**Leonard A. Smith**
Mathematics Institute, University of Oxford, Oxford, OX1 3LB, United Kingdom;
e-mail: lenny@maths.ox.ac.uk

# Does a Meeting in Santa Fe Imply Chaos?

*I have no data yet. It is a capital mistake to theorize before one has data.*

— Holmes to Watson in *A Scandal in Bohemia*
(see Doyle, 1930)

This chapter compares the success of several nonlinear prediction techniques applied to Data Set A of the Santa Fe Time Series Prediction and Analysis Competition (both `A.dat` and `A.cont`. The advantages of a new approach making predictions based on selective use of several different delay reconstructions are illustrated, and a comparison of both local linear and local nonlinear predictions is given. In addition, the phase coherence of the system and the self-consistency of the data is examined using the longer data set `A.cont`; the latter locates a possible sensor failure in this data set. Limitations due to the amount of data, the sampling rate, and the saturation in the data, in combination with the quality of the predictions achieved with very little information on the value of the initial condition

Times Series Prediction: Forecasting the Future and Understanding the Past, Eds. A. S. Weigend and N. A. Gershenfeld, SFI Studies in the Sciences of Complexity, Proc. Vol. XV, Addison-Wesley, 1993

(32 bits or less), suggest that, while the system is clearly nonlinear, evidence from `A.dat` for sensitive dependence on initial condition, if any, is slight.

## 1. INTRODUCTION

The theme of the workshop was the extraction of information from data. In this chapter, we will approach this task with a variety of nonlinear prediction techniques and an examination of the data itself. For all the prediction results presented, we will construct the predictor from the data in file `A.dat` and test it on the data in file `A.cont` which is a continuation of `A.dat`, thus all error statistics are out of sample. Because the predictors are constructed from `A.dat`, this data set will be called the learning set.

In addition to global predictors, in which each prediction is influenced by the entire learning set, and local predictors, based on a local (one hopes relevant) subset of points, we shall introduce predictors that take advantage of different reconstructions at different times (i.e., in different regions of phase space). This approach can provide better predictions than any single "optimal" reconstruction.

Many of the results presented at the conference, along with those reported here, illustrate that the system is very predictable. Examination of the data in the continuation set, especially following collapses (defined below), supports this view. We suggest that the dynamics of the underlying system might be modeled simply; while the series is clearly nonlinear, there is little evidence of sensitivity

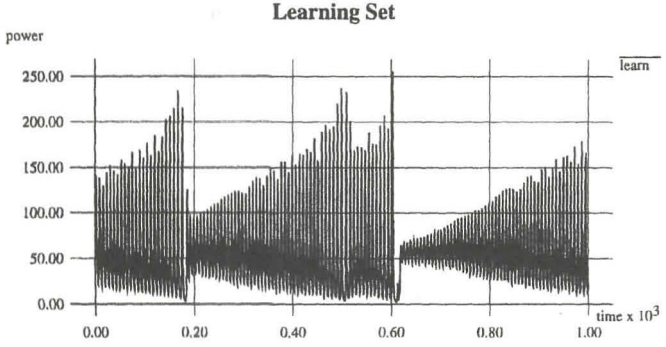


FIGURE 1 The original data set `A.dat`. These 1,000 points are used as the learning set for the predictors presented below.

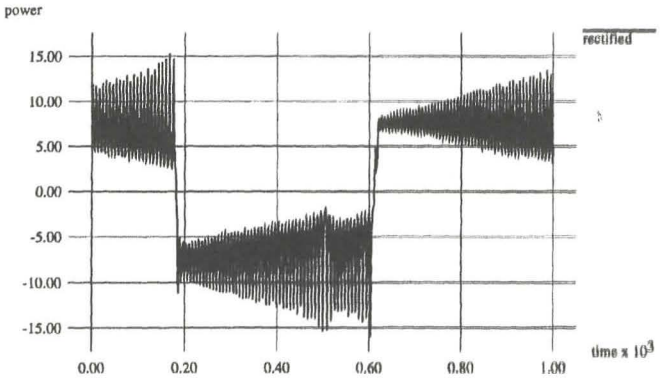


FIGURE 2 An artificially rectified version of `A.dat`; the regular growth of the oscillations appears more clearly after taking the square root of the data. In the second cycle, the negative square root has been arbitrarily taken to illustrate the possible degeneracy.

to initial condition. Although the physical system may be in fact chaotic, it seems premature to theorize this from the presented analyses based on the data in `A.dat` alone. Finally, we suggest modifications to the observations that underly this data set which could clarify these issues.

While we shall concentrate on prediction below, related techniques can be employed to detect anomalies in the data stream, which may result from either sensor failure or a major change in the system's dynamics. Such an anomaly is, in fact, detected in `A.cont`.

## 2. THE DATA

The initial data set `A.dat`, shown in Figure 1, consists of 1,000 observations, equally spaced in time and digitized to 8-bit integers in the range 0 to 255 (inclusive). We will call the high-frequency oscillations "cycles," the sudden large decreases in the amplitude of the cycles "collapses," and the packet of growing oscillations between two collapses "an event." The series samples three events, in two of these the collapse is observed; note that if this data set were to arise from motion on a chaotic attractor, it would provide a sparse image of the attractor since only one complete event (collapse through growth to collapse) is available. With an eye on prediction we note that, as only two collapses are observed, the prediction of collapses may be expected to prove difficult, even if the mechanism giving rise to them is straightforward.

### 2.1 THE PHYSICS: GENERIC?

The measurement of intensity is degenerate in the sense that, as the square of a physical variable (the electric field, $E$), it does not distinguish $+E$ from $-E$. To illustrate this possibility, we "rectify" the data, as shown in Figure 2, by taking the square root of each observation and arbitrarily taking the center set of cycles as negative. If, on physical grounds, there is a distinction in the dynamics between $+E$ and $-E$, then the two lobes of any "attractor" would not be perfectly symmetric; this would suggest the challenge of determining whether the techniques discussed at this meeting could identify which oscillations correspond to which lobe. If this is indeed the case, it would decrease the data density in the learning set still further and introduce projection effects into the learning set.

### 2.2 THE STATISTICS: STATIONARY?

We should also consider whether or not the statistics of `A.dat` are stationary or, more accurately, whether the statistics computed from `A.dat` have converged to those of an underlying stationary process as, technically, stationarity is defined for a process, not a data set. A necessary condition for such convergence is that the histogram of observed values be representative of that of the system. However, this is not sufficient for the reconstruction methods discussed below. These methods would require that the reconstruction in higher dimensions be "well explored."[1] Given that only two collapses occur in the learning set, this criterion could only be met for a very simple attractor.

