APPROPRIATE ALGORITHMS FOR NONLINEAR TIME SERIES ANALYSIS IN PSYCHOLOGY

CHRISTIAN SCHEIER
AILab, University of Zurich-Irchel
Winterthurerstrasse 190
8057 Zurich
E-mail: scheier@ifi.unizh.ch

WOLFGANG TSCHACHER
University Psychiatric Services Bern: Mid and West Sectors
Laupenstrasse 29
3010 Bern
E-mail: tschacher@spk.unibe.ch

Chaos theory has a strong appeal for psychology because it allows for the investigation of the dynamics and nonlinearity of psychological systems. Consequently, chaos-theoretic concepts and methods have recently gained increasing attention among psychologists and positive claims for chaos have been published in nearly every field of psychology. Less attention, however, has been paid to the appropriateness of chaos-theoretic algorithms for psychological time series. An appropriate algorithm can deal with short, noisy data sets and yields 'objective' results. In the present paper it is argued that most of the classical nonlinear techniques don't satisfy these constraints and thus are not appropriate for psychological data. A methodological approach is introduced that is based on nonlinear forecasting and the method of surrogate data. In artificial data sets and empirical time series we can show that this methodology reliably assesses nonlinearity and chaos in time series even if they are short and contaminated by noise.

1 Introduction

Chaos theory has a strong appeal for psychology for several reasons. First, it provides psychologists with a new vocabulary to conceptualise the complex and often seemingly random behavior of psychological systems. Psychotic episodes, for instance, can now be understood as manifestations of a chaotic system [1]. Second, the mathematical tools of chaos theory allow the formalisation and simulation of psychological theories and hence render the latter more transparent and stringent. Third, nonlinear time series analysis - to a large extent developed by the dynamical systems community - makes it possible to empirically investigate the nonlinearities observed in psychological phenomena. It is therefore not surprising that chaos-theoretic concepts and methods have recently received growing interest among psychologists [2]. In particular, nonlinear modeling techniques and algorithms have been widely used and claims of positive evidence for chaos in psychological data have been published in nearly every field of psychology[1,3,4,5].

However, less attention has been paid to the question of the appropriateness of these algorithms for psychological time series. One rather obvious reason to raise this question is that most of the classical nonlinear techniques have been developed in the physical sciences and hence make strong assumptions concerning the quality and length of the data[6]. More specifically, they often require thousands of data points with high resolution and a small amount of noise. Psychological time series, however, are typically contaminated by a high percentage of noise, are not very fine grained and rather short in length. An other problem with these algorithms is that it is often not possible to determine objectively the significance of the results which obviously renders the interpretation of the latter difficult if not impossible. If we look more closely to the most popular nonlinear method - dimension analysis[7] - even more problems arise. Dimension algorithms seem, for example, to indicate a finite dimension, even when presented with random noise[8]. Another bias is known as the 'edge effect'[9] due to finite data samples in high dimensions. This effect can make even white noise appear to have a finite dimension. Moreover, dimension estimates are sensitive to the amount of noise in the system[10] and require a large amount of data in order to yield reliable results[11]. Hence, it can be concluded that dimension analysis and other classical algorithms of dynamical systems theory such as Wolf's algorithm for calculating the Lyapunov-Exponent[12] are not appropriate for psychological data. Similar conclusions have been drawn in fields such as economics[13] or biology[14,15] that have to deal with related problems, namely the nonlinear analysis of small, noisy data sets.

Consequently, a lot of research has recently been devoted to the development of alternative approaches to nonlinear time series analysis which avoid the mentioned pitfalls of the traditional methods[16]. In this paper, we try to abstract from this very large literature those algorithms that we consider to be most pertinent for the nonlinear analysis of psychological data in that they can deal with short and noisy time series. Moreover, most of these algorithms have been developed outside the physical sciences and have been shown to yield reliable results when applied to short and noisy data. Even more importantly, our emphasis is on using several of these appropriate algorithms since when only one method is used to analyse a time series, the results are expected to be at best incomplete, and at worst wrong and misleading[17]. None of these algorithms have - to our knowledge - been applied to psychological data. Our overall goal is to develop a methodology for the nonlinear analysis of psychological time series which is appropriate in the sense that it satisfies the constraints imposed by psychological systems in general (e.g. very little is known about the dynamical system that generated the data) and psychological data in particular (e.g. short and noisy time series). In essence, two important new methodologies are introduced and applied to artificial and empirical time series: nonlinear forecasting and the method of surrogate data. The nonlinear forecasting of time series data has a relatively recent history [18,19]. However, it constitutes one of the most promising fields of research in nonlinear time series analysis and has recently gained a lot of attention [16]. The even more recently proposed method of surrogate data [20,21] serves the purpose of seeing whether certain classes of models for the data can be rejected. This involves generating many surrogate data sets satisfying the null hypothesis, and computing a statistic (such as the forecast error) for each. It has been shown that this technique is very successful in distinguishing chaos from various random processes even in short and noisy data sets [14]. Moreover, since it is inherently a statistical approach the surrogate data method allows the quantification of the results in a more rigorous manner than was possible with the traditional techniques.

