Penultimate Version of a paper accepted in the Journal of Mathematical Psychology

Book Review

An Introduction to the Art

R. A. Heath. **Nonlinear Dynamics: Techniques and Applications in Psychology**, Earlbaum: Mahwah, NJ, 2000, ix + 379 pp., $79.95, ISBN 0-8058-3199-1 (hard cover), and $36.00, ISBN 0—8058-3200-9 (paper).

Richard Heath is a Professor in the Discipline of Psychology at the University of Sunderland, England. He obtained his B.Sc. (Hons.) at the University of Newcastle, Australia in 1970 and a Ph.D. in Psychology as a Commonwealth Scholar with Prof. Stephen Link at McMaster University, Canada, in 1976. Heath's research has emphasized the role of new mathematical and computational techniques to the study and modeling of complex and nonlinear aspects of cognition. His experimental work has examined signal detection, categorization, memory, fatigue, handwriting and the detection of behavior change. He has also developed nonlinear system identification models of attention and interference, non-stationary versions of the random walk model of choice response time, and adaptive, novelty sensitive models of human memory.

Andrew Heathcote is an Associate Professor in the School of Behavioural Sciences, University of Newcastle, Australia. He obtained a B.Sc. (Hons.) at the University of Tasmania in 1984, a Ph.D. in Psychology as a Commonwealth Scholar with Prof. D. Mewhort at Queen's University, Canada, in 1991. His research interests include nonlinear analysis and modeling of response time, skill acquisition and recognition memory.

Reviewed by Andrew Heathcote

Richard Heath's book was conceived during a sabbatical visit to Purdue University in 1995 when Professor James Townsend suggested the need for a guide to nonlinear dynamical analysis (NDA) for experimental psychologists. The outcome, Nonlinear Dynamics: Techniques and Applications for Psychology, integrates a large and rapidly evolving literature, providing 385 references and a glossary of 117 terms. Around half of the references come from non-psychological sources, including physics, biology, chemistry and physiology journals, as well as sources dedicated exclusively to the study of nonlinear dynamics.

The approach taken in Heath (2000) is informal, using examples to develop practical skills, such as how to use NDA software, and how to interpret the results of NDA analyses. For the most part, the analyses are performed with freely available software, such as the excellent suite of programs provided by the TISEAN project (http://www.mpipks-dresden.mpg.de/~tisean; Hegger, Kantz & Schreiber, 1999). As indicated by the sub-title of the book, it emphasizes the application of NDA to psychological data, particularly behavioral measures. Heath does explore NDA for physiological measures such as EEG and heart inter-beat interval, which have been the primary subject of most applications in psychology and medicine, but mainly focuses on his own groundbreaking applications of NDA to behavioral measures such as response time, movement, and prediction. A second strand of the book describes computational and mathematical models of nonlinear psychological phenomena, with an emphasis on Heath's own original contributions.

The book is written is a clear and readable style that makes it suitable as a text for a lecture or laboratory course aimed at training senior undergraduate or graduate psychologists in NDA. Example data sets, software, and useful links, as well as errata, are provided on an accompanying web site. In general, only a basic knowledge of statistics and computing is assumed, although some of the chapters presenting Heath's original theoretical work assume a higher level of mathematical knowledge. Rather than being a "cookbook", Heath embeds his descriptions of NDA in a framework of real psychological data and relates the results to mathematical and computational models. In practice, NDA of empirical data is still as much an art as a science. The many examples in Heath's book, usually with computational methods spelled out and example data sets provided via the book's web page, help the novice analyst to learn to recognize important patterns of results, gain experience in selecting parameters, and apprehend the limitations in conclusions that can be drawn from results.

Given the breadth of areas covered, and the rapid ongoing development of these areas, some supplementary reading may be beneficial. Readers requiring a structured development and mathematical derivations should consult Abarbanel (1996) or Kantz and Schreiber (1997), the latter being particularly relevant for empirical researchers as it stresses the need to test for nonlinear dynamics rather than assuming it. Kaplan and Glass (1995) and Nusse and Yorke (1997) provide accessible introductions to the theory of nonlinear dynamical systems, which is not covered in any detail by Heath (2000). Readers particularly interested in EEG and other physiological measures should consult Gregson and Pressing (2000) and Pritchard and Duke (1992), and those interested in the dynamics of disease processes should consult Belair, Glass, van der Heiden and Milton

(1995). The next section of this review examines why NDA may be useful for psychologists, and establishes some basic concepts necessary for the following detailed examination of the contents of Heath (2000).

Why Nonlinear Dynamics?

Heath (2000) addresses two interrelated tensions for mathematical psychology that were identified by Luce (1995): dynamic versus static approaches, and structure versus noise in psychological data. Luce speculated that psychology, like other sciences, would increasingly use dynamic models as it matures. Low dimensional nonlinear dynamical or "chaotic" models of behavior are often intuitively appealing: "… qualitative facts about many nonlinear systems strike a receptive chord in behavioral and social scientists because much of the behavior under their scrutiny seems to undergo radical transitions and often has to be described as chaotic." (Luce, 1995, p. 16). Figure 1 is an example of mathematical chaos, a time series from the Lorenz (1969) equations[1], which will be used to illustrate nonlinear dynamics in this review. It consists of two sets of unstable periodic orbits, with transitions between the sets occurring at apparently irregular intervals.

Insert Figure 1 about here

Figure 1 illustrates Luce's (1995) interrelated tension, noise versus structure in data: "… the findings of the past 10 to 15 years about nonlinear dynamic systems call into question whether the actual source of the noise is randomness or ill-understood dynamics." (p.24). The unstable orbits of the Lorenz series might easily appear to be noise, especially if measurement adds genuine noise. Luce believed that accounting for apparent noise with nonlinear dynamic models may have a profound impact, but was cautious about the future of nonlinear dynamics in psychology, viewing it as no more than an interesting speculation at that time. He was also pessimistic about effective testing of this speculation because of the effects of measurement noise and the lack of principled nonlinear dynamical models. Heath (2000) reviews and develops theories and techniques that provide at least the beginnings of a way to address Luce's concerns.

Both systems of continuous differential equations and discrete recursive maps can generate the irregular but deterministic behavior called "chaos". Chaos can range from low dimensional, the primary concern of Heath's (2000) book, to high dimensional. Low dimensional chaos can, in principle, be distinguished from noise; however, obtaining data of sufficient quality can be difficult. The data time series must be long and stationary, and sampled at regular time intervals. Early NDA techniques were developed for the physical sciences where long stationary data series are more readily available. In the behavioral and social sciences, however, processes such as learning and fatigue, and an inability to control all influential factors, make such series difficult to obtain. Genuine noise, particularly linearly autocorrelated noise, is particularly problematic as it can be mistaken for chaos by some NDA techniques.

Fourier analysis provides a traditional approach to oscillatory signals in time series, allowing dominant frequencies to be identified, and broadband noise to be removed through band-pass filtering. However chaotic signals can be broadband, and so frequency domain filters are ineffective. Instead of a Fourier decomposition, which transforms from temporal to frequency domains, almost all NDA techniques rely on a representation that transforms temporal information into a geometrical representation, delay embedding. An m dimensional delay embedding converts a time series, $x(t)$, to a set of m dimensional points (x(t+δ), x(t+2δ), … x(t+mδ)), where δ is the time delay between samples.

A celebrated theorem due to Takens (1981) shows that a one to one image of the d dimensional set of points visited by a stationary dynamical system (its attractor) can be reconstructed from an embedding in a space with dimension $m > 2d$. For the Lorenz system, for example, $d \approx 2.05$, with the attractor being a fractal subset of the space defined by the three Lorenz variables (i.e., (x, y, z) of which only x is shown in Figure 1). The fractal nature of these sets for chaotic systems has led to them being described as "strange attractors". Since Takens' initial result, "embedology" has been an active area of research. Of most interest for applications, Sauer, Yorke and Casdagli (1991) showed that more general schemes than simple delay coordinates are allowed, including delay embeddings of series transformed by singular value decomposition and geometric filtering (Grassberger, Hegger, Kantz, Schaffrath & Schreiber, 1993).

