Applicability of Dimension Analysis to Data in Psychology

Arno Steitz, Wolfgang Tschacher, Klaus Ackermann, and Dirk Revenstorf

With 16 Figures

Abstract. This paper is motivated by the question of whether dimension analysis is a valid and practical method for the reduction of data in psychology. The paper presents a short introduction to the analysis of chaotic systems by the Grassberger–Procaccia algorithm. General aspects of this method are demonstrated; we tested the limits of dimension analysis depending on signal-to-noise ratio, length of time series, and resolution of measurement. For this purpose, the Hénon map was used as a basic model. The Grassberger–Procaccia algorithm was also applied to a simulated time series of group processes and an empirical time series of smoking behavior. To compensate for artefacts induced by local correlations a revised dimension analysis was performed with the group simulation data. Results suggest that neither group simulation nor cigarette consumption data can be reduced to a low-dimensional deterministic system.

1. Introduction

The study of various physical and mathematical models has shown that even simple nonlinear systems display very complex behavior within certain ranges of their parameters. Time series of these systems may look like irregular random series but actually are totally determined by only few variables, a phenomenon referred to as deterministic chaos. Therefore, one may search for the fingerprints of simple systems in observational data of noise-like complexity. In the biological and social sciences analyses for chaotic behavior have been applied, for example, to epidemiology (Schaffer & Kot, 1986), chronobiology (an der Heiden, this volume), cardiac electrophysiology (Glass et al., 1986) and, with special relevance to psychology,to EEG activity (Babloyantz, 1985; Graf & Elbert, 1989).

We are basically interested in analyzing dynamical systems within clinical psychology. The most convenient and thus most popular tool to find such fingerprints of nonlinear systems is the evaluation of the correlation dimension by means of the Grassberger–Proccacia algorithm (Grassberger & Procaccia, 1983). We studied some aspects of its application analyzing time series from three systems. The first one is the well-known analytical Hénon map. The second is a time series from a computer model simulating group interaction. The third time series consists of about 1000 observations concerning a person's daily cigarette consumption.

Springer Series in Synergetics, Vol. 58 Self-Organization and Clinical Psychology
Editors: W. Tschacher, G. Schiepek, and E.J. Brunner

2. An Introduction to Dimension Analysis of Chaotic Systems

Since a number of papers are available that give a good overview of the dimension analysis of chaotic systems (Simm et al., 1987; Mayer–Kress, 1987), we start with a short introduction to dimension analysis with the Grassberger–Proccacia algorithm.

The set of mutually independent variables necessary to describe the behavior of a system span the system's **state space** (the variable set of a particular state of the system corresponds to one coordinate point in the state space). In general, we do not know the number of variables governing the system nor do we know the effective dimension of its state space. Instead, we measured the evolution in time of a system by one or few projections of the system's state space onto our measure quantities. Time series that show no spectral structure, i.e. no clear–cut periodic pattern, may result either from projections in noisy directions, or from projections of high dimensional state surfaces, or from low dimensional but geometrically complex attractors. Dimension analysis will discriminate the latter from the two former ones (i.e. discriminate systems which are governed by few, from those which are governed by a large number of variables).

From the study of dissipative chaotic model systems we know that their evolution in time — at least on chaotic time scales — is unpredictable due to the exponential divergence of initially close trajectories. Nevertheless, the possible states of the system do not fill up the whole volume of the state space but are confined to the so–called chaotic attractor — a generalization of the concept of state surfaces to which the evolution of a system is bound. By describing the geometry of the attractor, dimension analysis can be one way to recognize and to characterize a chaotic system.

2.1 The Generalized Dimensions

How can we describe an attractor (i.e. the ensemble of state points in the state space of a chaotic system) by its geometrical properties? To begin with, we assume sufficiently long time series of a chaotic system in all its relevant variables. The information of all time series can be represented as an ensemble of state points in a state space of dimension equal to the number of relevant variables. Each point represents a state of the system at a given point in time.

The probability distribution of the attractor points, i.e. the probability of finding one out of N attractor points in cell k under a covering Ω of the state space by $M(r)$ cells of radius r is

$$p_k(r) = \frac{N_k(r)}{N(r)} \tag{1}$$

where N_k is the number of attractor points in cell k. (Every point of the state space lies in only one cell of the covering Ω.)

This enters in the definition of the so-called q-dimensions which describe the 'strange' geometrical structure of a chaotic attractor. The q-dimensions can be seen as the moments of an expansion of the probability distribution of the attractor points:

$$D_q = \frac{1}{q-1} \lim_{r \to 0} \frac{\log(\sum_{k=1}^{M(r)} p_k^q)}{\log(r)} \quad . \tag{2}$$

The Hausdorff dimension D_0 can easily be seen to correspond to the standard dimension for n-dimensional surfaces. It can be shown that in general $D_{q+1} \leq D_q$ (Hentschel & Procaccia, 1983).

2.2 The Correlation Dimension and Its Approximation by the Grassberger-Proccacia Algorithm

For practical purposes we are not able to calculate all q-dimensions (especially for large q) because of the computational burden or lack of sufficiently long time series. For D_2 a practical algorithm to approximate the above definition in (2) has been suggested by Grassberger and Proccacia. From (2) the D_2 or correlation dimension for a dense covering Ω with $r \to 0$ is written

$$D_2 = \lim_{r \to 0} \frac{\log(\sum_{k=1}^{M(r)} p_k^2)}{\log(r)} \tag{3}$$

where p_k^2 is the probability that two arbitrary points P_i, P_j are in cell k.

$\sum_{k=1}^{M(r)} p_k^2$ is the probability to find these points in any cell of the covering. This is approximately the probability $C(r)$ to find two points P_i, P_j with distance less or equal to the diameters of the cells r

$$\sum_{k=1}^{M(r)} p_k^2 \cong C(r) = \lim_{N \to \infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j \neq i=1}^{N} \Theta(r - \mid \vec{P_i} - \vec{P_j} \mid) \quad . \tag{4}$$

where Θ is the Heaviside function $\qquad \Theta(x) = \begin{cases} 1 & \text{for} \quad 0 \leq x < 1 \\ 0 & \text{else} \end{cases} \quad .$

Finally, the correlation dimension ν is defined as

$$D_2 \cong \nu = \lim_{r \to 0} \frac{\log(C(r))}{\log(r)} \quad . \tag{5}$$

2.3 The Reconstruction of the Attractor from One-Dimensional Time Series

A prerequisite for the application of the Grassberger-Proccacia algorithm is that we know the attractor points in state space or at least are able to reconstruct the attractor in a way that preserves its interesting geometrical structure (Packard, 1980; Takens, 1981).

For a given dimension of the state space, state vectors of that dimension have to be constructed from the time series. Since these vectors should reveal the effective dimension of the attractor, which is smaller than the dimension of the embedding space, the components of the vector should reflect this in a correlation among them. Several procedures for the construction of the vectors have been proposed. We refer to the one called time delayed reconstruction. The time series of length n yields $(n - m\tau)$ m–dimensional vectors $(x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau})$.

