Fast chaos versus white noise – entropy analysis and a Fokker-Planck model for the slow dynamics

Holger Kantz[+], Wolfram Just[*], Nilüfer Baba[+], Katrin Gelfert[+], Anja Riegert[+]
[*] Dept. of Theoretical Physics, University of Chemnitz, Germany
permanent address: Queen Mary / Univ. of London, UK.
[+] Max Planck Institute for the Physics of Complex Systems, Nöthnitzer Str. 38, D-01187 Dresden, Germany.

Abstract

We show in detail how methods of time series analysis such as dimension and entropy estimates support the idea that fast low-dimensional chaos can be modeled properly by noise. Motivated by this observation, we derive a formalism by which fast chaotic degrees of freedom in systems with time scale separation can be replaced by a suitable stochastic process. A Fokker Planck equation for the phase space density of the slow degrees of freedom is derived from the full deterministic system by a projection operator technique together with a perturbation expansion. We compare two different projection strategies and illustrate the resulting equations by specifying them for a simple model system.

1 Introduction

The distinction between chaos and noise from observed data is a notoriously difficult task. Recently, this issue has gained new attention when the potential verification of the presence of so called microscopic chaos from the observation of a physical diffusion experiment was discussed. More precisely, by data analysis, in Ref. [1] it is claimed to prove that the seemingly stochastic motion of an observed Brownian particle is due to the deterministic chaotic motion of the pool of small particles colliding with the large one. As a response to this work, a set of papers appeared [2] in all of which strong arguments against the correctness of the reasoning were made. Despite the fact that there is no good reason for any doubts about microscopic chaos, there are many reasons why it should be impossible to prove its existence on the basis of data analysis.

The present paper starts with a review of results which even show the exact opposite, namely, that fast chaotic motion is indistinguishable from a suitable stochastic process in a well defined sense. This serves as motivation and as justification for the second part, where we present a scheme by which fast chaos is replaced by a stochastic process. We will focus on continuous time dynamical systems with time scale separation, i.e., on ordinary differential equations where one set of variables can be called fast and the remaining set of variables is slow, where the ratio between the fast time scale and the slow time scale will be a small parameter κ. In the limit of $\kappa \to 0$, we will replace the full deterministic system in a formally exact perturbation theory by a stochastic differential equation for the slow variables alone, provided that the fast variables are chaotic. Our approach will yield as additional result the fact that fast chaotic degrees of freedom are in some sense small amplitude perturbations of the slow variables (regardless of the magnitude of the coupling terms from the fast to the slow variables), since their effective magnitude is of order of $\sqrt{\kappa}$. Moreover, for periodic fast variables, the stochastic term vanishes and the averaging principle [3] yields the full leading order approximation.

One property of a stochastic process which will be essential for the following discussion lies in the fact that a more precise knowledge about the present state will, asymptotically, not increase our knowledge about the future. As a consequence, the analogue of the Kolmogorov-Sinai (KS) entropy of a stochastic process is infinite. In physics, the concept of a heat bath of harmonic oscillators is a well established way of how to introduce randomness into a deterministic system through a thermodynamic limit. In this case, the infinite entropy comes from the infinite amount of information needed to specify the initial condition of the heat bath in this limit – transport of this information into the observables generates an effectively infinite KS-entropy of this non-chaotic system.

In the situation we are interested in, only a small, definitely finite number of fast degrees of freedom will be responsible for the generation of stochasticity. We require our fast variables to be chaotic themselves. Hence, they possess positive but finite KS-entropy, which, under a suitable rescaling of time, can diverge. To recall, KS-entropy is information production per unit time, such that it scales under $t \to \alpha t$ like $h_{KS} \to h_{KS}/\alpha$.

Before we proceed to a sketch of the derivation of the main result, the derivation of an effective stochastic model for a given deterministic system, and to its thorough discussion, we show in some detail in which sense chaotic deterministic dynamics and stochastic processes are (in-)distinguishable, by dwelling on the concept of ϵ-entropy per unit time. Let us emphasize that C. Beck has studied in great detail the transition from chaos to stochasticity in a variety of low-dimensional chaotic maps, where the observables indicating this transition have been higher-order correlation functions, invariant densities, attractor dimensions, and others [4, 5, 7, 6]. For specific systems he has derived rigorous results, whereas our more general and model free discussion focuses on qualitative aspects.

2 Entropy and dimension estimates

Let $\mathbf{x} \in \Gamma \subset \mathbf{R}^d$ be a state vector, subject to the equation of motion

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})\,. \tag{1}$$

We assume that $\mathbf{f}$ is Lipschitz continuous, such that the initial value problem $\mathbf{x}(0) = \mathbf{x}_0$ has a unique solution. A time series $\mathcal{S} = \{s_1, \ldots, s_N\}$ is obtained from a trajectory $\mathbf{x}(t)$ by recording the values of a measurement function $s_k = s(\mathbf{x}(t = k\Delta))$, where Δ is the sampling interval. If $\mathbf{x}(t)$ are confined to a bounded invariant set $\mathcal{A}$, then s_k are elements of an interval $\mathcal{I}$ on the real axis. The celebrated theorems of Takens and of Casdagli et al. [8] state that m-dimensional vectors $\mathbf{s}_k = (s_k, s_{k-1}, \ldots, s_{k-m+1})$ form an immersion of the invariant set $\mathcal{A}$ to which the trajectory $\mathbf{x}(t)$ is confined, if $m > 2D_f(\mathcal{A})$, i.e., if the dimension m of the embedding space is larger than twice the box-counting dimension of the set $\mathcal{A}$. This has two important implications for the distinction of chaos from noise: If a given time series $\mathcal{S}$ represents a trajectory on an D_f-dimensional invariant set, then also the set of delay vectors $\mathbf{s}_k$ is confined to an D_f-dimensional set, and the finite KS-entropy of the dynamical system Eq.(1) is uniquely related to the KS-entropy of the system in the delay embedding space, whose trajectory is represented by the sequence of $\mathbf{s}_k$.

The model free distinction between chaos and noise based on observed data consists in the verification or falsification whether a given time series represents a finite dimensional object and possesses a finite KS-entropy. The difficulty in this approach lies in the need for a suitable extrapolation, since formally the dimension of a finite number of points (as it is given by the time series) is zero. The benefit of this discussion, however, will be that we will not only be able to distinguish between chaos and noise, but we will also see the equivalence of chaos and a corresponding noise process in a certain sense. Hence, dimension and entropy estimates will be the tools for characterization of an invariant probability measure including all of its (possibly nonlinear) temporal correlations and its scale dependent properties.

