

Elsevier Editorial System(tm) for Physica D: Nonlinear Phenomena

Manuscript Draft

Manuscript Number:

Title: Non-uniform decay of predictability and return of skill in stochastic oscillatory models

Article Type: Full Length Article for Regular Issue

Section/Category: Regular Issue

Keywords: Stochastic processes; Markov processes; Relative entropy

Corresponding Author: Dr. Serguei Lapin,

Corresponding Author's Institution: University of Houston

First Author: Serguei Lapin

Order of Authors: Serguei Lapin; Kresimir Josic; Ilya Timofeyev; Nofil Barlas

Manuscript Region of Origin:

Abstract:

We examine the dynamical mechanisms that lead to the loss of predictability in low dimensional stochastic models that exhibit three main types of oscillatory behavior: damped, self-sustained, and heteroclinic. We show that the information that an initial ensemble provides about the state of the system decays non-uniformly in time. Long intervals during which the forecast provided by the ensemble does not lose any of its power are typical in all three cases. Moreover, the information that the forecast provides about the individual variables in the model may increase, despite the fact that information about the entire system always decreases. We analyze the fully solvable case of the linear oscillator, and use it to provide a general heuristic explanation for the phenomenon. We also show that during the intervals during which the forecast loses little of its power, there is a flow of information between the marginal and conditional distributions.

Professor C.K.R.T. Jones, Editor-in-Chief
Department of Mathematics, Phillips Hall, CB #3250,
University of North Carolina, Chapel Hill, NC 27599-3250

Dear Professor Jones,

Please consider the manuscript "Non-uniform decay of predictability and return of skill in stochastic oscillatory models" for publication in the *Physica D* journal. All correspondence regarding this manuscript should be addressed to

Serguei Lapin
University of Houston
Department of Mathematics
Houston, TX 77204-3008

email: slapin@math.uh.edu
Phone : 713-743-0246
Fax : 713-743-3505

Sincerely,
Serguei Lapin

Non-uniform decay of predictability and return of skill in stochastic oscillatory models

Nofil Barlas^a, Krešimir Josić^a, Serguei Lapin^{a,*},
Ilya Timofeyev^a

^a*Department of Mathematics, University of Houston, Houston TX 77204-3008,
USA*

Abstract

We examine the dynamical mechanisms that lead to the loss of predictability in low dimensional stochastic models that exhibit three main types of oscillatory behavior: damped, self-sustained, and heteroclinic. We show that the information that an initial ensemble provides about the state of the system decays non-uniformly in time. Long intervals during which the forecast provided by the ensemble does not lose any of its power are typical in all three cases. Moreover, the information that the forecast provides about the individual variables in the model may increase, despite the fact that information about the entire system always decreases. We analyze the fully solvable case of the linear oscillator, and use it to provide a general heuristic explanation for the phenomenon. We also show that during the intervals during which the forecast loses little of its power, there is a flow of information between the marginal and conditional distributions.

Key words: Stochastic processes, Markov processes, Relative entropy

PACS: 02.50.Ey, 02.50.Ga, 05.10.Gg

1 Introduction

Due to atmospheric uncertainty, statistical predictions involving Monte-Carlo simulations for an ensemble of trajectories are frequently used in weather and climate forecasting. The evolution of such ensembles, and, in particular, their

* Corresponding author.

Email address: `slapin@math.uh.edu` (Serguei Lapin).

spread can be used to quantify the reliability of predictions. Several measures that quantify “potential predictability” of dynamical systems in this sense have been utilized in atmosphere/ocean science. These include the Root Mean Squared Error [5], Anomaly Correlation Coefficient [4] and Potential Prediction Utility [2,6]. Other measures of predictability, inspired by dynamical systems theory, include Lyapunov exponents [3,20] and various notions of entropy [15,12].

A measure that appears to be particularly well suited to quantify the predictability of a stochastic dynamical system is *relative entropy* [14] (also called *Kullback-Leibler divergence*). The relative entropy can be interpreted as the amount of information provided by a particular prediction [7]. Unlike some other measures of utility, it reflects differences in all moments, including the mean and variance, of two distributions. In addition, relative entropy satisfies several important mathematical properties which make it a relatively unique measure of predictability.

Typically, the predictability properties of a given system are characterized by the behavior of the relative entropy averaged over the equilibrium distribution of the system obtained by Monte-Carlo simulations with an ensemble of ensembles. Each individual ensemble in the simulation describes the decay of the utility of prediction for an initial state. The mean of each initial state is chosen at random from the equilibrium distribution, and their variances reflect uncertainties due to imperfect measurements.

The overall predictability of the model can then be characterized by averaging the relative entropy over all initial states thus generated.

The goal of this article is to show that even for Markov systems the mechanisms that leads to the loss of predictability have surprising and counterintuitive aspects. While utility of ensemble predictions typically decays exponentially in time (see [14,19], for example), it can behave very differently for each individual ensemble. This implies that the predictability of a model can be considered as a functional dependent on the initial state used for the prediction, and significant information may be lost by averaging over all initial states [16].

In particular, we show that for a certain class of models and initial states there are large intervals of time during which the utility of the prediction remains nearly constant. Therefore, the predictability of these particular forecasts is very different from the exponentially decaying averaged predictability of the system. Oscillatory transport of the initial ensemble towards and away from peaks of a spatially nonuniform equilibrium distribution is the primary mechanism behind this behavior. During the times of extended predictability the ensemble mean resides in areas where the mass of the equilibrium distribution

is small, and predictability is lost during the transient returns of the ensemble mean.

Relative entropy decays monotonically in time for Markov processes, since information that is lost cannot be regained [11]. Therefore the utility of a prediction cannot increase in time for Markov models, and in the language of atmosphere/ocean science, there is no *return of skill* [1]. However, this result holds only for the full density of the system being modeled, and does not apply to the marginal densities. Indeed, we show that the marginal relative entropies of *all* variables in the model may increase simultaneously in time, while the relative entropy of their joint distribution decays. Therefore, while information about the total state of the system is necessarily lost over time, information about its attributes may be regained. Surprisingly, information about all attributes may be regained simultaneously. We illustrate that the mechanism leading to the return of skill and the near constancy of relative entropy over certain time intervals are closely related.

As the marginal relative entropies may change either in concert or in opposite directions, it may be difficult to make sense of the flow of information between the variables defining a Markov model. It is sometimes more natural to consider the flow of information between the different conditional and marginal distributions. We also show that the phenomena described above can be understood in terms of such a flow of information.

The rest of the paper is organized as follows. In Section 2 we present general properties of the relative entropy functional and discuss its relevance to other measures of predictability. In Section 3 we consider the stochastically perturbed linear oscillator. This example is solvable analytically, and is utilized here to explain the mechanism behind the non-uniform decay of relative entropy and return of skill in noisy oscillatory systems. We consider the stochastic perturbation of a nonlinear oscillator (obtained from the normal form of a Hopf bifurcation) and homoclinic cycle (obtained from the Duffing equation) in Sections 4.1 and 4.2, respectively. Conclusions are presented in Section 5.