The choice of the delay parameter τ is important. A too small delay with respect to the characteristic correlation length of the time series compresses the reconstructed attractor by inducing a high correlation among the components of the vector of the state point, thus yielding a dimension smaller than the real one. A too large delay decouples the components of the vector. The reconstruction then tends to fill up the embedding space resulting in a dimension of the reconstructed attractor corresponding to the number of components of the vector.

2.4 The Evaluation of the Correlation Dimension from $C(r)$

For a genuine embedding of the attractor (dimensions of the embedding space greater than the dimension of the attractor) the correlation dimension $\nu = \lim_{r\to 0} \frac{\log C(r)}{\log r}$ should be independent of the dimension of the embedding state space. In practice, $C(r)$ can only be calculated down to a small r_f because of the limited resolution of experimental data. Also, the finite length of the time series accounts for an incomplete and inhomogeneous reconstruction of the attractor and causes a noisy and unreliable evaluation of $C(r)$ for $r < r_f$. The correlation dimension will show up as a constant ratio of $\log(C(r))$ to $\log(r)$ across a range of embedding dimensions and a range of r. To evaluate $C(r)$ for a given r and a given m–dimensional reconstruction we compute the mean over all attractor points of the number of their neighbor points with distances less than r. Variations of $C(r)$ for different locations on the attractor are neglected.

In the case of reconstructions with large time delays, strongly correlated neighboring points of the time series should be given attention, since they will be neighbors in state space as well. Thus they induce spurious correlations and have to be carefully discarded in the evaluation of $C(r)$.

3. A Simple Analytical Model: the Hénon Map

We examined a 1500 point time series from the two–dimensional Hénon map given by the equations:

$$\begin{aligned} x_{n+1} &= y_n + 1 - 1.4x_n^2 \\ y_{n+1} &= 0.3x_n \ . \end{aligned} \tag{6}$$

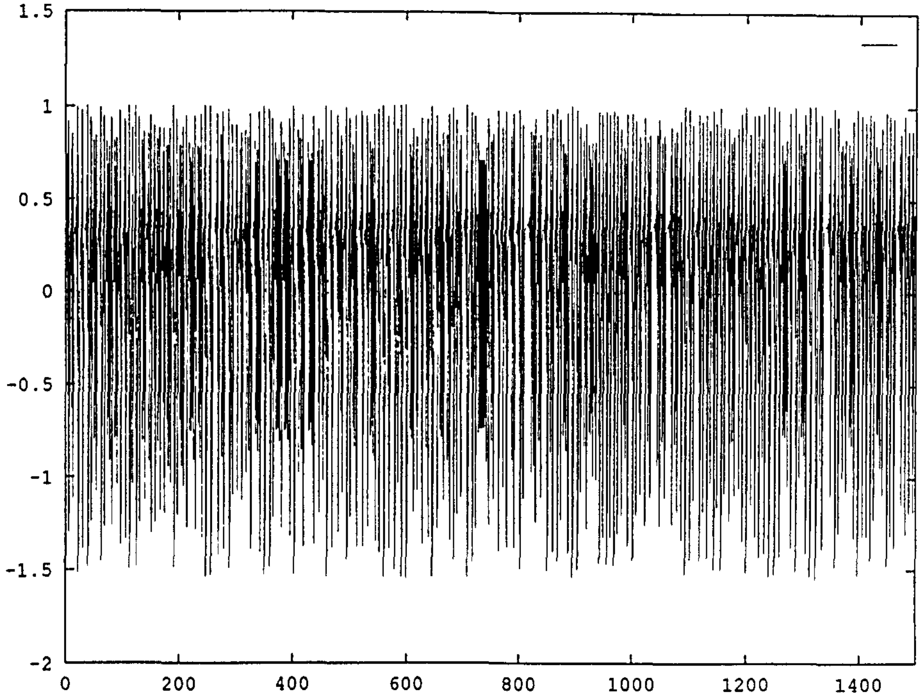

Fig. 1. Time series of Hénon map

The plot in Fig.1 shows clear oscillations around a mean value, which appears to remain constant throughout observation (stationarity). The oscillations have various amplitudes and give no evidence of regularity.

Fig.2 shows the results of a FFT (Fast Fourier Transform) of this time series. By this method, the signal is decomposed to harmonic oscillations at different frequencies. The contribution of each individual wave to the signal is measured by its power. The resulting power spectrum confirms that the frequencies of sine waves are not equally probable — waves of frequency around 300 and between 900 and 1000 (in units of the inverse sampling–time) are represented more prominently; still the power in between these frequency bands does not decrease significantly as it would in a quasiperiodic system.

The autocorrelation of the time series decreases considerably after a few iterations of the Hénon map. In order to reconstruct phase space with the method of time delays, we chose as a rule of thumb the first minimum of the autocorrelation function as suitable. In our case a phase space was reconstructed using a lag of 2. In two–dimensional phase space (Fig.3) the well–known structure of the Hénon attractor appears as expected (i.e. the reconstruction procedure yields the same structure as in a phase space spanned by the analytical dimensions of the map given in (6)).

The Grassberger–Proccacia method was applied with embedding dimensions $1 \leq m \leq 5$ (Fig.4). Quite clearly slopes converge to a value of around 1.25, which is consistent with the dimension 1.26 given in the literature.

