Automated embedding and the creep phenomenon in chaotic time series

by

Masayuki Otani and Antonia J. Jones

Abstract.

Embedding techniques represent a powerful advance in the development of experimental chaos. However there seems no universal method to find the best set of parameters to use. In this paper we present a new approach, an automated embedding method, to estimate a near optimum embedding dimension and delay time based on the Γ-test [Stefánsson 1997]. A strange attractor can be regarded as the union of an infinite number of unstable periodic orbits. In order to extract these unstable periodic orbits from a set of embedding space vectors (used for reconstruction) we introduce a technique to estimate a suitable jump time. We also describe the 'creep phenomenon', discovered using these techniques, which may allow us to make more accurate longer term predictions.

Keywords: Delay coordinates, Embedding dimension, Gamma test, Prediction.

Masayuki Otani

DEPARTMENT OF COMPUTING
IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE,
180 Queen's Gate
London SW7 2BZ

Telephone: +44-(0)-171-594-8281
Telefax: +44-(0)-171-581-8024

Antonia J. Jones

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF WALES, CARDIFF
PO Box 916, Cardiff CF2 3XF

Telephone: +44-(0)-122-287-4812
Telefax: +44-(0)-122-287-4598

Date/version: 14 October 2000

Automated embedding and the creep phenomenon in chaotic time series

CONTENTS

1 Introduction: embedding using delay coordinates 1
2 Choice of the delay time 2
3 Embedding dimension 4
3.1 The gamma test 5
3.2 Application of the gamma test 6
4 Automated embedding method 8
5 Extraction of unstable periodic points 9
5.1 Jump time 9
5.2 Choice of the jump time: the creep phenomenon 10
6 Summary 11
References 12

List of Figures

Figure 1 Irrelevance, $t_D = 0.82$. 2
Figure 2 Redundance, $t_D = 0.05$. 2
Figure 3 $\overline{\Gamma}$ versus m using the average displacement method. 6
Figure 4 $\overline{\Gamma}$ versus m using the autocorrelation method. 7
Figure 5 $\overline{\Gamma}$ versus m for the Rössler model. 7
Figure 6 The slope A versus m for the Rössler model. 7
Figure 7 Reconstructed Rössler attractor with a set of suitable embedding parameters. 10
Figure 8 Reconstructed Duffing attractor with a set of suitable embedding parameters. 10

List of Tables

Table 1 The delay times estimated by the average displacement and the autocorrelation based methods for the three continuous time chaotic systems. 4
Table 2 The delay times, the embedding dimensions, $\overline{\Gamma}$ and A estimated for the three chaotic models. 9
Table 3 Distinct unstable periodic orbits of orders up to 10 extracted - the Rössler model. 12

List of Algorithms

Algorithm 1 The automated embedding method. 8
Algorithm 2 The jump time for close returns. 11

Automated embedding and the creep phenomenon in chaotic time series

by

Masayuki Otani[1] and Antonia J. Jones[2]

[1] *Department of Computing, Imperial College of Science, Technnology and Medicine, 180 Queen's gate, London SW1 2BZ, England, U.K.*

[2] *Department of Computer Science, University of Wales, Cardiff, P.O. Box 916, Cardiff CF2 3XF, Wales, U.K.*

Abstract. Embedding techniques represent a powerful advance in the development of experimental chaos. However there seems no universal method to find the best set of parameters to use. In this paper we present a new approach, an automated embedding method, to estimate a near optimum embedding dimension and delay time based on the Γ-test [Stefánsson 1997]. A strange attractor can be regarded as the union of an infinite number of unstable periodic orbits. In order to extract these unstable periodic orbits from a set of embedding space vectors (used for reconstruction) we introduce a technique to estimate a suitable jump time. We also describe the 'creep phenomenon', discovered using these techniques, which may allow us to make more accurate longer term predictions.

Keywords: Delay coordinates, Embedding dimension, Gamma test, Prediction.

1 Introduction: embedding using delay coordinates.

Let x be a one of a number of scalar variables whose evolution through time is described by a d-dimensional dynamical system. Suppose that x is sampled discretely at at times $t = n\tau$ (n = 1, 2, ...), where τ is the minimimum sampling period. Let us write

$$\xi_i = \left\langle x(it_J),\ x(it_J + t_D),\ .....,\ x(it_J + (d_E - 1)t_D) \right\rangle \tag{1}$$

where t_D is the *delay time* (the time between successive components of the vectors) and t_J is the *jump time* (the time between successive vectors). We call such a construction a d_E *embedding* (of x). Necessarily both t_D and t_J will be (fixed) positive integral multiples of τ.

The use of successive samples of a single variable to generate an embedding with a view to reconstructing the details of an attractor for a higher dimensional dynamic system was suggested in [Packard 1980] and a frequently quoted embedding theorem [Takens 1980] establishes the existence of such models for homogeneous systems: if the underlying state space of a system has d-dimensions then the embedding space needs to have at most $2d + 1$ dimensions to capture the dynamics of the system completely. These results were later generalised and improved in [Levin 1993].

It is a remarkable fact that much of the dynamics of a high dimensional system can be recovered from a suitable embedding of a *single* variable, but in practice critical factors in the accuracy of such reconstructions are the delay time and the embedding dimension. Several techniques have been developed over the years in an attempt to estimate suitable values for these parameters in any particular case. However, it is fair to say that existing techniques suffer from several drawbacks, for example they are usually computationally expensive and often require considerable trial and error.

In this paper we propose an automated technique which is not computationally expensive and appears to produce

very satisfactory results. Using this technique we have found what appears to be a fairly general effect which we have called the *creep phenomenon*: *viz.*

> • Given an embedding dimension and delay time computed as being near optimal for reconstruction we observe that successive points on the section are not unpredictable in their location but have a tendency to be located in a more confined region (a 'cloud') which moves steadily around the attractor as time progresses.

