

Linear and Nonlinear Dynamical Systems Data Analytic Techniques
and an Application to Developmental Data

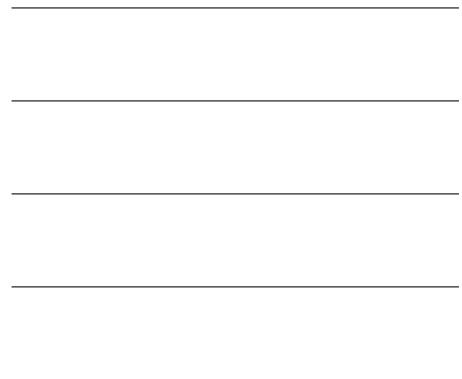
Steven Marshall Boker
Charlottesville, Virginia

B.S., University of Denver, 1972
M.A., University of Virginia, 1995

A Dissertation Presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia
May 1996



©Copyright by
Steven Marshall Boker
All Rights Reserved
May 1995

Abstract

This work presents a methodology for analyzing developmental and physiological time series from the perspective of dynamical systems. An overview of recent advances in nonlinear techniques for time series analysis is presented. Methods for generating a nonlinear dynamical systems analog to a covariance matrix are proposed. A novel application of structural equation modeling is proposed in which structural expectations can be fit to these nonlinear dependency matrices. A data set has been selected to demonstrate an application of some of these linear and nonlinear descriptive analyses, a surrogate data null hypothesis test, and nonlinear dependency analysis. The dynamical systems methods are evaluated in the light of (a) whether the techniques can be successfully applied to the example data and if so, (b) whether the results of these analyses provide insight into the processes under study which was not provided by other analyses.

Contents

1	General Introduction	1
2	Introduction to Nonlinear Methods	5
2.1	Time–Delay Embedding	8
2.2	Mutual Information	16
2.2.1	Determining an Appropriate Time Delay for Embedding	20
2.3	False Nearest Neighbors	27
2.4	Fractal Dimensions	35
2.5	Surrogate Data Tests for Nonlinearity	42
2.6	Nonlinear Prediction and Lyapunov Exponents	46
2.7	Nonlinear Noise Reduction	52
2.8	Nonlinear Dependency Matrices	54
3	Introduction to Dynamical Modeling	60
3.1	Linear Dynamical Modeling	61
3.2	Nonlinear Dynamical Modeling	64
4	An Application to Developmental Time Series	68
4.1	The Development of Postural Control	69
4.2	Experimental Methods	72
4.3	Analytic Methods	75

Dynamical Analyses of Development	iv
5 Results	78
5.1 Raw Time Series	78
5.2 Fourier Analysis	78
5.3 Surrogate Data Test for Nonlinearity	81
5.4 Mutual Information	86
5.5 False Nearest Neighbors Analysis	89
5.6 Nonlinear Dependencies	92
5.6.1 Developmental Change in Nonlinear Dependencies	107
6 Discussion	113
6.1 General Discussion	121
References	123

List of Figures

- 1 A diagram of the types of systems which may potentially underlie physiological and psychological phenomena. Note that there are many possibilities of which little is known. (After a diagram presented by Thomas Schreiber at the *Nonlinear Techniques in Physiological Time Series* conference in Dresden, October 1995.) 6
- 2 An example mutual information plot. The X axis plots the number of units of delay, and the Y axis plots the average mutual information as a percentage of the total mutual information for each corresponding delay. 18
- 3 (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y_{(t+2\pi)}$ 21
- 4 (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y_{(t+\pi)}$ 21
- 5 (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y_{(t+\pi/4)}$ 23
- 6 (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y_{(t+3\pi/4)}$ 23

7 (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y(t+\pi/2)$ 23

8 An example mutual information plot from measurements of center of pressure of a sitting, 4 month-old infant. The X axis plots the number of units of delay, and the Y axis plots the average mutual information as a percentage of the total mutual information for each corresponding delay. Note the first minimum of the function at a time delay of 5 and the evidence of periodicity from the peak at a time delay of 8. 26

9 A geometric explanation of the false nearest neighbors algorithm. When the red, green and purple points are projected in one dimension, they appear to be close to one another. When projected in two dimensions, the green point is no longer near the red and purple points. When projected in three dimensions, there is no further change in the relative distances between the points. 28

10 A false nearest neighbors plot of a 1024 sample time series of Gaussian noise. 31

11 False nearest neighbors curves for Gaussian noise added to a 1024 point time series of the Lorenz Equation. Note that as the percentage of noise increases, the slope increases for the lines in the right hand portion of the graph. 33

12 A geometric explanation of the box counting method for obtaining the fractal dimension of an attractor. The area of the attractor is measured by the number of boxes which contain points whereas the metric is length of one side of the box used to make the measurement. See the text for a more detailed explanation. 38

13 A geometric depiction of local nonlinear prediction methods. As a time series evolves, points which were near to each other in state space are expected to remain near to each other over short periods of time. The rate at which they diverge along each axis in state space is related to the Lyapunov exponents. Nonlinear prediction methods use the behavior of near neighbors in state space to predict the future evolution of a target point. See the text for a more detailed explanation. 48

14 A geometric depiction of local nonlinear prediction methods for estimating the Lyapunov exponents from a time series. Ellipsoidal neighborhoods of points around a so-called “fiducial trajectory” (labeled x in the figure) change shape as the time series evolves. An exponential curve fit to the change in the size of an ellipsoidal neighborhood along the axis of greatest change (labeled l_1 in the figure) yields an exponent known as the largest Lyapunov exponent. Other Lyapunov exponents may be estimated along the other axes of the neighborhood. Note that as the length of l_1 increases, the length of l_2 decreases. This is commonly the case for chaotic systems. 50

15 An example of a univariate first order linear difference equation as represented by a structural equation model. 61

16 An example of three first order linear difference equations with a restricted pattern of cross couplings as represented by a structural equation model. 62

17 An example of a second order linear difference equation as represented by a structural equation factor model. 63

18 Two example regression models. (A) A two variable regression model fit to a covariance matrix in the standard RAM path diagram notation. (B) A two variable regression fit to a nonlinear dependency matrix using the MIMSE nonlinear path diagram modification to RAM notation. Note that where there was one covariance double headed arrow in (A) there are now two directed dependency double headed arrows in (B). 67

19 Schematic drawing of the moving room. The child portrayed inside the room is falling backward due to perceived optic flow produced by movement of the room. Note that the floor does not move; the subjective perception of self-motion is created by moving the walls and ceiling of the room together. 73

20 Schematic drawing of the forceplate. The transducers P_1, P_2, P_3 and P_4 were sampled at 50 Hz and these data transformed to two center of pressure time series X and Y along two orthogonal axes aligned with the edges of the forceplate. 74

21 Example time series for the room movement from the first trial of all six experimental conditions for infant number 21. 79

22 Example time series for the fore-aft center of pressure from the first trial of all six experimental conditions for infant number 21. 80

23 Mean FFT plots for the room movement over all trials and all infants for each experimental condition. The 95% confidence interval for distinguishing the frequency components from white noise is indicated by the dashed line in each graph. 82

24 Mean FFT plots for for the fore–aft center of pressure over all trials and all infants for each experimental condition. The 95% confidence interval for distinguishing the frequency components from white noise is indicated by the dashed line in each graph. 83

25 Mutual information curves for the real fore–aft time series and for the surrogate time series for all time delays up to 2000 ms was calculated for each trial and then the resulting mutual information curves were averaged by taking the mean for each time delay across all of the trials. The dashed lines represent 95% confidence intervals for the mean curve ($1.96 \times$ the standard error of the mean). 86

26 Mean mutual information calculated within and between the room position time series, the fore–aft time series and the lateral time series. A mutual information curve for all time delays up to 2000 ms was calculated for each trial and then the resulting mutual information curves were averaged by taking the mean for each time delay across all of the trials. 88

27 False nearest neighbors curves for sitting infants for ages 5, 7, 9, and 13 months. Each line represents the False Nearest Neighbors curve for the data from two trials for one infant in the control condition. 91

28 Mean mutual information within the room position time series by experimental condition and across all trials and individuals. Means are calculated within experimental condition and across trials and individuals. The solid line represents the mean mutual information curve and the dashed lines above and below the mean mutual information curve represent 95% confidence intervals for the mean. 95

29 Mean mutual information of the room position time series predicting the fore–aft center of pressure time series. Means are calculated within experimental condition and across trials and individuals. The solid line represents the mean mutual information curve and the dashed lines above and below the mean mutual information curve represent 95% confidence intervals for the mean. 97

30 Mean mutual information of the fore–aft center of pressure time series predicting the room position time series. 100

31 Mean mutual information of the room position time series predicting the lateral center of pressure time series. 102

32 Mean mutual information of the lateral center of pressure time series predicting the room position time series. 103

33 Mean mutual information within the fore–aft center of pressure time series. 105

34 Mean mutual information within the lateral center of pressure time series. 106

35 Mean mutual information of the fore–aft center of pressure time series predicting the lateral center of pressure time series. 108

36 Mean mutual information of the lateral center of pressure time series predicting the fore–aft center of pressure time series. 109

37 Mean mutual information of the room position predicting the fore–aft center of pressure time series in the 0.6 Hz experimental condition. The four graphs aggregate the mean mutual information within each of the four age categories. The dashed lines are 95% confidence intervals around the mean mutual information curve. 111

38 Mean mutual information of the fore–aft center of pressure time series predicting the room position in the 0.6 Hz experimental condition. The four graphs aggregate the mean mutual information within each of the four age categories. The dashed lines are 95% confidence intervals around the mean mutual information curve. 112

39	Mean mutual information of the room position predicting the fore–aft center of pressure time series in the 0.3 Hz experimental condition. . . .	114
40	Mean mutual information of the fore–aft center of pressure time series predicting the room position in the 0.3 Hz experimental condition. . . .	115

List of Tables

- 1 Box counting results for the simulated example in Figure [12]. 39

- 2 The six frequency and amplitude room movement conditions. The baseline condition is labeled as 0 Hz frequency and 0 cm amplitude. “Multi” refers to a multiple frequency pseudorandom movement of the room. 75

- 3 Summary of results of the surrogate data tests for nonlinearity. Each cell lists the percentage of time series for which the null hypothesis of a linear stochastic system was rejected. Thus the hypothesis of a linear stochastic model being sufficient for these data was rejected in 80% of the time series for the free sitting control condition with no room movement. 85

- 4 Average false nearest neighbors slopes for each infant age group aggregated over all stimulus conditions. 92

- 5 Two mean cross dependency measures calculated over all experimental conditions. The three dependency measures are the mean over all time delays $-50 \leq \tau \leq 50$ between infant fore-aft center of pressure time series and room position time series. Values in parentheses are 95% confidence intervals for the associated statistic. Both mutual information is expressed in bits. 93

1 General Introduction

Development is a process.

This seemingly innocuous statement carries with it a large number of implications. If development is a process, then it cannot be represented as a single state. A record of development as it occurs can be represented as an ordered set of states. The development itself is not that record — a process is not an ordered set of states. However, the developmental process can be represented as the mapping between the sequence of states within this ordered set.

This is the dilemma with which developmental data analysis is confronted. Some developmental record has been measured: a set of measurements of balance over the first few years of life, a set of measurements of height over the period of childhood, a set of measurements of intelligence over a period of a lifetime. From these snapshots in time, a developmental data analyst must make some statement about the possible processes which may have generated such data sets. It is not enough merely to provide a prediction of the states. It is the nature and structure of the possible mappings between the states which must be understood if we are to understand developmental processes.

Dynamical systems analysis provides a theoretical framework in which develop-

mental data can be modeled as a developmental process. The vocabulary and techniques from dynamical systems were created in order to speak succinctly and precisely about the relationships which can be inferred to exist between ideal continuous processes and the discretely sampled records that we can observe. In the same fashion as Sherlock Holmes, one uses these techniques to eliminate the impossible; by setting bounds on the potential structure of processes some theories may be shown to be either internally inconsistent or inconsistent with the data. These techniques operate in a fashion akin to structural equation modeling, however rather than making statements about the likelihood of a structure of covariances, these dynamical systems techniques allow one to make statements about the likelihood of the topology of a dynamical system.

Development can sometimes appear to be discontinuous, first a behavior is not present and then at a later time it is present. However, the appearance of discontinuity is often dependent on a set of discrete measurements rather than a continuous measurement. In fact, the longer the interval between measurements, the more chance there is that a developmental process will be perceived as being discontinuous. Some development occurs at much shorter time scales than others, and there are many developmental processes which have periods of rapid change interleaved with periods of very relatively little change. When such a developmental process is sampled at a long time interval then apparent discontinuities can be observed which are only artifacts

of the sampling process.

One of the themes of this dissertation is that the measurement interval must be appropriate to the process under study. Sampling with too short an interval is wasteful of resources and sampling at too long an interval can miss critical structure in the developmental process of interest. We present some methods which allow one to determine whether the sampling interval used to measure a particular set of experimental data is appropriate for studying the process in question.

Simulations of nonlinear dynamical systems have shown that nonlinear time series can be entirely deterministic, that is generated without any random component, and yet exhibit behavior which appears to have an error variance when analyzed by linear statistical methods. This work will present a variety of techniques for the analysis of nonlinear time series which have the potential to be modeled as signal portions of time series that are often discarded as noise. While the methods demonstrated here are not fully developed, the first steps are taken toward a general methodology for recognizing and dealing with nonlinearity in developmental time series.

Sections 2 and 3 introduce a suite of dynamical systems descriptive and modeling techniques which represent the spectrum of linear and nonlinear methods that are currently in use in the fields of Physics and Electrical Engineering. We discuss these techniques and their applicability to developmental time series in general terms, hop-

ing to provide an introduction to these new methods and to make them accessible to behavioral scientists.

A new possibility for modeling dependency in nonlinear time series is introduced and discussed in general terms. Unfortunately, it is beyond the scope of the present work to provide a full exploration of this new modeling method, but some of the descriptive work leading up to the application of this form of nonlinear dependency modeling is covered in Section 5.6. This promising method is expected to be an active area of inquiry for the author over the next several years.

Sections 4 and 5 describe an experiment concerning the development of postural control in infants and present the results of applying dynamical systems techniques to the data from this experiment. Some of the results of linear analyses are presented as well as another set of results from nonlinear analyses. Finally, the differences between the linear and nonlinear analyses are discussed and compared.

Dynamical systems analysis carries with it a set of terms which have come to be used to describe time series and theoretical models for the phenomena observed. We will begin with a general introduction to these ideas, since some of the ideas and many of the terms may be unfamiliar to behavioral scientists.

2 Introduction to Nonlinear Methods

Over the past 20 years it has become fairly common knowledge that there exist simple, deterministic, mathematical systems whose behavior can look very random. These systems have come to be known as *chaotic* systems and are often represented as having so-called *strange attractors* (Ruelle & Takens, 1971). The lure of these chaotic systems is that natural phenomena which behave in seemingly random patterns may actually have some simple, deterministic explanation. Short-term intraindividual variability (Nesselroade, 1991) in a variety of psychological measures is one type of seemingly random fluctuation which may have some deterministic structure.

Chaotic systems are but one type of nonlinear systems, and nonlinear systems are but one type of dynamical systems; and one should not forget linear and stochastic systems (See Figure [1]). What the presence of chaotic systems has done is provide another axis along which deterministic systems could be distinguished from stochastic systems. The promise of a payoff from chaotic systems should not blind us to the fact that dynamical systems theory (both linear and nonlinear) has much to offer the analysis of developmental and physiological data.

In order to begin the inquiry, it is first necessary to define a few terms. A *system* will be taken to be a set of variables which change over time, and which have a some natural connectedness (such as all being measured from the same individual).

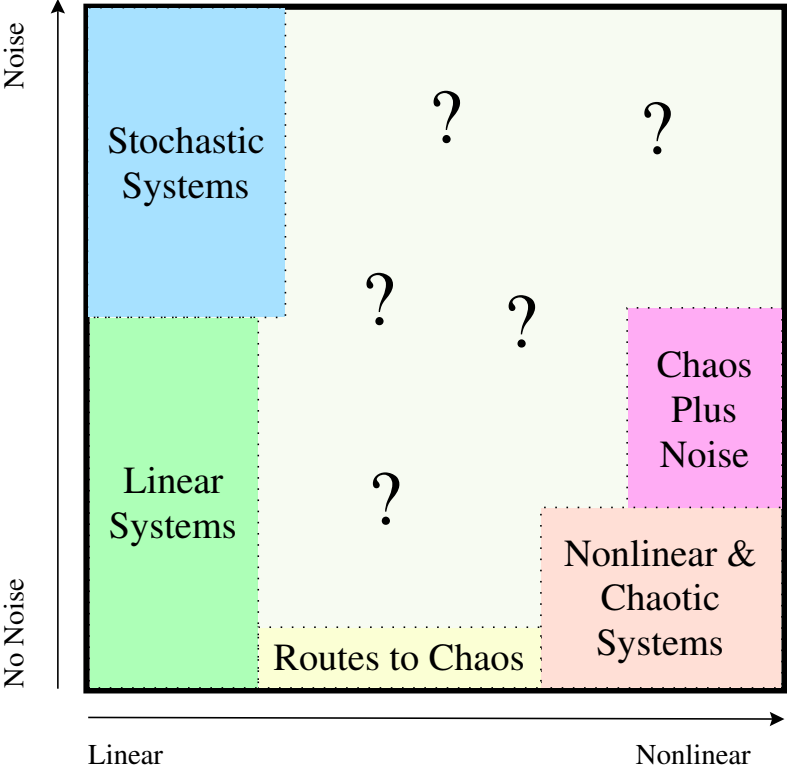


Figure 1: A diagram of the types of systems which may potentially underlie physiological and psychological phenomena. Note that there are many possibilities of which little is known. (After a diagram presented by Thomas Schreiber at the *Nonlinear Techniques in Physiological Time Series* conference in Dresden, October 1995.)

A *dynamical system* will be defined to be a system in which the present state (the values of all of the variables and all of their derivatives) is somehow dependent on previous states of the system. A *deterministic system* will be taken to be a system in which the present state is **entirely** dependent on previous states of the system. A *linear system* is a system in which all of the dependence of the current state on previous states can be expressed in terms of a linear combination. A *linear stochastic system* is a system in which all of the dependence of the current state on previous states can be expressed in terms of a linear combination and the residual unpredictable portions can be expressed as additive, independent, identically distributed, random variables.

A *nonlinear system* is a system in which the dependence of the current state on previous states cannot be expressed entirely as a linear combination; even if some of the dependence can be captured in a linear combination of the previous states, something extra is required to capture all of the dependence. That extra “something” is frequently an interaction between variables, i.e. a multiplicative term involving the previous states. A *nonlinear stochastic system* exists somewhere among the question marks in Figure [1]; very little is known about these systems, but this work will propose methods to begin to deal with these systems (See Section [2.8] and Section [3] for an introduction to these methods).

If an infinitesimal difference in the initial conditions (initial state) of two real-

izations of the same dynamical system causes an exponential divergence between the trajectories of the two systems, then that system is said to have *sensitive dependence on initial conditions*. A *chaotic system* is a nonlinear dynamical system which exhibits sensitive dependence on initial conditions. Linear dynamical systems never show sensitive dependence on initial conditions. Some nonlinear dynamical systems show sensitive dependence with some initial conditions and no sensitive dependence with other initial conditions. These systems are sometimes chaotic and sometimes not chaotic depending on their state, and can be changed from a non-chaotic regime to a chaotic regime and back again by external manipulations of their variables.

There has been a large body of work over the past 15 years which has dealt with methods for determining whether a system exhibits dynamical properties, whether a system is nonlinear, and whether a system is exhibiting chaotic behavior. The following seven subsections introduce some of the methods which may be use in characterizing systems from time series data.

2.1 Time-Delay Embedding

In order to measure the structure of a system that, according to some theory, might be a dynamical system, one must employ some method that will capture the theoretical dynamics from data which can be empirically gathered. Experimental data are almost

never continuous. They are gathered by intermittent repeated observations of the system in question. Thus, a practical method for analyzing a dynamical system from empirical observation must preserve the essential qualities of the underlying continuous dynamics given discrete data. One method which has been used to good effect is *time-delay embedding* (Packard, Crutchfield, Farmer, & Shaw, 1980). Time delay embedding uses short “snips” of a time series as its unit of analysis. It has been shown that under reasonable conditions, a time delay embedding preserves the quantities in which we are interested (Whitney, 1936; Takens, 1981; Sauer, Yorke, & Casdagli, 1991).

Suppose we are interested in some time-varying phenomenon which might have some underlying dynamical process. We measure this phenomenon using some instrument at equal intervals over some length of time. In so doing, we have generated a time series, an ordered set

$$X = \{x_t, x_{t+1}, x_{t+2}, \dots, x_{t+N}\} \quad (1)$$

where N is the total number of measurements and t is the time of the initial observation of the system. One of the first questions which must be addressed is, “How well does this time series represent the underlying dynamical system?” There are two main potential problems which may render ineffective a time series measurement of a dynamical system: the resolution of the measuring instrument may be insufficient to

capture important variation in the measured quantity, or the interval between measurements may be inappropriate with respect to the time constants of the dynamics of interest (Nesselroade & Boker, 1994).

These two measurement problems can be addressed using techniques from information theory (Shannon & Weaver, 1949), in which one asks the question, “How much is learned by making a measurement given what I know from the preceding measurement?” If one is entirely surprised by each new measurement, that is to say that each measurement contributes the same amount of information as it would if one had no knowledge of the preceding measurements, then the time series is reduced to a sequence of independent measurements. In this case there is no dynamical structure which can be discerned. If one is not at all surprised by each new measurement, that is to say that each measurement is exactly the same as the preceding measurement, then the time series is reduced to a sequence of repeating values. In this case once again there is no dynamical structure which can be estimated from the time series.

Thus, there is some optimum range of information which should be learned by each measurement in a time series: too much information per measurement and the time series cannot be distinguished from a random process, and too little information per measurement and the time series can not be distinguished from a static process. In Section [2.2], *mutual information* methods are presented which can estimate the information inherent in a time series and which can help answer the question of

whether a time series has been measured in such a way that it is possible to discern the structure of an underlying dynamical process.

Suppose that the time series from Equation [1] above has been determined to be a sufficient measurement of what might potentially be a dynamical system. Such dynamical systems are often represented as systems of differential equations. A differential equation represents the current state of some variable as a function of the previous value of the variable and first, second and perhaps higher order derivatives of the variable at a previous time step, where the time step becomes vanishingly small. In a time series, these derivatives must be estimated from discrete samples, so they reduce to taking first and second (and possibly higher order) differences. Consider the following hypothetical example: a time series X is thought to be a manifestation of the difference equation

$$x_t = b_0 x_{t-\tau} + b_1 \frac{x_{t-\tau} - x_{t-2\tau}}{\tau} + b_2 \frac{(x_{t-\tau} - x_{t-2\tau}) - (x_{t-2\tau} - x_{t-3\tau})}{\tau}. \quad (2)$$

Each possible state of this second order difference equation can be represented by a location in a 3 dimensional space in which the first axis is $x_{t-\tau}$, the second axis is $(x_{t-\tau} - x_{t-2\tau})/\tau$ and the third axis is $((x_{t-\tau} - x_{t-2\tau}) - (x_{t-2\tau} - x_{t-3\tau}))/\tau$. Such a representation is called the *phase space* of the system. Many deterministic systems occupy only a small portion of the possible phase space and often have characteristic

topologic and geometric properties within that phase space that help identify the form of equations to use when modeling the system. If, after being perturbed by some outside influence, a system returns to a stable, bounded region of phase space then that system is said to be dissipative. This stable, bounded region of phase space is called an *attractor* for the dynamical system. Many of the systems of interest to psychologists and physiologists fall into this category.

While the phase space is a convenient form for studying the behavior of differential equations, when working with discrete measurements such as are presented in time series, the phase space representation poses some problems. Note that each difference term incorporates all of the variables from the previous difference term except that an additional variable at a lag one time step longer is incorporated. Thus the first difference uses $x_{t-\tau}$ and $x_{t-2\tau}$, while the second order difference uses $x_{t-\tau}$, $x_{t-2\tau}$ and $x_{t-3\tau}$. If each of these $x_{t-i\tau}$ contain some error of measurement, then this error is compounded in the estimation of each of the difference terms: the higher order terms have more error than the lower order terms.

This increase in measurement error is one of the motivations for the development of another time series representation called a *state space*. A state space for the second order difference equation shown in Equation [2] represents each possible state by a location in a 3 dimensional space where the first axis is $x_{t-\tau}$, the second axis is $x_{t-2\tau}$ and the third axis is $x_{t-3\tau}$. The question then arises, “Does a state space preserve

the essential topologic and geometric properties of the attractor as represented in phase space?” The short answer to this question is that given certain assumptions, a properly constructed state space will almost always preserve all of the topologic properties and many of the geometric properties of interest.

A state space for a time series is constructed by *embedding* the time series into a higher dimensional space. The most widely used version of embedding is *time-delay embedding*. For example in a four dimensional embedding, a time series vector $X = \{x_t, x_{t+1}, x_{t+2}, \dots, x_{t+N}\}$ would be expanded into the matrix

$$\mathbf{X}^{(4)} = \begin{bmatrix} x_{(t+0)} & x_{(t+0)+\tau} & x_{(t+0)+2\tau} & x_{(t+0)+3\tau} \\ x_{(t+1)} & x_{(t+1)+\tau} & x_{(t+1)+2\tau} & x_{(t+1)+3\tau} \\ x_{(t+2)} & x_{(t+2)+\tau} & x_{(t+2)+2\tau} & x_{(t+2)+3\tau} \\ x_{(t+3)} & x_{(t+3)+\tau} & x_{(t+3)+2\tau} & x_{(t+3)+3\tau} \\ \vdots & \vdots & \vdots & \vdots \\ x_{(t+N)-3\tau} & x_{(t+N)-2\tau} & x_{(t+N)-\tau} & x_{(t+N)} \end{bmatrix}. \quad (3)$$

The idea of embedding was first proposed by Whitney (1936) who showed that a differentiable map from a d -dimensional differentiable manifold M to the Euclidean real space \mathfrak{R}^{2d+1} is a diffeomorphism on M . The consequence of this theorem is that if a dynamical system S is differentiable, there exists an embedding of S which will thus preserve the topology of its attractor. Whitney’s result is not useful for time series analysis, since the result required a differentiable map and this type of map is not available for a discrete time series.

The Whitney embedding theorem was extended to time–delay maps by Takens (1981), who showed that for a generic dynamical system, the time–delay map from a d –dimensional differentiable manifold M to the Euclidean real space \mathfrak{R}^{2d+1} is a diffeomorphism on M . Takens’ embedding theorem implies that if the attractor for a generic dynamical system is the manifold M , then there exists a time–delay map which will preserve the topology of the attractor. This time–delay map is the function which embeds the time series vector X into the matrix $\mathbf{X}^{(2d+1)}$ as shown in the example above in Equation 3. An alternative and more accessible proof of Takens’ embedding theorem was given by Noakes (1991).