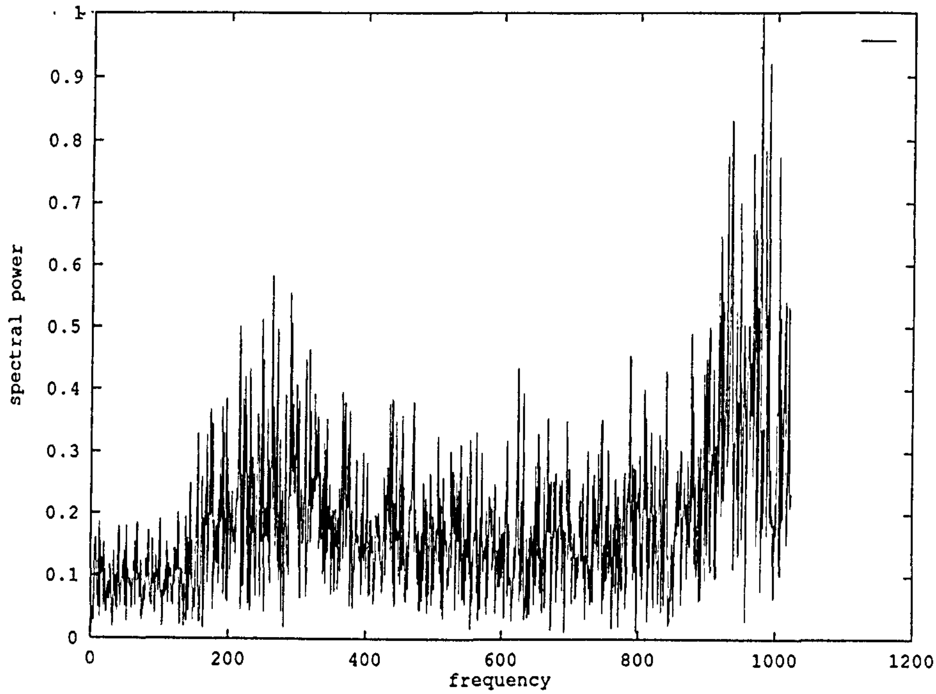

Fig. 2. Power spectrum of Hénon time series

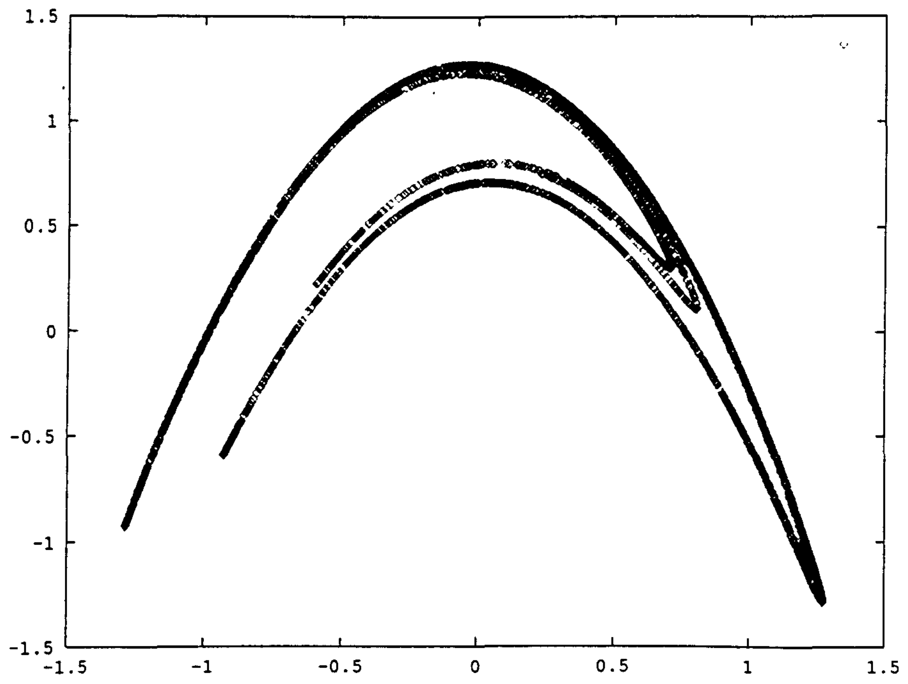

Fig. 3. Reconstructed 2D map from Hénon time series

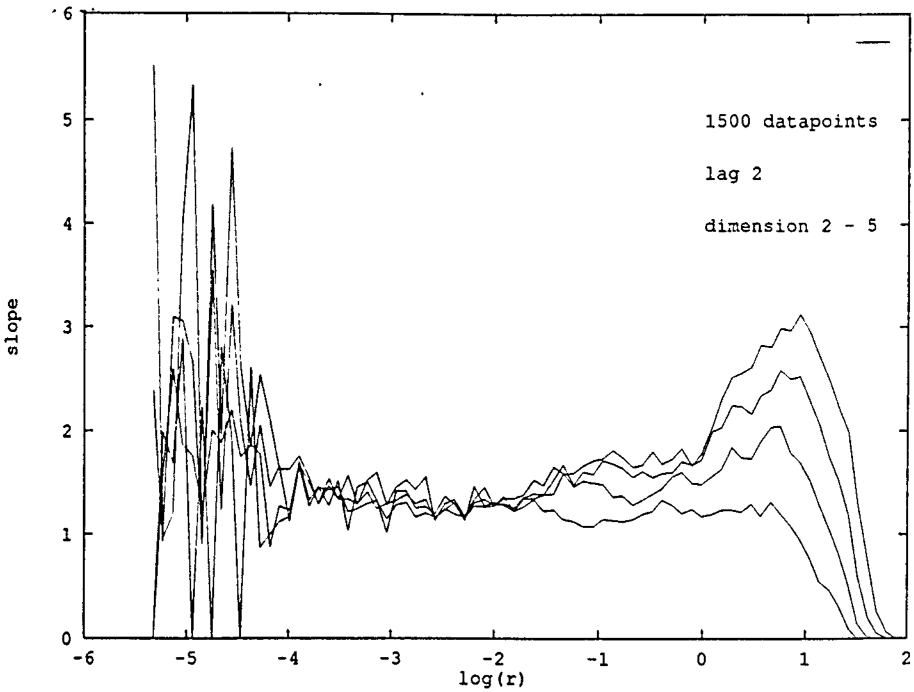

Fig. 4. Slope vs log(r) for Hénon time series with embedding dimensions 2–5, $\tau = 2$

This value appears as a plateau in all curves (except for embedding dimension $m = 1$) within the significant range of r; it is evidently not dependent on m as it would be in the case of white noise.

4. Evaluation of the Correlation Dimension Under Different Constraints for the Time Series

4.1 Experimental noise

Pure noise has no attractor; it fills up the state space. Thus, the correlation dimension of noise just measures the dimension of embedding space. In contrast, the correlation dimension of any reconstructed attractor is independent of embedding space (at least for embedding dimensions near that of the attractor). If noise is added it will tend to smear the attractor in embedding space. As a result the correlation dimension will couple to the embedding dimension. The results of application of noise to the Hénon map are shown in Fig.5. A noise level of 10% destroys the constancy of the correlation dimension for different dimensions of the embedding space, making it impossible to estimate safely the dimension of the attractor.

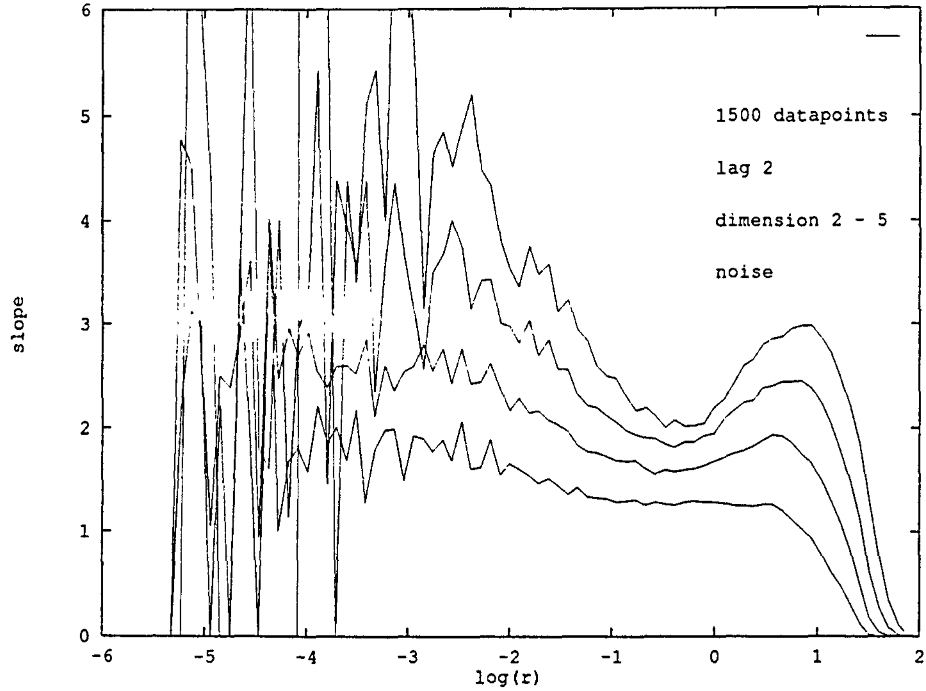

Fig. 5. Slope vs log(r) for Hénon time series (10% noise superposed) with embedding dimensions 2–5, $\tau = 2$

4.2 Length of the Time Series

The length of the time series and the dimension of the embedding state space define the resolution of the attractor reconstruction. As a general rule (see Mayer–Kress, 1987, for a derivation from error estimates on $C(r)$ and the procedure to find the range and the value of constant slope) the length of the time series should be at least $n = b^m$ where m is the dimension of the embedding space and b the minimal tolerable resolution in an coordinate direction. For the Hénon map with dimension $\cong 1.25$ it should be at least $m = 2$ and $b = 10$. In Fig.6 a dimension estimate for $n = 100$ corresponding to the above minimal length is shown. For $m = 2$ a plateau is still visible, but especially for higher embedding dimensions the estimation of the correlation dimension becomes difficult.