The data also appears to be undersampled, judging from the series (the beating of the sampling rate and the cycles is reflected in Figure 4). Thus the times and magnitudes of successive local maxima are poorly approximated; it is suggested below that an excellent model of this system could be constructed from these two parameters. We also note that the signal saturates at 255, this is clearly illustrated in the histogram of observations from the continuation set (not shown, but the effect is seen in Figure 9). This saturation tends to coincide with collapse, making it difficult to judge the true sensitivity to initial conditions.

## 3. THE PREDICTORS

In this section we consider a variety of methods to predict the short-term behavior of the system. The methods considered require that the single observable first

[1] This would, of course, depend on the length of the data set, the sampling rate, and the embedding dimension. It is quite reasonable to expect that a given series reconstructed in three dimensions could appear "stationary" while the same data reconstructed in ten dimensions would not. Note, however, the discussion of this issue in Sugihara and May (1990).

be embedded to form a higher dimensional reconstruction. The time ordering of points in this reconstruction is then used to turn the prediction problem into one of interpolation.

### 3.1 THE METHOD OF DELAYS

The first step in applying these methods of prediction is to *reconstruct* the time series in a geometrical framework. A trajectory, $\mathbf{x}(t)$, is reconstructed in $M$ dimensions from the single observable, $s(t)$, recorded with uniform sampling time, $\tau_s$, by the method of delays to yield a series of vectors

$$\mathbf{x}_i = (s_i, s_{i-j}, \ldots, s_{i-j(M-1)}) \tag{1}$$

where $j$ (or $j\tau_s$) is called the delay time, $\tau_d$. For a deterministic system and a generic observable, this reconstruction preserves many of the characteristics of the original system for sufficiently large $M$ (see, e.g., Casdagli et al., 1991; Packard et al., 1980; Sauer, Yorke, & Casdagli, 1991; and Takens, 1981). In this chapter, we will restrict attention to delay reconstructions of the full series where $M = 4$ with either $\tau_d = 1$ or $\tau_d = 4$. In addition to the prediction time, $\tau_p$, there remains one further time scale to be considered, the width of the "window" used to determine a point in the reconstruction space, $\tau_w$. For the method of delays, $\tau_w = ((M-1)j+1)\tau_s$. We have observed, however, that multivariate reconstructions of the maximum value and duration of each cycle (i.e., prediction of the value and time of occurrence of the next maxima, given the series of local maxima) provide excellent, if preliminary, results. This approach avoids some of the difficulties discussed below, particularly with respect to collapses. Predictions based on this approach should be significantly improved by additional data with higher resolution (in both time and intensity), so that the maxima are more sharply defined, and may provide the clearest evaluation of sensitivity to initial condition.

### 3.2 MODES OF PREDICTION

In this section, we clarify several details in the approach to making predictions that must be specified independently of the details of the predictor itself.

3.2.1 DIRECT AND ITERATIVE FORECASTS. In general, we will employ *direct forecasts* with a single-step predictor invoked once per prediction. We contrast this single prediction a time $\tau_p$ into the future, with *iterative forecast predictions* where the final prediction is based upon a number of predictions made at a smaller time step (e.g., $i$ forecasts, each advancing $\tau_p/i$ well into the future).

3.2.2 RUNAWAY AND UPDATE EXTENSION. For the competition, we were asked to take an initial condition and iterate it as far as possible, using the output of the predictor to continue the series. We will denote these as *runaway* extensions in contrast to *update* extensions, where the true value is incorporated after each step of the prediction. While the former are of interest when extending a true series into the future, the latter allow the evaluation of a predictor at a fixed $\tau_p$, which eases parameter determination and interpredictor comparisons. All the results in the paper, with the exception of Figures 6 and 7, are based on update predictors.

3.2.3 ENTRAINMENT. Finally, we note that for quantized data, the system may repeatedly produce the same exact string of observed data values and then diverge. If this string is as long as the window considered by a predictor, $\tau_w$, a deterministic *runaway* predictor cannot reproduce the behavior; it will become entrained within a periodic cycle quite unrelated to the primary dynamics of the system. A related problem distinguishes direct and iterated forecasts, namely that iterated predictors can't get through regions where, for example, the series appears to be constant for a time greater than $\tau_w$; direct predictors can effectively "jump" over such a data segment. Depending upon location in phase space it is possible to combine the advantages of both iterative and direct forecasts by selective usage. This approach, illustrated below, combines two reconstructions with different delay times.

## 3.3 GLOBAL PREDICTORS

Having produced a four-dimensional reconstruction of the ($n_L = 1000$) points in `A.dat` with $\tau_d = 1$ and $\tau_p = 1$, we first consider a global predictor, $F(\mathbf{x}) : R^4 \rightarrow R^1$, which estimates $s$ for any $\mathbf{x}$. $F(\mathbf{x})$ is constructed about $n_c$ centers

$$\mathbf{c_j}, \quad j = 1, 2, \ldots, n_c; \quad \mathbf{c_j} \in R^4 \tag{2}$$

chosen from the learning set in such a manner that they are spread out on the reconstruction. We will consider $F(\mathbf{x})$ of the form

$$F(\mathbf{x}) = \sum_{j=1}^{n_c} \lambda_j \phi(||\mathbf{x} - \mathbf{c_j}||) \tag{3}$$

where $\phi(r)$ are radial basis functions (Powell, 1987), in this contribution, either $\phi(r) = r^3$ or $\phi(r) = e^{-r^2/c^2}$ where the constant, $c$, is based on a multiple of the average distance between data points considered in the fit, $d_{nn}$. The $\lambda_j$ are constants determined by a least-squares fit to the observations in the learning set:

$$\mathbf{b} = \mathbf{A}\lambda \tag{4}$$

where $\lambda$ is a vector of length $n_c$ whose $j$th component is $\lambda_j$ and $\mathbf{A}$ and $\mathbf{b}$ are given by

$$A_{ij} = \omega_i \phi(||\mathbf{x_i} - \mathbf{c_j}||) \tag{5}$$

and

$$b_i = \omega_i s_i \tag{6}$$

where $i = 1, \ldots, n_L$ and $j = 1, \ldots, n_c$. Traditionally, the weights $\omega_i$ associated with each point in the learning set reflects its accuracy; alternatively, the $\omega_i$ may be tuned to improve the fit in under-represented regions of the reconstruction (e.g., near the collapses). For global reconstructions, we shall restrict attention to the case where all $\omega_i$ are equal (but note the discussion in L. A. Smith, 1992). Details