Stationarity is essential in establishing an embedding. For a stochastic process, x(t), stationarity can be defined as invariance of all finite-dimensional joint distribution functions over shifts on the temporal dimension (Rao & Gabr, 1984). In the context of NDA, Casdagli (1997) defined nonstationarity for practical purposes as: "… a time series $x_1, x_2, \ldots, x_N$ is nonstationary if, for low m, there are variations in the estimated joint distribution of $x_i, x_{i+1}, \ldots, x_{i+m-1}$ that occur on time scales of order N" (p. 12). Given stationarity, NDA can recover aspects of the dynamics using geometrical analyses of the embedded set. The analyses often involve estimating the properties of local "neighborhoods" of points, where a neighborhood is defined on a measure in the embedding space. If such analyses aggregate local statistics to estimate global properties of the attractor, they strongly rely on the assumption of stationarity. Aggregation is particularly important to counter the effects of measurement noise.

Two notes of caution should be sounded here. Even where stationarity holds, the finite number of points in a sample may mean that a particular data series is not suitable for NDA because all regions of the dynamics, and hence the full joint distribution, cannot be estimated. The problem is particularly evident with intermittent chaos or noisy periodic behavior when few intermittent events or periods are sampled. Second, Casdagli, Eubank, Farmer and Gibson (1991) showed that even arbitrarily small amounts of noise could destroy the embedding property in some circumstances. As Schreiber (1999) says: "bold interpretations of Takens' theorem, for example, that we can recover the full dynamics of the human body from a recording of a single variable, [are] not only in contradiction with common sense but also disproven by mathematical arguments." (p.17). This is not to deny that delay coordinates provide a generally useful representation. For example, delay coordinate representations are fundamental to methods of predicting nonlinear time series with both deterministic and stochastic dynamical structure (Weigend & Gershenfeld, 1993). Prediction has also become a core technique for NDA because it remains useful when noise levels are high.

Insert Figure 2 about here

Figure 2 illustrates the relatively smooth nonlinear dynamics that underlie the irregular behavior of the Lorenz time series, by plotting each value against an estimate of its first derivative, obtained by successive (lag one) differences. This phase plot shows that the Lorenz series tends to change quickly near the center of the positive and negative sets of orbits, and more slowly in the region between sets, and at the positive and negative extremes. The spikes to the left and right of the main body of Figures 2a and 2b are produced by very large changes, one large decrease for the positive set and three large increases for the negative set. The phase plot in Figure 2a shows that the initial 2000 observations in the series have different dynamics to the remainder of the series (Figure 2b). The series first exhibits an unusually large decrease then oscillates for an extended period around the center of the negative set. Once the series has settled into its attractor only large <u>increases</u> occur and longer oscillations occur only around the center of the positive set. Such initial nonstationarity is common in dynamical systems, which may take some time to settle into their attractor if started from a point off the attractor.

Because delay representations underpin most NDA techniques, the user must select an embedding dimension and delay, and often other parameters. Parameter selection is a difficult and uncertain task because the best choice depends both on the data set and the technique. A number of methods have been devised to select these parameters for a given data set, but these methods can also require parameters to be set, so bootstrapping an analysis can be difficult. Further, both dynamical analysis and parameter selection methods can fail when noise is present. Fortunately, Heath's book summarizes the field at a point when the applied investigator can make use of a readily available toolbox of analysis programs. However, it should be emphasized from the outset that is not possible to provide a definitive step-by-step guide that covers NDA for all types of series. The very complexity that makes nonlinear dynamics attractive for psychological modeling means that no one approach is optimal for all chaotic series. As described in the next section, Heath (2000) examines a broad range of approaches to NDA, and illustrates the selection of parameters using a variety of example data sets.

Techniques and Applications in Psychology

Chapter 1 illustrates linear dynamical analyses using ARMA (AR = autoregressive, MA = moving average) modeling (Box & Jenkins, 1976). ARMA modeling is a necessary preliminary to NDA, and is the default method for modeling time series, with a widely available set of efficient and robust tools. Autoregressive models create future values from linear combinations of past values and noise, producing linear autocorrelation between past and future observations. Most chaotic models have the same type of recursive dynamics, except that the combination also contains nonlinear components, and so chaos almost always produces strong linear autocorrelation. In Chapter 1, Heath demonstrates that averaging across time series generated by autoregressions with different parameters, which is analogous to averaging across time series for different subjects, is unwise when analyzing dynamical structure. This point underlines one of Luce's (1995) main concerns about the future of dynamical analysis; that an inability to average makes noise reduction difficult, and so denies access to the details of dynamics underlying noisy behavioral time series.

ARMA analysis selects an AR(12) model for the series in Figure 1a, which accounts for more than 99% of its variance. However, the residuals are patterned, as the linear model does not account for all of the increasing magnitude of oscillations within an unstable orbit, and four small sets of very large deviations occur, corresponding to the large differences evident in Figure 2. Such large failures, and fits requiring large numbers of parameters (i.e., 12 as opposed to the three parameters of the Lorenz equations), are characteristic of linear ARMA models of chaotic time series.

A confusing aspect of Chapter 1 for experienced ARMA users is Heath's failure to remove the strong mean trend evident in his first example data set before applying

ARMA analysis. De-trending, through subtracting a regression estimate of the mean function or transformations such as successive differences, is usually applied because ARMA analysis assumes second order or “weak” stationarity: constancy of the mean, variance and auto-covariance. The Lorenz series displays nonstationarity up to around 2000 observations, evident in an increase in the local mean of the series (the thick line in Figure 1a). However, the local mean continues to fluctuate appreciably throughout the series, despite a broad averaging window of 1000 observations. These later fluctuations do not reflect nonstationarity, but rather the nature of the underlying dynamical system, which spends varying amounts of time in the upper and lower orbits. Similar observations can be made about the local variance of the series, and neither the local means nor variances are very discriminating as detectors of nonstationarity in this case.

While de-trending is standard for ARMA analysis, it must be applied cautiously in the context of NDA. For example, removal of the local mean averaged at the scale illustrated in Figure 1a would remove information about the underlying dynamics. NDA should definitely not be carried out on residuals from ARMA analysis (Theiler & Eubank, 1993). Chaotic dynamics often produce linear components, and ARMA models with large number of parameters can fit some variation due to nonlinear components, so that residuals do not contain the information necessary for NDA (often called “bleaching”). However, tests for nonlinear structure do rely heavily on comparison with linear models. Instead of examining residuals, an experimental series is compared to “surrogate” series, formed by randomizing the experimental series while maintaining the stationary linear structure required by a null hypothesis (Theiler, Eubank, Longtin, Galdrikian & Farmer, 1992). Heath examines the topic of surrogate testing in Chapter 7.

Chapter 2 examines another linear technique, Kalman filtering, which can be used for the detection of violations of weak stationarity in a time series. Detection of nonstationarity is of interest not only as a check on departures from the assumptions of NDA, but also as an indicator of change in the mechanisms generating behavior. Heath uses the Kalman filter mainly as a device for detecting change on-line, in keeping with Heath (1984). It is important to note that the Kalman filter is optimally sensitive only to violations of weak stationarity, particularly changes in linear autocorrelation, so it may spuriously report nonstationarity in some stationary nonlinear series, or fail to detect nonstationarity in nonlinear aspects of the time series. For the Lorenz series, the Kalman filter programs accompanying the book consistently detected changes in the local autocorrelations at times corresponding to the large positive differences in Figure 2[2].

Chapter 3 examines nonlinear system identification (Marmarelis & Marmarelis, 1978), which applies Volterra-Wiener methods to produce global models of linear and nonlinear dynamics. Although predating much of the recent work on nonlinear dynamics, this technique fits multivariate models of a polynomial recursive form that can generate chaos. It is used by Heath to model choice RT data and selective attention effects in a dual task paradigm. In Chapter 4 Heath develops a new gradient descent search algorithm that obtains good fits of the nonlinear system identification model, with the starting point for search determined by the approximations described in Chapter 3. While univariate nonlinear dynamical modeling is a relatively mature area, much further work is needed on the multivariate case. Chapters 3 and 4 make both original technical and psychological contributions to this development.

