Model Misspecification Tests, Model Building and Predictability in Complex Systems.

Radu Manuca and Robert Savit
Physics Department
and
Program for the Study of Complex Systems
University of Michigan
Ann Arbor, MI 48109

ABSTRACT

In this paper we show that significant advantages can be realized by using functions of simple variables, such as time lags, as directions in reconstruction spaces for complicated time series. First, we show that more sensitive model misspecification tests can be constructed when the candidate model is used as one of the directions in the reconstruction space. This naturally results in misspecification tests based on the construction of conditional probabilities which can reveal misspecified models even when other powerful methods, such as the BDS test applied to a sequence of residuals, fails. Second, we show that model building and predictability of the time series can be substantially improved by using an informational criterion to determine the functions with which to associate the directions of the reconstruction space. This criterion is of the form of a conditional probability and is related to a measure of the short term predictability of the time series. We consider specifically an example in which the search space of functions is linear combinations of time-lags, and we compare the resulting model with a linear least squares fit to the data. We also show that this optimized reconstruction space can improve longer-term predictability. We demonstrate this by comparing longer-term predictions made using the optimized space with predictions made using a simple time-lagged reconstruction space.

I. Introduction

A number of methods for the analysis and description of nonlinear processes are based on the study of a data set produced by the system in a reconstruction space. Generally, one studies data sets in a reconstuctions space whose directions are simple lags (if the series is descrete) or derivatives (if the series is continuous), or some other equally simple quantity. Although there has been some work in choosing the best reconstruction directions for different purposes[1], the common practice is to rely on relatively simple quantities for the directions. This attitude is quite understandable, and follows, we believe, in no small part form the influence of the powerful results of Takens and Ruelle[2], which demonstrate that many of the important topological properties of an attractor, are independent of the details of the reconstruction space.

However, when studying real data sets which may be generated by noisy complex processes, one if often interested in aspects of the system other than the topological properties. The problem of extracting information from such data sets, and developing statistically sound conclusions may also be complicated if the system is strongly data-limited. In such cases, it may be advantageous to consider reconstruction spaces whose directions are other than simple variables such as lags or derivatives.

In this paper we describe the benefits of allowing the directions of the reconstruction space to be more complicated functions of the variables in the problem. In any reconstruction space, it is natural to consider probabilities for continuing closeness of sets of variables. Such probabililties have been used by Brock, Dechert and Scheinkman[3] to study residuals of time series for randomness, and by Savit and Green[4] to look for dependencies among sets of variables in a data set. Here we shall consider similar conditional probabilities, but in which some of the arguments are functions rather than simple variables. In particular, we shall describe the application of this technique to two common problems faced in the analysis of complex systems, namely, model misspecification testing and model building. In the case of model misspecification problems, we will show that very powerful discriminants of misspecification can be constructed by allowing one of the directions in the reconstruction space to be the candidate model itself. Conditional probabilities, related to combinations of Grassberger-Procaccia correlation integrals can be constructed which provide very strong and specific tests against model misspecification. In the case of model building, we show that the reconstruction space can be optimized according to an informational measure which is directly related to the predictability of the series. To demonstrate the efficacy of this reconstruction we shall study a discretized time series deduced from the Lorenz system and we shall show that when overlaid by a simple prediction algorithm, predictions made with this optimized reconstruction are significantly better than those made using a construction based on simple lags.

The rest of this paper is organized as follows: In the next section, we construct the conditional probabilistic indicators we shall use, and describe their meaning, in general. In Section III, we study the problem of model misspecification, by analyzing a time series generated by summing four independent Henon maps. This series was studied previously by Theiler and Eubank[5], who showed that tests of the series of residals from a linear AR fit failed to reject the null hypothesis of IID for the residuals. We show that our appoach is able to reject the linear AR fit as a good explanation of the data. In Section IV we turn to the problem of model building. Here we show that an improved reconstruction space (which we call optimized) can be defined by maximizing a set of conditional probabilities (related to predictability), whose arguments involve functions of the variables in the problem. Using as an example a discretized version of a time series generated by the Lorenz system, to show that predictions made using this optimized reconstruction space are superior to those made using a reconstruction space based on lags of the time series. We also show that, at least in this example, the maximum necessary embedding dimension for the series is reduced by one by using the otpmized variables, rather than simple lags. This is a significant advantage in data-limited situations. Section V consists of a a summary and discussion. Some technical and subsidiary issues are discussed in two appendices. In particular, in Appendix B we show that in a complete search space our method leads to the correct description of the data set, (as do a number of other approaches).