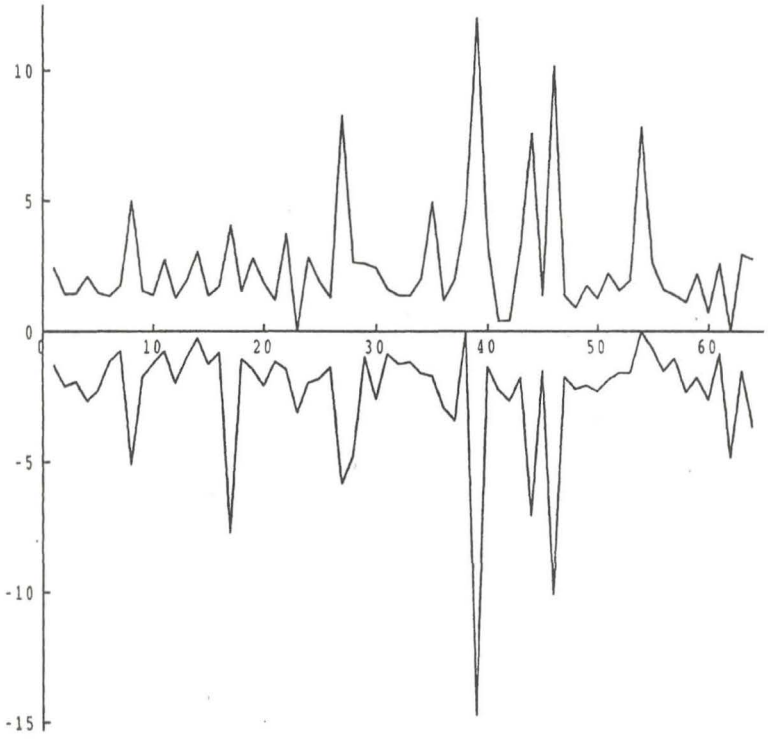


FIGURE 3 Variation of the average error with location in phase space as denoted by the nearest center. Positive and negative errors are averaged separately. Note the great variation at different locations on the reconstruction.

of the construction of this type of predictor may be found in Broomhead and Lowe (1988), Casdagli (1989), and Farmer and Sidorowich (1988). We call function $F$ constructed in this way an RBF predictor.

There are several advantages of global predictors. The reconstruction is smooth and the coefficients are fixed. Thus, since all the computational overhead is done at the outset, large test data sets are easily evaluated. The global reconstruction

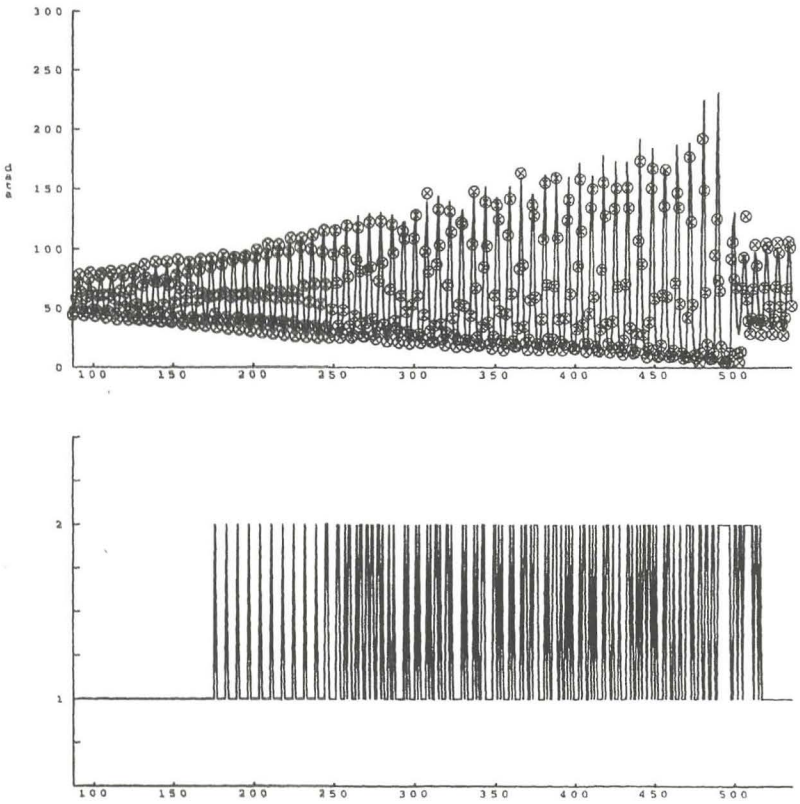


FIGURE 4 Global, multireconstruction predictor results in an update mode. The upper panel showed the observed (solid) and predicted (symbol) values. Note that, as the predictions are fairly accurate, they reveal the beating between the cycles and the sampling rate. The lower panel shows which of the two predictors (denoted by either 1 or 2) was used for the prediction at each given instant.

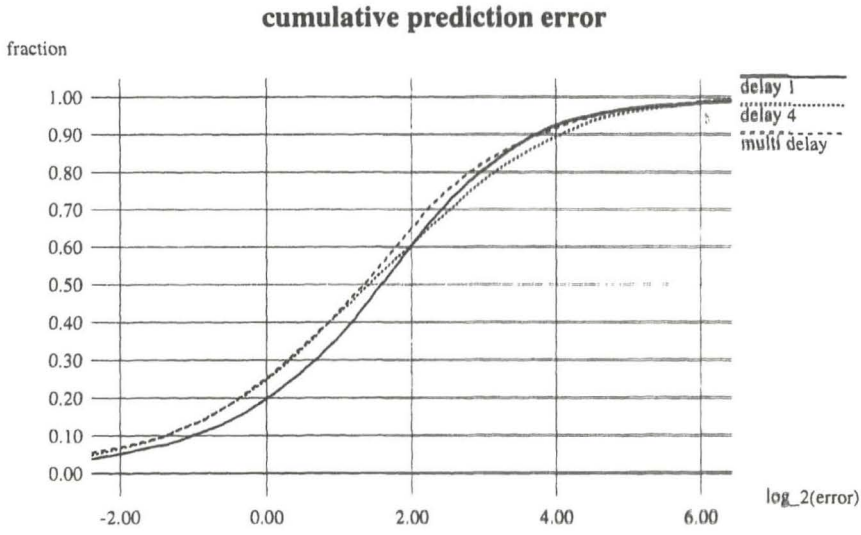


FIGURE 5 The prediction profiles of three global reconstructions with one-step ahead (update) predictions. The curves indicate the fraction of the test set that is predicted to within a given error; thus, for a given value of the error, the highest curve represents the best predictor. The solid and dotted lines reflect the profiles of two different fixed-delay predictors, while the dashed line is that of a (more accurate) third predictor combining the two, as described in the text.