It should be noted, though, that in order to observe convergence larger embedding dimensions must be considered as well. The lag τ that was chosen for reconstruction further increases demands for long time series. As was stated in Sect. 2.3, a time series of length n yields $(n - m\tau)$ attractor points. As an example, when the attractor is reconstructed with a lag of 5 and $C(r)$ is tested up to embedding dimension 10, it takes 150 time series points to reconstruct 100 attractor points!

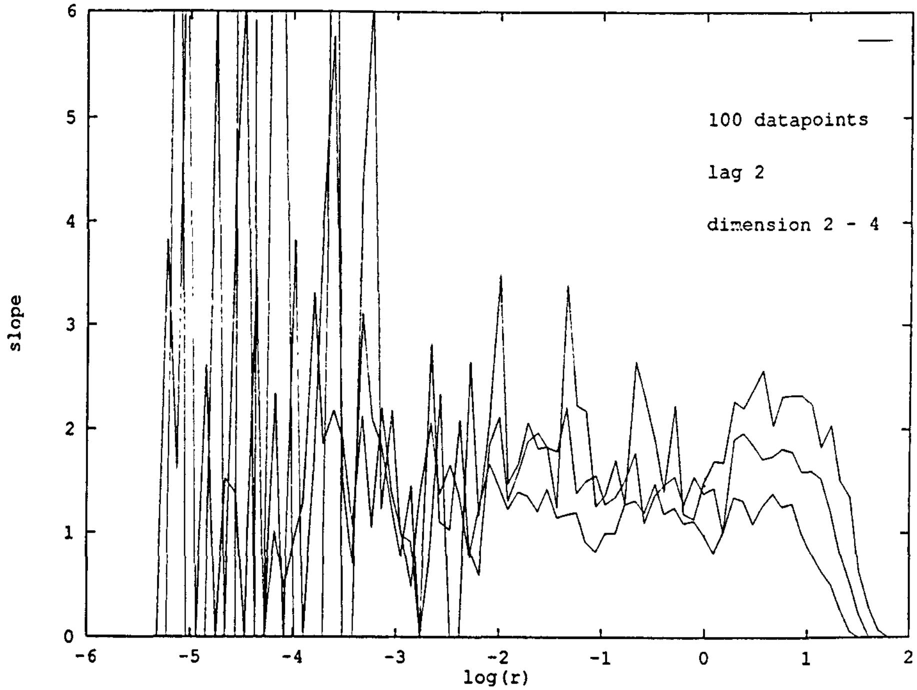

Fig. 6. Slope vs log(r) for Hénon time series (length of series reduced to 100 data points) with embedding dimensions 2–4, $\tau = 2$

4.3 Insufficient Resolution of the Time Series

The resolution of the data from the time series determines the resolution of the reconstructed attractor in the embedding state space. Too low a resolution constrains the evaluation of the correlation to large radii thus making it impossible to measure the filigrane local structure of the attractor. In Fig.7 the resolution of the Hénon time series has been diminished to only 6 bins. This simulates the effects of measuring the Hénon time series with a 6 point rating scale. As can be seen, the plateau in the slope versus log(r) plot has disappeared (with n= 1500 as above).

5. Two Time Series

5.1 Simulation of Group Processes

Figure 8 presents a time series from a simulation system designed in order to model group processes. The simulation computes 'social distances' between members of a group; these may be interpreted like a sociogram (Moreno, 1953) or a 'sculpture' (Schweitzer & Weber, 1983). Further descriptions are given by Tschacher et al. (this volume). The time series maps the spacial expansion of a configuration of eight 'persons' by the sum value of all persons'

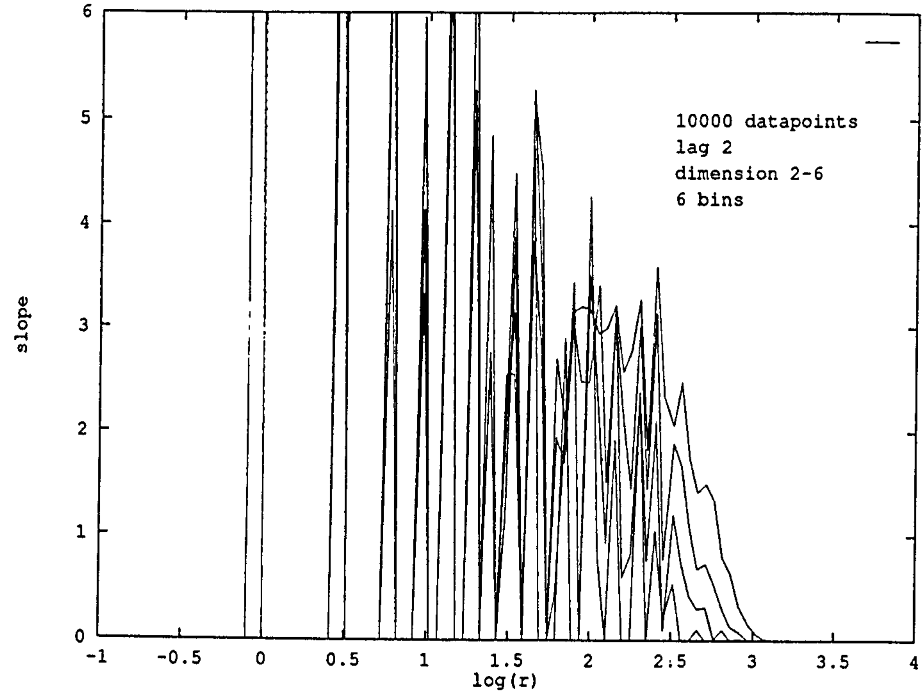

Fig. 7. Slope vs log(r) for Hénon time series (resolution reduced to 6 bins) with embedding dimensions 2–4, $\tau = 2$