II. Conditional Probabilities and Dynamical Models

In this section we shall introduce the quantities that we shall use to delimit a variable space and to test specific models. For simplicity, we shall first discuss the case of a single time series, and an *a priori* delimited search space of the lags of that time series. Extension of the method to more general cases will be obvious.

Consider a time series x(t). Suppose for the moment that t is discrete. Let σ_x be the standard deviation of the values of the time series x(t), and let μ be real number (typically, $0.05 \le \mu \le 2.0$) Define $\varepsilon_x = \mu\sigma_x$. Now consider the d+1-dimensional vector

$$v(i) = (x(i), x(i-1), \ldots, x(i-d)) = (v_0(i), v_1(i), \ldots, v_d(i)). \qquad (1)$$

Let t_k stand for the statement $|v_k(i)-v_k(j)| \le \varepsilon_x$.

Suppose we have a dynamical model, f, either deterministic or with additive noise, for the generation of the time series, such that f(i) estimates x(i). We assume that f depends on a set of variables, say lags of the time series, and is an autonomous function in the search space. In particular, we do not intend f to be a simple time-dependent fit to the data. So, for example, if x(t) were sinusoidal, we would seek to test f's which were functions of the first and second lags, x(t-1) and x(t-2), and *not* f's of the form $\sin\omega t$. Let F stand for the statement $|f(i)-f(j)| \le \varepsilon_f$, where $\varepsilon_f = \mu\sigma_f$, and σ_f is the standard deviation of the values of f(i). We emphasize that f(i) does not necessarily represent the true dynamics of the system, but is in general only a model for those dynamics.

An example of the kind of quantities we shall be concerned with is the conditional probability of the form $P(t_0 \mid F, t_k)$, where $k \le d$. In the following paragraphs, we shall use this example to introduce the ideas behind the methods we are presenting. Generalizations will be discussed below. In words, this expresses the probability that two values of x are within ε_x if the values of a model function for those x's are within ε_f and if the kth lags are also within ε_x. The *cognoscenti* will recognize in this quantity an underlying picture, in which the usual phase space embedding philosophy for the analysis of complex systems is modified to include embedding spaces in which one of the directions is the model itself, rather than a simpler variable such as a time lag. Depending on the choice of f, many of the powerful geometric and topological theorems of strange attractors may not apply in this space. That, however, is a secondary consideration for us here.

Suppose now that we want to test how well a dynamical model, f, captures the information in the data x(t). We shall present two kinds of measures. The first are global, and are direct measures of the additional explanatory power in f, compared with the absence of a model. The second are more specific measures that indicate additional dependence of the data on a subset of variables in the search space. This second class of measures can be used generically in two ways: First, one may check for additional dependence of the data on a variable that already appears in the model function, f. Used in this way, these measures are tools for helping to tune the parameters of the model. In addition one may check for dependencies on variables that do not appear in f. This application can tell us something about the relevant search space of variables for our problem. This second application raises a number of interesting problems which we shall relegate to a future work.

A. Global Measures

Consider the quantity

$$\Phi = \frac{P(t_0 | F) - P(t_0)}{P(t_0 | F)} = 1 - \frac{P(t_0)}{P(t_0 | F)} \qquad (2)$$

If the model function f has no explanatory power for the data x, then the conditional probability $P(t_0 | F)$ should equal the unconditional probability $P(t_0)$, and so in an appropriate statistical sense, $\Phi = 0$. The extent to which Φ deviates from zero is a measure of the extent to which f is a good explanation for the data. We shall return below to a discussion of the statistics of (2) and a statement of the null hypothesis.

Another quantity which also provides a global characterization of the function f, is

$$\tilde{\Phi} = \frac{P(t_0 | F, t_1, t_2, \ldots, t_d) - P(t_0 | F)}{P(t_0 | F, t_1, t_2, \ldots, t_d)} = 1 - \frac{P(t_0 | F)}{P(t_0 | F, t_1, t_2, \ldots, t_d)}. \qquad (3)$$

In $P(t_0 | F, t_1, t_2, \ldots, t_d)$, we have included all the variables upon which x(i) may depend. Although (2) and (3) are both global measures, they have different interpretations and are in some sense complementary. Φ is a measure of the amount of information contained in f, whereas $\tilde{\Phi}$ is a measure of the amount of information from the search space *not* contained in f. (2) measures the extent to which values of f that are close are associated with values of x that are close. And roughly speaking, (3) is zero if the dependence on the variables in the search space (in our case the first through d^{th} lags) is correctly included in f. Deviations of (3) from zero indicate that the inclusion of other variables in f, or a change in the parametric form of f involving those variables would improve the ability of f to explain the data.