Chapter 5 discusses graphical methods, in particular introducing delay plots and the important idea of recurrence. A delay plot graphs a time series, x(t), against a delayed version of itself, x(t+δ), where δ is the delay. Periodic and quasi-periodic attractors display exact recurrence, so that they return to the same state or states at regular intervals. Hence, at appropriately chosen delays, the series can be represented by a single point or small set of points in a delay plot. Chaotic systems do not display exact recurrence because they contain unstable periodic orbits, but delay plots still reveal regularities and can be useful in understanding the qualitative dynamics of a system. Delay plots and phase plots (e.g., Figure 2) contain the same information. For example, the lag-one derivative estimate for a point in a lag-one delay plot is just the difference between vertical and horizontal distances from the right diagonal. The phase plot has the advantage of a direct interpretation in terms of process dynamics, and can be extended using estimates of higher order derivatives.

Figure 2 shows that an appropriately chosen lag provides a smooth representation of temporal structure in Lorenz series. When the delay is short (e.g., Figure 2b) orbits are too closely packed to be discriminable, whereas when the delay is long, detail may be lost in some regions. Figure 2c, with a lag of 12, provides a good compromise, but it is important to note that any given lag may not provide the best detail of all regions of the dynamics. For example, a lag of 17 provides better resolution of the central region of the dynamics but worse resolution of the extremes than Figure 2c (Hegger et al.'s, 1999, Figure 1, provides an even clearer example of this phenomena with real data). The embedding theorems should apply for any delay; however, most NDA techniques require specification of an appropriate delay in order to avoid the redundancy evident in Figure

2b and the loss of structure produced by long delays. Series of phase or delay plots can be used to choose the lag producing the clearest structure.

For the Lorenz data, a delay of 12 provided the clearest overall visual structure in both phase and delay plots, agreeing with the order of the AR model for this time series. Other measures based on autocorrelation, such as the correlation time (the lag at which autocorrelation drops to 1/e of its initial value) or the first zero of the autocorrelation function, are sometimes recommended for choosing a delay. However these estimates were not useful for the Lorenz series. In Chapter 6, Heath favors the use of time delayed mutual information (i.e., the information shared by lagged versions of a time series) for estimating delay (Fraser & Swinney, 1986). Mutual information accounts for both linear and nonlinear structure, with the delay being set to its first minimum as a function of lag. For the stationary ($t > 2000$) Lorenz series, the resulting estimate was 17, somewhat longer than the linearly derived estimate of 12, reflecting the longer time scale of the nonlinear interactions. As with phase plots, little difference was observed in analyses of the Lorenz series when delays between 12 and 17 were used.

The main topic of Chapter 5 is recurrence plots, which were originally proposed by Eckmann, Oliffson-Kamphorst and Ruelle (1987). They are constructed by plotting points, on a grid defined by two time axes, where recurrence nearly occurs in an m dimensional embedding. Recurrence was defined as being a k-th nearest neighbor in the original version, but a definition based on a minimum distance is more often used (Koebbe & Mayer-Kress, 1991). Recurrence plots produce intriguing patterns that can be useful in detecting non-stationarity, but general guidelines for how to interpret the patterns are difficult to formulate. Heath's emphasis in Chapter 5 is on a number of statistical indices used to quantify recurrence plots, so called "recurrence quantification analysis" or RQA (Trulla, Giuliani, Zbilut & Webber, 1996; Webber & Zbilut, 1994). Some of the indices are related to measures derived from dynamical systems theory, such as the correlation sum and the maximum Lyapunov exponent, whereas others are more ad-hoc.

RQA has strong advocates who claim it "does not require assumptions about stationarity, length or noise" (Thomasson, Hoeppner, Webber & Zbilut, 2001, p. 94). However, the RQA indices are formed by aggregating local statistics across a time series, and aggregation does require stationarity, at least if any rigorous meaning is to be attached to the aggregate values. Both series length and noise levels are important in obtaining precise estimates, and, as discussed previously, an embedding may not even be possible due to noise, so it is difficult so see how noise and length are irrelevant to RQA. Although RQA does provide a summary of a recurrence plot that can prove useful, the lack of a rigorous derivation makes it is difficult to know how to interpret these indices (cf. Schreiber, 1999). Despite these drawbacks, RQA continues to develop (e.g., Zbilut, Giuliani & Webber, 1998) and easy-to-use RQA software is freely available (e.g., Eugene Kononov's VRA software, available at http://pw1.netcom.com/~eugenek). Casdagli (1997) provides an alternative and more rigorous approach to recurrence plots.

Chapter 6 deals with the estimation of quantitative indices derived from nonlinear dynamical system theory. As these indices are based on topological regularities in an embedding, an embedding dimension must be specified. According to Takens' (1981) theorem, the minimum embedding dimension is given by the smallest integer $m > 2d$, where d is the (possibly fractional) dimension of the attractor that generated the time series. The dimension of an attractor depends on the type of nonlinear dynamics that generates it. Simple periodic attractors (e.g., sine waves) have a dimension of one, and quasi-periodic attractors, such as tori (e.g., Ptolemaic epicycles), have higher integer dimensions. Chaotic dynamics have "strange" attractors with non-integer dimensions, because they fill the phase space in a fractal manner with an invariant distribution of points on the attractor at different length scales. For example, the Lorenz equations operate in a three dimensional space, but the Lorenz attractor has dimension, $d \approx 2.05$. For the example analyses of real data in Chapter 6, Heath favors the False Nearest Neighbour method (Kennel, Brown & Abarbanel, 1992) to determine a minimum embedding dimension.

Heath (2000) discusses a number of different attractor dimensionality (d) estimates based on both the correlation sum and information theory measures. For example, dimensionality may be estimated using D1, the information dimension, and the dynamics of two series may be compared using Kullback entropy (Schreiber, 1999). The most commonly used estimate of dimensionality, D2 (Grassberger & Procaccia, 1983), uses the fractal property of strange attractors to obtain an estimate. D2 is based on the correlation sum, $C(\varepsilon)$, which measures the proportion of embedded points that fall within a given distance (ε) of each other. The correlation sum measures recurrence at a given distance or scale, so the correlation sum is just the proportion of points marked as recurrent at a given distance on a recurrence plot.

The correlation dimension summarizes the way the correlation sum changes as a function of distance. It can be estimated from the exponent of a power law relationship between the correlation sum and distance. The exponent can be estimated by the slope of a log(C(ε))-log(ε) plot, and because the power law relationship never applies for all distances in a finite series (technically the dimension of a finite series is zero), the slope is usually assessed locally, that is for a restricted range of distances. For chaotic systems the slope increases with m, but then remains constant once a proper embedding is achieved. The constant slope provides the D2 estimate of dimensionality. In contrast, white noise, which is constituted of a series of independent random values, fills the phase space uniformly no matter what its dimension and so produces increasing slope estimates.

Initial applications of NDA to experimental data used finite D2 estimates as evidence for chaos. However, a finite D2 is a necessary but not sufficient condition for chaos. Colored noises also produce finite D2 estimates. As early as 1989, Osborne and Provenzale showed that an embedded random walk has a dimension of two. Provenzale, Smith, Vio and Murante (1992) provide an excellent discussion of these problems and provide examples of finite D2 estimates for both linear and nonlinear autoregressive models. Subsequently, it was determined that much supposed evidence for chaos relying on finite estimated dimensionality was spurious (see Theiler & Rapp, 1996).

D2 and other dimensionality estimates are best used as a method of characterizing rather than identifying nonlinear dynamics. Even where nonlinear dynamics are present, measurement noise is particularly detrimental to dimensionality estimation, as it smears out the fine details of an attractor and so denies access to short scales. As a result of noise, accurate absolute dimensionality estimates are almost impossible in practice. Relative measurements of dimensionality may, however, remain useful[3].

An important technical issue in D2 estimation, and most other estimates based on the correlation sum, is that they assume that pairs of points are drawn randomly and independently according to the scale invariant measure of the attractor. Independence will not apply for points occurring close in time, and if such points are included a spuriously low estimate of D2 results. To avoid the problem, points closer than some minimum time can be excluded from the correlation sum (Grassberger, 1987; Theiler, 1990). The number of points excluded, w (often called the "Theiler window"), can be chosen generously, up to say 10% of the series length, as the number of points available to the D2 algorithm increases as the square of the series length. Heath discusses two methods of determining the Theiler window, either setting it to three times the correlation time or using a "Space-Time Separation Plot" (Provenzale, et al., 1992).