also provides a natural partition of the phase space which can serve many uses. For instance, variation in the quality of predictions with location provides an estimate of the expected error in a given prediction (L. A. Smith, 1992). Figure 3 shows the average positive and negative error in the neighborhood of each center, illustrating the strong variations which occur in practice. In some regions there is a definite bias in the predictions (i.e., they are always too low), this can occur when there is not sufficient data (weight) in these regions and the least-squares fit to the entire data set simply under-fits them; alternatively, one cause of variation in the quality of "mean zero error" regions is the variation of projection effects within a given delay reconstruction. It is this second point we shall discuss here.

Once we have used a particular partition to classify each point in the data series, there is no need for the restriction to a single delay reconstruction for making predictions; given a single, global delay reconstruction to define partitions, we then pick the particular delay reconstruction that works best in each partition and use

it.[2] Such a multireconstruction predictor is given in the next section. First, we introduce a graphical method to compare different predictors through the cumulative distribution of prediction error, or predictor profile, for short (L. A. Smith, 1990). This graph represents the fraction of the test set that can be predicted within a given accuracy, and clarifies the effect of outlying "bad" predictions that may bias statistics like the average predictor error. Prediction profiles for two $m = 4, \tau_p = 4$ global RBF predictors are shown in Figure 5. The figure shows, for example, that the predictor corresponding to the solid line predicts 20% of the test set with an error of less than 1 bit ($\log_2(\text{error}) < 0.0$). In one, $\tau_d = 1$, while $\tau_d = 4$ in the other. Note that the $\tau_d = 1$ predictor has more very small errors and more large errors which are large; where it is accurate, its predictions are superior to the $\tau_d = 4$ predictor, yet where it is inaccurate, they are far inferior. The point is that these two predictors work best in different locations in phase space; by recording those locations and using the appropriate predictor for each initial condition, we obtain the third curve in the figure. Although this multiple-delay predictor can't make any prediction better than "both" of its composites, it tends to pick the better of the two and hence gives a better prediction profile than either of them.

3.3.1 MULTIRECONSTRUCTION PREDICTION: A GLOBAL EXAMPLE. We illustrate what is happening by examining the predictions and which predictor is used where. The upper panel of Figure 4 shows an out-of-sample segment of the observed (solid) and predicted (symbol) time series. The lower panel notes which of the two predictors was used to make each prediction. For the small amplitude oscillations (and near local minima), the $\tau_d = 4$ predictor is preferred; it provides a good estimate of phase and amplitude which is more easily obtained in a longer window. When the oscillations are large and more irregular, the shorter delay contains the more relevant information. This makes sense; it may be a bit naïve to attempt to compute an "optimal delay time" if there is no need to average over the entire reconstruction.

## 3.4 LOCAL PREDICTORS

An alternative to the single global predictor is to construct a local predictor from the points in the neighborhood of each data point. For local methods other than the nearest neighbor, this involves defining a local neighborhood, usually with a fixed radius or fixed number of $k$ neighbors. Neither is optimal; it would be nice, but computationally expensive, to choose a neighborhood with respect to the local characteristics of the underlying function. Casdagli and Weigend (this volume) investigate the variation of predictions with $k$ using local linear maps in a variety of circumstances.

[2] Multiple reconstruction local predictors could also be constructed, utilizing either the error in a withheld subset or the in-sample error to determine the best delay.

3.4.1 NEAREST NEIGHBOR(S). The nearest-neighbor predictor simply chooses the point in the learning set closest to the point to be predicted and uses its image as the prediction. As shown in Figure 6, runaway nearest-neighbor prediction works rather well in `A.dat`. For coarsely quantized data, these predictors are likely to become entrained in the learning set; once a series of predictions produces a point corresponding to a point in the learning set, the predictor will march through the learning set until it reaches the end of the learning set—producing results similar to the striking overlay predictions by Kostelich and Lathrop and others in this volume. When an entrained nearest-neighbor predictor reaches the end of the data

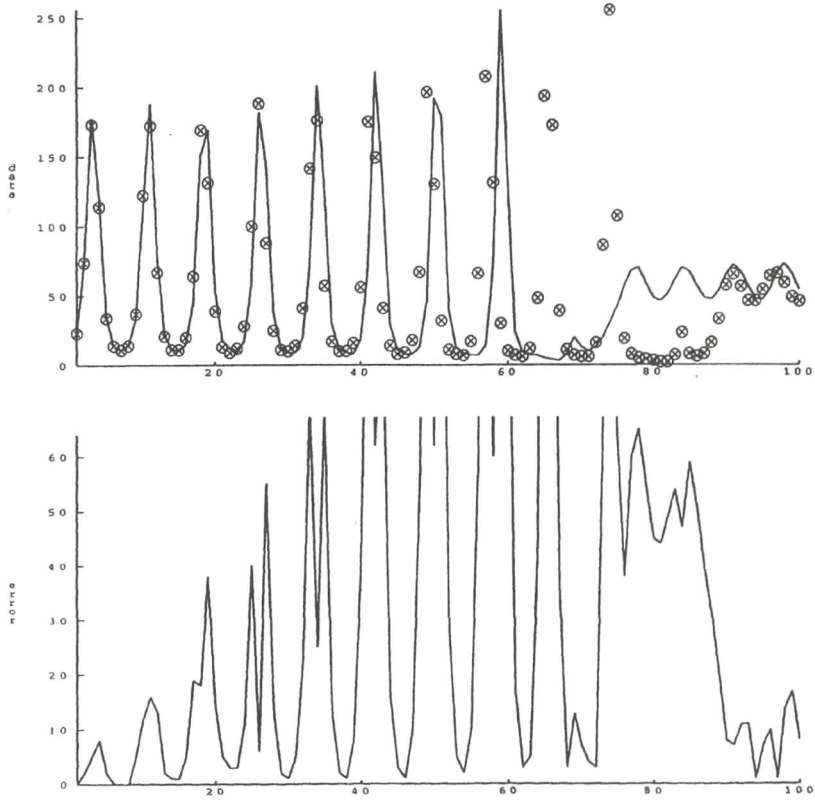


FIGURE 6 The results for runaway (single) nearest-neighbor prediction showing the observed (solid) and predicted (symbol) values. Here the image of the nearest neighbor of the point to be predicted was chosen where $M = 4, \tau_d = 1$. The lower frame shows the error as a function of time.