B. Local Measures

As stated above, we shall limit ourselves to measures of dependence on variables that are already contained in the function f. With this in mind, we will consider two objects. The first is

$$\xi_k = \frac{P(t_0 \mid F,t_k) - P(t_0 \mid F)}{P(t_0 \mid F,t_k)} = 1 - \frac{P(t_0 \mid F)}{P(t_0 \mid F,t_k)} \qquad (4)$$

If f contains an accurate parameterization of the dependence on the k-th lag, then we expect ξ_k to be zero in an appropriate statistical sense (see below). In fact, a more precise way to describe the information we can deduce from a calculation of (4) is to note that: 1.) If f is the correct explanation of the data, then surely, ξ_k will be statistically zero, and 2.) if (4) fails to be zero, then we expect that it will have specific power against a misspecification of the dependence on the kth lag in the function f. Equation (4) can also be used to study dependence on lags which are not explicitly contained in f.

An apparently closely related statistic to (4) is the function

$$\rho_k = \frac{P(R_0 \mid t_k) - P(R_0)}{P(R_0 \mid t_k)} = 1 - \frac{P(R_0)}{P(R_0 \mid t_k)} \qquad (5)$$

In (5), R_0 stands for the statement $|r(i)-r(j)| \leq \varepsilon_r$, where $\varepsilon_r = \mu\sigma_r$, and σ_r is the standard deviation of the values of the residuals of the fit: $r(i) = x(i) - f(i)$.

This quantity directly probes the additional dependence of the residuals on the k^{th} lag of the data. Although (4) and (5) are closely related, they have rather different numerical and statistical properties. These different behaviors have important implications for the process of searching for an optimal description of the process. Here we digress briefly to compare the typical behavior of ξ_k and ρ_k in a chaotic system with and without noise. The reader uninterested in these details may skip to the beginning of section III.

Let us suppose we have a model for x(i) which satisfies $x(i) = f(i) + r(i)$ with $\langle r \rangle \sim 0$ so that we may take σ_r as a measure of the typical magnitude of r(i). As an example, consider a set of data generated by the Henon map in the chaotic regime,

$$x(i) = 1 - 1.4x(i-1)^2 + 0.3x(i-2). \qquad (6)$$

Let us calculate ξ_2 and ρ_2 using the Henon map time series with 8,192 points with $\mu = 0.05$ and a function

$$f(i) = x(i-1)^2 - ax(i-2). \qquad (7)$$

(As we shall explain in section IV, the additive constant and an overall multiplicative factor in the model f are irrelevant for our results here.) The only important factor is the ratio of the coefficient of the $x(i-1)^2$ and the $x(i-2)$ terms. This is the value, a. In Fig. 1 we have plotted ξ_2 and ρ_2 as a function of a. Both ξ_2 and ρ_2 are essentially zero at the correct value of $a = a' \approx 0.22$, but their behavior away from a' is quite different. To a first approximation, ρ_2 is nearly constant away from a', while ξ_2 smoothly drops to zero. Notice also that for $0.18 < a < 0.26$, ξ_2 is essentially linear. This range of a corresponds to values of σ_r which are less than ε. Because a is the coefficient of a linear term in f, the linear dependence on a is roughly equivalent to a linear dependence on σ_r. In fact, it is not difficult to show in general (see Appendix A) that if $\sigma_r \sim O(\varepsilon)$, then $\xi_k \sim O(\sigma_r^1)$ and $\rho_k \sim O(\sigma_r^0)$.

Depending on the nature and goal of the search algorithm, ξ_2 or ρ_2 may be preferable as an indicator of additional dependence. For example, in a much larger search space, one might try to use genetic algorithms to find a best model for the data. In that case, ξ_2 might be preferable to ρ_2 as a fitness function, since the sharp dip in ρ_2 could be easily missed. The same comments, of course, apply to other search algorithms in which one is looking coarsely at the space. At the last stages of a search algorithm, where great sensitivity to parameter choices is required, ρ_2 may be the preferred indicator.