Sauer et al. (1991) recently extended Takens’ embedding theorem to the case where the attractor for the dynamical system is a smooth *fractal* (see Mandelbrot, 1977; Peitgen, Jürgens, & Saupe, 1992, for introductions to fractals). A smooth fractal maintains continuity at every point, but may not be differentiable at any point. Sauer et al. also showed that almost every time–delay map from a smooth manifold M with a box–counting fractal dimension of d into the Euclidean real space \mathfrak{R}^{2d+1} is a diffeomorphism on M . The estimation of box–counting fractal dimension and other fractal dimensions will be discussed in detail in Section [2.4]. Sauer et al.’s results are important for time series analysis both because they say that a lower bound on the required dimension of the embedding space can be estimating directly from the data, and because they say that every time–delay map of the time series

which exceeds that lower bound will, with a probability approaching 1.0, preserve the topology of the attractor. A notable exception is when a time delay is selected which is an exact multiple of one of the main periods of a periodic system – in that case, the embedding may not be topology preserving. An example of this problem is shown in Section [2.2.1] where a function whose attractor is a limit cycle can degenerate to a line if the time delay is improperly chosen.

All of the methods applied in the present work will use, in one way or another, a time-delay embedding to change a time series into an embedded matrix. With that in mind, the next three subsections will describe methods which pertain to choosing an optimal embedding for a particular time series. In order to perform an optimal embedding two parameters must be discerned from the data.

1. The time delay τ must be estimated which will best expand the attractor in the state space. One method to choose τ uses estimates of mutual information with respect to time-delay; this method will be discussed in Section 2.2.1.
2. The dimension, D_e , of the Euclidean real space in which to embed the time series must be determined. D_e is equivalent to the number of columns in the embedded matrix. A minimum sufficient D_e can be determined from the data either using the method of *false nearest neighbors* described in Section 2.3 or by using the box-counting dimension as described in Section 2.4.

2.2 Mutual Information

Information theory (Shannon & Weaver, 1949) provides a measure for nonlinear dependence within and between time series. When a sequence of measurements of a variable are taken over a period of time, one can estimate the uncertainty in the prediction of the next measurement given the preceding measurements (see Resnikoff, 1989, for an overview). If we have two time series, U and V , the uncertainty, $I_{UV}(u_i, v_j)$, about a measurement u_i given v_j is called the *mutual information* of the two measurements, and is expressed as

$$I_{UV}(u_i, v_j) = \log_2 \left[\frac{p_{UV}(u_i, v_j)}{p_U(u_i)p_V(v_j)} \right], \quad (4)$$

where $p_U(u_i)$ is the probability of observing u_i in the time series U , $p_V(v_i)$ is the probability of observing v_i in the time series V , and $p_{UV}(u_i, v_j)$ is the joint probability of observing u_i and v_i in the two time series (Abarbanel, Brown, Sidorowich, & Tsimring, 1993). The *average mutual information* of the two time series is the mean mutual information over all of the measurements in the two series:

$$I_{UV}(T) = \frac{1}{N} \sum_{i,j} p_{UV}(u_i, v_j) \log_2 \left[\frac{p_{UV}(u_i, v_j)}{p_U(u_i)p_V(v_j)} \right]. \quad (5)$$

The average mutual information within a time series can be defined similarly. If

we have a time series U , the average uncertainty about a measurement $u_{t+\tau}$ given a measurement u_t at time t over all t is the average mutual information $I(\tau)$,

$$I(\tau) = \frac{1}{N} \sum_{t=1}^T p(u_t, u_{t+\tau}) \log_2 \left[\frac{p(u_t, u_{t+\tau})}{p(u_t)p(u_{t+\tau})} \right] \quad (6)$$

in the series U for a lag τ .

Figure [2] shows an example plot of the average mutual information within a single time series over a range of time delays (or *lags*). Note there is a relatively large amount of information shared between measurements separated by a small time delay, but as the time delay increases the amount of information shared by the measurements decreases. In any continuously varying system it is to be expected that the amount of mutual information between two measurements would increase as the time between the measurements becomes small. However, notice that in Figure [2] there are some other time delays when the mutual information has a maximum. These maxima are signatures of periodicity in time series. Without further tests (see Section [2.5]), it remains to be seen whether these are linear or nonlinear periodicities.

The mutual information function applied within a single time series is very similar to the autocorrelation function in that a measurement of the degree of dependency is sought. Mutual information differs in that it continues to measure dependency even when the system is nonlinear or chaotic — a case when autocorrelation may suggest

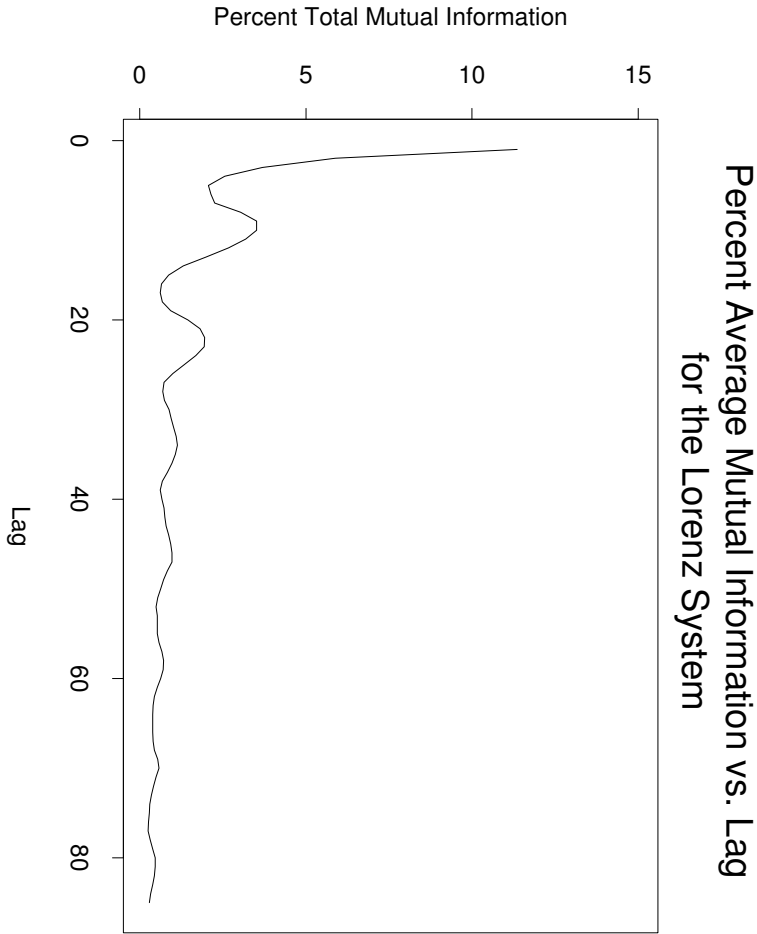


Figure 2: An example mutual information plot. The X axis plots the number of units of delay, and the Y axis plots the average mutual information as a percentage of the total mutual information for each corresponding delay.

an independent random process. The autocorrelation function also requires that the distributions of u_t and $u_{t+\tau}$ be jointly normal in order for it to make an accurate calculation of dependency, whereas the mutual information function requires no such assumption. The advantages of autocorrelation are that it can be calculated quickly and it does a good job of describing a linear system. The disadvantage is that the autocorrelation function always assumes that the underlying process is linear and calculates a value for dependency; if the underlying process is not linear, the value would be incorrect.

The mutual information function has proved to be useful in a wide variety of non-linear analyses, but its calculation can prove to be extremely computationally intensive for long time series. However, recent developments by Fraser and Swinney (1986) and Pompe (1993, 1994, 1995) have reduced the calculation to manageable proportions for the univariate case. Fraser and Swinney introduced a rapid method for estimating the required joint probabilities by a joint median cut sorting algorithm. Pompe's Generalized Mutual Information algorithm takes a different approach and uses the relationship of the mutual information function to the fractal dimension (Eckmann & Ruelle, 1985) to estimate mutual information within a long time series rapidly by calculating the correlation integral (Grassberger & Procaccia, 1983b) of the time series. This estimation technique can be applied to time series of at least 1000 time points, and generates an average mutual information plot in less than 1/100th of the

time required for the full average mutual information calculation.

2.2.1 Determining an Appropriate Time Delay for Embedding

The average mutual information function can be applied to determine reasonable values for the time delay to be used to create a time–delay embedding of a time series. The choice of time delay will influence the geometry of the attractor which is to be reconstructed within the embedding (Liebert & Schuster, 1989). If the dynamical system is periodic and a time delay is chosen which is an exact multiple of the period of the system, then the embedding may not preserve the topology of the attractor. In order to illustrate the effect of choosing the time delay, consider the case of a time series generated by the function $y(t) = \sin(t)$ as shown in Figure 3–A. If a time delay of $\tau = 2\pi$ is chosen and the time series is embedded into a two dimensional space, the state space portrait shown in Figure 3–B will result.

In order to see how the time series maps onto the state space, four points in state space have been chosen: one red, one green, one yellow and one purple. Consider the red point with a value of $y(t) = 1.0$ in the upper left of Figure 3–A. After a time delay $\tau = 2\pi$ has elapsed, the red point still has a value of $y_{(t+2\pi)} = 1.0$. These two values are plotted in Figure 3–B, $y(t) = 1.0$ on the X -axis and $y_{(t+2\pi)} = 1.0$ on the Y -axis. Similarly the purple dot starts with a value of $y(t) = -1.0$ and after $\tau = 2\pi$

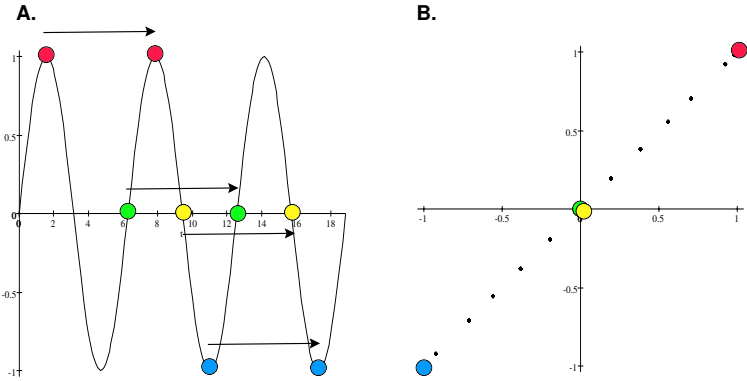


Figure 3: (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y(t+2\pi)$.

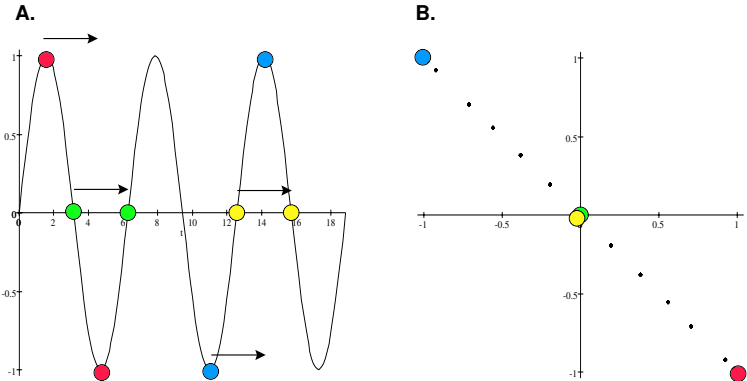


Figure 4: (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y(t+\pi)$.

has elapsed, has a value of $y_{(t+2\pi)} = 1.0$. These two values determine where in the state space the purple dot is plotted.

But now consider what happens with the green and yellow points. Each starts with a value of $y_{(t)} = 0.0$ and after $\tau = 2\pi$ has elapsed, has a value of $y_{(t+2\pi)} = 0.0$. Therefore they both map to the same point in the state space. However, the green and yellow points have very different derivatives from one another, thus they will map to two very different points in a phase plot. This illustrates a problem which can occur with time-delay embedding. Since in this case the time delay was chosen to be exactly the period of the function which generated the time series, the time-delay embedded state space has not preserved the topology of the attractor for the function.

When the time delay is chosen to be $1/2$ the period of the function, a similar problem results. In Figure [4] a time delay of $\tau = \pi$ has been chosen. Now the red dot evolves from $y_{(t)} = 1.0$ to $y_{(t+\pi)} = -1.0$ and the purple dot evolves from $y_{(t)} = -1.0$ to $y_{(t+\pi)} = 1.0$ after an elapsed delay of $\tau = \pi$. However, the green and yellow points still are mapped onto the same point in the time-delay embedded state space. In general one must avoid embedding time delays which are $1/2$ integer multiples of the period of a system.

Figure [5] and Figure [6] illustrate state space plots of the example time series with time delays of $\tau = \pi/4$ and $\tau = 3\pi/4$ respectively. Certainly these time delays

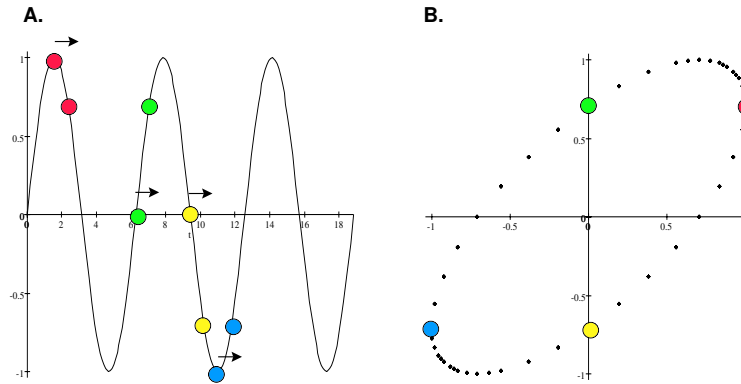


Figure 5: (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y(t+\pi/4)$.

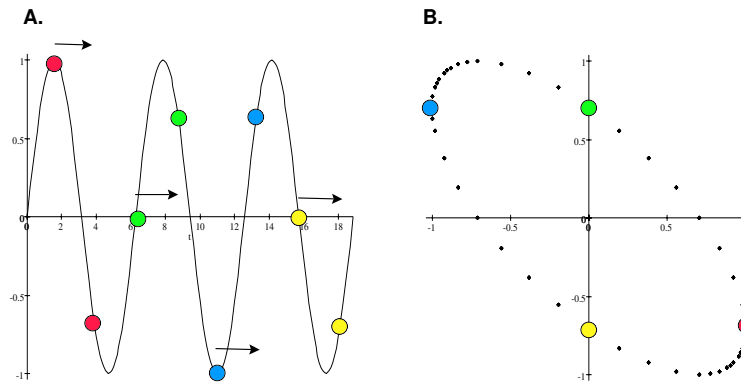


Figure 6: (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y(t+3\pi/4)$.

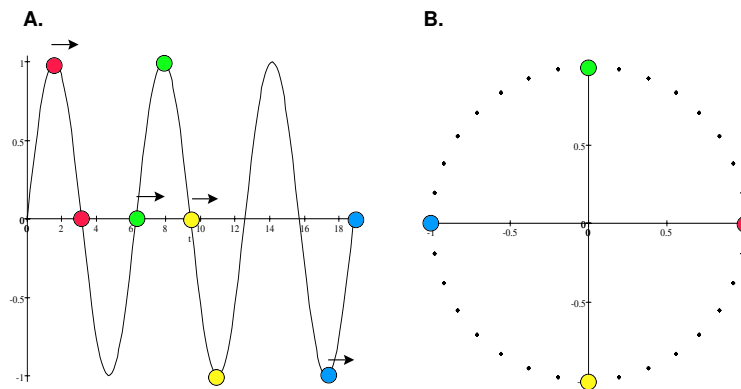


Figure 7: (A) Time series generated by the function $y(t) = \sin(t)$. (B) 2-D State space plot of $y(t)$ against $y(t+\pi/2)$.

are better than the first two time delays since the green and yellow points are now mapped onto two different points in the state space. It now becomes apparent that the attractor is a limit cycle and that its topology is the same as that of a circle. In this sense, a successful embedding has been constructed. However, note that although the points in these two state space plots were sampled at equal intervals in the time series, there appears to be a greater density of points near the ends of the ellipses in Figure [5-B] and Figure [6-B].

If a time delay of $\tau = \pi/2$ is chosen, the state space shown in Figure [7-B] results. In this case the attractor occupies a maximum area in the state space and the density of points is equal across the state space. This last point is subtle, but important. The density of points in state space can be an artifact of the time delay chosen for embedding the time series. The best time delay is one which minimizes the contribution of the time delay artifact to the density of points in the state space.

This example time series was linear and the optimum time delay can be calculated as the first zero crossing of the autocorrelation function as the time delay is increased. The first minimum of the mutual information function for this time series would produce the same value for the optimum time delay. In general, for nonlinear periodic series, a good choice for the embedding time delay is the first minimum of the mutual information function of the series.

In practice, the choice of a time delay must also provide a balance between eliminating autocorrelation and preserving nonlinear dependency. The time delay must be long enough to remove much of the autocorrelation in the time series. Starting from the left side of the plot in Figure [2], there is a large value for mutual information which falls rapidly as the the time delay approaches 5. This rapid decline in mutual information is generally due to autocorrelation in the data. A better way to say this is that a linear system would do a good job of prediction at time lags less than 5, so a time delay of at least 5 is required in this case in order that the nonlinear structure in the data is not overwhelmed by the linear predictability.

The longer the time delay, the more data which is lost from each time series during the embedding process, so the minimum time delay which can eliminate most of the autocorrelation is going to be a good choice when the time series are short. In addition, the longer the time delay, the less of the nonlinear structure is preserved if there is any noise or perturbation of the system during the total elapsed time of the series. In almost any psychological time series there are guaranteed to be external events which will perturb the system between measurements. Thus there is a strong argument that the shortest acceptable time delay should be used. For these reasons, it is recommended that the shortest time delay be used which eliminates most of the autocorrelation in the series. Unfortunately, this is not an exact science; there is some room for individual judgment in picking the time delay. However, the theorems of

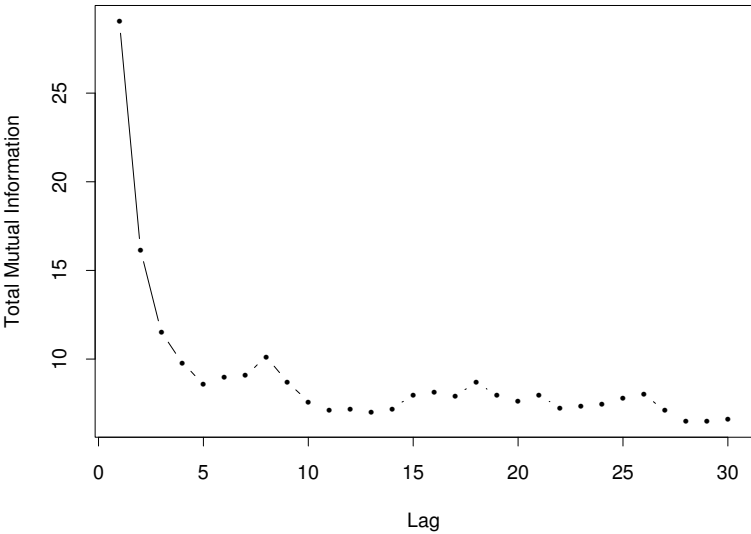


Figure 8: An example mutual information plot from measurements of center of pressure of a sitting, 4 month–old infant. The X axis plots the number of units of delay, and the Y axis plots the average mutual information as a percentage of the total mutual information for each corresponding delay. Note the first minimum of the function at a time delay of 5 and the evidence of periodicity from the peak at a time delay of 8.

Sauer et al. (1991) suggest that there is a very low probability that the choice of time delay will lead to an embedding which does not preserve the topology of potential attractors in the data.

2.3 False Nearest Neighbors

The method of *false nearest neighbors* is designed to determine how many dimensions are sufficient to embed a particular time series (Kennel, Brown, & Abarbanel, 1992). Recall that the number of dimensions in the embedding space is equivalent to the number of columns in the matrix which results from embedding the time series. Of course, it would be wasteful to use more columns than were necessary; wasteful not only of computer resources, but more importantly wasteful of the data. Frequently, collection of data is a demanding and expensive process and one wishes to be able to make the most of every precious sample. Adding an extra column onto the embedding matrix costs an extra sample from the end of each longitudinal sequence of data. In data sets where there are mixed cross-sectional and longitudinal data as well as missing data this can prove to be especially costly; each extra column will cause some reduction in the number of available rows. On the other hand, if the number of columns is not sufficient, the state space will not accurately reflect the true topology of the attractor for the system under study. An optimum value for the dimension of the embedding space then becomes the minimum value which will correctly represent

the topology of the attractor.

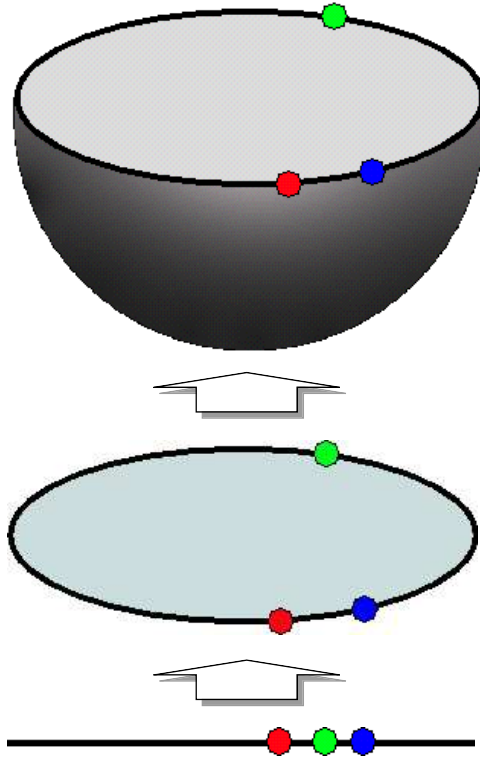


Figure 9: A geometric explanation of the false nearest neighbors algorithm. When the red, green and purple points are projected in one dimension, they appear to be close to one another. When projected in two dimensions, the green point is no longer near the red and purple points. When projected in three dimensions, there is no further change in the relative distances between the points.

The basic idea behind false nearest neighbors is that points in a state space should be close to each other because their dynamical state is similar, not because they have been projected close to each other as an artifact of constructing the embedding with a dimension which is too low. Figure [9] presents a geometric explanation of the concept that is at the core of the false nearest neighbors technique.

The time series is first embedded into a one dimensional state space. For each point in the state space, its nearest Euclidean neighbor is found. In the example, the line at the bottom of Figure [9] represents the one dimensional state space and the nearest neighbor of the red point is the green point. Next, the time series is embedded into a two dimensional state space as represented by the oval in the middle of Figure [9]. The red point and the green point are no longer near to each other, so the green point is labeled a “false nearest neighbor” because it was only near to the red point due to the projection of the time series onto the line.

Next, the nearest neighbor for each point in the two dimensional state space is found. Now the nearest neighbor to the red point is the purple point. The time series is now embedded into a three dimensional state space as represented in the hemisphere at the top of Figure [9]. The red point and the purple point are still near to each other and so the purple point is not a false nearest neighbor. This process continues until either there are no further false nearest neighbors or the data set becomes so sparse in a high dimensional space that no points can be considered to be near neighbors to begin with.

The false nearest neighbors algorithm can be summarized as follows:

1. Find the nearest neighbor for each point in an embedding of dimension D .
2. Find the percentage of those nearest neighbors which do not remain within a

ratio of their original distance in an embedding of dimension $D + 1$

$$\frac{|x_i^{(D+1)} - x_j^{(D+1)}|}{|x_i^{(D)} - x_j^{(D)}|} > R_T \quad (7)$$

where $|x_i^{(D)} - x_j^{(D)}|$ is the distance between the two points in dimension D and $|x_i^{(D+1)} - x_j^{(D+1)}|$ is the distance between the two points in dimension $D + 1$.

3. If there is no nearest neighbor x_j for a point x_i such that

$$\frac{|x_i^{(D)} - x_j^{(D)}|}{\sigma_X} < R_A \quad (8)$$

where σ_X is the standard deviation of x over the whole time series, then x_i is declared to have a false nearest neighbor.

The resulting percentage of false nearest neighbors for each embedding dimension is then plotted against the corresponding embedding dimension to create a false nearest neighbors plot as shown in Figure [10]. This plot shows how the false nearest neighbors algorithm reacts to a time series which is Gaussian noise. Note that there is no point where there are no false nearest neighbors and that as the embedding dimension becomes larger than 4, the number of false nearest neighbors begins to increase again.

This effect of additive noise on the false nearest neighbors plot can be exploited to examine the relative amount of additive noise mixed with a nonlinear dynamical

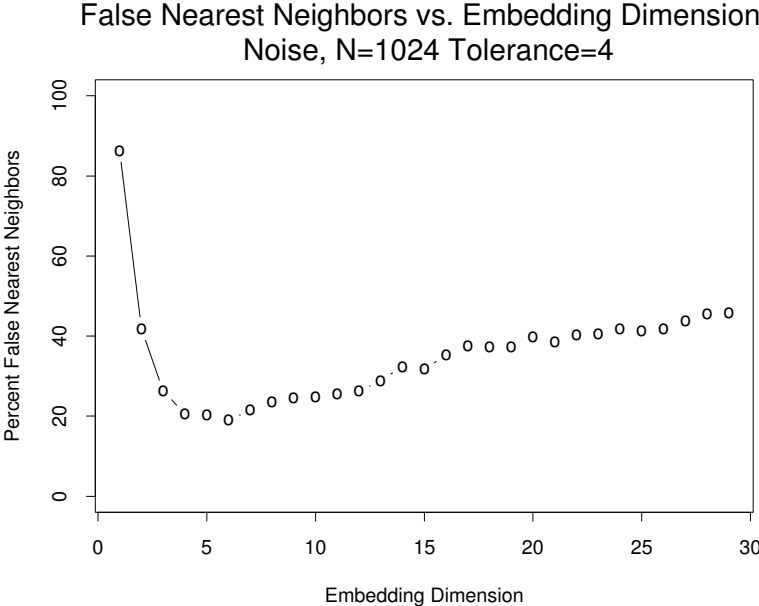


Figure 10: A false nearest neighbors plot of a 1024 sample time series of Gaussian noise.

signal. The simulation plotted in Figure [11] exemplifies this phenomenon. A non-linear signal, a time series of the Lorenz Equation, is mixed with increasing amounts of Gaussian noise. Note that as the amount of noise is increased, the slope increases on the right hand side of the plot. Also note that there is little difference in the lines to the left of the first minimum of the false nearest neighbors curve, the algorithm suggests that a minimum embedding dimension of three is required to represent the dynamical portion of the signal in this series. Further methods to separate additive noise from deterministic dynamics will be presented in Section [2.7]

The false nearest neighbors algorithm has been shown to be fallible as a detector of deterministic dynamics (Fredkin & Rice, 1995), so it is not recommended that the algorithm be used as such. In particular, short linear stochastic time series which have long time delay autocorrelations and nonlinear deterministic systems may have similar false nearest neighbors plots. The *surrogate data* methods presented in Section [2.5] are better detectors of nonlinear determinism than is the false nearest neighbors method. A recent extension which Kennel and Abarbanel (1995) call *false nearest strands* provides more rigid conditions for the existence of a near neighbor and is less susceptible to mistaking linear stochastic processes for nonlinear determinism. However, the conditions are so stringent that for the short (and potentially noisy) time series available to psychologists, the false nearest strands method may prove to be less useful than the false nearest neighbors method.

