created and S_Δ calculated. When all of the values of S_Δ^{surr} have been recorded over all randomized surrogates, then the Monte Carlo probability may be computed according to a two-tailed version of Equation 10 and the null hypothesis of stationarity accepted or rejected based on P_M .

Note that a useful signature of irreversible timescales within a time series may be obtained from a functional form of S_Δ , *i.e.*, evaluated over a range of τ .

3.4 Test for symbolic synchronization

Two or more physically coupled systems are defined to be *symbolically synchronized* if, within a finite period of observation, their symbolic dynamics are statistically similar, either with instantaneous or lagged simultaneity. Depending on the degree of coupling and on the regularity of each independent system, synchronization may be episodic or protracted. *Synchrograms* are graphical displays of synchronous behavior, and the symbolization-based synchrograms developed in this study are motivated by the lack of methodology in nonlinear time-series analysis practice for registering synchronization among more than two measurement series, either continuous or symbolized.

The symbolic synchrogram depends first on the ability to encode multiple (two or more) symbol series, each arising from a different measurement series. **Figure 4** illustrates the process of multi-series encoding. In the figure, there are three symbol series, A , B and C , each symbolized from three separate but simultaneously measured time series. The process of encoding relies on grouping symbols together. In single-series encoding, all elements of the symbol sequence occur sequentially from the same symbol series; in multi-series encoding, symbols are taken from different symbol series to form a “sequence”, which might not be in temporal order after the grouping.

To specify which symbol series and in what order symbols are to be grouped, two keys, the series key and the interval key, are needed. The *series key* specifies from which symbol series symbols are taken at each step in the sequence. The *interval key*, denoted K_i , specifies the interval of separation between steps in the sequence. Because multiple symbol series are used, it is possible for an interval to be negative as well as positive. **Figure 4** explains this process schematically. In the figure, the series key is $\{A A B B C C\}$, meaning that the first two steps in the sequence draw symbols from series A , the third and fourth steps from series B , and the fifth and sixth steps from series C . The interval key is $\{+1 -1 +1 -1 +1\}$, meaning that step two lags step one by an interval of one, but step three and step one are simultaneous (because the -1 decrements the pointer, and $+1 -1 = 0$). Similarly, step five and step one are simultaneous. The combined effects of the series key and the interval key are seen in the (red) order guide numbers 1–6. The symbols adjacent the guide numbers are combined, in order 1–6, into the specified sequence $0 1 0 1 0 1$, which then is encoded into code 21. In the resulting code series, sequence $0 1 0 1 0 1$ occurs twice — all three symbol series (A, B, C) exhibit self-anticorrelation ($0 1$) but

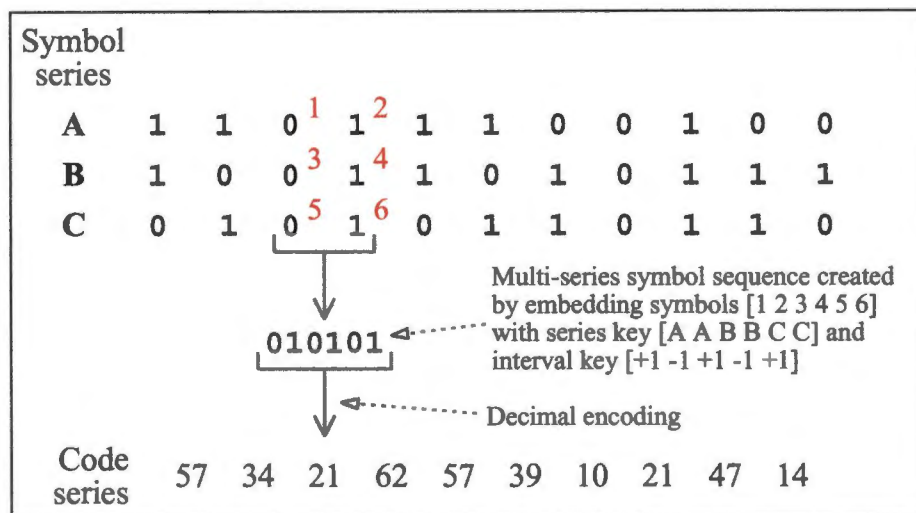


Figure 4: Construction of a multi-series code series used in the symbol synchrogram.

mutual synchronization (at the first timestep, all three series have a symbolic value of 0; at the next timestep, all three series have a symbolic value of 1).

The *symbol synchrogram* is a graphical display of the code series, with time on the abscissa and the code at each temporal index on the ordinate. When certain codes dominate the mutual behavior of the symbol series, “banding” (many proximate consecutive instances of the same code) will occur; for asynchronous or weakly synchronized behavior, any banding will be episodic and by chance.

An intrinsic advantage in the symbol synchrogram is the combined systemic state may be determined and viewed at any instant in time. Additionally, the combined systemic symbol statistics may also be tallied, so that the primary modes of mutual behavior, and their transitional states, may be recorded and interpreted.

The goal in constructing the synchrograms is to test the null hypothesis that there is no significant synchronization between time series. Because the synchrogram is a multivariate, temporally embedded code series, a natural, perhaps naïve, choice for statistic is the modified Shannon entropy. For instance, very strongly correlated behavior would result in a finite number of sequence codes predominating the synchrogram, so entropy would be expected to be low; weakly correlated or uncorrelated behavior would result in higher values of entropy. However, the Shannon entropy is nonspecific in which patterns it measures — it simply quantifies the degree of organization of the symbol statistics and cannot discriminate sequences 1, 2 and 3 from 4, 5 and 6.

Therefore, a second, specific statistic is employed, one which measures the relative frequencies of occurrence of sequences associated with the types of synchronization of interest. One statistic is S_+ , which is the sum of frequencies of occurrence for purely

correlated sequences; another is S_- , the sum of frequencies of occurrence for purely anticorrelated sequences. By using the targeted statistics S , only the sequences of interest are counted, thus allowing a much more powerful statistical test.

When significant synchronization occurs, it is expected that the modified Shannon entropy for the original synchrogram should be less than entropy values for the phase-randomized surrogate synchrograms; *i.e.*, the phase randomization destroyed real synchronization, and the resulting symbol statistics are more complicated. Algorithmically, the phase randomization is achieved by selecting the starting locations of each symbol series at random, so that each surrogate synchrogram should have a different phase relationship than the original data. When the synchronization is in the form of strong correlation, it is expected that the statistic S_+ should be higher than those for the phase-randomized surrogates. Similarly, when the synchronization is in the form of strong anticorrelation, it is expected that the statistic S_- should be higher than those for the phase-randomized surrogates. It should be noted that S_+ and S_- could fail to account for synchronization based on systematic phase slipping (switching between correlated and anticorrelated behavior), so an alternate statistic $S_{\pm} = S_+ + S_-$ could then be employed. The significance of these statistical tests can be measured with a Monte Carlo process employed in previous sections.

The test procedure is as follows. Given time series $\mathbf{x}_1 \dots \mathbf{x}_M$, code series \mathbf{c} is created by using a multivariate transformation described by the parameters K_s , K_n and K_i (see **Figure 4** for a schematic). Then, H_{SM} and/or S are calculated for \mathbf{c} . For each of N_{surr} (typically 50–200) surrogates, a phase-randomized code series $\tilde{\mathbf{c}}$ is created, from which H_{SM}^{surr} and/or S^{surr} are/is calculated. When all of the values of the statistical metrics have been recorded over all randomized surrogates, then the

Monte Carlo probability may be computed according to Equation 10 and the null hypothesis of stationarity accepted or rejected based on P_M .

As mentioned above, the appropriate surrogate transformation is to randomize the phase relationships of the time series used to construct the synchrograms. Algorithmically, the phase randomization is achieved by selecting the starting locations of each symbol series at random, so that each surrogate synchrogram should have a different phase relationship than the original data. In the present study, the starting locations are chosen at random (uniformly) in the interval of the first 10 percent of each time series. Each resulting code series, including that for the original data, is then limited to a length of 90 percent of the original time series. In this manner, all code series will have the same length, but each, to varying degrees, neglect portions of the first and last 10 percent of the original time series. The phase-randomization process, as implemented, loses some information in the original data, but for long, stationary data sets, this information loss should not be significant or bias the test.

Chapter 4

Results and discussion

4.1 Stationarity

This section presents application of the stationarity test presented in **Section 3.2** to measurement data from a fluidized-bed system.

4.1.1 Data

Data from a laboratory-scale fluidized bed system are examined using the stationarity metric and test. Two types of stationarity are measured: dynamical state matching and dynamical stationarity. The data were obtained by the author and by M. VASUDEVAN of the University of Tennessee.

4.1.1.1 Experimental apparatus and procedure

The experimental fluidized-bed data were obtained at the Fluidization Laboratory of the University of Tennessee. **Figure 5** displays the fluidized-bed system schematically. The fluidized-bed vessel was a 10.2-cm inner diameter by 260-cm tall Plexiglas tube with a perforated-plate distributor. House air was used as the fluidizing gas. Dynamic differential-pressure measurements were obtained from taps located 10 and 23 cm above the distributor using a Baratron differential-pressure transducer. The

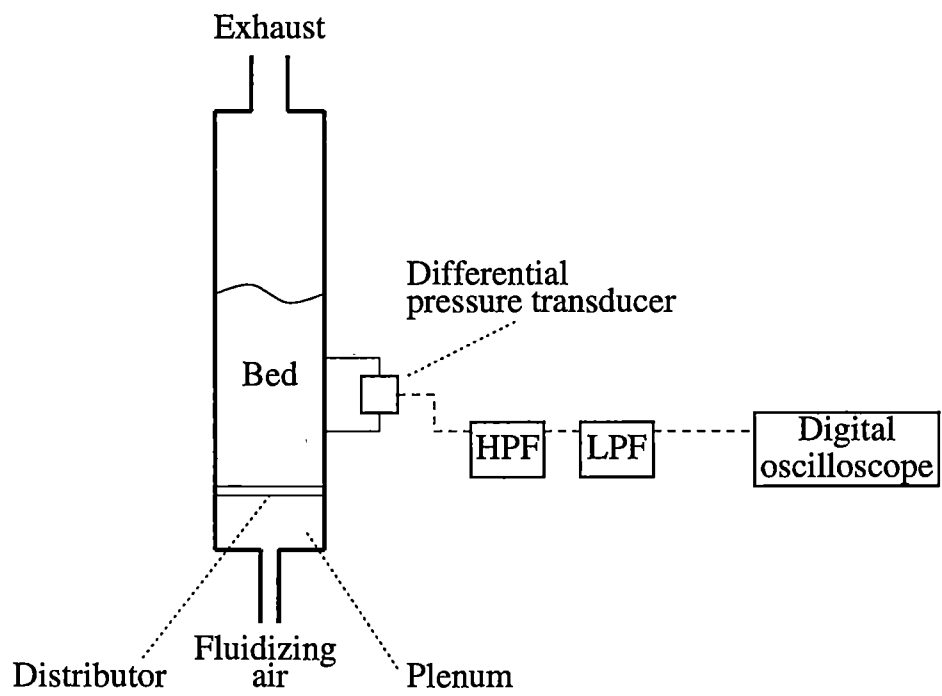


Figure 5: Schematic of the fluidized-bed system.

transducer signal was analog bandpass filtered between 0.1 and 40 Hz to remove long time-scale instrument drift and 60-Hz power-line contamination before being digitized with a 12-bit precision Nicolet digital recording oscilloscope. Time series were recorded at 200 samples per second for 300 seconds, for a total of 60000 records.

Glass ballotini with a mean diameter of 0.27 cm and a density of 2.6 gm/cm³ were used as the fluidizing solids; these particles are classified as Geldart Group D and reach minimum fluidization at approximately 142 cm/sec air velocity at ambient conditions (operating temperature of 298 K and pressure of 1 atm). For the complete measurement record, two groups of data were obtained. The first group, termed the *reference* group, was recorded at 16 different gas-flow conditions; at each flow condition, two time series (replicates) were recorded within several minutes of each other to test for long-term stationarity. The second group, termed the *test* group, was recorded at 8 different gas-flow conditions, with two time series per condition. **Figure 6** shows schematically the relationships between the reference and test flow conditions. The reference time series are labeled 1 through 16, and the test time series are labeled A through H. Additionally, the first replicate at each condition is labeled “a”, the second “b”.

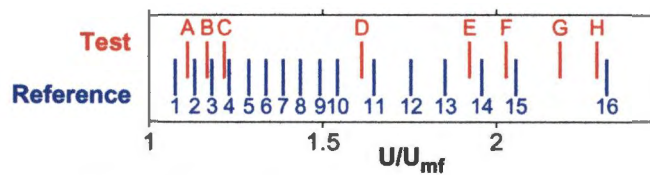


Figure 6: Relationship of the relative conditions for the reference and test sets for the fluidized-bed data.

4.1.1.2 Dynamical behavior of the fluidized-bed data

The fluidization behavior of the ballotini varies significantly over the range of conditions encompassed in the measurement record, with states changing from bubbling to slugging to turbulent transition in a manner consistent with that described by Vasudevan, Finney, Nguyen, van Goor, Bruns, and Daw (1995).

Figure 7 shows short time-series segments from the first replicates of all 16 conditions of the reference group; the segments are standardized (converted to zero mean and unit variance) for visual comparison. All flows are standardized with respect to the minimum fluidization velocity, U_{mf} , the gas flow at which incipient fluidization is observed. Near minimum fluidization (1), the bed is in a bubbling regime, characterized by irregular oscillations with flattened peaks in the pressure trace. Here, bubble size is quite variable and unstable. With further increase in gas flow (2), the bed transitions into the slugging regime, characterized by more regular oscillations in the pressure trace. The slugging regime is quite stable with further increases in flow (3–4), although variations in the pressure-cycle magnitudes and waveforms are discernible. With further increases in flow (5–12) past the maximum-stable slugging condition (4), the regularity of the slugging rhythm is interrupted occasionally with brief episodes of smaller-than-average pressure cycles; these interruptions become more frequent with increasing flow. The bed enters the turbulent transition regime (13–15) at still higher flows; here, slugging is very violent and punctuated by brief episodes of swirling of particles near the distributor before the resumption of slugging. Near turbulence is seen at the highest flow condition (16), at which the dominant mode of bed motion is vigorous mixing of particles in the lowest area of the bed; bubbles or slugs rarely

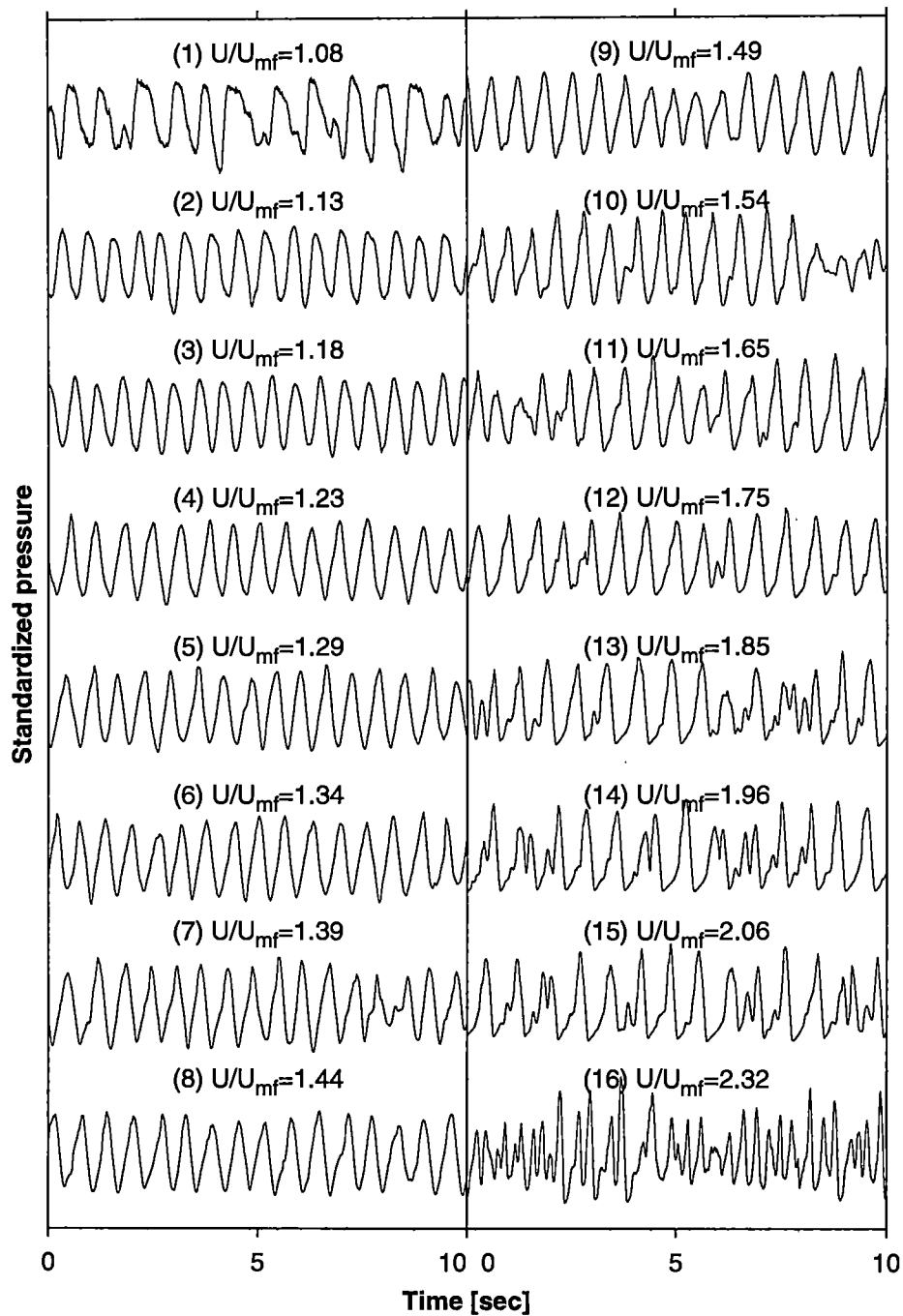


Figure 7: Short time-series segments for the reference group from the fluidized bed over a range of fluidization conditions. Segments are standardized to aid in visual comparison.

are observed.