2 Relative entropy for SDEs

Consider a stochastic dynamical system model of climate, which we assume to be Markov. Let $q(\vec{x})$ be the invariant (climatological) distribution, and let $p(\vec{x}, t)$ be the probability density corresponding to the ensemble of realizations predicting the state of the system at time t . The relative entropy, or *Kullback-*

Leibler divergence between these two distributions is defined as

$$R(p(\vec{x}, t), q(\vec{x})) = R(t) = \int p(\vec{x}, t) \log \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) d\vec{x}. \quad (1)$$

This can be thought of as a measure of “distance” between the distributions $p(\vec{x}, t)$ and $q(\vec{x})$ ¹. More precisely $R(t)$ corresponds to the amount of information that the distribution $p(\vec{x}, t)$ provides about the state of the system in excess of that given by the equilibrium distribution $q(\vec{x})$. It is therefore natural to interpret $R(t)$ as a measure of the utility of the prediction provided by an ensemble of particular realizations.

Relative entropy reflects differences in the mean and variance, as well as other moments of two distributions: an increase in the utility of a prediction may be due to the narrow spread of the ensemble (reflected in a difference between the variances of p and q), or the fact that this ensemble indicates a large departure from normal conditions (reflected in a difference between the means of p and q).

Relative entropy also satisfies three important mathematical properties:

- (1) it is invariant under well behaved non-linear transformations of state variables,
- (2) it is non-negative and,
- (3) it declines monotonically in time for Markov processes.

The fact that relative entropy decreases in time can be naturally interpreted as a decline in the utility of a prediction, or *skill*.

In this section we recall the relative entropy between multivariate Gaussian distributions, and provide an expression that will be used to analyze its decay in the model systems we consider subsequently.

2.1 *Relative entropy for Gaussian distributions*

Suppose that $q(\vec{x})$ and $p(\vec{x}, t)$ are n -dimensional multivariate Gaussian distributions with means μ_p and μ_q and correlation matrices σ_p and σ_q respectively. In this case a closed form for the relative entropy can be obtained [14]

$$R = \frac{1}{2} \left(\log \left(\frac{\det(\sigma_q^2)}{\det(\sigma_p^2)} \right) + \text{Tr}(\sigma_p^2(\sigma_q^2)^{-1}) + \underbrace{(\mu_p)^T (\sigma_q^2)^{-1} (\mu_p) - n}_{\text{signal term}} \right). \quad (2)$$

¹ Care needs to be taken in interpreting R as a distance since $R(p, q)$ does not in general equal $R(q, p)$.

For the Gaussian case the relative entropy can be naturally decomposed into two parts: the signal (the third term in the sum) and the dispersion (the remaining terms). The signal component accounts for the difference in the means of the two distributions, $q(\vec{x})$ and $p(\vec{x}, t)$, while the dispersion reflects the difference in their variances. Therefore, the signal and dispersion terms are analogous to the Anomaly Correlation Coefficient and Root Mean Squared Error, respectively.

2.2 Relative entropy as a Lyapunov functional for the Fokker-Planck equation

We next consider the Fokker-Planck equation

$$\frac{\partial p(\vec{x}, t)}{\partial t} = - \sum_i \frac{\partial}{\partial x_i} [A_i(\vec{x})p(\vec{x}, t)] + \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial x_i \partial x_j} [B_{ij}(\vec{x})p(\vec{x}, t)], \quad (3)$$

corresponding to a model stochastic differential equation [11]. Here $A(\vec{x}, t)$ is the drift vector, $B(\vec{x}, t)$ is the diffusion matrix, and the distribution $p(\vec{x}, 0)$ provides the initial data for the Fokker-Planck equation. We assume that the system under consideration has a unique equilibrium solution $q(\vec{x})$.

The relative entropy $R(p(\vec{x}, t), q(\vec{x})) = R(t)$ is dependent on the initial data $p(\vec{x}, 0)$, corresponding to the distribution of initial conditions of an ensemble in a Monte-Carlo simulation. We will show in the next section that the decay of relative entropy can vary markedly for different choices of initial ensembles.

In this case the relative entropy, defined in equation (1), is a Lyapunov functional for the Fokker-Planck equation [11]. Indeed, a direct calculation shows that the relative entropy decays monotonically in time. Using the definition of relative entropy we obtain

$$\frac{dR}{dt} = \int d\vec{x} \left[\frac{\partial p(\vec{x}, t)}{\partial t} (\log p(\vec{x}, t) + 1 - \log q(\vec{x})) - \frac{\partial q(\vec{x})}{\partial t} \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) \right]. \quad (4)$$

Let us assume that $q(\vec{x})$ is a stationary solution of (3), and that $q(\vec{x})$ is non-zero everywhere, except at infinity, where it and its first derivatives vanish. The contributions to dR/dt stemming from the drift ($\left(\frac{dR}{dt}\right)_{\text{drift}}$) and diffusion ($\left(\frac{dR}{dt}\right)_{\text{diff}}$) terms in the Fokker-Planck equation can also be obtained by the same calculation:

$$\left(\frac{dR}{dt}\right)_{\text{drift}} = \sum_i \int d\vec{x} \frac{\partial}{\partial x_i} \left[-A_i p(\vec{x}, t) \log \left(\frac{p(\vec{x}, t)}{q(\vec{x})} \right) \right], \quad (5)$$

$$\left(\frac{dR}{dt}\right)_{\text{diff}} = -\frac{1}{2} \sum_{i,j} \int d\vec{x} p(\vec{x}, t) B_{ij} \left[\frac{\partial}{\partial x_i} \log \frac{p(\vec{x}, t)}{q(\vec{x})} \right] \left[\frac{\partial}{\partial x_j} \log \frac{p(\vec{x}, t)}{q(\vec{x})} \right]. \quad (6)$$

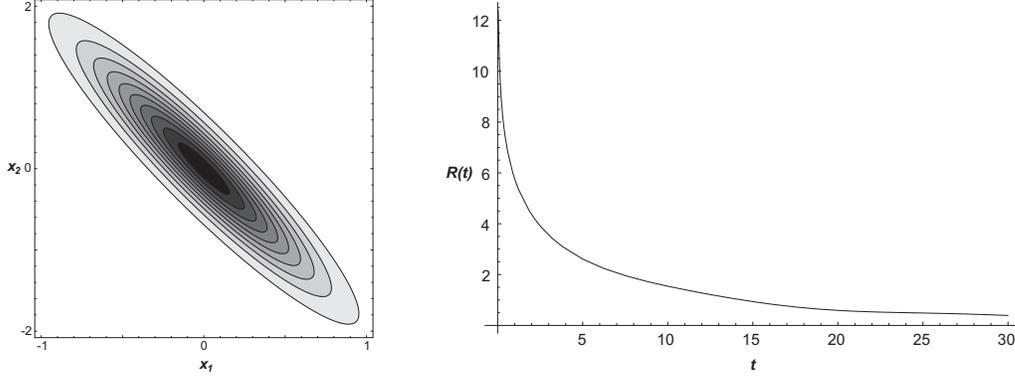


Fig. 1. Left: The equilibrium distribution $q(\vec{x})$ for the linear oscillator (8). Right: Average relative entropy in the case of the stochastic linear oscillator. Parameters: $\alpha = 0.4573, \beta = 0.2435, \gamma = \delta = -1.0852$, and $\varepsilon = 0.1$.