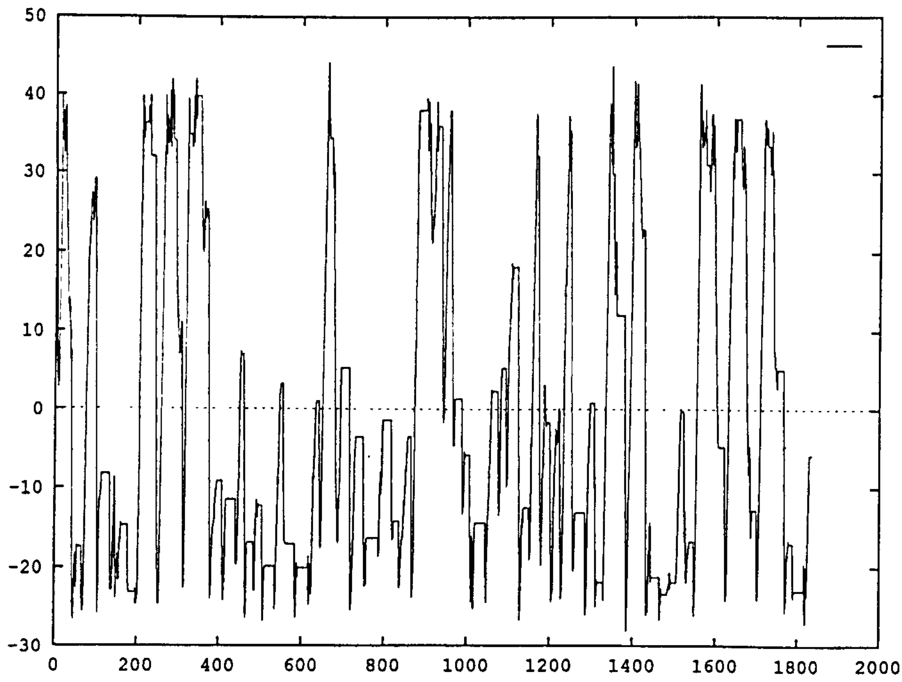

Fig. 8. Time series of simulation of group interaction

distances from a reference point. This value varies from iteration to iteration. In the time series analyzed here model parameters were set in a way corresponding to 'conflict' between group members — the simulation takes a strange course and does not approach any of several stationary states as usual. 'Persons' form a close turbulent unit instead, occasionally building up formations that resemble 'gliders' in cellular automata.

A total of $n = 1800$ values from this process (corresponding to 1800 iterations of the simulation) were sampled for analysis. The series plotted in Fig.8 does not appear to have any recognizable pattern.

The power spectrum is given in Fig.9. Several peaks can be observed, but the spectrum in between is continuous. Thus, the system is not a mere superposition of several different oscillations, even if some oscillations are represented more. As the signal is neither periodic nor white noise, it might be chaotic with some fractal dimension. In order to test for chaos, phase space was reconstructed. To establish coordinates we determined the first minimum of the autocorrelation function which yields an appropriate time delay of $\tau = 40$ (i.e. phase space is spanned by coordinates $x(t), x(t+40), x(t+80)$ etc.). The system in 2D phase space is presented in Fig.10.

Fig.11 gives the slopes of the Grassberger–Proccacia method. Any region of constant ratio of the log–log plot of the correlation dimension $C(r)$ to radius r should show up here as a horizontal plateau of slope curves irrespective of embedding dimension. Although the plot is not as clear as in the case