The time series in the test group were taken at flow conditions between those taken in the reference group. **Figure 8** shows short time-series segments from the first replicates of all 8 conditions of the test group; the segments are standardized (converted to zero mean and unit variance) for visual comparison. Visually, these time series appear consistent with their nominal flow value and with the corresponding reference sets. For instance, test set A is between reference sets 1 and 2, and its trace most closely resembles reference set 1; test set H is between reference sets 15 and 16, and its trace most closely resembles reference set 16.

4.1.2 Results of the stationarity test

An important validation of the stationarity metric is whether this metric can identify a given test case based only on the library of reference cases. The simplest form of this library system is a direct comparison of each reference time series against the test time series. **Figure 9** shows direct comparison of each test case against all reference cases using the stationarity metric T_{stat} . In the top plot, test sets A–C are compared; in the middle, test sets D–F; in the bottom, test sets G–H. In each case, the test case is correctly matched into its appropriate fluidization state (see **Figure 6** for comparison of states based on fluidization gas flow). For instance, test set D dynamically should be most similar to reference sets 10 and 11; the value of T_{stat} for test set D with respect to all the reference sets reaches a minimum at reference set 11, indicating best dynamical similarity. Tests sets G and H both lie between reference sets 15 and 16; for G, T_{stat} is minimal when compared with reference set 15, indicating closest dynamical agreement, whereas for H, best agreement is with reference set 16.

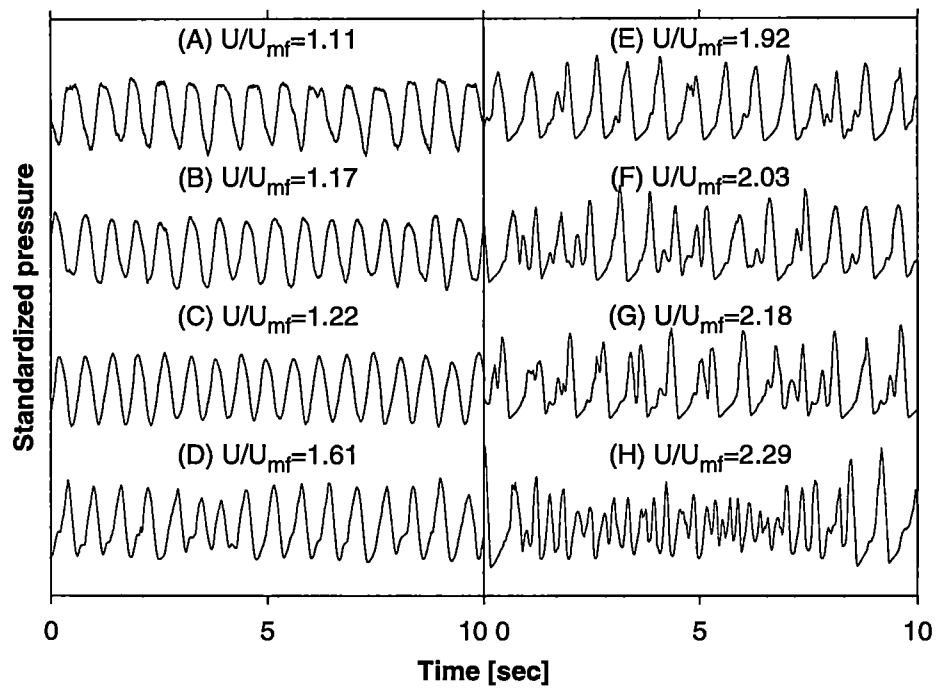


Figure 8: Short time-series segments for the test group from the fluidized bed over a range of fluidization conditions. Segments are standardized to aid in visual comparison.

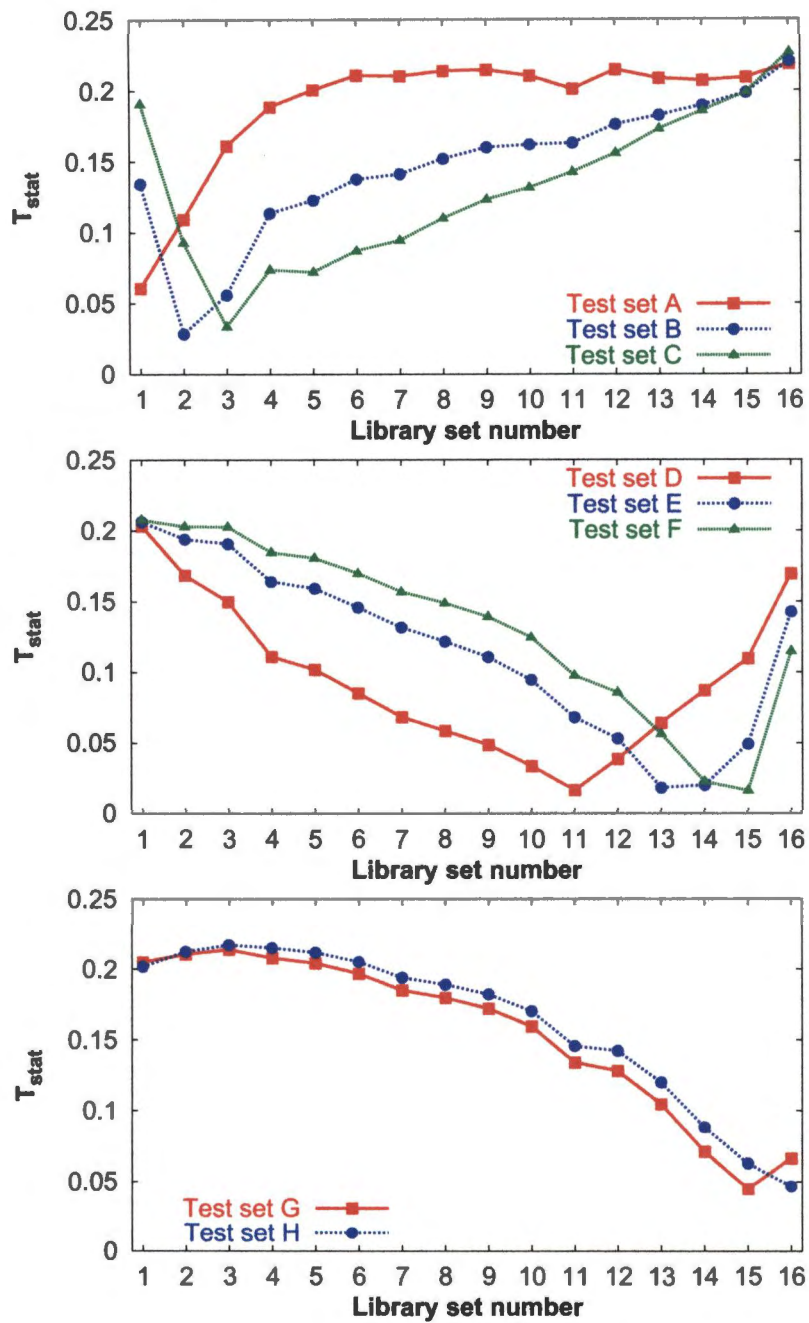


Figure 9: Comparison of the test group against the reference group for the fluidized-bed data using the T_{stat} statistic. In each plot, each test case is compared against each reference case; only the first replicates are used. Symbolization parameters were $K_n = \{4\ 4\ 4\}$ and $K_i = \{21\ 21\}$.

The simple comparisons displayed in **Figure 9** do not address statistical significance. By the test presented in **Section 3.2**, statistical significance can be estimated using Monte Carlo probabilities. In this case, surrogates are generated by randomly shuffling the symbol series and then calculating the stationarity statistic T_{stat} for each surrogate. Thus, the final test determines whether the observed differences in symbol statistics occur from real, dynamical differences or are within finite-sample variations.

An example of the stationarity test is shown in **Table 1**, which gives results for reference set 4 and test set C, its closest neighbor in fluidization state. For all tests, 30000 records from each time series were used so that all time series would be compared on a consistent basis. An important first question is whether a given time series exhibits autostationarity, specifically, whether the symbol statistics in the first half of the time series are statistically the same as those from the the second half. In the table, autostationarity results are shown for the two replicates of reference set 4 and for the two replicates of test set C. For each time series, the stationarity statistic T_{stat} is reported. Additionally, the upper bound for 95 percent confidence based on 200 surrogate trials (T_{95}) and the p -value of the statistic are reported. From the table, in the "Autostationarity" block, it is seen that time series 4a, 4b and Ca are statistically stationary, as their p -values exceed the defined significance of 0.05. However, set Cb is statistically symbolically nonstationary and is excluded from further analysis.

From the table, in the "Stationarity" block, time series 4a, 4b and Ca are all statistically different based on their symbol statistics. All trials strongly reject the null hypothesis ($p < 0.005$) of stationarity (in this case, dynamical equality), but the differences between 4a and Ca and 4b and Ca are much greater than that between 4a and 4b. In the former case, the primary difference is the fluidization state, determined

Table 1: Stationarity test results for the fluidized-bed data for reference set 4, replicates a and b, and test set C, replicates a and b. T_{95} is the 95 percent upper confidence bound on T_{stat} . Symbolization parameters were $K_n = \{4\ 4\ 4\}$ and $K_i = \{21\ 21\}$.

Set(s)	T_{stat}	T_{95}	p -value
Autostationarity			
4a	0.008038	0.008957	0.279
4b	0.008322	0.008912	0.184
Ca	0.007111	0.008783	0.642
Cb	0.012106	0.008871	<0.005
Stationarity			
4a & Ca	0.072131	0.008671	<0.005
4b & Ca	0.062969	0.008783	<0.005
4a & 4b	0.024364	0.008696	<0.005

by the gas flow; in the latter, weak differences between replicates measured at the same experimental conditions several minutes apart are suggestive of a long time-scale drift inherent in fluidized-bed dynamics (Daw 2000). Because of the Monte Carlo methodology employed here, acceptance probabilities are discrete, so that the differences between 4a and Ca are registered with the same strength as those for 4a and 4b. Had a continuous probability distribution (e.g., Gaussian) been employed, then the rejection strengths would have been registered as clearly as they are visible in the table.

It should be noted that the degree of measured stationarity or nonstationarity depends on the symbolization parameters used to construct the symbol statistics — in some instances, dynamics will appear similar, whereas in others, dynamics will appear dissimilar. The problem of choosing parameters for statistical tests is ubiquitous except in the simplest, and least informative, tests and should not be viewed as a deterrent in implementing a symbolization-based stationarity test. A reasonable practice would be to choose a small set of reasonable symbolization parameters and to apply the test multiple times, using aggregate results to determine acceptance overall acceptance or rejection of the null. Additionally, if other symbolization-based tests such as temporal reversibility are to be applied, then testing for stationarity based on the symbolization parameters used in the other test would be reasonable and would offer comparison on a consistent basis.

4.2 Temporal irreversibility — T_{irr}

This section presents application of the temporal-irreversibility test based on a static symbolization transformation presented in **Section 3.3** to measurement data from an internal combustion engine.

4.2.1 Data

Data from a production internal combustion engine are examined for evidence of temporal irreversibility. The data were supplied by F.T. CONNOLLY of the Ford Motor Company. These data are labeled *Ford*.

4.2.1.1 Experimental apparatus and procedure

The Ford data were obtained at the Science Research Laboratory of Ford Motor Company using a 1993-model V-8 engine manufactured by Ford Motor Company. The engine had 4.6 liter total displacement with production port fuel injection, a two-valve head, a bore of 9.02 cm, a stroke of 9.0 cm, and a compression ratio of 9. The nominal operating conditions were 1200 revolutions per minute and 20 degrees before-top-center spark timing. The fuel-air equivalence ratio was controlled by adjusting the fuel-injection pulse width and the throttle position while maintaining a 25 N·m load with an absorbing/motoring dynamometer.

Continuous cylinder-pressure time series were sampled from a head-mounted transducer in one cylinder. The pressure time series were sampled once per crank angle degree so that there were 720 pressure measurements per engine cycle. Estimates of

combustion heat release were computed from the pressure time series using an integral method similar to that of Rassweiler and Withrow (Heywood 1988), so that the resulting time series were discrete, with one heat-release estimate per engine cycle. The time series had approximately 2843 consecutive engine cycles.

4.2.1.2 Dynamical behavior of the engine data

For the types of measurements in the engine described above, the dynamical behavior of combustion quality varies dramatically with fueling changes. Specifically, cycle-to-cycle combustion variations change from Gaussian in nature at stoichiometric fueling to a noisy period-2 behavior at lean fueling (Daw, Finney, Green, Kennel, Thomas, and Connolly 1996; Finney, Green, and Daw 1998; Wagner, Drallmeier, and Daw 1998; Daw, Kennel, Finney, and Connolly 1998; Green, Daw, Armfield, Finney, Wagner, Drallmeier, Kennel, and Durbetaki 1999; Scholl and Russ 1999). In the lean-fueling régime, the period-2 behavior is the result of anticorrelated (alternating higher and lower) variations in combustion quality. As explained in the model of Daw *et al.*, the nonlinear nature of these oscillations is primarily the result of the sensitive dependence of combustion quality on equivalence ratio, which is a function of perturbations to the amount of fuel injected and in the amount of unburned fuel in the residual gas. The severity of the cyclic variability at lean fueling conditions limits the operating range of spark-ignition engines, so understanding the causes of this variability is important for design and control considerations.

The nature of the heat-release oscillations as equivalence ratio changes may be seen in **Figure 10**, which displays typical, short segments of the heat-release time series over a range of equivalence ratios. At the nearly stoichiometric fueling to

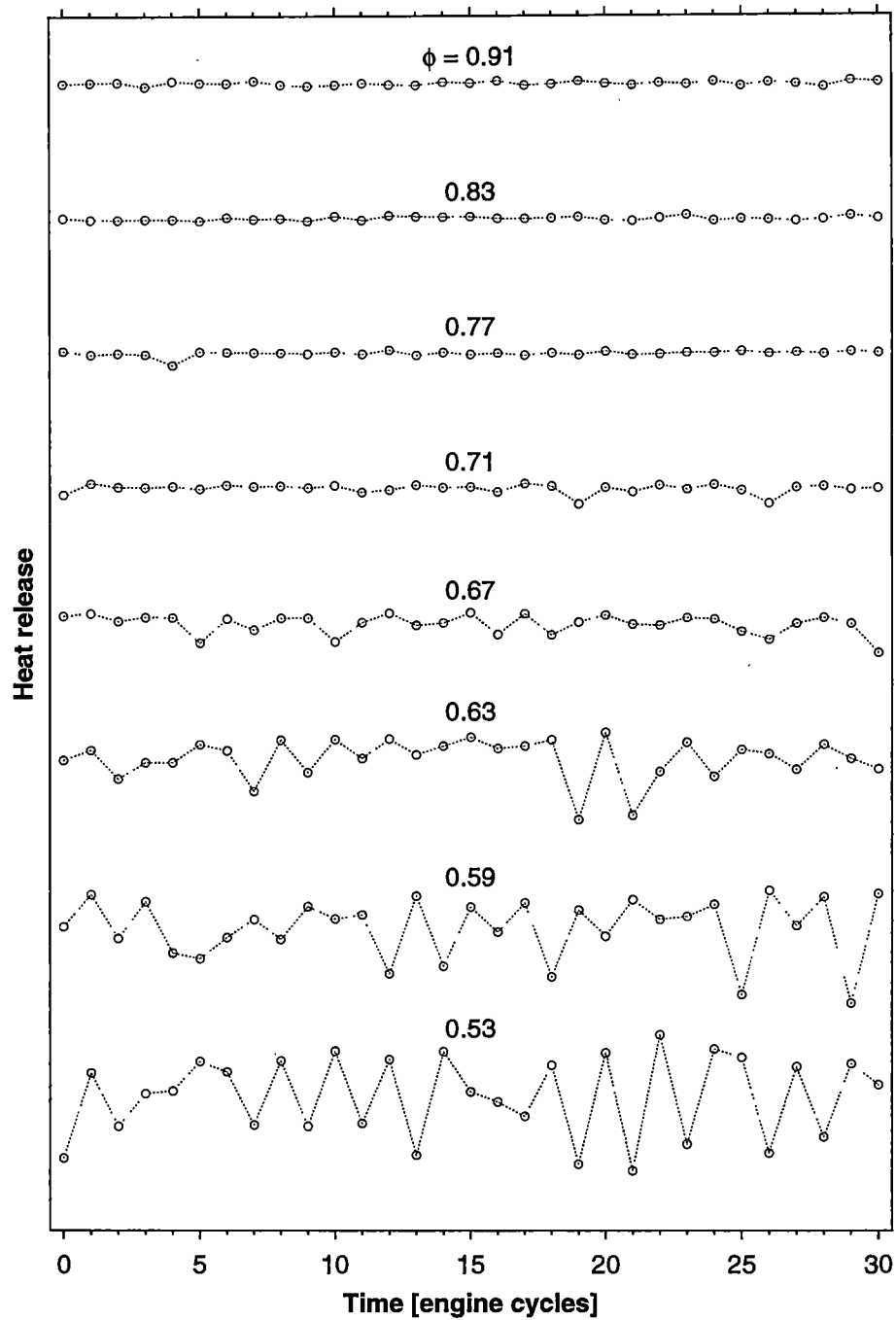


Figure 10: Short time-series segments from the Ford engine at a range of equivalence ratios (ordinate labels). Segments are plotted on the same scale.

moderately lean conditions (equivalence ratios 0.91–0.71), the range of cycle-to-cycle variability in heat release is small and without apparent temporal correlation; in this range, combustion efficiency is most linearly dependent on equivalence ratio, and combustion variations primarily depend on random parametric fluctuations. As the fueling is leaned further (equivalence ratios 0.71–0.63), the range of variability increases (larger variance), and some noticeable oscillations are apparent (e.g., for $\phi = 0.63$, cycles 6–14, with alternating higher and lower heat-release values). At very-lean fueling (equivalence ratios 0.59–0.53), the variability and anticorrelated structure (“sawtooth” appearance of the oscillograms) becomes greatest as the engine dynamics are best described as noisy period-2 in nature. Here, the residual-gas effect dominates the dynamics — unburned fuel left in the cylinder feeds a larger-than-average burn on the next cycle, after which, the fuel inventory being depleted, combustion is weaker than average, resulting in a weaker burn.