If we wish to predict the dynamical flow in situations where the underlying equations are not known then we might measure all variables of the system and build a non-linear model, such as a neural network to effect the prediction. Practical examples of constructing one-step predictive neural networks based on this idea can be found in [Dracopoulos 1993]. However, if not all system variables are accessible then it becomes necessary to construct an embedding. An embedding which produces a locally smooth mapping of successive points of section in a large region of the reconstructed state space, and which also has a relatively long jump-time, will provide a better non-linear predictive model than one for which the successive jump-time vectors are highly uncorrelated. Hence the existence of the creep phenomenom may enable us to construct more accurate predictive models on the jump-time scale than was hitherto possible.

2 Choice of the delay time.

The delay time is the time period between successive components of each of the embedding space vector. It is written as a multiple of the sampling time, $t_D = \theta\tau$ where $\theta \in \mathbb{N}$. There are various methods to estimate a value of t_D, none of which are universal.

In practice, there is always some noise in the data. Also with a finite amount of data the approximation of the dynamics in the reconstructed state space is never perfect. Therefore, it is important to choose a delay time which will reconstruct the original state space faithfully. It is clear from experiment that the choice of delay time can substntially affect the quality of the reconstructed attractor.

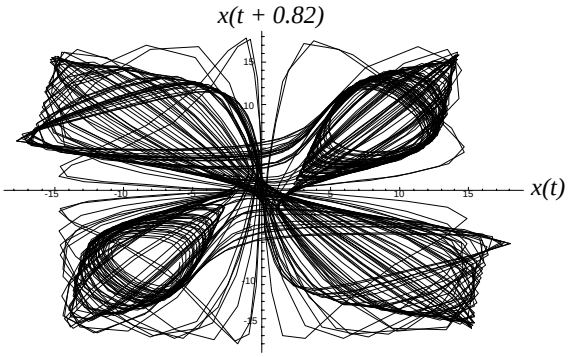

Figure 1 Irrelevance, $t_D = 0.82$.

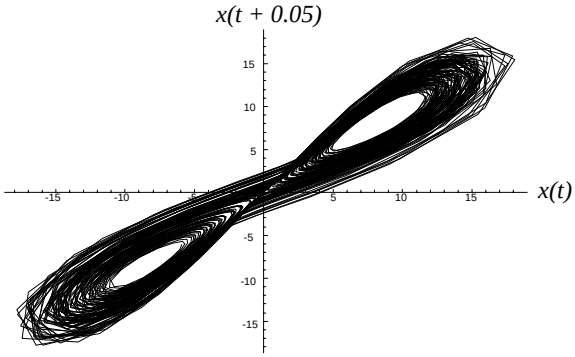

Figure 2 Redundance, $t_D = 0.05$.

If t_D is too large, in the presence of chaos and noise, the dynamics at one time step become disconnected from the dynamics at the next time step, so that even simple geometric objects look extremely complicated. This is called *irrelevance* [Casdagli 1991]. The components of the vectors may be completely decorrelated, resulting in an essentially random distribution of points in the embedding space, leading to an attractor dimension close to the dimension of the embedding space [Albano 1988]. Also, if t_D is close to some periodicity in the system, the component at that period will be under-represented in the reconstruction.

On the other hand if delay time is too small, each component in the vector will be indistinguishable and all trajectories will lie near a diagonal line of the embedding space, leading to a reconstructed attractor with a dimension close to one. This is called *redundance*. To avoid this, t_D should make each component in the vector independent.

Irrelevance and redundance are illustrated using the Lorenz model [Lorenz 1963]. In Figure 2, the delay time (t_D = 0.05) is taken to be too small so that the reconstructed attractor is squashed near to the line of identity. In Figure 1, the delay time is taken to be much larger (t_D = 0.82). This has caused the reconstructed attractor to bear little resemblance to the original attractor, by introducing irrelevant information. There must be an optimum delay time to use for the system reconstruction lying between t_D = 0.05 and t_D = 0.82. An important question is how to find such a delay time?

A number of techniques have been developed with this goal in mind. A comprehensive survey of various methods to estimate the optimum delay time can be found in [Casdagli 1991], see also [Buzung 1992] and [Rosenstein 1994]. It should be noted that none of the algorithms mentioned above or elsewhere are universal, and the estimated t_D usually requires some refinements which are proceeded by trial and error.

A common choice for t_D is to use the autocorrelation function, $g(l)$, which should provide a reasonable measure of the transition from redundance to irrelevance. Autocorrelation based methods have the advantage of short computation times. However, these methods tend to produce inconsistent results, although there is some evidence [Albano 1991] that autocorrelation methods can provide a good initial estimate.

In [Zeng 1991] t_D was chosen as the time at which $g(l)$ first falls to e^{-1}. Similarly, S. J. Schiff and T. Chang [Schiff 1992] have chosen t_D when $g(l)$ was not significantly different from zero for the first time. Yet another method [King 1987] is to choose t_D by locating the first inflection point of $g(l)$. Similarly in [Holzfuss 1986], t_D is chosen to be the time taken for $g(l)$ to reach the local minimum.

Using the method suggested by [Holzfuss 1986] and [King 1987], 0.82 was calculated to be the time to reach the first local minimum of autocorrelation for the Lorenz model. As illustrated in Figure 1, this choice of t_D was found to be inadequate.

Ideally a method to estimate the delay time should be computationally efficient, work well with noisy data and lead to consistent, accurate estimates of key descriptors of the original attractor. It is also important to choose the right overall *time span* covered by the *m*-dimensional embedding space vector which is given by $t_w = (m - 1)t_D$ [Broomhead 1986]. This is the length of the time interval spanned by the first and last components of the vector.

If the delay time is fixed for any embedding dimension, we expect the time span to increase with the dimension. This could cause irrelevance if the dimension is high. However, if we fix the time span with an appropriate delay time, we could limit the rate of expansion of the attractor. Most of the autocorrelation based methods do not make use of the time span, so that once found the delay time is fixed. This may cause the reconstructed attractor to be inaccurate in a multi-dimensional reconstruction.