The organisation of this paper is as follows. In Section 2 the nonlinear forecasting methodology and the surrogate data method are introduced. More specifically, a nonparametric simplex algorithm [15], a parametric forecasting algorithm [17], and two surrogate data methods [20,21] are presented. In Section 3 these methods are applied to artificial test data and empirical psychological time series. In Section 4 the results are discussed, the main problems are pointed out and possible further improvements of the proposed methodology are outlined.

2 Methods

2.1 Nonlinear forecasting

One essential property of chaos is its determinism, i.e. chaotic systems obey certain rules. Trajectories of chaotic systems can be predicted for short time scales. However, chaos amplifies noise exponentially and as a result this short-term determinism becomes long-term randomness. The idea behind the forecasting algorithms presented in this section is to use this characteristic to detect chaos in time series. In order to do this, a number of steps have to be performed. The first task is to reconstruct the system phase space from the time series, typically measurements of a single scalar observable at a fixed spatial point. This is of course standard practice in nearly every nonlinear algorithm and we therefore only present the main ideas [22]. From the orginal time series $x(t)$ an embedding space or state space representation of dimension

E is constructed whose points are

$$X(t) = (x(t), x(t-\tau), ..., x(t-(E-1)\tau) \tag{1}$$

$X(t)$ is called a *delay vector* with *embedding dimension* E and *delay time* τ. Takens [23] proved that if $E > 2d$, where d denotes the number of degrees of freedom of the dynamical system, then there is a smooth map $f^T : \Re^E \rightarrow \Re$ such that

$$X(t+T) = f^T(X(t)) \tag{2}$$

where the current state is $X(t)$ and $X(t+T)$ is a future state. The problem in nonlinear forecasting is to estimate $X(t+T)$. Since chaotic dynamics does not occur unless f is nonlinear, one has to build nonlinear models to approximate chaotic dynamics. Thus, the task is to find a good representation or approximation for the unknown function f. To avoid additional complexity, we now assume that our time series is embedded in an appropriate state space, and we have determined the embedding parameters E and τ. The problem now is to reconstruct the deterministic rule underlying the data. In the following, we discuss two nonlinear methods of local forecasting: first order and second order approximation.

The simplest (and earliest) forecasting procedure was suggested by Lorenz [24]. Suppose we want to predict the value of $X(k+1)$ knowing a long time series $X(j)$ for $j \leq k$. In first-order approximation [19] - sometimes also called 'method of analogs' - one finds the nearest neighbour to the current value of $X(k)$, say, $X(v)$, and then assumes that $X(v+1)$ is the predicted value for $X(k+1)$. Obviously, this is not really a representation or much of a model and the quality of this prediction can be improved in several ways. One possibility is to take a collection of near neighbours of the point $X(k)$ and take the averaged value of their images as the prediction.

Sugihara and May [15] have introduced a nonparametric simplex method (hereafter abbreviated as SM) that does exactly this. First, the time series is splitted in two parts. The first part is used as a library to predict the points on the second part. Hence, short-term predictions are made that are based on a library of past patterns in a time series as in the example above. Specifically, for each -dimensional point for which one wishes to make a prediction - for each 'predictee' -, all nearby points in the state space are selected. Of these points the simplex with the minimum diameter formed from $E+1$ nearest neighbours is constructed. To estimate the prediction, the evolution of the simplex after T time steps is calculated, weighting the actual distances by original distances from the respective, relevant neighbours. Finally, the correlation between the predicted and the raw data is computed. A decrease in this correlation with increasing prediction time steps suggests chaos whereas a non-decreasing prediction accuracy points to a random or some linear process (see Section 3).

The SM has been applied to biological[25,15], economical[26] and psychopathological [27] time series. In Section 3 it will be shown that the SM is a useful method to distinguish various random and linear processes from chaotic time series even in short and noisy data sets. Let us now turn to a further improvement of the first-order approximation scheme. The main idea is to fit an affine model to approximate the unknown function f of equation (2). This is called *second-order approximation*. The implementation of the 'method of analogs' in second-order approximation can be done in different ways. Polynomials are a good representation because their parameters can be linearly fit to minimise least square deviations (see below). Radial basis functions [17] or neural networks [26] provide other alternatives. One very interesting parametric forecasting method was recently proposed by Casdagli [17]. In this approach, nonlinear models are constructed with a variable smoothing parameter which at one extreme defines a nonlinear deterministic model, and at the other extreme defines a linear stochastic model. Specifically, piecewise-linear approximations to the function f of equation (2) are constructed by using a variable number of k neighbours. In essence, a small value of k corresponds to a deterministic approach to modelling as used in the SM (see above) whereas the largest value of k is equivalent to fitting linear stochastic autoregressive models. Finally, intermediate values of k correspond to fitting nonlinear stochastic models as, for example, proposed by Tong [28]. Since with this approach one compares deterministic with stochastic models, Casdagli calls his method deterministic versus stochastic modelling (DVS).

The DVS algorithm is implemented as follows [17]. First, as in the SM, the time series is divided in two parts: a *fitting set* containing the first N_f data points of the time series, and a *testing set* containing $N_t = N - N_f$ data points, where N is the length of the time series.Again, one chooses an embedding dimension E, a delay time τ, a forecasting step T and a predictee $X(i)$ with $i \geq N_f$ for a T-step-ahead forecasting test. Next, the distances of the predictee from the delay vectors $X(j)$,$1 + (E-1)\tau \leq j \leq N_f - T$, in the fitting set are computed and the k, $2(E+1) \leq k \leq N_f - T - (E-1)\tau$, nearest neighbours of the predictee are found to fit an affine model of the following form:

$$x_{j(l)+T} \approx \alpha_0 + \sum_{n=1}^{E} \alpha_n x_{j(l)-(n-1)\tau}, l = 1, ..., k \tag{3}$$

where the parameters $\alpha_0, ..., \alpha_E$ are computed by ordinary least squares.

