

Article

## Entropy Evolution and Uncertainty Estimation with Dynamical Systems

X. San Liang <sup>1,2</sup>

<sup>1</sup> School of Marine Sciences and School of Mathematics and Statistics, Nanjing University of Information Science and Technology (Nanjing Institute of Meteorology), 219 Ningliu Blvd, Nanjing 210044, China; E-Mail: sanliang@courant.nyu.edu; Tel.: +86-25-5869-5695; Fax: +86-25-5869-5698

<sup>2</sup> China Institute for Advanced Study, Central University of Finance and Economics, 39 South College Ave, Beijing 100081, China

Received: 4 May 2014; in revised form: 11 June 2014 / Accepted: 19 June 2014 /

Published: 30 June 2014

---

**Abstract:** This paper presents a comprehensive introduction and systematic derivation of the evolutionary equations for absolute entropy  $H$  and relative entropy  $D$ , some of which exist sporadically in the literature in different forms under different subjects, within the framework of dynamical systems. In general, both  $H$  and  $D$  are dissipated, and the dissipation bears a form reminiscent of the Fisher information; in the absence of stochasticity,  $dH/dt$  is connected to the rate of phase space expansion, and  $D$  stays invariant, *i.e.*, the separation of two probability density functions is always conserved. These formulas are validated with linear systems, and put to application with the Lorenz system and a large-dimensional stochastic quasi-geostrophic flow problem. In the Lorenz case,  $H$  falls at a constant rate with time, implying that  $H$  will eventually become negative, a situation beyond the capability of the commonly used computational technique like coarse-graining and bin counting. For the stochastic flow problem, it is first reduced to a computationally tractable low-dimensional system, using a reduced model approach, and then handled through ensemble prediction. Both the Lorenz system and the stochastic flow system are examples of self-organization in the light of uncertainty reduction. The latter particularly shows that, sometimes stochasticity may actually enhance the self-organization process.

**Keywords:** uncertainty estimation; entropy; second law of thermodynamics; Lorenz system; ensemble prediction; quasi-geostrophic flow; shear instability; Fisher information matrix; reduced model approach; self-organization

---

## 1. Introduction

Uncertainties are ubiquitous. Perhaps the best evidence that pertains to our daily life comes from the atmospheric system. The recent extremely freezing weather in North America that led to a loss of at least 5 billion US dollars is just such an example (*cf.* Wikipedia under the title “Early 2014 North American cold wave”). Having experienced the record cold in the past 20 years and a heavy, intense snow storm brought by a polar vortex during 7–9 January 2014, the City of Boston, Massachusetts, announced a state of emergency as the second polar vortex was by prediction about to approach 13 days later, with public school closures and a lot of air flight cancellations. However, on that day (22 January) the weather was not that bad at all as predicted; actually it was quite normal. It is the uncertainty, which may sit in the forecast model, or the weather system itself, or both, that prevents us from reaching an accurate forecast.

Uncertainties are measured by entropy, a physical notion introduced by Claude E. Shannon in 1948 [1]. It is connected to the thermodynamic entropy (originally introduced by Rudolf Clausius) through its microscopic version in statistical mechanics (defined by J. Willard Gibbs in 1878 after earlier work by Boltzmann (1872) [2]). This connection, however, is still in debate; pros and cons can be found respectively in, say, [3,4]. Here we are not intended to be involved into the debate; we will limit our discussion to Shannon entropy (or absolute entropy) and one of its derivatives namely Kullback-Leibler divergence (or relative entropy).

There is in the literature a large body of work on entropy production (e.g., [5–12]; refer to [13] for an excellent review and discussion). In comparison, investigation of entropy evolution with a given dynamical system is by far from enough. This is unfortunate, as entropy evolution is not only theoretically important *per se*, but also closely related to our daily life; the polar vortex-induced severe weather mentioned above is just such an example. In fact, uncertainty evolution prediction is a big issue in dynamic meteorology and oceanography and there is a long history of research ever since 60's [14–24], albeit in different fashions. Traditionally, a common practice in studying atmospheric/oceanic uncertainty is through probability prediction, with probability density function (pdf) estimated through bin counting). For ocean and atmosphere forecasts, this is in general very expensive. Considering the large dimensionality of the system, one forecast *i.e.*, one single realization in the sample space, could be time consuming, while the sample space is usually made of a huge number of forecasts. On the other hand, this traditional and common practice is actually prone to mistake; a demonstration with the Lorenz system will be given later in Section 5. In this study, we will show that, given a dynamical system, there actually exist some neat laws on entropy evolution. With these laws one can in many cases avoid estimating the pdfs and hence overcome the above difficulties. Considering the large body of atmosphere-ocean work on uncertainty studies, this research is expected to provide a timely help.

So far, in the setting of a Markov chain, Cover and Thomas [25] have proved that relative entropy will never increase, but absolute entropy does not have this property. If the system given is deterministic, it has been established that the time rate of change of absolute entropy is precisely equal to the divergence of the vector field of the system, followed by a mathematical expectation (e.g., [26,27]), and that the relative entropy is always conserved [28]. Particularly, the evolution law of absolute entropy has led to a rigorous formalism of information flow, a fundamental notion in general physics which has broad

applications in different disciplines [27,29]. For stochastic dynamical systems, an evolution equation is stated under a positivity assumption for relative entropy in exploring an approach to the Fokker–Planck equation [30]. As can be seen, these laws, though important, only partially exist and appear sporadically in different forms in the literature; even these limited investigations are more often than not sandwiched within other subjects. In this study, we want to give them a comprehensive introduction and a systematic derivation within a generic framework of dynamical systems, both deterministic and stochastic. As a verification, these laws are alternatively derived with linear systems in terms of means and covariances. Those already existing in the literature will be clearly remarked in the text; we include them here for completeness, not to claim originality.