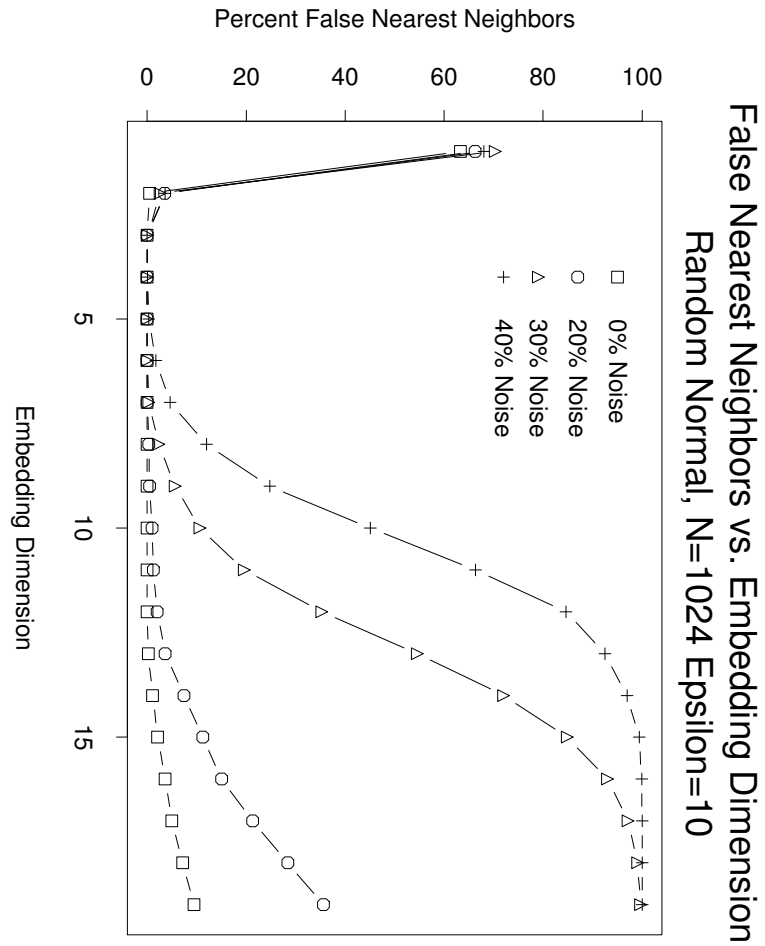


Figure 11: False nearest neighbors curves for Gaussian noise added to a 1024 point time series of the Lorenz Equation. Note that as the percentage of noise increases, the slope increases for the lines in the right hand portion of the graph.

There is another difficulty when using the false nearest neighbors method. This is the problem of how to choose the proper R_T and R_A for the time series to be analyzed. At this point there is no firm and fast answer to this question. The value of R_T has proved to be a relatively insensitive parameter in numerical simulations performed by Abarbanel et al. (1993), and the same authors suggest a value equal to the RMS value of the observations

$$R_A = \sqrt{\frac{1}{N} \sum_{k=1}^N (x_k - \bar{x})^2} . \quad (9)$$

In practice it is best to look at the sensitivity of the R_T and R_A parameters with respect to the time series in question rather than rely on a set rule.

Finally, the false nearest neighbors method can be computationally lengthy since it needs to locate a nearest neighbor for every vector in every embedding dimension. A variety of pre-sorting algorithms have been proposed to speed the search for near neighbors in an embedding, since several of the nonlinear analysis techniques such as nonlinear prediction, nonlinear smoothing and Lyapunov exponents calculations require the search for near neighbors (see Section [2.6] and Section [2.7]). Schreiber (1995) has evaluated several of these pre-sorting and tree building techniques and has concluded that pre-sorting can reduce nearest neighbor searches to less than 3% of the time required for a naive search.

2.4 Fractal Dimensions

One of the most interesting measurements which has been developed for analysis of nonlinear systems is the *fractal dimension* (Mandelbrot, 1977). Mandelbrot (1967) suggested that the coastline of Britain had a property of “ruggedness” or “complexity” which was similar at a wide range of scales and that could be measured. The method he proposed can be described as follows. If one uses a yardstick 10 kilometers long to measure the length of a coastline, one finds a length L_1 . If one uses a yardstick 1 kilometer long, one finds a second length $L_2 > L_1$ because one is now measuring smaller bays and inlets which had escaped the first measurement. If one uses a yardstick 1 meter long one finds another length $L_3 > L_2$, because one is now measuring around individual rocks. If one uses a yardstick 1 millimeter long one finds yet another length $L_4 > L_3$, because one is now measuring around individual grains of sand. In fact, as one decreases the length of the yardstick, the length of the coast increases without bound. Mandelbrot found the surprising result that when the log of the length of the ruler is plotted against the log of the length of the coastline, there is a linear relation between the two log lengths. The fractal dimension of the coastline is the slope of a regression line fit to this log–log plot.

It is important to note that fractal dimensions do not need to be integers, as dimensions have customarily been considered. The fractal dimension has the appeal-

ing property that carefully constructed systems can be shown to provide a smooth transition from a D dimensional process through fractional dimensions to a $D+1$ dimensional process. In this way, the fractal dimension provides a continuous measure of a quantity which is similar to the complexity of a system.

The definition of fractal dimension depends on scale invariant self-similarity within the data to which it is applied. This scale invariant self-similarity may only hold within some scales of the data (called *scaling regions*), and not in other scales. In that case, the regression line is only fit to the scaling region in which the log-log relationship between the metric (the yardstick) and the measurement appears linear.

If a data set is demonstrated to have a fractal dimension this does not mean that it is necessarily a nonlinear dynamical system. Many interesting linear stochastic systems exist that generate time series with stable and non-integer fractal dimensions. However, the fractal dimension has been used in surrogate data methods to distinguish between nonlinear deterministic dynamical systems and linear stochastic systems. These methods will be described in Section [2.5]. Much of the work on the fractal dimension of time series has been motivated by the discovery that the attractors for known chaotic systems had characteristic fractal dimensions (Grassberger & Procaccia, 1983a).

There are three main types of fractal dimension which are commonly estimated:

the *box counting dimension*, *correlation dimension* and *information dimension*. Hentschel and Procaccia (1983) showed that these dimensions are related to each other in such a way that the correlation dimension was the next higher moment of the box counting dimension and the information dimension was the next higher moment of the correlation dimension (for a more accessible treatment see Schroeder, 1991). Essentially, the box counting dimension uses Euclidean distance as its metric, the correlation dimension uses autocorrelation as its metric and the information dimension uses mutual information as its metric.

The box counting fractal dimension for a time series can be calculated by the following simple geometric method. Create a time–delay embedding of the time series into a 2 dimensional state space, taking care to choose a time delay which will spread the attractor over as much of the state space as possible, as described in Section [2.2.1]. In order to illustrate the method, a few points from a simulated state space are plotted in Figure [12–A]. Consider the length of one side of the box around the state space in Figure [12–A] to be 1; and thus the area of the attractor is also 1 because there is only one box and it has points from the state space within it. In Figure [12–B] the area of each of the boxes is now $1/4$ and there are 4 boxes which have points contained within them, so the area of the attractor is $4 \times 1/4 = 1$. In Figure [12–C] the area of each of the boxes is now $1/16$ and there are 13 boxes which have points contained within them, so the area of the attractor is $13/16$. In Figure [12–D] the area of each

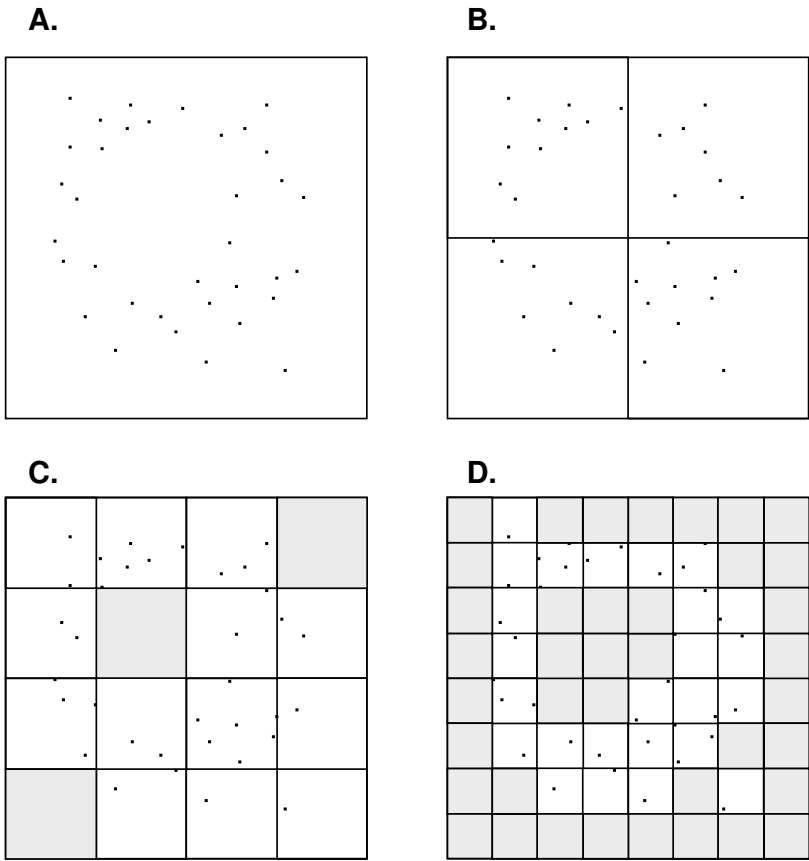


Figure 12: A geometric explanation of the box counting method for obtaining the fractal dimension of an attractor. The area of the attractor is measured by the number of boxes which contain points whereas the metric is length of one side of the box used to make the measurement. See the text for a more detailed explanation.

of the boxes is now $1/64$ and there are 25 boxes which have points contained within them, so the area of the attractor is $25/64$. Continue in this manner until there is only one dot in the majority of the boxes with any dots in them.

Table 1: Box counting results for the simulated example in Figure [12].

Length of Side	Area of Attractor	$\log(\text{Length of Side})$	$\log(\text{Area of Attractor})$
1.000	1.000	0.000	0.000
0.500	1.000	-0.301	0.000
0.250	0.812	-0.602	-0.090
0.125	0.391	-0.903	-0.408
\vdots	\vdots	\vdots	\vdots

Now construct a table as shown in Table [1] and plot the log of the length of one side of the box versus the log of the area of the attractor, then look for a linear region in the resulting curve. The slope of that linear region is the fractal dimension of the time series in an embedding dimension of 2. Now repeat the process in an embedding dimension of 3, 4, and so on. If the process generating the time series has a stable fractal dimension, the box counting dimension should asymptote to a fixed value as the embedding dimension is increased. The fractal dimension of white noise is in theory infinite, so the box counting dimension of a time series generated from white noise should not asymptote to a fixed value as the embedding dimension is increased.

A wide variety of methods have been proposed for estimating the fractal dimension of time series data. The most widely used of these was proposed independently by

Grassberger and Procaccia (1983a, 1983b) and Young (1982), and has since become commonly known as the Grassberger–Procaccia method. This method takes a time series $X = \{x_1, x_2, x_3, \dots, x_N\}$ which is time–delay embedded into a state space ξ and estimates the correlation dimension using a function called the *correlation integral*

$$C(l) = \lim_{N \rightarrow \infty} \frac{1}{N^2} P_l \quad (10)$$

where P_l is the number of pairs of state space vectors (ξ_i, ξ_j) , $i \neq j$ that are separated by a distance $\|\xi_i - \xi_j\| < l$. This reduces to

$$C(l) = \frac{2}{N(N-1)} \sum_{i \neq j}^N \Theta(l - \|\xi_i - \xi_j\|), \quad (11)$$

Where Θ is the Heavyside function. The Heavyside function is a simple function which maps one condition onto the value 1 and the complementary condition onto the value 0. The correlation dimension can then be calculated as

$$D_C = \lim_{l \rightarrow \infty} \frac{-\log C(l)}{\log l}. \quad (12)$$

A number of authors have extended and improved the efficiency of the Grassberger–Procaccia method of calculating fractal dimensions (Dvořák & Klaschka, 1990; Prichard, 1994). We recommend the method of Liebovitch and Toth (1989) for calculating the box–counting, correlation and information dimensions. Liebovitch and Toth’s method

is much faster than the Grassberger–Procaccia method and provides similar accuracy, although the Grassberger–Procaccia method tends to systematically underestimate the true fractal dimension and Liebovitch and Toth’s method tends to systematically overestimate the true fractal dimension for small data sets. Thus Liebovitch and Toth’s method is conservative with respect to claims of low fractal dimension. Other types of fractal dimensions have been proposed (Water & Schram, 1988; Dubuc, Quiniou, Roques-Carmes, Tricot, & Zucker, 1989), but none has been shown to be as robust or as accurate as the three discussed here (Farrell, Passamante, & Hediger, 1995).

There are a number of caveats which should be mentioned with regards to calculating the fractal dimension of time series such as might be generated by a physiological or psychological process (Kantz & Schreiber, 1995). The first warning is that the maximum value of the fractal dimension which can be calculated by most of the methods is directly related to the length of the time series such that

$$N \geq 10^{D/2} \tag{13}$$

where N is the length of the time series and D is the fractal dimension (Ruelle, 1990). Thus the fractal dimension which would be calculated from a time series of 1000 points is $D = 6$ even if the time series was white noise. If a time series has even a small amount of additive noise mixed with a known nonlinear chaotic system, the

true fractal dimension becomes quickly obscured from the current estimation methods (Schreiber, 1993).

The calculation of a fractal dimension is dependent on an assumption of stationarity within the time series in question. Certainly means and linear trends should be removed before calculating a fractal dimension. But there are more subtle types of stationarity problems which have to do with changes in the attractor due to perturbations from the environment. At this point there are no fool-proof methods for testing for or correcting this type of stationarity problem, although one solution is to calculate a running fractal dimension using a moving window along a longer time series.

2.5 Surrogate Data Tests for Nonlinearity

Given an arbitrary time series generated by an unknown process, it is important to test whether the time series needs to be modeled as a deterministic system and given determinism, whether a linear deterministic system is sufficient to model the measured data or if a nonlinear analysis is required. If there is no evidence of deterministic behavior, there is little point in pursuing a dynamical systems model; a correlational model will fit the measured data just as well. If there is no evidence of nonlinearity, then there is little point in pursuing a nonlinear dynamical systems model when a

linear model will fit the measured data just as well. In almost every case, the simpler model is to be preferred over the more complex.

Making these determinations requires some sort of test of significance. Normal statistical theory will not help in this case, since the calculation of standard errors requires some model of the process which generated the data and in this case the model is not known. A variant on the bootstrap method (Efron, 1979b, 1979a) for empirically determining the distribution of a statistic has been used in order to overcome this problem (Horn, 1965). Theiler, Eubank, Longtin, and Galdrikian (1992) have called this the method of *surrogate data* and it was independently proposed by Kennel and Isabelle (1992). The basic idea is to generate a population of null hypothesis data sets (surrogates) appropriate to the test of interest and then use the distribution of some nonlinear invariant (such as the fractal dimension of a time delay embedding) of these surrogates to estimate a confidence interval around the mean of the invariant. Then if the nonlinear invariant of the measured data lies outside the confidence interval of the surrogates, the null hypothesis is rejected.

Constructing the surrogate data sets can take many forms and will naturally vary depending on the particular null hypothesis that one desires to test. The first class of null hypotheses that need to be tested is that there is no dynamical process operating in the data. One possible surrogate data set considers the null hypothesis that the data are independent, identically Gaussian distributed random variables. A

pseudorandom number generator can be used to generate a sample of surrogate time series that each have the same mean and variance as the measured time series.

Another variant of this null hypothesis of no dynamical system preserves the amplitude distribution of the original signal, but destroys the time relations between the samples. In this case, the easiest method to generate a sample of surrogate time series is to shuffle the ordering of the samples in the measured time series. Each surrogate so generated will have all of its time dependencies removed, but the original amplitude distribution of the measured time series will be preserved.

The second class of null hypotheses are that the data were generated by a linear stochastic system. One method of generating a surrogate data set for this null hypothesis is to fit an autoregressive model to the data and then generate a sample of surrogate time series by iterating that model. There is a problem with this method: there are many possible autoregressive models which would need to be fit in order to find the best fitting alternative. A clever way around this problem was suggested by Osborne, Kirwin, Provenzale, and Bergamansco (1986). The Fourier transform of the time series is applied, a uniform random number between 0 and 2π is added to the phase spectrum of the Fourier series, and then the inverse Fourier transform is applied. The effect of this method is to generate a surrogate which shuffles the time ordering of the data while preserving the linear autocorrelations in the time series. The resulting surrogate fits the null hypothesis that the time series is a linear

stochastic process (colored noise).

A refinement of this method involves iteratively phase-randomizing the Fourier series and then correcting the resulting amplitude distribution for minor anomalies introduced by the phase-randomization process (Schreiber & Schmitz, 1996). This process is called *polished surrogates* and is the method which will be used to generate the linear stochastic null hypotheses for the present work.

Surrogate data methods have begun to be used in physiological applications (Schiff & Chang, 1992; Collins & De Luca, 1993, 1994) and have been recently extended to the multivariate case by Prichard and Theiler (1994) and Paluš (1995). In this case the surrogates must mimic not only the autocorrelations within each time series, but also all of the cross correlations between the series. This method simply adds the same uniform random number between 0 and 2π to the phase spectrum of each Fourier series. Prichard and Theiler cite an extension of the Weiner-Khintchine theorem as proof that this method preserves the desired correlations.

A cautionary note about null hypothesis testing is always in order. It is wise to remember that the null hypothesis can only be rejected, it cannot be accepted. Thus the surrogate data methods cannot rule out the presence of a nonlinear dynamical system as the generating process for a measured time series. If the null hypothesis cannot be rejected, then the conclusion is only that the measured time series does not

support the conclusion of a (nonlinear) dynamical system. This may be because the underlying process is truly not a (nonlinear) dynamical system, or it could be because the surrogate data test lacks sufficient power in the measured time series. Since there is no model for the standard errors of the test, one cannot properly calculate the power of these surrogate data tests.

2.6 Nonlinear Prediction and Lyapunov Exponents

In most cases, it is not enough to just provide descriptive statistics about a time series whether these are linear invariants such as mean and variance, or nonlinear invariants such as fractal dimension. Instead, one wants to be able to predict the future state of a system given the past state. Predictions come in two basic categories: (1) predictions which do not require a model (sometimes called nonparametric prediction in linear statistics), and (2) predictions in which a model is proposed and fit to the data. The second category involving linear and nonlinear dynamical modeling will be dealt with in detail in Section [3].

Nonlinear prediction (sometimes called nonlinear forecasting) techniques do not require an a priori model for the dynamical process, but rather attempt to make local predictions in state space using simple linear or polynomial least squares fitting. Nonlinear prediction can be used as a forecasting technique when only the outcome is of

interest and there is no desire to understand the underlying dynamical process. This is rarely the case in psychological or physiological inquiry. On the other hand, nonlinear prediction techniques can produce descriptive invariants that provide information about the predictability of a time series given a set of prior conditions. This provides a similar measure to that of mutual information, but whereas mutual information measures nonlinear dependency, nonlinear prediction measures predictability. This distinction is subtle and may not be actual. It may turn out that these two methods result in calculating the same invariants through different computational algorithms.

The nonlinear prediction methods are based on techniques suggested independently by Farmer and Sidorowich (1987) and Casdagli (1989), in which near neighbors in state space are used to generate a prediction for a selected point. Figure [13] illustrates the idea behind nonlinear prediction as follows. Given a point $x_{(t)}$ in state space, find the k near neighbors $\{x_{(t)}^{(1)}, x_{(t)}^{(2)}, \dots, x_{(t)}^{(k)}\}$ defined as those points within ϵ of $x_{(t)}$ which do not belong to the trajectory of $x_{(t)}$. Find the points at the next i time steps $(t + 1, t + 2, \dots, t + i)$ for each of the neighbor points $\{x_{(t+i)}^{(1)}, x_{(t+i)}^{(2)}, \dots, x_{(t+i)}^{(k)}\}$. Use these points to construct a best fitting trajectory through the target points $\{x_{(t)}, x_{(t+1)}, \dots, x_{(t+i-1)}\}$ and use this best trajectory to predict the position of $x_{(t+i)}$ in state space.

A variety of methods have been proposed to find best fitting trajectories, but the simple method of a least square fit polynomial seems to do as well as other more com-

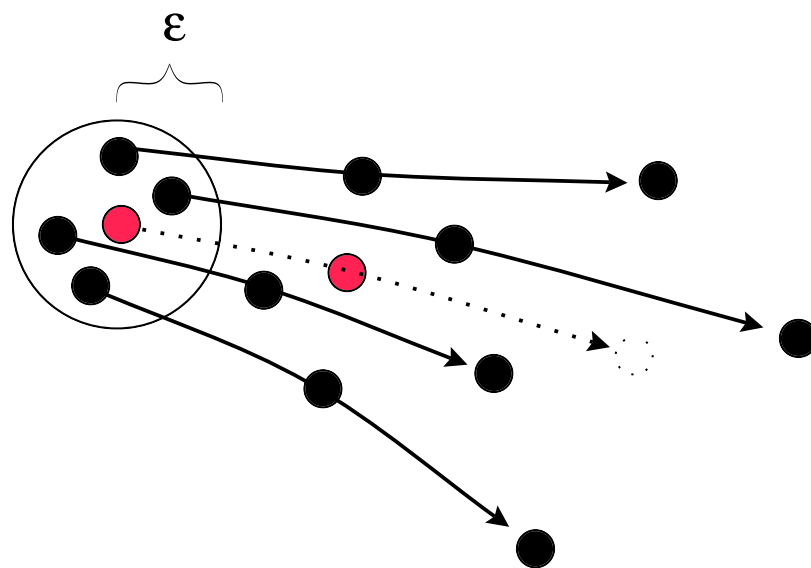


Figure 13: A geometric depiction of local nonlinear prediction methods. As a time series evolves, points which were near to each other in state space are expected to remain near to each other over short periods of time. The rate at which they diverge along each axis in state space is related to the Lyapunov exponents. Nonlinear prediction methods use the behavior of near neighbors in state space to predict the future evolution of a target point. See the text for a more detailed explanation.

plicated methods such as radial basis functions and neural networks. In fact, Farmer and Sidorowich (1987) suggest that a first order (linear) polynomial works quite well. Recently Abarbanel, Carroll, Pecora, Sidorowich, and Tsimring (1994) have published work which indicates that there can be a significant improvement in the predictions if a quadratic trajectory is fit. Certainly, the choice of polynomial degree will depend on the data at hand; the short and noisy time series available to psychologists will probably require the stability imposed by linear trajectory estimation.

As mentioned earlier, prediction of the next point in a time series is not the only reason that one might use nonlinear prediction techniques. It may be very useful to know the predictability of a time series either in general over the entire state space, or locally within regions of interest on the attractor. Predictability can be thought of as the rate at which near neighbors in a region of state space diverge (or converge) as the dynamical system is iterated forward in time. If one fits an exponential to each of the principal axes along which a neighborhood of points is diverging (or converging) the resulting exponents are the Lyapunov exponents (Lyapunov, 1949) of the system. This method for determining the largest Lyapunov exponent was originally suggested by Wolf, Swift, Swinney, and Vastano (1985) and has become a relatively common procedure since that time.

A similar algorithm was devised by Eckmann, Kamphorst, Ruelle, and Ciliberto (1986) which can be used to estimate the second and sometimes third as well as

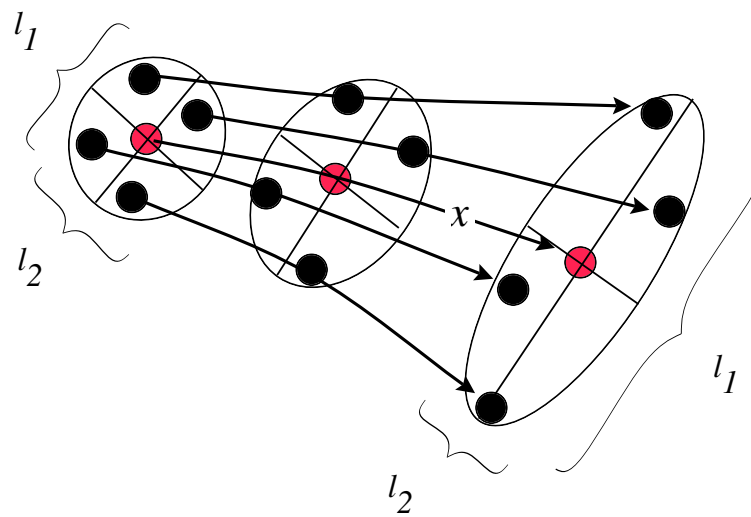


Figure 14: A geometric depiction of local nonlinear prediction methods for estimating the Lyapunov exponents from a time series. Ellipsoidal neighborhoods of points around a so-called “fiducial trajectory” (labeled x in the figure) change shape as the time series evolves. An exponential curve fit to the change in the size of an ellipsoidal neighborhood along the axis of greatest change (labeled l_1 in the figure) yields an exponent known as the largest Lyapunov exponent. Other Lyapunov exponents may be estimated along the other axes of the neighborhood. Note that as the length of l_1 increases, the length of l_2 decreases. This is commonly the case for chaotic systems.

the largest Lyapunov exponent directly from a time series (A succinct comparison of the two methods appears in Vastano & Kostelich, 1986). Further refinements of these methods have been proposed by Zeng, Eykholt, and Pielke (1993), Brown, Bryant, and Abarbanel (1991), and Rosenstein, Collins, and De Luca (1993). Each of these methods relies on the basic methods of nonlinear prediction to construct a set of prediction trajectories and estimate the expansion of the set as time is iterated forward in the system. This is equivalent to an estimate of the prediction error in nonlinear prediction, although it allows different axes to have different prediction error values.

Lyapunov exponents are important because they provide a well defined signature for the existence of chaos within a time series. If the largest Lyapunov exponent in a time series is positive and the area of its attractor is bounded, then nearby trajectories diverge at an exponential rate and thus the system exhibits sensitive dependence on initial conditions (Abarbanel, Brown, & Kennel, 1991). However, in time series which are likely to be available to psychologists, there is little chance that Lyapunov exponents can be accurately measured. Current methods require long ($N > 5000$) time series and low amounts of measurement noise in order to obtain even the largest exponent with reasonable reliability (Abarbanel et al., 1993).

It should be noted that if a system is driven by a nonlinear dynamical process, the predictability will generally not be the same forward in time as it will be backward in

time. This is not the case for linear stochastic systems. This asymmetry of prediction has been advanced as one method for testing for nonlinearity (Sugihara & May, 1990) and can be used with the surrogate data methods described in Section [2.5] to test for the presence of nonlinear dynamics.

The nonlinear prediction methods can be generalized to the multivariate case by using a multivariate embedding rather than an time–delay embedding of a single variable. Hybrids which use time–delay embedding of each of several variables are also possible. The questions which are of interest when performing a multivariate time–delay nonlinear prediction concern limits of predictability across and within variables. If the coupling between two chaotic variables is strong, the cross prediction may be stronger than each within–variable prediction. Relative strengths of coupling in response to varying stimuli can be estimated from these local nonlinear prediction models, making them particularly interesting when a system is perturbed in a measurable way over time.