The space-time plot is related to the correlation sum and the recurrence plot in that it shows equal probability contours for the distribution of distances between pairs of points as a function of time. For a chaotic series the contours initially rise then oscillate around a constant value, whereas for colored noise they continue to rise. The Theiler window is set at a time beyond which the constant behavior is operating. For the stationary section of the Lorenz series a value of 100 worked well, which approximately equals the estimate from three times the correlation time (3 x 31) suggested by Heath. Values in the range 50-300 were found to have a similar effect.

Figure 3 shows, for the stationary Lorenz series and a Theiler window of 100, local slope estimates of $\log(C(\varepsilon)) \sim \log(\varepsilon)$ as a function of $\log(\varepsilon)$ for 1 to 10 embedding dimensions. The slope in a "scaling region", a range of distances in the middle of the plot with a constant slope that is the same for all larger embedding dimensions, provides an estimate of dimensionality. Heath uses plots of $\log(C(\varepsilon)) \sim \log(\varepsilon)$ directly for estimating dimensionality. The local slope plot is an alternative representation that is useful in that it makes evident the extent of, and deviations from, scaling behavior. Dimensionality cannot be estimated unless scaling behavior occurs; the authors of the TISEAN programs (Hegger et al., 1999) emphasize this point so strongly that they purposely do not provide any automatic estimate of dimensionality, only outputs for constructing plots. In Figure 3 the scaling region is achieved at an embedding dimension of 5 and only slightly overestimates the true value, despite the relatively short series.

Insert Figure 3 about here

The pointwise dimension measure, PD_2i (Skinner, Molnar, Vybiral & Mitra, 1992), provides an alternative measure of dimensionality that was favored in Gregson and Pressing's (2000) review of NDA for physiological measures. PD_2i uses a set of rules to determine if a scaling region is present and explicitly reports when scaling does not hold. Estimates are provided for contiguous sections of a time series, so nonstationarity can be detected by changes in estimated values. Where pointwise estimates are defined and relatively consistent across the time series they can be averaged to give a global estimate of dimensionality. However, Heath reports analyses of noisy experimental data where PD_2i was not very useful because it was often undefined.

As well as measures based on fractal topology, Chapter 6 examines techniques based on the defining characteristic of chaos: sensitive dependence to initial conditions. In chaotic systems nearby trajectories diverge exponentially, so that initially close points quickly move far apart. Divergence occurs at a rate described by the Lyapunov spectrum. Globally stable chaotic systems are bounded overall so the sum of their Lyapunov exponents is negative, but trajectories diverge in one or more dimensions at rates determined by the set of positive exponents. Heath (2000) describes a method of estimating the full Lyapunov spectrum using neural network based software, but notes that often noise makes this difficult. He achieves more success by estimating only the maximum Lyapunov exponent, with a positive maximum exponent indicating chaos.

The False Nearest Neighbour method (Kennel et al., 1992), used to determine a minimum embedding dimension, is also based on the idea of examining the divergence of neighboring points. For a range of embedding dimensions, the distance in the future (usually just one step) between each data point and its nearest neighbor is compared. If the ratio of the distance apart after one step to the original distance exceeds a criterion, the point is declared a false neighbor. The process is repeated for a range of embedding dimensions and the percentage of false neighbors plotted. The dimension at which percentage first approaches zero provides an estimate of the minimum embedding dimension. A Theiler window must be specified as the false nearest neighbor technique makes similar assumptions to the dimensionality estimation techniques. A minimum embedding dimension can also be estimated from a D2 plot by the minimum number of dimensions necessary to attain a scaling region.

Chapter 7 examines three relatively recent developments in NDA: nonlinear prediction, geometric filters and inferential tests for nonlinearity based on surrogate series. These techniques are important for applications because they allow NDA to be applied to noisy data series. Sugihara and May (1990) suggested predictability could be used to measure the nonlinear dynamics in a time series. Although the deterministic nature of chaotic series makes them predictable in the short term, small differences due to measurement error are rapidly magnified by sensitive dependence, so that they cannot be predicted in the long term. White noise, in contrast, is not predictable on any time scale.

Sugihara and May suggested that decreasing predictability with time provides evidence for chaos. However, as with other NDA measures, stochastic temporal dependencies (e.g., colored spectra produced by linear and nonlinear autoregressive processes) can confound results as they also cause a gradual decrease in predictability.

To achieve prediction without knowledge of the underlying determining equations, Sugihara and May (1990) used a principle suggested by Lorenz (1969); similar present states in a deterministic system should evolve into similar future states. To compensate for the effects of noise, the evolution of a point in an embedding is predicted by the aggregate evolution of a set of neighboring points, rather than just the evolution of the single most similar point (Lorenz's "analogue"). Local aggregation can range from simple averaging to more sophisticated combinations, and the size of the neighborhood varied to match the level of noise. Barahona and Poon (1996) suggested using global polynomial prediction, based on Volterra-Wiener methods similar to those used by Heath in Chapters 3 and 4, for detecting chaos in short, noisy, time series. However, Schreiber and Schmitz (1997) report that when noise is high, local average techniques are better for this purpose.

Estimators for most quantitative indices assume nonlinear dynamics are present, and can produce misleading results when they are not. Hence, it is important to test for, rather than assume, nonlinear dynamics in a time series. However, most measures of nonlinear dynamics are also sensitive to linear dynamics and variations in the static distribution of the data. A bootstrap solution to these problems is provided by "surrogate series testing" (Theiler et al., 1992; see also Davison & Hinkley, 1997). Surrogate series are random variations on an experimental series that are constrained to preserve structure in the experimental series dictated by a null hypothesis. Permutations of the order of the experimental series, for example, produce surrogates that instantiate the null hypothesis of no temporal structure and the same static distribution of values as the experimental series. Theiler et al.'s (1992) Amplitude Adjusted Fourier Transform (AAFT) surrogates test a more interesting null hypothesis, a stationary Gaussian linear autoregressive process observed through a static and monotonic measurement function. Allowing for a monotonic measurement function provides considerable flexibility to accommodate whatever static distribution characterizes the measured series. Schreiber and Schmitz (2000) developed an iteratively refined AAFT surrogate that has improved accuracy in matching both the spectrum and distribution of a finite data set.

A test is constructed by comparing the experimental and surrogate series on a measure sensitive to nonlinearity. Significance is determined by the percentile attained by the nonlinear measure for the experimental series in the surrogate distribution. In a survey of measures of nonlinearity, Schreiber and Schmitz (1997) found that nonlinear prediction error was the most consistent in discriminating a range of noisy chaotic times series (see also Tong, 1990 for a review of tests of nonlinearity). For nonlinear prediction error, the test is one tailed, with prediction error for the experimental series at the p-th percentile rejecting the null hypothesis with confidence 100-p. Heath also describes a two-tailed test using a cubic time reversal index, which measures the asymmetry of the distribution of differences using a measure related to their third cumulant around zero (e.g., the sum of cubes of the differences divided by their sum of squares). This statistic is based on the theory of polyspectra (Rao & Gabr, 1984), which generalizes stationary Gaussian linear models to include moments of order greater than two. Stationary linear

Gaussian processes have a symmetric difference distribution and zero cumulants of orders greater than two; thus either significantly positive or negative time asymmetry relative to the surrogate distribution indicates the presence of nonlinearity or nonstationarity.

Schreiber and Schmitz (1997) found the cubic time reversal index to have low power with noisy Lorenz x series data, although it performed well for other chaotic series. For the stationary portion of the Lorenz series in Figure 1, the lag one difference distribution contains three very large positive values (see Figure 2b), which dominate the time reversal index because of the cubing operation. The presence of such large outliers exacerbates the low efficiency of higher order cumulants for estimating distribution properties (cf. Ratcliff, 1979), perhaps explaining some of the low power found by Schreiber and Schmitz. Diks, van Houwelingen, Takens and DeGoede (1995) provide a useful discussion of indices based on the reversibility of linear time series and suggest a potentially more efficient kernel based symmetry measure. No matter how efficient the statistic, however, assymetry is a sufficient indicator of nonlinearity, but is a not a necessary condition; hence, some nonlinear dynamics may display little asymmetry.