Another objective of this study is to demonstrate how these evolutionary laws may be applied to real physical problems. We will focus on two cases: one low-dimensional system, another large-dimensional system. For the former, we will investigate how entropy changes with the Lorenz system as the butterfly pattern appears. For the latter, the instability of a quasi-geostrophic flow is examined. The biggest issue with this problem is how a statistically complete but computationally tractable ensemble is formed. Consider a system of dimensionality  $n$ . For each dimension even if only 2 random draws are made, the ensemble members will total to  $2^n$ . Computationally this is feasible only when  $n$  is very small (say  $n < 15$ ), while in reality  $n$  is usually huge (larger than 10,000)! Whether the afore-mentioned laws can be made applicable to large-dimensional systems relies heavily on the solution of this issue, which we will be demonstrating in Section 6. Besides the technical aspect, we also would like to explore what the impact it would be if stochasticity sneaks into a large-dimensional deterministic system. Stochastic modeling is hot these days in oceanography and atmospheric science, partly reflecting the attempt to recover the unresolved small-scale processes, *i.e.*, to assess the effect of the tiny deviations from average between the grid points that the computer cannot see (which may multiply and eventually deteriorates the forecast), and characterize the inaccuracies of the model *per se*. But how the addition of stochasticity may impact the otherwise deterministic partial differential equation is far from being investigated. As an application of the established laws, this study will show some preliminary results on this impact in terms of entropy evolution.

In the following the evolutionary laws for absolute and relative entropies are systematically derived. Specifically, Section 2 is for deterministic systems, and Section 3 for stochastic systems. To validate the obtained formulas, in Section 4 the linear case, thank to its elegant form, is independently derived. This is followed by two applications: one with the renowned Lorenz system (Section 5), another with a quasi-geostrophic shear flow instability problem (Section 6). In section 7 the obtained formulas are summarized, and a discussion of the results are given.

## 2. Deterministic Systems

First consider an  $n$ -dimensional deterministic system with randomness limited within initial conditions:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t), \quad (1)$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$  are the state variables. Associated with  $\mathbf{x}$  there is a joint probability density function,  $\rho = \rho(t; \mathbf{x}) = \rho(t; x_1, x_2, \dots, x_n)$ , and hence an absolute entropy

$$H = - \int_{\mathbb{R}^n} \rho \log \rho \, d\mathbf{x}, \tag{2}$$

and a relative entropy

$$D = \int_{\mathbb{R}^n} \rho \log \frac{\rho}{q} \, d\mathbf{x} = -H - \int_{\mathbb{R}^n} \rho \log q \, d\mathbf{x}, \tag{3}$$

with some reference density  $q$  of  $\mathbf{x}$ . We are interested in how  $H$  and  $D$  evolve with respect to Equation (1). For this purpose, assume that  $\rho$ ,  $q$ , and their derivatives are all compactly supported; further assume enough regularity for  $\rho$ ,  $q$ ,  $D$ , and  $H$ . The mathematics involved here is neglected for a broad readership; those who feel interested may consult [25] for a detailed discussion. Note the choosing of the reference density  $q$  is slightly different from what people are using these days [21] in applications, particularly in predictability studies, who usually choose it to be some constant distribution (initial distribution, for example). Here  $q$  is not fixed; it is just a different distribution. Although  $\rho$  and  $q$  do not commute in  $D$ , there is an effective symmetry in the sense that we can consider  $\rho$  and  $q$  as two initial densities. The relative entropy reports the discriminability of them in terms of their divergence. This means that if relative entropy is conserved (*cf.* Theorem 2 below), the two densities are then equally discriminable as time progresses.

Corresponding to Equation (1) there is a Liouville equation

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{F}) = \frac{\partial \rho}{\partial t} + \frac{\partial(\rho F_1)}{\partial x_1} + \frac{\partial(\rho F_2)}{\partial x_2} + \dots + \frac{\partial(\rho F_n)}{\partial x_n} = 0 \tag{4}$$

governing the evolution of the joint density  $\rho$ . Here  $\nabla$  is the gradient operator with respect to the  $n$ -dimensional vector  $\mathbf{x}$ . From the equation we derive the time rate of change of  $H$  and  $D$ . These results have been obtained before; we include them here for completeness.

**Theorem 1.** *For the deterministic system (1), the joint absolute entropy of  $\mathbf{x}$  evolves as*

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}), \tag{5}$$

where the operator  $E$  stands for mathematical expectation.

Remark: This equation was first derived in [26], expressed then as  $\frac{dH}{dt} = \iint \rho \nabla \cdot \mathbf{F} d\Gamma$ . The above neat form (5) was independently obtained in [27], which allows for the establishment of a rigorous formalism of information flow or information transfer, a fundamental notion in general physics which has wide applications.

**Proof.** We follow [27] to establish this equation. Multiply (4) by  $-(1 + \log \rho)$  to get

$$-\frac{\partial \rho \log \rho}{\partial t} = \nabla \cdot (\mathbf{F} \rho \log \rho) + \rho \nabla \cdot \mathbf{F}.$$

Integrate over the sample space  $\mathbb{R}^n$ . Since  $\rho$  is compactly supported, the first term on the right hand side vanishes, and hence

$$\frac{dH}{dt} = \int_{\mathbb{R}^n} \rho \nabla \cdot \mathbf{F} dx = E(\nabla \cdot \mathbf{F}). \quad \square$$

In arriving at this formula, originally in [27] it is assumed that extreme events have a probability of zero, which corresponds to our above compact support assumption. This makes sense in practice and has been justified in [27], but even this assumption may be relaxed, and the same formula follows [31].

Parallel to Equation (5), the evolution law for  $D$  is also very neat and, actually, simpler.

**Theorem 2.** *As a state evolves with the deterministic system (1), its relative entropy or Kullback-Leibler divergence is conserved.*

Remark: This property was first discovered by Plastino and Daffertshofer [28].