2.7 Nonlinear Noise Reduction

Techniques similar to those used in nonlinear prediction have proven to be remarkably effective in identifying and removing additive error in simulated nonlinear dynamical systems. These processes also have roots in the nonlinear filtering techniques which

were proposed more than 30 years ago by Gabor, Wilby, and Woodcock (1960). Noise reduction based on nonlinear prediction was proposed by Hammel (1990) who suggested that there are stable and unstable directions along the manifold of a chaotic attractor in time–delay state space. These are equivalent to the directions labeled l_1 (unstable) and l_2 (stable) in Figure [14]. Hammel noticed that if the system was iterated forward in time, the predictability was high along the stable axes and if the system was iterated backward in time the predictability was high along the unstable axes. The details of Hammel’s algorithm have been superseded by more recent advances, but the critical observation that the dynamics are not symmetric with respect to time along different directions within the state space of a nonlinear system has proven to be central to many attempts at nonlinear noise reduction (Abarbanel et al., 1993).

Schreiber and Grassberger (1991) propose a method which uses a local linear filter and an embedding which explicitly incorporates information about future evolution of the system so that no iteration forward and backward in time is necessary. This method was generalized to the multivariate case and shown to be superior to univariate nonlinear noise reduction (Hegger & Schreiber, 1992). Kostelich (1992) proposes a method which focuses on the dynamics near recurrent points in a time series. This method requires a relatively high degree of oversampling and is unlikely to be of benefit in psychological or short physiological time series.

There is much work still proceeding in the area of nonlinear noise reduction. Two recent overviews make numerical comparisons between the extant algorithms and suggest modifications (Grassberger, Hegger, Kantz, Schaffrath, & Schreiber, 1993; Schreiber & Kantz, 1995). The modified method attempts to make orthogonal projections onto the manifold of the attractor for the time series. This method appears to be the most successful to date.

2.8 Nonlinear Dependency Matrices

One of the strengths of linear systems analysis is that there are well known methods for decomposing and fitting a model to the structure of covariance between the variables in a multivariate system. It seems reasonable that some variant of these powerful techniques could be applied to nonlinear dependency if a suitable analog to the covariance matrix could be worked out. In this section I propose three possible nonlinear substitutes for the covariance matrix, two based on mutual information and one based on nonlinear prediction.

Consider a dynamical system which has been measured by three simultaneous instruments at N equally spaced intervals, producing three time series

$$X = \{x_{(t)}, x_{(t+1)}, x_{(t+2)}, \dots, x_{(t+N)}\} \quad (14)$$

$$Y = \{y_{(t)}, y_{(t+1)}, y_{(t+2)}, \dots, y_{(t+N)}\} \quad (15)$$

$$Z = \{z_{(t)}, z_{(t+1)}, z_{(t+2)}, \dots, z_{(t+N)}\}. \quad (16)$$

In order to construct an embedding for these time series, a single time delay τ must be chosen which will be adequate for spreading out each of the attractors for the internal structure of X , Y and Z , while also spreading out the attractor that represents each of the cross-dependencies between the three series. For this purpose, I propose calculating the average mutual information versus time delay within each time series and across each pair of time series. The time delay to be used will be one which most nearly satisfies the following four criteria.

1. The time delay must not be equal to half-integer multiples of strong periodic structure in either the within or between series average mutual information. A strong periodic structure can be identified in a plot of the average mutual information versus time delay by a large peak which repeats at a fixed number of time-delay units.
2. The time delay should not be so short as to be on the first steep descent of the average mutual information versus time delay curve. This will help prevent linear autocorrelation from masking nonlinear structure within the time series.
3. The time delay should not be so long that there is no remaining mutual infor-

mation between time points. This criteria may not be sensible for all pairs. If some pairs quickly deteriorate to near zero mutual information, it may well be that there is no nonlinear structure between them. This is a property of the system under study and should not be ignored. On the other hand, the linear effects between one pair of variables may have a time constant longer than the nonlinear effects between a second pair of variables. In this case, the question becomes substantive. Which variables are important to the theory will drive the choice of which cross dependencies will be analyzed effectively linearly and which will be analyzed nonlinearly.

4. Finally, the time delay should be as close as possible to the first minimum of the average mutual information curve for all variables. This criterion is a strong one and will not generally be able to be satisfied, but if it can be satisfied it suggests that there is a unifying time constant which is common to all of the variables.

Once a time delay τ has been chosen for the embedding, a dimension for the embedding must be determined. I propose to use the method of false nearest neighbors within each time series, using the same time delay τ for each time series. The maximum dimension d found to be required to embed any one of the time series will be used as the embedding dimension for all of the time series.

Next, d -dimensional time delay embeddings $\mathbf{X}^{(d)}$, $\mathbf{Y}^{(d)}$, and $\mathbf{Z}^{(d)}$ are constructed for each of the time series X , Y and Z respectively

$$\mathbf{X}^{(d)} = \begin{bmatrix} x_{(t+0)} & x_{(t+0)+\tau} & \cdots & x_{(t+0)+d\tau} \\ x_{(t+1)} & x_{(t+1)+\tau} & \cdots & x_{(t+1)+d\tau} \\ x_{(t+2)} & x_{(t+2)+\tau} & \cdots & x_{(t+2)+d\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(t+N)-d\tau} & x_{(t+N)-(d-1)\tau} & \cdots & x_{(t+N)} \end{bmatrix}, \quad (17)$$

$$\mathbf{Y}^{(d)} = \begin{bmatrix} y_{(t+0)} & y_{(t+0)+\tau} & \cdots & y_{(t+0)+d\tau} \\ y_{(t+1)} & y_{(t+1)+\tau} & \cdots & y_{(t+1)+d\tau} \\ y_{(t+2)} & y_{(t+2)+\tau} & \cdots & y_{(t+2)+d\tau} \\ \vdots & \vdots & \ddots & \vdots \\ y_{(t+N)-d\tau} & y_{(t+N)-(d-1)\tau} & \cdots & y_{(t+N)} \end{bmatrix}, \text{ and} \quad (18)$$

$$\mathbf{Z}^{(d)} = \begin{bmatrix} z_{(t+0)} & z_{(t+0)+\tau} & \cdots & z_{(t+0)+d\tau} \\ z_{(t+1)} & z_{(t+1)+\tau} & \cdots & z_{(t+1)+d\tau} \\ z_{(t+2)} & z_{(t+2)+\tau} & \cdots & z_{(t+2)+d\tau} \\ \vdots & \vdots & \ddots & \vdots \\ z_{(t+N)-d\tau} & z_{(t+N)-(d-1)\tau} & \cdots & z_{(t+N)} \end{bmatrix}. \quad (19)$$

Finally, I propose to use the three embedding matrices $\mathbf{X}^{(d)}$, $\mathbf{Y}^{(d)}$, and $\mathbf{Z}^{(d)}$ to calculate two matrices, each of which could be considered to be a nonlinear analog to a covariance matrix. The first matrix will be constructed from all of the pairwise average mutual information values between embedded state vectors and τ delayed time series elements within and between the series. I propose calling this matrix of pairwise average mutual information coefficients the *nonlinear dependency matrix*. The calculation of each coefficient will be adapted from the definition of bivariate mutual information in Equation [5] as follows. Consider as an example the calculation

of average mutual information between the embeddings $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$

$$I_{XY}^{(d)}(\tau) = \sum_{i=1}^{N-\tau} p(\mathbf{X}_{(i)}^{(d)}, Y_{(i+\tau)}) \log_2 \left[\frac{p(\mathbf{X}_{(i)}^{(d)}, Y_{(i+\tau)})}{p(\mathbf{X}_{(i)}^{(d)})p(Y_{(i+\tau)})} \right], \quad (20)$$

where τ is the time delay used to construct the embeddings, $\mathbf{X}_{(i)}^{(d)}$ is the i th row of the d -dimensional embedding matrix $\mathbf{X}^{(d)}$, and $Y_{(i+\tau)}$ is the $(i + \tau)$ th element of the time series Y . The calculation of $I_{XY}^{(d)}(\tau)$ is made in exactly the same fashion, but with the time series values being retrieved from X .

The second matrix will be constructed from all of the pairwise average mutual information values between embedded state vectors and time-delayed embedded state vectors within and between the embeddings. I propose calling this matrix of pairwise average mutual information coefficients the *nonlinear state dependency matrix*. The calculation of each coefficient will be adapted from the definition of bivariate mutual information in Equation [5] as follows. Consider as an example the calculation of average mutual information between the embeddings $\mathbf{X}^{(d)}$ and $\mathbf{Y}^{(d)}$

$$I_{XY}^{(d)}(\tau) = \sum_{i=1}^{N-\tau} p(\mathbf{X}_{(i)}^{(d)}, \mathbf{Y}_{(i+\tau)}^{(d)}) \log_2 \left[\frac{p(\mathbf{X}_{(i)}^{(d)}, \mathbf{Y}_{(i+\tau)}^{(d)})}{p(\mathbf{X}_{(i)}^{(d)})p(\mathbf{Y}_{(i+\tau)}^{(d)})} \right], \quad (21)$$

where τ is the time delay used to construct the embeddings, $\mathbf{X}_{(i)}^{(d)}$ is the i th row of the d -dimensional embedding matrix $\mathbf{X}^{(d)}$, and $\mathbf{Y}_{(i+\tau)}^{(d)}$ is the $(i + \tau)$ th row of the d -dimensional embedding matrix $\mathbf{Y}^{(d)}$. The calculation of $I_{XY}^{(d)}(\tau)$ is made in exactly

the same fashion, but with both values being retrieved from the same embedding matrix $\mathbf{X}^{(d)}$.

One important difference between the nonlinear dependency matrices and the covariance matrix is that the nonlinear dependency matrices are not necessarily symmetric. The dependency of X on Y is not necessarily the same as the dependency of Y on X . For example, suppose every time a state $\mathbf{X}_{(q)}^{(3)} = \{2, 3, 4\}$ occurs then τ delay units later the value $Y_{(q+\tau)} = 5$ always occurs. If the state $\{2, 3, 4\}$ occurs with some high frequency, at random intervals in $\mathbf{X}^{(3)}$, then the values in Y will have a considerable dependency on $\mathbf{X}^{(3)}$. However, since the occurrence of the state $\{2, 3, 4\}$ in $\mathbf{X}^{(3)}$ is random, this dependency says nothing about whether the values in X are dependent on the states in $\mathbf{Y}^{(3)}$. This asymmetry is one consequence of the nonlinear and time-delayed nature of the calculations which form the nonlinear dependency matrices.

A third matrix could be constructed from all of the pairwise nonlinear predictability calculations within and between the embeddings. I propose calling such a matrix of pairwise nonlinear predictabilities the *nonlinear predictability matrix*. The calculation of this matrix would follow the calculation of divergence of trajectories outlined in Section [2.6]. The matrix would be formed in the same way as the nonlinear dependency matrices, but the calculation involved would be an approximation of the divergence of trajectories as outlined in Brown et al. (1991) and Rosenstein et al.

(1993), but retaining only the major axis of the expansion for each pairwise calculation.

3 Introduction to Dynamical Modeling

At some point, the description and characterization of experimental data becomes insufficient to gain understanding about the potential mechanisms operating within the systems that generated the data. One wishes to test a theory and see whether it can be rejected, or whether it performs better or worse than some other theory. Modeling involves describing a precise prediction for some invariants of the behavior of a system and then some method for determining whether those predictions are realized within the data derived from the results of observations of the behavior of the system in question.

The models discussed here will fall into two basic categories: linear dynamical models and nonlinear dynamical models. There are two main differences between these two categories: (1) the linear models will analyze covariance matrices while the nonlinear models will analyze nonlinear dependency matrices, and (2) the linear models estimate derivatives (and thereby states) after the calculation of the covariances whereas the nonlinear models use state space embedding techniques prior to the calculation of the dependencies.

3.1 Linear Dynamical Modeling

An example linear dynamical structural equation model is shown in the path diagram in Figure [15]. The path diagrams illustrated here will all use the conventions of RAM notation for representing structural equation models (McArdle & McDonald, 1984; McArdle & Boker, 1990).

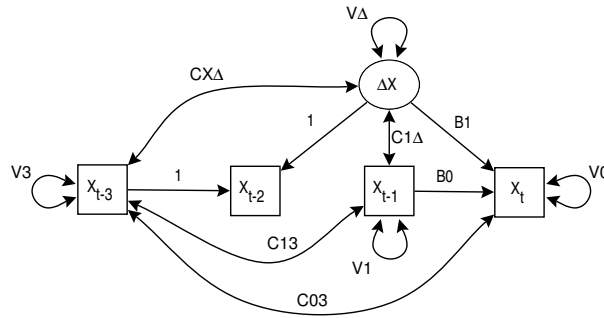


Figure 15: An example of a univariate first order linear difference equation as represented by a structural equation model.

The model shown in Figure [15] meets the requirement that any dynamical systems model must in some sense attempt to fit the phase space or state space of the time series. This model estimates the first difference $\Delta x_{t-2} = x_{t-3} - x_{t-2} + e_{\Delta x}$ and uses the first difference and previous value of x_t to predict x_t

$$x_t = B0x_{t-1} + B1\Delta x_{t-2} + V0 \tag{22}$$

Systems of these linear dynamical models may be estimated simultaneously as in

the example shown in Figure [16]. Here three time series are tested for structural relations in the coupling between the series. Multiple models may be fit and the goodness of fit of each model may be compared with the sample of models. The outcome variables at the bottom of the figure are predicted by some restricted set of the previous time points and first differences. This type of model may fit very well when the underlying dynamical system is a linear stochastic system and the time series is sufficiently oversampled that the autocorrelations and cross-correlations are high. In such cases, this model may not necessarily fit better than a standard autoregression / cross-regression model, but the interpretation of the parameters is more straightforward from a dynamical systems perspective.

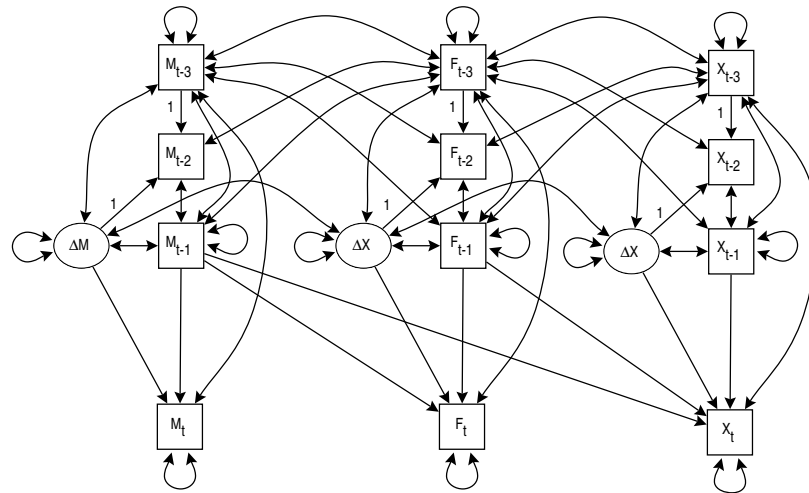


Figure 16: An example of three first order linear difference equations with a restricted pattern of cross couplings as represented by a structural equation model.

When the system is oversampled and exhibits more complex behavior than can be

accounted for with first differences, a higher order factor difference model may be used to estimate the second differences. When the data are sufficiently oversampled, the model shown in Figure [17] may be able to estimate the parameters for the differential form of the dynamical system. Simulations of univariate dynamical systems from differential equations need to be performed in order to test this claim.

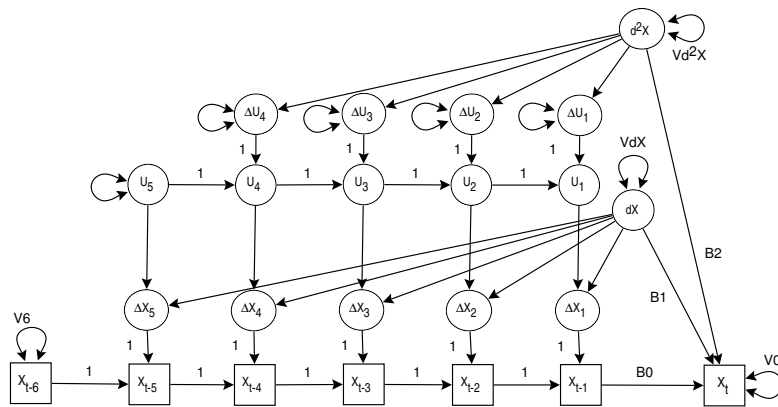


Figure 17: An example of a second order linear difference equation as represented by a structural equation factor model.

The linear dynamical modeling methods have the advantages that they can be easily fit using well established software and that they are interpretable within the framework of linear stochastic systems. However, if nonlinear surrogate techniques suggest that there are nonlinear dynamics within any one of a set of multivariate time series, then the linear modeling techniques may give incorrect parameter estimates. The existence of strong coupling may be masked, or coupling could be overestimated by using linear techniques when the data are nonlinear. In these cases, I propose the use of nonlinear dynamical modeling.

3.2 Nonlinear Dynamical Modeling

Structural equation models which are fit to one of the forms of nonlinear dependency matrices as defined in Section [2.8] will be here referred to as *nonlinear dynamical models*. These models may take a number of forms depending on the requirements of the theory and data to which they may be applied.

The major differences between these models and the linear dynamical models described in the previous section are due to the differences between a covariance matrix and a nonlinear dependency matrix. There is nothing in a covariance matrix which inherently carries information about the dynamical state of the system on which it is calculated. The only dynamical relationships in a covariance matrix are those which occur as structural components between elements in the matrix. Furthermore, these dynamical relationships are only those which can be discerned after aggregation over the data; aggregation may destroy critical components of exactly the dynamical relationships that one is seeking to understand.

On the other hand, each element in a nonlinear dependency matrix carries information about the dynamical relationship between two variables. This is due to the fact that the statistical aggregation in a nonlinear dependency matrix occurs after the dynamics have been embedded from the time series. Thus, although aggregation may still blur the nonlinear dynamical relationships between the time series, the

“distance” between two variables in the nonlinear dependency matrix is one which reflects their co-occurring dynamic structure rather than merely the co-occurrence of values.

Thus dynamical modeling using nonlinear dependency matrices can be represented by much simpler models than the linear dynamical models discussed previously. For instance, a structural single factor model applied to a nonlinear dependency matrix expresses a hypothesis that there is a common factor of nonlinear dependence between the variables. Since the nonlinear dependency matrix is calculated using mutual information, I propose to call this class of nonlinear dynamical models *Mutual Information Modeling with Structural Equations* (MIMSE).

However, there is one complication which will require some algebraic changes to the normal structural equations. This complication arises because a nonlinear dependency matrix may be asymmetric. In McArdle’s RAM formulation, the expected covariance equation is

$$C = F(I - A)^{-1} S ((I - A)^{-1})' F' , \quad (23)$$

where C is the expected covariance matrix, F is the filter matrix, A is the asymmetric parameter matrix, I is identity matrix, and S is the symmetric parameter matrix (McArdle & McDonald, 1984; McArdle & Boker, 1990). Since the S matrix is

symmetric, the expected covariance matrix is guaranteed to be symmetric.

In order to calculate an asymmetric matrix of nonlinear dependencies, I propose a relatively simple change to the S matrix which carries the values of the covariance parameters in the RAM expected covariance equation. This matrix will now need to potentially be asymmetric, and since it will now carry the dependency parameters, I propose to rename this matrix the D matrix. Thus the MIMSE calculation for the expected nonlinear dependency equation becomes

$$N = F(I - A)^{-1} D ((I - A)^{-1})' F' , \quad (24)$$

where N is the expected nonlinear dependency matrix, F is the filter matrix, A is the asymmetric parameter matrix, I is identity matrix, and D is the dependency parameter matrix. Such an expectation equation can be fit using standard structural equation modeling software such as Mx (Neale, 1994).

Due to the fact that the dependency matrix is asymmetric, a new path diagramming convention will need to be defined. Since the dependency parameter matrix is taking the place of the covariance parameter matrix, since many people think of the covariance elements in a path diagram to be “two-headed arrows”, and since the dependency parameters now are directional rather than non-directional, I propose that a dependency parameter be depicted as a “two-headed arrow” with both of its

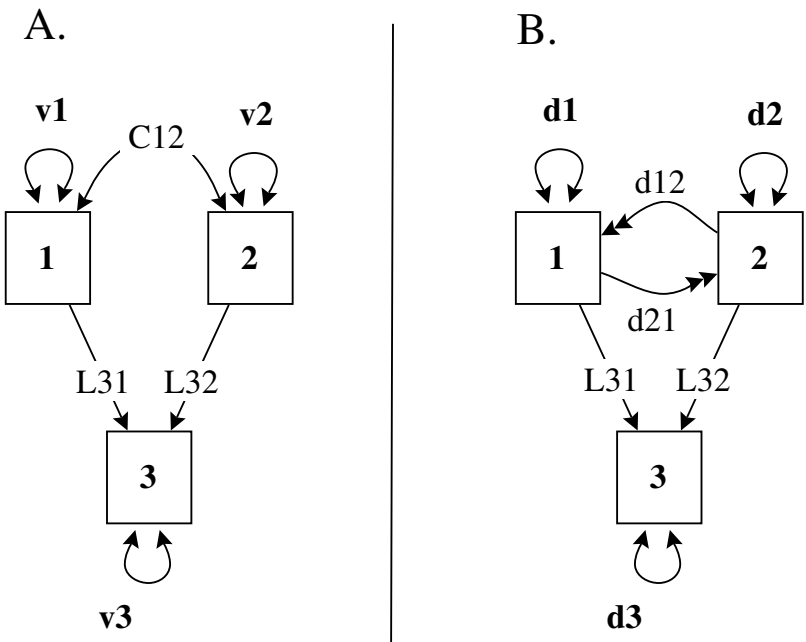


Figure 18: Two example regression models. (A) A two variable regression model fit to a covariance matrix in the standard RAM path diagram notation. (B) A two variable regression fit to a nonlinear dependency matrix using the MIMSE nonlinear path diagram modification to RAM notation. Note that where there was one covariance double headed arrow in (A) there are now two directed dependency double headed arrows in (B).

arrowheads at one end. Figure [18] shows the MIMSE modification to the RAM notation. It should be noted that all of the RAM tracing rules remain intact except that directionality of the dependency double-headed arrows must be taken into account during the tracing.

4 An Application to Developmental Time Series

Some of the techniques described in the previous sections were applied to an example data set taken from an experiment examining of the development of postural control in infants (Bertenthal, Rose, & Bai, 1997). These data were chosen since there are theoretical reasons why a dynamical system would be expected to be involved in their generation. These data do not comprise one or two very long time series as would typically be used in dynamical systems analysis as practiced in Physics. Instead we are confronted with many shorter time series which have been generated under specific experimental conditions: a case much more typical of the behavioral sciences. One of the main lessons to be learned from this application is that nonlinear techniques can be used even when the data are composed of many shorter time series if appropriate care is taken in their aggregation.

It should be noted that these data are not longitudinal in nature so that an assumption must be explicitly made that observed changes in aggregated invariants

across age categories in some sense reflect the developmental change within an individual. This assumption should especially be kept in mind when cross-sectional data are analyzed using dynamical systems methods. Here we are examining postural control; a system which involves dynamical processes on the time scale of tens of milliseconds. Recall that there may be developmental processes which are unfolding on the time scale of months and that these also may be able to be mapped to dynamical processes. By aggregating cross-sectional data we have lost the ability to make a strong model of the month-by-month developmental process as an evolving dynamical system. Therefore, this work will be making a statement about the development of the dynamics of postural control but will not be making a statement about the dynamics of the development of postural control.

4.1 The Development of Postural Control

Upright bipedal stance involves controlling musculature in such a way that balance can be maintained under varying environmental conditions. A control mechanism must exist which maps visual, auditory, vestibular, and proprioceptive perceptual input onto the appropriate muscle groups. This control mechanism must take into account both changes in the environment and changes in the state of the postural system in order to quickly and flexibly adapt to new conditions. For example, if the apparently solid rock on which one is standing were to suddenly begin to move, then

one's posture must respond immediately to this perturbation of the environment in order to remain standing.

Since there is a time delay between the perception of a change in the environment and the possible adaptive control which can be applied in order to respond to that change, it would be sensible for the control mechanism to attempt to anticipate the changes in the environment in order to make prospective adaptation that will result in a close temporal match between perception and action. This can be illustrated by thinking about how one catches a ball; the hand is extended so that the ball and the hand will arrive at the same point at the same time.

For these reasons we are inclined to think of postural control as a dynamical system; a system that must integrate input from several perceptual modalities in order to maintain an equilibrium. However, this equilibrium is not a fixed point; even when we attempt to stand motionless we sway slightly. This motion during quiet standing has been studied by Collins and De Luca (1994) who have concluded that the postural control system can be modeled as a linear stochastic system. Collins and De Luca (1993, 1995) also argue that posture utilizes two linear control mechanisms: an open-loop system over short time scales and a closed-loop system over longer time scales.

Vision and proprioception are two principal inputs to the postural control sys-

tem (Howard, 1986). This can be demonstrated using a *moving room* experimental paradigm (Lishman & Lee, 1973; Lee & Aronson, 1974). The moving room, or “Phantom Swing” as it was called in the 19th century, effectively dissociates the input from proprioception and vision. In this experimental paradigm, the subject stands in a room whose walls and ceiling are not fixed to the floor, but instead are mounted on wheels. When the experimenter moves the room, the subject receives visual input indicating self motion while the proprioceptive input indicates no self motion. A slowly oscillating room produces a significant swaying response in subjects, including infants (Bertenthal, 1990).

This work examines data from a moving room experiment performed on a group of infants (Bertenthal et al., 1997). These infants were selected to have ages which straddled the average age of onset of self-supported sitting: 6.5 months (Bayley, 1969). The data from this experiment holds interest for examination using dynamical systems methods since changes in the dynamics of the postural control mechanism can be studied simultaneously with the coupling between perception and action. By studying developmental changes in the coupling between visual perception and postural action we hope to better understand the nature of the postural control system.

4.2 Experimental Methods

Forty infants participated in the study, 10 in each of 4 age groups: 5, 7, 9 and 13 months. The moving room consisted of a $1.2\text{ m} \times 1.2\text{ m} \times 2.1\text{ m}$ open-ended enclosure as shown in Figure 19. The walls and ceiling of the enclosure were constructed of fiberboard covered with green and white vertically striped cloth and mounted on small rubber wheels that rolled on tracks fixed to the floor. The floor was padded and covered with white muslin cloth. Two fluorescent lights mounted at the top of the two side walls illuminated the moving room. A small window in the middle of the front wall of the moving room provided a view to a small electronically activated toy dog that was used to fix the gaze of the infant at the beginning of each trial. A potentiometer was attached to one wall of the moving room such that the position of the room could be measured by the voltage drop through the potentiometer. The position of the room was sampled at 50 Hz and converted to a digital time series using an 8 bit A/D converter thereby creating a time series $R = \{r_1, r_2, r_3, \dots, r_N\}$ representing the room movement.