From the figure (**Figure 10**), it is difficult to discern the nature of the oscillations at the nearly stoichiometric conditions because their variabilities are so small compared with those at the leanest conditions. To aid visual inspection, the segments may be standardized (converted to zero mean and unit variance), as is shown in **Figure 11**. When viewed in standard units, the nearly stoichiometric conditions also contain frequent oscillations, but it is not obvious whether these oscillations are generated by largely random sources such as gas turbulence and charge mixing or because of the fuel-inventory effect believed to dominate at very lean conditions. This question may be answered by testing for presence of significant irreversibility, which would rule out a Gaussian model, or static transformations of a Gaussian, as the most likely source for the dynamics. Resolution of this question is important to

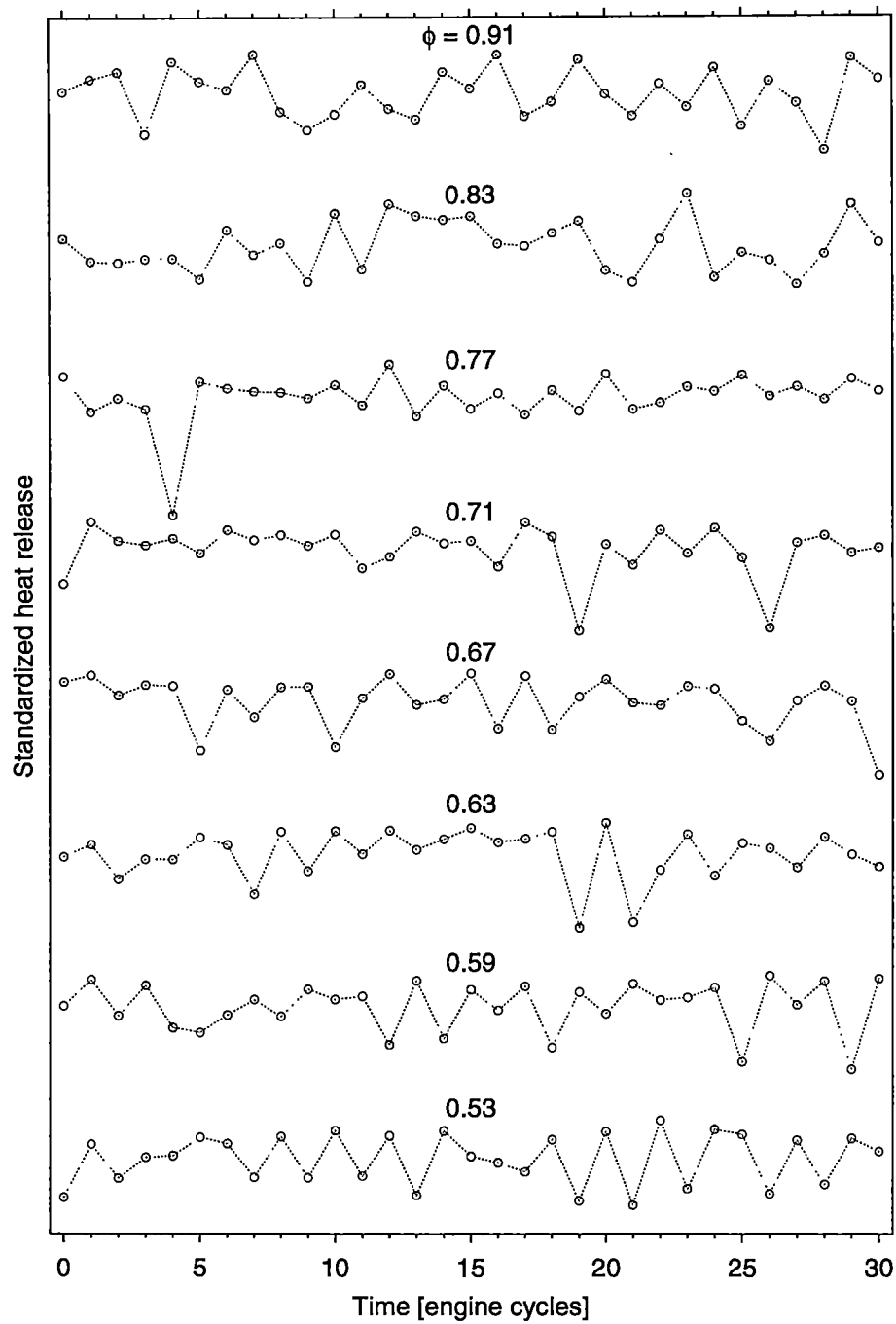


Figure 11: Short time-series segments from the Ford engine at a range of equivalence ratios (ordinate labels). Segments are standardized to aid in visualization of the combustion oscillations.

resolve which type of model, a noisy nonlinear dynamics model or an anticorrelated oscillation (ACO) model¹, best describes (or is prohibited from describing) observed combustion variability.

4.2.2 Results of the symbolic temporal-irreversibility test

Figure 12 displays the results for the temporal-irreversibility test based on the statistic T_{irr} for 200 surrogate trials. In each plot, the T_{irr} statistic for the original time series at each equivalence ratio is displayed with red lines and symbols, and the blue broken line represents the 95 percent confidence interval based on Monte Carlo probabilities (values of T_{irr} greater than the confidence bound signify irreversibility). For each plot, a unit symbolization interval was used, and symbolization keys for each case were $K_n = \{2\ 2\ 2\ 2\ 2\ 2\}$ (a), $K_n = \{4\ 4\ 4\}$ (b), and $K_n = \{8\ 8\}$ (c).

With the binary symbolization (a), at all equivalence ratios the data appear to be temporally reversible². However, with the quaternary (b) and octonary (c) symbolizations, significant temporal irreversibility is registered at equivalence ratios at and below 0.71. The maximal degree of irreversibility occurs at an equivalence ratio of 0.59, where the noisy anticorrelated behavior is most pronounced. For this reason, the anticorrelated-oscillation model is confidently rejected. The noisy nonlinear dynamics model accounts for the temporal irreversibility at very fuel-lean conditions, so it at least is not invalidated based on the lack of temporal irreversibility. This topic

¹In the ACO model, acoustic fluctuations in the intake manifold and fuel-intake system account to the source of the anticorrelated combustion variations. The ACO model is inherently temporally reversible (Green et al. 1999; Daw et al. 2000).

²As discussed in Section 3.3, binary symbolization cannot distinguish temporal irreversibility — only symbol-set sizes greater than two can — as is verified in the present example.

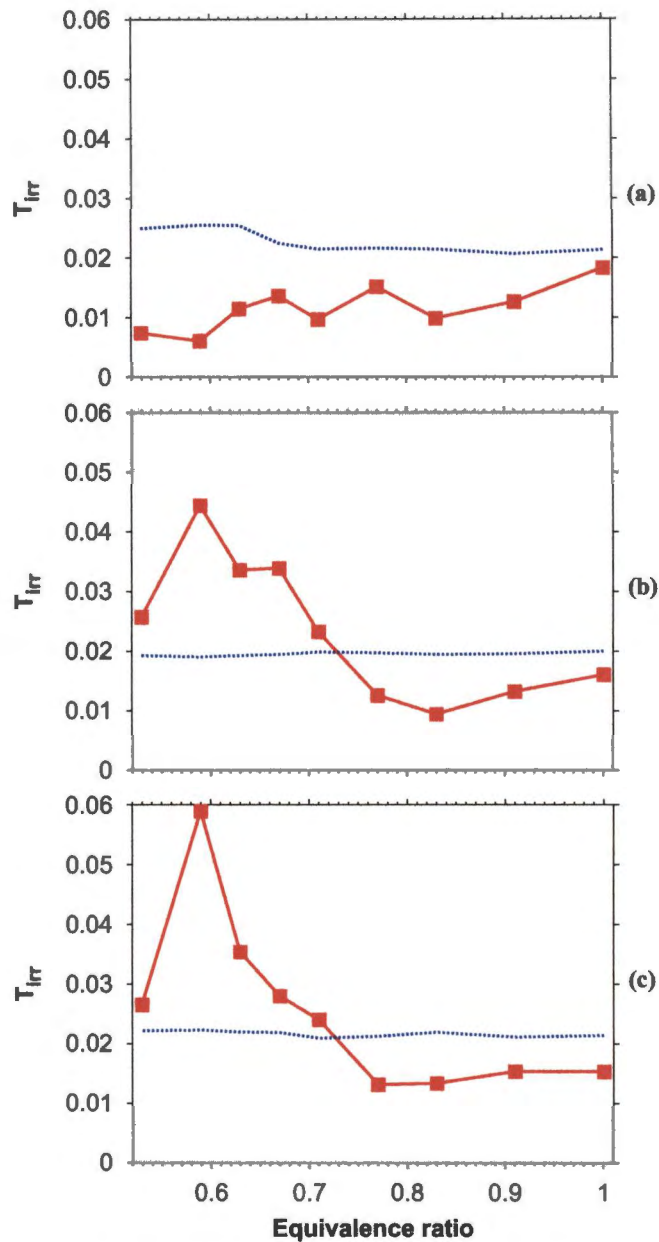


Figure 12: Results of the symbolic temporal-reversibility test for the Ford engine data over a range of equivalence ratios for 200 surrogate trials. Symbolization parameters were $K_n = \{2\ 2\ 2\ 2\ 2\}$ (a), $K_n = \{4\ 4\ 4\}$ (b), and $K_n = \{8\ 8\}$ (c). The broken curves represent the upper bounds for 95 percent confidence for temporal reversibility.

is addressed comprehensively by Green, Daw, Armfield, Finney, Wagner, Drallmeier, Kennel, and Durbetaki (1999) and Daw, Finney, and Kennel (2000).

4.3 Temporal irreversibility — S_{Δ}

This section presents application of the temporal-irreversibility test based on a dynamic symbolization transformation presented in **Section 3.3** to measurement data from a thermal pulse combustor.

4.3.1 Data

Data from a laboratory thermal pulse combustor are examined for evidence of temporal irreversibility. The data were supplied by K.D. EDWARDS of the University of Tennessee (Edwards 2000b).

4.3.1.1 Experimental apparatus and procedure

The data consist of combustion-chamber pressure time series measured from a laboratory-scale thermal pulse combustor operating over an equivalence-ratio range from nearly stoichiometric to flameout in fuel-lean conditions.

Figure 13 displays the pulse-combustor system schematically. The combustor consists of a 0.295-liter, 5-cm inner diameter cylindrical combustion chamber coupled with a 91-cm long tailpipe. A ceramic flameholder is installed in the combustion chamber to provide an ignition heat source and to help stabilize the flame. Air and propane enter the combustor past swirl-inducing vanes at a constant flow rate through separate choked-flow orifices. A spark plug is used to initiate; once ignited,

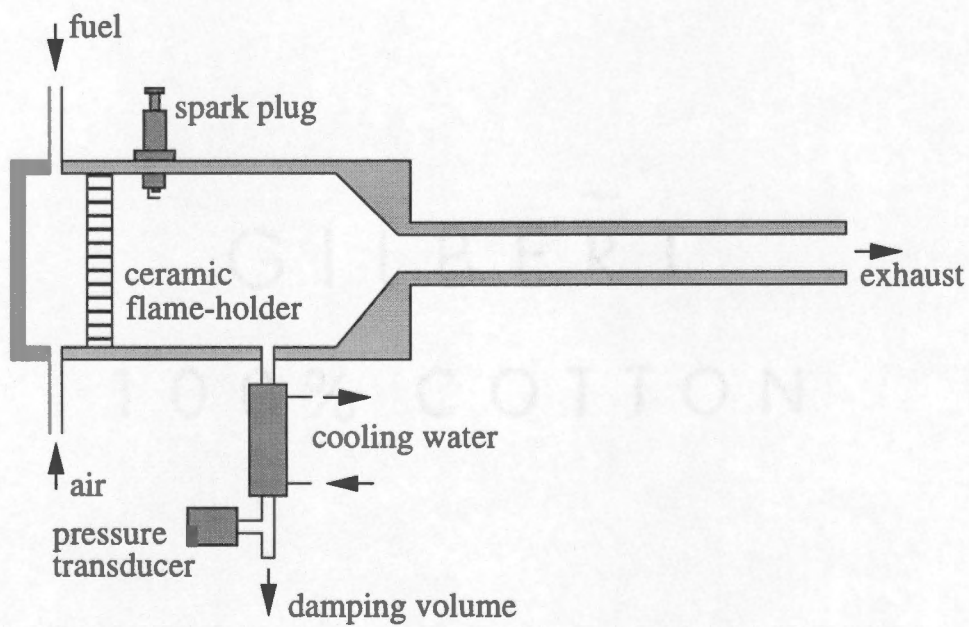


Figure 13: Schematic of the thermal pulse combustor. *Source:* (Edwards et al. 1998).

the combustion reaction is self-sustaining. Although unnecessary for combustor operation, the spark plug is left running to prevent excessive fuel buildup in the case of flameout.

A piezoelectric pressure transducer was used to measure the pressure fluctuations inside the combustion chamber. A damping volume attached to the pressure tap reduces acoustic coupling between the tap and the combustion chamber by lowering the natural acoustic frequency of the tap. The transducer signal was conditioned using a dual-mode amplifier and bandpass filtered from 0.1 to 2000 Hz. Data were digitally recorded at 5000 Hz.

The global equivalence ratio of the fuel-air mixture is controlled by varying the pressure drop across the choked-flow orifices. By monitoring this pressure drop and measuring the orifice flow area, the global equivalence ratio is calculated using compressible-flow relations.

The combustion pulses because it is driven by acoustical oscillations in the combined chamber-tailpipe system. A fraction of the exhaust energy leaving the tailpipe is reflected at the exit boundary, creating a standing compression wave at the natural acoustic frequency of the tailpipe. This acoustic wave propagates back into the combustion chamber, compressing and igniting the fuel-air mixture. Expansion and exhaust of the combustion gases through the tailpipe completes the cycle.

For the purposes of this study, the two experimental parameters which control the pulse-combustor operating state are the global equivalence ratio — the quotient of the actual fuel-air ratio and the stoichiometric fuel-air ratio — and the gas residence time — the time for a unit mass of fuel-air mixture to pass from the gas inlets out of the combustion chamber. The residence time is equal to the quotient of the

combustion-chamber volume and the combined fuel-air volumetric flow rate. A series of experimental data typically is measured by setting a fixed residence time and then adjusting the fuel-air ratio to achieve different equivalence ratios. Edwards (2000a) provides more information about the relationship of equivalence ratio and residence time on the nature of pulse-combustor operation.