It is also useful to see how the sensitivity of our indicators is affected by the presence of noise in the data set. To this end we compute Φ, ξ_2, and ρ_2 for a data set generated by the Henon map with additive noise:

$$x(i) = 1 - 1.4\, x^2(i-1) + 0.3\, x(i-2) + \sigma\eta(i), \qquad (8)$$

where $\eta(i)$ is an IID sequence chosen from a flat probability distribution, centered about zero with a standard deviation equal to that of the Henon map without noise. σ is a numerical factor which scales the strength of the noise, thus $\sigma=1$ means a signal to noise ratio of 1. In Figs. 2-4 we plot Φ, ξ_2, and ρ_2 for various strengths of the noise as a function of a. Although there is degradation in these indicators with increasing noise, we see very clearly, a maximum in Φ and minima in ξ_2, and ρ_2 even for σ on the order of 1/2.

Finally, we comment that we have not yet discussed any of the statistical properties of the indicators presented in this section. An appropriate statistical interpretation for these quantities depends on the context in which they are used. In particular, one will often wish to compare the values of the indicators computed on a data set, with values of the same quantities, but associated with a data set satisfying a null hypothesis. These issues will be discussed in context in sections III and IV below, and in Appendix B.

III. Model Misspecification

In this section, we demonstrate the sensitivity of the indicators described above to misspecifications of a model. (As before, we have in mind deterministic models or those with additive noise. However, we believe that these techniques can readily be adapted to the study of more general models.) In the following example we make no attempt to deduce the correct model (indeed, in a search space of lags of the time series no simple map describes this process), but rather show that some of the statistics of section II correctly identify a linear fit to the time series as a misspecification, even when other techniques fail to do so. As we shall see, this example is helpful in its clarity, but is unfortunate in that the candidate model is a linear function. Strictly speaking, we are testing for linear misspecification in this example, for which there are other existing methods which may be useful. However, it should be clear that our approach applies equally well to nonlinear candidate models, and is therefore more generally applicable than methods designed to test only for linear misspecification.

With these comments in mind let us consider a time series generated by summing together four independent Henon maps in the chaotic region. This example was discussed in a nice paper by Theiler and Eubank[5], and we shall compare with their results in what follows.

We consider four independent samples of the chaotic Henon map generated by

$$x_j(i) = 1 - 1.4x_j(i-1)^2 + 0.3x_j(i-2)\ ;\ \ j=1,2,3,4. \qquad (9)$$

The time series to be studied is given by

$$y(i) = \sum_{j=1}^{4} x_j(i). \qquad (10)$$

As pointed out in Ref. 5, if one attempts to model this series (with 512 data points) with a linear auto regressive fit, the best fit, f(i), according to several criteria involves six terms.

The series of residuals of this fit are, according to several tests, statistically indistinguishable from white noise. The residuals also pass a modified version of the BDS statistic as IID. Theiler and Eubank use these results to argue that "bleaching" chaotic data, i.e., studying the residuals of a chaotic series after subtracting a model fit can be misleading. In this case, even a test as sensitive as BDS fails to identify the residual series as non-IID.

In the tables below, we report the results of applying the statistics discussed in the previous section to the analysis of this series. In each of these tables, we report the value of a statistic applied to the data series, as well as the mean, $\hat{S}$, and standard deviation, σ, of that statistic when applied to an ensemble of bootstrapped data sets[6] which are constructed from the best linear AR fit. The kth bootstrapped time series is given by

$$z_k(i) = \tilde{f}(i) + [y(i)-f(i)]_R \qquad (11)$$

Here $\tilde{f}(i)$ is the same functional form as f, the best linear AR fit to the original data, but the arguments here are lags of the bootstrapped data set. The subscript R on the last term indicates that the elements in the square brackets have been randomized to produce an IID series with the same probability distribution as the residuals of the original data set. Thus, the bootstrapped data set, z, are correctly described by the linear AR model with IID residuals. This process is repeated 1000 times, and the mean and standard deviation of the statistic on the collection of bootstrapped data sets is computed. We then compute the significance, Σ

$$\Sigma = \frac{S - \hat{S}}{\sigma}. \qquad (12)$$

This is a measure of the statistical significance of the statistic on the original data set. We look for values of Σ greater than about three as a rough indication of statistical significance. That is, if Σ is greater than about three, we conclude it is unlikely that the original data set could be produced by a linear auto regressive process.[7] As a further measure of statistical significance, we report in the last row of these tables, ν, the number of bootstrapped realizations out of 1000 which have a value of S exceeding the value of S for the original data set. In Tables I-III the statistics are calculated with a value of μ=0.4.