Finally, the fitted model (3) is used to estimate a T-step-ahead forecast for the predictee. This procedure is repeated for all data points in the test set. The prediction accuracy can then be computed by using correlation between the predicted and the actual values of the time series. If the correlation increases with the number of E-dimensional nearest neighbours, the model is a linear stochastic process. In contrast if a small number of nearest neighbours shows high predictive power, this points to a nonlinear deterministic process. The DVS algorithm has been successfully applied to short and noisy biological time series [17,29]. In Section 3 the DVS it will be applied to artificial test data and psychological time series.

2.2 Surrogate data

Although the SM and the DVS are superior to traditional approaches in that they can detect chaos in short and noisy data sets they have one major drawback: their results cannot be obtained statistically, i.e. it is still necessary to rely on 'face validity' to be able to conclude if a time series is chaotic or not. In order to overcome this problem, several approaches have been developed, the most promising one being the method of surrogate data (MSD), proposed by Theiler *et al.* [21]. A related algorithm called 'noise versus chaos' (hereafter abbreviated as NVC) was introduced by Kennel and Isabelle [20], the major difference being in the NVC's use of predictor errors from many points in phase space instead of the average prediction error as a distinguishing statistic (see below).

Both algorithms have been successfully applied to short, noisy data from biology [29,21], economics [26] and psychopathology [27]. In the rest of this section, the two methods will be presented. The procedure consists of three parts: First, some linear process is specified as a null hypothesis. Second, suitable random data - the surrogate data - are generated, normalised to the mean and the variance of the original data. Third, a discriminating statistic for the original and for each of the surrogate data sets is computed. The basic idea is that, if the value computed for the original time series is significantly different than the ensemble of values computed for the surrogate data, then the null hypothesis is rejected and nonlinearity is detected. In the following, these three issues will be addressed in more detail.

The *null hypothesis* typically specifies that certain characteristics of the original time series are preserved (e.g. mean and variance) but that there is no further structure in the data. The surrogate data set is then generated to mimic these preserved trivial properties of the raw data but otherwise be random. In this paper, the following null hypotheses were used:

1. *temporally uncorrelated noise*$(H_0^{(TUN)})$. The simplest null hypothesis

one can devise about a time series is that there is no evidence for any dynamics at all; that is, that the data are fully described by independent and identically distributed (IID)random variables. To generate the surrogate data, two algorithms can be used. First, if the distribution is assumed to be gaussian, then the surrogate data can be generated from a standard pseudorandom generator. Second, to test the hypothesis of uncorrelated noise with arbitrary amplitude distribution, surrogate data can be generated by shuffling the time-order of the raw data. We will refer to these procedures as *Sur(rand)* and *Sur(shuffle)*, respectively.

2. *linearly autocorrelated gaussian noise*($H_0^{(LAGN)}$).The null hypothesis in this case is that all the structure in the time series is given by the Fourier power spectrum or, equivalently, by the autocorrelation function (ACF). The simplest case is that given by the Ornstein-Uhlenbeck process, which can be produced by

$$x(t) = \phi_0 + \phi_1 x(t-1) + \theta\epsilon(t) \tag{4}$$

were $\epsilon(t)$ is uncorrelated gaussian noise of unit variance. The coefficients ϕ_0, ϕ_1 and θ determine mean, variance, and autocorrelation time of the time series, respectively. The more general case can be implemented by fitting an AR(K)MA(L)-model of the form

$$x(t) = \sum_{k=1}^{K} \phi_k x(t-k) + \sum_{l=1}^{L} \theta_l \epsilon(t-l) \tag{5}$$

where the coefficients $\{\phi_k\}$ and $\{\theta_l\}$ are to be determined by fits to the data, typically using a least squares or an information theoretic criterion, and the $\epsilon(t)$ are some stochastic forcing terms which are specified by the modeler. Surrogate data can then be generated by iterating (4) or (5), where the coefficients have been fit to the original data. We will refer to these procedures as *Sur(Uhl)* and *Sur(ARMA)*, respectively.

An alternative algorithm, based on the phase randomisation of the fast fourier transform, is described in Theiler et. al. [21]. In this case, the surrogate data are generated in two steps:

1. taking the discrete fourier transform

$$\Xi(k) = \sum_{t=0}^{N-1} x(t) e^{2\pi i n k/N} \tag{6}$$

and

2. generating surrogate data sets $s = 1, ..., S$

$$\Xi^s(k) = \Xi(k)[\zeta(k) + i\eta(k)] \tag{7}$$

where ζ and η are independent, real Gaussian random numbers with mean 0 and variance 1/2 and $\eta(k) = -\eta(N-k)$ ensuring $x^s(k)$ to be real. We will refer to this algorithm as $Sur(Fourier)$. There are, of course, many other null hypotheses that can be tested for. For instance, one could test for noisy periodicity as proposed by Kennel and Isabele [20] or the null hypothesis of a static nonlinear filter of linearly correlated noise as used by Theiler *et al.* [21]. In the following, however, we restrict ourselves to the two null hypotheses described above.