Under the given assumptions on $q(\vec{x})$ it can be shown that

$$\left(\frac{dR}{dt}\right)_{\text{drift}} = 0 \quad \text{and} \quad \left(\frac{dR}{dt}\right)_{\text{diff}} \leq 0. \quad (7)$$

It follows that the decrease in relative entropy is due only to diffusion terms. This is not surprising from an information theoretic viewpoint since it is the diffusion terms that correspond to the stochastic components of the equation that lead to information loss. While the most immediate effect of these terms is to increase the spread in the ensemble forecast, they interact in a nontrivial way with the drift terms to determine the rate of this decrease. We illustrate this point in the next section.

3 The stochastic linear oscillator

We next consider the fully solvable example of the damped stochastic linear oscillator which has been used as a prototype model of the El Niño/Southern Oscillation [13]. Even in this simple model relative entropy decays non-monotonically, and the marginal relative entropies can oscillate. The results of this section were obtained using analytical expressions for the relative entropy which we do not report in full due to their complexity.

The model is given by the following two-dimensional stochastic differential equation:

$$\begin{aligned} dx_1 &= \alpha x_1 dt + \beta x_2 dt, \\ dx_2 &= \gamma x_1 dt + \delta x_2 dt + \varepsilon dW, \end{aligned} \quad (8)$$

where W is a Wiener process, ε is the noise level, and the remaining parameters are chosen so that with $\varepsilon = 0$ the system exhibits damped oscillations. Figure 1 (Left) shows the contour plot of the equilibrium distribution in one specific case. Although we use the same parameters for all subsequent simulations in this section, we show below that our observations hold under very general conditions.

The Fokker-Planck equation describing the evolution of an initial density is given by equation (3) with

$$A = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 0 \\ 0 & \varepsilon \end{pmatrix}.$$

This equation can be solved analytically assuming a deterministic initial condition $p(\vec{x}, 0) = \delta_{\vec{x}^0}(\vec{x})$ where $\vec{x}^0 = (x_1^0, x_2^0)$ [11]. The solution at time t is a Gaussian with mean

$$\begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = e^{At} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix}, \quad (9)$$

and covariance matrix

$$\begin{bmatrix} \langle x_1, x_1 \rangle & \langle x_1, x_2 \rangle \\ \langle x_2, x_1 \rangle & \langle x_2, x_2 \rangle \end{bmatrix} = \int_0^t dt' e^{A(t-t')} \begin{pmatrix} 0 & 0 \\ 0 & \varepsilon^2 \end{pmatrix} e^{A^T(t-t')}. \quad (10)$$

Since the equilibrium distribution $q(\vec{x})$ can also be computed explicitly, equation (2) can be used to obtain the relative entropy analytically in this case.

3.1 The decay of relative entropy

The overall predictability of a model is determined by the average rate of decay of the relative entropy, where the average is taken over all deterministic initial conditions weighted by the stationary probability density. In general, this requires Monte-Carlo simulations with a collection of ensembles [14]. Each ensemble consists of a number of initial conditions distributed according to a narrow Gaussian or a delta function centered at a point sampled randomly from the equilibrium distribution. As noted earlier, averaging over the equilibrium distribution provides a measure of the average predictability of a given model. The average relative entropy for the linear oscillator (8) averaged over initial ensembles corresponding to densities $\delta_{\vec{x}^0}(\vec{x})$ is shown in Figure 1 (Right).

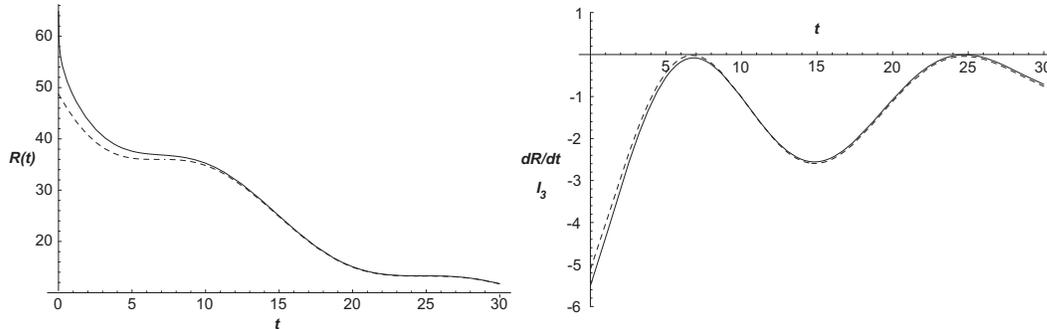


Fig. 2. Left: Relative entropy for a particular initial condition $(x_1^0, x_2^0) = (1, 1)$ (solid), and the contribution to the relative entropy due to the signal term in equation (2) (dashed). Right: Behavior of $(\frac{dR}{dt})_{\text{diff}}$ (solid) and I_3 (dashed) in time for an initial condition $(x_1^0, x_2^0) = (1, 1)$. Parameters as in Figure 1.

One of the essential properties of the relative entropy is its monotonic decay in time reflecting the loss of information due to the stochastic forcing as each ensemble of initial conditions is propagated forward in time. Simple estimates show that initially diffusive terms dominate, and relative entropy has a logarithmic singularity at the origin. Therefore, there is a boundary layer around $t = 0$ during which the relative entropy decays as $-\log(t)$. This is followed by a long interval during which relative entropy decays exponentially, as shown in Figure 1 (Right).

However, the situation can be different for any particular initial ensemble whose mean is sufficiently far from the mean of the equilibrium distribution. Rather than decreasing exponentially, there are intervals during which the relative entropy remains nearly constant. This is illustrated in Figure 2 (Left), using the relative entropy for an ensemble of trajectories with the initial condition $(x_1^0, x_2^0) = (1, 1)$ so that $p(\vec{x}, 0) = \delta_{(1,1)}(\vec{x})$. The ensemble is generated utilizing independent realizations of the Wiener process.

In particular, during the time intervals $[5 \dots 10]$, and $[22 \dots 27]$ the rate of relative entropy decay is nearly zero. During these intervals, the forecast skill remains nearly constant, and our confidence in the prediction based on an ensemble of possible projections does not decrease.

Relative entropy decays in a similar manner for any initial density whose mean differs sufficiently from that of the equilibrium distribution. However, the plateaus in relative entropy do not occur at the same time, and we show below that their position in time depends on the phase of the mean of the initial ensemble. Therefore, the average over different initial ensembles provides a somewhat misleading picture: compared to the rate of decay of relative entropy for a *particular ensemble* the rate corresponding to the average is much larger during the plateaus, or much smaller between the plateaus.

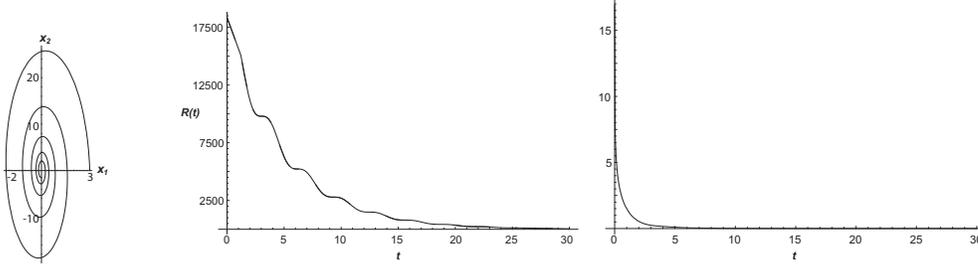


Fig. 3. Case with matrix A defined by (11) for an initial condition $(x_1^0, x_2^0) = (3, 0)$. Left: Trajectory of the mean of the distribution. Middle: Relative entropy behavior. Right: Difference between the full relative entropy and the signal term.