**Proof.** Differentiation of (3) with respect to  $t$  gives

$$\begin{aligned} \frac{dD}{dt} &= -\frac{dH}{dt} - \int \frac{\partial \rho}{\partial t} \log q dx - \int \frac{\rho}{q} \frac{\partial q}{\partial t} dx \\ &\equiv -\frac{dH}{dt} + (I) + (II). \end{aligned}$$

The integrals are all understood to be over  $\mathbb{R}^n$ , and this simplification will be used hereafter, unless otherwise indicated. The two shorthands are:

$$\begin{aligned} (I) &= \int_{\Omega} [\nabla \cdot (\rho \mathbf{F}) \log q] dx = \int_{\Omega} \nabla \cdot (\rho \mathbf{F} \log q) - \int_{\Omega} \rho \mathbf{F} \cdot \nabla (\log q) \\ &= -E(\mathbf{F} \cdot \nabla \log q), \\ (II) &= -E\left(\frac{\partial \log q}{\partial t}\right). \end{aligned}$$

So

$$\begin{aligned} \frac{dD}{dt} &= -\frac{dH}{dt} - E\left(\frac{\partial \log q}{\partial t} + \mathbf{F} \cdot \nabla \log q\right) \\ &= -\frac{dH}{dt} - E\left[\frac{1}{q} \left(\frac{\partial q}{\partial t} + \mathbf{F} \cdot \nabla q\right)\right]. \end{aligned} \tag{6}$$

Recall that  $q$  is also a joint density of  $\mathbf{x}$ , so its evolution must follow the same Liouville equation, *i.e.*,

$$\frac{\partial q}{\partial t} + \mathbf{F} \cdot \nabla q = -q \nabla \cdot \mathbf{F}.$$

The relative entropy evolution (6) thus becomes

$$\frac{dD}{dt} = -\frac{dH}{dt} + E(\nabla \cdot \mathbf{F}). \tag{7}$$

Substitution of Equation (5) for  $\frac{dH}{dt}$  gives

$$\frac{dD}{dt} = 0. \tag{8}$$

That is to say, relative entropy is conserved.  $\square$

The concise laws stated in Theorems 1 and 2 have clear physical interpretations. Equation (5) tells that that the time rate of change of absolute entropy actually depends on the expansion or contraction of

the phase space of concern; indeed, it is essentially equal to the average expansion rate per unit volume. To interpret Equation (8), recall that relative entropy may be taken as a measure of the distance between the two densities  $\rho$  and  $q$  in some function space  $L^1(\mathbb{R}^n)$  (i.e., space of integrable functions) [25], albeit it does not meet all the axioms for a metric. The conservation law (8) hence tells that the separation of  $\rho$  and  $q$  never changes as time moves on, even with highly nonlinear and chaotic systems.

### 3. Stochastic Systems

#### 3.1. Absolute Entropy

We now generalize the above results to systems with stochasticity included. Let  $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$  be an array of  $n$  standard Wiener processes, and  $\mathbf{B}$  a matrix which may have dependency on both  $\mathbf{x}$  and  $t$ . The system we are about to consider bears the form:

$$d\mathbf{x} = \mathbf{F}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{w}. \tag{9}$$

Correspondingly the density evolves according to a Fokker–Planck equation

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{F}) + \frac{1}{2} \nabla \nabla : (\rho \mathbf{G}), \tag{10}$$

where  $\mathbf{G} = \mathbf{B}\mathbf{B}^T$  is a nonnegatively definite matrix. The double dot product here is defined such that, for column vectors  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ , and  $\mathbf{d}$ ,

$$(\mathbf{ab}) : (\mathbf{cd}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}).$$

A dyadic  $\mathbf{ab}$  in matrix notation is identified with  $\mathbf{ab}^T$ .

**Theorem 3.** For system (9), the joint absolute entropy evolves as

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) - \frac{1}{2} E(\mathbf{G} : \nabla \nabla \log \rho). \tag{11}$$

If  $\mathbf{B}$  is a constant matrix, it can be expressed in a more familiar form:

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) + \frac{1}{2} E(\nabla \log \rho \cdot \mathbf{G} \cdot \nabla \log \rho). \tag{12}$$

Remark: We learned later on that recently Friston and Ao [32] arrived at a formula in the form:  $\frac{dH}{dt} = \int \log \rho \nabla \cdot (\mathbf{F}\rho) d\mathbf{x} + \int \frac{\nabla \rho \cdot \mathbf{\Gamma} \cdot \nabla \rho}{\rho} d\mathbf{x}$ , if  $\mathbf{\Gamma} = \gamma \mathbf{I}$  ( $\gamma > 0$ ) is diagonal.

**Proof.** Multiplication of Equation (10) by  $-(1 + \log \rho)$ , followed by an integration over the entire sample space  $\mathbb{R}^n$ , yields an evolution of the absolute entropy

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) - \frac{1}{2} \int (1 + \log \rho) \nabla \nabla : (\rho \mathbf{G}) d\mathbf{x}. \tag{13}$$

In arriving at the first term on the right hand, the previous result (i.e., Equation (5)) with the Liouville equation has been applied. For the second term, since  $\int \nabla \nabla : (\rho \mathbf{G}) d\mathbf{x} = 0$  by the compact support assumption, it results in

$$\begin{aligned} -\frac{1}{2} \int \log \rho \nabla \nabla : (\rho \mathbf{G}) d\mathbf{x} &= \frac{1}{2} \int \nabla \cdot (\rho \mathbf{G}) \cdot \nabla \log \rho d\mathbf{x} \\ &= -\frac{1}{2} \int \rho \mathbf{G} : \nabla \nabla \log \rho d\mathbf{x} = -\frac{1}{2} E(\mathbf{G} : \nabla \nabla \log \rho), \end{aligned}$$

where integration by parts has been used. This yields Equation (11). If  $\mathbf{B}$  is a constant matrix,  $\mathbf{G}$  may be taken out of the expectation. Integrating by parts,