One recent technique to estimate the optimum delay time has been proposed by M. T. Rosenstein and his colleagues [Rosenstein 1994]. The method is known as the *average displacement method* and it estimates the optimum expansion of the reconstructed attractor from the line of identity in the reconstruction space.

The average displacement function is a measure of an attractor's expansion as a function of delay in *m*-dimensional space. Using the scalar time series it is defined as

$$S_m(n\tau) = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{j=1}^{m-1}(x(i\tau + jn\tau) - x(i\tau))^2} \qquad (2)$$

where N is the number of data points used for the estimate and $t_D = n\tau$. As the delay time increases from zero, the reconstructed trajectory expands from the diagonal and the average displacement increases accordingly until it reaches a plateau. With larger values of m, reconstruction expansion reaches a plateau at smaller value of the delay time, which maintains the time span approximately constant.

It is ideal to recalculate the delay time once the embedding dimension is decided, as the embedding works better if we consider the time span of the vectors rather than just the delay time.

Table 1 The delay times estimated by the average displacement and the autocorrelation based methods for the three continuous time chaotic systems.

Model	τ	Average displacement method gives	Autocorrelation to reach zero	Autocorrelation to reach local minimum point
Lorenz R = 144 b = 8/3 σ = 10	0.01	m = 2, t_D = 0.28 m = 3, t_D = 0.14 m = 4, t_D = 0.11 m = 5, t_D = 0.09 m = 6, t_D = 0.07 m = 7, t_D = 0.06	0.48	0.82
Rössler a = 0.2 b = 0.4 c = 5.7	0.1	m = 2, t_D = 2.2 m = 3, t_D = 1.4 m = 4, t_D = 1.0 m = 5, t_D = 0.8 m = 6, t_D = 0.7 m = 7, t_D = 0.6	1.4	2.9
Duffing oscillator d = 0.2 ω = 0.665, f = 36	0.1	m = 2, t_D = 0.7 m = 3, t_D = 0.5 m = 4, t_D = 0.4 m = 5, t_D = 0.3	2.3	4.7

Table 1 contains different delay times calculated by three different methods for three chaotic models, Lorenz, Rössler [Rössler 1976] and Duffing oscillator [Parlitz 1985]. As before τ is the sampling time. The average displacements were calculated for a number of different dimensions.

3 Embedding dimension

Another difficulty with the reconstruction is the estimation of the dimension to use for the embedding space vectors. In order to successfully implement the delay coordinates method, we must choose the number of dimensions to use in the embedding space.

As mentioned before, it has been shown in [Takens 1981] that a faithful reconstruction can be achieved by at most $2d + 1$ embedding dimension for a d-dimensional system. However, for a dissipative system the effective dimensionality for the long term behaviour is that of the attractor, which is estimated by the Hausdorff dimension. This dimensionality may be considerably smaller than that of the original state space. Thus, we could use the dimension of the attractor, say D, instead of d for the reconstruction.

Once the dimension of the attractor is estimated, the minimal requirement for the embedding dimension is $d_E \geq D$ [Farmer 1987]. Combining this result with that of Takens's theorem, we have

$$D \leq d_E \leq 2D + 1 \tag{3}$$

Therefore, we may need to estimate the Hausdorff of the attractor in order to choose the embedding dimension. A common approach to estimate the Hausdorff dimension (often called, with some degree of imprecision, the

fractal dimension) and subsequently the embedding dimension, is to use the correlation dimension [Grassberger 1983] or the Lyapunov dimension [Kaplan 1979].

A recently developed method called the Γ-test [Stefánsson 1997], [Končar 1997] can also be used to find the embedding dimension, but without estimating the Hausdorff dimension. It can be very effective and less computationally expensive when compared to the use of correlation dimension. The Γ-test algorithm runs in O(*M*log *M*), where *M* is the number of sample data points.

Using the condition in (3), we could limit the range of the embedding dimension once the Hausdorff dimension is estimated. A problem arise as the search area increases with the Hausdorff dimension. For example, if a given model has a Hausdorff dimension of (say) 3.6, then (3) predicts that there is an optimum embedding dimension in the range, $3.6 \le d_E \le 2(3.6) + 1$. As d_E must be a natural number, the range can be interpreted as $4 \le d_E \le 9$. Once this range is known, we must examine the set of embedding space vectors for each dimension within the range which yields the best result. This is a time consuming process. Using the Γ-test, we can predict the embedding dimension without defining the search range.

3.1 The gamma test.

On the assumption of a continuous or smooth underlying input/output model, given the input/output data set, the Γ-test estimates the best mean squared output error that can be achieved without overfitting. Applied in the correct context, in effect the Γ-test estimates that part of the output variance which cannot be attributed to variations due to a continuous or smooth model, i.e. it estimates that part of the output variance due to noise. If the Γ-test result is close to zero, this signifies that the model is deterministic. We denote the input vector by $\mathbf{x} = (x_1, ..., x_n)$ and the scalar output by *y*. In cases where there is more than one output the Γ-test can be applied to each separately with very little extra computational cost. The Gamma test software has been implemented in *C* using (optionally using long double precision arithmetic) to allow accurate processing of very large high precision data sets. It can be downloaded with instuctions from http://www.cs.cf.ac.uk/ec

Suppose the samples are generated by a continuous function f: $\mathbb{R}^m \to \mathbb{R}$ and let *y* be defined as

$$y = f(x_1, ..., x_m) + r \tag{4}$$

where *r* represents an indeterminable part, which may be due to noise or lack of functional determination in the input/output relationship.

The variance of *r*, Var(*r*) provides a lower bound for the mean squared error of the output *y*. If *f* is continuous and there is no noise then Var(*r*) is zero. The Γ-test is a method for estimating Var(*r*).