Discriminating statistic and technique of comparison. Although, in principle, any discriminating statistic can be used we chose to use a phase-space predictor to obtain the results presented in Section 3. More specifically, the correlation between predicted and raw data was used. Using a predictor has the advantage that it directly investigates one of the essential differences between chaos and randomness, namely determinism. Moreover, it is algorithmic simple and free of any parameters. Theiler et al. also consider correlation dimension as a statistic, but - as mentioned in the introductory section - dimension analysis has several drawbacks in respect to the analysis of short and noisy data. Similarly, as pointed out by Daemmig [30], numerical estimation of even the largest Lyapunov exponent can be problematic in the presence of noise.

Once one has computed the value of the discriminating statistic for the original and the surrogate data, the significance of the difference in the MSD is defined as follows. In our case the statistic of interest is the correlation coefficient ρ. In this case the resulting mean correlation ρ_H is compared to ρ_{raw} and the result is normalised by the standard deviation of the surrogates. Thus the significance with which one can reject the null hypothesis is

$$S \equiv \frac{\rho_H - \rho_{raw}}{\sigma_H} \tag{8}$$

A significant rejection of the null hypothesis occurs when $S \geq 2$ [21]. In the NVC, the set of prediction errors - which are decimated by the factor of the prediction error autocorrelation - on the real data set (A), and the ensemble of prediction errors on surrogate data sets (B) are compared using the Mann-Whitney rank-sum statistic

$$U = \sum_{i=1}^{card(a)} \sum_{j=1}^{card(B)} \Theta(A_i - B_j) \tag{9}$$

Type of process	SM $\langle\Omega\rangle$	DVS $\langle p_{nl}\rangle$ $\langle p_l\rangle$	MSD $\Psi(H_0^{TUN})$ $\Psi(H_0^{LAGN})$	NVC $\langle z\rangle$
non-linear deterministic	$\gg 0$	$\gg$	>2 >2	<-3.72
linear stochastic e.g. AR(M)	≈ 0	$\ll$	>2 <2	>-3.72
MA(L)	$\gg 0$	$\ll$	>2 <2	>-3.72
white noise	≈ 0	≈ 0	<2	$\gg -3.72$

Figure 1: A framework for the nonlinear analysis of psychological time series. $\langle\Omega\rangle$: mean forecasting accuracy one day ahead *minus* mean forecasting five days ahead;$\langle\rho_{nl}\rangle$,$\langle\rho_l\rangle$: mean forecasting accuracy in the nonlinear deterministic end (nl) and the linear stochastic end (l); $\Psi(H_0^{TUN})$, $\Psi(H_0^{LAGN})$: mean MSD significance S for the two null-hypotheses outlined in the text. $\langle z\rangle$: mean z-value. All means are computed using $E = 1...10$; for SM, DVS, and MSD $\tau = 1$, for NVC $\tau = 1...10$

where $\Theta(\cdot)$ is the Heaviside step function: $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x \leq 0$. For large sets a standard normal distributed Z-statistic is obtained

$$z = \frac{U - card(A)card(B)/2}{\sqrt{\frac{1}{12}card(A)card(B)(card(A) + card(B) + 1)}} \tag{10}$$

differentiating whether the distribution of the prediction errors of A or B are significantly smaller or larger. With K repetitions for different time delays τ the limits on z

$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi e^{-z^2}dz}} = \alpha/K \tag{11}$$

are more trenchant and in our case with $K = 10$ at -3.72 for a one-variable confidence limit$\alpha = 0.01$.We have now reached a point where we can integrate these algorithms into a coherent framework, depicted in Figure 1. In essence, it allows us to distinguish empirically between nonlinear (possibly chaotic) deterministic systems, linear systems such as an ARMA-process and white noise.

Obviously, this is a rather general differentiation. However, it is easily possible to add new features in this scheme. One can, for instance, integrate more elaborated hypotheses such as a static nonlinear filter of linearly correlated noise into the MSD or the NVC. An other alternative would be to use a nonlinear stochastic predictor[28] as a discriminating statistic. In any case we believe that the reliable distinction between the three classes of processes listed in Figure 1 constitutes an important step towards our overall goal, namely the development of a coherent and appropriate methodology for the nonlinear analysis of psychological systems. In Section 4 we will show how this framework can be combined with other time series analysis tools. First, however, let us turn to its application to artificial and empirical time series.

3 Results

3.1 *Artificial time series*

To test the algorithms, two experiments were conducted. The first experiment served to answer the question whether they can differentiate between white noise, linear and nonlinear (in our case chaotic) processes consisting of very few data points. In the second experiment robustness with respect to noise level was tested. Remember that we analyse the following methods: the simplex method (SM), the 'deterministic versus stochastic'-algorithm (DVS), the method of surrogate data (MSD), and the 'noise versus chaos'-technique (NVC).