$$\begin{aligned} E(\mathbf{G} : \nabla \nabla \log \rho) &= \mathbf{G} : \int \rho (\nabla \nabla \log \rho) \, d\mathbf{x} \\ &= \mathbf{G} : \left( - \int \nabla \rho \nabla \log \rho \, d\mathbf{x} \right) \\ &= -\mathbf{G} : \int \rho \nabla \log \rho \nabla \log \rho \, d\mathbf{x} \\ &= -\mathbf{G} : E(\nabla \log \rho \nabla \log \rho) \\ &= -E(\nabla \log \rho \cdot \mathbf{G} \cdot \nabla \log \rho). \end{aligned}$$

and Equation (12) thus follows.  $\square$

In reality, a large portion of noise is additive. That is to say, the stochastic perturbation amplitude  $\mathbf{B}$  in Equation (9), and hence  $\mathbf{G}$ , is indeed a constant matrix. We thus may have more chance to use the formula (12). Notice that  $\mathbf{G} = \mathbf{B}\mathbf{B}^T$  is nonnegatively definite,  $\nabla \log \rho \cdot \mathbf{G} \cdot \nabla \log \rho \geq 0$ . That is to say, in this case, systems without phase volume expansion/contraction in the deterministic limit (such as the Hamiltonian system), stochasticity always functions to increase absolute entropy.

It is interesting to note that the above formula (12) may be linked to Fisher information if the parameters, say  $\mu_i$ , of the distribution are bound to the state variables in a form of translation such as that in a Gaussian process. In this case, one can replace the partial derivatives with respect to  $x_i$  by that with respect to  $\mu_i$ . And, accordingly,

$$E(\nabla \log \rho \nabla \log \rho) = \mathbf{I},$$

is the very Fisher information matrix. We hence have

**Corollary 1.** *In the stochastic system (9), if  $\mathbf{B}$  is constant, and if  $\mathbf{x}$  is bound to some parameter vector  $\boldsymbol{\mu}$  in a translation form, then*

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) + \frac{1}{2} (\mathbf{B}\mathbf{B}^T) : \mathbf{I}, \tag{14}$$

where  $\mathbf{I}$  is the Fisher information matrix with respect to  $\boldsymbol{\mu}$ .

### 3.2. Relative Entropy

**Theorem 4.** *For system (9), let the probability density function of  $\mathbf{x}$  be  $\rho$ , and let  $q$  be a reference density. Then the evolution of the relative entropy  $D(\rho||q)$  follows*

$$\frac{dD}{dt} = -\frac{1}{2} E \left[ \nabla \left( \log \frac{\rho}{q} \right) \cdot \mathbf{G} \cdot \nabla \left( \log \frac{\rho}{q} \right) \right], \tag{15}$$

where  $\mathbf{G} = \mathbf{B}\mathbf{B}^T$ .

Remark: A similar formula was stated in [30] under a positivity assumption. Here the assumption is relaxed.

**Proof.** We have shown before that the evolution of density  $\rho$  is governed by the Fokker–Planck Equation (10). For the reference density  $q$ , it is also governed by a Fokker–Planck equation, which reads

$$\frac{\partial q}{\partial t} = -\nabla \cdot (q\mathbf{F}) + \frac{1}{2}\nabla\nabla : (q\mathbf{G}). \tag{16}$$

Substituting Equations (10) and (16) into the identity

$$\frac{\partial(\rho \log q)}{\partial t} = \frac{\partial \rho}{\partial t} \log q + \frac{\rho}{q} \frac{\partial q}{\partial t}$$

for  $\frac{\partial \rho}{\partial t}$  and  $\frac{\partial q}{\partial t}$ , and then integrating over  $\mathbb{R}^n$ , we get

$$\begin{aligned} -\frac{d}{dt} \int_{\Omega} \rho \log q \, d\mathbf{x} &= - \int \left( \frac{\partial \rho}{\partial t} \log q + \frac{\rho}{q} \frac{\partial q}{\partial t} \right) d\mathbf{x} \\ &= - \int \log q \left[ -\nabla \cdot (\rho\mathbf{F}) + \frac{1}{2}\nabla\nabla : (\rho\mathbf{G}) \right] d\mathbf{x} - \int \frac{\rho}{q} \left[ -\nabla \cdot (q\mathbf{F}) + \frac{1}{2}\nabla\nabla : (q\mathbf{G}) \right] d\mathbf{x} \\ &= \int \left[ \log q \nabla \cdot (\rho\mathbf{F}) + \frac{\rho}{q} \cdot (q\mathbf{F}) \right] d\mathbf{x} - \frac{1}{2} \int \left[ \log q \nabla\nabla : (\rho\mathbf{G}) + \frac{\rho}{q} \nabla\nabla : (q\mathbf{G}) \right] d\mathbf{x} \\ &= \int \left[ \log q \nabla \cdot (\rho\mathbf{F}) + \nabla \log q \cdot \rho\mathbf{F} + \rho \nabla \cdot \mathbf{F} \right] d\mathbf{x} - \frac{1}{2} \int \left[ \log q \nabla\nabla : (\rho\mathbf{G}) + \frac{\rho}{q} \nabla\nabla : (q\mathbf{G}) \right] d\mathbf{x} \\ &= E(\nabla \cdot \mathbf{F}) - \frac{1}{2} \left[ \int \log q \nabla\nabla : (\rho\mathbf{G}) d\mathbf{x} + \int \frac{\rho}{q} \nabla\nabla : (q\mathbf{G}) d\mathbf{x} \right]. \end{aligned} \tag{17}$$

Subtracting Equation (13) from above gives the time evolution of the relative entropy:

$$\frac{dD}{dt} = \frac{1}{2} \int \left[ \left( \log \frac{\rho}{q} \right) \nabla\nabla : (\rho\mathbf{G}) + \nabla\nabla : (\rho\mathbf{G}) - \left( \frac{\rho}{q} \right) \nabla\nabla : (q\mathbf{G}) \right] d\mathbf{x}. \tag{18}$$