A forceplate was set in the middle of the floor of the moving room and an infant's bicycle seat was mounted on the center of the forceplate. The forceplate consists of a rigid metal plate suspended from 4 pressure transducers as shown in Figure 20. The transducers were each sampled at 50 Hz synchronously with the room position

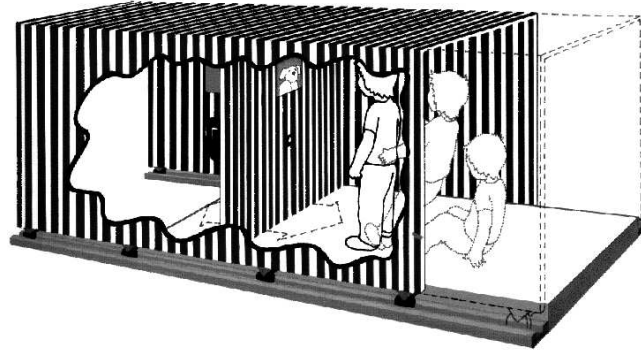


Figure 19: Schematic drawing of the moving room. The child portrayed inside the room is falling backward due to perceived optic flow produced by movement of the room. Note that the floor does not move; the subjective perception of self-motion is created by moving the walls and ceiling of the room together.

and converted to 4 digital signals with an 8 bit A/D converter. The four time series $P1, P2, P3$ and $P4$ from the four transducers were transformed to two center of pressure (COP) time series $X = \{x_1, x_2, x_3, \dots, x_N\}$ and $Y = \{y_1, y_2, y_3, \dots, y_N\}$ where

$$x_i = 2D_x \frac{(p1_i + p2_i) - (p3_i + p4_i)}{p1_i + p2_i + p3_i + p4_i}$$

and

$$y_i = 2D_y \frac{(p1_i + p4_i) - (p2_i + p3_i)}{p1_i + p2_i + p3_i + p4_i}.$$

An infant sat in the infant bicycle seat which in turn was sitting on the forceplate. The first trial followed a short time in which the infant was allowed to acclimate to the room. At the beginning of each trial, the toy was activated to direct the infant's

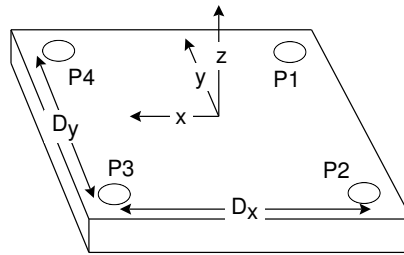


Figure 20: Schematic drawing of the forceplate. The transducers $P1$, $P2$, $P3$ and $P4$ were sampled at 50 Hz and these data transformed to two center of pressure time series X and Y along two orthogonal axes aligned with the edges of the forceplate.

attention to the front of the room. The walls then oscillated in one of six movement conditions for a period of approximately 12 seconds during which time the infant's center of pressure was measured. Between successive trials the infant was allowed a brief interval in which to restabilize to the motionless room.

During each trial, the force plate measured center of pressure along two axes: the fore-aft axis (X) aligned with the movement of the room, and the lateral axis (Y) orthogonal to the movement of the room. As the room began to move, the two axes of center of pressure and the unidirectional position of the room were simultaneously sampled at 50 Hz (20 msec sampling interval) for 10.25 seconds. Thus each trial generated three synchronized time series containing 512 samples: X , Y and R .

Infants were tested in a sequence of 12 trials, two trials for each of six room movement conditions as shown in Table 2. Four conditions consisted of sinusoidal

Table 2: The six frequency and amplitude room movement conditions. The baseline condition is labeled as 0 Hz frequency and 0 cm amplitude. “Multi” refers to a multiple frequency pseudorandom movement of the room.

	0 Hz	Multi	0.3 Hz	0.6 Hz
0 cm	X			
Multi		X		
9 cm			X	X
18 cm			X	X

movement with one of two amplitudes (9 cm or 18 cm) combined with one of two frequencies (.3 Hz or .6 Hz). The remaining two conditions were a motionless control condition and a pseudorandom low frequency movement pattern. Each session started with the motionless control trial followed by the moving room trials presented in random order and then followed by the motionless condition again as the last trial.

4.3 Analytic Methods

Four nonlinear techniques were applied in the analysis of these data: mutual information, surrogate data, false nearest neighbors and nonlinear dependencies. The raw time series were first centered: about their means. This is appropriate for these time series because the mean is dependent on exactly where the bicycle seat was set on the force plate, an artifact which is unrelated to the phenomenon under investigation. Each of the centered time series was then plotted and inspected in order to determine if there were transients at the beginning or end of the time series. These transients

may have occurred due to the sampling beginning or ending at an inappropriate time, starting before the infant was settled or ending after the infant was being picked up by its mother. There were no apparent large transients observed in these data, so no attempt was made to remove transients.

Before the nonlinear techniques were applied, the data was described using a linear technique: Fourier analysis. This technique tests whether there is some form of periodicity in the data, although it doesn't tell us whether these are linear or nonlinear periodicities. The Fourier technique also provides a convenient measure to see whether the nonlinear methods provide further insight into the data than standard linear techniques.

The first nonlinear technique to be applied was a surrogate data test for nonlinearity. If these data cannot be distinguished from linear surrogates then there is no point in pursuing further nonlinear analyses. A null hypothesis of a linear stochastic system was used and the resulting surrogate time series were tested against the real time series using mutual information. The results are presented in two different ways: first using individual trials in the hypothesis test and second using aggregates and testing 95% confidence intervals around the mean mutual information for the surrogate and real time series.

The second nonlinear technique was a mutual information analysis which was

used to determine an optimum time delay for creating an embedding of the time series. This technique helped the other techniques focus on the dynamics of the time series without becoming overwhelmed with the autocorrelation in the data due to the momentum of an infant's body.

The third technique used was a false nearest neighbors analysis. This analysis was used to ascertain if a qualitative change coincident with the onset of self-supported sitting might have changed the minimum required dimension for embedding the time series. The false nearest neighbors analysis was also undertaken to examine whether there is a developmental change in the proportion of noise present in time series.

Finally, a set of analyses was performed which examined the nonlinear dependencies both within the center of pressure time series and between the room position time series and the center of pressure time series. These analyses test whether there are significant changes in the cross dependencies with changes in experimental condition and will also look for developmental changes in the cross dependencies. The nonlinear dependency analyses have the potential for answering further questions such as what time delays are present in the entrainment of the infants' center of pressure to the room and whether there are developmental differences in these time delays.

5 Results

5.1 Raw Time Series

Figures [21] and [22] plot prototypical time series for the room movement and fore–aft sway for three trials from a single seven month–old infant. Figure [21] plots the room movement for the six different stimulus conditions and Figure [22] plots the corresponding fore–aft center of pressure for the infant during the same 10.25 second interval. In Figures [22]–B, [22]–C and [22]–D the movement of center of pressure for the infant bears little visible resemblance to the corresponding room movement, but Figures [22]–E and [22]–F exhibit a noticeable periodicity which appears similar to that of the room movement in Figure [21]–E and [21]–F.

5.2 Fourier Analysis

In order to quickly determine if the infants are responding to the movement of the room, the Fourier spectrum was calculated for the time series resulting from each trial. The Fourier spectra were aggregated for all trials within each condition. Since four of the moving room stimulus conditions represented sinusoidal oscillations, the Fourier spectrum of the moving room time series is represented by one peak at the frequency of the oscillation. It follows that if the infant’s behavior was entrained

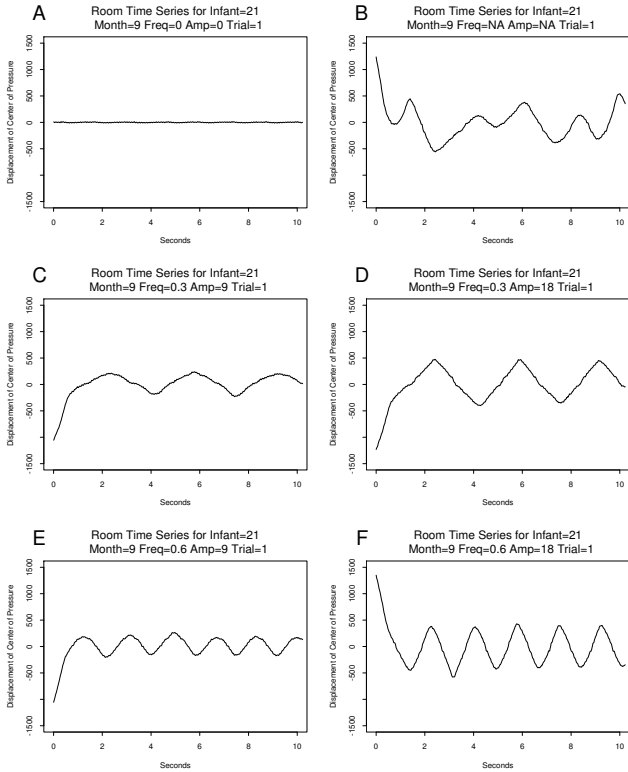


Figure 21: Example time series for the room movement from the first trial of all six experimental conditions for infant number 21.

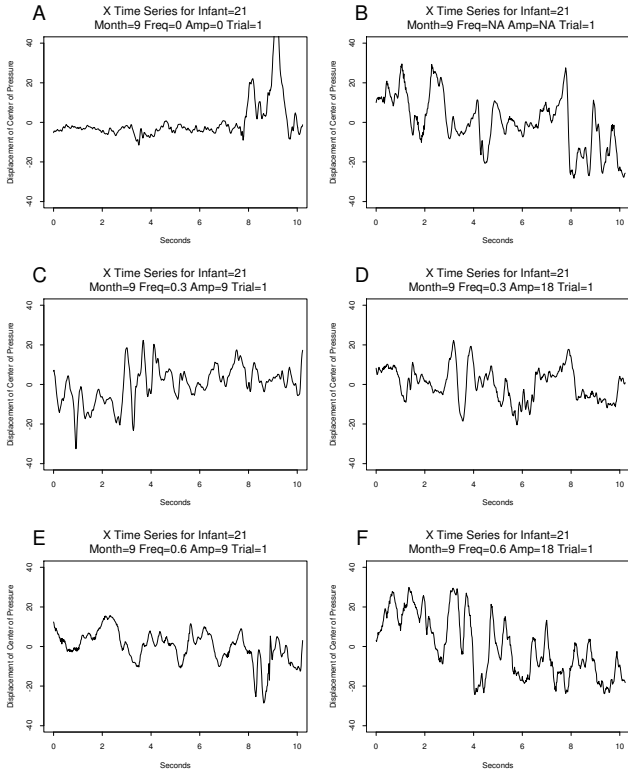


Figure 22: Example time series for the fore-aft center of pressure from the first trial of all six experimental conditions for infant number 21.

to the linear oscillation of the room, there should be a peak in the mean Fourier spectrum of the infant center of pressure at the corresponding frequency of the four sinusoidal stimulus conditions as shown in Figure [23].

Mean Fourier spectrum plots for the fore–aft center of pressure time series under each condition are shown in Figure [24]. Note that for the 0.6 Hz conditions there is a peak in the distribution corresponding to a frequency of 0.6 Hz. In a similar fashion in the 0.3 Hz, 18 cm condition there is a peak in the distribution at 0.3 Hz. It is evident by inspection that these data exhibit some form of coupling between the visual environment and postural control.

5.3 Surrogate Data Test for Nonlinearity

The time series from two of the experimental conditions were tested to see if a linear stochastic model would be sufficient to describe the center of pressure data. We chose to test the control condition where no room movement occurred and the 0.6 Hz condition in which maximum coupling between the moving room and the infant's center of pressure was expected to occur (Bertenthal et al., 1997). In this way we could examine both the question of whether there was evidence of nonlinearity in the center of pressure of a free sitting infant, and if so, whether there was a decrease in nonlinearity when the postural control system was coupled to a sinusoidally oscillating

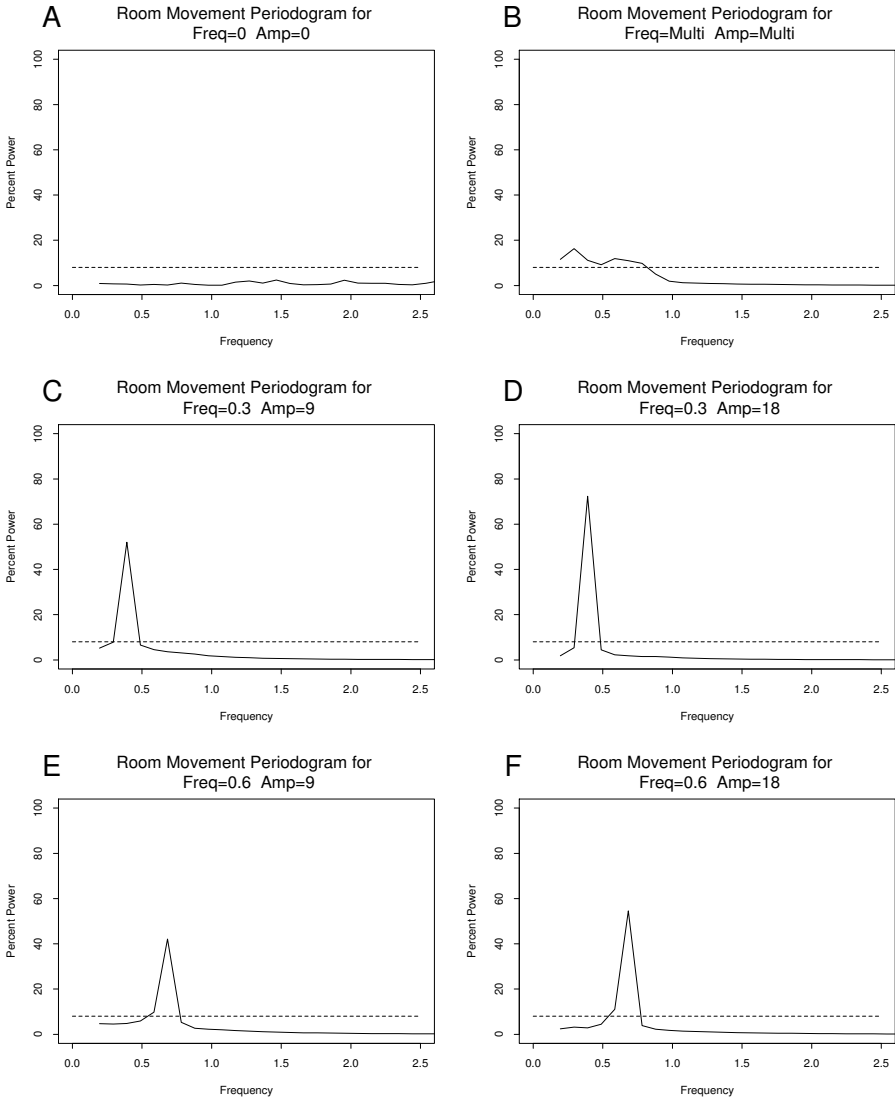


Figure 23: Mean FFT plots for the room movement over all trials and all infants for each experimental condition. The 95% confidence interval for distinguishing the frequency components from white noise is indicated by the dashed line in each graph.

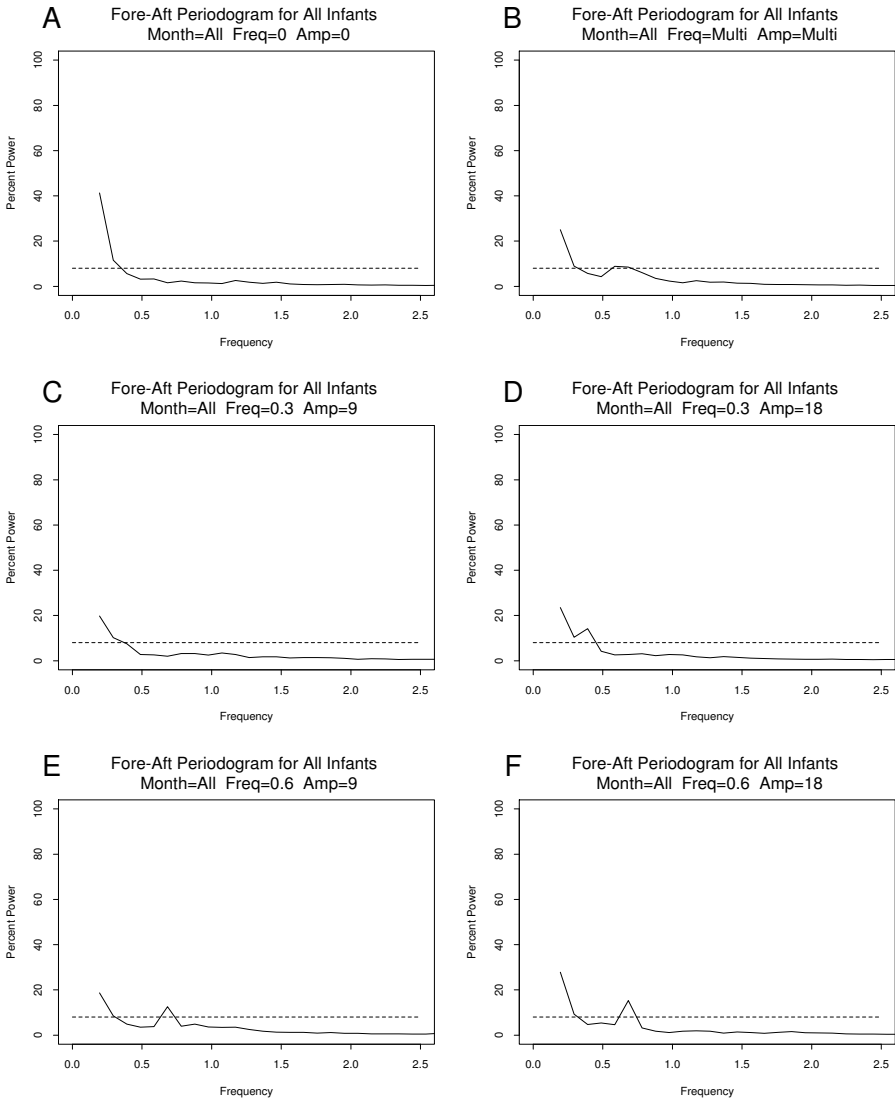


Figure 24: Mean FFT plots for for the fore-aft center of pressure over all trials and all infants for each experimental condition. The 95% confidence interval for distinguishing the frequency components from white noise is indicated by the dashed line in each graph.

visual stimulus.

For each time series within each selected experimental condition, twenty surrogate time series were generated using the *polished surrogates* method. This method involves iteratively phase-randomizing the Fourier series and then correcting the resulting amplitude distribution for minor anomalies introduced by the phase-randomization process (Schreiber & Schmitz, 1996). This process broadens the null hypothesis to include simple nonlinear rescalings of a Gaussian random process, whereas simply phase-randomizing the Fourier series includes the restriction that the linear process is Gaussian.

We then calculated the average mutual information $I(\tau)$ for a time-delay of $\tau = 200$ ms for each of the surrogate time series and for the source time series. If the mutual information of the source time series was larger than the largest surrogate mutual information, then the null hypothesis of no nonlinearity was rejected. If the null hypothesis is rejected then the hypothesis that a linear model is sufficient is rejected at the ($p = .05$) level.

Table 3 summarizes the results of the surrogate data tests for nonlinearity. Recall that at the ($p = .05$) level we should expect that 5% of the null hypothesis conditions should be rejected by chance alone. However, in the free sitting control condition the null hypothesis is rejected 83% of the time. We consider this to be strong evidence

of nonlinearity in these data.

Table 3: Summary of results of the surrogate data tests for nonlinearity. Each cell lists the percentage of time series for which the null hypothesis of a linear stochastic system was rejected. Thus the hypothesis of a linear stochastic model being sufficient for these data was rejected in 80% of the time series for the free sitting control condition with no room movement.

	0 Hz ($N = 80$)	0.6 Hz ($N = 160$)
All ages	83%	59%
5 Months	75%	65%
7 Months	85%	55%
9 Months	75%	60%
13 Months	95%	55%

Note that in the free sitting condition the evidence for nonlinearity is greatest in the oldest infants. It is also interesting to note that when the center of pressure of the infant is coupled to the 0.6 Hz sinusoidally oscillating room, the percentage of trials in which the null hypothesis is rejected is reduced to 59%. This reduction in the apparent nonlinearity of the center of pressure time series is expected since in this experimental condition the center of pressure is coupled to a linear oscillator.

A second method for testing the difference between the real time series and the surrogate time series is shown in Figure 25. A full mutual information curve was calculated for each real fore–aft time series in the control condition and the mean mutual information at each lag was calculated and is plotted in the upper solid line

in Figure 25. The dashed lines around this mean mutual information curve represent 95% confidence intervals for this mean curve. The lower solid line in Figure 25 plots the mean mutual information curve for the surrogates and the dashed lines around the lower curve are the 95% confidence intervals for the surrogate mean. From this figure, we can conclude that the null hypothesis of no nonlinearity can be rejected for all time-delays between 100 ms and 2000 ms.

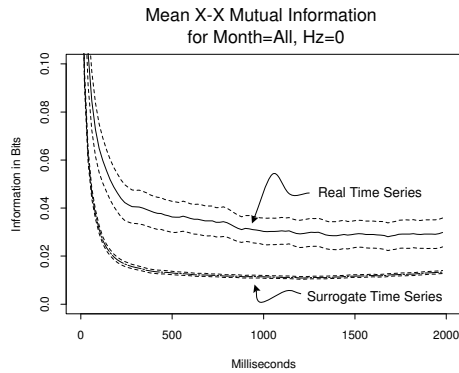


Figure 25: Mutual information curves for the real fore–aft time series and for the surrogate time series for all time delays up to 2000 ms was calculated for each trial and then the resulting mutual information curves were averaged by taking the mean for each time delay across all of the trials. The dashed lines represent 95% confidence intervals for the mean curve ($1.96 \times$ the standard error of the mean).

5.4 Mutual Information

Mutual information was calculated for each time series: the room position series, the fore–aft center of pressure series and the lateral center of pressure time series.

Figure 26 plots the grand mean mutual information curves for time delays up to 100

samples. These curves were constructed by calculating the mutual information for each trial and then averaging within time delay across all trials.

For instance, Figure 26–E plots the grand mean mutual information within the fore-aft center of pressure time series. For each trial ($N=480$), a mutual information curve was calculated. Then for each time delay, the mean value across all trials was calculated. These mean values were then plotted, thereby creating Figure 26–E.

In calculating grand mean mutual information curves, we wish to find the minimum lag that can be used to construct an embedding space without artificially compressing the space due to autocorrelation. The two graphs in Figures 26–E and 26–I provide an estimate of this minimum lag. The mutual information curves in these two graphs fall rapidly as the lag increases from 0 to approximately 100 ms. At delays longer than 200 ms the mutual information curves remain relatively flat. Thus according to this analysis, delays shorter than 100 ms may be contaminated with autocorrelation. Since each sample is 20 ms, we suggest that a time–delay embedding where $\tau < 5$ would be likely to result in embedding spaces which were artificially compressed. A time–delay embedding with $\tau > 10$ would be wasteful of the data. For further analysis, have decided to use a time–delay of $\tau = 5$ as the minimum delay warranted in these data.

Some other features can be noted in the graphs from Figure 26. For instance,

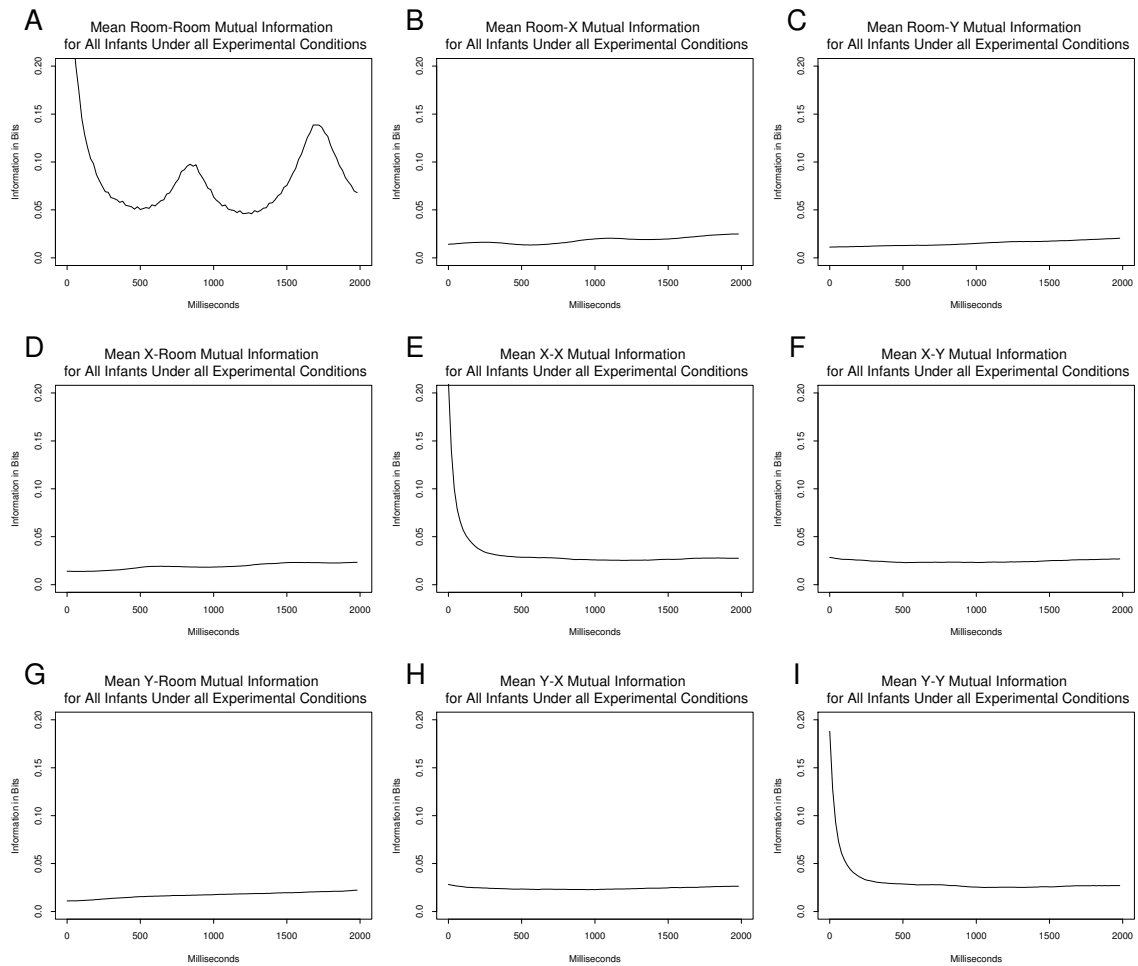


Figure 26: Mean mutual information calculated within and between the room position time series, the fore–aft time series and the lateral time series. A mutual information curve for all time delays up to 2000 ms was calculated for each trial and then the resulting mutual information curves were averaged by taking the mean for each time delay across all of the trials.

the periodicity of the moving room time series is clearly evident in the two peaks in Figure 26–A. Recall how in Figures 3 and 4 on page 21, when we chose just the wrong delay for embedding a sinusoid oscillation, the embedding space would turn out to be projected onto a line. These wrong choices of time delays map onto the two peaks in the Figure 26–A, one where the delay is half the wavelength of the sinusoid and one where the delay is equal to the wavelength of the sinusoid. The reason that one peak is higher than the other is due to the fact that we are averaging over two sinusoidal conditions, one of which has a frequency twice that of the other. Thus the taller peak is one full period of the faster frequency (0.6 Hz has a period of 1667 ms) and one half period of the slower frequency (0.3 Hz has a half period of 1667 ms).

Finally, it should be noted that none of the cross dependencies have a grand mean mutual information greater than 0.05 bits. Thus, when we look closer at the cross dependencies in subsequent sections we will plot mutual information in a range of 0.00 to 0.05 bits.