Experiment 1: Differentiation. Three categories of processes were generated: (1) random processes, (2) linear processes, and (3) chaotic processes. The main interest was to check whether each of the algorithms was able to differentiate between them. Note that with this length of the data none of the classical algorithms could be used. Figure 2a shows the results obtained by applying the simplex algorithm (SM) to these data. It can been seen that, indeed, the forecasting accuracy - quantified by using the correlation coefficient ρ - decreases with increasing forecasting time step T only for the chaotic time series. This signature of ρ decreasing with T does not arise when the erratic time series is in fact a noisy limit cycle or a first order autoregressive process. Not unexpectedly, no significant forecasting accuracy can be obtained for the IID process. Hence, the SM can discriminate chaotic from linear and random processes. Similar conclusions can be drawn for the DVS algorithm (Fig. 2b). It correctly classifies the chaotic time series in the deterministic end (high correlation with small k values) and the other processes in the stochastic end (high correlation with high k values). However, these conclusions rely on

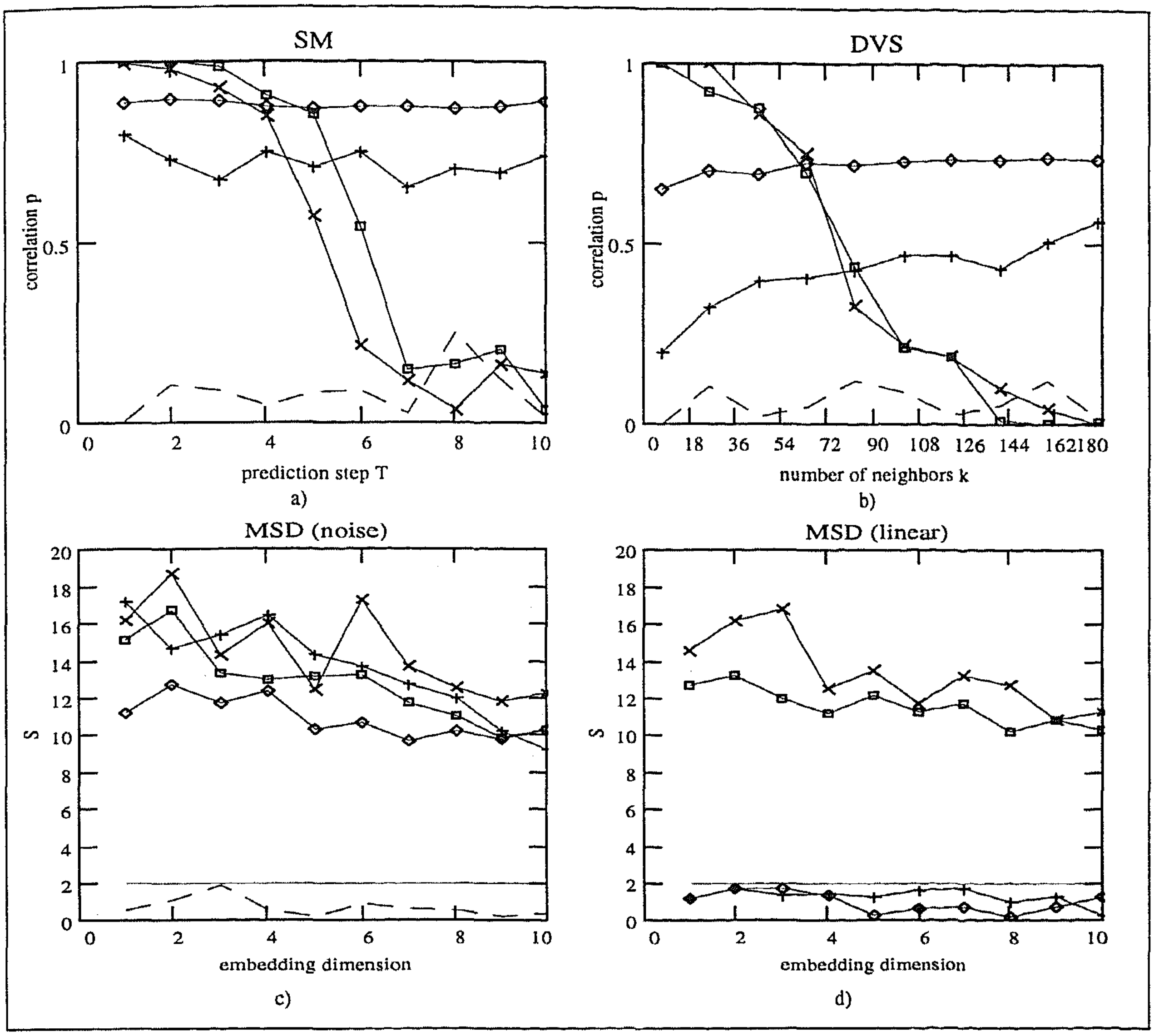

Figure 2: Differentiating chaotic, linear and random processes with the algorithms presented in the text. a) SM, b) DVS, c) and d) MSD. The following processes were used: *crosses*: Henon-Map in the chaotic regime ($x(t+1) = y(t) + 1 - 1.4x^2(t)$;$y(t) = 0.3x(t)$);*squares*: Logistic-map in the chaotic regime ($x(t+1) = 4x(t)(1-x(t))$); *diamonds*: Sine wave with 50% noise ($x(t) = sin(0.5t) + \epsilon(t), \epsilon \in [-0.5, 0.5]$); *plusses*: first order autoregressive process (AR(1), $x(t+1) = 0.4x(t) + \epsilon(t), \epsilon \in [-1, 1]$); *dashed*: random process generated with a standard pseudorandom number generator. For all calculations $E = 3, \tau = 1$