set, a sudden, rather disorganized-looking behavior may ensue as it suddenly finds a sharp increase in the distance to nearby points; this behavior may continue until the predictor "drifts" near a region where the data density is higher and, perhaps, becomes re-entrained. Other "overtrained" predictors may exhibit similar behavior. In the instance shown here, the nearest-neighbor predictor became entrained in less than ten steps.

Note that this predictor does collapse (two cycles late) and that the phase of the observed and predicted data nearly coincide *after* the collapse. This suggests that phase information is preserved through the collapse, a question investigated below.

3.4.2 LOCAL LINEAR AND QUADRATIC PREDICTION. Given the coordinates of points within a neighborhood, a local linear predictor finds the best linear combination of these distances for interpolating the future observation. A local quadratic predictor works similarly but includes the quadratic terms. For small data sets, determining the correct size for each neighborhood is crucial; if it is too large, higher order nonlinear effects will be included while, if it is too small, the quadratic and RBF predictors may overfit the data. (Techniques to avoid overfitting with neural nets are discussed in Weigend and Rumelhart (1991).) A major difficulty is that "large" will vary with location over the reconstruction (in terms of either the $k$ nearest neighbors or some fixed distance $d$ in reconstruction space). While considering the average ratio of the observed error to the expected (local, in-sample) error can provide an idea of whether the local predictor is overfitting the data, it also suffers from averaging over the space. A truly local solution based on information theoretic grounds, suggested by A. Mees, currently under investigation in collaboration with K. Judd. In trials based on `A.dat`, the local quadratic map often yields the better results, although there is enough variation that the best approach might well be to generate an ensemble of predictions (with uncertainty estimates) and form a suitable weighted average. Beven gave an example of this approach in a hydrological context (Beven & Binley, 1992).

3.4.3 LOCAL RADIAL BASIS FUNCTION PREDICTION. Local RBF predictors are similar to the global RBF predictor above but use only local data. This has the advantage of not assuming that the neighborhood will be small enough to be linear, but care must be taken not to overfit the data in the locally linear case. Runaway predictions using this method are shown in Figure 7; they provide quite reasonable estimates of the behavior until the collapse, but do not predict the collapse. Indeed, it is very difficult to make this predictor collapse; this is due, in part, to the low weight that the two collapses in the learning set have (low because there are only two realizations out of many observed cycles). This can be changed either by taking smaller neighborhoods or, independently, by increasing the weights $w_i$ of the points in the regions of reconstruction space corresponding to a collapse, thereby forcing an improved fit to that region. Alternatively, one could try to detect (by, say, a simple threshold method) when the next collapse should occur and then

to extend predictions through the collapse by visual inspection, following Wan's example (Wan, this volume).

We repeat that only Figures 6 and 7 correspond to the requirements of the contest (i.e., runaway predictors starting at the end of `A.dat` and predicting the first 100 points of `A.cont`). Both these sets of predictions were based on 32 bits (an initial vector of 4 data points, each of 8 bits). The quality of these predictions and the observation that they, along with the striking *visual pattern matching* predictions presented at the conference, "fail" only at the transition, suggests that a simple model might explain the majority of the dynamics. Such a model would

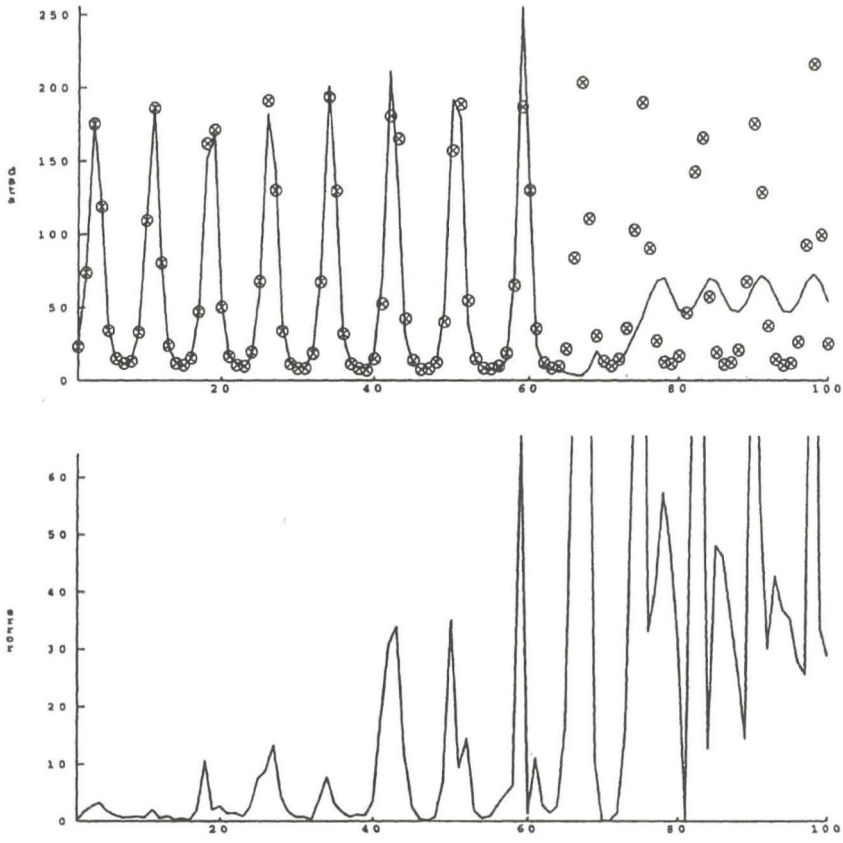


FIGURE 7 Runaway local nonlinear RBF predictions using $\phi = r^3$ with 16 centers distributed between the 32 nearest neighbors of the point to be predicted