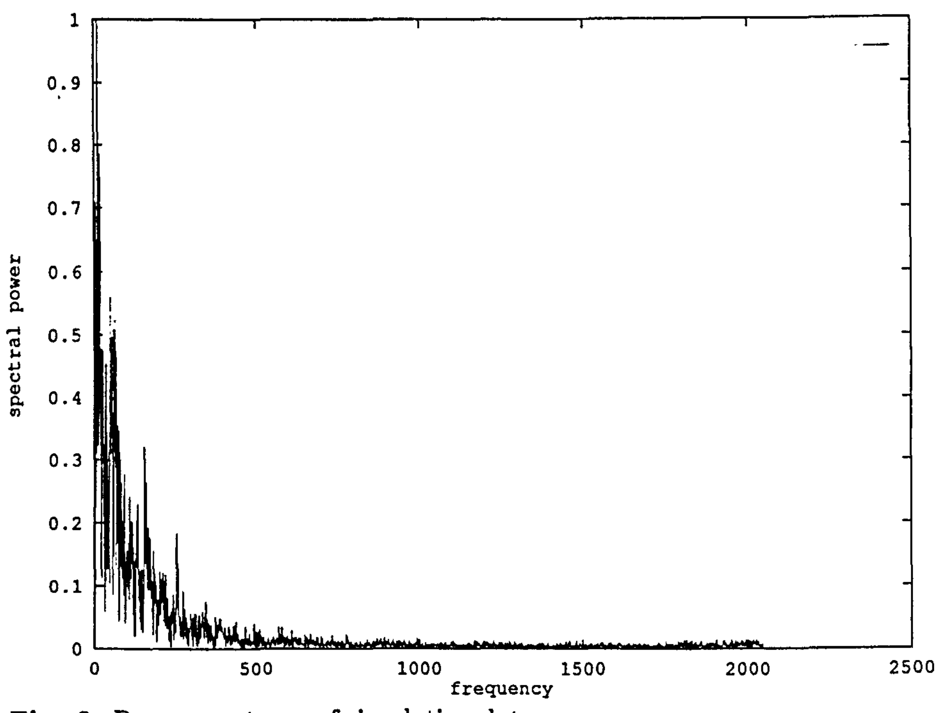

Fig. 9. Power spectrum of simulation data

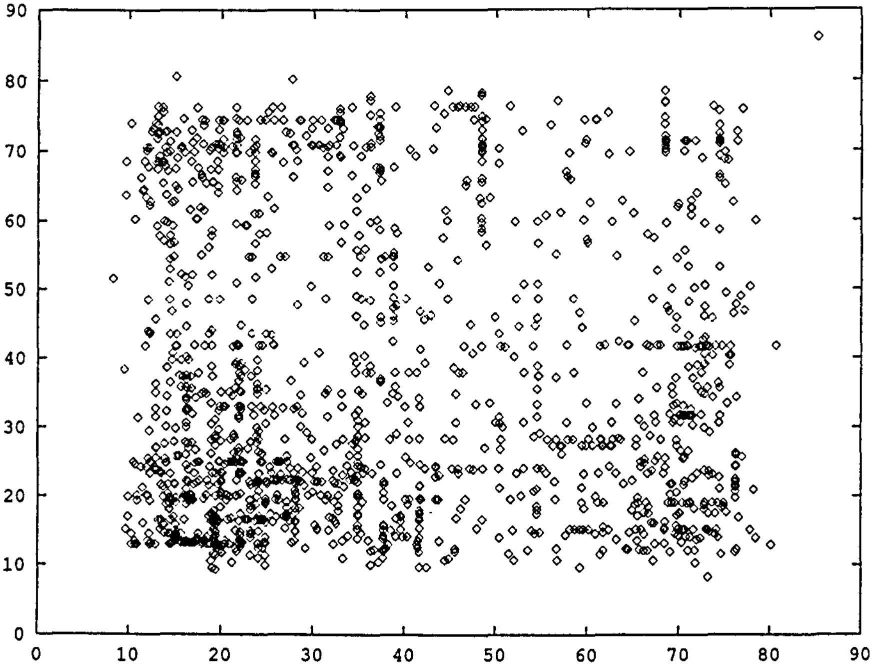

Fig. 10. Reconstructed 2D map from simulation data

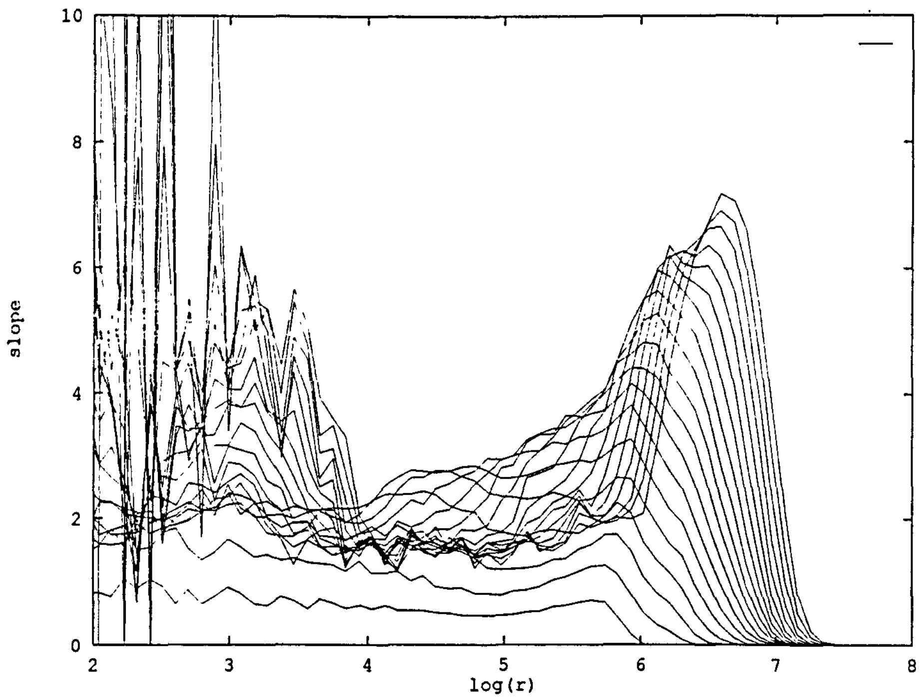

Fig. 11. Slope vs $\log(r)$ for simulation data with embedding dimensions 1–20, $\tau = 40$

of the Hénon map, a plateau of this sort can be derived from the diagram. This points to a fractal attractor of dimension $\nu \cong 1.6$. But peculiarly, the plateau is established only from embedding dimensions $m > 12$. The fact that convergence is not recognizable in lower dimensional embeddings casts doubt on this result (see Sect. 2.4).

Therefore a revised dimension analysis was employed in order to check for artefacts induced by local correlations. For the computation of the correlation $C(r)$ for each attractor point all points with time distances less than τ in the time series were discarded. This procedure yields Fig.12. There is no common plateau left in the slope versus $\log(r)$ diagram. Evidence for a low dimension has disappeared. This indicates that the correlation dimension derived from the Grassberger–Proccacia algorithm was largely due to local effects caused by points on the same segment of the trajectory; the global structure of the hypothesized attractor given by many foldings of the trajectory was not grasped by the standard method.

5.2 An Empirical Time Series of Smoking Behavior

The time series depicted in Fig.13 comprises 1555 observations of an individual's cigarette consumption. Being trained in clinical psychology and behavior therapy the 28-year-old male student took these daily counts to monitor his smoking behavior. Fig.13 indicates that major periods of the present se-

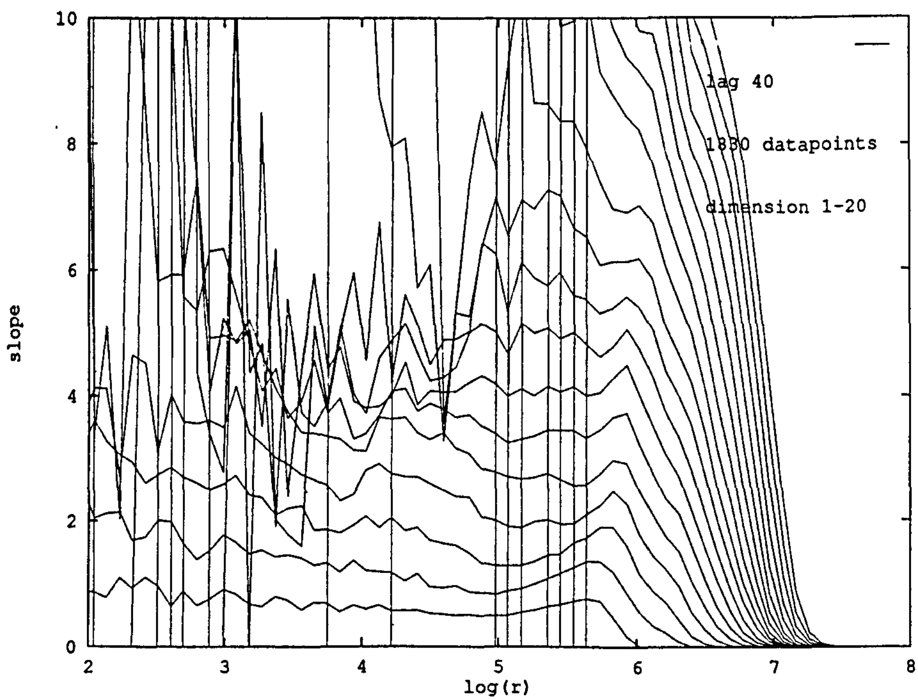

Fig. 12. Corrected slope vs $\log(r)$ for simulation data with embedding dimensions 1–20, $\tau = 40$