Type of process	SM	DVS		MSD		NVC
	$\langle\Omega\rangle$	$\langle p_{nl}\rangle$	$\langle p_l\rangle$	$\Psi(H_0^{TUN})$	$\Psi(H_0^{LAGN})$	$\langle z\rangle$
Hénon	0.87	0.21	0.99	17.23	15.89	-16.47
Logistic map	0.85	0.17	0.99	14.77	14.12	-13.38
AR(1)	0.08	0.78	0.34	15.61	1.21	- 0.01
Sinus	0.02	0.81	0.45	12.34	0.98	0.21
Random	0.04	0.12	0.09	0.89	---	2.34

Figure 3: Differentiating chaotic, linear and random processes with the algorithms presented in the text.

the inspection of the plots in Figs. 2a and 2b and hence on 'face validity'.

In order to get quantitative results, the MSD and the NVC algorithm were applied to the test data. The results for the MSD are depicted in Figs. 2c and 2d. First, the null hypothesis of temporally uncorrelated noise was tested and surrogate data were generated with the shuffling algorithm *Sur(shuffle)* described in section 2 (Fig. 2c). The null hypothesis can be rejected for all time series except the IID process. Second, surrogate data were generated using algorithm *Sur(Fourier)*, hence testing for the null hypothesis of linearly autocorrelated gaussian noise (Fig. 2d). In this case, the null hypotheses can be rejected only for the chaotic time series. Thus, the MSD correctly identifies the linear processes and discriminates them from the nonlinear (chaotic) ones. In Figure 3 these results are integrated into the framework presented in the last section. It is easy to verify that we can correctly and reliably distinguish the time series in question. This is an important result since we use very short data sets.

Experiment 2: Discretization In the second experiment, it was tested whether nonlinear determinism can be detected even when the continuous dynamical system which generated the time series is discretized. Discretization of continuous dynamics occurs frequently in psychology where rating-scales are often the only means to obtain relevant data. It is thus important to know if the algorithms presented here are appropriate for discretized dynamical systems. In a first step, the Henon-Map time series used in the first experiment was mapped onto the intervals $[1...I], I \in [5...9]$. Next, the algorithms were applied to these discretized time series. In figure 4 the results of this experi-

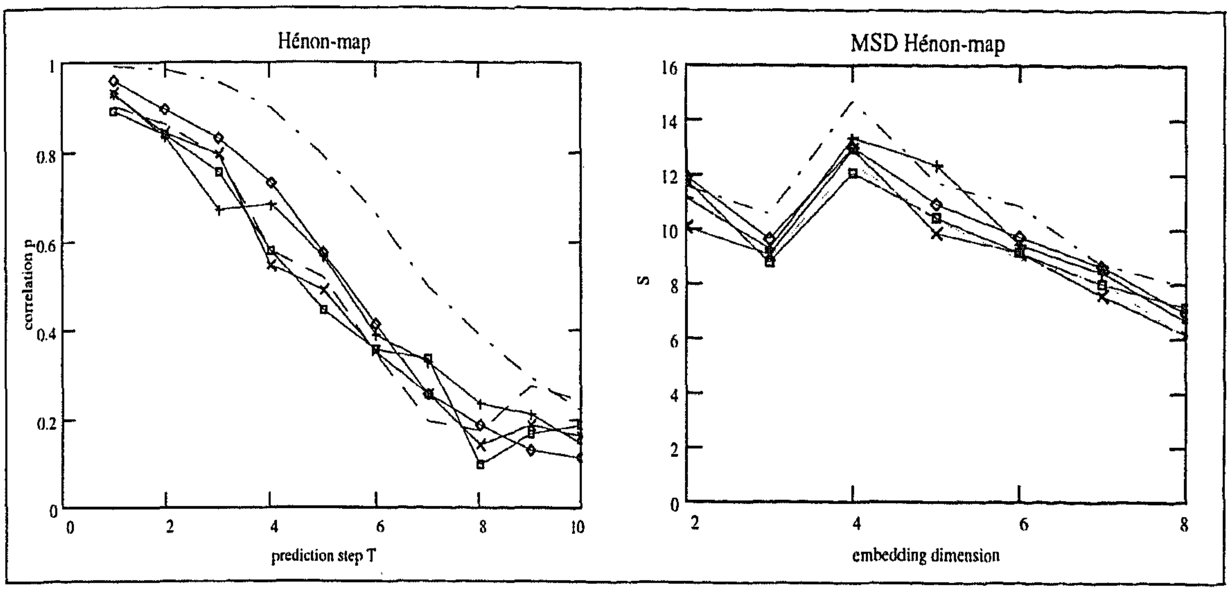

Figure 4: nonlinear forecasting (left) and surrogate statistic (right) for orignial and discretized Henon-Map data. *Dash-dotted*: origninal data; *crosses*: 5-point scale; *boxes*: 7-point scale; *diamonds*: 8-point scale; *dashed*: 9-point scale. $E = 3$,$N = 400$,$\tau = 1$. Explanations see text.

ment are shown for the SM and the MSD. Two conclusions can be drawn from this experiment: First, the results obtained for the discretized data are very similar to the ones obtained for the continuous data ($\rho \approx 0.97$). Second, less significant results are obtained for the discretized time series. This is an important outcome since we get more *conservative* results for the discretized data. Hence, results obtained by applying these algorithms to rating-scales data (like the ones presented below) are expected to yield reliable results which tend to be *less significant* compared to results from the underlying continuous system.