Suppose $(\mathbf{x}, y)$ is a data sample. Let $(\mathbf{x}', y')$ be a data sample such that $|\mathbf{x}' - \mathbf{x}| > 0$ is minimal. Here $|.|$ denotes Euclidean distance and the minimum is taken over the set of all sample points different from $(\mathbf{x}, y)$. Thus $\mathbf{x}'$ is the nearest neighbour to $\mathbf{x}$ (in any ambiguous case we just pick one of the several equidistant points arbitrarily).

It is well to observe that y' may well *not* be the nearest neighbour of *y* in output space. The dash notation is not ideal but it leads to less complicated expressions than many alternatives.

The Γ-test is based on the statistic

$$\gamma = \frac{1}{2M} \sum_{i=1}^{M} (y'(i) - y(i))^2 \tag{5}$$

Given data samples $(\mathbf{x}(i), y(i))$, where $\mathbf{x}(i) = (x_1(i), ..., x_n(i))$, $1 \le i \le M$, let $\mathbf{x}(N(i, p))$ be the p th nearest neighbour to $\mathbf{x}(i)$. It can be shown that lim γ converges in probability to Var(r) as M -> ∞ and nearest neighbours distances tend to zero. Even the crude measure provided by (5) often proves useful. However, if one is prepared to assume that f is smooth with bounded first partial derivatives we can improve this estimate by making a first order approximation to $f(\mathbf{x}')$ based on Taylor's theorem. By computing a regression line based on some 20 or 30 nearest neighbours one can estimate the intercept $\overline{\Gamma}$ = lim Γ by extrapolating the regression line to the limiting case where the mean-squared nearest neighbour distance $\Delta = 0$.

Given data samples $(\mathbf{x}(i), y(i))$, where $\boldsymbol{x}(i) = (x_1(i), ..., x_m(i))$, $1 \le i \le M$, let $\boldsymbol{x}(N(i, p))$ be the p th nearest neighbour to $\boldsymbol{x}(i)$. Nearest neighbour lists for p*th* nearest neighbours ($1 \le p \le p_{max}$), typically we take p_{max} in the range 20-50, can be found in O(Mlog$M$) time using techniques developed by Bentley, see for example [Friedman 1977].

We write

$$\Delta(p) = \frac{1}{p} \sum_{h=1}^{p} \frac{1}{M} \sum_{i=1}^{M} |\mathbf{x}(N(i, h)) - \mathbf{x}(i)|^2 \tag{6}$$

and

$$\Gamma(p) = \frac{1}{p} \sum_{h=1}^{p} \frac{1}{2M} \sum_{i=1}^{M} (y(N(i, h)) - y(i))^2 \tag{7}$$

then Δ(p) is the mean square distance of the $h \le p$ nearest neighbours and Γ(p) is an estimate for the statistic γ (defined in (5)) based on the $h \le p$ nearest neighbours. We perform a least squares fit on the coordinates $(\Delta(p), \Gamma(p))$ to obtain a regression line in the form of y = Ax + $\overline{\Gamma}$. If the inputs are normalised to a bounded range then the value A is an approximate estimate of the complexity of the input/output model surface.

3.2 Application of the gamma test.

If the inputs used, say x_t, are the past history values of an observable system output sampled with the delay time t_D we can apply the Γ-test to the problem of estimating the embedding dimension.

Let an input vector be defined as

$$\xi_i = \left(x(it_J),\ x(it_J + t_D),\ ...,\ x(it_J + (m-1)t_D)\right) \tag{8}$$

and the output as $x((i+1)t_J)$. Typically as the embedding dimension is increased, $\overline{\Gamma}$ first decreases, reaches a minimum, and then increases. To find the optimum embedding dimension, we can start with m = 2 in (8). The dimension m is increased by one until the minimum of $\overline{\Gamma}$ is attained, say at $n = d_E$, the final estimate for the embedding dimension.

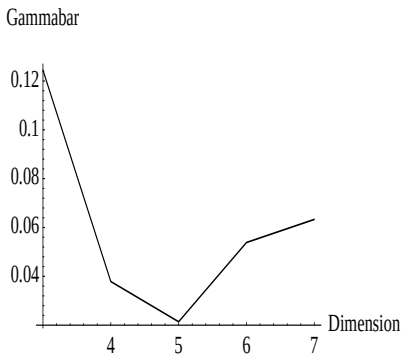

Figure 3 $\overline{\Gamma}$ versus m using the average displacement method.

Figure 3 illustrates a plot of $\overline{\Gamma}$ versus m. It was obtained by using a time series of the x-coordinate for the Lorenz model, with the delay times calculated by the average displacement method. In this case d_E was found to be five with $\overline{\Gamma} \approx 0.021$.

If $\overline{\Gamma}$ is not close to zero, the data set is non-deterministic, therefore we cannot hope to reconstruct the attractor accurately. This may happen if the signal-to-noise ratio is high, or the choice of delay time is poor. It is

interesting to note that the smaller the signal-to-noise ratio the smaller the required embedding dimension. However, the reduction of noise is less when compared to a model with a higher signal-to-noise ratio requiring a higher embedding dimension.

It is very important to use a suitable value of delay time, as the test is sensitive to it. For a deterministic continuous time system, a poor choice of delay time could mean very high values of $\overline{\Gamma}$. With a suitable choice of delay time we would expect $\overline{\Gamma}$ to be small and initially decrease with increasing dimension. For the Lorenz model, the smallest $\overline{\Gamma}$ was found to be approximately 21.52 when the delay time was estimated as the time for the autocorrelation function to reach zero, which was 0.48. Furthermore $\overline{\Gamma}$ did not decrease at all. This is illustrated in Figure 4. This figure of 21.52 should be contrasted with the significantly improved $\overline{\Gamma}$ of 0.021. The large difference between these values highlights the importance of choosing a correct delay time.