2.1 Dimension estimates

In the field of nonlinear time series analysis, numerical dimension analysis has some tradition. The most suitable generalized dimension is the correlation dimension, since it can be computed without systematic finite sample effects [14] and with the largest scaling range employing the Grassberger-Procaccia correlation sum [9] for suitable $k > 0$ (to suppress temporal correlations between vectors $\mathbf{s}_i$ and $\mathbf{s}_j$)

$$C_2(m,\epsilon) = \frac{2}{(N-m-k)(N-m-k-1)} \sum_{i,j:j>i+k,i>m-1}^{N} \Theta(\epsilon - ||\mathbf{s}_i - \mathbf{s}_j||) \; , \quad (2)$$

where m is the dimensionality of the delay vectors $\mathbf{s}_i$. If the unobserved state vectors $\mathbf{x}(t = k\Delta)$ are confined to a subset with correlation dimension D_2 of the state space, then for $m > D_2$ the correlation sum $C_2(m,\epsilon)$ typically exhibits a scaling range for small ϵ with $C_2(m,\epsilon) \propto \epsilon^{D_2}$. On the large scales, finite size effects of the set (edge effects) destroy the scaling evidently, but in addition entropic folding effects can occur for either high entropic signals or for unsuitably large sampling intervals of the data. In order to keep the presentation short, we refer the interested reader to Ref.[10] for details.

As an illustration of the effect, we recall some features of a very simple 2-dimensional map:

$$\begin{aligned} x_{n+1} &= e^{-\mu} x_n + \sqrt{\mu} y_n \\ y_{n+1} &= 1 - 2y_n^2 \; . \end{aligned} \quad (3)$$

C. Beck called this "dynamical system of Langevin type" and showed [4] that in the limit $\mu \to 0$ the dynamics of the variable x converges to an Ornstein-Uhlenbeck process. In [11], the detailed entropy and dimension analysis of this family of systems was performed. For vanishing μ, the entropy per correlation time of x diverges. In Fig.1 the dimension estimates for one fixed small value of μ are shown: Whereas on the small scales the dimension $D_2 = 2$ is clearly recovered, on the large scales the data seem to fill volumes in any arbitrarily high-dimensional embedding space, which is the typical signature of a stochastic system. For even smaller values of μ, the deterministic regime is almost inaccessible and the data appear stochastic for any numerical analysis, even if they were created from the same low-dimensional deterministic system.

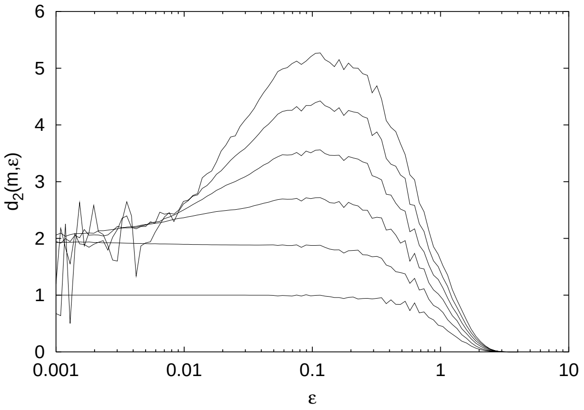

Figure 1: Numerical dimension estimates for a time series of x_n of the system Eq.(3) with $\mu = 1/16$, embedding dimension $m = 1, \ldots, 6$ from bottom to top (technical remark: time lag 3 in the delay vectors, i.e., $\mathbf{s}_j = (s_j, s_{j-3}, \ldots, s_{j-3m+3})$). Due to folding effects, on the large length scales the data seem to be stochastic, since they fill volumes in arbitrarily high embedding dimensions m.

In summary, data from systems with high entropy per unit time appear stochastic on the large length scales when analyzed in terms of dimension. This is neither related to the particular algorithm nor to the choice of the correlation dimension but a consequence of the folding of the invariant set in the time delay embedding space.

2.2 Coarse grained dynamical entropies

The Kolmogorov-Sinai entropy of a dynamical system can be estimated through the computation of the positive Lyapunov exponents and the usage of the Pesin identity. In time series analysis, however, it is typically necessary to go back to its proper definition. Let $\mathcal{P}$ be a partition of the interval $\mathcal{I} \subset \mathbf{R}$ to which the time series data s_k are confined, i.e., $\mathcal{P}$ is a set of $N_{\mathcal{P}}$ disjoint cells satisfying $\cup_i c_i \supset \mathcal{I}$. One defines the joint probabilities $p_{i_1,\ldots,i_m}$ that a subsequence of m successive observations fulfils $s_k \in c_{i_1}$, $s_{k+1} \in c_{i_2}$, etc.. Block entropies of block length m are defined through

$$H_m = -\sum_{i_1,\ldots,i_m} p_{i_1,\ldots,i_m} \ln p_{i_1,\ldots,i_m} \tag{4}$$

and conditional entropies as

$$h_m = H_{m+1} - H_m \ . \tag{5}$$

The Kolmogorov-Sinai entropy of a dynamical system is given by

$$h_{KS} = \sup_{\mathcal{P}} \lim_{m\to\infty} h_m(\mathcal{P}) \ , \tag{6}$$

where $\sup_{\mathcal{P}}$ denotes the supremum over all possible partitions $\mathcal{P}$. For time series data, the probabilities $p_{i_1,\ldots,i_m}$ are estimated through the relative frequencies of the occurrence of the corresponding events (maximum likelihood estimators).

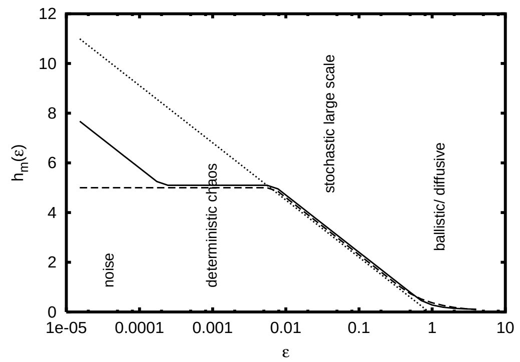

Figure 2: Prototypic behavior of the ϵ-entropy for $m > D$ for a chaotic process plus measurement noise (continuous line), for the same process without noise (dashed line) and the stochastic large scale behavior (dotted line). For more explanation see text.

Consequently, the strict upper limit in m for which a meaningful block entropy can be computed is when ${N_\mathcal{P}}^m > N$ (N: length of the time series, $N_\mathcal{P}$: number of cells). Hence, the $\lim_{m\to\infty}$ is unaccessible, but also the supremum over all possible partitions cannot be computed.