3.2 Empirical time series

We present results on clinical data. These results are meant to be illustrative. Our goal is to show that the framework developed in the last sections can indeed yield reliable and meaningful results when applied to psychological time series. We studied 14 patients treated at the 'Soteria Bern'. The Soteria as a small residential clinic specialised for persons experiencing first psychotic manifestations is based on ideas of milieu therapy and affect logic [1,31]. The prerequisite for inclusion in our sample was that the daily manifestations of psychotic symptomatology of a patient could be observed almost completely for a long enough period of time (at least 200 days). The longitudinal course was mapped by the daily rating of a patient's psychoticity by Soteria staff members. A seven-point scale was used as described in Aebi et. al. [31]. At

Patient (sex)	Model	SM	DVS		MSD		NVC
		$\langle\Omega\rangle$	$\langle p_{nl}\rangle$	$\langle p_l\rangle$	$\Psi(H_0^{TUN})$	$\Psi(H_0^{LAGN})$	$\langle z\rangle$
2 (m)	nonlin	0.578	0.53	0.87	9.26	6.66**	-8.32 **
1 (m)	nonlin	0.757	0.32	0.64	4.59	3.42**	-5.12 **
5 (f)	nonlin	0.358	0.43	0.71	3.91	11.34**	-7.12 **
3 (f)	nonlin	0.698	0.55	0.81	15.27	2.18*	-12.55 **
13 (f)	nonlin?	0.671	0.34	0.48	5.13	2.33*	-3.45
9 (m)	nonlin?	0.479	0.44	0.49	11.64	2.28*	-1.88
11 (f)	nonlin?	0.472	0.21	0.20	4.72	2.18*	-3.16
14 (m)	linear	0.790	0.65	0.21	12.22	0.98	-0.23
4 (f)	linear	0.696	0.54	0.11	17.13	1.23	0.47
10 (m)	linear	0.852	0.48	0.01	11.97	0.87	1.09
6 (f)	linear	0.920	0.74	0.32	10.84	1.90	-2.90
8 (f)	linear	0.661	0.44	0.18	15.27	1.72	0.66
12 (m)	noise	0.477	0.12	0.19	1.66	---	2.33
7 (f)	noise	0.174	0.21	0.01	0.80	---	0.77

Figure 5: Results for empirical time series. Explanations see text.

the focus of our interest is the course of psychotic derealisation measured with this scale. We thus restrict ourselves to statements about a specific systems level (namely the level of psychopathological time courses of 200 to 800 days); under these constraints we think it is possible to differentiate types of psychotic dynamics using the nonlinear time series analysis framework presented above.

The identification of different psychotic dynamics will eventually have direct clinical implications. For instance, stochastic systems whose time series do not show serial structure point to a high sensitivity for fluctuating environmental stimuli. These systems are suggested by behavioral theories which take behavior largely under the control of external stimuli. In our study they pose a fundamental null hypothesis since environmental influences on psychoticity are not controlled for in our field data. Nonlinear deterministic (possibly chaotic) dynamics on the other hand point to the existence of an internally controlled, low-dimensional system unfolding relatively autonomously from environmental fluctuations. Moreover, empirical evidence of such dynamics would be a validation of the dynamical disease concept for psychoses[32]. In the context of psychotherapy nonlinear deterministic systems seem compatible with psychoanalytical or system theories.

In figure 5 the results for these data are depicted. The figure shows that in 4 (28%) patients out of 14 all 4 algorithms suggest nonlinear dynamics since we find a rapid decay of the forecasting accuracy $\langle \Omega \rangle$, a higher predictive power for nonlinear models in the DVS, and a significant rejection of the null hypothesis of a linear stochastic process in both the MSD and the NVC. In another 3 patients we can tentatively conclude that their psychotic paths were generated by a nonlinear deterministic system. There is a significant rejection of the 'linear' null hypothesis in the MSD and a steep decay in the forecasting accuracy in the SM. Hence, we find in 7 (50%) out of 14 patients strong evidence for nonlinear determinism. Five time series are best modelled as an autoregressive linear process. Two cases are classified as random. The results of significance tests are summarised under the heading 'model' in Figure 5.

These results suggest that the framework presented in this paper is indeed useful to differentiate several important process models and to reliably assess nonlinear determinism in psychological time series. For further interpretations of these results, see Tschacher et. al. [27].

4 Discussion and conclusions

Our main goal in this paper was to introduce appropriate algorithms for nonlinear time series analysis in psychology. An appropriate algorithm can deal with short, noisy data sets and yields 'objective' results. We presented two new approaches - nonlinear forecasting and the method of surrogate data - and integrated them into a framework that allows for the reliable assessment of nonlinearity and determinism in psychological data. Application of this framework to artificial and empirical data showed that it can differentiate between three important classes of dynamical systems: (1) random walks, (2) linear stochastic systems, and (3) nonlinear deterministic systems. Moreover, the presented framework satisfies constraints imposed by psychological data since it is designed for the analysis of short and noisy time series. The approach outlined in this paper can be improved in several ways. First, it should be complemented with stationarity tests and linear methods. In a first step, one tests whether the time series in question is stationary and otherwise differentiates it. Then one tests for temporal structure in the data. If no structure is detected, the analysis is stopped since the data are white noise. Otherwise, the structure is examined in more detail. It can be linear or nonlinear. If linearity is detected, linear time series methodology is used to further analyse the data. Similarly, if nonlinearity is detected, the nonlinear algorithms presented in this paper can be used to further specify the detected nonlinearity. One can, for instance, test for chaos. It is here where the classical techniques can be helpful.