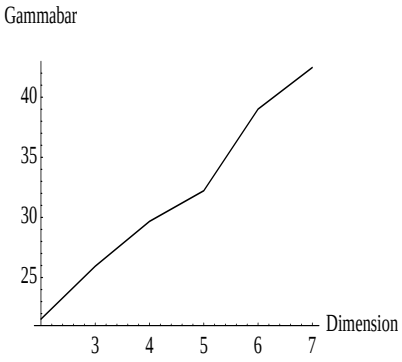

Figure 4 $\overline{\Gamma}$ versus *m* using the autocorrelation method.

Using an estimation of the embedding dimension provided by the Γ-test for the Lorenz, the Rössler and the Duffing oscillator models, the delay times found by the autocorrelation method always performed poorly when compared to ones found by the average displacement method. A reason for this is that the delay time is independent of the dimension, i.e. the delay time is fixed, for the autocorrelation method. Using the average displacement method, delay times were calculated for each dimension as the delay times depended on the dimension.

For both models, the dimension of the attractor was first calculated from the Lyapunov exponents obtained from the time series, see [Sano 1985] for detail. This determined ranges $2.06 \leq d_E \leq 5.12$ for the Lorenz model and $2.01 \leq d_E \leq 5.02$ for the Rössler model. Next, without using this knowledge, t_D and d_E were calculated in each case (using the method described above based on the Γ-test and average displacement). We obtained $d_E = 5$ and $d_E = 4$ for the Lorenz and Rössler model model respective, i.e. in both cases the range conditions were satisfied.

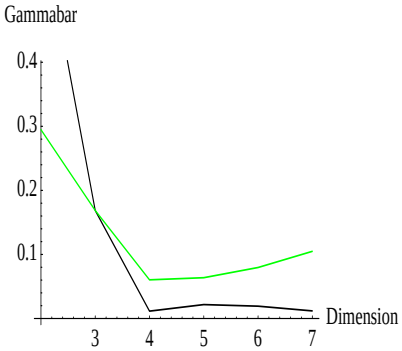

Figure 5 $\overline{\Gamma}$ versus *m* for the Rössler model.

For smaller values of the Gamma test slope estimate *A*, we expect to require fewer data points to reconstruct the state space. For the Rössler model, the same embedding dimension was predicted by the autocorrelation and the average displacement methods. This is illustrated in Figure 5. The plot obtained for the average displacement method is in black and the autocorrelation in grey. Notice that $\overline{\Gamma}$ values are consistently lower for the average displacement method for dimensions of greater than three.

In Figure 6, the plot of the slope versus the dimension reveals that the slopes are also consistently lower for the average displacement method. This strongly suggests that the choice of delay time chosen by the average displacement method is superior to that found by the autocorrelation method.

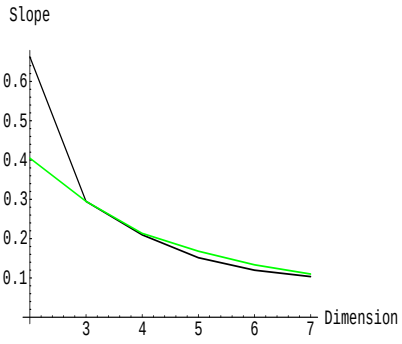

Figure 6 The slope *A* versus *m* for the Rössler model.

We could also use the Γ-test to validate the delay time. If $\overline{\Gamma}$ never reaches zero, or does not initially decrease with increasing dimension for a deterministic system, the choice of delay time must be changed. For continuous time systems with no noise, we expect $\overline{\Gamma}$ to be close to zero if a suitable dimension is used and the delay time is very small. Therefore decreasing the delay time may assist $\overline{\Gamma}$

to be smaller. Of course, we should be careful not to make the delay time too small as this may introduce a high redundance error.

If the time series is obtained from a physical experiment where the underlying mathematical model is unknown, we should at first check if the system is chaotic by the usual tests, see for example [Otani 1997]. A chaotic dissipative system has an attractor and thus the Γ-test should yield $\overline{\Gamma}$ close to zero.

4 Automated embedding method

In studying various techniques, we have devised an approach, an automated embedding method, to obtain a reasonable set of embedding space vectors for the state space reconstruction.

Procedure *Automated embedding* (*X*)

(* *X* is a real time series $\{x(i\tau) : i = 1, 2, 3, ...\}$ of a scalar variable of the dynamic system *)

$dim = 1$

Repeat
  $dim = dim + 1$
  Estimate the delay time from the sequence X by the average displacement (plateau) method as in (2). Call the result $T = T(dim)$
  Create a time series *DTS* of vectors $\xi_i = (x(iT), x((i+1)T), ..., x((i+dim-1)T))$ sampled at $T = T(dim)$, using *X*.
  Create an input/output data set $\{(\xi_i, x((i+dim)T)) : i = 1, 2, 3, ...\}$ *GammaData* for the Γ-test.
  Estimate $\overline{\Gamma}$ by the Γ-test using *GammaData*. Call this *GammaValue*(*dim*, *T*)
Until *GammaValue*(*dim*-1, *T*(*dim*-1)) < *GammaValue*(*dim*, *T*(*dim*)) AND
  GammaValue(*dim*-1, *T*(*dim*-1)) ≈ 0

$t_D = T(dim-1)$
$t_J = t_D$
$d_E = dim-1$

Algorithm 1 The automated embedding method.

The preferred method uses a single coordinate time series of the model to be reconstructed. We set the initial embedding dimension to one. The dimension is increased by one and the delay time is estimated by the average displacement method for that dimension. Delay time may be estimated by other methods which are dependent on the dimension. A time series of a few hundred data points is usually sufficient to calculate the average displacement, typically 500 data points for low dimensional models. We then create the set of input/output pairs using the dimension and the delay time. The statistic $\overline{\Gamma}$ is estimated by the Γ-test using the input/output pairs. The estimations of the delay time and $\overline{\Gamma}$ are repeated for a number of dimensions until a local minima of $\overline{\Gamma}$ (which hopefully is close to zero) is found. We choose the optimum embedding dimension d_E, and delay time t_D, as the dimension and the corresponding delay time which gave the local minima of $\overline{\Gamma}$.