As a first attempt to uncover model misspecification, let us look for structure in the sequence of residuals. In Table I we apply the Savit-Green statistics to the series of residuals. These statistics are related to the BDS statistics, but are differentially indicative of dependence on specific variables. In this table, δ_j is a measure of the additional dependence of an element of the residual time series on the jth lag of the residual time series, given that the information in the intervening j-1 lags has been used.[8] Looking at the first two columns of this table, we see that the significances for this dependence on the first lag, and on the second lag, given that we have used the information in the first lag are not large, and so, consistent with the general conclusions of Ref. 5, the residuals pass this test as IID. In the third column of the table we also calculate δ_1',the dependence of the residual series on the second lag *without* including the dependence on the first lag. Again, we see no statistically significant dependence.

In Table II, we compute the ρ_k of equation (5), which measures the dependence of the residuals on lags of the original time series. Here we see very significant dependence on the first lag. This is a clear indication that there is significant remaining structure in the residuals and that the linear auto regressive fit is not, therefore, a good explanation of the original time series.

To make the point still clearer, we compute, in Table III, the functions Φ, (eq. (2)) and ξ_k (eq. (4)). The first column in this table compares Φ for the original and bootstrapped data sets. We see that Φ is roughly the same for the original and bootstrapped data sets. That this should be so is not obvious, but neither is it completely unexpected. If the size of the residuals are fairly small, and if ε is not too small, then we might expect that Φ, which is a global measure of the predictability of the linear fit, could be roughly the same for the original and bootstrapped data. Turn now to columns 2-4. Here we compute ξ_k with k=1,2 and 3 for the original data set and its bootstraps. Recall that the ξ_k are differentially indicative of missing dependence in the model function. The significance in column 2 shows that there is additional structure in the original data set which is not captured by the linear auto regressive fit. Thus, we conclude again that there is significant nonlinear structure in this data set that is not captured by the linear auto regressive model. Moreover, the significances in these columns suggest that a description of this nonlinearity could involve, primarily additional dependence on the first lag of the time series. Note that this is consistent with the results of Table II in which we saw that the residuals were primarily dependent on the first lag of the original time series.

There are also other methods that one can use to test for model misspecification that do not require a strict analysis of residuals. For example, one can consider applying the δ's or the

BDS statistic, to the original time series, and then constructing a set of bootstrapped data to the time series as in equation (11) above. One could then compare values of these statistics on the original series and the set of bootstrapped data. Although neither the BDS nor Savit-Green statistics were originally designed to be used in this way, a statistically significant difference between their values on the original and bootstrapped data sets would indicate that the proposed model was inadequate. It is interesting to explore the relative power of this approach and the methods proposed here. As an example, we have computed δ_1 with $\mu=0.4$ on the original time series given in equation (10), and compared it to the value computed on a set of bootstrapped data constructed according to (11). We find that the significance of the statistic in this case is only 1.6. This should be compared with Table III in which we see that the significance of ξ_1 for the same value of μ and similarly constructed bootstrapped data is 3.8. Thus, for this parameter setting ξ_1 is able to distinguish a poor fit to the series (7), while this use of δ_1 is not. We have not made a systematic comparison of these two methods, but it would certainly be interesting to do so.

At this point we digress briefly from our main discussion to describe an interesting observation about this time series. The reader who wishes to focus only on the main points in the paper may skip to Section IV.

In Table IV we present the values of the quantity

$$\xi' = 1 - \frac{P(t_0(\mu) | F(\mu), t_1(\mu'))}{P(t_0(\mu) | F(\mu), t_1(\mu'), t_2(\mu))} \qquad (13)$$

where we have explicitly indicated the fact that t_0, F and t_2 are computed with a different value of μ than is t_1. This is an obvious generalization of expression (4), which is sensitive to dependence on the second lag of the time series given the dependence contained in the first lag and in the model function. (For more details about such indicators see Ref. 4.)

Now, by increasing μ' with fixed μ, we can study the additional nonlinear dependence on the second lag in the series, as we lose the information about the nonlinear dependence in the first lag. (Remember that in all of this, the linear auto regressive model is used to "filter" the linear dependence.) As Table IV shows, the dependence on the second lag decreases if we ignore the dependence on the first lag. This is somewhat counter-intuitive: In a more typical case (such as that arising from a chaotic map), one might expect the remaining information in the second lag to increase if we ignore the first lag. One kind of structure that could give rise to the behavior seen in Table IV is one in which the value of the first lag determines a sector or branch of the process, such that there is clear deterministic dependence on the second lag within each branch.