4.3.1.2 Dynamical behavior of the pulse combustor

The dynamical behavior of the thermal pulse combustor varies dramatically over a range of global equivalence ratios and residence times. **Figure 14** illustrates typical time-series segments over a range of equivalence ratios at a fixed residence time. For a set residence time, combustion dynamics are observed to vary with equivalence ratio. At nearly stoichiometric fueling (in the figure, $\Phi = 1.004$), pressure oscillations are fairly regular, with small variations in peak pressure related to cycle-to-cycle combustion variations. As equivalence ratio decreases into the fuel-lean operating régime ($\Phi = 0.888$ to $\Phi = 0.700$), cycle-to-cycle variability actually decreases, largely because in this region the combustion efficiency is very stable with respect to equivalence ratio (at leaner or richer conditions, there is a nonlinear relation between combustion efficiency and equivalence ratio). Less discernibly, the dominant frequency of oscillation decreases from 117 Hz at $\Phi = 1.004$ to 102 Hz at $\Phi = 0.700$. This frequency change is caused primarily by the operating temperature of the combustor decreasing with decreasing fuel-air equivalence, and secondarily by the gaseous composition changing to include more air than propane (because air and propane behave similarly as ideal gases with similar molecular weights, this effect is negligible compared with the

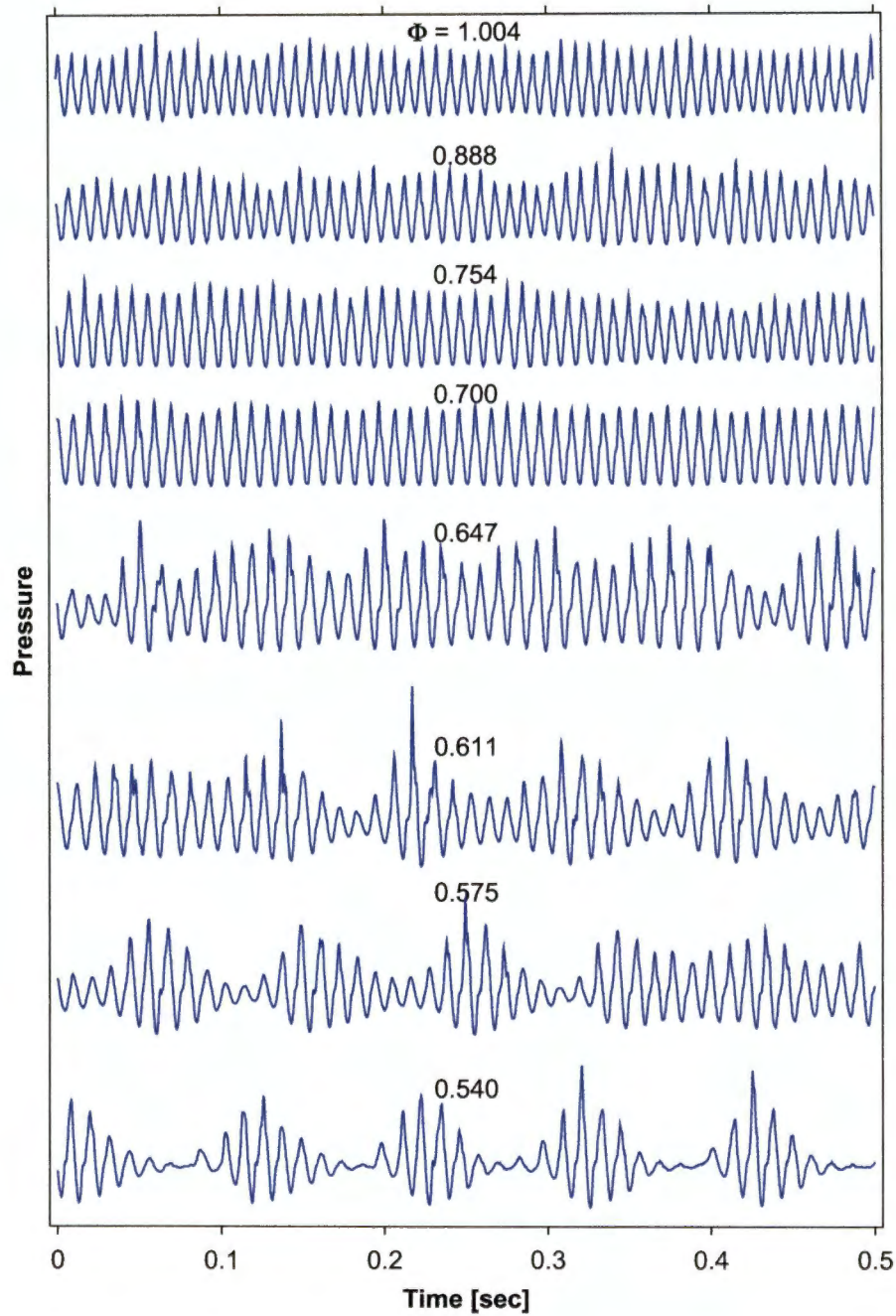


Figure 14: Short pressure time-series segments from the thermal pulse combustor at a residence time of 50 msec. Segments are listed in decreasing equivalence ratio, from nearly stoichiometric (topmost) to very fuel lean, approaching flameout, (bottommost) and are plotted on the same scale.

temperature decrease). As fueling is leaned further ($\Phi = 0.647$), combustion variability becomes more structured and pronounced, with modulation of trough values as well. The amplitude-modulation becomes more pronounced and almost “regular” as equivalence ratio is decreased ($\Phi = 0.540$). At these lowest equivalence ratios, there is a competing ignition-extinction process, in which, because of the overall limited amount of fuel in the mixture, the combustor must build up an inventory of fuel to burn well, but once the inventory is depleted, the combustion is very poor (or negligible). As a result, cyclic pressure oscillations vary from nearly zero to very large in packets of approximately eight acoustic cycles, and the maximal peak-to-trough variations exceed those of the more regularly behaving conditions at higher equivalences. This ignition-extinction process is thought to occur in many chemically reactive systems and has been observed in other combustion systems such as internal combustion engines (Daw, Finney, and Kennel 2000) and utility burners (Daw 1999). Because operating in and extending the usability of the fuel-lean régime is an apposite engineering concern, understanding the nature — linear or nonlinear — of the combustion-magnitude oscillations is important for choosing design and control parameters.

To study the amplitude modulation of the pressure-measurement time series and thus the variability of combustion strength, the continuous time series depicted in **Figure 14** are discretized by recording the peak-to-trough magnitude of each pressure cycle. In the pressure signal, a cycle is defined as the upward transversal of the signal through the time-averaged signal mean to a peak point, followed by a downward transversal of the signal through the mean value to a trough point, and ending with an upward approach of the signal to the mean value. A cycle magnitude is then the

algebraic difference of the peak value and the trough value. By recording the cycle magnitude for each cycle throughout the pressure time series, the discrete time series is created.

Figure 15 shows short segments of the cycle-magnitude time series created from the full pressure time series partially displayed in **Figure 14**. In the segments for equivalence ratios from 1.004 to 0.700, the cycle-magnitude variability is relatively small and without apparent structure. In the segments for equivalence ratios from 0.647 to 0.540, the cycle-magnitude variability becomes markedly larger and assumes structured cyclic behavior of apparent limit-cycle nature.

A distinguishing feature of whether these discrete time series are linear or nonlinear in nature is asymmetry in time, particularly regarding trends in rise and fall times. To facilitate visual inspection, the segments in **Figure 15** are rescaled to have unit variance and plotted in **Figure 16**. In the segments for equivalence ratios from 1.004 to 0.700, deviations from point to point do not follow clear trends of bias toward longer rise versus fall times, or vice versa. In the segments for equivalence ratios from 0.647 to 0.540, a bias of shorter rise times (three-to-four successive points upward) and longer fall times (four-to-five points downward) becomes apparent. Because of the overall variability of combustion patterns at these fuel-lean equivalences, this bias is not wholly systematic. It is therefore desirable to quantify the overall trends of rise and fall times to characterize the nature of the cycle-magnitude variations of the range of equivalence ratios.

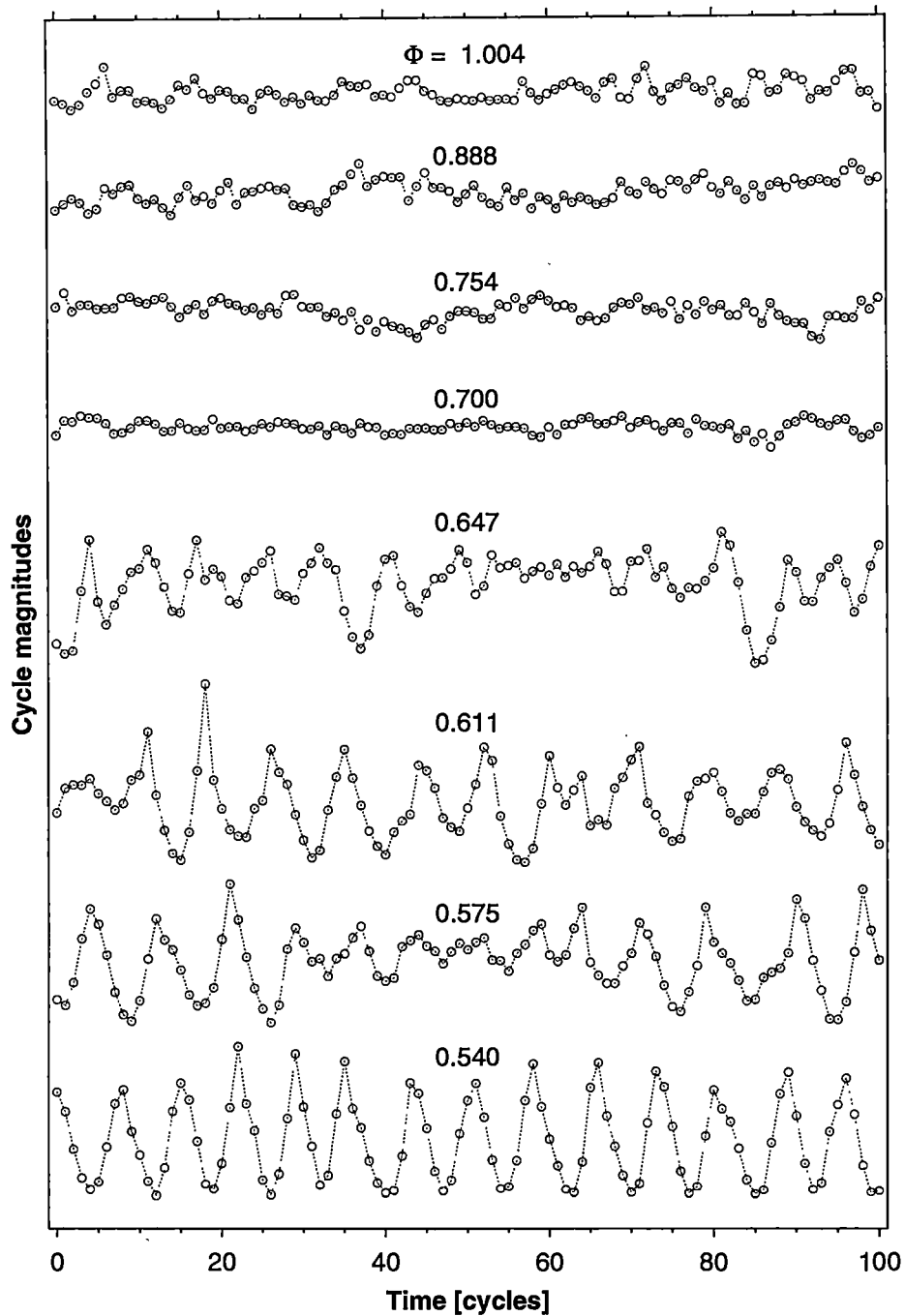


Figure 15: Segments of peak-to-trough cycle-magnitude time series from the thermal pulse combustor at a residence time of 50 msec. Segments are displayed in decreasing equivalence ratio, from nearly stoichiometric ($\Phi = 1.004$) to very fuel-lean, approaching flameout ($\Phi = 0.540$). Segments are plotted on the same scale.

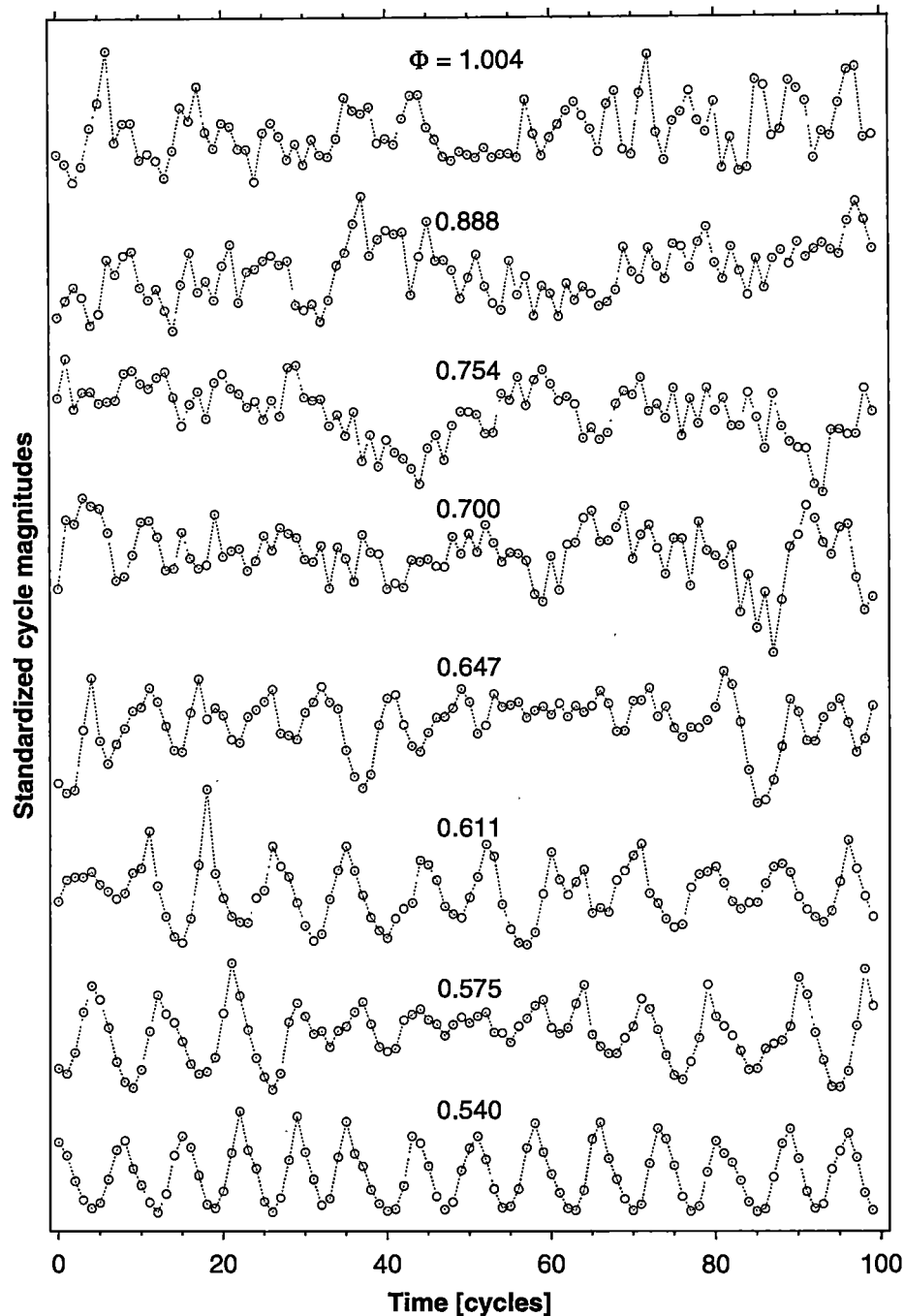


Figure 16: Segments of peak-to-trough cycle-magnitude time series from the thermal pulse combustor at a residence time of 50 msec. Segments are displayed in decreasing equivalence ratio, from nearly stoichiometric ($\Phi = 1.004$) to very fuel-lean, approaching flameout ($\Phi = 0.540$). Segments are standardized to unit variance to highlight cycle-to-cycle rise and fall behavior.

4.3.2 Results of the symbolic difference statistic test

The difference symbolization statistic S_{Δ} presented in Section 3.3 is designed to tally the overall difference of data rises and falls. Under the null hypothesis, $S_{\Delta} = 0$ for temporally reversible time series. To quantify the variance in S_{Δ} and establish a confidence interval on the test, the method of surrogate data is employed, and confidence is determined based on Monte Carlo probabilities. For the given statistic and null, a proper surrogate transformation is to shuffle the original time series before dynamically symbolizing. For reversible data, the shuffle transformation should have no effect on S_{Δ} , within the finite-sample error bounds. For irreversible data, the transformation should remove any temporal asymmetry, and S_{Δ} should tend to zero, within the finite-sample error bounds.

Figure 17 displays the difference symbolization statistic S_{Δ} for peak-to-trough cycle magnitudes over a range of equivalence ratios. In the figure, the filled points on the plot correspond to the time series depicted in Figure 16. The broken lines correspond to 95 percent confidence bounds based on Monte Carlo probabilities obtained from 200 surrogate trials.

From equivalence ratios 1.07 leaning to 0.62, the cycle-magnitude oscillations do not exhibit significant temporal irreversibility. In this range, combustion variations appear to be Gaussian in nature and result from random fluctuations such as chamber and tailpipe turbulence. From equivalence ratios 0.61 to 0.54, the cycle-magnitude oscillations exhibit statistically significant temporal irreversibility, implying that they do indeed form a limit cycle, which is inherently nonlinear. In this range, the nonlinear ignition-extinction process of the in-chamber fuel inventory dominates the combustion

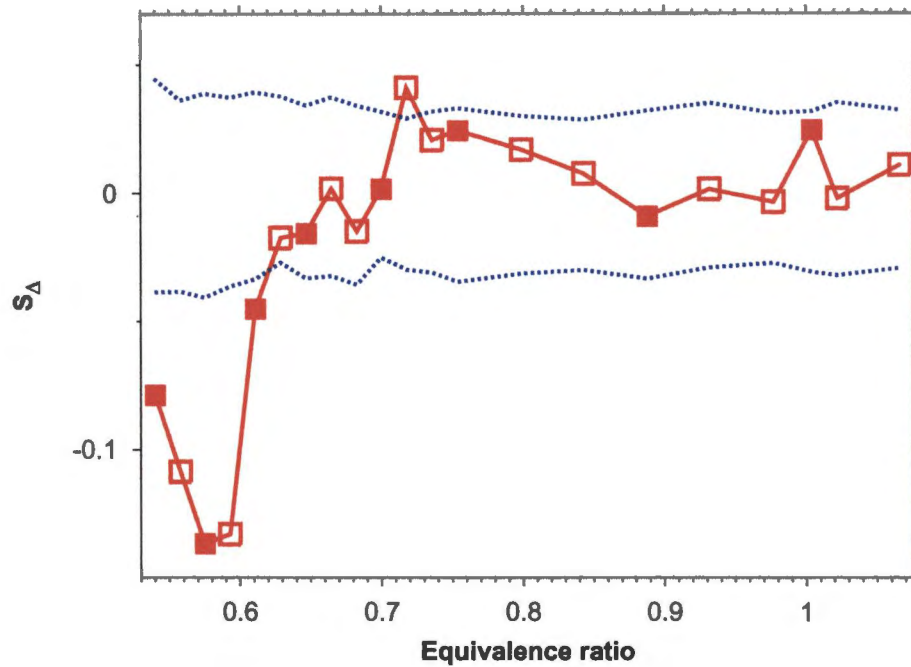


Figure 17: Behavior of the difference statistic S_{Δ} for peak-to-trough cycle magnitudes for the pulse-combustor data over a range of equivalence ratios. The filled points on the plot correspond to the time series depicted in **Figure 16**. The broken lines correspond to 95 percent confidence bounds based on Monte Carlo probabilities obtained from 200 surrogate trials.

variations. Verification that this combustion-quality modulation is nonlinear has important implications for control, because it defines a range in which nonlinear (chaos) control might be feasible and validates an assumption of a physical model for the process. Indeed, in this fuel-lean operating régime, a chaos-based feedback control of pulse-combustor dynamics has been implemented (Edwards, Finney, Nguyen, and Daw 2000; Edwards 2000a), with dramatic improvement in emissions quality.