We have found that this method has a significant speed advantage when compared to methods based on estimation of the Hausdorff dimension. Table 2 contains the delay times and the embedding dimensions found to be suitable by the automated embedding method for the three chaotic systems.

The automated embedding procedure is summarised in Algorithm 1.

Table 2 The delay times, the embedding dimensions, $\bar{\Gamma}$ and A estimated for the three chaotic models.

Model	Delay time, t_D	Embedding dimension, d_E	$\bar{\Gamma}$	A
Lorenz	0.09	5	0.021393	0.283369
Rössler	1.00	4	0.011508	0.209624
Duffing's oscillator	0.40	4	0.010151	0.410428

5 Extraction of unstable periodic points.

The orbit of any point p on the strange attractor will make arbitrarily close returns to p. It has been suggested [Gunaratne 1989] that nearly periodic orbits are dense on the strange attractor and that since the motion on the attractor is chaotic, these orbits are unstable. Thus, a chaotic system can be visualised as a collection of many different nearly-periodic behaviours. On the surface of section, an unstable orbit of period k appears as k points. A point belonging to such a set is referred to as an unstable periodic point of period k. Unstable periodic orbits have been successfully extracted from experimental time series [Auerbach 1987], [Belmonte 1988], [Lathrop 1989], [Gunaratne 1989].

We are interested in extracting sets of points representing the unstable periodic orbits of a strange attractor, described by a set of embedding space vectors with the parameters estimated by the automated embedding method.

5.1 Jump time

The jump time t_J, is the time interval between successive vectors. It is also written as a multiple of the sampling time, $t_J = \Phi\tau$ where $\Phi \in \mathbb{N}$. The topology of the reconstructed attractor is independent of t_J, however the sequence of points generated for the attractor and the number of embedding space vectors required to obtain the attractor plot are affected by the choice of t_J.

The effect of jump time is such that for small values, typically t_J several times smaller than t_D, the successive embedding space vectors will be similar and thus the trajectory is well defined. The disadvantage in taking t_J small is that we need to obtain significantly more embedding space vectors to reconstruct topology of the attractor. If t_J is large, typically several times larger than t_D, we need fewer embedding space vectors for the reconstruction, but it will not be as well defined. Also the time taken to obtain each embedding space vector increases proportionally to t_J in a physical experiment.

If we wish to use the embedding space vectors to reconstruct the attractor with a reasonable number of *almost* periodic points, it is ideal to set t_J as the average time for the system to go around the original state space once. However, when we try to reconstruct the attractor from a time series of a physical experiment, in most cases we do not know if there is a forcing frequency or the average time for the system to go around the original state space once. In such a situation, estimation of t_J can be difficult.

When the system is driven by an external periodic force, the period of the external drive establishes a natural period for sampling the dynamics of the system. In such a case, we sample some variable of the system at a

particular phase of the external force to form a surface of section. In this case the sampling time τ is the period of the external force. It is then reasonable to use τ as both the t_D and t_J since the variable samples are already reasonably separated in time.

5.2 Choice of the jump time: the creep phenomenon

When using embedding techniques to model a chaotic system, we must be careful in choosing a value of t_J, so that there are a reasonable proportion of successive vectors which are close to each other.

One could estimate a reasonable t_J by comparing successive embedding space vectors, say ξ_i and ξ_{i+1}, for increasing values of t_J (in steps of τ) for $t_J > t_W$. If there are many vectors such that $|\xi_i - \xi_{i+1}| < r_1$, where r_1 is typically two to three orders of magnitude smaller than the signal, the jump time used to create the set of embedding space vectors should be used. This is because any faithful reconstruction of the attractor should contain many close returning orbits. In the attractor, the vectors near an unstable periodic point of period one should stay nearby for a number of iterations. Therefore we should observe a good proportion of the vectors say 10%, satisfying the condition $|\xi_i - \xi_{i+1}| < r_1$, assuming the set of the vectors used is large. We have typically used 500 vectors for this estimation. If the system has a natural forcing frequency, we should be able to estimate it by the method explained above and successive vectors should remain nearby.

In Figure 7 the Rössler attractor is reconstructed by a set of 2,500 embedding space vectors with $d_E = 4$, $t_D = 1.00$ and $t_J = 17.9$. The plot is a two-dimensional representation of the four-dimensional state space, $x(t_J)$ versus $x(t_J + t_D)$. The dark region in the forth quadrant of the plot represents 50 successive points falling on the attractor. The iterates of points move clockwise and after 500 points, reach the dark region in the first quadrant of the plot, represented by another 50 successive points. As can be seen, the successive points remain nearby. We have called this effect the *creep phenomenon*.

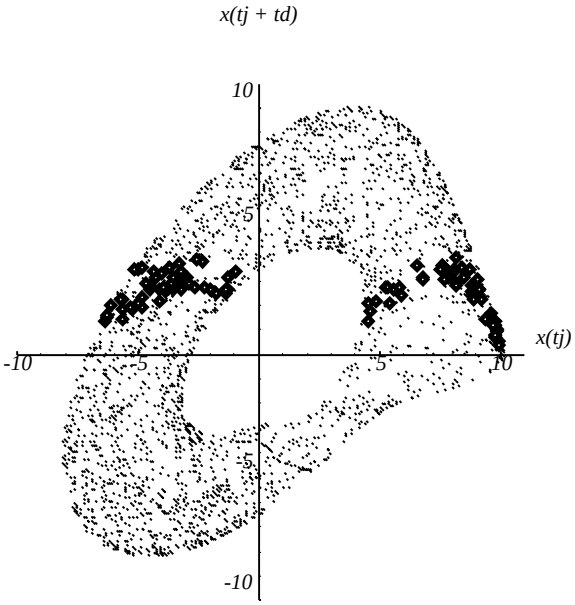

Figure 7 Reconstructed Rössler attractor with a set of suitable embedding parameters.

