I. THE KS-ENTROPY

The unpredictability of a sequence of observations, let it be a time series representing a chaotic system or an encoded message, can be characterized by the Kolmogorov-Sinai (KS) entropy. It is the average additional information obtained by observing the state of the system at a certain instant, provided one has observed the whole past.\(^1\) This amount of information per step trivially is zero for a periodic sequence. Since uncorrelated random noises can not be predicted, the KS-entropy for such data equals the information transport rate through the observation channel, i.e. the logarithm of the number of possible different states (e.g., \(\ln 2\) for bit strings). Nontrivial values can be found for the output of systems with memory like Markov processes, but also for deterministically chaotic signals. Numerically, one computes entropies of finite order \(m\), \(h_m\) which in the limit of \(m \to \infty\) converge to the KS-entropy. Unfortunately, as soon as \(m\) becomes large with respect to the length of the time series, the estimates of \(h_m\) are systematically underestimated and decay towards zero for even larger \(m\), such that in many realistic chaotic systems \(h_\infty\) cannot be estimated with reasonable precision. Moreover, certain definitions of the complexity of a sequence\(^2,3\) rely on the way the \(h_m\) converge towards \(h_\infty\). Ebeling\(^3\) postulates for certain systems a power law \(h_m = h_\infty + am^{-\beta}\) with \(\beta=0.2-0.5\). For a numerical determination of \(h_\infty\) and \(\beta\) extremely large \(m\) are required.

In this part the paper we report on a slight modification of the standard algorithm which yields enlarged scaling ranges for the KS-entropy and the information dimension. The realization of the supremum over all partitions generally is nontrivial. In favourable cases generating partitions are known\(^4-6\) and the string of real numbers can be converted into a symbol sequence. One can determine the probabilities of neighbours, yields improved entropy estimates. In the same way, the scaling range for estimates of the information dimension can be extended considerably. The improvement is demonstrated for experimental data.

\[ H_m = -\sum p_{i_1, \ldots, i_m} \ln p_{i_1, \ldots, i_m} \]  

and the Kolmogorov-Sinai entropy

\[ h = \sup_{\mathcal{P}} \lim_{m \to \infty} h_m = \sup_{\mathcal{P}} \lim_{m \to \infty} H_{m+1} - H_m. \]

The realization of the supremum over all partitions generally is nontrivial. In favourable cases generating partitions are known\(^4-6\) and the string of real numbers can be converted into a symbol sequence. One can determine the probabilities of neighbours, yields improved entropy estimates. In the same way, the scaling range for estimates of the information dimension can be extended considerably. The improvement is demonstrated for experimental data.
\[ h_m(e) = H_{m+1}(e) - H_m(e) \]
\[ = \frac{1}{M} \sum_{i=1}^{M} \left( \ln n_i^{(m)}(e) - \ln n_i^{(m+1)}(e) \right). \quad (4) \]

One has to perform the limit \( m \to \infty \) and \( \epsilon \to 0 \) in order to find the asymptotic value of the KS-entropy. The inclusion of finite sample corrections obtained in Ref. 11 essentially amounts to substituting \( \ln n \) by \( \Psi(n+1) \), where \( \Psi \) is the derivative of the gamma function. In particular, this allows for the possibility of \( n \) to be zero. Alternatively, it is sometimes suggested to omit points without neighbours in Eq. (3), but this leads to wrong estimates of the block entropies.

By discussing the problem of reference points without neighbours we are already at the crucial point: The finite sample corrections regularize Eq. (3) converges towards a constant and the estimate of \( h \) becomes zero.

The last line of Eq. (4) shows the problem in a different light: As soon as a given reference point has no \( \epsilon \)-neighbor for some \( m \) (i.e., \( n_i^{(m)} = 0 \)), it will not have one for \( m+1 \), and the two terms cancel each other. Thus such a point does not contribute to the sum, but it is included in the normalization \( 1/M \). Therefore, the right hand side of Eq. (4) yields a systematically too small value for \( h_m(e) \) for large \( m \) and/or small \( \epsilon \), which are the relevant limits. Obviously, we have the same problem when substituting the logarithm by the \( \Psi \)-function. Although our numerical results presented below always are based on the \( \Psi \)-functions, for convenience of notation we suppress in the following the finite sample corrections and deal with the logarithms. In all of the following expressions one can simply replace \( \ln n \) by \( \Psi(n+1) \) to incorporate them.