Apart from the practical impossibility to compute h_{KS}, there are also many systems with interesting scale dependent structure. This makes entropy expressions with an explicit scale dependence desirable. If a partition $\mathcal{P}_\epsilon$ is labeled by the scale $\epsilon = \max\{\text{diam } c_i : c_i \in \mathcal{P}\}$, Gaspard and Wang [13] show that

$$h(\epsilon) = \inf_{\mathcal{P}_\epsilon} \lim_{m\to\infty} h_m(\mathcal{P}_\epsilon) \tag{7}$$

is a meaningful definition of the ϵ-entropy. The ϵ-entropy is, unlike the KS-entropy, not invariant under coordinate transforms. Still, Eq.(7) cannot be exploited with a finite amount of data.

It is reasonable to introduce two approximations: First, one replaces the infimum over all ϵ-partitions by a unique ϵ-covering. Second, one replaces the Shannon entropy $-\sum p_i \ln p_i$ by the second order Renyi entropy $-\ln \sum p_i^2$ (which lacks the additivity property) in order to eliminate systematic finite sample errors [14]. In the limit $\epsilon \to 0$ this entropy is known to be a lower bound of the KS-entropy [15]. An estimate of the ϵ-entropy with block length m is given by [13, 12]:

$$h_m(\epsilon) = \ln C(m,\epsilon) - \ln C(m+1,\epsilon)\ , \tag{8}$$

where $C(m,\epsilon)$ is the normalized correlation sum Eq.(2).

The behavior of $h_m(\epsilon)$ as a function of the spatial resolution ϵ yields additional insight into the nature of the process underlying the data. The typical behavior for deterministic chaotic and for stochastic processes is summarized in the following table, where the constants $H_m^{(c)}$ and $h_m^{(c)}$ will be discussed later:

process	$H_m(\epsilon)$	$h_m(\epsilon) = H_{m+1}(\epsilon) - H_m(\epsilon)$
deterministic, $m \gg D_2$, ϵ sufficiently small	$H_m^{(c)} - D_2 \ln\epsilon$	$H_{m+1}^{(c)} - H_m^{(c)} \approx h_{KS}$
stationary stochastic deterministic $m < D_2$ or large ϵ	$H_m^{(c)} - m\ln\epsilon$	$h_m^{(c)} - \ln\epsilon$

The KS-entropy of a deterministic system can only be seen at sufficiently small scales. On the large length scales, where $h_m(\epsilon) < h_{KS}$, deterministic data are indistinguishable from stochastic data, i.e. $h_m(\epsilon) \propto -\ln\epsilon$. This is a consequence of the trivial bound $h_m(\epsilon) \leq -\ln\epsilon$ together with the nontrivial entropy reduction by non-uniformity of the coarse grained measure. Deviations from the $-\ln\epsilon$ behavior on even larger length scales can have different origins. In deterministic data, these are edge effects because of the finite range of the invariant set and hence the finite range of the observable. In stochastic data, when no such limitations exists (e.g., Gaussian random variables), the lack of recurrence governs the large scale cross-over regime, which we will neglect in the following.

Coarse grained dynamical entropies hence contain nontrivial information for both deterministic and stochastic data, since they can possess a rich structure as a function of the length scale. The lack of invariance of $h_m(\epsilon)$ under coordinate changes can be turned into a virtue: For smooth probability densities $\rho(\mathbf{s})$ (the density in the m-dimensional embedding space), the constant $h_m^{(c)} := H_{m+1}^{(c)} - H_m^{(c)}$ is given by the continuous entropy

$$H_m^{(c)} = -\ln \int \rho^2(\mathbf{s}) d^m s \; . \tag{9}$$

If the probability distribution factorizes into identical one-dimensional distributions, one has evidently $H_m^{(c)} = mH_1^{(c)}$ and hence $h_m^{(c)} = h_0^{(c)} = const.$. Equally straightforward considerations show that for a Markov-chain of order m, all continuous entropies for $m' > m$ are identical and so are the corresponding $h_{m'}(\epsilon)$-curves. If the m-dimensional density $\rho(\mathbf{s})$ is a multivariate Gaussian with the co-variance matrix $\mathbf{C}_{ij}^{(m)} := \langle s_i s_j \rangle$, the continuous entropies read

$$H_m^{(c)} = \frac{m}{2}\ln\pi + \ln[\det\, \mathbf{C}^{(m)}] \; . \tag{10}$$

The co-variance matrix is fully determined by the auto-correlation function $c(\tau) = \langle s(t)s(t-\tau)\rangle$ of the signal $s(t)$, hence one finds

$$\det\, \mathbf{C}^{(1)} = c(0) \tag{11}$$

$$\det\, \mathbf{C}^{(2)} = c(0)^2 \left(1 - \frac{c(\Delta)^2}{c(0)^2}\right) \tag{12}$$

$$\det\, \mathbf{C}^{(3)} = c(0)^3 \left(1 - 2\frac{c(\Delta)^2}{c(0)^2} + \frac{c(2\Delta)}{c(0)}\right)\left(1 - \frac{c(2\Delta)}{c(0)}\right) \; . \tag{13}$$

In the special case of exponentially decaying correlations one finds $c(2\Delta) = c(\Delta)^2/c(0)$, and hence again $h_2^{(c)} = h_1^{(c)}$ as expected.

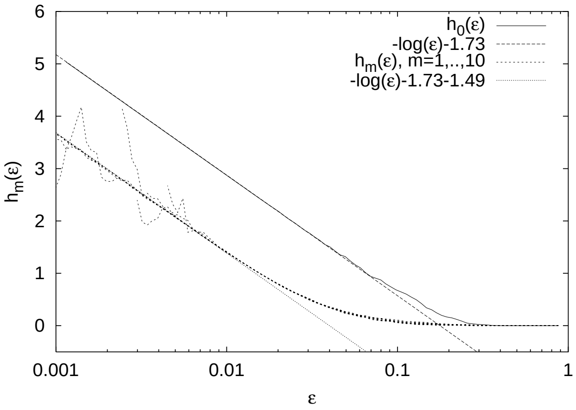

Figure 3: The ϵ-entropies obtained numerically from a time series of the Ornstein-Uhlenbeck process. The straight lines are the theoretical predictions with the constants h_0^c and h_1^c replaced by their numerical values (no fitting parameters), see text. The curves for $m = 1$ to $m = 10$ are almost perfectly superimposing each other and almost indistinguishable. Fluctuations of the curves for $m > 7$ and small ϵ are statistical finite sample effects.