Heath discusses and compares surrogate testing using a range of nonlinearity measures in Chapter 7, including measures based on the correlation sum, such as dimensionality estimates. He also examines an alternative approach that combines prediction and time asymmetry without the need for surrogates (Stam, Pinj & Pritchard, 1998). Prediction is compared for the original (x(t), t = 1…N) and a time reversed version of the original (y(t) = x(N-t+1)), with any difference being attributable to nonlinearity. Stam et al. found that their time asymmetry technique produced strong discrimination with the chaotic Rossler equations, even with 50% additive Gaussian noise. Barahona and Poon's (1996) global polynomial methods can also be used to test for nonlinearity without surrogate generation (see also Poon & Barahona, 2000).

Although surrogate testing is potentially very useful, results should be interpreted with caution. While chaos does imply nonlinearity as measured by appropriate surrogate tests, nonlinearity does not necessarily imply chaos. Nonlinear stochastic processes can also cause a positive finding, as can a nonmonotonic measurement functions. For example, Schreiber and Schmitz (2000) showed that taking successive differences, which is commonly used in time series analysis to achieve stationarity, produces a positive surrogate test result when applied to a stationary second order linear autoregressive process. Spurious findings of nonlinearity can also occur if a time series is unevenly sampled or has missing values (Schmitz, & Schreiber, 1999). However, the surrogate generation method of Schreiber (1998) may be applied to the latter cases. This method can create surrogate series for any null hypothesis that can be expressed as a cost function. Although the method is quite general, it can be very computationally expensive because it uses a simulated annealing algorithm.

Nonstationary, such as the slow baseline drifts and strong narrow band periodicities often seem in EEG data (e.g., Stam et al., 1998), can also cause false positive findings in surrogate tests. Perhaps the chief omission from Heath (2000) is a test that can detect nonstationarity in nonlinear processes by checking the constancy of the full joint probability distribution, rather than just its lower order moments. Schreiber's (1997) cross-prediction method was found to work well for the Lorenz series in Figure 1. This method determines the degree to which information about one segment of a time series allows prediction of another segment. If stationary holds, and each segment is sufficiently long, equal cross prediction should be found, because each segment will be governed by the same joint probability distribution.

Insert Figure 4 about here

Figure 4 illustrates the mean squared error for one-step prediction for the Lorenz series broken into 10 consecutive segments of length 1000. The algorithm used locally constant prediction with delay 17 five-dimensional coordinates and averaged over a minimum of 30 neighbors to generate one-step-ahead predictions. Larger embedding dimensions and neighborhood sizes produced a similar pattern of results, as did smaller delays. The cross prediction method correctly detected the initial nonstationarity as the series settled into its attractor, and stationary behavior thereafter. Nonstationarity is most striking for the first 1000 values, with the next 1000 showing only a slight effect. An advantage of Schreiber’s (1997) method is that the degree of averaging can be adjusted to suit the level of noise. Schreiber claims that this approach can work with sub-sequences as short as 300-400, making it suitable for use with short behavioral series.

Even with the available prediction and surrogate testing methods, noise remains extremely problematic for NDA. As discussed earlier, traditional approaches to noise removal, such as frequency domain filters, cannot be applied to chaotic signals because they are, like noise, broadband. Instead, geometric filters (Grassberger et al., 1993) are required that smooth local neighborhoods of the attractor. In each neighborhood the signal is reconstructed using just the first few principal components of a linearization, usually achieved by singular value decomposition to ensure numerical stability. Because a chaotic signal will mainly exist on the larger components, whereas noise will be distributed evenly across all components, the reconstructed signal has an improved signal to noise ratio. Corrections are usually also made for the effects of curvature and sensitive dependence (Schreiber, 1999) and the filter may be iterated.

Like surrogate testing, local projective geometric filters should be used cautiously, and in particular the filter should not be iterated too many times (Mees & Judd, 1993). When deterministic dynamic structure is present it may be distorted by over-filtering. For example, Mees and Judd found that a circular attractor contracted substantially after 20 iterations. Over-filtering can also create apparent structure from noise with finite length time series. While in theory noise is evenly distributed in phase space, sampling error can cause inhomogeneities that are expanded by the filtering process. Hegger et al. (1999) provide an example where pure Gaussian noise produces a structured delay plot after 10 iterations. They advise that little more than three iterations should be required if chaos is present. As was the case for prediction, larger noise levels can be dealt with by using larger neighbourhoods, and by projecting onto a low dimensional manifold.

The techniques examined in earlier chapters are mainly based on relatively simple equations for chaotic dynamics. Chapters 8 and 9 address more complex models that are likely more representative of the mechanisms underlying nonlinear psychological phenomena. As illustrated by Figure 1, chaotic dynamics can be characterized as being constituted of a number of unstable periodic orbits. In the “control of chaos” approach, a feedback mechanism perturbs the chaotic trajectory so that a particular orbit is stabilized, resulting in limit cycle behavior. In one of the earliest application of chaos to brain processes, Skarda and Freeman (1987) suggested that olfactory memories were stored as limit cycles that were retrieved from chaotic background activity in olfactory cortex under the control of a familiar smell. The appeal of this idea is that complexity of the dynamics affords large storage capacity while the chaotic trajectory is never far from any state, and so retrieval is rapid. More recently, techniques summarized in Ott (1993) have provided effective algorithms for chaotic control. In Chapter 8, Heath develops a new modular neural network model that implements control of chaos and examines some applications to decision-making and handwriting.

Chapter 9 is concerned mainly with the evolution of complex behavior in the randomly connected Boolean networks studied by Kaufman (1993) and in Heath's (1993) novelty sensitive adaptive memory model (EAM). A case is made that cognitive systems live on the "edge of chaos", where both stability and flexibility are maximized. This chapter is in many ways the least self-contained. The definitions of some important concepts, such as "maximum mean fitness", are difficult to follow, as was the connection between an optimal memory stability parameter and short-term memory capacity. However, the proposed novelty sensitive memory model is interesting and clearly Heath has taken the opportunity in this chapter for a more speculative and integrative approach.

The final chapter describes the application of nonlinear dynamical techniques and models to a very broad array of psychological paradigms, ranging from low-level physiology, such as cardiac activity and EEG, through perception, motor skill and simple decision making to the fluctuation of mood states and applications in social psychology. Heath also reports some interesting new work on people's ability to forecast chaotic time series. The experimental design is an ingenious application of surrogate series, requiring participants to predict both chaotic series and their surrogates. Differences in prediction performance between surrogate and chaotic series provided evidence that at least some individuals could predict the nonlinear dynamics of a time series. It is interesting to speculate whether participants who can predict nonlinear dynamics use a local similarity based method related to the nonlinear prediction methods used in NDA. NDA captures similarity between sequences by neighborhood relations in time delay coordinates, then examines the evolution of neighbors to achieve prediction. Perhaps people can also use similarities between the present situation and past situations to achieve prediction, although clearly more work is needed to test this possibility, and to explicate the processes and representations involved.

In summary, Heath (2000) demonstrates that NDA has come a long way since the first naive applications of concepts from the theory of nonlinear dynamical systems to experimental data. Although it will probably never be possible to prove chaos in measured time series, robust and powerful algorithms are now available to test for and quantify nonlinear structure in time series, rather than simply assuming it. Whether any underlying structure detected by such tests can be eventually attributed to true chaos: “At least, chaos theory has inspired a new set of useful time series tools and provides a new language to formulate time series problems – and to find their solutions.” (Schreiber, 1999, p. 1). Heath (2000), therefore, is a timely invitation to NDA. While NDA techniques are still evolving, many important problems have satisfactory solutions, and many of the common pitfalls have been explicated. Much work remains to be done on the multivariate case, and on spatio-temporal (so called “extensive”) chaos. However, at least for one-dimensional time series, psychologists can have confidence in results if NDA is carefully applied.