As we will see in the next section, this effect is due to the fact that after a transient the value of relative entropy is mainly determined by the location of the mean of an ensemble, and the means corresponding to different initial data can oscillate in- or out of phase.

3.2 Analysis of the rate of decay of $R(t)$

We can explain the nonuniform decay of the relative entropy by considering equation (2). After an initial quick increase in the variance of the transient distribution $p(\vec{x}, t)$ resulting in a logarithmic singularity of $R(t)$, and $\left(\frac{dR}{dt}\right)_{\text{diff}} \sim -1/t$, further changes in the variance occur at a timescale slow compared to that of the oscillations (see the top row of Figure 4). Therefore the signal term in expression (2) can be expected to dominate. That this is indeed the case is illustrated in Figure 2 (Left).

We illustrate the analysis in the case

$$A = \begin{pmatrix} \frac{1}{k} & \frac{1}{b} \\ -b & \frac{1}{k} \end{pmatrix}, \quad (11)$$

B as above, and $x_2^0 = 0$. The general case is very similar, but more tedious. The mean of the solution of equation (3) with this initial data is, by (9),

$$\bar{x}_1(t) = x_1^0 e^{-t/k} \cos t, \quad \bar{x}_2(t) = b x_1^0 e^{-t/k} \sin t,$$

so that k is a damping coefficient, and b determines how the solutions are stretched in the y direction. For fixed values of k and b of the same order, $k \gg 1$, and $b \gg 1$, small noise and $x_1(0)$ sufficiently large, the signal term dominates all other terms in equation (2) (an example with $k = 10$, $b = 10$, $\varepsilon = 0.1$, and $x_1^0 = 3$ is shown in Figure 3).

The signal term of the relative entropy has the form

$$R_{\text{signal}}(t) = e^{-\frac{2t}{k}} \frac{2b^2 (x_1^0)^2 (1 + k^2 + \cos(2t) + k \sin(2t))}{\varepsilon k^3}.$$

In the parameter regime of interest the term proportional to $e^{-\frac{2t}{k}} \sin(2t)$ determines the non-uniformities in the decay of relative entropy.

The plateaus in relative entropy therefore occur at the times in which $\sin(2t)$ is increasing. Those intervals correspond to the time during which $|x_2(t)|$ increases from 0 to b , and the mean of the transient distribution moves away from the mean of the stationary distribution (see the left panel of Figure 3). Similarly, information is lost rapidly during the times at which $|x_2(t)|$ decreases from b to 0.

Intuitively, this is a consequence of the fact that information is gained as the means of the transient and equilibrium distribution move apart, and this gain balances the loss of information due to the increase in the variance of the transient distribution through diffusion. As the means of the two distributions move together, *i.e.* during the times during which $\sin(2t)$ is decreasing, information is lost due to changes in both mean and variance. These intervals correspond to the rapid loss of information following the plateaus. Similar estimates and arguments apply to the example of the previous section. The interval between the first two panels in the top row of Figure 4 correspond to a plateau, while the interval between the last two panels corresponds to the sharp drop following a plateau.

According to the discussion in Section 2, the decay of relative entropy is entirely due to diffusion. Therefore the fact that the decay of relative entropy is dominated by the behavior of the mean whose evolution is completely determined by the drift appears somewhat counterintuitive. We next explain this apparent contradiction.

For the stochastic linear oscillator in (8) the diffusion part of the corresponding Fokker-Planck equation (6) reduces to

$$\left(\frac{dR}{dt} \right)_{\text{diff}} = -\frac{\varepsilon}{2} \int d\vec{x} p(\vec{x}, t) \left[\frac{\partial}{\partial y} (\log(p(\vec{x}, t)/q(\vec{x}))) \right]^2. \quad (12)$$

This can be evaluated using a straightforward, but lengthy calculation. The

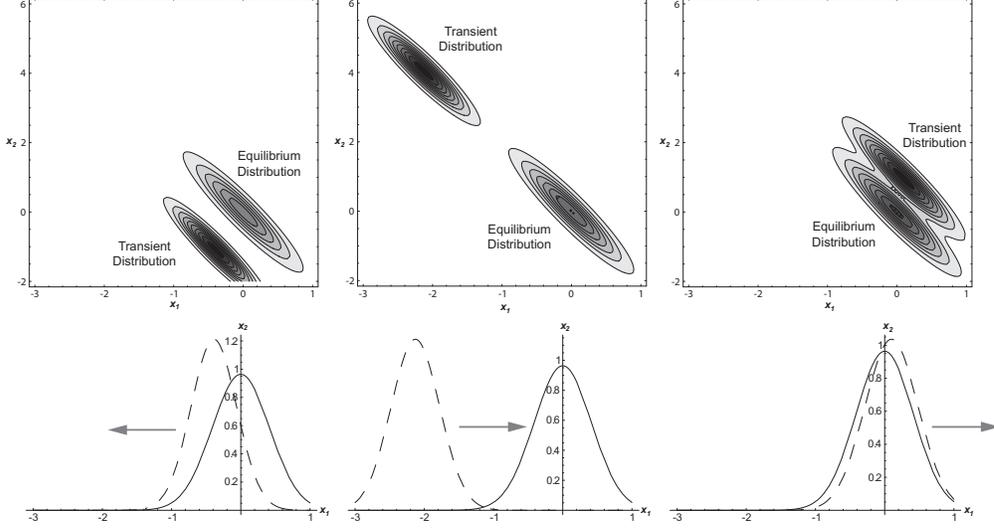


Fig. 4. Probability density functions of the transient distribution (dashed line) and the equilibrium distribution (solid line) for times $t=17.5$ (left), $t=25$ (center) and $t=35$ (right). The arrows indicate the direction in which the transient distribution moves at the time of the snapshot.

expression (12) can be rewritten as

$$\begin{aligned}
 \left(\frac{dR}{dt}\right)_{\text{diff}} = & -\frac{\varepsilon}{2} \left[\underbrace{\int d\vec{x} p(\vec{x}, t) \left(\frac{\partial \log p(\vec{x}, t)}{\partial y}\right)^2}_{I_1} - 2 \underbrace{\int d\vec{x} p(\vec{x}, t) \frac{\partial \log p(\vec{x}, t)}{\partial y} \frac{\partial \log q(\vec{x})}{\partial y}}_{I_2} \right. \\
 & \left. + \underbrace{\int d\vec{x} p(\vec{x}, t) \left(\frac{\partial \log q(\vec{x})}{\partial y}\right)^2}_{I_3} \right] \equiv I_1 + I_2 + I_3.
 \end{aligned} \tag{13}$$

A direct computation shows that only the third integral (I_3) depends on the mean of the transient distribution μ_p , while the first two integrals I_1 and I_2 depend only on the variances σ_p^2 and σ_q^2 . Moreover, this integral is exactly the time-derivative of the signal part of the relative entropy, *i.e.* $(I_3) = \frac{d}{dt}[(\mu_p)^T (\sigma_q^2)^{-1} (\mu_p)]$. Thus, since the relative entropy is dominated by the signal term, the behavior of $\left(\frac{dR}{dt}\right)_{\text{diff}}$ nearly equals I_3 , as depicted in Figure 2 (Right). Therefore, although the time-decay of the relative entropy is entirely due to diffusive terms in the equation, the magnitude of $\left(\frac{dR}{dt}\right)_{\text{diff}}$ is almost completely determined by the mean of the ensemble forecast.