5.5 False Nearest Neighbors Analysis

A false nearest neighbors analysis was performed in order to address two questions. First, is there an observable change in the required dimension of an embedding space for the center of pressure time series coincident with the onset of self-supported

sitting? If such a dimensional change were observed, it would provide evidence that there may be a qualitative change in the postural control system at the time of the qualitative change in sitting behavior. Second, is there an observable developmental change in the amount of noise mixed with the postural control signal in the center of pressure time series? If this developmental change in noise mixture is observed, it would provide evidence that there is a *quantitative* change in the postural control system in which an existing mechanism improves its performance.

The false nearest neighbors analysis was applied to each infant's time series from all experimental conditions. The two trials from each infant were appended into a single time series, the algorithm was applied, and cases of neighbors including the boundary between the pair of trials were excluded from the pool of potential neighbors. Figure 27 plots the false nearest neighbor curves for every infant within each age category for the control stimulus condition.

Figures 27-A, 27-B, 27-C and 27-D are essentially identical for dimensions less than or equal to 3, falling from approximately 60% false nearest neighbors at embedding dimension 1 to approximately 0% false nearest neighbors at embedding dimension 3. Thus no developmental change in the required dimension of an embedding space for these time series is observed using this analysis. This does not rule out a qualitative change in the postural control system at the time of the onset of self supported sitting, but no evidence supporting such a qualitative change was observed.

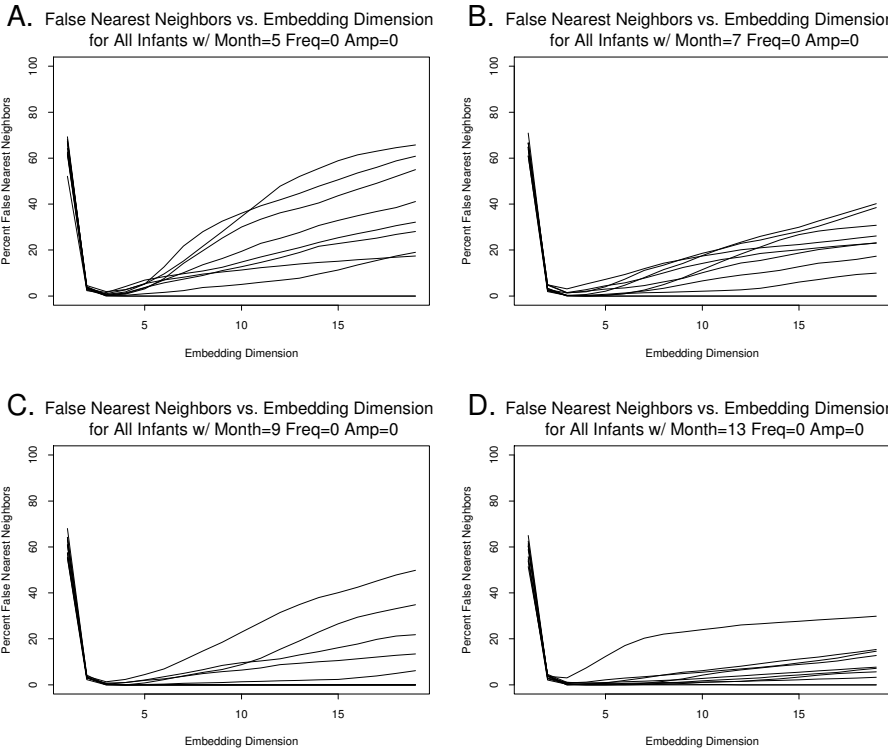


Figure 27: False nearest neighbors curves for sitting infants for ages 5, 7, 9, and 13 months. Each line represents the False Nearest Neighbors curve for the data from two trials for one infant in the control condition.

On the other hand when embedding dimensions greater than 4 are considered, there is an obvious difference in the four graphs in Figure 27. The slope of a regression line fit to the false nearest neighbors curve as the curve begins to ascend from 0% false nearest neighbors can be evidence for the amount of noise in one time series relative to other time series in the set. In Figure 27, these slopes are visibly smaller as age increases. A regression line was fit to each of the infant's false nearest neighbors curves for embedding dimensions greater than 4 and the mean values for these slopes are presented in Table 4.

Table 4: Average false nearest neighbors slopes for each infant age group aggregated over all stimulus conditions.

	Age in Months			
	5	7	9	13
Slope	2.09	0.57	0.33	0.34

5.6 Nonlinear Dependencies

One set of questions that arise with respect to the moving room experiment have to do with the dependency between the room position and the infant center of pressure. In the past, these questions have usually been addressed using linear measures of cross correlation. The present analysis calculates mutual information as a measure of these dependencies and compares the results of that calculation with the results

of calculating squared cross correlation across the two time series (Boker, Schreiber, Pompe, & Bertenthal, 1998). Table 5 presents the summary results from these two methods of calculating the mean dependency between the time series. The values within one column in Table 5 can be compared with one another, whereas the values cannot be compared across the two different columns since the two methods produce values in different metrics. What can be learned from these comparisons is how each of the two methods differ in measuring the developmental change occurring during the span of ages covered by the experiment.

Table 5: Two mean cross dependency measures calculated over all experimental conditions. The three dependency measures are the mean over all time delays $-50 \leq \tau \leq 50$ between infant fore-aft center of pressure time series and room position time series. Values in parentheses are 95% confidence intervals for the associated statistic. Both mutual information is expressed in bits.

	Mean Square Cross Correlations	Mutual Information
5 Months	0.028 (0.021,0.036)	0.0119 (0.0113,0.0126)
7 Months	0.070 (0.059,0.082)	0.0157 (0.0150,0.0163)
9 Months	0.092 (0.077,0.107)	0.0187 (0.0178,0.0195)
13 Months	0.096 (0.079,0.113)	0.0182 (0.0173,0.0192)

Note that both of these methods suggest that between the ages of 5 months and 7 months there is a significant change in the mean cross dependency of the room position and the infants' fore-aft center of pressure. During the interval that separates ages 9 and 13 months, neither of the methods detect a significant change in the mean

cross dependency. Over these two developmental intervals the two methods provide essentially the same results. However note the difference between the methods when comparing the values at 7 months and the values at 9 months. The mutual information method provides a greater degree of discrimination during this critical developmental time immediately following the onset of self-supported sitting in the infants.

Figure 28 presents the mutual information within the room movement time series for each of the four room movement frequencies: 0 Hz, 0.3 Hz, 0.6 Hz and the pseudorandom movement labeled as “multi” in Figure 28–B. Note that the mutual information curves show well defined peaks at the expected half-period intervals in the 0.3 Hz and 0.6 Hz experimental conditions, demonstrating the time delays where the room signal is 180° and 360° out of phase with itself. Also note that there is no such apparent peak in the pseudorandom “multi” condition, suggesting that the method for producing pseudorandom movement did not induce an accidental base frequency.

It is also interesting to note that in the control condition where the room was not supposed to move, there does appear to be a small regular periodic signal. If there was a small 60 Hz contamination in the sampling, since the sampling frequency was 50 Hz we might expect an apparent beat frequency of approximately 5 Hz. This would translate into small regular peaks approximately 100 ms apart. Such peaks are apparent in the 0 Hz graph in Figure 28–A, suggesting how sensitive the mutual

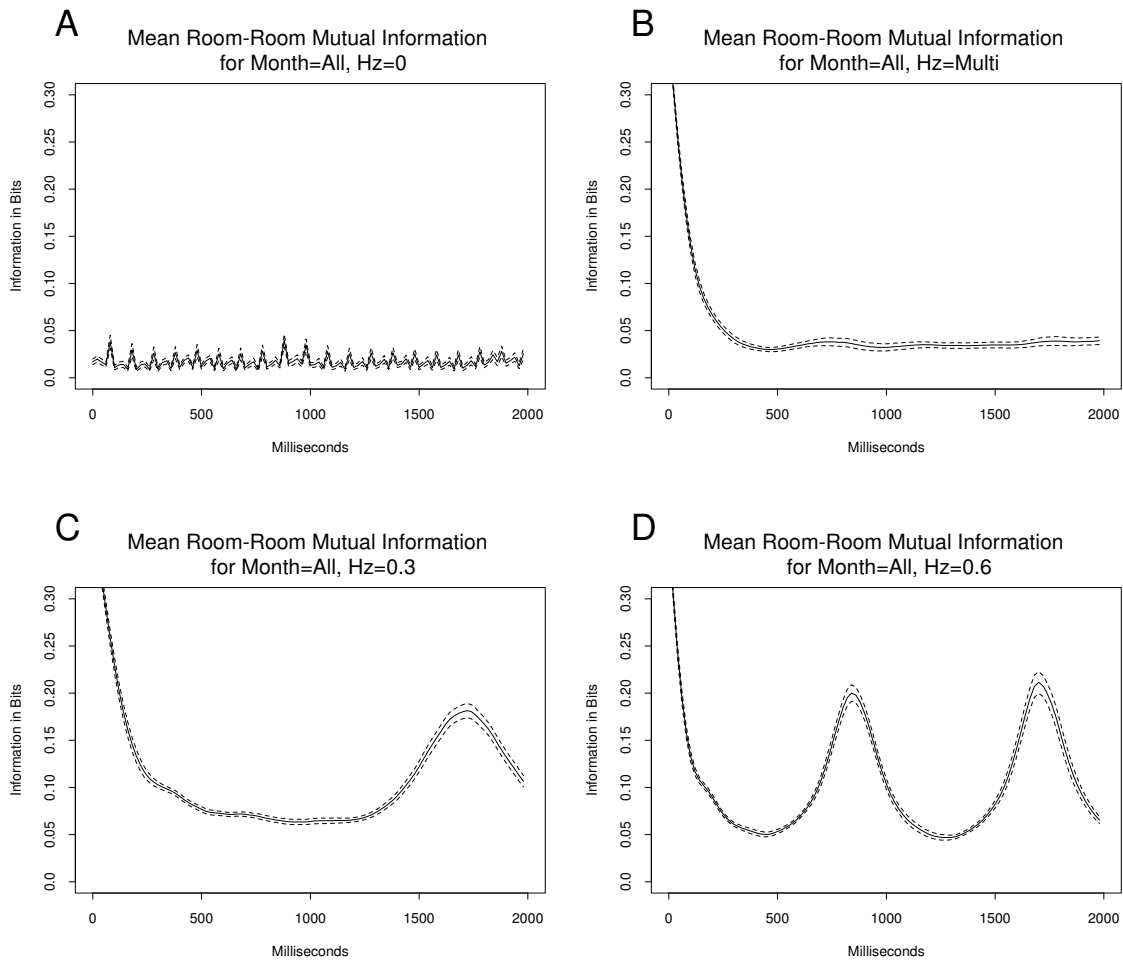


Figure 28: Mean mutual information within the room position time series by experimental condition and across all trials and individuals. Means are calculated within experimental condition and across trials and individuals. The solid line represents the mean mutual information curve and the dashed lines above and below the mean mutual information curve represent 95% confidence intervals for the mean.

information method can be.

Figure 29 presents the nonlinear dependency of the fore–aft center of pressure on the room position as calculated by mutual information. Note that when the room is not moving it is a poor predictor of the infants' position, although there is a small upward slope to the mutual information curve in Figure 29–A. This small upward slope is an artifact of the short length of the time series since as the time delay becomes larger, the number of samples which can participate in the mutual information calculation become smaller. As the number of non–populated cells in the joint probability distribution estimation become more numerous, more of the calculation is based on values near the center of the distribution, the position which is predicted by the non–moving room.

The nonlinear dependency of the infants' fore–aft center of pressure on the pseudorandom room movement is plotted in Figure 29–B. The general trend to greater mutual information at longer lags can be noted in this graph, but there is a significantly greater amount mutual information at all delays between the moving room and the center of pressure than was shown at any delay in the control condition. There are also two significant peaks in the mutual information in this graph, one centered at a delay of 250 ms and one centered at a delay of 1000 ms.

These two peaks suggest that there is some periodicity in the prediction of the

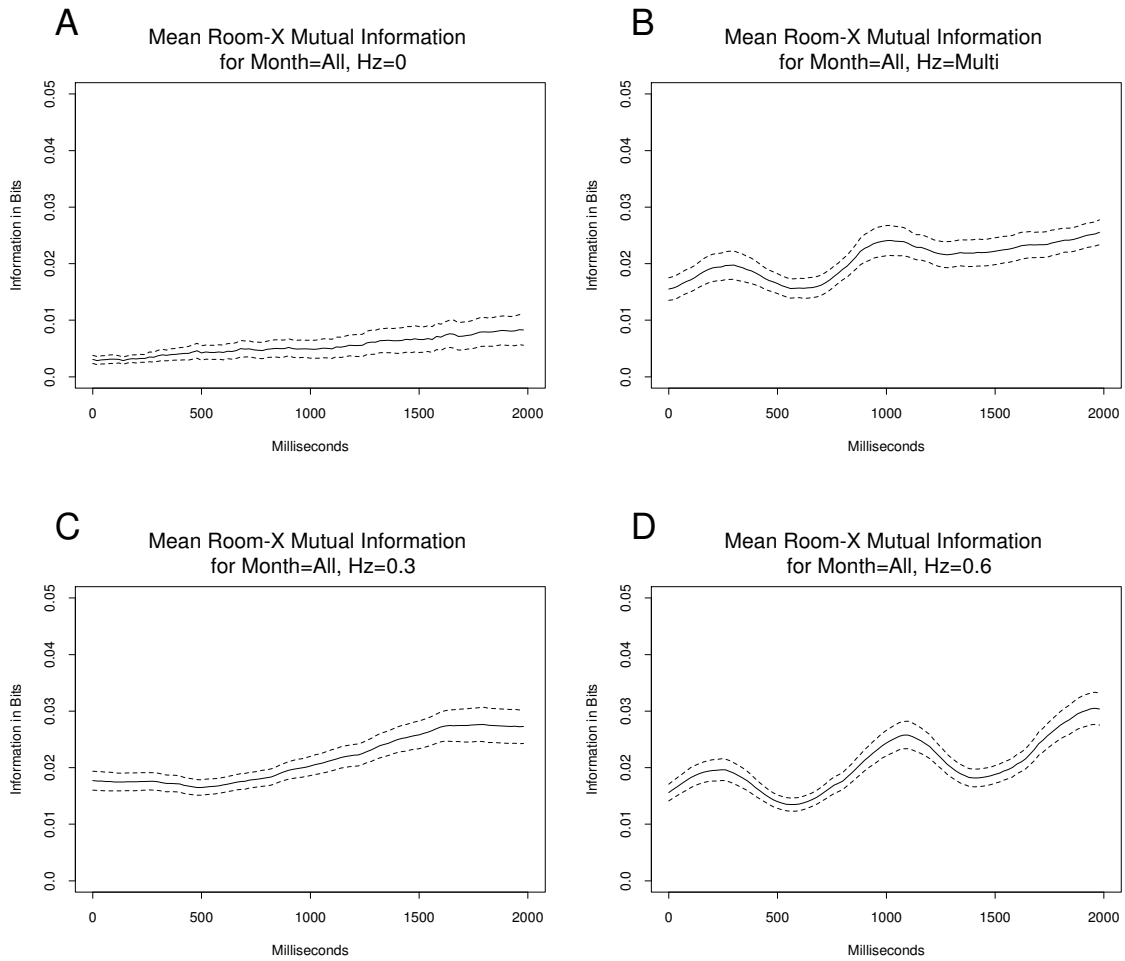


Figure 29: Mean mutual information of the room position time series predicting the fore-aft center of pressure time series. Means are calculated within experimental condition and across trials and individuals. The solid line represents the mean mutual information curve and the dashed lines above and below the mean mutual information curve represent 95% confidence intervals for the mean.

infants' fore-aft center of pressure by the pseudorandom moving room. There are two mechanisms that may have generated periodicity in this cross dependency. First of all, the pseudorandom room movement may have a greater than chance number of frequency components near a frequency of 0.65 Hz. Examining the Fourier transform of the pseudorandom room movement signals shown in Figure 23-B shows a small, but significant hump in the frequency components centered around 0.7 Hz. The second possibility which may underlie this observed periodicity may be due to some sort of resonance in the infants' response. The evidence presented here cannot distinguish between these two possibilities and a follow-up experiment would be advised to test these two hypotheses.

Figure 29-C plots the nonlinear dependency of the infants' fore-aft center of pressure on the room position for the 0.3 Hz condition. For this condition, the mutual information is significantly greater at all time-delays than for the control condition, and there is a significant increase in the mutual information as the time-delay increases. There should be a peak in the mutual information at a time delay a little longer than 1667 ms, the half period of the 0.3 Hz stimulus frequency. However, since the mutual information was only plotted to a time delay of 2000 ms it is impossible to tell whether or not there is a peak in the mutual information just short of 2000 ms.

The final graph in Figure 29 plots the same nonlinear dependency for the stimulus

condition of a room moving at 0.6 Hz. Here the evidence for the entrainment of the infant to the stimulus frequency can be seen in the two significant peaks centered at 250 ms and 1100 ms. The 850 ms delay between these two peaks suggests a periodicity at 1700 ms or 0.59 Hz, almost exactly the driving frequency of the stimulus moving room. The first peak at 250 ms suggests that on average, the infants are lagging behind the room approximately 250 ms in the 0.6 Hz condition.

The four graphs in Figure 30 plot the nonlinear dependency of the room position on the infants' fore-aft center of pressure. Whereas in a cross correlation analysis these would be expected to be symmetric with the graphs in in Figure 29, when plotting the nonlinear dependency these graphs are not necessarily symmetric due to the high predictability in the room position time series and much lower predictability in the fore-aft center of pressure time series. For instance, note the mutual information curve shown in Figure 30-B which plots the mutual information for the pseudorandom stimulus condition. Where in Figure 29-B there were two significant peaks in the mutual information curve of the room predicting the infants' center of pressure, now when the infants' center of pressure is used to predict the room these indicators of periodicity are not present.

On the other hand, there is a good deal of correspondence between the Figure 29-D and Figure 30-D where in both graphs there exist two significant peaks. In the case where the room is predicting the infant, there is a 250 ms delay to the first peak

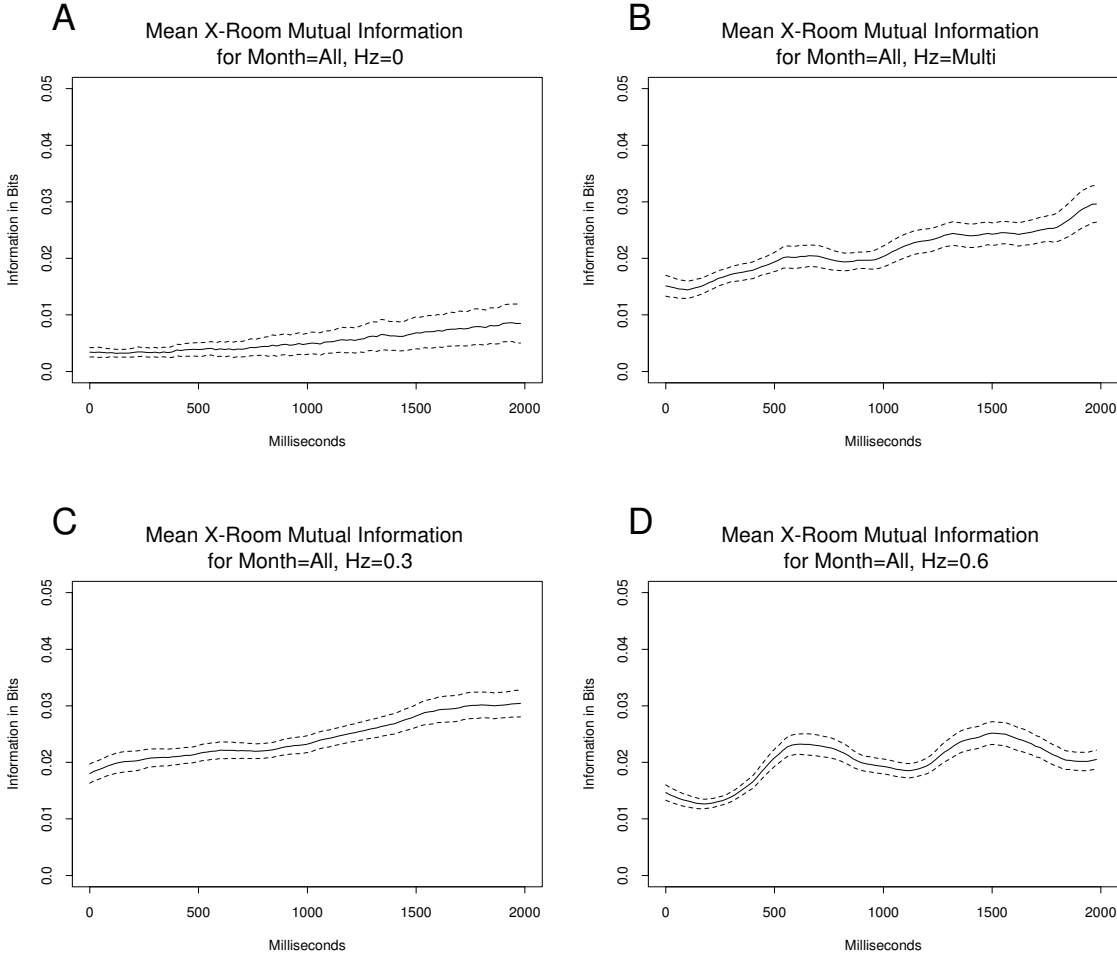


Figure 30: Mean mutual information of the fore-aft center of pressure time series predicting the room position time series.

indicating that the room is leading the infant by about 250 ms. In Figure 30–D where the infants' fore–aft center of pressure is predicting the room's position, we find the first peak at a delay of about about 600 ms which is about 250 ms before the half period of the room. Thus this calculation also suggests that the room position is leading the infants by about 250 ms in the 0.6 Hz moving room condition.

Figures 31 and 32 plot the nonlinear dependencies between the room position and the lateral center of pressure time series. When like experimental conditions are compared, the nonlinear dependency curves in these figures are not significantly different from each other. Thus we conclude that there is no apparent asymmetry between the prediction of the room position by the infants' lateral center of pressure and the prediction of the infants' lateral center of pressure by the room position.

The overall height of the nonlinear dependency curves in each of the moving room experimental conditions in Figures 31 and 32 is significantly greater than that of the control condition. This suggests that there are significant dependencies between the moving room time series and the lateral center of pressure time series.

However, it should be noted that the overall height of the nonlinear dependency curves plotted in Figures 31 and 32 are significantly less than the corresponding curves in Figures 29 and 30. Thus we conclude that there is significantly less dependency between moving room and the lateral center of pressure time series than there is

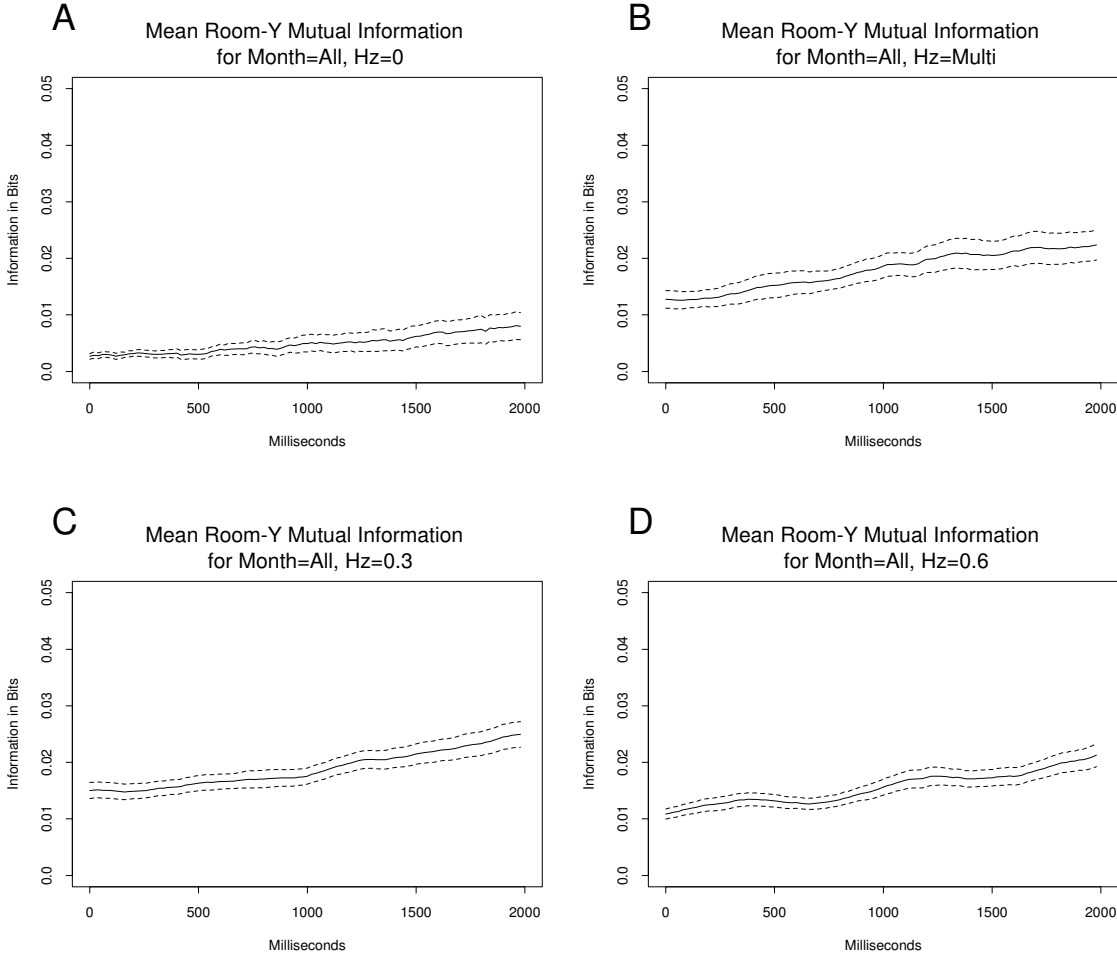


Figure 31: Mean mutual information of the room position time series predicting the lateral center of pressure time series.