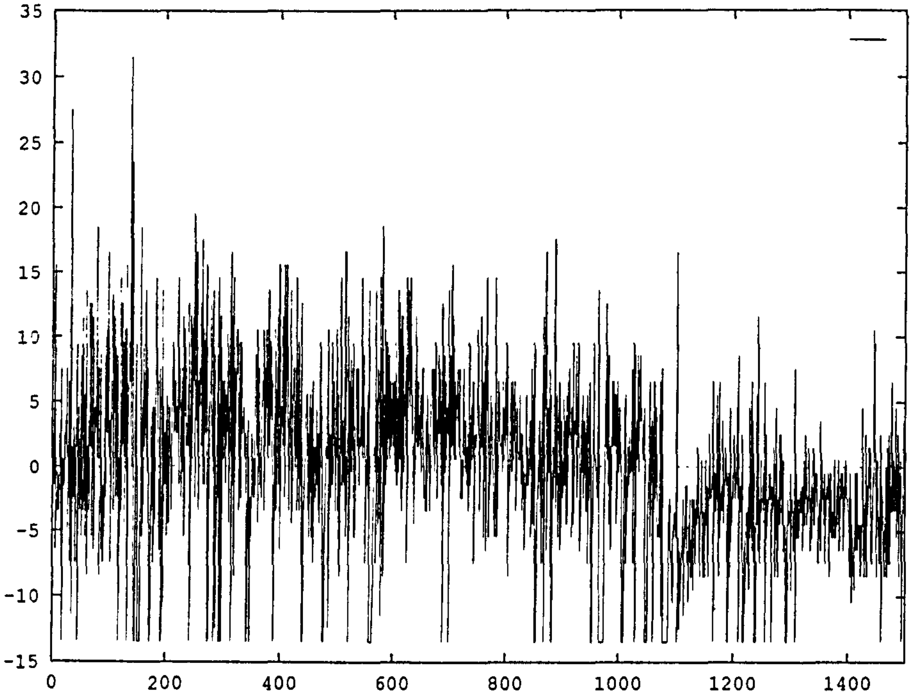

Fig. 13. Time series of smoking data

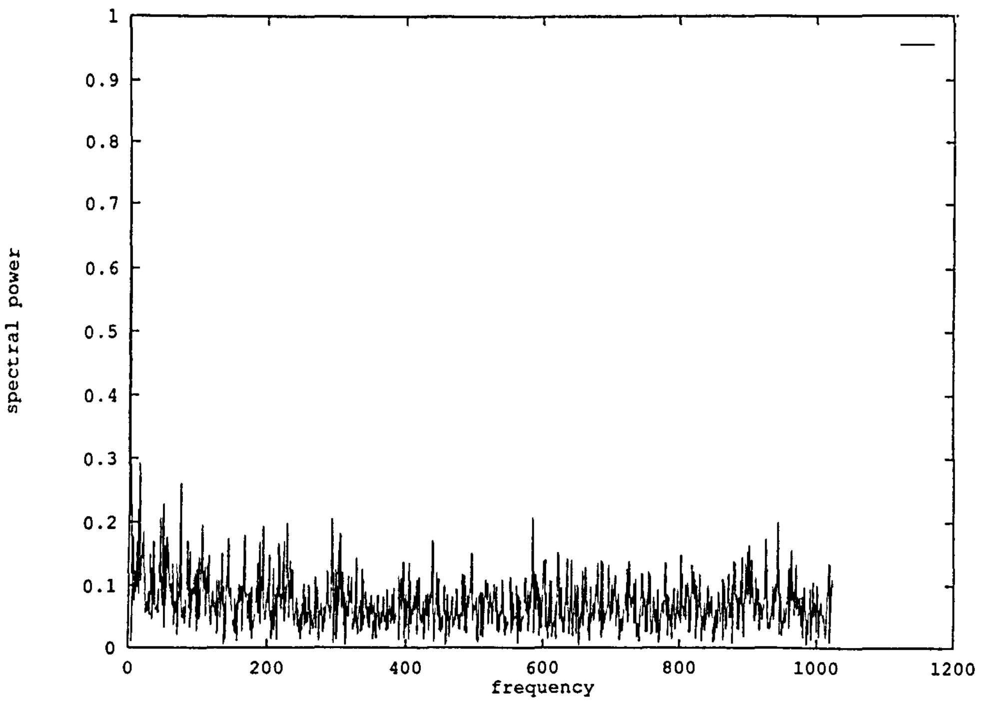

Fig. 14. Power spectrum of smoking data

ries appear to have means different from the means of other periods. There is a distinct level shift after approximately 1100 observations. Though stationarity could have been accomplished by appropriate transformations such as differencing, i.e. the calculation of successive changes in the data values, we preferred to restrict our statistical analysis to the initial 1000 observations and ignored the subsequent data. The power spectrum of the resulting time series given in Fig.14 is smooth and therefore all cycles are assumed to have occurred at an approximately equal intensity. Thus the FFT gives no indication of any hidden periodicity or regularity inherent in the time series.

With respect to the identification of an optimal time delay τ for the reconstruction of an appropriate attractor we again determined the first local minimum of the autocorrelation function, in the present example of $\tau = 4$. Provided that the corresponding embedding dimension is greater than the correlation dimension any range of r with a constant ratio of $\log(C(r))$ versus $\log(r)$ should show up as a common horizontal plateau of various slope curves in the slope versus $\log(r)$ diagram (Fig.16). However, this diagram clearly indicates that there is no such common plateau. Moreover, slopes apparently increase with higher embedding dimensions, consequently the points within successive reconstructions of the attractor exhibit a noise-like pattern.

In general, cigarette smoking is viewed as resulting from the complex interactions of environmental, physiological and psychological processes (Lichtenstein & Brown, 1982). If behavioral time series like the present one could