TABLE I

	δ_1	δ_2	δ_1'
statistic (S)	0.0215	0.0229	0.0020
$\bar{S}$ (bootstrapped data)	-0.0014	-0.0022	-0.0026
standard deviation	0.0145	0.0233	0.0145
significance	1.57	1.08	0.317
ν -value	62/1000	138/1000	350/1000

The Savit-Green statistics applied to the series of residuals. $\mu=0.4$. ν-value is defined in the text.

TABLE II

	ρ_1	ρ_2	ρ_3
statistic (S)	0.0577	-0.020	-0.0036
$\bar{S}$ (bootstrapped data)	-0.0025	-0.0015	-0.0020
standard deviation	0.0146	0.0154	0.0151
significance	4.12	-1.183	-2.104
ν -value	0/1000	108/1000	465/1000

Dependence of the residuals on lags of the original time series. $\mu=0.4$. ν-value is defined in the text.

TABLE III

	Φ	ξ_1	ξ_2	ξ_3
statistic (S)	0.2179	0.0666	-0.0027	-0.0018
$\bar{S}$(bootstrapped data)	0.220	-0.0010	-0.0022	-0.0024
standard deviation	0.0279	0.0178	0.0206	0.0189
significance	-0.078	3.80	-0.0227	0.0293
ν -value	472/1000	1/1000	493/1000	469/1000

Additional dependence of the original time series on lags thereof, given the linear auto regressive fit as a functional filter. $\mu=0.4$. ν-value is defined in the text.

TABLE IV

μ'	ξ'	significance, Σ
0.1	0.189	7.1
0.2	0.139	7.7
0.4	0.098	7.5
0.8	0.050	4.5
1.6	0.036	4.4
3.2	0.035	4.6

Values of the expression (13) for different values of μ'. Here $\mu=0.1$ and a time series of 8,192 points is used.

IV. Model Building

In addition to their use in model misspecification tests, the quantities described in section II can also be used to more efficiently generate a model for the data. To begin the discussion, it is important to distinguish two cases: one in which the search space is complete and the other in which the search space is incomplete. By a complete search space, we mean that the system can, in principle, be completely described by functions contained in the space. In a complete search space, a number of procedures, including ones based on the quantities in section II, will converge to the correct model. An example of model construction in a complete search space is contained in appendix B.

The case of an incomplete search space is considerably more interesting. Unlike the case of the complete search space, our methods lead to a different, and for many purposes superior description of the system than do other procedures, including those designed to minimize residuals such as a least squares approach. We shall show, in particular, that, when restricted to the same search space, our method leads to better medium-term predictions than does a least squares fitting procedure or even a simple Markov approach, as explained below.

The philosophies behind a straight forward data fitting approach and one based on the use of the methods of section II are quite different. Residual minimizing procedures generate models which track the data as closely as possible (in one sense or another), without regard to the structural differences between the data and the model. The approach we shall describe here is based on the construction of a set of pointers associated with the original data set. These pointers are determined by optimizing the conditional probabilities described in section II. Consequently, the relationship between the set of pointers and the data is dictated by an information criterion, rather than a criterion based solely on the size of residuals. As we shall argue, our method makes efficient informational use of the search space.

In some sense, our approach to modeling is related to a Markov process and may be thought of as an optimized generalization thereof. To specifically demonstrate one sense in which our procedure improves upon a Markov approach, we shall show that using our methods as a basis for a simple prediction algorithm gives superior results over the same prediction algorithm used in conjunction with the Markov description of the system. Moreover, as we shall argue, the use of a model function as a variable in a phase space reduces the dimensionality of the reconstruction, resulting in a more efficient statistical use of the data.

To illustrate our points, it is easiest to consider an example. To this end we consider the Lorenz system generated by the equations

$$\begin{aligned} \frac{dx}{dt} &= 10\,(y - x) \\ \frac{dy}{dt} &= 28x - y - xz \\ \frac{dz}{dt} &= xy - 1.6z \end{aligned} \tag{14}$$

The Lorenz system is quite simple which is both an advantage and a disadvantage. Pedagogically it is advantageous to have a simple system to study. On the other hand, many methods will do fairly well at describing and predicting such a simple system, so that the marginal advantages of our method are not as pronounced. Nevertheless, the advantages are significant, and should be even more marked on higher dimensional systems.

To construct our time series, we compute the integral time scale for x(t), and sample that series at intervals of 0.1 times the integral time scale to produce a discrete time series, x'(n). This series is then normalized by its standard deviation to produce a time series x(n) whose terms roughly lie in the range (-3,3). Our search space will be defined by functions