2.3 Linearly filtered Lorenz system

Let us consider a simple extension of the Lorenz system,

$$\begin{aligned}
\dot{x} &= -\alpha x + y_1 \\
\dot{y_1} &= \frac{1}{\kappa} s(y_2 - y_1) \\
\dot{y_2} &= \frac{1}{\kappa}(y_1(r - y_3) - y_2) \\
\dot{y_3} &= \frac{1}{\kappa}(y_1 y_2 - b y_3)
\end{aligned} \tag{14}$$

with the standard parameters $s = 10$, $r = 28$, and $b = 8/3$, and $\kappa = 1$ here. We record a time series of the variable $x(t)$ with a sampling interval which is arbitrarily chosen to be $\Delta = \alpha/10$, hence we record 10 samples per correlation time $1/\alpha$. Since the negative Lyapunov exponent of the Lorenz system for the standard parameters is about -10, for $\alpha > -10$ the linear degree of freedom increases the attractor dimension roughly by unity [16]. Hence, we expect in the numerical dimension and entropy analysis a dimension slightly above three and an entropy given by the positive Lyapunov exponent of the Lorenz system times the sampling interval.

Before we study the ϵ-entropies of this system, we show in Fig.3 those for data from the Ornstein-Uhlenbeck process, i.e., data generated by $\dot{x} = -\alpha x + \xi$, where ξ is white Gaussian noise of unit variance. We observe $h_m(\epsilon) \propto -\ln \epsilon$, where the offsets $h_0^{(c)} = H_1^{(c)} = \ln(\sqrt{\pi}\sigma)$ (where the variance of the data is σ^2), and $h_{m>1} = h_1^{(c)} = H_2^{(c)} - H_1^{(c)} = \ln(\sqrt{\pi}\sigma) + \ln\sqrt{1 - c(\Delta)/\sigma^4}$ ($c(\Delta)$ is the value of the auto-correlation function for lag Δ, i.e. in this case $c(\Delta) = \sigma^2 \exp(-\Delta\alpha)$). In summary, $h_1^{(c)} \approx h_0^{(c)} + \frac{1}{2}\ln(2\alpha\Delta)$. These values are in fact recovered with high accuracy by the numerics.

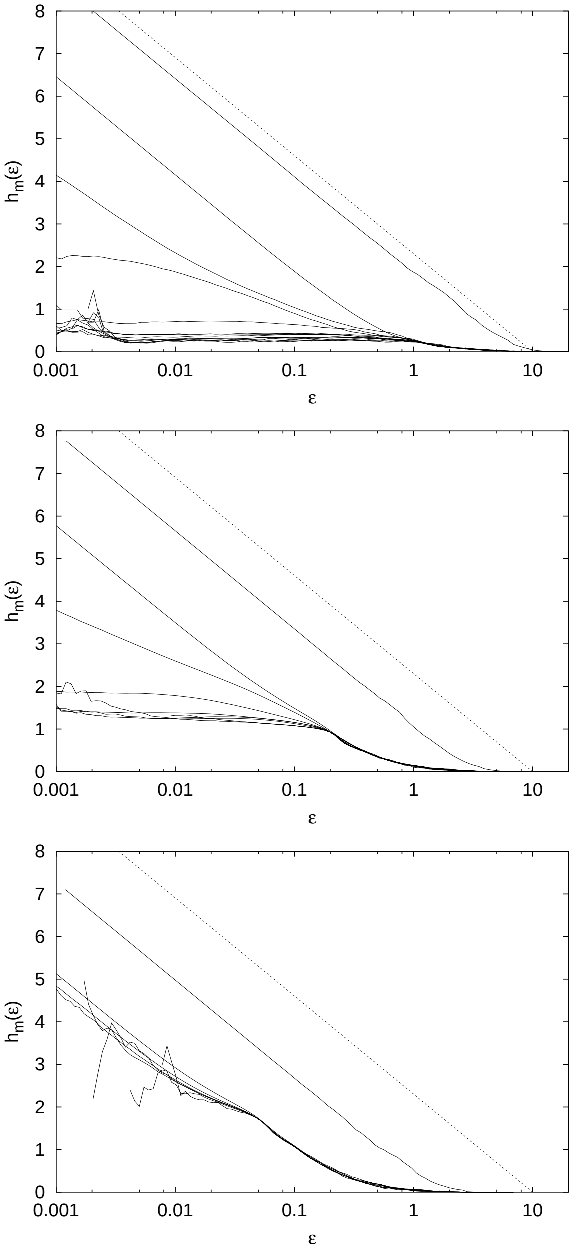

Figure 4: The ϵ-entropies $h_m(\epsilon)$ for the linearly filtered Lorenz system for (a) $\alpha = 0.1$, (b) $\alpha = 0.5$, (c) $\alpha = 2$. Apart from statistical fluctuations, the continuous curves show $m = 0$ to $m = 10$ from top to bottom.

Having fully interpreted the curves of the Ornstein-Uhlenbeck-process, we will now discuss the curves of the deterministic processes Eq.(14) for different α shown in Fig.4. For $\alpha = 0.1$ (panel (a)) the deterministic nature of the whole system is clearly visible, since for $m > 4$ the convergence of $h_m(\epsilon)$ to a constant is evident (compare the table following Eq.(8)). We can estimate an entropy of about 0.25 per time unit, which is in fact the value we expect from the knowledge of the maximal Lyapunov exponent of the Lorenz system and the way how we have sampled the data. For $m < 3$, the block length is too short to yield the deterministic properties and we see the signature of a stochastic process, as argued before. On the large length scales, also the ($m > 3$)-curves possess an unspecific crossover regime. For $\alpha = 0.5$ (panel (b)), a finite value of the entropies is still found. Since we sampled the time continuous process Eq.(14) with a sampling interval proportional to α, larger α implies a larger entropy per sampling interval. Consequently, the asymptotic value of the ϵ-entropy for large m and small ϵ has increased by a factor of 5 with respect to the case $\alpha = 0.1$. The relevant feature which illustrates our statements from the last section can be seen on the large length scales: A stochastic regime starts to emerge, where the different h_m for $m > 0$ cannot be distinguished from each other. For $\alpha = 2$, this stochastic regime covers about one order of magnitude. Within the given spatial resolution, joint probabilities $p_{i_1,\ldots,i_m}$ factorize into the following product of marginal and two-point conditional probabilities: $p_{i_1,\ldots,i_m} \approx p(i_m|i_{m-1}) \ldots p(i_2|i_1) p_{i_1}$. Hence, on these large scales, the system looks like a Markov chain of order one, which is a property of the white noise driven Ornstein-Uhlenbeck process.