Integrating by parts, and using the compact support assumption, this becomes

$$\begin{aligned} \frac{dD}{dt} &= -\frac{1}{2} \int \left[ \nabla \log \frac{\rho}{q} \cdot \nabla \cdot (\rho\mathbf{G}) - \nabla \left( \frac{\rho}{q} \right) \cdot \nabla \cdot (q\mathbf{G}) \right] d\mathbf{x} \\ &= -\frac{1}{2} \int \left[ \frac{q}{\rho} \nabla \left( \frac{\rho}{q} \right) \cdot (\nabla \rho \cdot \mathbf{G} + \rho \nabla \cdot \mathbf{G}) - \nabla \left( \frac{\rho}{q} \right) \cdot (\nabla q \cdot \mathbf{G} + q \nabla \cdot \mathbf{G}) \right] d\mathbf{x} \\ &= \frac{1}{2} \int \frac{1}{\rho} \nabla \left( \frac{\rho}{q} \right) \cdot \mathbf{G} \cdot (\rho \nabla q - q \nabla \rho) d\mathbf{x} \\ &= \frac{1}{2} \int \rho \nabla \left( \frac{\rho}{q} \right) \cdot \mathbf{G} \cdot \nabla \left( \frac{q}{\rho} \right) d\mathbf{x} \\ &= \frac{1}{2} E \left[ \nabla \left( \frac{\rho}{q} \right) \cdot \mathbf{G} \cdot \nabla \left( \frac{q}{\rho} \right) \right] \\ &= \frac{1}{2} E \left[ \left( \frac{q}{\rho} \nabla \left( \frac{\rho}{q} \right) \right) \cdot \mathbf{G} \cdot \left( \frac{\rho}{q} \nabla \left( \frac{q}{\rho} \right) \right) \right] \\ &= \frac{1}{2} E \left[ \nabla \left( \log \frac{\rho}{q} \right) \cdot \mathbf{G} \cdot \nabla \left( \log \frac{q}{\rho} \right) \right] \\ &= -\frac{1}{2} E \left[ \nabla \left( \log \frac{\rho}{q} \right) \cdot \mathbf{G} \cdot \nabla \left( \log \frac{\rho}{q} \right) \right]. \quad \square \end{aligned} \tag{19}$$

Notice that, because of the nonnegative definiteness of  $\mathbf{G} = \mathbf{B}\mathbf{B}^T$ , the right hand side of Equation (15) is always smaller than or equal to zero. That is to say,  $D$  can never increase. This is what is shown

in Cover and Thomas [25] with a Markov chain, a property that has been connected to the second law of thermodynamics. (Notice the negative sign in the definition of  $D$ ; that is to say, increase in  $H$  corresponds to decrease in  $D$ ).

#### 4. Validation with Linear Systems

In this section, we re-derive the entropy evolution laws using a different approach for the linear version of Equation (9), *i.e.*,

$$dx = \mathbf{A}x + \mathbf{B}dw, \tag{20}$$

where  $\mathbf{x}$  is an  $n$ -vector,  $\mathbf{A} = (a_{ij})$  and  $\mathbf{B} = (b_{ij})$  are  $n \times n$  constant matrices. In this case, the vector field  $\mathbf{F} = \mathbf{A}x$  is in a linear form, and the noise is additive. The reason to specially pick Equation (20) for consideration is two-fold. Firstly, linear systems are important *per se*. Though seemingly highly idealized, they have been actually extensively exploited in applied sciences such as climate change for real problem studies. It would be of interest to find the evolution laws, hopefully in a simplified form, for this particular case; Secondly, linear systems preserve Gaussianity, allowing probability density functions to be easily solved in terms of means and covariances. Hence both  $H$  and  $D$ , and subsequently  $\frac{dH}{dt}$  and  $\frac{dD}{dt}$ , actually can be directly obtained. The resulting formulas will serve to validate what we have established before in the preceding sections.

**Theorem 5.** For the  $n$ -dimensional linear system (20),

$$H = \log \left[ (2\pi)^{n/2} (\det \mathbf{C})^{1/2} \right] + \frac{n}{2}, \tag{21}$$

$$\frac{dH}{dt} = \frac{d}{dt} \log \sqrt{\det \mathbf{C}} = \text{Tr}(\mathbf{A}) + \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : \mathbf{C}^{-1}, \tag{22}$$

where  $\mathbf{C}$  is the covariance matrix, and  $\text{Tr}(\mathbf{A})$  the trace of  $\mathbf{A}$ . In the absence of stochasticity, *i.e.*, when  $\mathbf{B} = \mathbf{0}$ ,

$$\frac{dH}{dt} = \text{Tr}(\mathbf{A}). \tag{23}$$

**Proof.** For system (20), if the state is initially a Gaussian, then it will be Gaussian forever. Thus it suffices to compute the mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{C} = (c_{ij})$  for the pdf evolution:

$$\frac{d\boldsymbol{\mu}}{dt} = \mathbf{A}\boldsymbol{\mu}, \tag{24}$$

$$\frac{d\mathbf{C}}{dt} = \mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T. \tag{25}$$

Solution of these equations determines the density

$$\rho(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} [\det \mathbf{C}]^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \tag{26}$$

The entropy is then

$$H = -E(\log \rho) = \log \left[ (2\pi)^{n/2} (\det \mathbf{C})^{1/2} \right] + \frac{1}{2} E \left[ (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right].$$

Notice that  $E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})]$  is just the covariance matrix  $\mathbf{C}$ , so

$$E[(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})] = \mathbf{C} : \mathbf{C}^{-1} = Tr(\mathbf{C}\mathbf{C}^{-1}) = n,$$

and hence

$$H = \log \left[ (2\pi)^{n/2} (\det \mathbf{C})^{1/2} \right] + \frac{n}{2}.$$

Differentiating,

$$\frac{dH}{dt} = \frac{d}{dt} \log \left[ (2\pi)^{n/2} (\det \mathbf{C})^{1/2} \right] = \frac{d}{dt} \log \sqrt{\det \mathbf{C}}.$$