In Fig. 1 this well known problem of underestimation is demonstrated for a time series of a chaotic map (for details see sec. IV) with a KS-entropy of about 0.27 per time step. This decay towards zero extinguishes the scaling range [i.e., a regime where \( h_m(e) \approx \text{const} \)] completely for those \( \epsilon \) which are small enough to expect the true KS-entropy.

**II. ACCESSIBILITY SAMPLING**

Equation (4) can be written as an expectation value:

\[ h_m(e) = \langle \ln n_i^{(m)}(e) - \ln n_i^{(m+1)}(e) \rangle_M, \quad (5) \]

where \( \ln n_i^{(m)} - \ln n_i^{(m+1)} \) is the logarithm of a decay rate. Thus the KS-entropy is the average of the decay (or escape) rate of neighbours when increasing the time window from \( m \) to \( m+1 \).

As the main reason for the underestimation we isolated the problem that the difference \( \ln n_i^{(m)}(e) - \ln n_i^{(m+1)}(e) \) is trivially bounded by \( \ln n_i^{(m)}(e) \) (requiring \( n_i^{(m+1)} \geq 1 \)). Thus the lack of neighbours does not predominantly introduce statistical fluctuations which could partially cancel each other, but a systematic underestimation, as soon as \( \ln n_i^{(m)} < h \) [or \( \Psi(n_i^{(m)}(e) + 1) < h \)], i.e. as soon as for a given point less than about \( e^h \) neighbours are found.

The remedy we suggest is as simple as effective: Ignore all points without sufficiently many neighbours and compute

\[ \tilde{h}_m^{\min}(e) = \langle \ln n_i^{(m)}(e) - \ln n_i^{(m+1)}(e) \rangle_{n_i^{(m)} > n_{\min}} \quad (6) \]

for reasonable \( n_{\min} \). Thus the importance sampling is substituted by an accessibility sampling. The improvement is demonstrated in Fig. 1 for the above used time series. Notice that this reduced sampling is different from just discarding contributions from Eq. (3).

Our modification can be criticized since it systematically changes the sample on which the average of the decay rates is obtained, in the way that all “thin” parts are ignored. In the following we want to show that this error is small, predominantly of statistical nature and can be controlled empirically.

It is obvious that in the limit of large \( T \), \( \tilde{h}_m^{\min}(e) \) in Eq. (6) converges pointwise towards the true asymptotic value and in this limit is unbiased. However, it might be possible that for every finite \( T \) there are still some systematic deviations [of course much smaller than in Eq. (3)] from the true value, since one ignores the decay rates \( n_i^{(m)} / n_i^{(m+1)} \) on all parts of the space of \( m \)-vectors which are sampled insufficiently. The difference \( \ln n_i^{(m)} - \ln n_i^{(m+1)} \) is the logarithm of the estimate of the local decay rate, whereas \( n_i^{(m)} \) itself, which determines which reference points are excluded from Eq. (6), is \( T \) times the estimate of the invariant measure in the space of \( m \)-vectors. A systematic error can only occur, if these two quantities are correlated. In fact they might be, since the lower the rate by which a reference point loses its
neighbours when increasing \( m \), the more likely this point remains in Eq. (6) for large \( m \). Thus we cannot exclude that our way of sampling still contains some underestimation of the entropy when too many reference points are discarded. However, this error could be easily detected numerically: If the entropy is smaller than one bit per iteration, our way of sampling still contains some underestimation of the entropy when too many reference points are discarded. They turn out to be scattered unsystematically, thus rejecting the suspicion that one might sample the attractor systematically wrong.

Finally, we present in Fig. 3 the strongest argument in favour of our modification: We compute the averages of \( n_{\min} \) as soon as \( n_{\min} \geq 4 \) (as we indeed do) indicates that the effect of ignoring the "thin" parts of the attractor is negligible [of course we observe statistical fluctuations if the number of remaining points in Eq. (6) becomes too small].

In Fig. 2 we present another test: On the attractor, the reference points which were discarded by our way of sampling are indicated. They turn out to be scattered unsystematically, thus suggesting that one might sample the attractor systematically wrong.