A closer look at Fig.4(c) reveals additional details. The ϵ-entropies for $\alpha = 2$ show two different stochastic regimes: on length scales $\epsilon > 0.05$ one finds the Ornstein-Uhlenbeck behavior (superposition of all curves for $m > 0$), whereas for smaller ϵ, we see a stochastic process with some longer memory (no clear evidence for a low order Markov chain due to lack of perfect superposition of the curves, see the discussion following Eq.(9)). The crossover to the deterministic behavior $h_m(\epsilon) \approx h_{KS}$ cannot be reached numerically and would be around $\epsilon = 5 \cdot 10^{-4}$. Hence, the loss of correlations in the Ornstein-Uhlenbeck regime is a nontrivial feature and not simply related to the fact that entropies have a trivial upper bound $\propto -\ln \epsilon$.

In summary, by studying scale dependent dynamical entropies, we have been able to establish several facts which are relevant for the following section: The ϵ-entropy of a high entropic but deterministic signal converges to a finite value in the limit of small ϵ and large m. On length scales ϵ which are large compared to $\sigma \exp(-h_{KS}\Delta)$ (σ^2 denotes the variance of the data), instead the behavior typical of stochastic data is found: $h(\epsilon) \propto -\ln \epsilon$. Moreover, it is intuitively clear that the effective noise introduced by the high entropic chaos is in good approximation white noise, as seen by the lack of correlations in Fig.4(c) for large ϵ, since the large value of the entropy was reached through an implicit compression of time, which renders the exponential decay of the auto-correlation function of the mixing, chaotic signals almost δ-shaped in time. However, the fact that already for finite time scale separation correlations are absent on the large scales is a nontrivial observation.

3 A Fokker-Planck equation for the slow degrees of freedom

In the last section we have reported some observations which strongly and in a quantitative way suggest that high entropic chaos acts, at least on a well defined range of length scales, as white noise. In fact, our linearly filtered Lorenz system together with our way of sampling is a system with time scale separation, since by rescaling the time in Eq.(14) by $t \to \alpha t$, we end up in the class of the following two time scale systems, to which our analysis will be restricted:

We assume that we can decompose the set of phase space variables into two groups of dimensionality d_x and d_y, where x are called the slow variables and y the fast ones (from now on, vectors will not any more be denoted by particular symbols, and in order to avoid heavy indexing, we will write the expressions for the slow dynamics for $d_x = 1$; the generalization is straightforward). If we assume that the right hand sides of the following differential equations, f and g, are of the order of unity, the time scale separation can be mediated by a parameter κ through:

$$\dot{x} = f(x,y) \tag{15a}$$

$$\dot{y} = \frac{1}{\kappa} g(x,y) \quad , \tag{15b}$$

where $0 < \kappa \ll 1$.

The ultimate goal of this section will be to derive a stochastic differential equation for x alone,

$$\dot{x} = \widetilde{f}(x,\xi) \; , \tag{16}$$

where ξ is a multidimensional white noise process. Instead of continuing the line of thought of the preceding section, we will derive an evolution equation for the phase space density of the slow variables. This will supply the relevant answer to three questions which could not be generally answered in the last section, namely, how the deterministic part of the evolution equation for the slow variables should look like, which distribution the effective noise process on the large scales will assume, and with which amplitude it will be coupled to the slow variables. Previous studies have proven [4, 6] that for a particular realization of the fast chaotic process in terms of time discrete maps a modeling of the fast motion by Gaussian white noise becomes possible in certain scaling limits. Within our approach we are able to deal with quite general systems. We will estimate the validity of our perturbation expansions by comparison of different approximation schemes. Since we do not focus on a particular model we can contribute to the question which properties of the slow vector field f and of the fast motion g determine the diffusion matrix and how the renormalization of the slow vector field by fast fluctuations looks like.

The time evolution of phase space densities $\rho_t(x,y)$ corresponding to the trajectory-wise description of Eq.(15) is given by the corresponding Liouville–like equation

$$\frac{\partial \rho_t}{\partial t} = -\mathcal{L}\rho_t \quad , \tag{17}$$

where

$$\mathcal{L} := \frac{1}{\kappa}\mathcal{L}^{(0)} + \mathcal{L}^{(1)} := \frac{1}{\kappa}\frac{\partial}{\partial y} g(x,y) + \frac{\partial}{\partial x} f(x,y) \tag{18}$$

denotes the generator of the dynamics. We assume for simplicity that all densities are smooth and can be treated like ordinary functions[1]. We are interested in the distribution of the slow variables

$$\bar{\rho}_t(x) = \int dy\, \rho_t(x,y) \quad . \tag{19}$$

It is the goal to derive a closed equation of motion for this distribution (19) from the full equation of motion (17). The spirit of such a procedure consists in replacing the distribution $\rho(x,y)$ by a relevant density

$$\rho_t^{rel}(x,y) = \widetilde{\rho}(y|x)\bar{\rho}_t(x) \quad . \tag{20}$$

Regardless of the special form of $\widetilde{\rho}(y|x)$ it is possible to write down an exact equation of motion for $\bar{\rho}_t$ by using well established projection operator techniques [17]. The usefulness and the evaluation of such equations in terms of a perturbation expansion depends crucially on the particular choice for $\widetilde{\rho}$.

Two different choices of $\widetilde{\rho}$ are appealing. One would guess that as long as y is fast enough it adjusts almost instantaneously according to the invariant distribution of the fast equation (15b) with x being considered as a fixed parameter. Thus the fast variables are modeled by the invariant density of eq.(15b), $\rho_{ad}(y;x)$, with x being considered as a fixed parameter. Such a choice has the advantage that averages with respect to y can be computed as time averages provided certain ergodic properties are met. If $\eta_x[t,y]$ denotes the solution of eq.(15b) with initial condition $\eta_x[0,y] = y$, then

$$\langle h \rangle_{ad} := \int dy\, h(x,y)\rho_{ad}(y;x) = \lim_{T\to\infty} \frac{1}{T}\int_0^T dt\, h(x,\eta_x[t,y]) \tag{21}$$

holds for typical initial conditions and for observables $h(x,y)$.