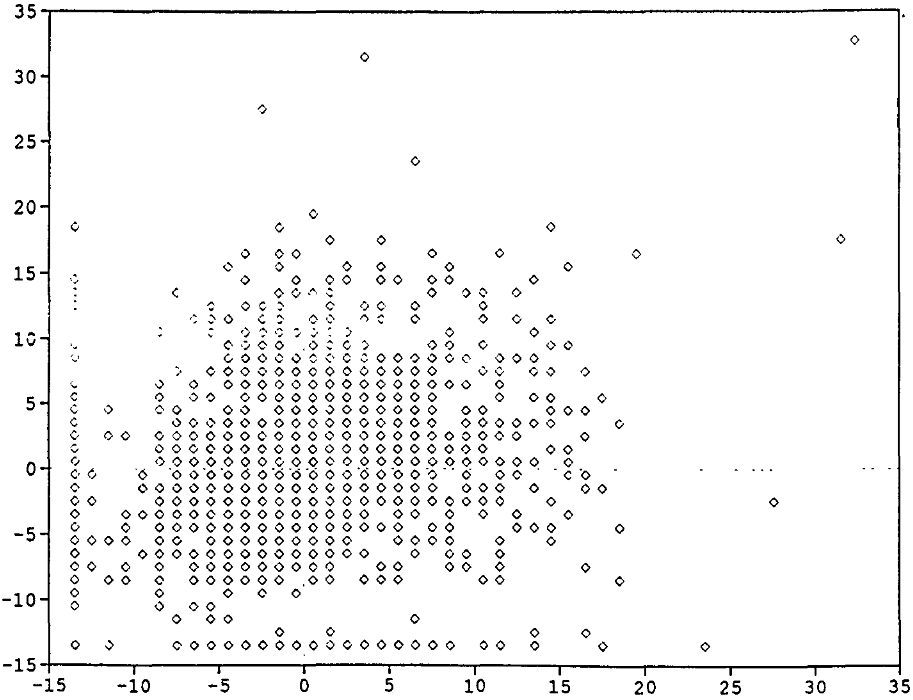

Fig. 15. Reconstructed 2D map from smoking data

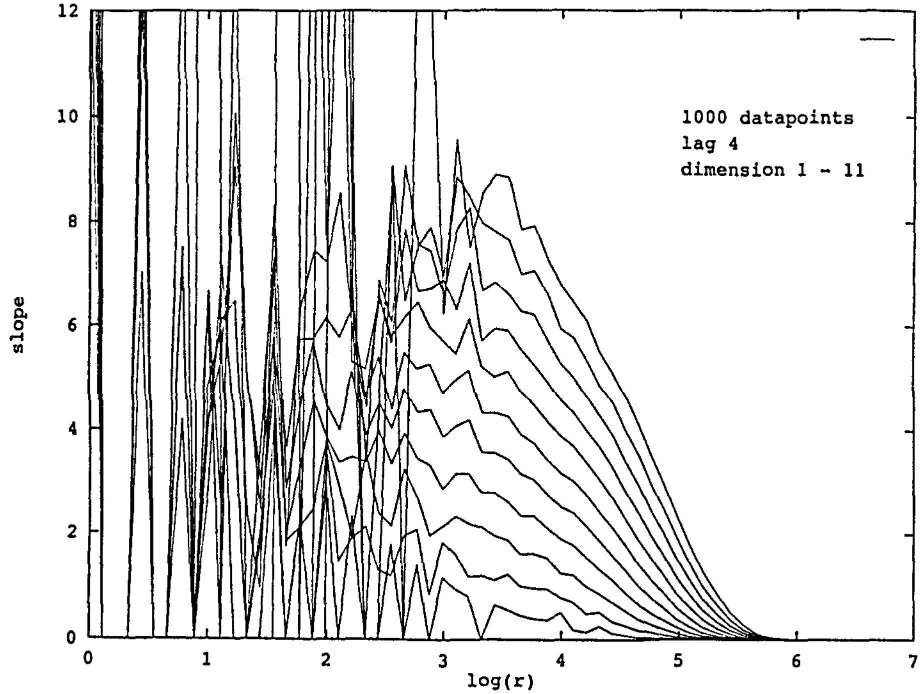

Fig. 16. Slope vs log(r) for smoking data with embedding dimensions 1–11, $\tau = 4$

be reduced to a finite, possibly small number of variables this could be of considerable relevance for clinical theory and therapeutic practice. But this is not yet supported by our data.

6. Options and Restrictions of Dimension Analysis in Psychology

The options of this method have already been stated in Sect. 1: it is possible to evaluate seemingly erratic behavior by determining the dimension of the underlying dynamical regime (if there is any such regime of just a few variables). In psychology, the multicausality of mental and social events has often been emphasized. This is accepted as the reason for either interpreting field data massively or controlling all circumstances by laboratory methods. In either case, complexity is reduced rather artificially. On the other hand, self-organization theory has shown in many instances that under certain conditions very complex systems are governed by but a few order parameters. These conditions seem to be met by most psycho-social systems, so that self-organizing processes should be expected in the field of psychology, too (Schiepek & Tschacher, this volume). Dimension analysis may be a suitable method to detect self-organized systems in apparently nonperiodic time series generated by psycho-social systems.

Even some basic assumptions about the etiology and maintainance of behavior mapped by the time series may be put to test: if our observations reflect an internally controlled process unfolding no matter what environmental influences exist at the time, a simple attractor or an attractor of a finite correlation dimension should be effective. This result would be compatible with psychological theories stressing cognitive 'inner' dynamics as primary causes of behavior. On the other hand, if behavior is to a high degree steered by 'external' control parameters, the fluctuation of these parameters renders any determination of the system's characteristics impossible. The time series then maps many instantiations of quite different systems, so that no overall values of attractor dimensions can be found. In order to map the system, a rigorous control of environmental parameters should be accomplished. The smoking behavior data rather point to this conclusion.

But methodological restrictions also have to be considered following the above discussion of the influence of noise, length of the time series, and resolution of data points. From the discussion some minimum quality standard for empirical time series must be achieved. The resolution of the variable measured and the length of measurement are independent indicators of data quality. They are also limiting factors, i.e. insufficient resolution cannot be remedied by a longer time series and vice versa. For research in psychology, high resolution of measurement is a demand which is quite difficult to satisfy. It will most certainly not be accomplished by the application of usual rating scales since only very simple structures can be mapped in a low-grained, discrete phase space — fractal attractors for one are not simple structures. The implementation of dynamical methods in psychology is largely a question of finding and measuring appropriate observables (Tschacher, 1990).

All of the factors listed in Sect. 4 severely limit the applicability of dimension analysis, maybe even to the point that while there is evidence of complex low-dimensional homeostatic mechanisms theoretically, the accompanying attractors may never be revealed empirically. Still, in our opinion it may prove worthwhile to test further time series from psycho-social systems, bearing in mind these restrictions. To this end observations of (socio-)physiological variables and variables of spacial behavior in restricted settings (such as therapy settings) may be useful.

Acknowledgment

The research reported in this article was in part supported by a grant from the Deutsche Forschungsgemeinschaft (DFG), Grant No. RE402/9-3.

We would like to thank Dipl.-Psych. Uwe Prudlo for permitting us to make use of the cigarette consumption data.

References

BABLOYANTZ, A. 1985. Strange Attractors in the Dynamics of Brain Actitivity. In H. Haken (Ed.): Complex Systems – Operational Approaches to Neurobiology, Physics, and Computers. Berlin: Springer, 116-122.

GLASS, L., SHRIER, A., & BÉLAIR, J. 1986. Chaotic Cardiac Rhythms. In A.V. Holden (Ed.). Chaos. Princeton: University Press, 237-256.

GRAF, K.-E. & ELBERT, T. 1989. Dimensional Analysis of the Waking-EEG. In E. Başar & T.H. Bullock (Eds.). Brain Dynamics. Progress and Perspectives. Berlin: Springer, 174-191.

GRASSBERGER, P., & PROCACCIA, I. 1983. Characterization of Strange Attractors. Physical Review Letters, **50**, 346-349.

HENTSCHEL, H.G.E. & PROCACCIA, I. 1983. The Infinitiv Number of Generalized Dimensions of Fractals and Strange Attractors. Physica **8D**, 435-444.

LICHTENSTEIN, E. & BROWN, R.A. 1982. Current Trends in the Modification of Cigarette Dependence. New York: Plenum.

MAYER-KRESS, G. 1987. Application of Dimension Algorithms to Experimental Data. In H. Bai-lin (Ed.). Directions in Chaos. Singapore: World Scientific.

MORENO, J.L. 1953. Who Shall Survive? Foundations of Sociometry, Group Psychotherapy, and Sociodrama. New York: Beacon House.

PACKARD, N.H, CRUTCHFIELD, J.P., FARMER, J.D., & SHAW, R.S. 1980. Geometry From a Time Series. Phys.Rev.Lett., **45**, 712-716.

SCHAFFER, W.M. & KOT, M. 1986. Differential Systems in Ecology and Epidemiology. In A.V. Holden (Ed.). Chaos. Princeton: University Press, 158-178.

SCHWEITZER, J. & WEBER, G. 1983. Beziehung als Metapher. Die Familienskulptur als diagnostische, therapeutische und Ausbildungstechnik. Familiendynamik, **8**, 113-128.

SIMM, C.W., SAWLEY, M.L., SKIFF, F. & POCHELON, A. 1987. On the Analysis of Experimental Signals for Evidence of Deterministic Chaos. Helvetica Physica Acta, **60**, 510-551.

TAKENS, F. 1981. Detecting Strange Attractors in Turbulence. In D.A. Rand & L.S. Young (Eds.). Lecture Notes in Mathematics. New York: Springer.

TSCHACHER, 1990. Interaktion in selbstorganisierten Systemen. Grundlegung eines dynamisch–synergetischen Forschungsprogramms in der Psychologie. Heidelberg: Asanger.