We have shown that the relative entropy of the full distribution decreases monotonically to zero, although at a non-uniform rate. As we will see next, the situation is quite different for the relative entropies of the marginal distributions $p(x_1, t)$ and $p(x_2, t)$ which may increase in time.

3.3 Return of skill for marginal entropies

We next consider the marginal entropies of the stationary and transient distributions. The relative entropy $R_{x_1}(t)$ and $R_{x_2}(t)$ for the two marginal distributions are again defined using equation (1), and can be interpreted as the amount of information that the marginal distribution $p(x_1, t)$ provides about the state of the variable x_1 at time t in excess of the information provided by the marginal stationary distribution $q(x_1)$. We emphasize that marginal entropies are not invariant under coordinate changes, so that the results of this section are highly coordinate dependent.

Note that in contrast to the full relative entropy, the marginal relative entropies do not necessarily decay in time. However, we can relate their behavior using *conditional relative entropies* which are defined by the following equation

$$\begin{aligned} R_{x_2|x_1}(t) &= R(p(x_2|x_1, t), q(x_2|x_1)) \\ &= \int p(x_1, t) \int p(x_2|x_1, t) \log \frac{p(x_2|x_1, t)}{q(x_2|x_1)} dx_1 dx_2. \end{aligned} \quad (14)$$

Here $p(x_2|x_1, t)$ denotes the conditional distribution of x_2 at time t given x_1 , and $R_{x_1|x_2}(t)$ is the excess information provided by the marginal distribution $p(x_2|x_1, t)$ over $q(x_2|x_1)$.

The *chain rule for relative entropy* [11] relates the full, marginal and conditional relative entropy

$$R(t) = R_{x_2|x_1}(t) + R_{x_1}(t). \quad (15)$$

We start by calculating the relative entropy of the marginal distributions which can be obtained analytically using equation (2). The evolution of the marginal relative entropies is shown in Figure 5. For the initial condition $(x_1^0, x_2^0) = (1, 1)$ the oscillations are well pronounced. This observation implies that the information about the variables x_1 and x_2 taken separately can increase in time, while information about their joint distribution must always decrease.

The top and bottom panels of Figure 4 compare the evolution of the full and marginal distributions. The increases in marginal relative entropy correspond to the times at which the mean of the marginal distribution moves away from the mean of the stationary distribution, *i.e.* the plateaus in the full relative entropy. The main factors contributing to this behavior can be identified as in the previous section, and here we provide an equivalent intuitive explanation.

The bottom panels of Figure 4 shows that the variance of the distribution $p(x_1|x_2, t)$ remains nearly constant during one oscillation. However, during the

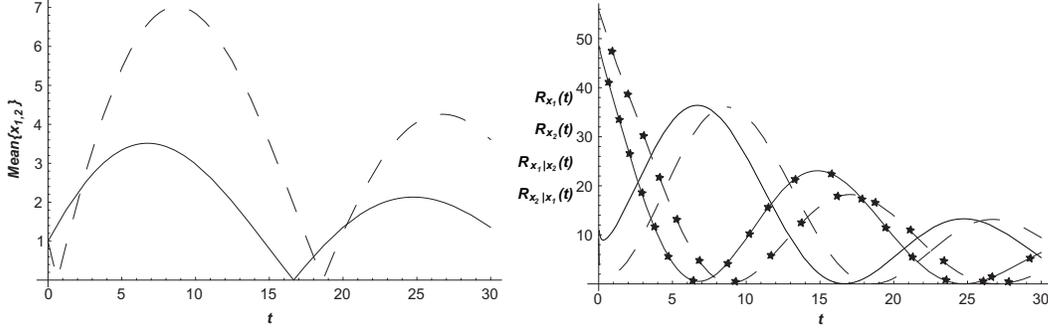


Fig. 5. Left: means of the variables $|x_1|$ (solid line) and $|x_2|$ (dashed line). Right: marginal relative entropies $R_{x_1}(t)$ (solid line) and $R_{x_2}(t)$ (dashed line) and $R_{x_1|x_2}(t)$ (solid-star line) and $R_{x_2|x_1}(t)$ (dashed-star line) for an initial condition $(x_1^0, x_2^0) = (1, 1)$.

time between the first two panels the mean of the distribution moves away from 0 which leads to an increase in $R_{x_1}(t)$. Similarly, the movement of the transient to the stationary marginal distribution during the period between the last two panels leads to a decrease in $R_{x_1}(t)$. The fact that both marginal relative entropies $R_{x_1}(t)$ and $R_{x_2}(t)$ increase at the same time, is a consequence of the fact that for the solution of the corresponding deterministic system both $\bar{x}_1(t)$ and $\bar{x}_2(t)$ can increase at the same time. Note that this would not be true in a different coordinate system. In particular, for the matrix A in (11), $\bar{x}_1(t)$ and $\bar{x}_2(t)$ and the marginal relative entropies oscillate out of phase.

It is natural to ask how the information contained in the marginal distributions of x_1 and x_2 is generated. Equation (15) provides the answer: With an increase in information about the marginals comes a decrease in information about the conditional distribution, that is, a decrease in the excess of information that a knowledge of x_1 provides about the state of x_2 over that provided by the stationary distribution $q(x_2|x_1)$ (see Figure 5 (Right)).

We also note that there is no direct “flow of information” between the variables x_1 and x_2 . However, one can think of a flow of information between the marginal and conditional distributions when the full relative entropy is approximately constant, since during that time the sum of the two is approximately constant as well.

4 Nonuniform decay of $R(t)$ in general systems

The results of the previous section extend to much more general stochastic systems. The nonuniform decay of relative entropy occurs whenever the main mass of the distribution $p(\vec{x}, t)$ approaches, and then diverges from the main mass of the stationary distribution $q(\vec{x})$. Oscillations in the marginal relative

entropies occur when such divergence occurs in the marginal distributions. The following two examples show that such behavior can be expected both in the case of stochastic oscillators, when the mass of the stationary distribution is distributed non-uniformly around the limit cycle, and in the case of stochastically perturbed homoclinic and heteroclinic cycles. Such dynamical behavior is often credited for the complex evolution in various prototype atmospheric models [8,9].