It is important to note that the difference statistic employed here measures the summed bias of rises versus falls over the entire time series. Physically, it is not expected that the transition from turbulence-based to inventory-based combustion variations should be as sharp as implied in the preceding text. Rather, there is expected to be a gradual increase in the influence of the inventory effect as equivalence ratio is decreased. Thus, the statistic is not altogether sensitive to marginally irreversible time series such as is seen at equivalence ratio of 0.647. In this case, as is seen in the time-series plot in **Figure 16**, there are episodes of apparently irreversible behavior (e.g., cycles 0–50 and 65–100) and reversible behavior (e.g., cycles 50–65). Overall, within the confidence bands, the time series is temporally reversible, but clearly the effects of the nonlinear oscillations are beginning to become visible. Additionally, the anomalous point in **Figure 17** at $\Phi = 0.72$ is not explained physically, specifically whether the algorithm indiscriminantly highlights dynamics not of interest or whether there indeed is significant temporal irreversibility at this condition. In this case, to quantify irreversibility, it might be more advantageous to quantify the irreversibility based on a static symbolic transformation using the T_{irr} statistic developed in **Section 3.3** or using a sensitive statistic such as Targeted False Flipped Symbols (Daw, Finney, and Kennel 2000). However, for the purposes of this study,

the difference symbolization statistic adequately registers the temporally irreversible behavior which was conjectured to exist based on a physical model for the process.

4.4 Synchronization

This section presents application of the synchronization test presented in **Section 3.4** to two series of measurement data from internal combustion engines.

4.4.1 Data

Two groups of measurement data from spark-ignition internal combustion engines are examined for evidence of intercylinder synchronization. The data were supplied by J.G. GREEN, JR. of the Oak Ridge National Laboratory (Green 2000). The different data groups are labeled *Expedition* and *Quad-4*.

4.4.1.1 Experimental apparatus and procedure

The Expedition data were obtained at the Science Research Laboratory of Ford Motor Company using a 1999-model production Expedition V-8 engine manufactured by Ford Motor Company. The engine had 4.6 liter total displacement with production port fuel injection, a two-valve head, a bore of 9.02 cm, a stroke of 9.0 cm, and a compression ratio of 9. The cylinder firing order, used to label the cylinders in the succeeding text, was 1-3-7-2-6-5-4-8; firing order is labeled from 0 to 7. The nominal operating conditions were 1200 revolutions per minute and 20 degrees before-top-center spark timing. The fuel-air equivalence ratio was controlled by adjusting the fuel-injection pulse width and the throttle position while maintaining a 25 N·m load

with an absorbing/motoring dynamometer.

The Quad-4 data were obtained at the Advanced Propulsion Technology Center of the Oak Ridge National Laboratory from a 1995-model production Quad-4 I-4 engine manufactured by General Motors Corporation. The engine had 2.3 liter total displacement with port fuel injection, a bore of 9.2 cm, a stroke of 8.50 cm, and a compression ratio of 9.5. The cylinder firing order, used to label the cylinders in the succeeding text, was 1-3-4-2; firing order is labeled from 0 to 3. The nominal operating conditions were 1200 revolutions per minute and 23 degrees before-top-center spark timing. The fuel-air equivalence ratio was controlled by adjusting the fuel injection pulse width with an adjustable load on an absorbing dynamometer.

For both engines, continuous cylinder-pressure time series were sampled from head-mounted transducers in each cylinder. The pressure time series were sampled once per crank angle degree so that there were 720 pressure measurements per engine cycle. Estimates of combustion heat release were computed from the pressure time series using an integral method similar to that of Rassweiler and Withrow (Heywood 1988), so that the resulting time series were discrete, with one heat-release estimate per engine cycle. Because of data-acquisition system limitations, data segments from the Expedition engine were limited to approximately 354 contiguous engine cycles; the Quad-4 engine time series had between 3000 and 10000 consecutive engine cycles.

4.4.1.2 Dynamical behavior of the engine data

For the types of measurements in the engines described above, the dynamical behavior of combustion quality varies dramatically with fueling changes. Specifically,

cycle-to-cycle combustion variations change from Gaussian in nature at stoichiometric fueling to a noisy period-2 behavior at lean fueling (Daw, Finney, Green, Kennel, Thomas, and Connolly 1996; Finney, Green, and Daw 1998; Wagner, Drallmeier, and Daw 1998; Daw, Kennel, Finney, and Connolly 1998; Green, Daw, Armfield, Finney, Wagner, Drallmeier, Kennel, and Durbetaki 1999; Scholl and Russ 1999). In the lean-fueling régime, the period-2 behavior is the result of anticorrelated (alternating higher and lower) variations in combustion quality. As explained in the model of Daw *et al.*, the nonlinear nature of these oscillations is primarily the result of the sensitive dependence of combustion quality on equivalence ratio, which is a function of perturbations to the amount of fuel injected and in the amount of unburned fuel in the residual gas. The severity of the cyclic variability at lean fueling conditions limits the operating range of spark-ignition engines, so understanding the causes of this variability, particularly whether there is interplay among cylinders, is important for design and control considerations. Specifically, random phase relationships result in smoother operation, as strong combustion events would be distributed more evenly along the crankshaft cycle, whereas synchronous operation would result in power surges and rougher operation. As a result, synchronization among cylinders enhances the coefficient of variability (Daw, Green, Wagner, Finney, and Connolly 2000), so disrupting synchronization would facilitate lean-mode operation.

The nature of the heat-release oscillations in the lean-fueling régime may be seen in **Figures 18** and **19**. In **Figure 18**, segments from all eight cylinders from the first 30 engine cycles at a nominal equivalence ratio of 0.59 are shown. These segments are typical of the behavior of each cylinder for the entire 354 engine cycles which were recorded. A common feature to all cylinders is the “sawtooth”-shaped

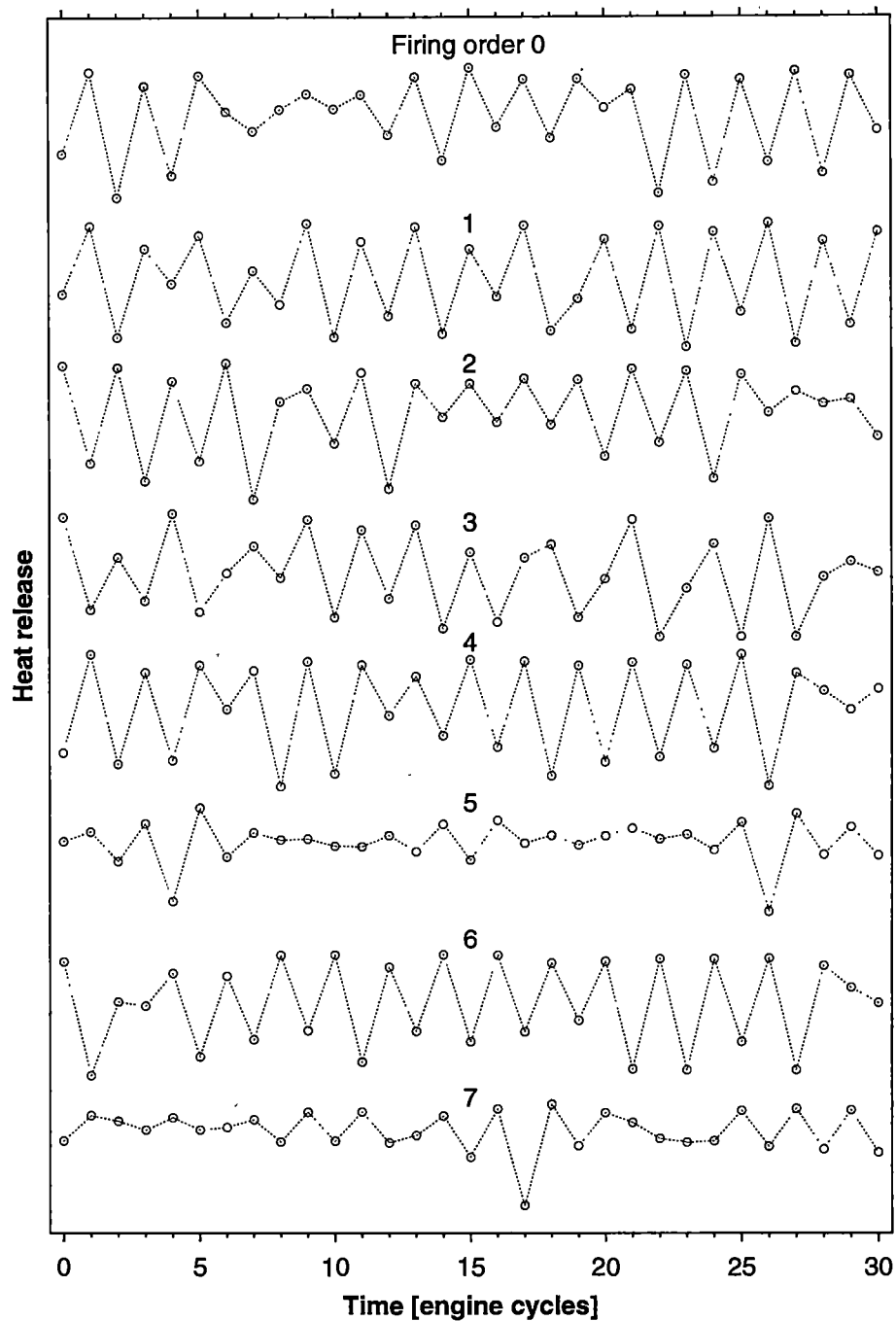


Figure 18: Short time-series segments from the Expedition engine at an equivalence ratio of 0.59. Segments are listed in firing order from top to bottom and are plotted on the same scale.

anticorrelated oscillations from cycle to cycle. These oscillations represent alternating cycles of stronger and weaker combustion quality. Although the anticorrelation is fairly consistent, the amplitudes of combustion heat release vary noticeably from cycle to cycle, and the rhythm of the anticorrelation is sporadically interrupted (e.g., in cylinder 0 during cycles 6–9). This behavior is characteristic of “noisy period-2” oscillation. Also noteworthy is that although all cylinders are nominally at the same fueling level and presumably at similar stoichiometries, the severity of oscillation differs from cylinder to cylinder. For instance, cylinders 5 and 7 are less variable than the other cylinders, probably caused by fueling effects such as fuel-injector malfunction or fuel or air intake difference caused by manifold acoustical oscillations. In nonlinear-dynamics terminology, the magnitude and character of the variability is largely determined by the degree of bifurcation that each cylinder is in; cylinders 5 and 7 are less bifurcated than the others.

In **Figure 19**, segments from all four cylinders of the Quad-4 engine from the first 30 engine cycles at a nominal equivalence ratio of 0.536 are shown. These segments are typical of the behavior of each cylinder for the entire 3000 engine cycles which were recorded. As with the Expedition engine, all cylinders exhibit behavior characteristic of noisy period-2 oscillations. A brief episode of correlation and anticorrelation between cylinders is seen in in cycles 22–29. In this episode, cylinders 1 and 3 are correlated in their behavior: at cycles 22, 24, 26 and 28, they both exhibit relatively strong combustion, whereas at cycles 23, 25, 27 and 29, they both exhibit relatively weak combustion. Conversely, cylinder 2 is exactly out of phase with cylinders 1 and 3, meaning they are anticorrelated. Of interest is whether this episode of correlation and anticorrelation is the statistically normal behavior or is a chance occurrence. The

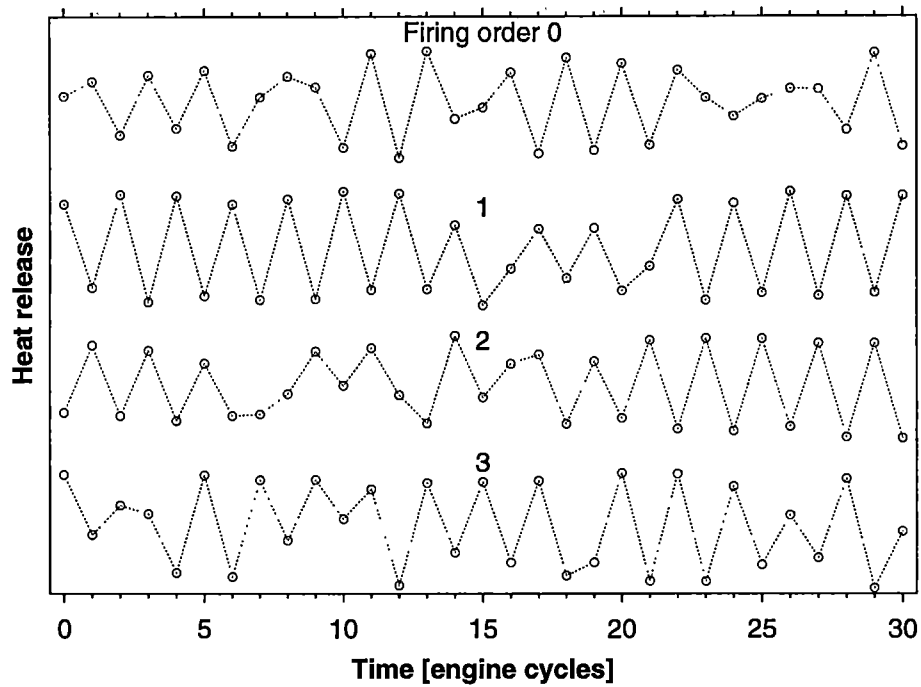


Figure 19: Short time-series segments from the Quad-4 engine at equivalence ratio of 0.536. Segments are listed in firing order from top to bottom and are plotted on the same scale.

symbolic synchrogram, with the Monte Carlo probability test described in **Section 3.4**, quantifies intercylinder dynamical relationships.

4.4.2 Two-cylinder synchrograms

Two-cylinder synchrograms are constructed to verify that the patterns of simple correlation and anticorrelation between cylinders as described above do exist and have a physical basis. Because each cylinder's behavior naturally oscillates in a noisy period-2 mode, a finite degree of correlation would be expected to be seen by chance, even if the cylinders were not physically coupled. The goal in constructing the synchrograms is to test the null hypothesis that there is no significant synchronization between cylinders. Because the test proposed in **Section 3.4** relies on the method of surrogate data to estimate significance, the appropriate surrogate transformation is to randomize the phase relationships of the two time series used to construct the synchrograms. Algorithmically, the phase randomization is achieved by selecting the starting locations of each symbol series at random, so that each surrogate synchrogram should have a different phase relationship than the original data. By constructing symbol synchrograms with $K_n = \{2\ 2\ 2\ 2\ 2\ 2\}$, $K_i = \{0\ 1\ 0\ 1\ 0\}$ and $K_s = \{1\ 2\ 1\ 2\ 1\ 2\}$, correlated sequences are 0 0 1 1 0 0 (code 12) and 1 1 0 0 1 1 (code 51), and anticorrelated sequences are 0 1 1 0 0 1 (code 25) and 1 0 0 1 1 0 (code 38). In this manner, coarse-grained patterns of three-cycle behavior are cataloged.

Symbol synchrograms for the Expedition engine are shown in **Figure 20**. The top synchrogram (a) shows the relationships between cylinders 0 and 3, judged to be relatively uncorrelated. Although sequences 12, 25, 38 and 51 appear to have slightly elevated frequencies of occurrence, no one sequence appears to dominate. In

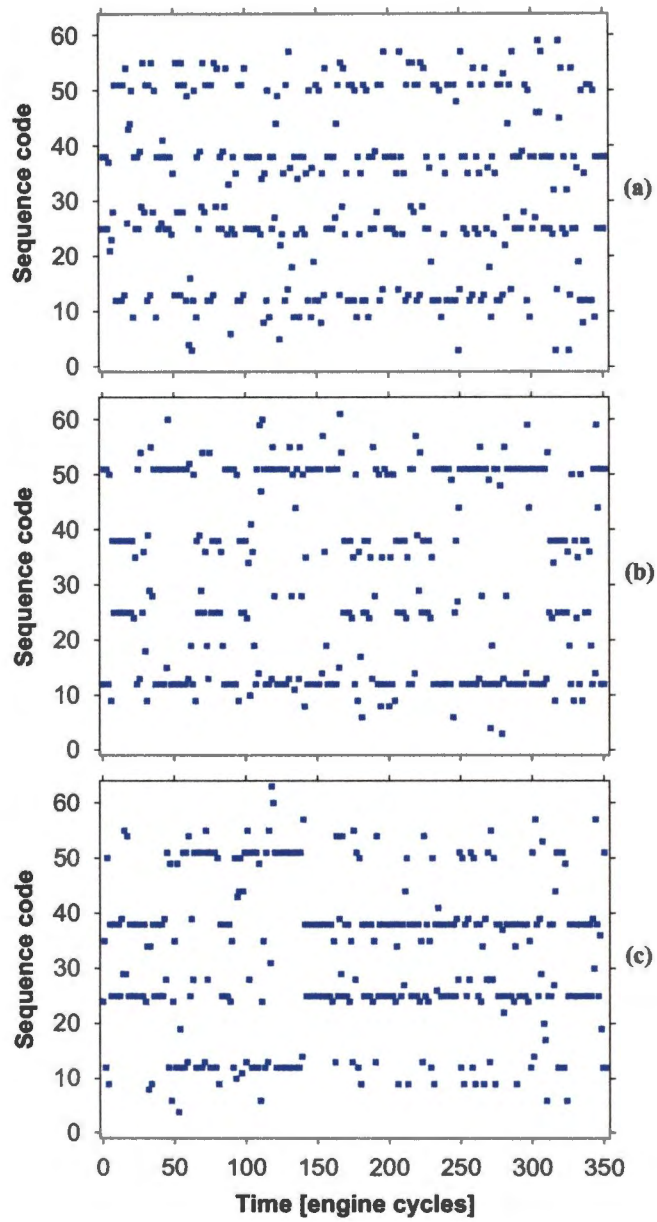


Figure 20: Symbol synchrograms for the Expedition data at an equivalence ratio of 0.59. Cylinders 0 and 3 (labeled in firing order) are relatively uncorrelated (a); 0 and 5 are correlated (b); 0 and 1 are anticorrelated (c).

contrast, the middle synchrogram (b) shows the relationships between cylinders 0 and 5, judged to be correlated. Here, sequences 12 and 51 predominate, resulting in "banding" (dark stripes) of the synchrogram. Similarly, the bottom synchrogram (c) shows the relationships between cylinders 0 and 1, judged to be anticorrelated. Here, sequences 25 and 38 predominate (except between cycles 45–140), also resulting in "banding".