With an increasing database of empirical results, theories based on nonlinear dynamics also become more viable. In the physical sciences where dynamics have been so successful, the underlying determining equations are usually known and can often be studied in parameterizations producing simple fixed point or limit cycle behavior before undertaking an analysis of the complexities of chaotic regimes. The theoretical strand of Heath's book contributes to the development of nonlinear dynamical theory of behavior (c.f. Gregson, 1988, 1992, 1995; Guastello, 1995; Kelso, 1995; Newell, Liu & Mayer-Kress, 2001). In this "Decade of Behavior", Heath (2000) breaks new ground in illustrating not only that the study of chaos has provided useful new tools for the analysis of behavioral time series, but also a rich source of inspiration for models of complex behavioral processes. The ultimate success of nonlinear dynamics in both behavioral and other areas of psychology will rest on the further development of specific dynamical models and the adaptation of NDA techniques to test them.

References

Abarbanel, H. D. I. (1996). Analysis of observed chaotic data, New York: Springer.

Barahona, M. & Poon, C.-S. (1996). Detection of nonlinear dynamics in short, noisy time series, Nature, 381, 215-217.

Belair, J., Glass, L., van der Heiden, U. & Milton, J. (1995). Dynamical disease, mathematical analysis of human illness, New York: AIP Press.

Box, G. E. P. & Jenkins, G. M. (1976). Time series analysis: Forecasting and Control. San Francisco: Holden Day.

Casdagli, M. (1997). Recurrence plots revisited, Physica D, 108, 12-44.

Casdagli, M., Eubank, S., Farmer, J. D. and Gibson, J. (1991). State space reconstruction in the presence of noise. Physica D, 51, 52-98.

Cleveland, W. S. (1985). The Elements of Graphing Data, Monterey, CA: Wadsworth.

Davison, A.C. & Hinkley, D.V. (1997) Bootstrap methods and their application. Cambridge: Cambridge University Press.

Diks, C., van Houwelingen, J. C., Takens, F. & DeGoede, J. (1995). Reversibility as a criterion for discriminating time series, Physics Letters A, 201, 221-228.

Eckmann, J. P., Oliffson-Kamphorst, S. & Ruelle, D. (1987). Recurrence plots of dynamical systems. Europhysics Letters, 4, 973

Fraser, A. M. & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information, Physical Review A, 33, 1134-1140.

Grassberger, P. (1987). Evidence for climatic attractors: Grassberger replies, Nature, 326, 524.

Grassberger, P. & Procaccia, I. (1983). Characterization of strange attractors. Physical Review Letters, 50, 189-208.

Grassberger, P., Hegger, R., Kantz, H., Schaffrath, C. & Schreiber, T. (1993). On noise reduction methods for chaotic data. Chaos, 3, 127-141.

Gregson, R. A. M. (1988). Nonlinear psychophysical dynamics. Hillsdale NJ: Lawrence Erlbaum Associates.

Gregson, R. A. M. (1992). n-dimensional nonlinear psychophysics. Hillsdale NJ: Lawrence Erlbaum Associates.

Gregson, R. A. M. (1995). Cascades and fields in perceptual psychophysics. Singapore: World Scientific.

Gregson, R. A. M., & Pressing, J. (2000). Dynamic modelling. In L. G. Tassianry, J. G. Cacioppo & G. Bernston, Principles of Psychophysiology: Physical, Social and Inferential Elements, New York: Cambridge University Press.

Guastello, S. J. (1995). Chaos, catastrophe, and human affairs: Applications of nonlinear dynamics to work, organizations, and social evolution. Hillsdale, NJ: Lawrence Erlbaum Associates.

Heath, R. A. (1984). The detection of change in physiological measures using an adaptive Kalman filter algorithm. Psychological Bulletin, 96, 581-588.

Heath, R. A. (1993). A nonlinear model for human associative memory based on error accumulation. In P. Leong & M. Jabri (Eds.) Proceedings of the Fourth Australian Conference on Neural Networks (pp. 130-133). Sydney, Australia: Department of Electrical Engineering, University of Sydney.

Hegger, R., Kantz, H. & Schreiber, T. (1999). Practical implementation of nonlinear time series methods: The TISEAN package, Chaos, 9, 413-435.

Kantz, H. & Schreiber. T. (1997). Nonlinear time series analysis, Cambridge, England: Cambridge University Press.

Kaplan, D. & Glass, L. (1995). Understanding nonlinear dynamics. New York: Springer-Verlag.

Kauffman, S. A. (1993). The origins of order: Self-organization and selection in evolution. New York: Oxford University Press.

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, 3403-3411.

Kelso, J. A. S. (1995). Dynamic patterns: The self-organization of brain and behaviour, Cambridge, MA: MIT Press.

Koebbe, M. & Meyer-Kress, G. (1991). Use of recurrence plots in the analysis of time-series data, M. Casdagli & S. Eubank (Eds.) Nonlinear Modelling and Forecasting, Reading, MA: Addison-Wesley, 361-376.

Lorenz, E. N. (1969). Atmospheric predictability as revealed by naturally occurring analogues. Journal of Atmospheric Science, 26, 636

Luce, R.D. (1995). Four tensions concerning mathematical modelling in psychology. Annual Review of Psychology, 46, 1-26.

Marmarelis, P. Z. & Marmarelis, V. Z. (1978). Analysis of physiological systems: The white-noise approach. New York: Plenum Press.

Mees, A.I. & Judd, K. (1993). Dangers of geometric filtering. Physica D, 68, 427-436.

Newell, K. M., Liu, Y-T., & Mayer-Kress, G. (2001). Time scales in motor learning and development, Psychological Review, 108, 57-82.

Nusse, H. E. & Yorke, J. A. (1997). Dynamics: Numerical Explorations, 2nd Edition, New York: Springer-Verlag.

Osborne, A. R., & Provenzale, A. (1989). Finite correlation dimension for stochastic systems with power-law spectra, Physica D, 35, 357-381.

Ott, E. (1993). Chaos in dynamical systems, Cambridge, England: Cambridge University Press.

Poon, C-S. & Barahona, M. (2000). Titration of chaos with added noise, Proceedings of the National Academy of Science, 98, 7107-7112.

Pritchard, W. S. & Duke, D. W. (1992). Measuring chaos in the brain: A tutorial review of nonlinear dynamical EEG analysis. International Journal of Neuroscience, 67, 31-80.

Provenzale, A., Smith, L. A., Vio, R. & Murante, G. (1992). Distinguishing between low-dimensional dynamics and randomness in measured time series, Physica D, 58, 31-49.

Rao, T. S., & Gabr, M. M. (1984). An introduction to bispectral analysis and bilinear time series models, New York, Springer-Verlag.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. Psychological Bulletin, 86, 446-461.

Sauer, T., Yorke, J. & Casdagli, M. (1991). Embedology. Journal of Statistical Physics, 65, 579-616.

Schmitz, A. & Schreiber, T. (1999). Testing for nonlinearity in unevenly sampled time series, Physical Review E, 59, 4044-4047.

Schreiber, T. (1997). Detecting and analysing nonstationarity in a time series using nonlinear cross predictions, Physical Review Letters, 78, 843-846.

Schreiber, T. (1998). Constrained randomisation of time series data, Physical Review Letters, 80, 2105-2108.

Schreiber, T. (1999). Interdisciplinary application of nonlinear time series methods, Physics Reports, 308, 1-64.

Schreiber, T. & Schmitz, A. (1997). Discrimination power of measures for nonlinearity in a time series, Physical Review E, 55, 5443-5447.

Schreiber, T. & Schmitz, A. (2000). Surrogate time series, Physica D, 142, 346-382.

Skarda, C. & Freeman, W. (1987). How brains make chaos in order to make sense of the world. Behaviour and Brain Sciences, 10, 161-195.

Skinner, J. E., Molnar, M., Vybrial, T. & Mitra, M. (1992). Application of chaos theory to biology and medicine. Integrative Physiological and Behavioural Science, 27, 39-53.

Stam, C. J., Pijn, J. P. M. & Pritchard, W. S. (1998). Reliable detection of nonlinearity in experimental time series with strong periodic components. Physica D, 112, 361-380.

Sugihara, G. & May, R. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, Nature, 344, 734-741.

Takens, F. (1981). Detecting strange attractors in turbulence: Lecture Notes in Mathematics, 898, New York: Springer.