4.1 Decay of relative information for nonlinear oscillators

In this section we show that the behavior of relative entropy described in the previous section can be observed in the case of nonlinear oscillators. In particular, we consider a planar, stochastic system with a limit cycle arising from a supercritical *Hopf bifurcation*:

$$\begin{aligned} dx_1 &= \mu x_1 dt - c\omega x_2 dt + \Theta x_1(x_1^2 + c^2 x_2^2) dt + \varepsilon dW_1, \\ dx_2 &= \frac{1}{c} \left(\omega x_1 + c\mu x_2 + c\Theta x_2(x_1^2 + c^2 x_2^2) \right) dt + \varepsilon dW_2, \end{aligned} \tag{16}$$

here $W_{1,2}$ are independent Wiener processes. Since the corresponding Fokker-Planck solution cannot be solved analytically, we examine the system numerically using the parameter values $\mu = 0.5$, $\omega = 1.0$, $c = 0.6$, and $\Theta = -1.0$. Similar behavior can be observed over a wide range of parameters. In the absence of white noise the system in (16) has a stable periodic orbit

$$x_1(t) = \sqrt{-\frac{\mu}{\Theta}} \cos(\omega t + \phi_0), \quad x_2(t) = \frac{1}{c} \sqrt{-\frac{\mu}{\Theta}} \sin(\omega t + \phi_0),$$

with period $T_{per} = 2\pi/\omega$. For small noise, the invariant measure is concentrated sharply around the vertical extrema of the unperturbed orbit. Similar to the linear oscillator, the invariant measure is stretched in the x_2 direction to better illustrate the non-uniform decay of relative entropy. Note that the speed at which a trajectory moves around the attracting periodic orbit of the deterministic system is at a minimum at the top and the vertical extrema of the orbit. These are therefore the places at which the equilibrium distribution will have local maxima. Similar behavior can be observed for other values of c for which the equilibrium measure is distributed non-uniformly along the limit cycle.

The relative entropy is evaluated numerically by discretizing the phase space into a uniform mesh. Stochastic Euler method is used to integrate the equation. The equilibrium distribution $q(\vec{x})$ is estimated utilizing bin-counting from

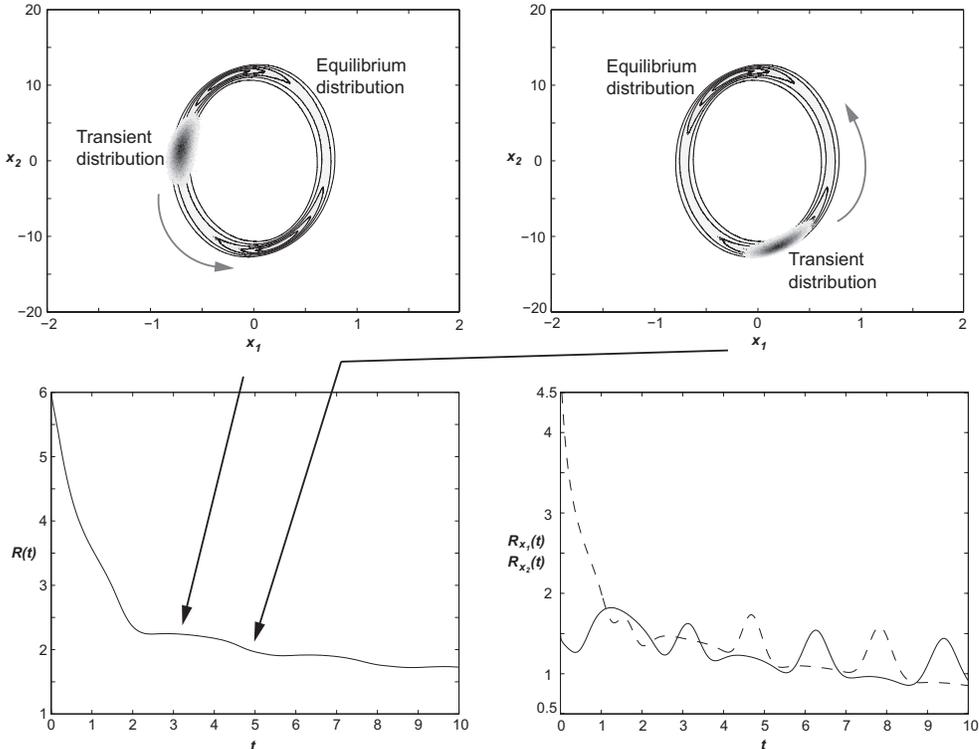


Fig. 6. Top: Probability density function for time=3 (left) and time=5 (right). Bottom Left: Full relative entropy $R(t)$. Bottom Right: marginal relative entropies $R_{x_1}(t)$ (solid) and $R_{x_2}(t)$ (dashed) in simulations of (16) with initial ensemble centered at $(x_1^0, x_2^0) = (0.5, 0)$ and $\varepsilon = 0.1$.

a single long realization. The initial non-equilibrium ensemble is a 250,000-member ensemble generated from the uniform distribution with width 0.3×0.3 centered at $(x_1, x_2) = (0.5, 0)$, away from the mean of the equilibrium distribution. Numerical estimates for relative entropy $R(t)$ and marginals relative entropies are shown in Figure 6.

The non-uniform decay of relative entropy and oscillations in marginal relative entropies are clearly visible after a short transient period. Since the periodic orbit given in (17) is stable, the transient period is due to the fast initial transition of the initial ensemble to the vicinity of this orbit.

After this initial phase, the relative entropy decays more slowly. As in the case of the damped linear oscillator, the variance of the transient distribution increases slowly compared to the time of the oscillations. Figure 6 shows that the plateaus in relative entropy correspond to the times during which the mass of the transient distribution moves away from a peak in the mass of the stationary distribution. Therefore the plateaus occur at twice the frequency ω , and the nonuniform decay of relative entropy is due to the mechanism discussed in the previous section.

4.1.1 Return of skill for marginal entropies

Marginal entropies depicted in Figure 6 exhibit strong out of phase oscillatory behavior with frequency 2ω , consistent with the frequency of plateaus of the full relative entropy. The marginal stationary distribution $q(x_1)$ is approximately unimodal. As in the case of the linear oscillator, the minima of the marginal entropies $R_{x_1}(t)$ occur at the times at which the mean value of the transient marginal distribution coincides with the mean value of the equilibrium distribution (top right of Figure 6).

Since the marginal distribution $q(x_2)$ is strongly bimodal, the situation is somewhat different. The minima of $R_{x_2}(t)$ occur at the times when the mean of the transient distribution $p(x_2)$ is between the two peaks in the stationary distribution $q(x_2)$. Since this occurs exactly when the distribution $p(x_2)$ is at its farthest distance from $q(x_1)$, the marginal relative entropies $R_{x_1}(t)$ and $R_{x_2}(t)$ oscillate out of phase.

4.2 Stochastically Perturbed Duffing Equation

We next consider a system of nonlinear stochastic differential equations exhibiting coherence resonance [17,10]. Although the deterministic analog of this system is very different from both previous examples, stochastic perturbations lead to intervals of extended predictability and the return of skill for marginal distributions.

The model is given by the Duffing equations driven by white noise [18]

$$\begin{aligned} dx_1 &= x_2 dt + \varepsilon dW_1, \\ dx_2 &= (x_1 - x_1^3 - \gamma x_2 + \beta x_1^2 x_2) dt + \varepsilon dW_2, \end{aligned} \tag{17}$$

where $W_{1,2}$ are independent Wiener processes, and γ, β and ε are parameters. For $\varepsilon = 0$ and parameters $\gamma = 0.4$ and $\beta = 0.497$ this system has an attracting double homoclinic cycle (Figure 7 (Left)) to the saddle point at the origin. The signature of the homoclinic connection is clearly visible in the invariant density of the stochastic system in (17) shown in Figure 7 (Right).