The patterns seen in the synchrograms occur because of two causes: physical coupling and/or chance. By chance, the period-2 oscillations would result in episodes of apparent synchronization. As an illustration of the effect of removing the influences of the period-2 behavior, **Figure 21** shows a synchrogram between cylinders 0 and 1 when order of the symbol series was shuffled before the synchrogram was constructed. No longer are sequences 12, 25, 38, and 51 remarkable or predominant, as any significant synchronization was destroyed during the shuffle. Thus, the natural period-2 behavior plays an important rôle in appearance of correlation patterns in sequences 12, 25, 38, and 51. The crux of the problem then is to quantify how significant the occurrence of the correlation patterns are given that each time series is inherently anticorrelated within itself.

For quantification of synchronization in the synchrograms, two statistics are used. First, because the synchrogram is simply a multivariate code series, the modified Shannon entropy is an obvious measure of the degree of organization of its symbol-sequence histogram. For instance, very strongly correlated behavior would result in a finite number of sequence codes predominating the synchrogram, so entropy would be expected to be low; weakly correlated or uncorrelated behavior would result in higher values of entropy. However, the Shannon entropy is nonspecific in which

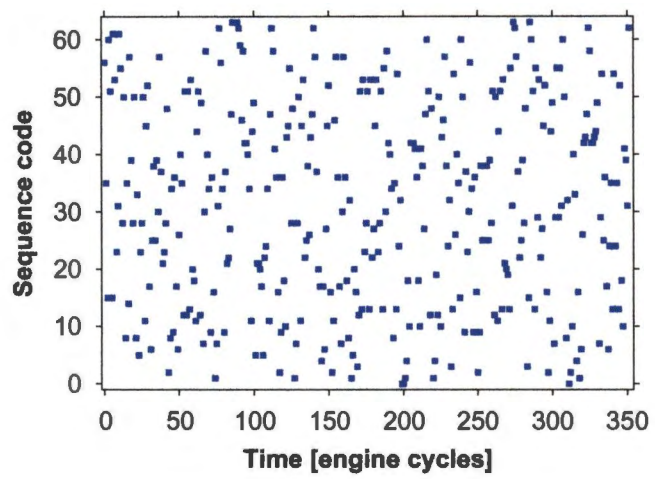


Figure 21: Randomization of the symbol series for cylinders 0 and 1 (firing order) destroys patterns of anticorrelation in the synchrogram.

patterns it measures — it simply quantifies the degree of organization of the symbol statistics and cannot discriminate sequences 12, 25, 38, and, 51 from, say, 0, 17, 43, 62. Therefore, a second type of statistic is employed. S_+ represents the sum of frequencies of occurrence for purely correlated sequences (12 and 51, and also the very rare instances of 0, 21, 42, and 63), and S_- represents the sum of frequencies of occurrence for purely anticorrelated sequences. By using the targeted statistics S , only the sequences of interest are counted, thus allowing a much more powerful statistical test.

When significant synchronization occurs, it is expected that the modified Shannon entropy for the original synchrogram should be less than entropy values for the phase-randomized surrogate synchrograms; *i.e.*, the phase randomization destroyed real synchronization, and the resulting symbol statistics are more complicated. When the synchronization is in the form of strong correlation, it is expected that the statistic S_+ should be higher than those for the phase-randomized surrogates. Similarly, when the synchronization is in the form of strong anticorrelation, it is expected that the statistic S_- should be higher than those for the phase-randomized surrogates.

Figure 22 shows the entropy values for the shuffled surrogates for the synchrograms depicted in **Figure 20**. In each plot, the entropy value for the original synchrogram is marked with a horizontal line from the ordinate intersecting the probability function of the surrogates and projected to the abscissa.

The top plot (**a**) shows results for cylinders 0 and 3, which were judged to be uncorrelated based on visual inspection of their synchrogram. The entropy value of the original synchrogram is 0.813, corresponding to Monte Carlo probability of approximately 0.36, based on 200 surrogate trials. Thus, 64 percent of the surrogate trials

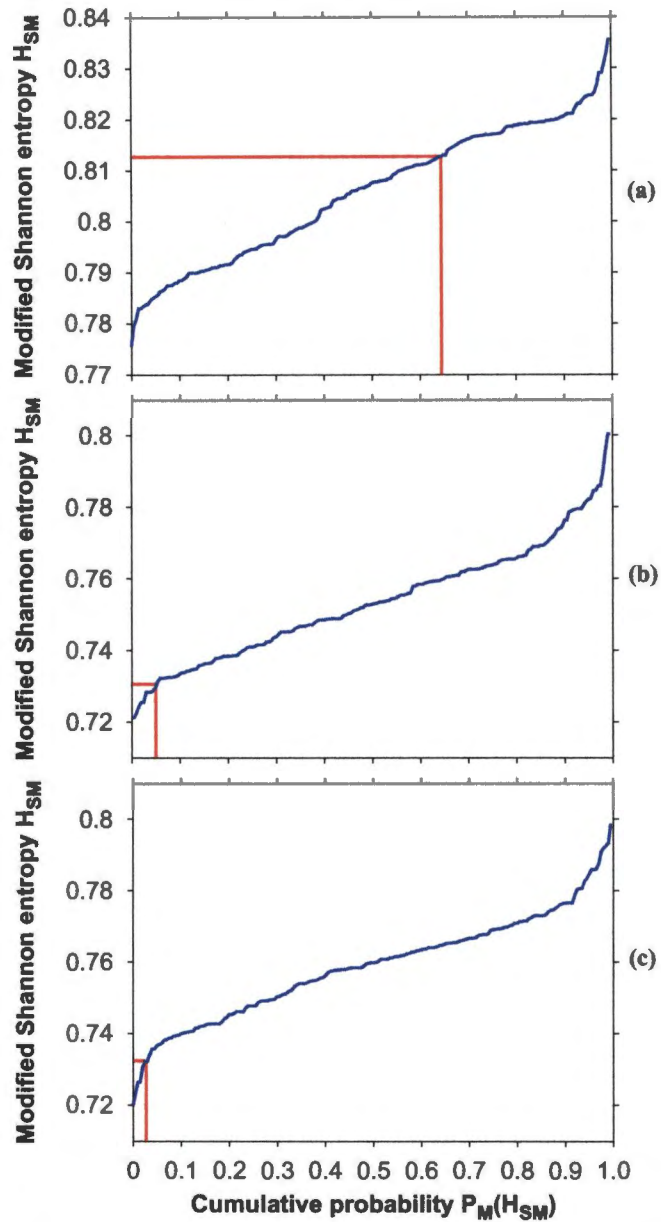


Figure 22: Modified Shannon entropy of the symbol synchrograms depicted in **Figure 20**. The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data. Cylinder groupings (labeled by firing order) are: 0 and 3 (a), 0 and 5 (b), 0 and 1 (c).

exhibited lower Shannon entropy than the original synchrogram, which is contrary to what would be expected had there been significant organization of the symbol statistics in the original synchrogram. For a generous significance level of $\alpha = 0.1$, the p -value of approximately 0.64 is much greater, meaning there is no evidence to reject the null of no synchronization.

The middle plot (b) shows results for cylinders 0 and 5, which were judged to be correlated based on visual inspection of their synchrogram. The entropy value of the original synchrogram is 0.731, corresponding to Monte Carlo probability of approximately 0.95, based on 200 surrogate trials. Thus, only 5 percent of the surrogate trials exhibited lower Shannon entropy than the original synchrogram. The p -value is approximately 0.05, meaning there is sufficient evidence to reject the null of no synchronization, strongly with $\alpha = 0.1$ and marginally with $\alpha = 0.05$.

The bottom plot (c) shows results for cylinders 0 and 1, which were judged to be anticorrelated based on visual inspection of their synchrogram. The entropy value of the original synchrogram is 0.732, corresponding to Monte Carlo probability of approximately 0.975, based on 200 surrogate trials. Thus, only 2.5 percent of the surrogate trials exhibited lower Shannon entropy than the original synchrogram. The p -value is approximately 0.025, meaning there is sufficient evidence to reject the null of no synchronization, both with reasonable significance levels of $\alpha = 0.1$ and $\alpha = 0.05$.

Figure 23 shows the frequency-summation statistic S values for the shuffled surrogates for the synchrograms depicted in **Figure 20**. In each plot, the S value for the original synchrogram is marked with a horizontal line from the ordinate intersecting the probability function of the surrogates and projected to the abscissa.

The top plot (a) shows results for cylinders 0 and 3, which were judged to be

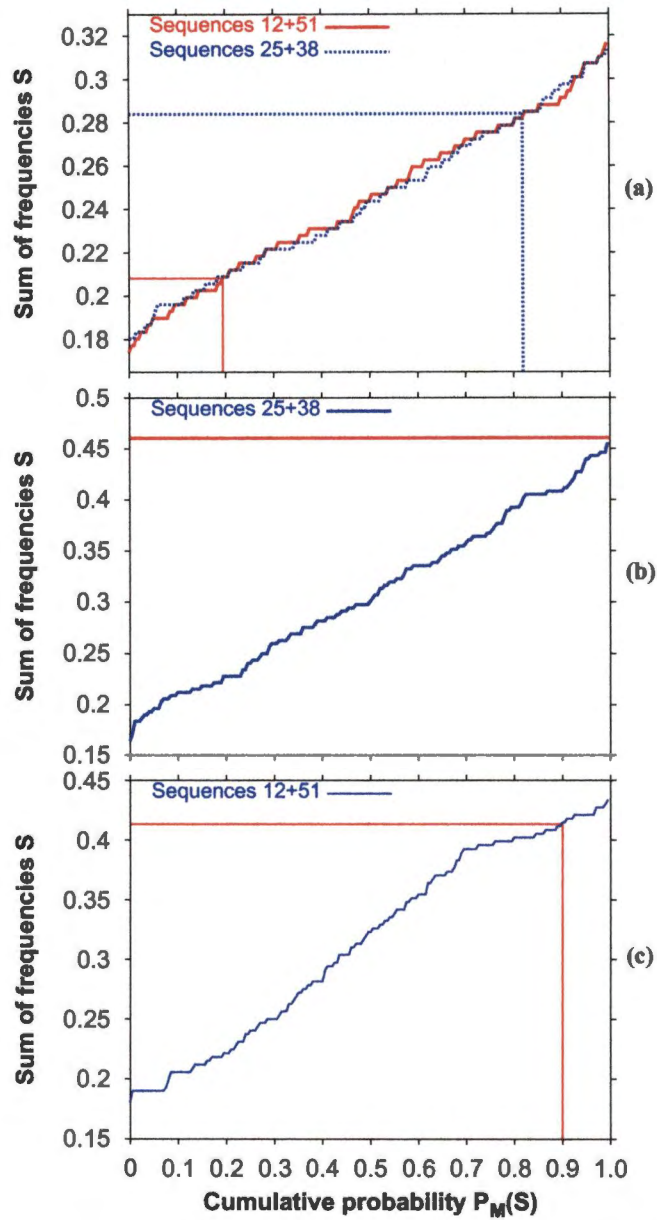


Figure 23: Correlation statistic S of the symbol synchrograms depicted in **Figure 20**. The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data. Cylinder groupings (labeled by firing order) are: 0 and 3 (a), 0 and 5 (b), 0 and 1 (c).

uncorrelated based on visual inspection of their synchrogram. In this case, because no correlation implies neither correlation- nor anticorrelation-dominant synchronization, both S_+ and S_- are examined. For the original synchrogram, $S_+ = 0.208$ ($p = 0.796$) and $S_- = 0.284$ ($p = 0.184$), based on 200 surrogate trials. Neither S_+ nor S_- , tested independently, offers significant basis for rejecting the null of no synchronization, a conclusion which confirms the test based on a metric of modified Shannon entropy.

The middle plot (b) shows results for cylinders 0 and 5, which were judged to be correlated based on visual inspection of their synchrogram. In this case, only S_+ is examined. For the original synchrogram, $S_+ = 0.461$ ($p \leq 0.005$), based on 200 surrogate trials. Such a low p -value suggests that the null of no synchronization may be rejected confidently, a conclusion which confirms the test based on a metric of modified Shannon entropy.

The bottom plot (c) shows results for cylinders 0 and 1, which were judged to be anticorrelated based on visual inspection of their synchrogram. In this case, only S_- is examined. For the original synchrogram, $S_- = 0.413$ ($p = 0.1$), based on 200 surrogate trials. The resulting p -value would offer only marginal rejection of the null at a high ($\alpha = 0.1$) significance level, suggesting that the null of no synchronization should not be rejected. This conclusion conflicts with that based on a metric of modified Shannon entropy. One reason for this inconsistency is the limited number of data (354 records) in the time series — for analysis of engine data, 1000 cycles is generally considered acceptable. Another, less probable reason for the inconsistency is that entropy is a more generic metric and S_- is a targeted metric, implying that entropy detects organization in the symbol statistics not related to the type of synchronization patterns of interest. For resolution of the question of limited data-set

size, the Quad-4 data, with 3000 records per time series, were expected to be more conclusive and consistent.

Symbol synchronograms for the Quad-4 engine are shown in **Figure 24**. The top synchronogram (a) shows the relationships between cylinders 0 and 2, judged to be relatively uncorrelated. The middle synchronogram (b) shows the relationships between cylinders 0 and 3, judged to be correlated. The bottom synchronogram (c) shows the relationships between cylinders 2 and 1, judged to be anticorrelated. Visual inspection of the synchronograms is not as conclusive as with those for the Expedition, partly because of the plotting scale (3000 instead of 350 cycles) and partly because the dynamical relationships between the cylinders appear more complex (the strength of the period-2 bifurcation is less), leading to more occurrences of sequences other than the targeted sequences described above (12, 25, 38, and 51).

Figure 25 shows the entropy values for the shuffled surrogates for the synchronograms depicted in **Figure 24**. In each plot, the entropy value for the original synchronogram is marked with a horizontal line from the ordinate intersecting the probability function of the surrogates and projected to the abscissa.

The top plot (a) shows results for cylinders 0 and 2, which were speculated to be uncorrelated based on visual inspection of their synchronogram. The entropy value of the original synchronogram is 0.74, corresponding to Monte Carlo probability of approximately 0.955, based on 200 surrogate trials. Thus, 4.5 percent of the surrogate trials exhibited lower Shannon entropy than the original synchronogram, which is a sufficiently low p -value to suggest rejecting the null of no synchronization. Visual inspection of the synchronogram (**Figure 24**) reveals that other activity other than synchronization of period-2 behavior might account for the low entropy value. For

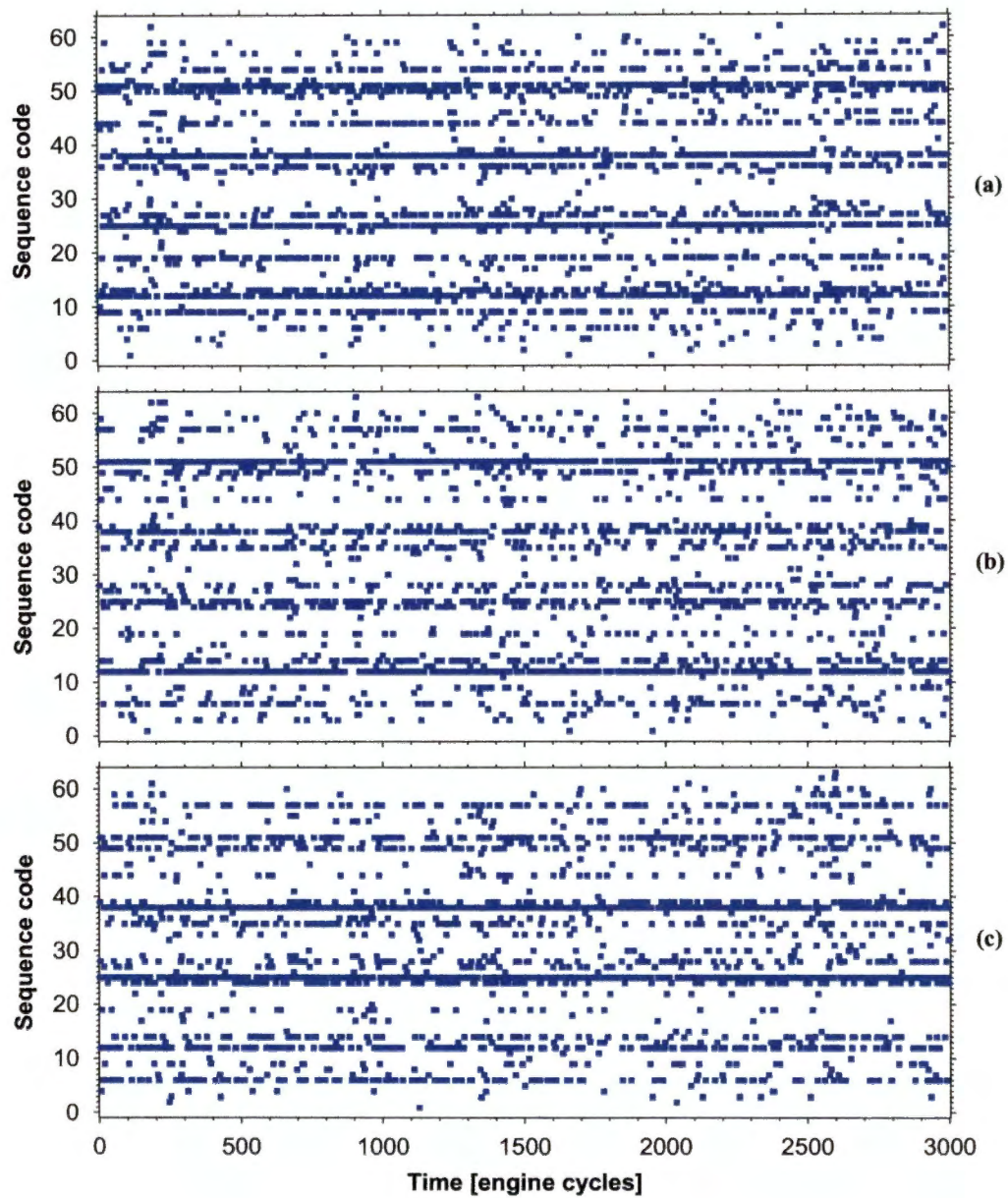


Figure 24: Symbol synchrograms for the Quad-4 data at an equivalence ratio of 0.536. Cylinders 0 and 2 (labeled by firing order) are relatively uncorrelated (a); 0 and 3 are correlated (b); 2 and 3 are anticorrelated (c).