Another possible improvement of the current framework concerns the amount of variables that can be analysed. Most (if not all) nonlinear algorithms presently allow only univariate analysis of the data. In psychology, however, one often has to use several variables to empirically investigate a given phenomena. In psychotherapy research, for instance, it is obviously important to analyze the interaction between therapeut and client. It is thus necessary to develop multivariate algorithms. This is described in Scheier and Tschacher [33]

All of the presented methods use the deterministic model (2) to analyze the data. The assumption of a purely deterministic process may, however, not be adequate for psychological time series. It is not very likely that all fluctuations in a time series can be explained by chaotic dynamics. Thus, it would be useful to improve the model (2) with respect to dynamical noise. A very interesting first step in this direction has recently been proposed by Nychka et. al. [10].

Finally, it will be important to gain more knowledge about psychological dynamics. Psychological research has to a large extend been tied to cross-sectional studies and thus has little to say about psychological processes [34]. In the future a more elaborated set of process models will have to developed (e.g. different kinds on nonlinearities dependent on the psychological system). We hope to have stimulated such an endeavor with the present paper.

Acknowledgments

We are grateful for the laborious documentations done by Soteria team members. Data acquisition was made possible by the pilot project Soteria (grant of the Swiss National Fund) headed by Luc Ciompi.

References

1. L. Ciompi, System Familie **5**, 133–145 (1992).
2. W. Tschacher*et al.*, *Self-organization in clinical psychology* (Springer, Berlin, 1992).
3. M. Koukkou *et al.*, Biological Psychiatry **33**, 397–404 (1993).
4. R. M. Warner, Behavioral Science **37**, 128-138 (1992).
5. A. Gottschalk, Arch. Gen. Psychiatry **52**, 947–959 (1995).
6. D. Ruelle,*Chance and chaos* (Princeton University Press, New Yersey, 1991)
7. J. Theiler, J. O. Soc. Am. **4**, 1055-1073 (1990)
8. A. R. Osborne *et al.*, Physica D **35**, 357-381 (1989)
9. I. Dvorak and J. Klaschka, Physics Letters A **145**, 225–231 (1990)
10. D. Nychka *et al.*, Journal of the Royal Statistical Society B **54**, 399-426 (1992)
11. D. Ruelle, Proc. Royal Soc.London **427A**, 241–248 (1990)

12. A. Wolf *et al.*, Physica D **16**, 285-299 (1985)
13. C. G. Gilmore, Int. J. of Bifurcation and chaos **3**, 583–589 (1993)
14. F. Mitschke and M. Daemmig, Int. J. of Bifurcation and Chaos **3**, 693–703 (1993)
15. G. Sugihara and R. M. May, Nature **344**, 734-741 (1990)
16. A. Weigend and T. Gershenfield *Predicting the future and understanding the past* (Addison-Wesley,Redwood City,1994)
17. M. Casdagli, Journal of the Royal Statistical Society **54**, 303–328 (1992)
18. J. P. Crutchfield and B.S. McNamara, Complex Systems **1**, 417–452 (1987)
19. J. D. Farmer and J. J. Sidorowich, Physical Review Letters **59**, 845–848 (1987)
20. M. B. Kennel and S. Isabelle, Physical Review A **46**, 234-258 (1993)
21. J. Theiler*et al.* , in *Nonlinear modeling and forecasting*, ed. M. Casdagli and S. Eubank (Addison-Wesley, Redwood City, 1992).
22. , E. Tsonis *Chaos. From theory to applications*, (Plenum Press, New York, 1992).
23. F. Takens in *Dynamical Systems and Turbulence*, ed. D.A. Rand and L.S. Young (Springer, Berlin, 1981).
24. E. N. Lorenz, J. Atmos. Sci. **26**, 636 (1969)
25. M. Dumon *et al.*, in *Proceedings of the first European Conference on Artificial Life ECAL95*314–3241993
26. D. Meier *et al.*, in *Proceedings Proceedings of 1st Int. Conference on Neural Networks in Capital Markets*210–2171993
27. W. Tschacher *et al.*, Biological Psychiatry (, i (n) press).
28. H. Tong *Nonlinear time series: A dynamical systems approach*(Oxford University Press, Oxford, 1990).
29. A. Longtin, Int. J. of Bifurcation and Chaos **3**, 651–663 (1993)
30. M. Daemmig*et al.*, Applied Physics **855**, 121-125 (1992)
31. E. Aebi *et al.*, Zeitschrift für Klinische Psychologie, Psychopathologie und Psychotherapie **41**, 18-30 (1993).
32. L. Glass and M. Mackey *From clocks to chaos. The rhythms of life* (University Press, New York, 1988).
33. C. Scheier and W. Tschacher, *AILab Technical Report AI.96.3*1996.
34. C. Scheier and W. Tschacher System Familie **7**, 11-21 (1994)