Theiler, J. (1990). Estimating fractal dimension. Journal of the Optical Society of America A, 7, 1055-1073.

Theiler, J. & Eubank, S. (1993). Don't bleach chaotic data. Chaos, 3, 335-341.

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

Theiler, J. & Rapp, P. E. (1996). Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. Electroencephalographic Clinical Neurophysiology, 98, 213-222.

Thomasson, N., Hoeppner, T. J., Webber, C. L. & Zbilut, J. P. (2001). Recurrence quantification in epileptic EEGs, Physics Letters A, 279, 94-101.

Tong, H. (1990). Non-linear time series, a dynamical system perspective. Oxford: Oxford University Press.

Trulla, L. L., Giuliani, A., Zbilut, J. P. & Webber, C. L. (1996). Recurrence quantification analysis of the logistic equation with transients, Physical Letters A, 223, 255-260.

Webber, C. L. & Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. Journal of Applied Physiology, 76, 965-973.

Weigend, A. S. & Gershenfeld, N. A. (1993). Time series prediction: Forecasting the future and understanding the past. Reading, MA: Addison-Wesley.

Zbilut, J. P., Giuliani, A. & Webber, C. L. (1998). Recurrence quantification analysis and principal components in the detection of short complex signals. Physics Letters A, 237, 131-135.

Figure Captions

Figure 1. 10000 x values (thin line) for the Lorenz differential equations: $dx/dt = 10(y - x)$, $dy/dt = x(28 - z - y)$, and $dz/dt = xy - z\,8/3$, starting from (x, y, z, t) = (0,1,0,0), with a Lowess (Cleveland, 1985) smooth using a 1000 element moving window (thick line).

Figure 2. A plot of x(t) from Figure 1 against its derivative estimated at lag one for (a) t = 1 … 2000, (b) t = 2001 … 10000, and (c) at lag 12 for t = 2001 … 10000.

Figure 3. Local slopes of log(C(ε)) vs. log(ε) plots for a range of distances, ε, and embedding dimensions from 1 to 10 in a delay 17 embedding with a Theiler window of 100 for the Lorenz series for t = 2001-10 000. Lower lines indicate lower embedding dimensions, and the dashed line indicates the true dimension, 2.05.

Figure 4. One-step cross prediction errors obtained using 10 sequential 1000 element segments (separate lines) to obtain predictions for all segments (abscissa) for the Lorenz series in Figure 1.

Acknowledgements

I would like to thank Alice Kelly, Mitchell Longstaff, Scott Brown, Doug Mewhort and Richard Chechilie for help in the preparation of this review.

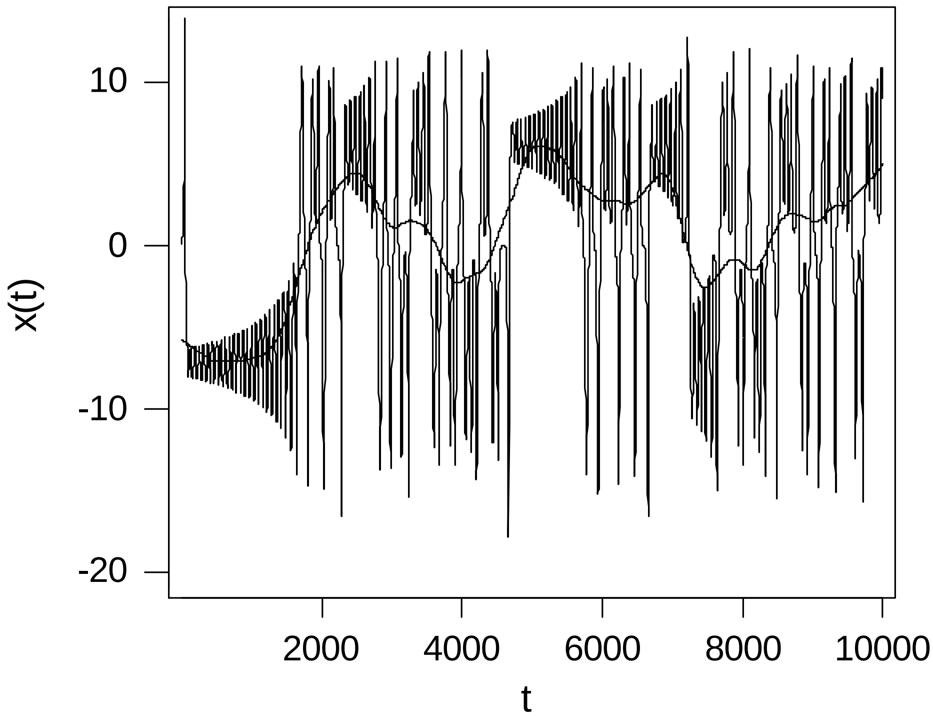
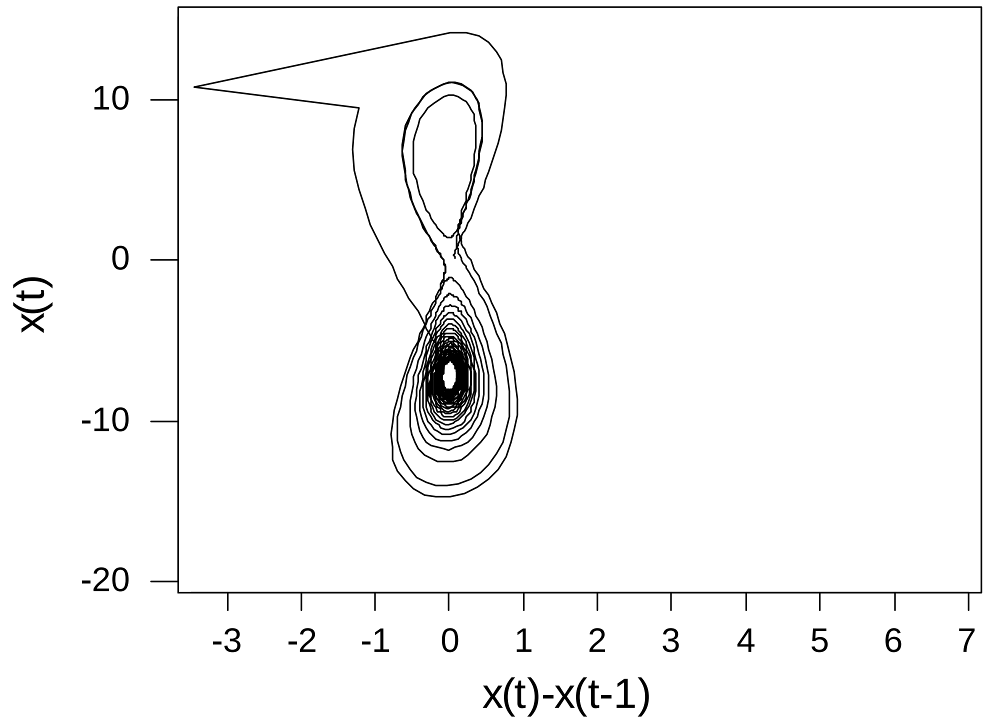
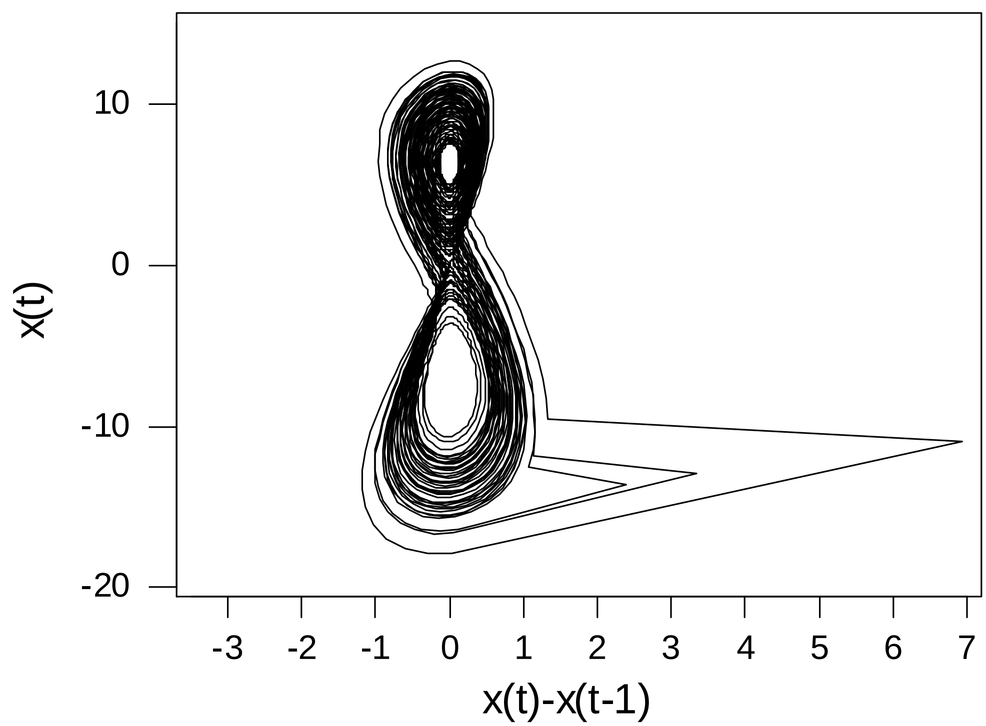