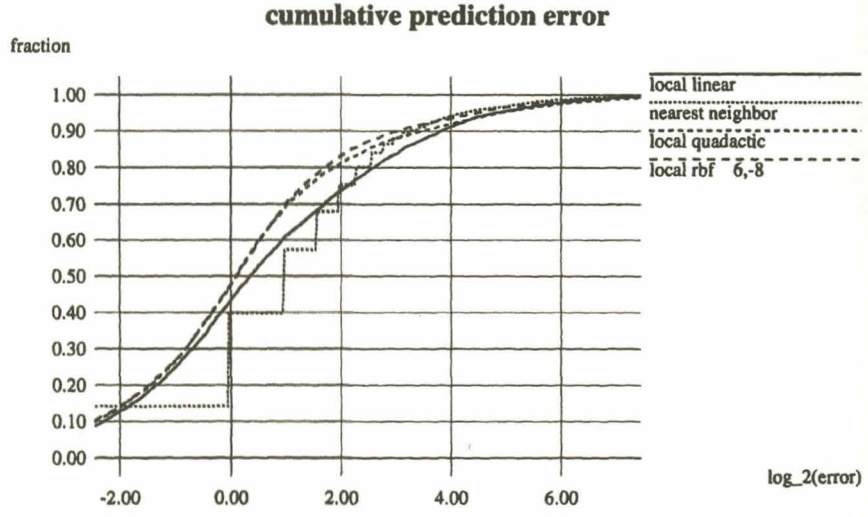


FIGURE 8 Local predictor results for RBF, quadratic, linear, and nearest-neighbor predictors.

be nonlinear, of course (we are measuring $|E^2|$), but quite possibly not chaotic. It would be interesting to determine just how well the amplitude of the cycles after a collapse could be predicted, and also whether those occasions where the amplitude appears to continue to decrease for a few cycles after the collapse could be identified, although the saturation of the signal noted above might make this difficult.

Before pursuing that idea, we compare the local predictors for the case $M = 4$, $\tau_d = 1$, $\tau_p = 4$; the prediction profiles are shown in Figure 8. These graphs reflect the quality of direct update predictions at a fixed step (four samples) into the future. Since the nearest-neighbor predictions are quantized, so are their errors, hence the steplike appearance in the figure. Note that while this predictor gives the fewest accurate predictions (40% of the predictions are accurate to 1 bit or better), it also gives slightly fewer large errors crossing to the left of the other three curves at approximately $\log_2(\text{error}) = 3.5$. The local RBF predictor yields slightly better predictions than the local quadratic, which is significantly better than the local linear. These results are all based on a neighborhood of $k = 32$ nearest neighbors with $n_c = 16$.

We conclude this section by admitting that no motivation for the choice of the particular RBF was given. An advantage of $r^3$ is that it introduces no additional free parameter. In practice, it is often the case that the exponential, $\alpha \approx 1$ with the constant $c = \alpha d_{nn}$, provides a fit of similar quality and that, by adjusting $\alpha$ and observing the predictor profiles, a better fit can be achieved. (Note that $d_{nn}$

is a local quantity.) It is also observed that such a fit is often more parsimonious than the original for, while we have introduced an additional parameter, more of the $\lambda_j$ of the fit are linearly related (due to zeros introduced in a singular value decomposition), thus reducing the total number of degrees of freedom employed in the predictor.

## 4. THE OVERLYING ENGINEERING

### 4.1 NOISE, SENSOR FAILURE, AND ERROR DETECTION

How similar techniques also can be applied to detect sensor failure and error detection is discussed by L. A. Smith et al. (1991). The basic idea is straightforward and, of course, related to our earlier comments on stationarity. If a deterministic system has explored the phase space of a given reconstruction and a reasonably good predictor has been constructed, then the observation that the expected error is persistently, unexpectedly large would indicate either a change in the systems dynamics or an error in the sensor. When a good predictor cannot be constructed, because the required dimension is too great or the underlying system is stochastic, errors can be suspected when the nearest center distance grows very large. This test is applicable to stochastic systems which, while filling regions in any given reconstruction completely, generally do not fill all regions of the embedding space. In the current application an instance is the detection of the string of zeroes in the continuation set; the series goes to zero for seven time steps near data point 6454.

We also note that, as suggested at the conference (Hubner et al., this volume), if the primary source of noise in this system arises from quantization effects, methods of noise reduction (see Grassberger, Schreiber, & Schaffrath, 1991) should be applied here.

## 5. THE UNDERLYING PHYSICS

The contributions presented have demonstrated that stunningly simple and remarkably complex methods of prediction can be applied to this data set with interesting results. But have we suggested anything about the physics or dynamics of the underlying system? The length of time ahead that we were able to predict based on only 32 bits (4 data points of 8 bits each) makes the system appear very predictable; the difficult predictions appear at the collapses, but even they may not be too complex.

Examining a series of local maxima (not shown), we note that they tend to rise until hitting a threshold near 255 and then suddenly collapse. Figure 9 reveals just how linear this progression is by plotting the local maxima of successive cycles

as $\max_{i+1}$ against $\max_i$; the bubble at larger values is due, in large part, to poor estimates of the larger maxima resulting from the large sampling time. We have suppressed a scatter of small maxima (as small as four) which usually arise immediately after a collapse, some of which are due to beating between small cycles and the sampling rate. While the majority of values fall above the diagonal (indicating growing amplitude), a significant number fall below, denoting a different dynamical behavior which can be identified in the series. Note that the saturation at 255 is also visible in Figure 9. Plotting the time between maxima as a function of maxima (see Figure 14 in the Appendix) suggests that the frequency of the oscillation is a decreasing function of amplitude; this may be related to the physical processes occurring in the experiment.