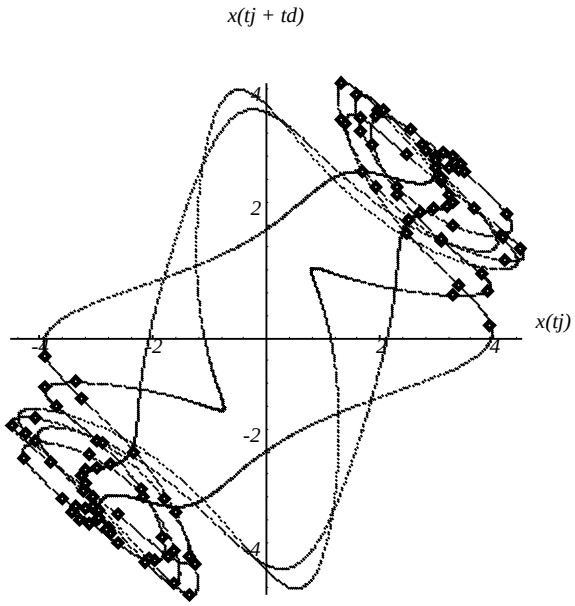

Figure 8 Reconstructed Duffing attractor with a set of suitable embedding parameters.

In Figure 8, the Duffing attractor is reconstructed by a set of 5,000 embedding space vectors with $d_E = 2$, $t_D = 0.7$ and $t_J = 9.4$. The model has natural oscillations in units of $2\pi/\omega$. Using the set of parameters as in Table 1, the forcing frequency is one cycle per 9.449 secs. Our method successfully predicted the jump time $t_J \approx 9.4$ very close to this time. The method could have estimated a more accurate jump time if we had used a smaller sampling time, for example 0.01 instead of 0.1. The same can be said for the case of the Rössler attractor. The creep phenomenon is also observed in Figure 8. Here, the first 50 successive points are in the third quadrant. After 100 points, the second set of 50 points moved to the first quadrant of the plot.

These are particularly interesting observations in that a relatively long term prediction, 17.9 for the Rössler and 9.4 for the Duffing oscillator models, can be made quite accurately with a good choice of the jump time.

Procedure *Jump time* (X, t_D, d_E)

(* X is a real time series $\{x(i\tau) : i = 1, 2, 3, ...\}$ of a scalar variable of the dynamic system *)

(* t_D and d_E are estimated by the automated embedding - Algorithm 1 *)

Choose r = two to three orders of magnitude smaller than the values in X

$t_J = t_D - \tau$

Repeat
  $t_J = t_J + \tau$
  Create a sequence of embedding space vectors ξ_n from X using t_D, t_J and d_E as in (1)
  counter = 0
  N = Length of ξ
  For n = 1 to N-1
    If $| \xi_n - \xi_{n+1} | < r$
      then *counter* = *counter* + 1
  endfor n
Until *counter*/N > 0.1
 (* at least 10% of the successive embedding space vectors are close returns *)

Return (t_J)

Algorithm 2 The jump time for close returns.

One method to predict the next point from the current point is to assume local linearity and estimate a Jacobian matrix of the dynamics at nearby points and then apply this to the current point. In general one could create a non-linear model using a neural network. Generally, with a poor choice of t_D it is difficult to successfully construct a smooth non-linear model which accurately predicts where the next point will fall. Furthermore, we may not be able to locate any unstable periodic points. A procedure to estimate a value of jump time which yields sufficient number of close returns is summarised in Algorithm 2.

Using the Rössler model we have created two sets of 10,000 and 2,500 embedding space vectors with d_E = 4, t_D = 1.00 and t_J = 17.9 to extract unstable periodic orbits of periods up to 10. We have used the extraction method proposed in [Auerbach 1987] where r_1 (above) was set to be 0.125. The number of distinct unstable periodic orbits extracted are as in Table 3.

6 Summary

We have described an automated method for choosing a delay time and embedding dimension which facilitates an accurate reconstruction of the high dimensional dynamics. This technique is based on the Γ-test and the average displacement method the combination of which is computationally relatively inexpensive. Of course, delay time may be estimated by other methods which are however dependent on the embedding dimension. Since embedding techniques are widely used to model a physical system in cases where the mathematical description is unknown, such an automated reconstruction method has wide applicability.

The choice of delay time t_D is important, as a good choice can reduce both the amount of data required and the effect of noise. Different techniques briefly described were autocorrelation based methods and the average displacement method. Throughout our experiments, we have consistently found the average displacement method yields better results than autocorrelation based methods.

The choice of the embedding dimension d_E is also an important parameter, as no matter what delay time we

choose, if the embedding dimension is too small we will not be able to reproduce the full system behaviour.

Table 3 Distinct unstable periodic orbits of orders up to 10 extracted - the Rössler model.

Period (Multiples of t_J)	Number of orbits extracted using 10,000 embedding space vectors	Number of orbits extracted using 2,500 embedding space vectors
1	1	1
2	62	17
3	95	30
4	80	25
5	25	9
6	19	3
7	13	6
8	11	0
9	5	3
10	3	0
Total	315	95

The trajectory of the reconstructed attractor is well defined for a small jump time t_J, but more vectors are required for the reconstruction.

A method to estimate a suitable jump time to locate unstable periodic orbits from a set of embedding space vectors was discussed in detail and these studies led to the discovery of the creep phenomenon. The creep phenomenon may assist us in making longer term predictions of the system behaviour given a chaotic time series of a single variable. The extraction method was used to locate unstable periodic orbits with periods up to 10 for the Rössler model using two different sized sets of embedding space vectors employing the parameters t_D and d_E estimated by the automated embedding method and the technique to find a suitable jump time t_J.

References

[Albano 1988] A. M. Albano, J. Muench, C. Schwartz, A. I. Mees and P. E. Rapp. *Singular-Value Decomposition and the Grassberger-Procaccia Algorithm*, Physical Review A **38**:3017-3026, 1988.