On the other hand one may choose for $\widetilde{\rho}(y|x)$ the stationary conditional distribution

$$\rho_{cond}(y|x) := \frac{\rho_*(x,y)}{\bar{\rho}_*(x)} \quad , \tag{22}$$

where $\rho_*(x,y)$ and $\bar{\rho}_*(x)$ denote the corresponding stationary densities. Surprisingly, this choice leads to formally consistent results. To be able to distinguish the two ways of computing averages with respect to y, we define

$$\langle h \rangle_{cond} := \int dy\, h(x,y)\rho_{cond}(y|x) \quad . \tag{23}$$

We will first employ $\widetilde{\rho}(y|x) = \rho_{cond}(y|x)$ and averages according to Eq.(23). If we apply the standard projection operator technique to Eq.(17) we obtain finally a closed equation of motion for the density (19). In lowest order perturbation theory such an equation reduces to (cf.[18])

$$\frac{\partial \bar{\rho}_t}{\partial t} = -\frac{\partial}{\partial x}\langle f \rangle_{cond}\bar{\rho}_t(x) + \frac{\partial}{\partial x} D^{(2)}_{cond}(x)\bar{\rho}_*(x)\frac{\partial}{\partial x}\frac{\bar{\rho}_t(x)}{\bar{\rho}_*(x)} \quad . \tag{24}$$

[1] One may add a small diffusive like contribution to Eq.(18) to ensure smoothness and consider the limit of vanishing diffusion at the end. In addition, a generalization in terms of corresponding measures seems to be possible, but would increase the amount of notation considerably.

Here the diffusion coefficient is determined by the fluctuation of the vector field

$$\delta_{cond} f(x,y) = f(x,y) - \langle f \rangle_{cond} \tag{25}$$

in terms of the autocorrelation function as

$$D^{(2)}_{cond}(x) = \int_0^\infty dt \langle \delta_{cond} f(x, \eta_x[t,y]) \delta_{cond} f(x,y) \rangle_{cond} \quad . \tag{26}$$

As long as the dynamics of Eq.(15b) is (exponentially) mixing, the diffusion coefficient is well defined since the correlations decay sufficiently fast.

Thus, in lowest order perturbation expansion the deterministic equations of motion reduce to a stochastic system with the Fokker–Planck equation being given by Eq.(24). We rewrite the rightmost term of Eq.(24) in a form compatible with standard notation (e.g.[19]):

$$\frac{\partial \bar{\rho}_t}{\partial t} = -\frac{\partial}{\partial x} D^{(1)}_{cond}(x) \bar{\rho}_t(x) + \frac{\partial^2}{\partial x^2} D^{(2)}_{cond}(x) \bar{\rho}_t(x) \quad . \tag{27}$$

where

$$D^{(1)}_{cond} = \langle f \rangle_{cond} + D^{(2)}_{cond}(x) \frac{\partial}{\partial x} \ln \bar{\rho}_*(x) \tag{28}$$

denotes the drift coefficient. Hence, the drift term not only consists of the y-averaged slow vector field, but we find an additional term involving the slow invariant density $\bar{\rho}_*(x)$. The diffusion term now represents a white noise process with suitable amplitude, given by the autocorrelations of the fluctuations of the vector field, acting on the variable x. It is straightforward to rewrite this as a Langevin equation for x.

When we employ the adiabatic density $\rho_{ad}(y;x)$ in the projection, we encounter more difficulties in our perturbation expansion. As we were able to derive in [20], the resulting Fokker Planck equation which is formally correct up to first order in κ reads as follows:

$$\frac{\partial \bar{\rho}_t}{\partial t} = -\frac{\partial}{\partial x} D^{(1)}_{ad}(x) \bar{\rho}_t(x) + \frac{\partial^2}{\partial x^2} D^{(2)}_{ad}(x) \bar{\rho}_t(x) \tag{29}$$

with the drift term

$$D^{(1)}_{ad}(x) = \langle f \rangle_{ad} + \int_0^\infty dt' \left\langle f(x,y) \frac{\partial \delta_{ad} f(x, \eta_x[t'/\kappa, y])}{\partial x} \right\rangle_{ad} \tag{30}$$

Here, we used $\delta_{ad} f(x,y) = f(x,y) - \langle f \rangle_{ad}$. The diffusion matrix $D^{(2)}_{ad}(x)$ is determined as in Eq.(26) except for replacing $\langle \ldots \rangle_{cond}$ by $\langle \ldots \rangle_{ad}$ throughout.

Hence, the results of the two different projection schemes seemingly differ in their structure since eqs.(28) and (30) contain different drift renormalisations. As our example in the next section will reveal, the seemingly structurally similar parts, the averages of the vector field $\langle f \rangle_{cond}$ and $\langle f \rangle_{ad}$, can differ even more. But one has to keep in mind that for a meaningful comparison of the two Fokker–Planck equations (24) and (29), i.e. of the associated stochastic systems, only the comparison of the corresponding drift and diffusion coefficients is relevant. In fact, the disussion of the example in the next section reveals that the two drift coefficients $D^{(1)}_{cond}$ and $D^{(1)}_{ad}$ as well as the diffusion coefficients $D^{(2)}_{cond}$ and $D^{(2)}_{ad}$ coincide in the leading order of the perturbation expansion. Thus both expansion schemes yield consistent results, and such a feature is a first hint for the validity of our expansion schemes. As already mentioned the scheme using adiabatic densities may be simpler to evaluate since it boils down to the computation of plain time averages in the fast system (cf. eq.(21)).

4 Two-scale Ornstein-Uhlenbeck process

We consider the following linear system, where mixing of y is enforced by an explicit noise process right away:

$$\dot{x} = -\alpha x + y \tag{31a}$$

$$\dot{y} = -\frac{\widetilde{\beta}}{\kappa} y + \sqrt{\frac{\widetilde{\delta}}{\kappa}} \xi \,, \tag{31b}$$

where $\langle \xi(t)\xi(t')\rangle = 2\delta(t-t')$ is Gaussian white noise. Formally the noise is scaled by $\kappa^{-1/2}$ in order to ensure that the evolution operator (cf. eq.(18)), which is already an operator of Fokker–Planck type, shows a suitable decomposition in a fast and a slow part. For a moment, we ignore the κ-dependence and write $\beta = \widetilde{\beta}/\kappa$, $\delta = \widetilde{\delta}/\kappa$.

It is easy to verify that the invariant density of Eq.(31) is

$$\rho_*(x, y) = N \exp\left(-(m_{11}x^2 + 2m_{12}xy + m_{22}y^2)\right) \tag{32}$$

with

$$m_{11} = \frac{\alpha(\alpha+\beta)^2}{2\delta} \quad m_{12} = -\frac{\alpha(\alpha+\beta)}{2\delta} \quad m_{22} = \frac{\alpha+\beta}{2\delta}$$

with a proper normalization factor N. Every level line of this distribution is an ellipse centered at zero and with semi-axes which, since $m_{12} \neq 0$, are not aligned with the axes of the coordinate system.