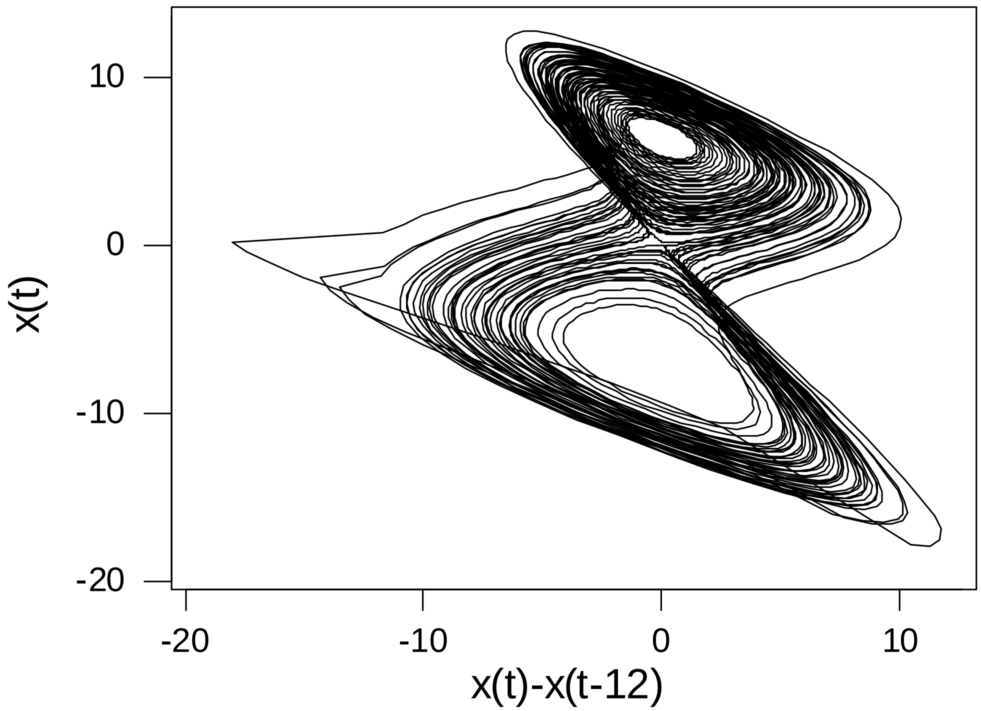
Figure 1


10
0
-10
-20
x(t)
-3
-2
-1
0
1
2
3
4
5
6
7
x(t)-x(t-1)

(a)


10
0
-10
-20
x(t)
-3
-2
-1
0
1
2
3
4
5
6
7
x(t)-x(t-1)

(b)



(c)

Figure 2

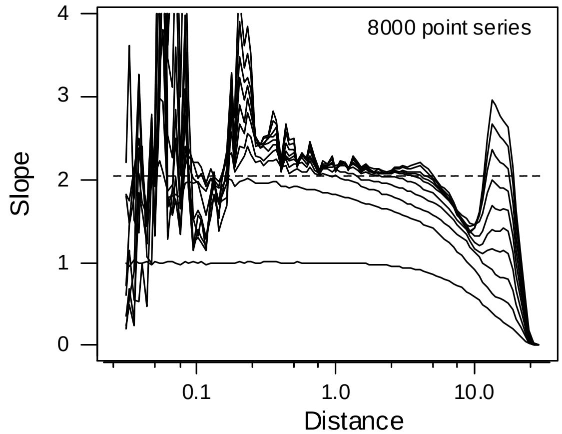


Figure 3

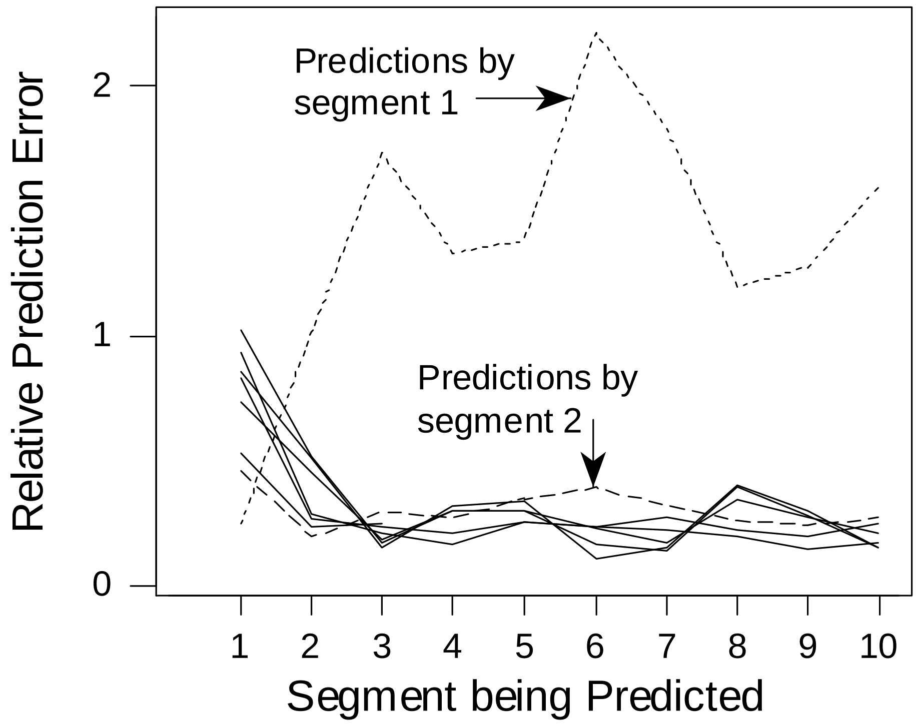

Predictions by
segment 1
Predictions by
segment 2
Relative Prediction Error
0
1
2
1
2
3
4
5
6
7
8
9
10
Segment being Predicted

Figure 4

Footnotes

---

[1] The DOS program Smalldyn (http://www.ipst.umd.edu/dynamics/) was used to generate the Lorenz series in Figure 1a. Smalldyn is a free version of the software accompanying Nusse and Yorke (1997) that provides a selection of chaotic equations and graphing facilities to examine the time series they produce.

[2] The website for Heath (2000) provides DOS command-line programs `kalman0` and `kalman1`, and document that describes their use. Unfortunately, a maximum of 10th order model is allowed whereas ARMA analysis indicated a 12th order model is required for this series. Failure to detect the early nonstationarity may have occurred because the Kalman filter works on-line, and so takes some time to accumulate information.

[3] Estimates should be obtained with the same measurement function, $\mathbf{y}=f(\mathbf{x})$, where $\mathbf{y}$ is the observed time series, $\mathbf{x}$ is the series produced by the dynamics and f is a monotone function. Under ideal circumstances dimensionality is invariant with respect to the measurement function but when only relative measurements are possible it may not be invariant. Hence, comparisons may be confounded if the measurement functions differ. Similarly, changes in the level and type of measurement noise may change dimensionality estimates, and thereby confound comparisons.