To demonstrate the existence of extended regions of predictability and the return of skill for marginal distributions we chose a particular 250,000-member initial ensemble centered at $x_1 = 0.25$, $x_2 = 0.25$, and estimate the relative entropy numerically as in the previous example. The distributions $p(\vec{x}, t)$ are computed utilizing the Monte-Carlo simulations with the initial ensemble generated from the uniform distribution on $[0.3] \times [0.3]$. (see Figure 7 (Left)).

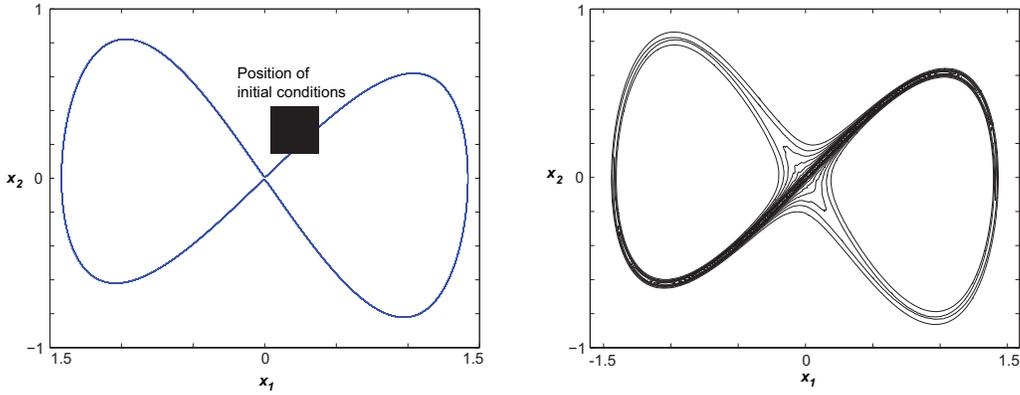


Fig. 7. Left: Homoclinic loop for Duffing equation with $\varepsilon = 0$. Right: Contour plot of probability density function for $\varepsilon = 0.01$.

The computed relative entropy for $\varepsilon = 0.01$ is presented in the bottom left panel of Figure 8. As in the linear oscillator case the relative entropy is almost constant over several time intervals. There are four plateaus in the graph of relative entropy, although the nature of the first plateau (at times $[4 \dots 8]$) is somewhat different from the subsequent ones.

Recall, that the decay in relative entropy is only due to the diffusion in equation (6). For the model (17) the diffusion term becomes

$$\left(\frac{dR}{dt}\right)_{\text{diff}} = -\frac{\varepsilon}{2} \int dx_1 dx_2 p(x_1, x_2) \sum_{i=1,2} \left[\frac{\partial}{\partial x_i} \left(\log \frac{p(x_1, x_2)}{q(x_1, x_2)} \right) \right]^2. \quad (18)$$

For the stochastic Duffing equation the behavior of the $\left(\frac{dR}{dt}\right)_{\text{diff}}$ is more complicated than in the case of linear oscillator. Namely, the value of $\left(\frac{dR}{dt}\right)_{\text{diff}}$ depends not only on the means, but also on all terms involving variances of the invariant and transient distributions.

Since the initial ensemble is chosen on one side of the heteroclinic loop, different trajectories do not separate during the first passage along the heteroclinic loop (see the first two top panels in Figure 8). As the cluster of initial conditions moves away from the origin we observe the long plateau in the graph of relative entropy, since the transient distribution moves away from the origin where the main mass of the equilibrium distribution is located. Indeed, Figure 9 illustrates that the mean of transient distribution is largest at times $[4 \dots 8]$, coinciding with the first plateau in relative entropy $R(t)$.

After the first transition, individual realizations return close to the origin, but separate following the two different branches of the homoclinic loop. Therefore, the mean of the ensemble is approximately zero (see Figure 9). Due to coherence resonance [10,17,18], most of the mass of the transient distribution

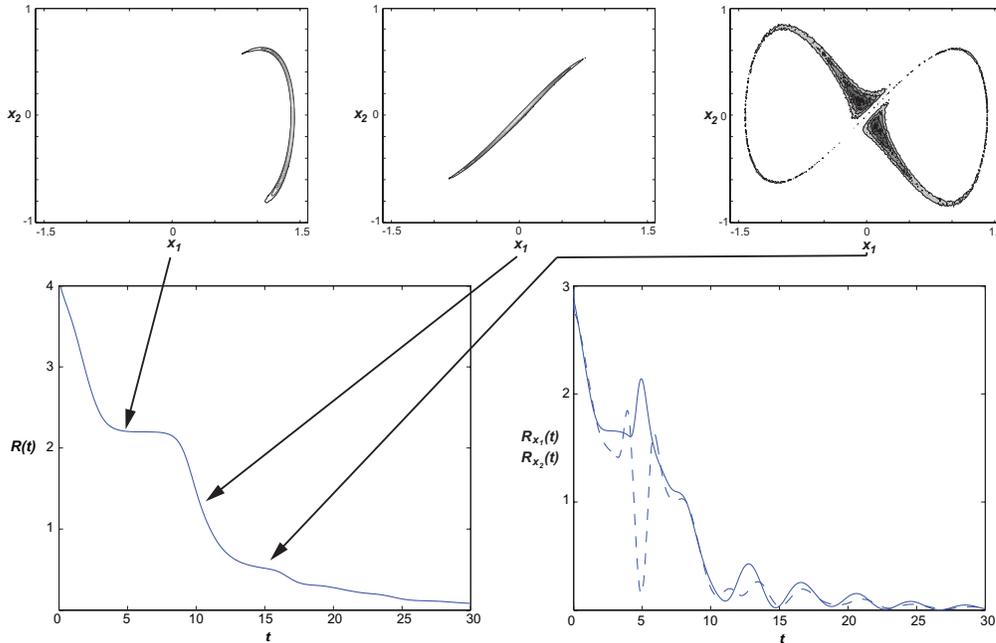


Fig. 8. Top: Probability density function for time=5 (left), time=10 (middle) and time=15 (right). Bottom Left: Full relative entropy $R(t)$, Bottom Right: marginal relative entropies $R_{x_1}(t)$ (solid) and $R_{x_2}(t)$ (dashed) in simulations of (17) with initial ensemble centered at $(x_1^0, x_2^0) = (0.25, 0.25)$ and $\varepsilon = 0.01$.

is ejected from the vicinity of the origin around the same time. The second plateau in the graph of relative entropy occurs when the two main portions of the transient distribution are at their farthest distance from the main mass of the equilibrium distribution at times [15...17]. The bimodality of the transient distribution during this time implies that the oscillatory behavior is manifested strongly through the variance of the ensemble (see Figure 9).

Although the details are somewhat different from the previous examples, the non-uniform decay in relative entropy is again due to the fact that the stationary distribution is concentrated in one area of the phase space, and oscillations in the system that take the transient distribution recurrently close to the main mass of the stationary distribution.