The invariant density of the x-variable is obtained by integration over y and reads:

$$\bar{\rho}_*(x) = \frac{\sqrt{(\alpha+\beta)\alpha\beta}}{\sqrt{2\pi\delta}} \exp\left(-\frac{\alpha\beta(\alpha+\beta)}{2\delta}x^2\right) \,, \tag{33}$$

so that we obtain the conditional density

$$\rho_{cond}(y|x) := \frac{\rho_*(x,y)}{\bar{\rho}_*(x)} = \frac{\sqrt{\alpha+\beta}}{\sqrt{2\pi\delta}} \exp\left(-\frac{\alpha+\beta}{2\delta}(y-\alpha x)^2\right) \,. \tag{34}$$

The conditioned mean value of y depends on x, in contrast to intuition when considering the skew nature of Eq.(31).

We can now compute $\langle f\rangle_{cond}$, for which we find:

$$\langle f\rangle_{cond} := \int \rho_{cond}(y|x) f(x,y)\, dy = -\alpha x + \alpha x \equiv 0 \,. \tag{35}$$

Hence, in this example the effective drift $D^{(2)}_{cond}(x)\frac{\partial}{\partial x}\ln\bar{\rho}_*(x)$ is the only drift in the Fokker Planck equation (27). Glancing back at Eq.(24), shows that this in fact has to be the case in order for $\bar{\rho}_*(x)$ to be the invariant density for which the right hand sides of Eqs.(24) and (27) have to vanish.

For the computation of the effective drift we need the diffusion term $D^{(2)}_{cond}(x)$, which we obtain from Eq.(26), the integral over the fluctuations of the vector field. Since $\langle f\rangle_{cond} = 0$, we have $\delta_{cond} f(x,y) \equiv f(x,y) = -\alpha x + y$. The integration over paths $\eta_x[t,y]$ is in this example the mean with respect to the white noise ξ, which leads to exponentially decaying correlations $e^{-\beta t}$. Since the variance of the conditional average of y is given by $\langle y^2\rangle_{cond} = \delta/(\beta+\alpha)$, one finds:

$$D^{(2)}_{cond}(x) = \frac{\delta}{\beta(\alpha+\beta)} \tag{36}$$

Hence, Eq.(24) for this example reads

$$\frac{\partial \bar{\rho}_t}{\partial t} = \frac{\partial}{\partial x}\alpha x \bar{\rho}_t(x) + \frac{\delta}{\beta(\alpha+\beta)}\frac{\partial^2}{\partial x^2}\bar{\rho}_t(x) \quad . \tag{37}$$

The invariant density Eq.(33) is the exact stationary solution of this equation for any values of κ. This is not surprising, since the system Eq.(31) fulfils already the basic approximations of our formalism: Gaussian distributions and exclusively linear correlations.

From a practical point of view our result may be unsatisfactory: In order to obtain Eq.(37) form Eq.(24), we need the complete solution of the full system. Moreover, the drift term of Eq.(24) vanishes identically. Since by construction the rightmost term of Eq.(24) vanishes for every system if $\bar{\rho}_t(x) = \bar{\rho}_*(x)$, the vector field $\langle f(x)\rangle_{cond}\rho_*(x)$ has to be divergence free. Hence, the drift term which counter-balances the diffusion such that the invariant density is in fact invariant is the term $D^{(2)}_{cond}(x)\frac{\partial}{\partial x}\ln \bar{\rho}_*(x)$, which is still dominant when the system is close to its stationary solution. In summary, in order to exploit Eq.(24), one always needs the solution of the full system, otherwise $\bar{\rho}_*(x)$ is unknown and so is the dominant part of the equation.

The way out is to return to the other natural choice of the relevant density with $\widetilde{\rho}(y|x) = \rho_{ad}(y;x)$. For the Ornstein-Uhlenbeck system Eq.(31) it is easy to verify that

$$\langle f(x,y)\rangle_{cond} + D^{(2)}_{cond}(x)\frac{\partial}{\partial x}\ln\bar{\rho}_*(x) = \langle f(x,y)\rangle_{ad} = -\alpha x \; . \tag{38}$$

Thus the drift terms of Eqs. (24) and (29) coincide in the order κ. Computing the diffusion coefficient according to Eq.(26) with $\langle . \rangle_{ad}$ yields $D^{(2)}_{ad}(x) = \delta/\beta^2$. We now re-introduce the parameter κ and find

$$D^{(2)}_{ad}(x) = D^{(2)}_{cond}(x)(1 + O(\kappa)) \; . \tag{39}$$

Hence, the example shows that the Fokker Planck equation (29) is the more useful one, since it can be specified for a given problem without beforehand knowing its invariant distribution. However, the Fokker Planck equations (24) and (29) coincide in the lowest nontrivial order of κ.

In order to test the accuracy of our analytic expressions numerically, we return to the Ornstein-Uhlenbeck-like system Eq.(14) of Sec.2.3. The specification of the terms involved in Eq.(29) is straightforward. Due to the fact that this is a skew system, we need a single long run of the Lorenz system in order to perform all necessary averages according to $\rho_{ad}(y;x)$. Its numerically computed auto-correlation yields the diffusion constant $D^{(2)}_{ad}$, whereas $\langle f\rangle_{ad} = -\alpha x$. In

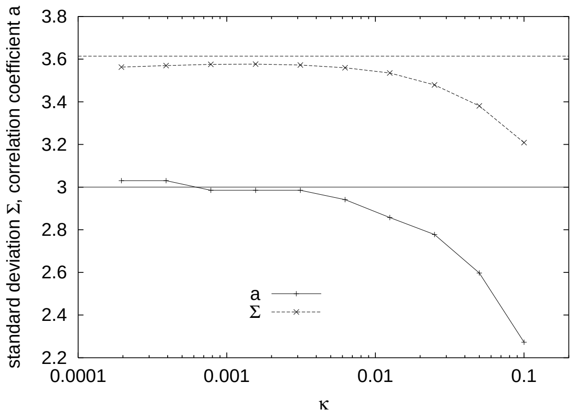

Figure 5: Numerically obtained rescaled standard deviation $\Sigma = \sqrt{\langle x^2 \rangle / \kappa}$ and correlation coefficient a, $\langle x(t)x(t-\Delta)\rangle \propto \exp(-a\Delta)$, of the observable x of the system Eq.(14) for $\alpha = 3$ as a function of the time scale separation parameter κ, together with the theoretically predicted asymptotic values (horizontal lines) obtained with the averages according to Eq.(21). For small κ, numerical values are affected by finite sample errors.