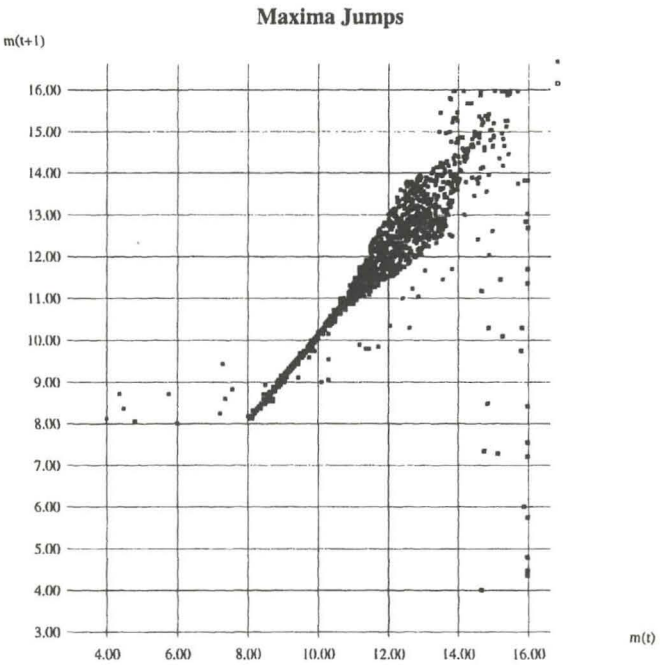


FIGURE 9 A plot of the square root of each local maximum against the one previous showing a clear linear trend. The ballooning at large maxima arises, in part, from variation due to sampling error. The scatter of small local maxima arise just after collapse; however, the interesting sparse band of values just below the diagonal indicates a "slow collapse" ($\max_i > \max_{i+1}$, are real and can be identified in the next two figures).

Assuming that the growing cycles can be modeled, we consider the collapses where most of the prediction results presented at the workshop (which made it that far) went wrong. Phase coherence and amplitude prediction can be considered separately. To investigate whether or not the phase is randomized during collapse, we compare the evolution of data strings that actually saturate at 255. If we line up 11 consecutive instances where the observed maximum was 255, we see a great organization in both the forward and backward time (see Figures 10, 11, and 12). Figures 10 and 11 reveal that the first maxima after the collapse tend to coincide; further, the relation between larger amplitudes and longer period cycles is revealed in Figure 10 as the larger amplitude cycles are clearly displaced to the right as the time since collapse increases. Figure 12, where the direction of time is reversed, shows an extraordinary coherence in phase 9 $\tau_s$ before saturation.

From these observations we conjecture that a predictor based on the values of successive maxima could accurately predict the time and amplitude of the next maximum; initial trails show this breaks down at the collapse, in part, because the saturation at 255 makes these points appear identical.

These observations suggest, as a model, a simple, linearly growing oscillation whose period increases slightly with amplitude. Noting that the signal tends to maintain its phase through the collapse, we have neither determined whether the behavior of the amplitude is random or deterministic, nor explained the interesting segments for which the $\max_i > \max_{i+1}$ just after a collapse. In any event, the

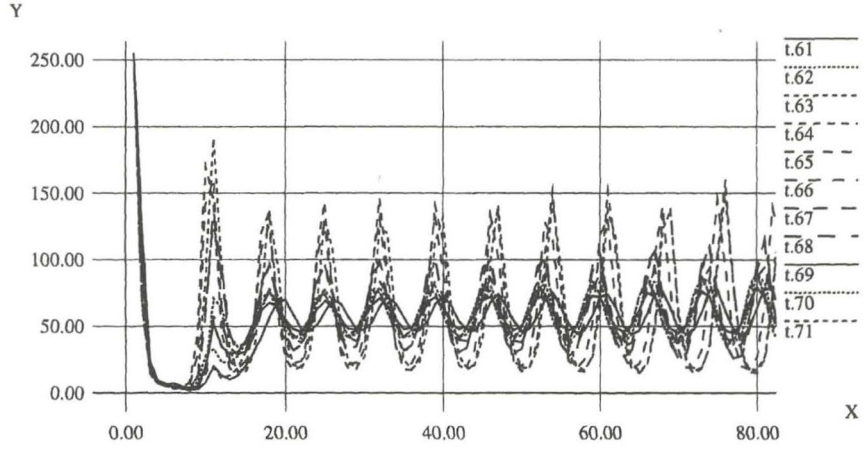


FIGURE 10 Eleven superimposed segments of the time series in `A.cont` aligned by an observation of value 255. Note the phase coherence after the collapse and that the loss of phase coherence toward the right can be attributed to the larger amplitude oscillations having a longer period (as suggested in the text).

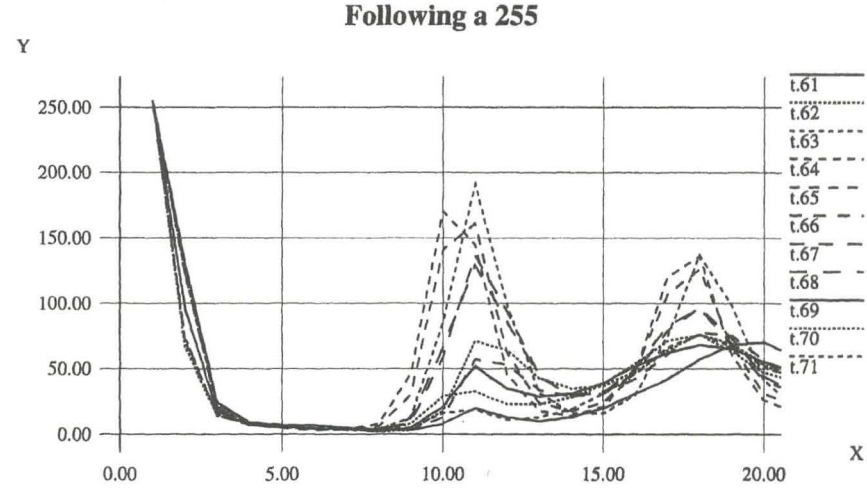


FIGURE 11 An enlargement of the inital section of the previous figure showing the first of two cycles after an observation of 255 in `A.cont`.

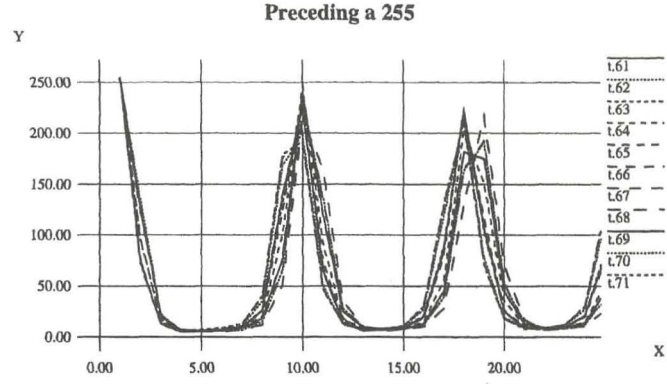


FIGURE 12 Here we show in reversed time (i.e., time increases to the left) the behavior of the series just prior to an observation of 255. This figure is intended to illustrate the coherence of signal just prior to a collapse.