To derive Equation (22), observe that, by the definition of determinant,

$$\det \mathbf{C} = \sum_{\sigma \in P_n} \text{sgn}(\sigma) \prod_{i=1}^n c_{i,\sigma_i}$$

where  $P_n$  is the totality of permutations of the set  $\{1, 2, \dots, n\}$ ,  $\sigma$  a permutation with  $\sigma_i$  being the  $i^{\text{th}}$  element, and  $\text{sgn}(\sigma)$  the signature of  $\sigma$ . Taking derivative with respect to time, we have

$$\frac{d}{dt} \det \mathbf{C} = \sum_{\sigma \in P_n} \text{sgn}(\sigma) \left( \prod_{i=1}^n c_{i,\sigma_i} \right) \left( \sum_{j=1}^n \frac{1}{c_{j,\sigma_j}} \frac{dc_{j,\sigma_j}}{dt} \right).$$

Here  $c_{j\sigma_j}$  has to be nonzero to legitimize the fraction expression. This is, of course, not true in general. But when it is zero, the whole sub-term is zero and hence disappears in the summation. So we need only consider this nonzero case. By Equation (25),

$$\begin{aligned} \frac{d}{dt} \det \mathbf{C} &= \sum_{\sigma \in P_n} \text{sgn}(\sigma) \left( \prod_{i=1}^n c_{i,\sigma_i} \right) \left[ \sum_j \frac{1}{c_{j,\sigma_j}} \sum_k (a_{j,k} c_{k,\sigma_j} + c_{j,k} a_{\sigma_j,k} + b_{jk} b_{\sigma_j,k}) \right] \\ &= \sum_{\sigma} \text{sgn}(\sigma) \prod_i c_{i,\sigma_i} \cdot \sum_j \left[ \frac{1}{c_{j,\sigma_j}} (a_{j,j} c_{j,\sigma_j} + c_{j,\sigma_j} a_{\sigma_j,\sigma_j}) + \frac{1}{c_{j,\sigma_j}} \left( \sum_{k \neq j} a_{j,k} c_{k,\sigma_j} + \sum_{k \neq \sigma_j} c_{j,k} a_{\sigma_j,k} \right) \right] \\ &\quad + \sum_{\sigma} \text{sgn}(\sigma) \prod_i c_{i,\sigma_i} \sum_j \frac{1}{c_{j,\sigma_j}} b_{jk} b_{\sigma_j,k} \end{aligned}$$

notice that within the square bracket (of the first summation over  $\sigma$ ), the first part is simply  $2a_{jj}$ . in the summations of the second part, each term must have exactly one element of  $\mathbf{C}$  that repeats a multiplier within the product  $\text{sgn}(\sigma) \prod_i c_{i,\sigma_i}$ , and hence the alternating sign summation over  $\sigma$  vanishes. (This means two rows/columns are the same in a matrix, and hence the determinant is zero). For the second summation over  $\sigma$ , it is the adjoint or adjugate (*i.e.*, the transpose of the cofactor matrix) of  $\mathbf{C}$  double dot with  $\mathbf{B}\mathbf{B}^T$ . So

$$\frac{d}{dt} \det \mathbf{C} = \det \mathbf{C} \cdot 2 \sum_j a_{jj} + \mathbf{B}\mathbf{B}^T : \text{Adj}(\mathbf{C}) = 2Tr(\mathbf{A}) \cdot \det \mathbf{C} + \mathbf{B}\mathbf{B}^T : \mathbf{C}^{-1} \cdot \det \mathbf{C}.$$

Therefore, for a linear system, its entropy change is

$$\frac{dH}{dt} = \frac{d}{dt} \log \sqrt{\det \mathbf{C}} = Tr(\mathbf{A}) + \frac{1}{2} \mathbf{B}\mathbf{B}^T : \mathbf{C}^{-1}.$$

Particularly, if the system is deterministic, *i.e.*, if  $\mathbf{B} = \mathbf{0}$ , it is simply

$$\frac{dH}{dt} = Tr(\mathbf{A}). \quad \square$$

Equation (23) tells that, for a deterministic linear system, the time rate of change of absolute entropy can be found without solving the system, and, moreover, it is a constant—the trace of  $\mathbf{A}$ . This is precisely what one would expect from the entropy evolution formula (5). The latter is hence verified.

Now turn to relative entropy. We need to verify the formula (15) in general, and Equation (8) in particular.

**Theorem 6.** Consider the  $n$ -dimensional linear system (20). Suppose there is a Gaussian distribution  $\rho$ , with  $\boldsymbol{\mu}$  and  $\mathbf{C}$  being its mean vector and covariance matrix, respectively. Also suppose there is a reference Gaussian distribution  $q$  with  $\boldsymbol{\mu}_q$  and  $\mathbf{C}_q$  as its mean vector and covariance matrix. Then, the relative entropy with respect to  $q$  is

$$D = \frac{1}{2} \log \left( \frac{\det \mathbf{C}_q}{\det \mathbf{C}} \right) - \frac{n}{2} + \frac{1}{2} \mathbf{C} : \mathbf{C}_q^{-1} + \frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_q), \quad (27)$$

and

$$\frac{dD}{dt} = (\mathbf{B}\mathbf{B}^T) : \mathbf{C}_q^{-1} - \frac{1}{2} (\mathbf{B}\mathbf{B}^T) : (\mathbf{C}^{-1} + \mathbf{C}_q^{-1} \mathbf{C} \mathbf{C}_q^{-1}) - \frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_q). \quad (28)$$

In particular, when  $\mathbf{B} = \mathbf{0}$ ,  $\frac{dD}{dt} = 0$ .

To prove, we need a lemma about the double dot.

**Lemma 1.**

- (1)  $\mathbf{A} : (\mathbf{B}\mathbf{C}) = (\mathbf{C}\mathbf{A}) : \mathbf{B}$ ,
- (2)  $\mathbf{A} : \mathbf{B} = Tr(\mathbf{B}\mathbf{A})$ .