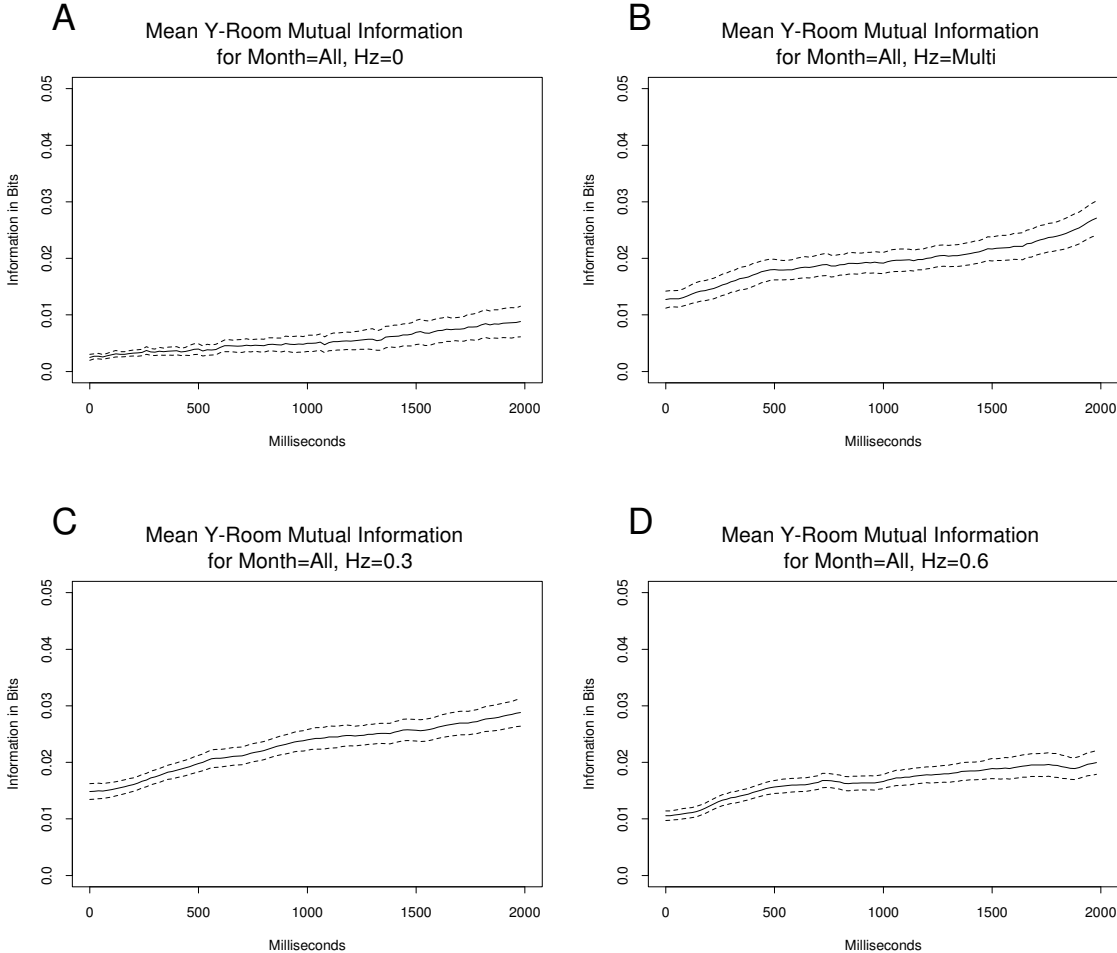


Figure 32: Mean mutual information of the lateral center of pressure time series predicting the room position time series.

between the moving room and the fore–aft center of pressure time series.

There is also something notably absent from the graphs in Figures 31 and 32. There is no evidence of periodicity in these mean nonlinear dependency curves between the room position and lateral center of pressure, even though there *is* significant dependency between the two signals. This may be due to the fact that these are mean dependency curves and the possibility that the infants are behaving in the following manner. Suppose that each infant on each trial swayed laterally in sync with the movement of the room. However, further suppose that there was a random association between left/right sway and front/back movement of the room. Practically this would mean that the infant might sway a little to the left as it moved backward on one trial whereas on the next trial it might sway a little to the right as it moved backward. If this were the case, then due to phase cancelation the infants' mean dependency curves would show no periodicity.

Figures 33 and 34 plot the mean mutual information within the fore–aft time series and lateral center of pressure time series respectively. These time series do not differ significantly from one another except for two features. First, in the range of time delays between 250 ms and 700 ms, there is significantly more mutual information in the control condition than in the three experimental conditions. This holds for both the fore–aft and lateral series in all conditions except the 0.3 Hz lateral time series. The second significant feature is that for the fore–aft time series there is a significant

hump in the mutual information curve centered around 1650 ms. This feature can be taken as evidence of periodicity within the fore–aft center of pressure time series; a periodicity with a frequency of about 0.3 Hz.

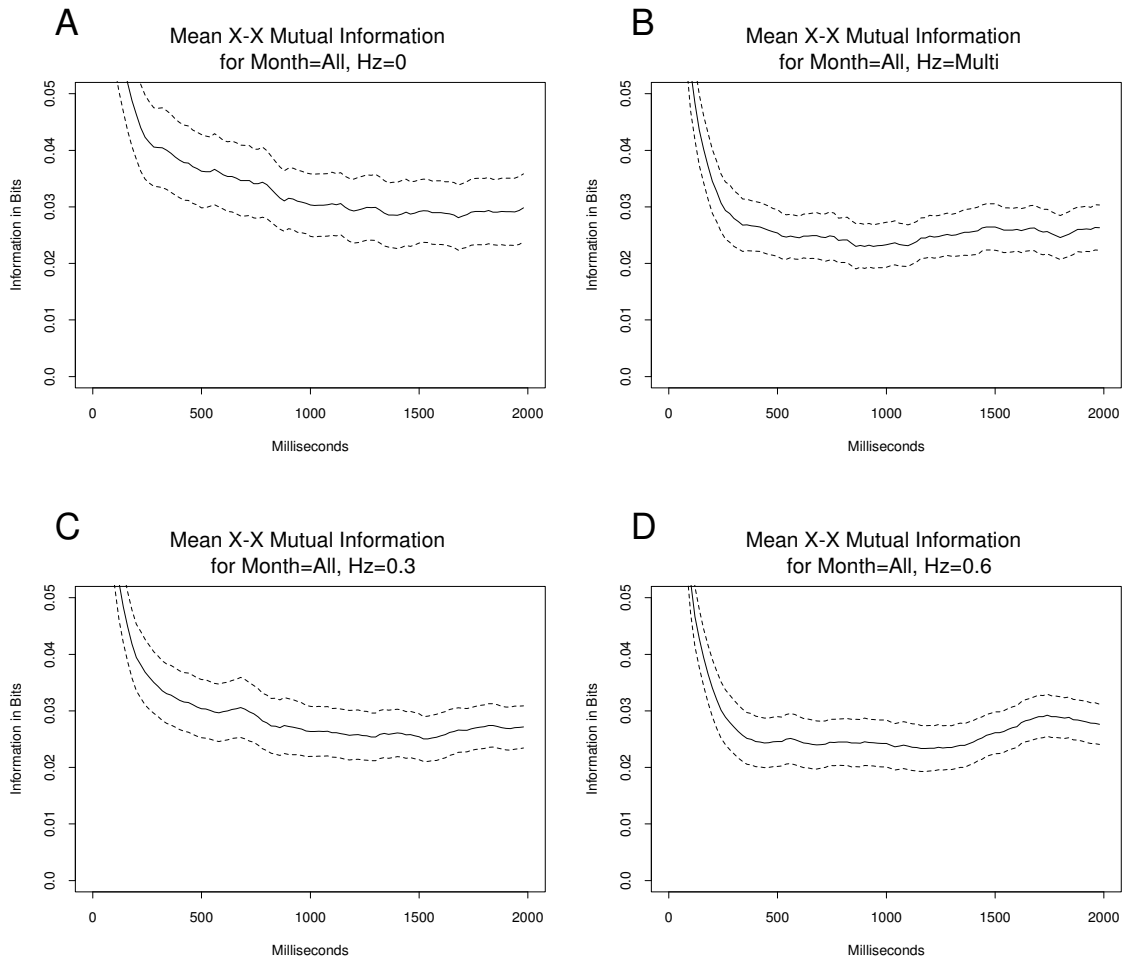


Figure 33: Mean mutual information within the fore–aft center of pressure time series.

Figures 35 and 36 plot the mutual information between the fore–aft center of pressure and the lateral center of pressure. Here there is slightly more mutual infor-

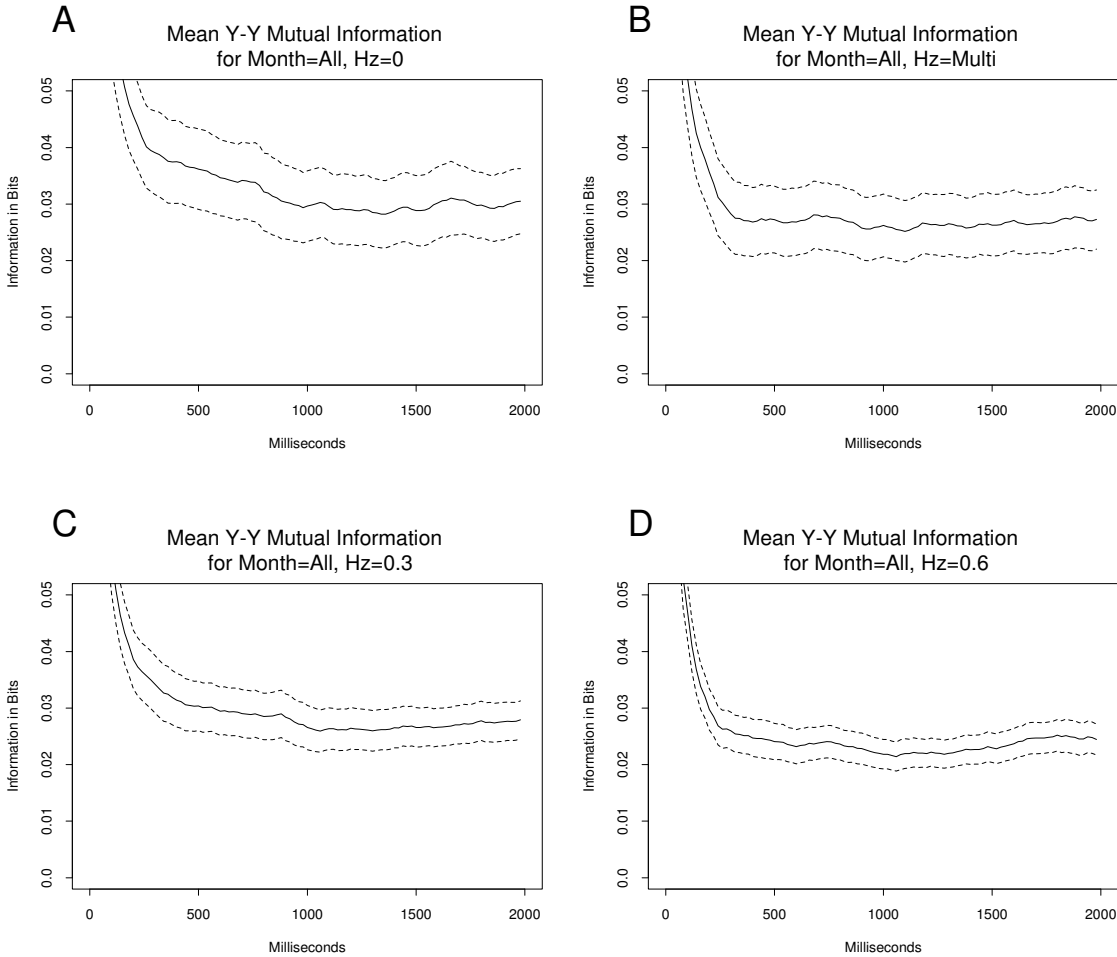


Figure 34: Mean mutual information within the lateral center of pressure time series.

mation between the fore–aft and lateral time series in the control condition, but it only achieves significance in comparison to the “multi” frequency and 0.6 Hz experimental conditions and then only for the time–delays between 0 ms and 1200 ms. It is difficult to make sense of why there would be more mutual information between the fore–aft and lateral time series in the control condition than there would be in the experimental condition when the infant is most strongly entrained to the room movement. This unexpected finding needs to be explored further.

5.6.1 Developmental Change in Nonlinear Dependencies

We next examined questions relating to the developmental change in the nonlinear dependencies between the room and the infants’ center of pressure time series. Figure 37 plots the mean nonlinear dependency of the infants’ fore–aft center of pressure on the room position according to the four age categories in the study. Note that even in Figure 37–A which plots the mean mutual information curve for the 5 month–olds, there is already some evidence of periodicity. As the age category increases from 7 months to 13 months, the evidence for periodicity becomes stronger and stronger.

It is also interesting to note that in each of the four mutual information curves in Figure 37, the curve is not a time–symmetric sinusoid. The time delay between a peak in the curve and the subsequent valley is shorter than the time delay between

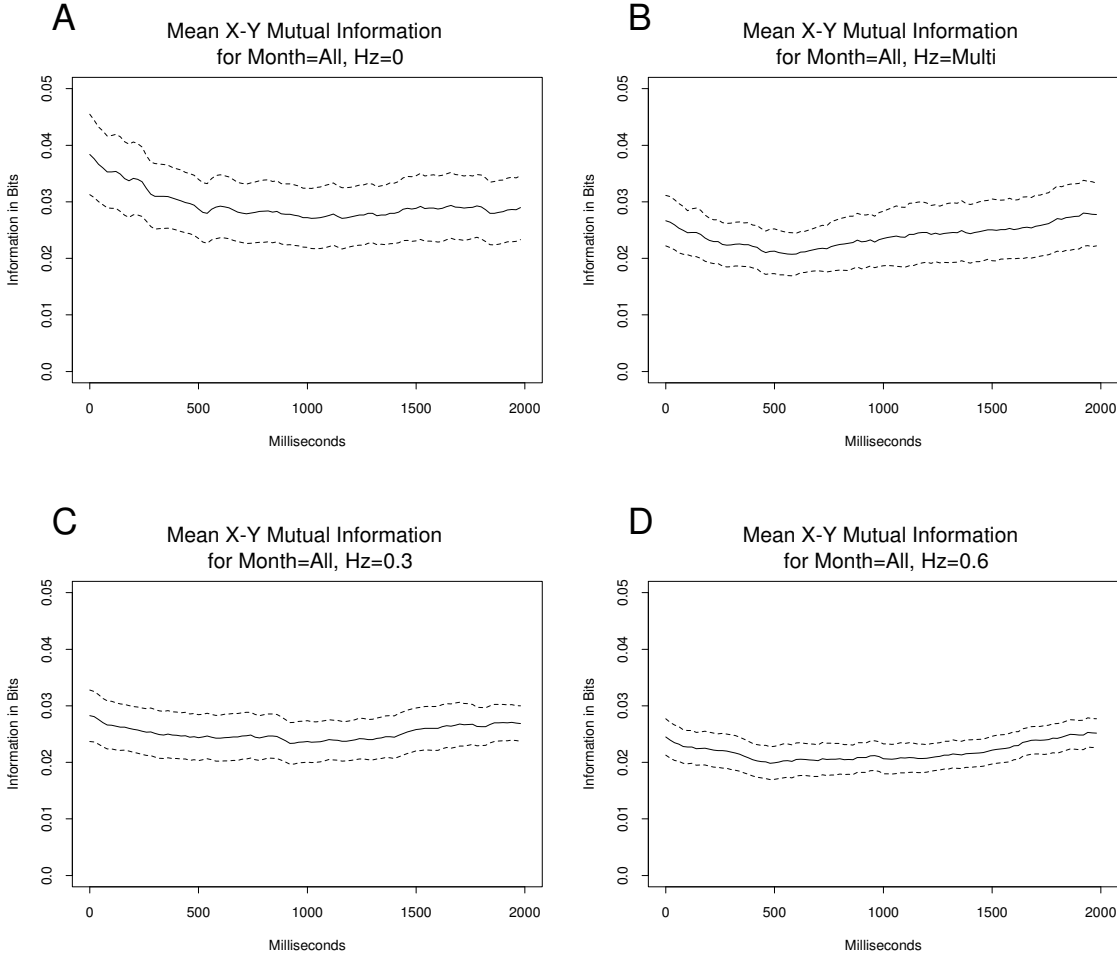


Figure 35: Mean mutual information of the fore-aft center of pressure time series predicting the lateral center of pressure time series.

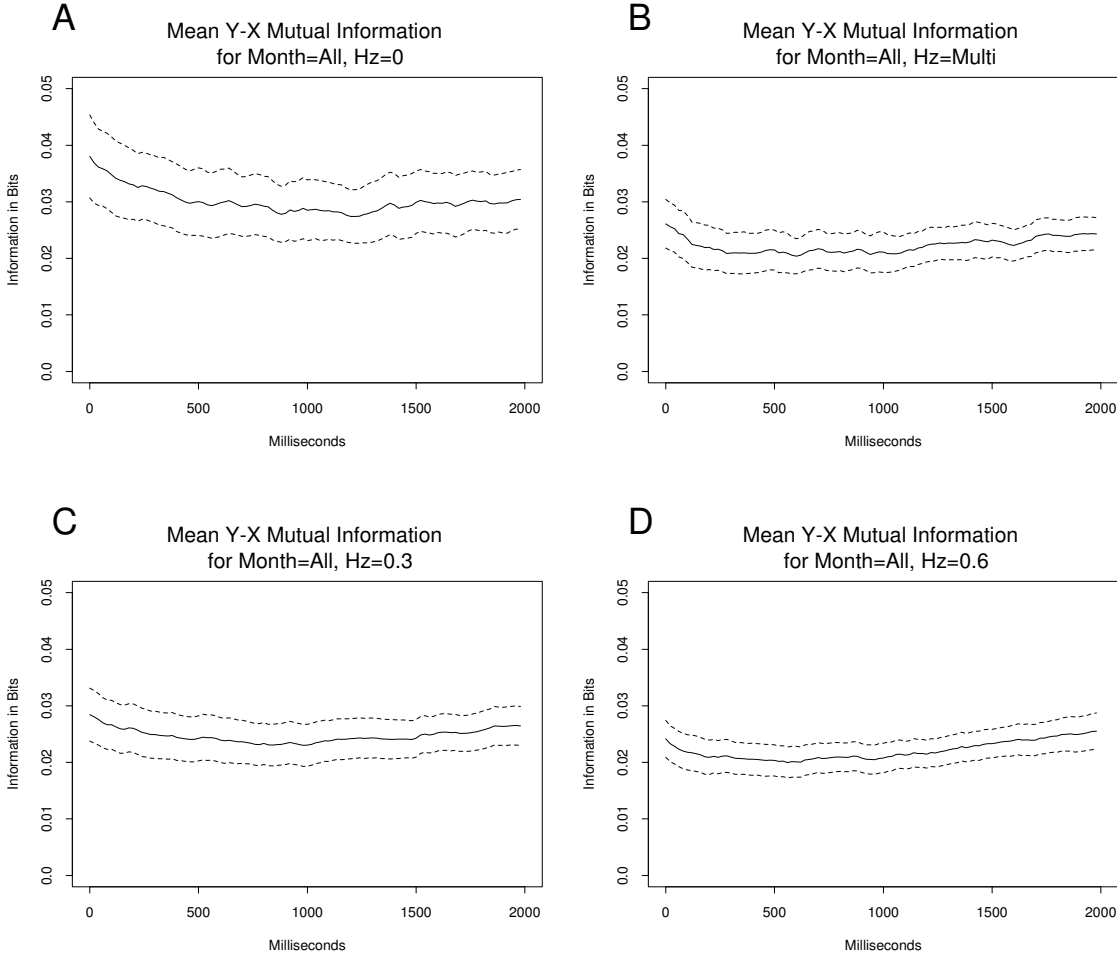


Figure 36: Mean mutual information of the lateral center of pressure time series predicting the fore-aft center of pressure time series.

a valley and the subsequent peak. This means that these data contain evidence for time asymmetry, one of the hallmarks of nonlinear dynamics. In order to test this hypothesis, a surrogate data test for time asymmetry should be performed.

Figure 38 plots the nonlinear dependency of the room position on the infants' fore-aft center of pressure for the 0.6 Hz room movement condition. These graphs show the same periodicities as in Figure 37, but the effect is not quite as strong in this prediction. Other than the reduced strength of the effect of periodicity, these graphs appear to be time reversals of the prediction of the fore-aft center of pressure by the room position. Note that the peak to valley time-delay asymmetry is reversed in these graphs from that of the previous figure.

Figures 39 and 40 plot the nonlinear dependency between the room position and the fore-aft center of pressure for the 0.3 Hz experimental condition. In this condition the entrainment of the infant to the room movement is less than in the 0.6 Hz condition. We see less evidence of periodicity in the younger infants, although there is a pronounced periodic peak and valley in the 13 month old infants' mutual information curve.

It should also be noted that there is a significant increase in the general height of the nonlinear dependency curves in Figure 39 as the age category increases from 5 months to 7 months and again as the age category again increases to 9 months.

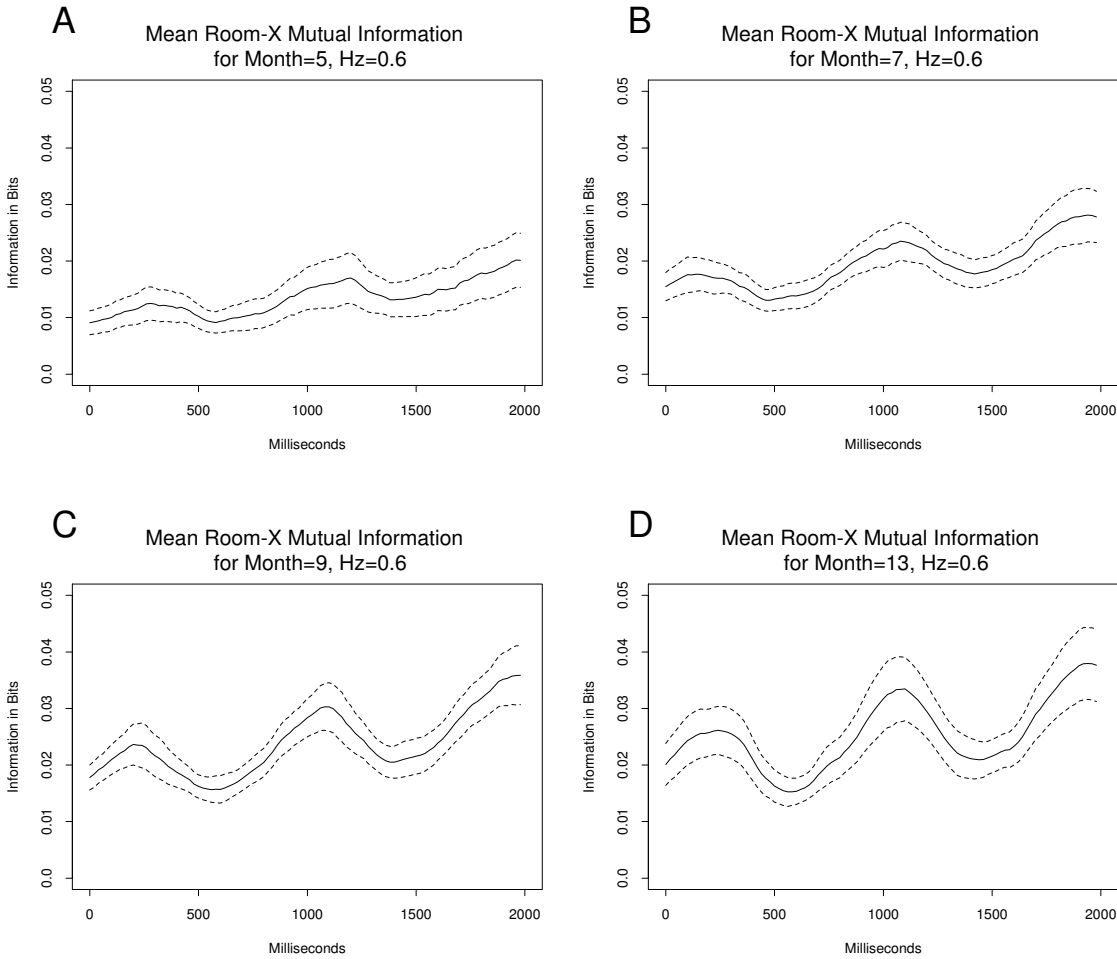


Figure 37: Mean mutual information of the room position predicting the fore–aft center of pressure time series in the 0.6 Hz experimental condition. The four graphs aggregate the mean mutual information within each of the four age categories. The dashed lines are 95% confidence intervals around the mean mutual information curve.

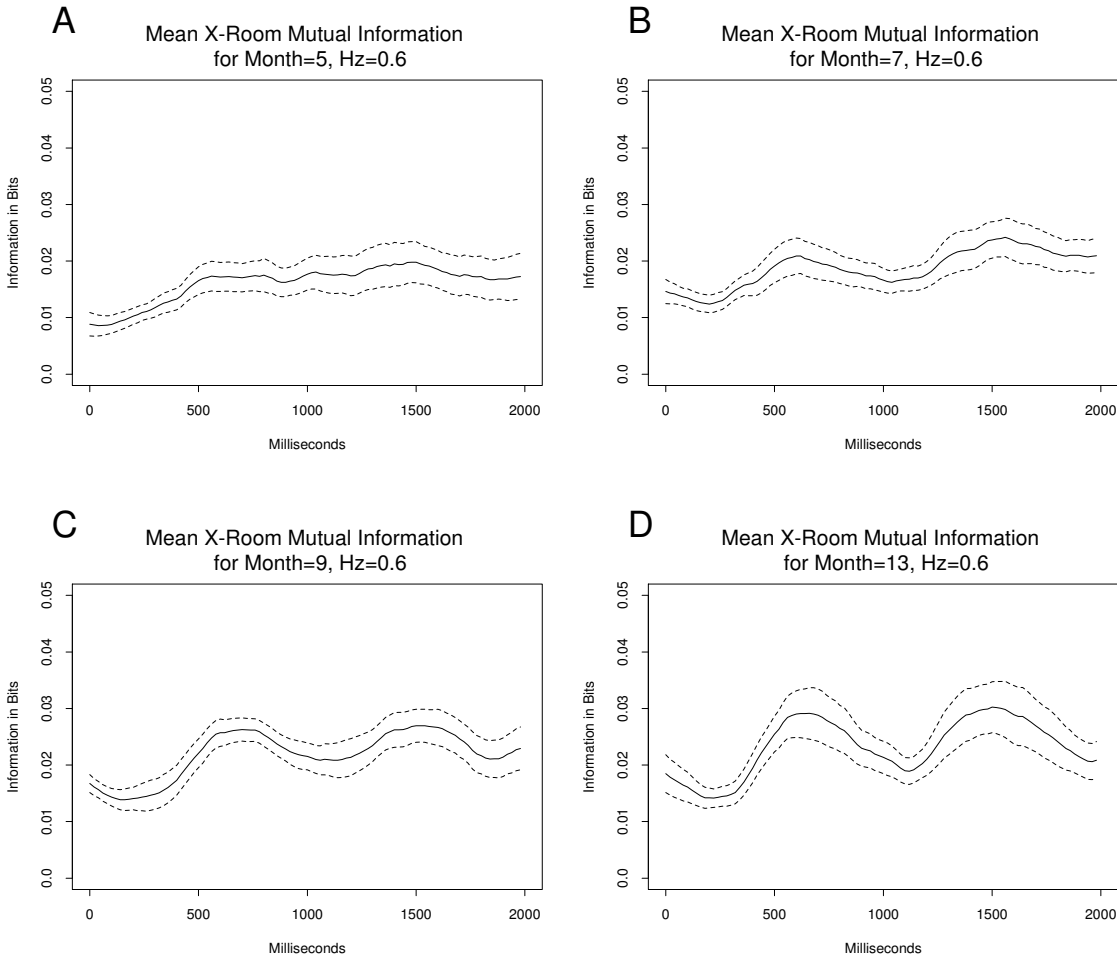


Figure 38: Mean mutual information of the fore–aft center of pressure time series predicting the room position in the 0.6 Hz experimental condition. The four graphs aggregate the mean mutual information within each of the four age categories. The dashed lines are 95% confidence intervals around the mean mutual information curve.