4.2.1 Return of skill for marginal entropies

Marginal relative entropies for x_1 and x_2 are shown in the bottom right panel of Figure 8. The mechanism leading to the oscillations in both marginal entropies is similar to the one described in the preceding examples. An inspection of Figure 8 shows that the marginal entropies are at a maximum at the times during which the main mass of the marginal transient distribution diverges maximally from the marginal of the stationary distribution. The fact that the transient and stationary distribution for x_1 are bimodal and trimodal, respectively,

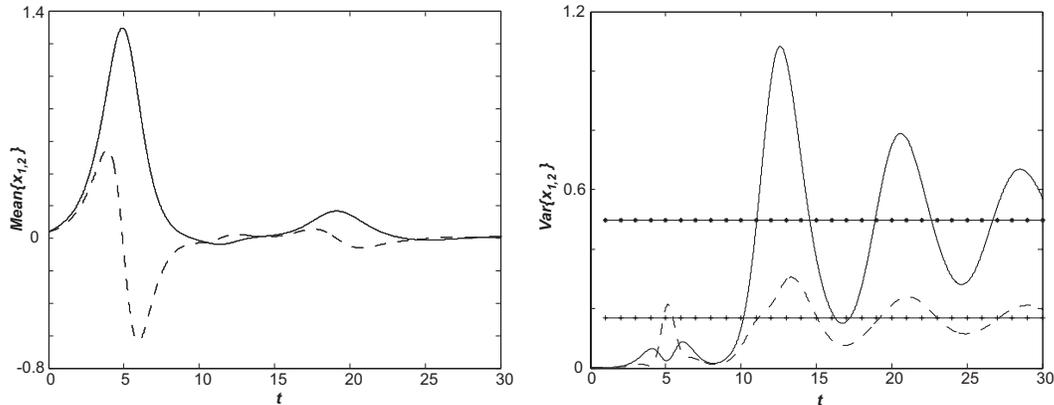


Fig. 9. Left: mean in x_1 (dashed line), mean in x_2 (solid line), Right: variance in x_1 (dashed line), variance in x_2 (solid line) in simulations of the stochastic Duffing equation in (17) with $\varepsilon = 0.01$. (Horizontal lines show equilibrium variances).

somewhat complicates the description. However, the animation provided at http://www.math.uh.edu/~ilya/research/predict_stoch_osc illustrates the entire process.

5 Conclusions

We considered the predictability of three models with particular emphasis on the non-uniform decay of the utility of predictions and return of skill (oscillations in marginals) for dynamic variables. These models were constructed as stochastic perturbations of linear oscillator, non-linear oscillator (Hopf normal form), and homoclinic cycle (Duffing equation), and are representative of a wide class of stochastic oscillators.

Relative entropy is utilized to characterize the predictability properties of these prototype systems. The averaged (with respect to many initial ensembles) predictability of all three systems decays exponentially in time. Nevertheless, as a result of the oscillatory-like behavior, two related phenomena emerge in the behavior of the relative entropy functional and marginal entropies for each particular ensemble simulation. In particular, (i) the full relative entropy decays at a non-uniform rate, and (ii) there is return of skill (oscillatory behavior) for the marginal entropies of all three systems.

Interestingly, we can also think of the return of skill as a flow of information from the conditional to the marginal non-equilibrium distribution. Both of these phenomena are driven by oscillations of the mean of the non-equilibrium (forecast) ensemble, and an increase in the variance of the non-equilibrium ensemble that is slow compared to the frequency of oscillation.

The leading order effect in this case is the transport of the non-equilibrium distribution in phase space by the underlying oscillatory dynamics. This results in a slower rate of decay for the relative entropy when the mean of the non-equilibrium ensemble is moving away from an area in which the invariant measure is concentrated. The same mechanism causes oscillations of marginal distributions and return of skill in each dynamic variable.

While the quantitative details differ between the oscillatory mechanisms considered, the qualitative behavior of the relative entropy functional is similar in all three cases. The oscillatory behavior is manifested strongly for initial ensembles concentrated in the tails of the invariant measure, but can also be detected for other initial data. This suggests that similar behavior of various predictability metrics can be detected in more complex systems, especially for initial ensembles concentrated around rare events.

Acknowledgements

The authors would like to thank Professor Richard Kleeman for valuable comments and suggestions. This research is partially supported by the NSF Grant ATM-0417867. I. Timofeyev is also partially supported by the NSF Grant DMS-0405944 and by the DOE Grant DE-FG02-04ER25645.

References

- [1] J. Anderson and H. van den Dool. Skill and return of skill in dynamic extended-range forecasts. *Mon. Wea. Rev.*, 122:507–516, 1994.
- [2] G. Boer. A study of atmosphere-ocean predictability on long time scales. *Clim. Dyn.*, 16:469–472, 2000.
- [3] G. Boffetta, M. Cencini, M. Falcioni, and A. Vulpiani. Predictability: a way to characterize complexity. *Physics Reports*, 356:367–474, 2002.
- [4] G. Branstator. The variability in skill of 72-hour global-scale NMC forecasts. *Mon. Wea. Rev.*, 114(12):2628–2639, 1986.
- [5] P. C. Chu, L. M. Ivanov, L. H. Kantha, O. V. Melnichenko, and Y. A. Poberezhny. Power law decay in model predictability skill. *Geophys. Res. Lett.*, 29 (15):1748–1753, 2002.
- [6] M. Collins. Climate predictability on interannual to decadal time scales: the initial value problem. *Clim. Dyn.*, 19:671–692, 2002.
- [7] T. M. Cover and J. Thomas. *Elements of information theory*. John Wiley and Sons, 1991.

- [8] D. T. Crommelin. Homoclinic dynamics: a scenario for atmospheric ultra-low-frequency variability. *J. Atmos. Sci.*, 59:1533–1549, 2002.
- [9] D. T. Crommelin. Regime transitions and heteroclinic connections in a barotropic atmosphere. *J. Atmos. Sci.*, 60:229–246, 2003.
- [10] R. E. L. DeVelle, E. Vanden-Eijnden, and C. B. Muratov. Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Phys. Rev. E*, 72:031105, 2005.
- [11] C. W. Gardiner. *Handbook of stochastic methods, 3rd edition*. Springer Verlag, Berlin, 2004.
- [12] P. Gaspard. *Chaos, scattering and statistical mechanics*, volume 9 of *Cambridge Nonlinear Science Series*. Cambridge University Press, Cambridge, 1998.
- [13] T. Kestin, D. Karoly, J.-I. Yano, and N. A. Rayner. Time frequency variability of enso and stochastic simulations. *J. Clim.*, 11:2258–2272, 1998.
- [14] R. Kleeman. Measuring dynamical prediction utility using relative entropy. *J. of the Atmospheric Sciences*, 59:2057–2072, 2002.
- [15] E. Ott. *Chaos in dynamical systems*. Cambridge University Press, Cambridge, second edition, 2002.
- [16] T. Palmer. Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Met. Soc.*, 74(1):49–65, 1993.
- [17] A. S. Pikovsky and J. Kurths. Coherence resonance in a noise-driven excitable system. *Phys. Rev. Lett*, 78(5):775–778, 1997.
- [18] E. Stone and P. Holmes. Random perturbations of heteroclinic attractors. *SIAM Journal on Applied Mathematics*, 50(3):726–743, 1990.
- [19] J. J. Tribbia and D. P. Baumhefner. Scale interactions and atmospheric predictability: An updated perspective. *Mon. Wea. Rev.*, 132:703–713, 2004.
- [20] C. Ziehmann, L. A. Smith, and J. Kurths. Localize Lyapunov exponent and the prediction of predictability. *Phys. Lett. A*, 271(4):237–251, 2000.