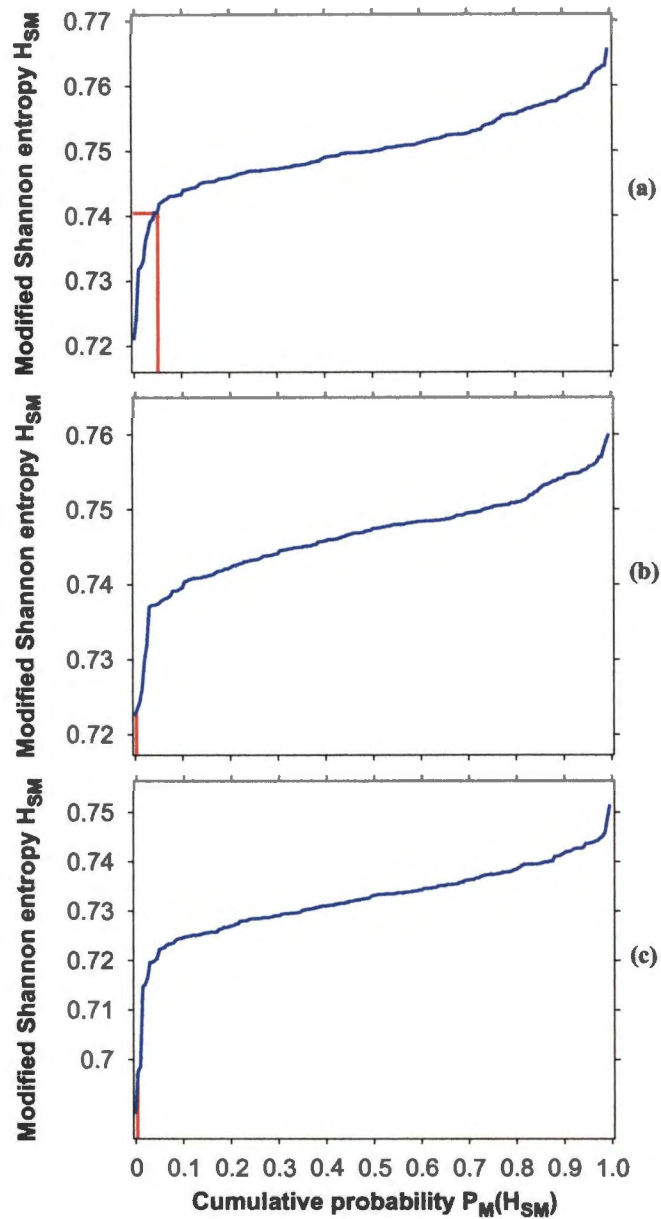


Figure 25: Modified Shannon entropy of the symbol synchrograms depicted in **Figure 24**. The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data. Cylinder groupings (labeled by firing order) are 0 and 2 (a), 0 and 3 (b), and 2 and 3 (c).

instance, sequence 9 (0 0 1 0 0 1) is prevalent, which could be viewed as a transition sequence from correlated (0 0) to anticorrelated (1 0 0 1) behavior. The increased prevalence of the transition sequences suggests that the cylinders were phasing in and out of different degrees of synchronization, with neither correlation nor anticorrelation favored — in the sum, the cylinders were not intrinsically synchronized, a conclusion not directly obtainable from the modified Shannon entropy of their synchrogram.

The middle plot (b) shows results for cylinders 0 and 3, which were judged to be correlated based on visual inspection of their synchrogram. The entropy value of the original synchrogram is 0.723, corresponding to Monte Carlo probability of approximately 0.995, based on 200 surrogate trials. Thus, only 0.5 percent of the surrogate trials exhibited lower Shannon entropy than the original synchrogram. The p -value is approximately 0.005, meaning there is strong evidence to reject the null of no synchronization.

The bottom plot (c) shows results for cylinders 2 and 3, which were judged to be anticorrelated based on visual inspection of their synchrogram. The entropy value of the original synchrogram is 0.697, corresponding to Monte Carlo probability of approximately 0.995, based on 200 surrogate trials. Thus, only 0.5 percent of the surrogate trials exhibited lower Shannon entropy than the original synchrogram. The p -value is approximately 0.005, meaning there is strong evidence to reject the null of no synchronization.

Figure 26 shows the frequency-summation statistic S values for the shuffled surrogates for the synchrograms depicted in **Figure 24**. In each plot, the S value for the original synchrogram is marked with a horizontal line from the ordinate intersecting the probability function of the surrogates and projected to the abscissa.

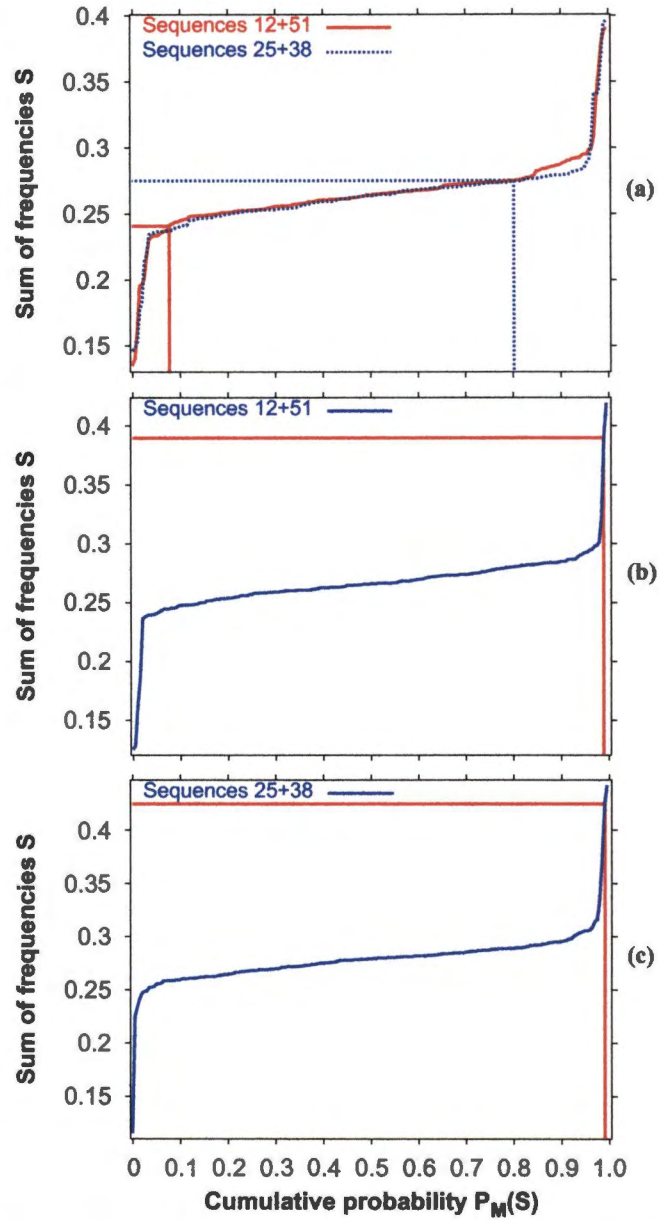


Figure 26: Correlation statistic S of the symbol synchrograms depicted in **Figure 24**. The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data. Cylinder groupings (labeled by firing order) are 0 and 2 (a), 0 and 3 (b), and 2 and 3 (c).

The top plot (a) shows results for cylinders 0 and 2, which were speculated to be uncorrelated based on visual inspection of their synchrogram. For the original synchrogram, $S_+ = 0.241$ ($p = 0.91$) and $S_- = 0.275$ ($p = 0.174$), based on 200 surrogate trials. Neither S_+ nor S_- , tested independently, offers significant basis for rejecting the null of no synchronization, a conclusion which contradicts the test based on a metric of modified Shannon entropy. As discussed previously, the reason for this discrepancy is that entropy, being a generic metric, is nonspecific and to some degree is ill-suited as a statistic for testing synchronization in symbol synchrograms. Conversely, the targeted summation statistics are specific to the specific question of whether certain types of correlated behavior (synchronization) occur.

The middle plot (b) shows results for cylinders 0 and 3, which were judged to be correlated based on visual inspection of their synchrogram. For the original synchrogram, $S_+ = 0.39$ ($p = 0.01$), based on 200 surrogate trials. Such a low p -value suggests that the null of no synchronization may be rejected confidently, a conclusion which confirms the test based on a metric of modified Shannon entropy.

The bottom plot (c) shows results for cylinders 2 and 3, which were judged to be anticorrelated based on visual inspection of their synchrogram. For the original synchrogram, $S_- = 0.425$ ($p = 0.005$), based on 200 surrogate trials. Such a low p -value suggests that the null of no synchronization may be rejected confidently, a conclusion which confirms the test based on a metric of modified Shannon entropy.

4.4.3 Three-cylinder synchrograms

Three-cylinder synchrograms are constructed to quantify the degree and type of synchronization among three engine cylinders. Because of the limited nature of the

Expedition data as described above, discussion here will focus on the Quad-4 data. In the noisy period-2 mode of present interest, a finite degree of correlation would be expected to be seen by chance, even if the cylinders were not physically coupled. The goal in constructing the synchrograms is to test the null hypothesis that there is no significant synchronization between cylinders. Because the test proposed in **Section 3.4** relies on the method of surrogate data to estimate significance, the appropriate surrogate transformation is to randomize the phase relationships of the three time series used to construct the synchrograms. Algorithmically, the phase randomization is achieved by selecting the starting locations of each symbol series at random, so that each surrogate synchrogram should have a different phase relationship than the original data. For three-cylinder synchrograms, symbolization parameters are $K_n = \{2\ 2\ 2\ 2\ 2\ 2\}$, $K_i = \{0\ 0\ 1\ 0\ 0\}$ and $K_s = \{1\ 2\ 3\ 1\ 2\ 3\}$, meaning that three time series are examined for temporal patterns over two consecutive engine cycles. It should be noted that more cylinders could be included over a longer span of consecutive engine cycles but here are not to facilitate visualization of the synchrograms, specifically to limit the number of possible sequences to 64.

With the symbolization scheme described immediately above, purely correlated sequences are 0 0 0 1 1 1 (code 7) and 1 1 1 0 0 0 (code 56). In these two instances, all three cylinders oscillate in phase. Other purely correlated sequences are 0 0 0 0 0 0 (code 0) and 1 1 1 1 1 1 (code 63), but these represent the rare events not of interest presently. With three cylinders, pure anticorrelation does not exist. Instead, mixed correlation and anticorrelation is of interest. These hybrid correlated sequences are 0 0 1 1 1 0 (code 14) and 1 1 0 0 0 1 (code 49) (cylinders *A* and *B* correlated, *C* anticorrelated); 0 1 0 1 0 1 (code 21) and 1 0 1 0 1 0 (code 42) (cylinders *A* and *C*

correlated, B anticorrelated); 0 1 1 1 0 0 (code 28) and 1 0 0 0 1 1 (code 35) (cylinders B and C correlated, A anticorrelated). In this manner, coarse-grained patterns of two-cycle behavior are cataloged.

A symbol synchrogram for the Quad-4 engine is shown in **Figure 27**. Because of the complicated nature of each cylinder's individual behavior (even in the noisy period-2 régime), this synchrogram visually is more complicated than the two-cylinder synchrograms. Of note is the "banding" at the hybrid sequences of 14, 21, 28, 35, 42, and 49. This hybrid would be expected given the observations that cylinders 1 and 3 are correlated and 1 and 2 are anticorrelated. The three-cylinder synchrogram, then, is a composite of the different two-cylinder synchrograms but also contains additional information. For instance, sequence 10 (0 0 1 0 1 0) is more prevalent early in observational record (cycles 1–2000) and gradually becomes less frequent. This sequence reveals that cylinders 2 and 3 continue an anticorrelated relationship whereas cylinder 1 somewhat misbehaves (two consecutive 0 cycles).

The quantification of how significant the synchronization patterns described above may be seen in the modified Shannon entropy and the frequency-summation statistics, as employed in the previous section. **Figure 28** shows the entropy values for the shuffled surrogates for the synchrogram depicted in **Figure 27**. The entropy value for the original synchrogram is marked with a horizontal line from the ordinate intersecting the probability function of the surrogates and projected to the abscissa. The entropy value of the original synchrogram is 0.769, corresponding to Monte Carlo probability of >0.995 , based on 200 surrogate trials. Thus, the p -value is 0.005, meaning there is strong evidence to reject the null of no synchronization.

Figure 29 shows the frequency-summation statistic S values for the shuffled

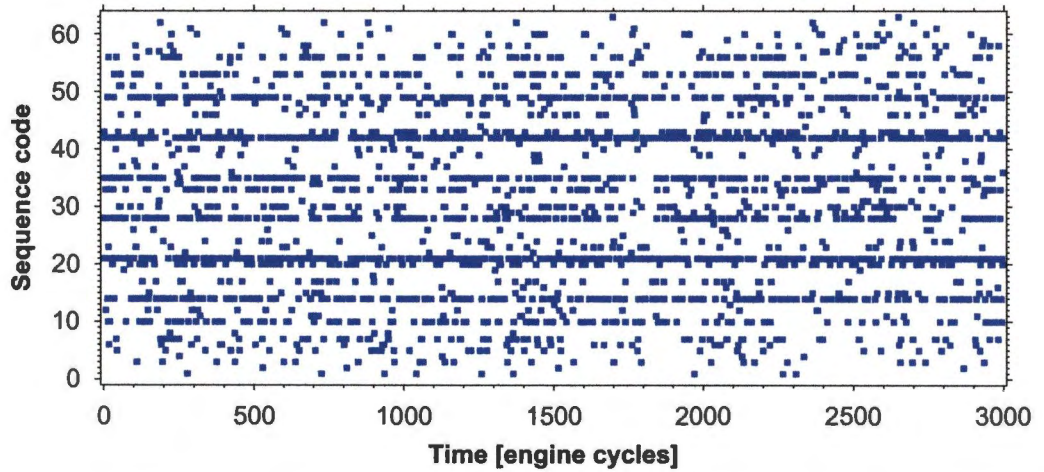


Figure 27: Symbol synchronogram depicting correlation patterns between cylinders 1, 2 and 3 (firing order) for the Quad-4 data at an equivalence ratio of 0.536.

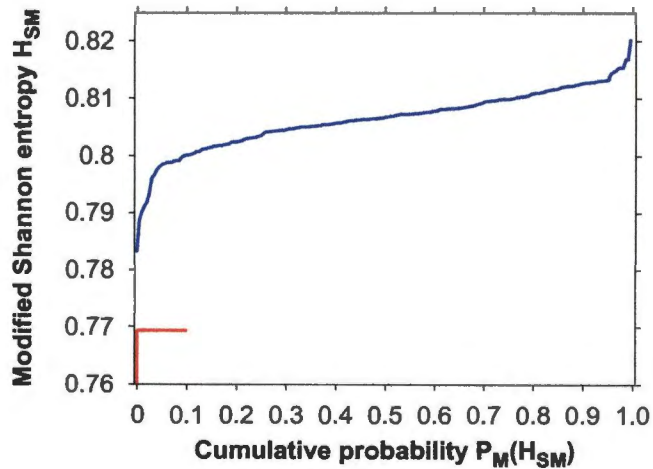


Figure 28: Modified Shannon entropy of the symbol synchronograms for cylinders 1, 2, and 3 (firing order) depicted in **Figure 27**. The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data.