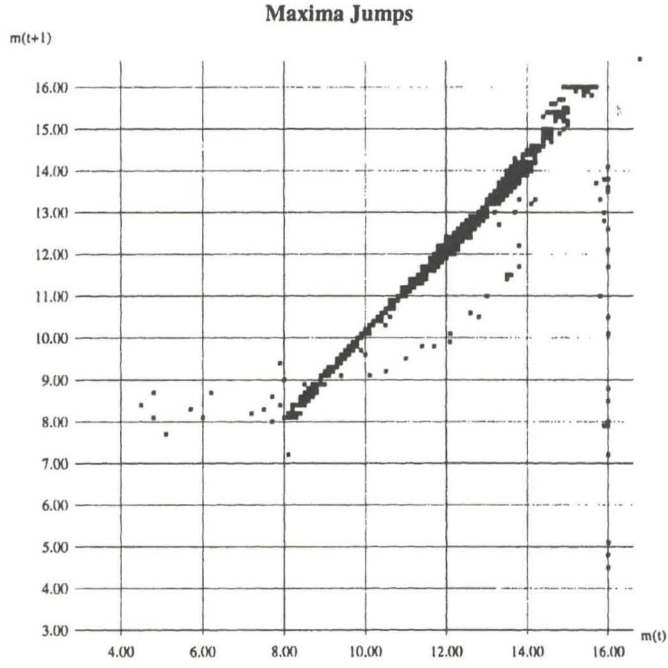


FIGURE 13 A plot of the square root of each local maximum against the previous one for the original (higher resolution) data set. The ballooning at large maxima is decreased. The scatter of small local maxima arise just after collapse; however, the sparse band of values just below the diagonal are real.

phase coherence is interesting and the increase in period may suggest a physical mechanism (such as loss of resonance) that triggers the collapse. It would also be interesting to treat the output of such a simple model (squared and sampled) as surrogate data and to determine whether the actual observations can be distinguished from such series (L. A. Smith, 1992; Theiler et al., 1992a).[3] This approach could be used to discriminate the dynamics of the amplitude after collapses, for instance whether to collapse (retaining phase information) to a random part of the cycle and repeat, or to determine the amplitude after collapse based on the data just before the collapse. While the short-term predictability of this system has been clearly established at this meeting (e.g., the contributions of Sauer (this volume)

[3] Alternatively, one could shuffle the order of segments of the data between collapses; while the majority of statistics computed would remain unchanged (collapses are rare events), long-term determinism would clearly be lost in these surrogates.

and Wan (this volume)), it remains a capital mistake to theorize on the dynamics over the collapse without sufficient data.

## APPENDIX: A CLOSER LOOK

After the conference and discussions with James Theiler and Tim Sauer, some doubts were raised about interpreting the balloon of Figure 9 as due to a long sampling time. To address these doubts, the original data (Hübner et al., 1989) were obtained, thus providing a longer data series at twice the sampling rate. Figure 13 corresponds to Figure 9 for the higher resolution data set. Although not resolving the problem of saturation, we see that the balloon is effectively deflated and this suggests that a piecewise linear map could model the series quite well. Then one could then extend such a model to extrapolate true values corresponding to saturated observations. It would be interesting to see if this would provide a method

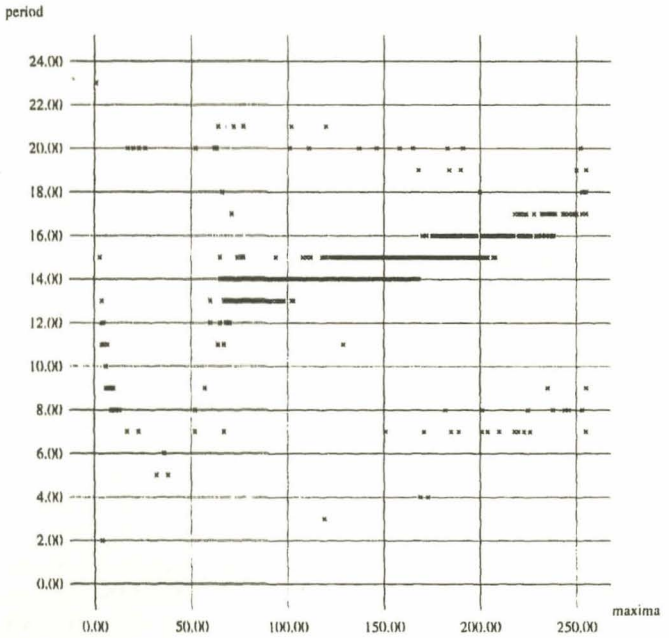


FIGURE 14 A scatter plot of the period between local maximum against the maximum from the full $NH_3$ data set. The majority of that data clusters near periods 13 to 17, with the greater maxima associated with longer periods.

to determine the height of the first maximum *after* a collapse, in particular, to distinguish the collapses of small, increasing amplitude oscillations from those of intermediate amplitude which decline before growing.

In addition, a reviewer requested an illustration of the claim that the period of a cycle tends to increase with increasing amplitude (see Section 5). This effect, also clearer in the higher resolution series, is reflected in the distributions shown in Figure 14. In this figure, there was no attempt to remove the spurious local maxima due to either beating or noise.

## ACKNOWLEDGMENTS

I would like to acknowledge several discussions and ales with M. Muldoon, in particular those regarding stationarity, and an ongoing exchange with J. Theiler vaguely focused on surrogate data. Some of the work reported here was completed during a visit to the UWA during which I enjoyed the hospitality and collaboration of A. Mees and K. Judd. I am grateful for conversations with and data manipulation suggestions by D. Drysdale and D. DeBeer, and many discussions with (as well as the presentations of) other participants at the workshop which contributed to this paper, in particular B. LeBaron's comments on the prediction profiles. Thanks are also due to M. Casdagli, C. Lanone, G. Sugihara, J. Theiler, and A. Weigend for their critical reading of the text. Finally, I owe a special thanks to the organizers, A. Weigend and N. Gershenfeld, for an excellent workshop and for assisting an American in England. This research was supported by a Senior Research Fellowship from Pembroke College, Oxford.