[Albano 1991] A. M. Albano, A. Passamante and M. E. Farrell. *Using higher-order correlation to define an embedding window*, Physica D **54**:85-97, 1991.

[Auerbach 1987] Ditza Auerbach, Predrag Cvitanovic, Jean-Pierre Eckmann, Gemunu Gunaratne and Itamar Procaccia. *Exploring Chaotic Motion Through Periodic Orbits*, Physical Review Letters **58**:23, 2387-2389, 1987.

[Belmonte 1988] A. L. Belmonte, M. J. Vision, J. A. Glazier, G. H. Gunaratne and B. G. Kenny. *Trajectory Scaling Functions at the Onset of Chaos: Experimental Results*, Physics Review letters **61**(5):539-542, 1988.

[Broomhead 1986] D. S. Broomhead and G. P. King. *Extracting qualitative dynamics from experimental data*, Physica D **20**:217-???, 1986.

[Buzung 1992] Th. Buzung and G. Pfister. *Comparison of algorithms calculating optimal embedding parameters for delay time coordinates*, Physica D **58**:127-137, 1992.

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

[Dracopoulos 1993] D. C. Dracopoulos and Antonia J. Jones. *Neuromodels of analytic Dynamic Systems*. Neural Computing & Applications, **1(4)**:268-279, 1993.

[Farmer 1987] J. D. Farmer and J. J. Sidorowich. *Predicting Chaotic Time Series*, Physical Review Letters **59(8)**:845-848, 1987.

[Friedman 1977] J.H. Friedman, J.L. Bentley and R.A. Finkel. *An algorithm for finding best matches in logarithmic expected time.* ACM Transactions on Mathematical Software, **3**(3):200-226, 1977.

[Grassberger 1983] P. Grassberger and I. Procaccia. *Characterization of Strange Attractors*, Physical Review Letters 50(5):346-349, 1983.

[Gunaratne 1989] G. H. Gunaratne, P. S. Linsay and M. J. Vision. *Chaos beyond Onset: A Comparison of Theory and Experiment*, Physics Review Letters **63(1)**:1-4, 1989.

[Holzfuss 1986] J. Holzfuss and G. Mayer-Kress. *An approach to error estimation in the application of dimension algorithms*, in Dimensions and Entropies in Chaotic Systems, editor G. Mayer-Kress, Springer-Verlag, New York, 114-122, 1986.

[Kaplan 1979] J. Kaplan, J. A. Yorke in *Chaotic behavior of multidimensional difference equations*, H. O. Peitgen et al., Eds., "Springer Lecture, Notes in Mathematics", Springer-Verlag, Berlin, **730**:204-227, 1979.

[King 1987] G. P. King, R. Jones, D. S. Broomhead. *Phase portraits from a time series: a singular system approach*, Nuclear Physics B **2**:379-???, 1987.

[Končar 1997] *Optimisation methodologies for direct inverse neurocontrol*. N. Končar. Ph.D. Thesis, Department of Computing, Imperial College of Science, Technology and Medicine, London 1997.

[Lathrop 1989] Daniel P. Lathrop and Eric J. Kostelich. *Characterization of an experimental strange attractor by periodic orbits*, Physical Review A **40(7)**:4028-4031, 1989.

[Levin 1993] A. U. Levin. *Recursive Identification Using Feedforward networks.* Department of Computer Science and Engineering, Oregon Graduate Institute, 2000 NW Walker Road. PO Box 91000, Portland, Oregon OR 97291-1000.

[Lorenz 1963] E. N. Lorenz. *Deterministic Nonperiodic Flow*, J. Atmospheric Sciences **20**:130-141, 1963.

[Otani 1997] M. Otani and Antonia J. Jones, *Guiding Chaotic Orbits*, Research Report, Department of Computer Science, University of Wales, Cardiff, P.O. Box 916, Cardiff, Wales, U.K. 1997.

[Packard 1980] N. H. Packard, J. P. Crutchfield, J. D. Farmer and R. S. Shaw. *Geometry from a time series*, Physics Review Letters **45(9)**:712-16, 1980.

[Parlitz 1985] U. Parlitz and W. Lauterborn. *Superstructure in the bifurcation set of the Duffing equation*, Physics

Letters A **107(8)**:351-355, 1985.

[Rosenstein 1994] M. T. Rosenstein, J. J. Colins, C. J. De Luca. *Reconstruction expansion as a geometry-based framework for choosing proper delay times*, Physica D **73**:82-98, 1994.

[Rössler 1976] O. E. Rössler. *An Equation for Continuous Chaos*, Physics Letter A **57**:397-398, 1976.

[Sano 1985] M. Sano and Y. Sawada. *Measurement of the Lyapunov Spectrum from a Chaotic Time Series*, Physical Review Letters **55(10)**:1082-1085, 1985.

[Schiff 1992] S. J. Schiff and T. Chang. *Differentiation of linearly correlated noise from chaos in a biologic system using surrogate data*, Biological Cybernetics **67**:387-393, 1992.

[Stefánsson 1997] Aðalbjörn Stefánsson, N. Koncar and Antonia J. Jones, *A note on the Gamma test*, Neural Computing & Applications, **5**:131-133, 1997.

[Takens 1980] F. Takens. *Detecting strange attractors in turbulence.* In A. Dold and B. Eckmann, editors, Dynamical Systems and Turbulence, pp 366-381. Springer-Verlag, 1980.

[Takens 1981] F. Takens. *Dynamical Systems and Turbulence.* Lecture Notes in Mathematics 898, D. A. Rand and L. S. Young, eds, Springer-Verlag, 1981.

[Zeng 1991] X. Zeng, R. Eykholt and R. A. Pielke. *Estimating the Lyapunov-Exponent Spectrum from Short Time Series of Low Precision*, Physical Review Letters **66(25)**:3229-3232, 1991.