**Proof.** For (1), the left hand side is  $\sum_i \sum_j a_{ij} (\sum_k b_{jk} c_{ki})$ , while

$$r.h.s. = \sum_m \sum_n (\sum_l c_{ml} a_{ln}) b_{nm} = \sum_l \sum_n a_{ln} \sum_m c_{ml} b_{nm} = \sum_l \sum_n a_{ln} \sum_m b_{nm} c_{ml}$$

which precisely equals the left hand side. Lemma (2) can be obtained from (1). In fact,

$$\mathbf{A} : \mathbf{B} = \mathbf{A} : (\mathbf{I}\mathbf{B}) = (\mathbf{B}\mathbf{A}) : \mathbf{I} = Tr(\mathbf{B}\mathbf{A}).$$

( $\mathbf{I}$  is the identity matrix.)  $\square$

**Proof of the theorem.**

By definition,

$$\begin{aligned} D(\rho||q) &= E_\rho \log \left\{ \frac{(2\pi)^{3/2} (\det \mathbf{C}_q)^{1/2}}{(2\pi)^{3/2} (\det \mathbf{C})^{1/2}} e^{-\frac{1}{2} [(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x}-\boldsymbol{\mu}) - (\mathbf{x}-\boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} (\mathbf{x}-\boldsymbol{\mu}_q)]} \right\} \\ &= \log \left[ \frac{\det \mathbf{C}_q}{\det \mathbf{C}} \right]^{1/2} - \frac{1}{2} E_\rho [(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \boldsymbol{\mu}) - (\mathbf{x} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} (\mathbf{x} - \boldsymbol{\mu}_q)]. \end{aligned}$$

We have shown that  $E_\rho [(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})] = n$ . In the same way, it is easy to see

$$E_\rho [(\mathbf{x} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1}(\mathbf{x} - \boldsymbol{\mu}_q)] = \mathbf{C} : \mathbf{C}_q^{-1} + (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_q);$$

Equation (27) thus follows.

To obtain  $\frac{dD}{dt}$ , we need to take time derivatives of the first, third, and fourth terms of the right hand side of Equation (27). (The second term is a constant.) Denote these terms as (I), (III), and (IV), respectively. By what we have established in the previous theorem,

$$\begin{aligned} \frac{d(I)}{dt} &= Tr(\mathbf{A}) + \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : \mathbf{C}_q^{-1} - Tr(\mathbf{A}) - \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : \mathbf{C}^{-1} \\ &= \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : (\mathbf{C}_q^{-1} - \mathbf{C}^{-1}). \end{aligned}$$

To find the derivatives of the third and fourth terms, we first need to find  $d\mathbf{C}_q^{-1}/dt$ . Since  $\mathbf{I} = \mathbf{C}_q \mathbf{C}_q^{-1}$ , we have, together with Equation (25),

$$0 = \mathbf{C}_q \frac{d\mathbf{C}_q^{-1}}{dt} + \frac{d\mathbf{C}_q}{dt} \mathbf{C}_q^{-1} = \mathbf{C}_q \frac{d\mathbf{C}_q^{-1}}{dt} + (\mathbf{A}\mathbf{C}_q + \mathbf{C}_q \mathbf{A}^T + \mathbf{B}\mathbf{B}^T) \mathbf{C}_q^{-1}$$

Solving for  $d\mathbf{C}_q^{-1}/dt$ ,

$$\frac{d\mathbf{C}_q^{-1}}{dt} = -\mathbf{C}_q^{-1} \mathbf{A} - \mathbf{A}^T \mathbf{C}_q^{-1} - \mathbf{C}_q^{-1} \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1}.$$

Thus

$$\begin{aligned} \frac{d(III)}{dt} &= \frac{1}{2} \frac{d}{dt} [\mathbf{C} : \mathbf{C}_q^{-1}] = \frac{1}{2} \frac{d\mathbf{C}}{dt} : \mathbf{C}_q^{-1} + \frac{1}{2} \mathbf{C} : \frac{d\mathbf{C}_q^{-1}}{dt} \\ &= \frac{1}{2}(\mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T) : \mathbf{C}_q^{-1} - \frac{1}{2} \mathbf{C} : (\mathbf{C}_q^{-1} \mathbf{A} + \mathbf{A}^T \mathbf{C}_q^{-1} + \mathbf{C}_q^{-1} \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1}) \\ &= \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : \mathbf{C}_q^{-1} - \frac{1}{2} \mathbf{C} : (\mathbf{C}_q^{-1} \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1}) \end{aligned}$$

where Lemma (1) has been used. Similarly,

$$\begin{aligned} \frac{d(IV)}{dt} &= \frac{1}{2} \left[ \left( \frac{d\boldsymbol{\mu}}{dt} - \frac{d\boldsymbol{\mu}_q}{dt} \right)^T \mathbf{C}_q^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_q) + (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \frac{d}{dt} \mathbf{C}_q^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_q) + (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} \left( \frac{d\boldsymbol{\mu}}{dt} - \frac{d\boldsymbol{\mu}_q}{dt} \right) \right] \\ &= \frac{1}{2} [(\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{A}^T \mathbf{C}_q^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_q) + (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} \mathbf{A}(\boldsymbol{\mu} - \boldsymbol{\mu}_q)] \\ &\quad - \frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T (\mathbf{C}^{-1} \mathbf{A} + \mathbf{A}^T \mathbf{C}_q^{-1} + \mathbf{C}_q^{-1}) \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_q) \\ &= -\frac{1}{2} (\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_q). \end{aligned}$$

Add these derivatives together and Equation (28) follows. In particular, when  $\mathbf{B} = 0$ ,

$$dD/dt = 0. \quad \square$$

As a concrete example, consider a 3D system with

$$\mathbf{A} = \begin{bmatrix} -\sigma & \sigma & 0 \\ \gamma - z_0 & -1 & -x_0 \\ y_0 & x_0 & -\beta \end{bmatrix}. \tag{29}$$