Finally, we present in Fig. 3 the strongest argument in favour of our modification: We compute the averages of \( n_{\min} \) as soon as \( n_{\min} \geq 4 \) (as we indeed do) indicates that the effect of ignoring the "thin" parts of the attractor is negligible [of course we observe statistical fluctuations if the number of remaining points in Eq. (6) becomes too small].

In principle, one therefore should compare the continuous curves to the fractal measure, one expects to find a range of \( \epsilon \) and a value \( m_0 \), such that for \( m > m_0 \), \( D_1(\epsilon, m) = d_1 = \text{const.} \). On the large scales generally no scaling can be found, since it is destroyed by the cutoff introduced by the finite extension of the attractor. Like the KS-entropy, \( D_1 \) decays to zero [again one should substitute \( \ln n \) by \( \Psi(n+1) \) to regularize the expression for vanishing \( n \) in the limit of small \( \epsilon \), such that for short data sets the scaling range may disappear. Rewriting Eq. (7) as

\[
D_1 = \frac{1}{M(\ln \epsilon - \ln \epsilon')} \left( \sum_{i=1}^{M} \ln n_i^{(m)}(\epsilon) - \sum_{i=1}^{M} \ln n_i^{(m)}(\epsilon') \right)
\]

reveals the same problem as discussed in connection with the KS-entropy: The relevant quantity determining the information dimension is the rate by which the number of neighbours decays when reducing \( \epsilon \). Again, if the number of neighbors found for some \( \epsilon \) is already too small, this decay rate cannot be computed any more and converges towards zero for points without neighbors. Thus we write Eq. (8) as

\[
\]
an expectation value and base the average only on points with sufficiently many neighbors. Sufficiently many means $n_{\text{min}} > (e/e')^{d_1}$, where of course $d_1$ is not known a priori, but may be estimated from the correlation dimension $d_2$. Thus we claim that

$$D_1(e, m) = \frac{1}{\ln e - \ln e'} \left( \ln n_i^{(m)}(e) - \ln n_i^{(m)}(e') \right)_{n_i > n_{\text{min}}}$$

yields by far enlarged scaling ranges for $D_1$.

The fixed mass method offers a powerful alternative to the plain correlation method described by Eq. (8). However, by the introduction of the accessibility sampling, the correlation method becomes superior, since it allows one to proceed towards smaller length scales (see Fig. 4).

### IV. EXPERIMENTAL DATA

The improvement of entropy and dimension estimates becomes relevant when the data base is limited as in the analysis of experimental time series. Both quantities are of particular interest, since they allow for cross-checks, if one or more Lyapunov exponents of the system are also computed (see, e.g., Ref. 15). The information dimension should coincide with the Kaplan-Yorke dimension, $D_{KY} = k + (\sum_{i=1}^{k} |\lambda_i| + |\lambda_{k+1}|)$, where $k$ is the largest integer such that $\sum_{i=1}^{k} |\lambda_i| > 0$. The KS-entropy should equal the sum of the positive Lyapunov exponents (Pesin’s identity). Substituting the KS-entropy and the information dimension by the correlation entropy and dimension, for whose computation very stable algorithms exist, weakens the above relations to inequalities.

Thanks to the help of the experimental group around Professor Brun at the ETH Zürich we possess a very fine time series of the output of an NMR laser. In a proper Poincaré section these are 38000 points on an attractor of a dimension around 1.3 and a maximal Lyapunov exponent of about 0.3. After nonlinear noise reduction the data are the best one can hope to find under realistic conditions to treat them by nonlinear diagnostics.

In Fig. 4 we present the estimates of the information dimension with and without our reduced sampling ($n_{\text{min}} = 2$).
avoids a prominent source for a systematic underestimation, and in all tested examples (logistic equation, Ikeda map, Hénon map, Lorenz system) yields promising results. This way of sampling requires one to two orders of magnitude less data than the conventional methods, thus allowing for estimates also in experimental time series. We presented arguments and numerical evidence that the fact that averages are not carried out over the correct sample does not introduce a too large bias. This way of reduced sampling can be easily employed also in box-counting algorithms, which for entropy estimates become important if one knows a binary generating partition. In this case one has to ignore sequences \( i_1, i_2, \ldots, i_m \) which occur not often enough.