Fig.5 we show, for different values of κ, the numerically obtained standard deviation and correlation time of the observable x of the system Eq.(14). They are to be compared to $\sqrt{D_{ad}^{(2)}/\alpha}$ and to α. We see that for $\kappa < 10^{-3}$ the agreement is indeed very good.

A slightly more complex example was presented in Ref.[18], where the stochastic resonance scenario was mimicked by a driving Lorenz system instead of white noise. Also in this case, making use of Eq.(29) the agreement was excellent for $\kappa < 0.01$.

5 Summary

A length scale dependent dimension and entropy analysis of systems with a fast chaotic component revealed that on large scales, fast chaos is indistinguishable from noise. Even more, in situations where the slow dynamics is sufficiently simple, this analysis shows a quantitative agreement between the full deterministic process and a suitable stochastic process driven by white noise, which includes a verification that long range memory effects do not exist, when only large scale properties are of interest. These results motivate the work presented in Section 3. There, we have presented two Fokker Planck equations, Eq.(24) and Eq.(29), for the slow variables, where the diffusion term models fast chaotic motion. Both results are formally exact of order κ in perturbation theory and therefore have to agree in this order, even though their structure is very different. This shows the possibility to replace fast chaos by noise and gives the optimal evolution equations of the slow variables. Eq.(29) is more useful in practical applications, since it contains only terms which can be computed when the fast system alone is analyzed, for frozen values of the slow variables. The examples that have been treated here in a quantitative way have been skew systems. If the slow variable itself couples to the fast dynamics then we will in general obtain effective diffusion coefficients which are not constant but depend on the slow variables themselves.

If the fast variables are (quasi-)periodic, the theoretical derivation of our results does apply only after a careful treatment of the correlations in the diffusion terms. Nonetheless, there is some evidence that in this case the diffusion vanishes and the result reduces to the result from the averaging principle. In agreement with numerical experiments, our treatment then predicts effects of the fast (quasi-)periodic variables of order κ.

As a side remark, this theory supplies an answer to the issue of whether it is better to run a car fast or slowly over a rough road paved with cobblestones. If we assume the vertical position of the center of mass of the car to be the slow variable, the amplitude of vibrations felt from the cobblestones will be the smaller the faster the car moves. Thus, the influence of the fast degrees of freedom on the slow ones becomes smaller as the speed of the fast variables increases.

References

[1] P. Gaspard, M. E. Briggs, M. K. Francis, J. V. Sengers, R.W. Gammon, J.R. Dorfman and R. V. Calabrese, *Experimental evidence for microscopic chaos*, Nature **394**, 865 (1998).
P. Gaspard, M.E. Briggs, M.K. Francis, J.V. Sengers, R.W. Gammon, J.R. Dorfman and R. V. Calabrese, *Gaspard et al. reply*, Nature **401**, 876 (1999).

[2] C. Dettman, E. Cohen, and H. van Beijeren, *Statistical mechanics: microscopic chaos from brownian motion?*, Nature **401**, 875 (1999).
P. Grassberger and T. Schreiber, *Statistical mechanics: microscopic chaos from brownian motion?*, Nature **401**, 875 (1999).
M. Cencini, M. Falcioni, H. Kantz, E. Olbrich, and A. Vulpiani, *Chaos or Noise – Difficulties of a Distinction*, Phys. Rev. **E 62**, 427-437 (2000).

[3] Y. Kifer, *Averaging in dynamical systems and large deviations*, Invent. Math. **110**, 337–370 (1992).

[4] C. Beck and G. Roepstorff, *From dynamical systems to the Langevin equation*, Physica A 145 (1987) 1–14,
C. Beck. *Dynamical systems of Langevin type*, Physica A 233 (1996) 419–440.

[5] C. Beck, *Brownian motion from deterministic systems*, Physica 169A, 324 (1990).

[6] C. Beck, *From the Perron–Frobenious equation to the Fokker–Planck equation*, J. Stat. Phys. **79**, 875 (1995).

[7] A. Hilgers and C. Beck, *Approach to Gaussian stochastic behaviour for systems driven by deterministic forces*, Phys. Rev. **E 60**, 5385 (1999).

[8] F. Takens, *Detecting strange attractors in turbulence*, in "Dynamical Systems and Turbulence", eds. D.A. Rand, L.-S. Young, Lecture Notes in Mathematics, Vol.898, p.366-381, Springer, Berlin 1981.

T. Sauer, J.A. Yorke and M. Casdagli, *Embedology*, J. Stat. Phys. **65**, 579-616 (1991).

[9] P. Grassberger and I. Procaccia, *Characterization of strange attractors*, Phys. Rev. Lett. **50** (1983) 346-349.

[10] E. Olbrich and H. Kantz, *Inferring chaotic dynamics from time series: On which length scale determinism becomes visible*, Phys. Lett. **A 232** (1997) 63-69.

[11] H. Kantz and E. Olbrich, *The transition form deterministic chaos to a stochastic process*, Physica A 153 (1998) 105-117.

[12] H. Kantz and E. Olbrich, *Coarse grained dynamical entropies - investigation of high-entropic dynamical systems*, Physica **A 280**, 34-48 (2000).

[13] P. Gaspard and X.-J. Wang, *Noise, chaos, and (ϵ, τ)-entropy per unit time*, Physics Reports 235 (1993) 291.

[14] P. Grassberger, *Finite sample corrections to entropy and dimension estimates*, Phys. Lett. A 128 (1988) 369-373.

[15] P. Grassberger and I. Procaccia, *Estimation of the Kolmogorov entropy from a chaotic signal.*, Phys. Rev. A 28 (1983) 2591.

[16] F. Mitschke, M. Möller and W. Lange, *Measuring filtered chaotic signals*, Phys. Rev. **A 37** (1988) 4518-4521.
R. Badii, G. Broggi, B. Derighetti, M. Ravani, S. Ciliberto, A. Politi, and M.A. Rubbio, *Dimension increase in filtered chaotic signals*, Phys. Rev. Lett. **60** (1988) 979-982.

[17] R. Zwanzig, *Ensemble method in the theory of irreversibility*, J. Chem. Phys. **33**, 1338–1341 (1960).

[18] W. Just, H. Kantz, C. Rödenbeck, M. Helm, *Stochastic Modelling: Replacing fast degrees of freedom by stochastic processes*, J. Phys. A: Math. Gen. **34**, 3199 (2001).

[19] H. Risken, *The Fokker–Planck equation*, (Springer, Berlin, 1989).

[20] W. Just, K. Gelfert, N. Baba, A. Riegert, H. Kantz *Elimination of fast chaotic degrees of freedom: On the accuracy of the Born approximation* J. Stat. Phys. **112**, 277 (2003).