We interpret this to mean that there is a significant increase in entrainment over this period. However, in the 13 month old infants we see a significant peak and valley form where there was no such feature in the younger age categories. We interpret this to mean that there is a significant change in phase synchronization to the room movement in the 13 month–old infants: the time delays of the infants in the sample of 13 month–olds is significantly more consistent than that of the 9 month–olds.

There is not a significant increase in the height of the nonlinear dependency curves in Figure 39, suggesting that there is a smaller developmental change in the coupling of the infant to the room than there is in the coupling of the room to the infant. This hypothesis can be tested directly using a MIMSE model for the cross–dependencies. Unfortunately that is beyond the scope of the present work, but will be actively pursued in the coming year.

6 Discussion

There are several notable results from the present nonlinear analyses of the data from the moving room experiment. The first result comes from the mutual information analysis within the fore–aft and lateral center of pressure time series. At time delays of less than 100 ms the center of pressure time series is strongly autocorrelated. This effect is expected due to the momentum of the infant’s body and must be taken into

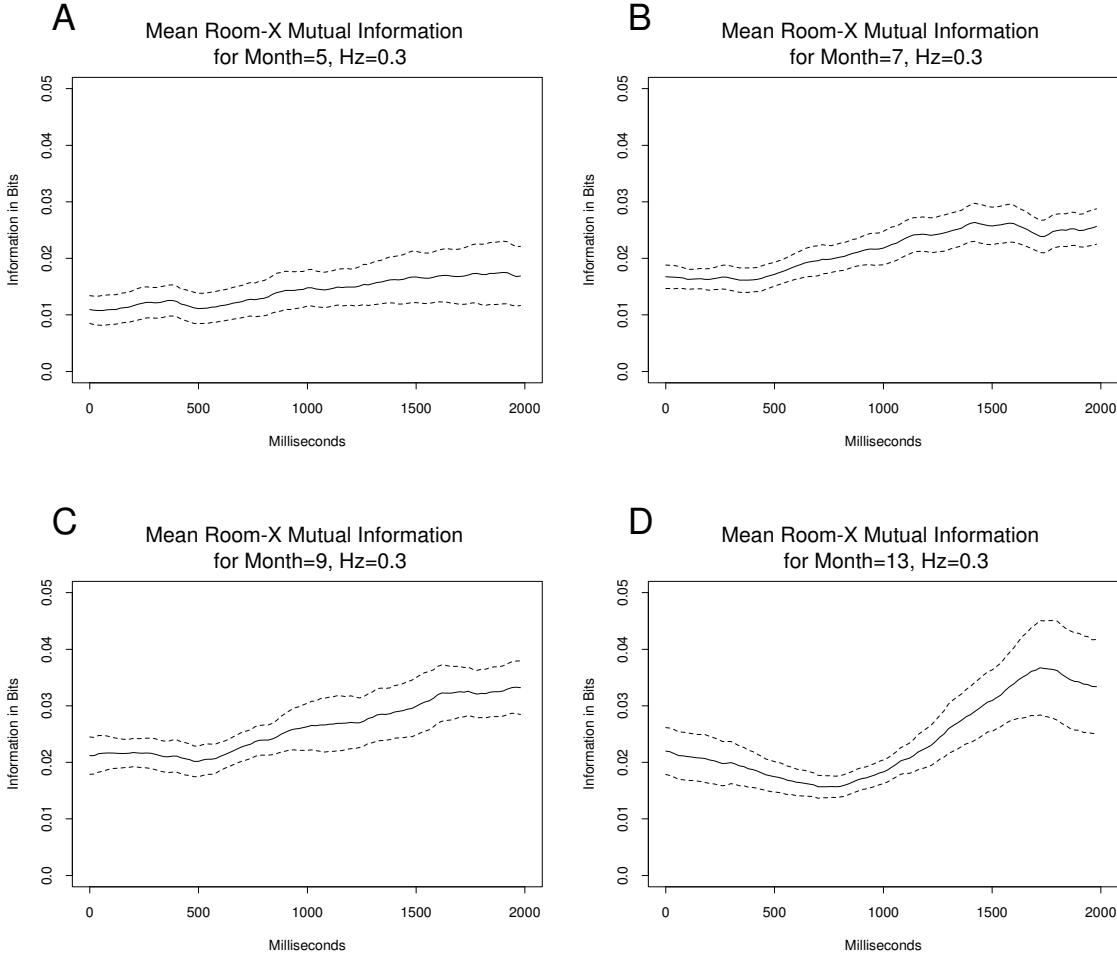


Figure 39: Mean mutual information of the room position predicting the fore-aft center of pressure time series in the 0.3 Hz experimental condition.

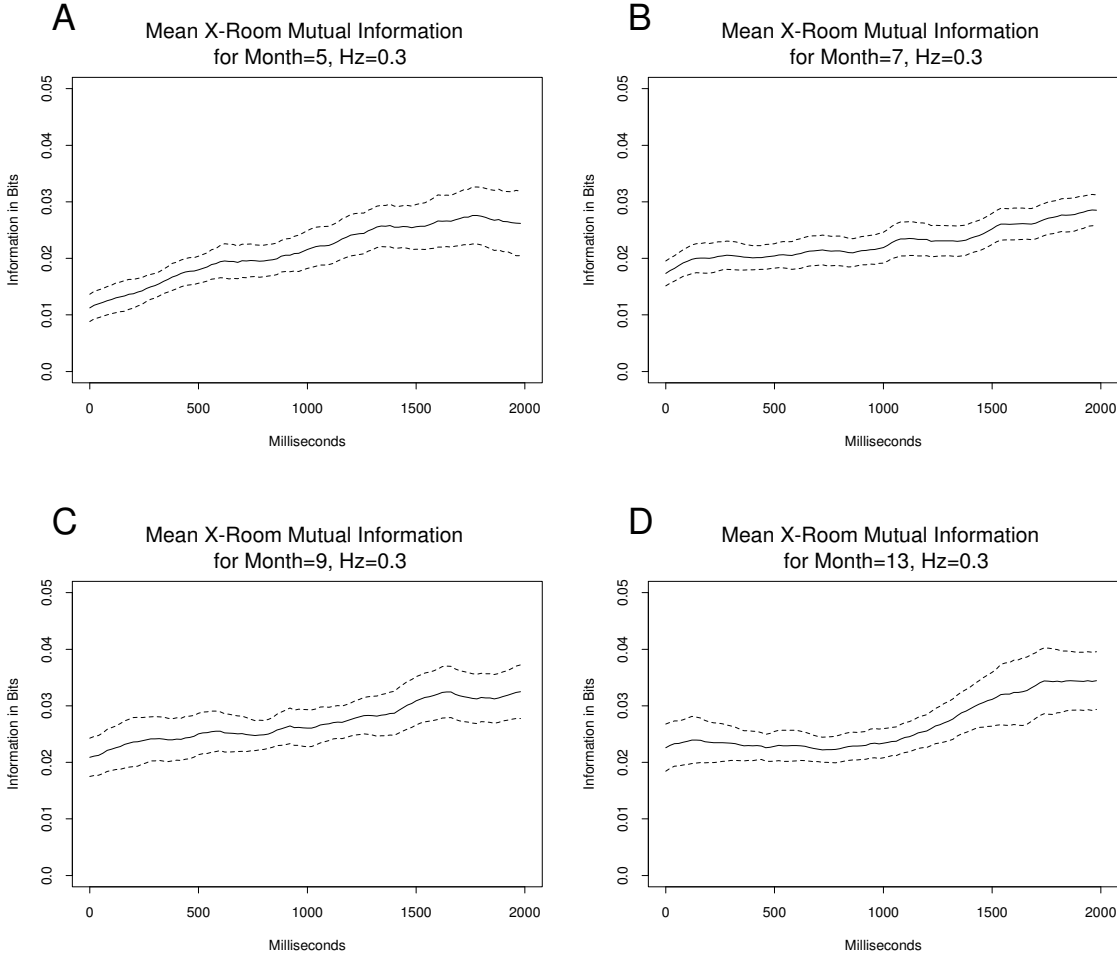


Figure 40: Mean mutual information of the fore-aft center of pressure time series predicting the room position in the 0.3 Hz experimental condition.

account by the postural control mechanism, but is not a direct effect of postural control. As such, we conclude that a sampling interval of one half this interval of 100 ms would be sufficient to capture the dynamics of the control signal. Sampling at a higher rate generates more data, but it does not generate more information about the control mechanism. Therefore an optimal sampling strategy would sample at 50 ms intervals for a longer period of time rather than sampling at shorter intervals for a shorter period.

Choosing an appropriate sampling rate is critical since many of the nonlinear techniques currently in use have strict limits on the total number of data points needed in a time series. Sampling at a rate higher than one half the first minimum of the mutual information function may generate highly correlated samples. Thus when sampling at very short intervals one may be generating many more data points, however one is not necessarily generating more degrees of freedom. It is important to sample at a rate which will capture the salient characteristics of the system and sample for a long enough period that the length of the time series satisfies the requirements of the chosen method of analysis. We recommend using a mutual information plot of pilot data as a guide to appropriate repeated measures experimental design.

The second notable result presented in this work is that a surrogate data analysis of the center of pressure time series exhibits strong evidence that a linear stochastic model of the postural control of infants' sitting behavior is insufficient to capture

the dynamical properties of these data. Another way of expressing this conclusion is that we have found strong evidence of nonlinearity in the center of pressure time series. This conclusion includes the possibility that the dynamics of the control mechanism itself may have a nonlinear component. This finding is not what would be predicted by Collins and De Luca's (1994) analysis of adult center of pressure data from quiet standing. There are two primary explanations for this inconsistency: the postural control mechanism underlying infants' self-supported sitting may be fundamentally different than the postural control mechanism for adults' quiet standing, or the method which Collins and De Luca used to test for nonlinearity may have been insufficiently sensitive to detect the difference between their real and surrogate data. Since Collins and De Luca sampled at 10 ms intervals and did not select an appropriate time delay for their analysis, we speculate that their surrogate data test failed to detect evidence of nonlinearity due to a large degree of autocorrelation present at such a short sampling interval.

The third main finding presented here derives from the false nearest neighbors analysis of the sitting infants in the control condition, which found that there is a decrease in the slope of the false nearest neighbors curve over the age range of 7 to 9 months with no further decrease in the slope during the age range 9 to 13 months. One way of interpreting this finding is that the decrease in slope of the false nearest neighbors curves is a measure of a decrease in stochastic or high-dimensional noise in

the infants' postural control mechanism. The surrogate data tests for nonlinearity in the same time series over the interval 5 to 13 months showed an increase in percentage of trials which rejected the null hypothesis of no nonlinearity. These two findings together suggest that within the center of pressure signal for sitting infants there may be an additive stochastic component mixed with a nonlinear postural control signal for the youngest infants, and as the infants develop, the stochastic component is reduced and is replaced by additional nonlinear dynamical control.

The false nearest neighbors analysis tested for, but did not find evidence for a change in the required minimum embedding dimension for the center of pressure time series. Thus no evidence was found for a qualitative change in postural control coincident with the onset of self-supported sitting. This may mean that there is no qualitative change in the postural control mechanism at this critical developmental time, or it may be due to insufficient sensitivity of the false nearest neighbors algorithm to dimensional changes in the center of pressure time series, or perhaps it may be that whatever qualitative shift may occur is not manifested in a change in the dimension of the center of pressure time series. At the very least, we can report that there is not a large or obvious change in the dimensionality of the center of pressure signal coincident with self-supported sitting behavior.

The nonlinear dependency analyses provided several noteworthy findings. The mean nonlinear dependency between the moving room and the infants' fore-aft center

of pressure provided a greater sensitivity in discriminating developmental change as evidenced by smaller error bars with respect to developmental differences than did the linear measure of mean square cross correlation. This finding is not surprising given that strong evidence of nonlinearity in the center of pressure signal was detected. The linear measure ignores the nonlinearity in the center of pressure time series and thus if there is any developmental change in the nonlinearity, the linear measure would fail to detect it.

The nonlinear dependency analysis of the room movement predicting the infants' fore-aft center of pressure detected significant dependency in all experimental conditions when compared to the control condition. Also, significant periodicity was detected in the multiple frequency and 0.6 Hz experimental conditions. The nonlinear dependency analysis of the 0.6 Hz experimental condition produced an estimate of a 250 ms average lag of the infants' position behind the movement of the room. The analyses presented here were unable to produce an estimate of the lag of the infant in the 0.3 Hz experimental condition due to the fact the the total time series included only three full periods of the 0.3 Hz moving room. Further analysis may be able to produce an estimate of the lag of the infant in the 0.3 Hz condition, although this is near the limit of these data to be able to answer this question.

When the nonlinear dependency between the room position and the infants' lateral center of pressure was calculated, significant dependency was found in all exper-

imental conditions when compared to the control condition. However, no significant periodicity was found in any of the experimental conditions. We attribute this to a random pairing of left–right movement with fore–aft movement. Although within a trial these movements exhibit periodic dependency, across trials the periodic dependency cancels out.

When the nonlinear dependency analysis was aggregated by age category, a strong effect of age was found in the dependency between the room movement and the fore–aft center of pressure for both the 0.6 Hz and 0.3 Hz experimental conditions. The 0.6 Hz experimental condition showed evident periodicity even for the 5 month olds, and there was a significant change in the periodicity between 5 and 7 months and again between 7 and 9 months. This effect was evident both for the dependency of the infant on the room as well as for the dependency of the room on the infant. The 0.3 Hz experimental condition only exhibited a developmental effect in the periodicity between 9 months and 13 months. This result suggests that 13 month old infants produced a center of pressure signal in the 0.3 Hz condition which was significantly less variable than younger infants.

6.1 General Discussion

When all of the findings from the present investigation are viewed together, a picture of the development of the coupling between visual perception and postural control begins to emerge. The evidence suggests that visually guided postural control is present in 5 month old infants and does not undergo a major qualitative change concurrent with the onset of self-supported sitting. The analyses presented here suggest that visually guided postural control may involve a nonlinear dynamical control mechanism. Furthermore the evidence suggests that this mechanism is mixed with noise in younger infants and that much of that noise is somehow “converted” to low dimensional nonlinearity during the development that occurs between 5 and 9 months.

There are a number of analyses which are planned for these data in order to further understand the development of the coupling of visual perception to postural control. In a particularly promising analysis, the structure of the lagged nonlinear dependencies within and between the lateral, fore-aft and room signals will be modeled using the MIMSE modeling technique presented in Section 3.2. The author expects that these models will provide further insight into the development of the nonlinear coupling between vision and postural control.

Behavioral research has depended heavily on relatively simple linear models for over a hundred years. As our discipline continues to evolve, interest grows in us-

ing promising but more complex and elaborate modeling procedures suited to the complexity of our data. As behavioral scientists confront the problem of modeling complex intraindividual variability they will increasingly turn to nonlinear dynamical systems representation (West, 1985). The analyses presented in this dissertation have helped to demonstrate the utility of nonlinear dynamical systems techniques. Moreover, they highlight the fact that one does not need extremely long time series in order to perform nonlinear analyses. If one is willing to make the assumption that measurements of individuals are representative of a general underlying process, then multiple time series can be used to examine developmental changes in that process. The results presented here suggest that the application of nonlinear techniques to developmental and physiological time series is a promising area of process-oriented research that is rapidly becoming feasible for implementation by behavioral and social scientists.

References

- Abarbanel, H., Brown, R., & Kennel, M. B. (1991). Lyapunov exponents in chaotic systems: Their importance and their evaluation using observed data. *International Journal of Modern Physics B*, 5(9), 1347–1375.
- Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., & Tsimring, L. S. (1993). The analysis of observed chaotic data in physical systems. *Reviews of Modern Physics*, 65(4), 1331–1392.
- Abarbanel, H. D. I., Carroll, T., Pecora, L. M., Sidorowich, J. J., & Tsimring, L. S. (1994). Predicting physical variables in time-delay embedding. *Physical Review E*, 49(3), 1840–1853.
- Bayley, N. (1969). *Manual for the Bayley Scales of Infant Development*. New York: Psychological Corporation.
- Bertenthal, B. I. (1990). Application of biomechanical principles to the study of perception and action. In H. Bloch & B. I. Bertenthal (Eds.), *Sensory-motor organization and development in infancy and early childhood* (pp. 243–260). Dordrecht: Kluwer.
- Bertenthal, B. I., Rose, J. L., & Bai, D. L. (1997). Perception-action coupling

in the development of the visual control of posture. *Journal of Experimental Psychology: Human Perception and Performance*, 23(6), 1631–1643.

Boker, S. M., Schreiber, T., Pompe, B., & Bertenthal, B. I. (1998). Nonlinear analysis of perceptual–motor coupling in the development of postural control. In H. Kantz, J. Kurths, & G. Mayer-Kress (Eds.), *Nonlinear analysis of physiological data* (pp. 251–270). Berlin: Springer Verlag.

Brown, R., Bryant, P., & Abarbanel, H. D. I. (1991). Computing the Lyapunov spectrum of a dynamical system from an observed time series. *Physical Review A*, 43(6), 2787–2806.

Casdagli, M. (1989). Nonlinear prediction of chaotic time series. *Physica D*, 35, 335–356.

Collins, J. J., & De Luca, C. J. (1993). Open–loop and closed–loop control of posture: A random–walk analysis of center–of–pressure trajectories. *Experimental Brain Research*, 95, 308–318.

Collins, J. J., & De Luca, C. J. (1994). Random walking during quiet standing. *Physical Review Letters*, 73(5), 764–767.

Collins, J. J., & De Luca, C. J. (1995). Upright, correlated random walks: A statistical–biomechanics approach to the human postural control system. *Chaos*, 5(1), 57–63.

- Dubuc, B., Quiniou, J., Roques-Carmes, C., Tricot, C., & Zucker, S. (1989). Evaluating the fractal dimension of profiles. *Physical Review A*, *39*(3), 1500–1512.
- Dvořák, I., & Klaschka, J. (1990). Modification of the Grassberger-Procaccia algorithm for estimating the correlation exponent of chaotic systems with high embedding dimension. *Physics Letters A*, *145*(5), 225–231.
- Eckmann, J.-P., Kamphorst, S. O., Ruelle, D., & Ciliberto, S. (1986). Liapunov exponents from time series. *Physical Review A*, *34*(6), 4971–4979.
- Eckmann, J.-P., & Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, *57*(3), 617–656.
- Efron, B. (1979a). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*, 1–26.
- Efron, B. (1979b). Computers and the theory of statistics: Thinking the unthinkable. *SIAM Review*, *21*, 460–480.
- Farmer, J., & Sidorowich, J. (1987). Predicting chaotic time series. *Physical Review Letters*, *59*(8), 845–848.
- Farrell, M. E., Passamante, A., & Hediger, T. (1995). Comparing a nearest-neighbor estimator of local attractor dimensions for noisy data to the correlation dimension. *Physical Review A*, *41*(12), 6591–6595.

- Fraser, A., & Swinney, H. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, *33*, 1134–1140.
- Fredkin, D. R., & Rice, J. A. (1995). Method of false nearest neighbors: A cautionary note. *Physical Review E*, *51*(4), 2950–2954.
- Gabor, D., Wilby, F., & Woodcock, R. (1960). A universal non-linear filter, predictor and simulator which optimizes itself by a learning process. *Proceedings of the IEE*(108B), 422–438.
- Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C., & Schreiber, T. (1993). On noise reduction methods for chaotic data. *Chaos*, *3*, 127–140.
- Grassberger, P., & Procaccia, I. (1983a). Characterization of strange attractors. *Physical Review Letters*, *50*, 346–349.
- Grassberger, P., & Procaccia, I. (1983b). Measuring the strangeness of strange attractors. *Physica D*, *9*, 189–208.
- Hammel, S. M. (1990). A noise reduction method for chaotic systems. *Physics Letters A*, *148*, 421–428.
- Hegger, R., & Schreiber, T. (1992). A noise reduction method for multivariate time series. *Physics Letters A*, *170*, 305.
- Hentschel, H., & Procaccia, I. (1983). The infinite number of generalized dimensions of fractals and strange attractors. *Physica D*, *8*, 435–444.

- Horn, J. L. (1965). *Fluid and crystallized intelligence*. Unpublished doctoral dissertation, University of Illinois, Urbana. (Unpublished doctoral dissertation)
- Howard, I. P. (1986). The perception of posture, self motion, and the visual vertical. In K. R. Boff, L. Cauffman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. sensory processes and perception* (pp. 1–62). New York: Wiley.
- Kantz, H., & Schreiber, T. (1995). Dimension estimates and physiological data. *Chaos*, 5, 143–154.
- Kennel, M. B., & Abarbanel, H. D. I. (1995). *False neighbors and false strands: A reliable minimum embedding dimension algorithm*. (Unpublished Manuscript)
- Kennel, M. B., Brown, R., & Abarbanel, H. D. I. (1992). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6), 3403–3411.
- Kennel, M. B., & Isabelle, S. (1992). Method to distinguish possible chaos from colored noise and to determine embedding parameters. *Physical Review A*, 46(6), 3111–3118.
- Kostelich, E. J. (1992). Problems in estimating dynamics from data. *Physica D*, 58, 138–152.

- Lee, D. N., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception & Psychophysics*, *15*, 529–532.
- Liebert, W., & Schuster, H. (1989). Proper choice of the time delay for the analysis of chaotic time series. *Physics Letters A*, *142*(2), 107–111.
- Liebovitch, L. S., & Toth, T. (1989). A fast algorithm to determine fractal dimensions by box counting. *Physics Letters A*, *141*(8,9), 386–390.
- Lishman, J. R., & Lee, D. N. (1973). The autonomy of visual kinaesthesia. *Perception*, *2*, 287–294.
- Lyapunov, M. A. (1949). General problems in the stability of motion (translation from russian original of 1893 in communications of the society of mathematics, kharkow). *Annals of Mathematical Studies*, *17*.
- Mandelbrot, B. B. (1967). How long is the coastline of Britain? Statistical self-similarity and fractional dimension. *Science*, *156*, 636–638.
- Mandelbrot, B. B. (1977). *Fractals: Form, chance and dimension*. San Francisco: W. H. Freeman & Sons.
- McArdle, J. J., & Boker, S. M. (1990). *Ramopath*. Hillsdale, NJ: Lawrence Erlbaum.
- McArdle, J. J., & McDonald, R. P. (1984). Some algebraic properties of the Reticular Action Model for moment structures. *British Journal of Mathematical and Statistical Psychology*, *87*, 234–251.

- Neale, M. C. (1994). *Mx: Statistical modeling*. (Box 710 MCV, Richmond, VA 223298: Department of Psychiatry. 2nd Edition)
- Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. S. Palermo (Eds.), *Visions of aesthetics, the environment, and development: The legacy of Joachim F. Wohlwill* (pp. 213–240). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Nesselroade, J. R., & Boker, S. M. (1994). Assessing constancy and change. In T. F. Heatherton & J. L. Weinberger (Eds.), *Can personality change?* (pp. 121–147). Washington, DC: American Psychological Association.
- Noakes, L. (1991). The Takens embedding theorem. *International Journal of Bifurcation and Chaos*, 4(1), 867–872.
- Osborne, A. R., Kirwin, A. D., Provenzale, A., & Bergamansco, L. (1986). A search for chaotic behavior in large and mesoscale motions in the pacific ocean. *Physica D*, 23, 75.
- Packard, N. H., Crutchfield, J. P., Farmer, J. D., & Shaw, R. S. (1980). Geometry from a time series. *Physical Review Letters*, 45(9), 712–716.
- Paluš, M. (1995). *Detecting nonlinearity in multivariate time series* (Tech. Rep. Nos. 95–07–059). Santa Fe Institute.

- Peitgen, H.-O., Jürgens, H., & Saupe, D. (1992). *Chaos and fractals: New frontiers of science*. Berlin: Springer-Verlag.
- Pompe, B. (1993). Measuring statistical dependences in a time series. *Journal of Statistical Physics*, 73, 587.
- Pompe, B. (1994). On some entropy methods in data analysis. *Chaos, Solitons & Fractals*, 4(1), 83–96.
- Pompe, B. (1995). A tool to measure dependencies in data sequences. In M. Cox & F. Pavese (Eds.), *Proceedings of the international euroconference on advanced mathematical tools in meteorology*. Singapore: World Scientific.
- Prichard, D. (1994). The correlation dimension of differenced data. *Physics Letters A*, 191(7), 245–250.
- Prichard, D., & Theiler, J. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*, 73(7), 951–954.
- Resnikoff, H. L. (1989). *The illusion of reality*. New York: Springer-Verlag.
- Rosenstein, M. T., Collins, J. J., & De Luca, C. J. (1993). A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D*, 65, 117–134.

- Ruelle, D. (1990). Deterministic chaos: the science and the fiction. *Proceedings of the Royal Society London A*, *427*, 241–248.
- Ruelle, D., & Takens, F. (1971). On the nature of turbulence. *Communications of Mathematical Physics*, *20*, 167–192.
- Sauer, T., Yorke, J., & Casdagli, M. (1991). Embedology. *Journal of Statistical Physics*, *65*(3,4), 95–116.
- Schiff, S. J., & Chang, T. (1992). Differentiation of linearly correlated noise from chaos in a biologic system using surrogate data. *Biological Cybernetics*, *67*(5), 387–93.
- Schreiber, T. (1993). Determination of the noise level of chaotic time series. *Physical Review E*, *48*, 13–16.
- Schreiber, T. (1995). Efficient neighbor searching in nonlinear time series analysis. *International Journal of Bifurcation and Chaos*, *5*(2), 349–358.
- Schreiber, T., & Grassberger, P. (1991). A simple noise-reduction method for real data. *Physics Letters A*, *160*, 411–418.
- Schreiber, T., & Kantz, H. (1995). Noise in chaotic data: diagnosis and treatment. *Chaos*, *5*, 133–142.
- Schreiber, T., & Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, *77*, 635.

Schroeder, M. (1991). *Fractals, chaos and power laws*. New York: W.H. Freeman.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: The University of Illinois Press.

Sugihara, G., & May, R. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, *334*, 734–741.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence, warwick 1980: Proceedings of a symposium held at the university of warwick* (pp. 366–381). Berlin: Springer-Verlag.

Theiler, J., Eubank, S., Longtin, A., & Galdrikian, B. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, *58*, 77–94.

Vastano, J. A., & Kostelich, E. J. (1986). Comparison of algorithms for determining Lyapunov exponents from experimental data. In G. Mayer-Kress (Ed.), *Dimensions and entropies in chaotic systems: Quantification of behavior*. New York: Springer Verlag.

Water, W. van der, & Schram, P. (1988). Generalized dimensions from near-neighbor information. *Physical Review A*, *37*(8), 3118–3125.

West, B. J. (1985). *An essay on the importance of being nonlinear*. Berlin: Springer-Verlag.

- Whitney, H. (1936). Differentiable manifolds. *Annals of Mathematics*, *37*, 645–680.
- Wolf, A., Swift, J. B., Swinney, H. L., & Vastano, J. A. (1985). Determining Lyapunov exponents from a time series. *Physica D*, *16*, 285–317.
- Young, L.-S. (1982). Dimension, entropy and Liapunov exponents. *Ergodic Theory of Dynamical Systems*, *2*, 109.
- Zeng, X., Eykholt, R., & Pielke, R. A. (1993). Estimating the Lyapunov–exponent spectrum from short time series of low precision. *Physical Review Letters*, *66*(25), 3229–3232.