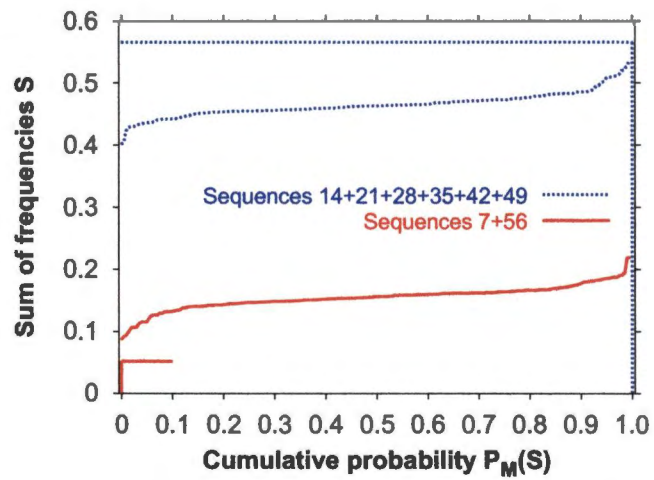


Figure 29: Correlation statistic S of the symbol synchrograms for cylinders 1, 2, and 3 (firing order) depicted in **Figure 27**. The curve is the cumulative distribution of the surrogates; the lines indicate the value for the original data.

surrogates for the synchrogram depicted in **Figure 27**. The S value for the original synchrogram is marked with a horizontal line from the ordinate intersecting the probability function of the surrogates and projected to the abscissa. Here, S_+ is the sum of the observed frequencies of sequences 7 and 56, representing pure correlation, and S_{\pm} is the sum of the observed frequencies of sequences 14, 21, 28, 35, 42 and 49, representing hybrid correlation. For the original synchrogram, $S_+ = 0.052$ ($p > 0.995$) and $S_{\pm} = 0.565$ ($p < 0.005$), based on 200 surrogate trials. Alone, S_{\pm} offers significant basis for rejecting the null of no synchronization, a conclusion which confirms the test based on a metric of modified Shannon entropy. In other words, the synchronization patterns are very unlikely to have occurred solely by chance.

The synchrogram is very telling in an additional, unexpected aspect — pure correlation among cylinders 1, 2, and 3 is avoided innately by the system. In all 200 surrogate trials, randomly shifting the phases of the three time series yielded higher frequencies of sequences 7 and 56 than were present in the original data. It is speculated that acoustical effects within the intake manifold are a primary mode of coupling the cylinders, and if there were pressure oscillations within the manifold, the air/fuel intake of any three cylinders would not be in phase because they would be at different locations in relation to the waveform nodes. More rigorous experimentation could help to identify the modes of systematic synchronization and whether manifold acoustics or some other phenomenon is most influential.

Chapter 5

Conclusions and recommendations

This work has presented methods of symbolization-based analysis to measurement time series from systems of engineering relevance. Tests have been developed to indicate stationarity, temporal irreversibility and synchronization with provisions for estimating statistical confidence. All the confidence measures presented in this study were obtained from a bootstrapping approach to Monte Carlo probability. In doing so, a unified, straightforward and uncomplicated methodology has been employed at the expense of the computational simplicity and without the distribution-theoretical basis of tests with *a priori* significance. However, the Monte Carlo approach should not be discounted — its power lies in the ability to gage the variability and range of a statistic when its distribution is unknown, and in cases in which the distribution is known, it provides a good approximation to the theoretical value. Because it is discrete in nature (limited by the number of surrogate trials), good agreement with any true distribution depends largely on the number of surrogate trials, the larger of which yields better agreement but which is more taxing computationally and which can behave badly for limited data (when some surrogate trials are too similar). Thus, the approach presented in this study is generic but should be viewed as a foundation for other, perhaps more sophisticated, methods.

The stationarity tests presented herein are applicable to a range of problems in

engineering systems. Besides the general question of ergodicity, the stationarity methods developed in this study are useful in systems of dynamical libraries, in which the objective is to match a test case against a library of reference cases. Metrics based on symbol statistics are natural choices to classify temporal data patterns, just as the Fourier spectrum is a natural choice to classify frequency content. A useful extension of the methodology presented in this work would be a multiple test using two or more difference metrics, which would allow a more robust determination of matching than the singular testing presently employed.

The temporal-irreversibility tests presented herein have utility in selecting a proper or consistent model for observed dynamics. For instance, with the internal combustion engine, temporal irreversibility has been shown to discriminate between two competing models, one reversible and the other irreversible. The overwhelming presence of irreversibility in the observed record discounts the reversible model. In the pulse combustor, the presence of temporal irreversibility in fuel-lean amplitude modulation suggests that the hypothesized fuel-inventory effect, known to be nonlinear and temporally irreversible, is a better model than the Helmholtz-resonator model as a source of the modulation.

The synchronization test presented herein provides important information regarding the strength and significance of correlated activity among different measurement signals. In the case of the internal combustion engines, synchronization becomes an important contributor to degraded engine performance at fuel-lean conditions, so understanding the phasic relationships between cylinders is important for design and control. However, because of the limited nature of the data analyzed in this study, no conclusions regarding a model for manifold acoustic oscillations can be drawn and

should be addressed in future studies with more targeted experimentation.

In this study, data symbolization and the proposed derivative tests are presented without comparison with other, non-symbolization-based tests, if available. In some respects, direct comparison is often not available (most synchrograms are based on only two signals, whereas the present work extends capability to more than two signals) or not meaningful. For testing generic concepts such as stationarity, tests are often catered to the definition at hand, so tests are chosen based on what is desired to be measured. In this sense, the present work extends the corpus of nonlinear time-series analysis methodology, with applicability being decided by the analyst. Future studies might include a rigorous comparison of the tests developed in this study against other available tests, but this is not really necessary or even desired except in the context of sensitivity to noise and data-set sample size. The general framework presented in this study provides a basis for such future study and for applicability to engineering time series.

References

References

- Angeli P. and Hewitt G.F. (1996). Phase-distribution measurements in liquid-liquid pipeline flows using an impedance probe. In *Proceedings of the ASME Heat Transfer Division*, Volume HTD 334-3, pp. 149–156. American Society of Mechanical Engineers, ISBN 0-7918-1521-8.
- Barnard G.A. (1963). Discussion on Professor Bartlett's paper. *Journal of the Royal Statistical Society, Series B* **25**, 294.
- Campbell M.J. and Walker A.M. (1977). Reply to discussion of papers by Campbell & Walker, Morris, and Tong. *Journal of the Royal Statistical Society, Series A* **140**, 462.
- Cox D.R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* **8**, 93–115.
- Crutchfield J.P. and Packard N.H. (1983). Symbolic dynamics of noisy chaos. *Physica D* **7**, 201–233.
- Cunningham L.B.C. and Hynd W.R.B. (1946). Discussion to 'Symposium on autocorrelation in time series'. *Supplement to the Journal of the Royal Statistical Society* **8**(1), 96–97.
- Daniels H.E. (1946). Discussion to 'Symposium on autocorrelation in time series'. *Supplement to the Journal of the Royal Statistical Society* **8**(1), 87.
- Daw C.S. (1999). Personal communication.
- Daw C.S. (2000). Personal communication.
- Daw C.S., Finney C.E.A., Green J.B., Kennel M.B., Thomas J.F., and Connolly F.T. (1996). A simple model for cyclic variations in a spark-ignition engine. Technical Paper 962086, Society of Automotive Engineers.
- Daw C.S., Finney C.E.A., and Kennel M.B. (2000). A symbolic approach for measuring temporal irreversibility. *Physical Review E* **62**.
- Daw C.S., Finney C.E.A., Nguyen K., and Halow J.S. (1998). Symbol statistics: a new tool for understanding multiphase flow phenomena. In *Proceedings of the ASME Heat Transfer Division*, Volume HTD 361-5. American Society of Mechanical Engineers, ISBN 0-7918-1597-8.
- Daw C.S., Green J.B., Wagner R.M., Finney C.E.A., and Connolly F.T. (2000). Synchronization of combustion variations in multi-cylinder spark-ignition engines. In *Proceedings of the 2000 Global Powertrain Congress*.

- Daw C.S., Kennel M.B., Finney C.E.A., and Connolly F.T. (1998). Observing and modeling nonlinear dynamics in an internal combustion engine. *Physical Review E* **57**(3), 2811–2819.
- Diks C., van Houwelingen J.C., Takens F., and DeGoede J. (1995). Reversibility as a criterion for discriminating time series. *Physics Letters A* **201**, 221–228.
- Diks C., van Zwet W.R., Takens F., and DeGoede J. (1996). Detecting differences between delay vector distributions. *Physical Review E* **53**(3), 2169–2176.
- Edwards K.D. (2000a). **Application of occasional feedback trajectory correction to enhance lean combustion quality in a pulsed combustor.** Ph. D. thesis, University of Tennessee.
- Edwards K.D. (2000b). Personal communication.
- Edwards K.D., Finney C.E.A., Nguyen K., and Daw C.S. (1998). Use of symbol statistics to characterize combustion in a pulse combustor operating near the fuel-lean limit. In *Proceedings of the 1998 Technical Meeting of the Central States Section of the Combustion Institute*, pp. 385–390.
- Edwards K.D., Finney C.E.A., Nguyen K., and Daw C.S. (2000). Application of nonlinear feedback control to enhance the performance of a pulsed combustor. In *Proceedings of the 2000 Technical Meeting of the Central States Section of the Combustion Institute*, pp. 249–254.
- Enochson L.D. and Otnes R.K. (1968). **Programming and Analysis for Digital Time Series Data.** The Shock and Vibration Information Center, United States Department of Defense.
- Finney C.E.A., Green J.B., and Daw C.S. (1998). Symbolic time-series analysis of engine combustion measurements. Technical Paper 980624, Society of Automotive Engineers.
- Finney C.E.A., Nguyen K., Daw C.S., and Halow J.S. (1998). Symbol-sequence statistics for monitoring fluidization. In *Proceedings of the ASME Heat Transfer Division*, Volume HTD 361-5, pp. 405–411. American Society of Mechanical Engineers, ISBN 0-7918-1597-8.
- Fisher R.A. (1935). **The Design of Experiments.** Hafner Publishing Company.
- Green J.B. (2000). Personal communication.
- Green J.B., Daw C.S., Armfield J.S., Finney C.E.A., Wagner R.M., Drallmeier J.A., Kennel M.B., and Durbetaki P. (1999). Time irreversibility and comparison of cyclic-variability models. Technical Paper 1999-01-0221, Society of Automotive Engineers.

- Halow J.S. and Daw C.S. (1994). Characterizing fluidized-bed behavior by decomposition of chaotic phase-space trajectories. *AIChE Symposium Series* **90**(301), 69–91.
- Heywood J.B. (1988). **Internal Combustion Engine Fundamentals**. McGraw-Hill (ISBN 0-07-028637-X).
- Hoekstra B.P.T., Diks C.G.H., Allesie M.A., and DeGoede J. (1997). Nonlinear analysis of the pharmacological conversion of sustained atrial fibrillation in conscious goats by the class Ic drug cibenzoline. *Chaos* **7**(3), 430–446.
- Hope A.C.A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society, Series B* **30**, 582–598.
- Huygens C. (1673). **The Pendulum Clock**. Translated by R.J. Blackwell. Iowa State University Press (1986), ISBN 0-8183-0933-9.
- Kennel M.B. (1997). Statistical test for dynamical nonstationarity in observed time-series data. *Physical Review E* **56**(1), 316–321.
- Kennel M.B. and Mees A.I. (2000). Testing for general dynamical stationarity with a symbolic data compression technique. *Physical Review E* **61**(3), 2563–2568.
- Kugiumtzis D. (1999). Test your surrogate data before you test for nonlinearity. *Physical Review E* **60**(3), 2808–2816.
- Kurths J., Schwarz U., Witt A., Krampe R.Th., and Abel M. (1996). Measures of complexity in signal analysis. In Katz R.A. (Ed.), *Chaotic, Fractal, and Nonlinear Signal Processing*, Volume 375 of *AIP Conference Proceedings*, pp. 33–54. American Institute of Physics: AIP Press.
- Kurths J., Voss A., Saparin P., Witt A., Kleiner H.J., and Wessel N. (1995). Quantitative analysis of heart rate variability. *Chaos* **5**(1), 88–94.
- Lawrance A.J. (1991). Directionality and reversibility in time series. *International Statistical Review* **59**(1), 67–79.
- Lehrman M., Rechester A.B., and White R.B. (1997). Symbolic analysis of chaotic signals and turbulent fluctuations. *Physical Review Letters* **78**(1), 54–57.
- Miller R.G. (1966). **Simultaneous Statistical Inference**. McGraw-Hill Book Company.
- Morse M. and Hedlund G.A. (1938). Symbolic dynamics. *American Journal of Mathematics* **60**, 815–866.
- Paluš M. (1996). Nonlinearity in normal human EEG: cycles, temporal asymmetry, nonstationarity and randomness, not chaos. *Biological Cybernetics* **75**(5), 389–396.

- Pomeau Y. (1982). Symétrie des fluctuations dans le renversement du temps. *Journal de Physique* **43**, 859–867.
- Ramsey J.B. and Rothman P. (1996). Time irreversibility and business cycle asymmetry. *Journal of Money, Credit, and Banking* **28**(1), 1–21.
- Rapp P.E., Albano A.M., Zimmerman I.D., and Jiménez-Montaña M.A. (1994). Phase-randomized surrogates can produce spurious identifications of non-random structure. *Physics Letters A* **192**, 27–33.
- Rayleigh J.W.S. (1896). **The Theory of Sound**. Dover Publications (1945), ISBN 0-486-60292-3 (vol. 1) and ISBN 0-486-60293-1 (vol. 2).
- Saparin P.I., Gowin W., Kurths J., and Felsenberg D. (1998). Quantification of cancellous bone structure using symbolic dynamics and measures of complexity. *Physical Review E* **58**(5), 6449–6459.
- Sauer T., Yorke J.A., and Casdagli M. (1991). Embedology. *Journal of Statistical Physics* **65**, 579–616.
- Schäfer C., Rosenblum M.G., Abel H.-H., and Kurths J. (1999). Synchronization in the human cardiorespiratory system. *Physical Review E* **60**(1), 857–870.
- Scholl D. and Russ S. (1999). Air-fuel dependence of random and deterministic cyclic variability in a spark-ignition engine. Technical Paper 1999-01-3513, Society of Automotive Engineers.
- Schreiber T. (1997). Detecting and analyzing nonstationarity in a time series using nonlinear cross predictions. *Physical Review Letters* **78**(5), 843–846.
- Schwarz U., Benz A.O., Kurths J., and Witt A. (1993). Analysis of solar spike events by means of symbolic dynamics methods. *Astronomy and Astrophysics* **277**, 215–224.
- Stam C.J., Pijn J.P.M., and Pritchard W.S. (1998). Reliable detection of nonlinearity in experimental time series with strong periodic components. *Physica D* **112**, 361–380.
- Stone L., Landan G., and May R.M. (1996). Detecting Time's Arrow: a method for identifying nonlinearity and deterministic chaos in time-series data. *Proceedings of the Royal Society of London, Series B* **263**, 1509–1513.
- Takens F. (1980). Detecting strange attractors in turbulence. In D. A. Rand and L.-S. Young (Eds.), *Dynamical Systems and Turbulence (Warwick 1980)*, Volume 898 of *Lecture Notes in Mathematics*, pp. 366–381. Berlin: Springer-Verlag, ISBN 3-540-11171-9 and 0-387-11171-9.
- Tang X.Z. and Tracy E.R. (1998). Data compression and information retrieval via symbolization. *Chaos* **8**(3), 688–696.

- Tang X.Z., Tracy E.R., Boozer A.D., deBrauw A., and Brown R. (1995). Symbol sequence statistics in noisy chaotic signal reconstruction. *Physical Review E* **51**(5), 3871–3889.
- Tass P., Rosenblum M.G., Weule J., Kurths J., Pikovsky A., Volkmann J., Schnitzler A., and Freund H.-J. (1998). Detection of $n:m$ phase locking from noisy data: application to magnetoencephalography. *Physical Review Letters* **81**(15), 3291–3294.
- Theiler J., Eubank S., Longtin A., Galdrikian B., and Farmer J.D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58**, 77–94.
- Timmer J., Ganert C., Deuschl G., and Honerkamp J. (1993). Characteristics of hand tremor time series. *Biological Cybernetics* **70**, 75–80.
- Timmer J., Schwarz U., Voss H.U., Wardinski I., Belloni T., Hasinger G., van der Klis M., and Kurths J. (2000). Linear and nonlinear time series analysis of the black hole candidate Cygnus X-1. *Physical Review E* **61**(2), 1342–1352.
- Tong H. (1990). **Non-linear Time Series: a Dynamical System Approach**. Oxford University Press.
- van der Heyden M.J., Diks C., Pijn J.P.M., and Velis D.N. (1996). Time reversibility of intracranial human EEG recordings in mesial temporal lobe epilepsy. *Physics Letters A* **216**, 283–288.
- van der Welle R. (1985). Void fraction, bubble velocity and bubble size in two-phase flow. *International Journal of Multiphase Flow* **11**(3), 317–345.
- vander Stappen M.L.M. (1996). **Chaotic Hydrodynamics of Fluidized Beds**. Ph. D. thesis, Delft University of Technology.
- Vasudevan M., Finney C.E.A., Nguyen K., van Goor N.A., Bruns D.D., and Daw C.S. (1995). Stabilization and destabilization of slugging behavior in a slugging fluidized bed. In *Proceedings of the 13th International Conference on Fluidized Bed Combustion*, pp. 1001–1014.
- Voss H. and Kurths J. (1998). Test for nonlinear dynamical behavior in symbol sequences. *Physical Review E* **58**(1), 1155–1158.
- Wagner R.M., Drallmeier J.A., and Daw C.S. (1998). Prior-cycle effects in lean spark ignition combustion — fuel/air charge considerations. Technical Paper 981047, Society of Automotive Engineers.
- Weiss G. (1975). Time-reversibility of linear stochastic processes. *Journal of Applied Probability* **12**, 831–836.

- Witt A., Kurths J., and Pikovsky A. (1998). Testing stationarity in time series. *Physical Review E* **58**(2), 1800–1810.
- Yu D., Lu W., and Harrison R.G. (1999). Detecting dynamical nonstationarity in time series data. *Chaos* **9**(4), 865–870.

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, with minor in American Government, in 1991. He next studied mechanical engineering at the University of Tennessee, obtaining a B.S.M.E. in 1993, M.S. in 1995, and Ph.D. in 2000.