V. CONCLUSIONS

Ignoring points without sufficiently many neighbours in estimates of the KS-entropy and the information dimension avoids a prominent source for a systematic underestimation, and in all tested examples (logistic equation, Ikeda map, Hénon map, Lorenz system) yields promising results. This way of sampling requires one to two orders of magnitude less data than the conventional methods, thus allowing for estimates also in experimental time series. We presented arguments and numerical evidence that the fact that averages are not carried out over the correct sample does not introduce a too large bias. This way of reduced sampling can be easily employed also in box-counting algorithms, which for entropy estimates become important if one knows a binary generating partition. In this case one has to ignore sequences \( i_1, i_2, \ldots, i_m \) which occur not often enough.

ACKNOWLEDGMENT

This work was supported by DFG within the Graduiertenkolleg “Feldtheoretische und numerische Methoden in der Elementarteilchen- und Statistischen Physik”.

1. P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965).
2. P. Grassberger, “Toward a quantitative theory of self-generated complexity,” Int. J. Theor. Phys. 25, 907–938 (1986).
3. W. Ebeling and G. Nicolis, “Entropy of symbolic sequences: the role of correlations,” Europhys. Lett. 14, 191–196 (1991).
4. P. Grassberger and H. Kantz, “Generating partitions for the dissipative Hénon map,” Phys. Lett. A 133, 235–238 (1985).
5. F. Giovanini and A. Politi, “Homoclinic tangencies, generating partitions, and curvature of invariant manifolds,” J. Phys. A 24, 1837–1887 (1991).
6. F. Christiansen and A. Politi, “Generating partition for the conservative standard map,” Phys. Rev. E 51, R3811–R3814 (1995).
7. P. Gaspard and X.-J. Wang, “Noise, chaos, and (e, r)-entropy per unit time,” Phys. Rep. 235, 291–345 (1993).
8. P. Grassberger and I. Procaccia, “Characterization of strange attractors,” Phys. Rev. Lett. 50, 346–349 (1983).
9. P. Grassberger and I. Procaccia, “Measuring the strangeness of strange attractors,” Physica D 9, 189–208 (1983).
10. P. Grassberger and I. Procaccia, “Estimation of the Kolmogorov entropy from a chaotic signal,” Phys. Rev. A 28, 2591–2593 (1983).
11. A. Cohen and I. Procaccia, “Computing the Kolmogorov entropy from time signals of dissipative and conservative dynamical systems,” Phys. Rev. A 31, 1872–1882 (1985).
12. P. Grassberger, “Finite sample corrections to entropy and dimension estimates,” Phys. Lett. A 128, 369–373 (1988).
13. P. Grassberger, “Generalized dimensions of strange attractors,” Phys. Lett. A 97, 227–234 (1983).
14. P. Grassberger, “Generalizations of the Hausdorff dimension of fractal measures,” Phys. Lett. A 107, 101–105 (1985).
15. R. Badii and A. Politi, “Statistical description of chaotic attractors: the dimension function,” J. Stat. Phys. 40, 725–750 (1985).
16. J.-P. Eckmann, S. Oliffson Kamphorst, D. Ruelle, and S. Ciliberto, “Liapunov exponents from time series,” Phys. Rev. A 34, 4971–4979 (1986).
17. P. Boesiger, E. Brun, and D. Meier, “Solid-state nuclear spin flip maser pumped by dynamic nuclear polarization,” Phys. Rev. Lett. 38, 602–605 (1977).
18. H. Kantz, T. Schreiber, I. Hoffmann, T. Buzug, G. Pfister, L. G. Flepp, J. Simonet, R. Badii, and E. Brun, “Nonlinear noise reduction: A case study on experimental data,” Phys. Rev. E 48, 1529–1538 (1993).
19. D. S. Broomhead and D. Lowe, “Multivariate functional interpolation and adaptive networks,” Complex Syst. 2, 321–355 (1988).
20. M. Casdagli, “Chaos and deterministic versus stochastic nonlinear modeling,” J. R. Stat. Soc. B 54, 303–328 (1991).
21. J.-P. Eckmann and D. Ruelle, “Ergodic theory of chaos and strange attractors,” Rev. Mod. Phys. 57, 617–656 (1985).