This is the linearized version of the Lorenz system around  $(x_0, y_0, z_0)^T$ . It is unstable, as we will study in the next section. Choose

$$\sigma = 10, \beta = 8/3, \gamma = 28,$$

and  $(x_0, y_0, z_0) = (1, 2, 0)$ . Let the stochastic perturbation amplitude be

$$\mathbf{B} = b \times \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \equiv b\tilde{\mathbf{B}}$$

with  $b$  a tunable parameter. When  $b = 0$ , the stochasticity is turned off. Initialize (24) and (25) with  $\boldsymbol{\mu} = (0, 1, 2)^T$ , and  $c_{11} = 4, c_{22} = 9, c_{33} = 16, c_{ij} = 0$  for all  $i \neq j$ . The solutions of  $\boldsymbol{\mu}$  and  $\mathbf{C}$  are shown in Figure 1a,b. Correspondingly the absolute entropy  $H$  is plotted in Figure 1c,d is the entropy evolution of the corresponding deterministic system, *i.e.*,  $b = 0$ . In this case, it is a straight line, with intercepts at  $t = 0$  and  $t = 0.1$  being 7.4335 and 6.0668, respectively, yielding a slope of

$$\frac{6.0668 - 7.4335}{0.1} = -13.667.$$

On the other hand,

$$\nabla \cdot \mathbf{F} = \text{Tr}(\mathbf{A}) = -(\sigma + 1 + \beta)$$

and

$$\nabla(\log \rho) = -\frac{1}{2} \nabla(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = -\mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

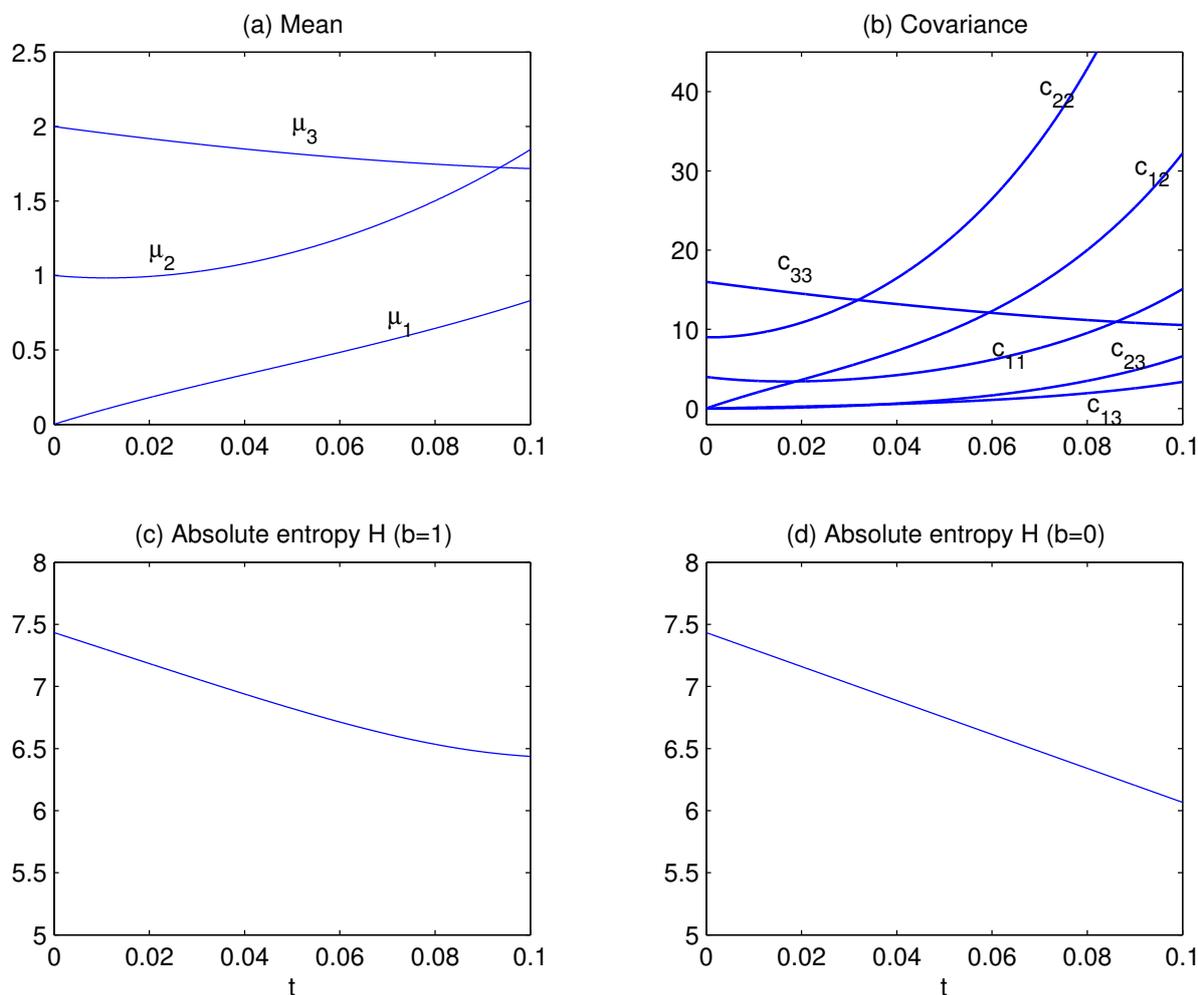
which gives  $\nabla[\nabla(\log \rho)] = -\mathbf{C}^{-1}$ . So, by the formula (11),

$$\frac{dH}{dt} = -(\sigma + 1 + \beta) + \frac{b^2}{2} (\tilde{\mathbf{B}}\tilde{\mathbf{B}}^T) : \mathbf{C}^{-1}.$$

With the computed covariances,  $H$  can be integrated out, and the result is the same as that in Figure 1c (not shown). Particularly, when  $b = 0$ , the slope of  $H$  is a constant, *i.e.*,  $-(\sigma + 1 + \beta) = 13.667$ , which is precisely what we obtained above in Figure 1d. With this linear system, the formula (11) and, particularly, Equation (5), is thence verified.

It should be noted that, in Figure 1d, the absolute entropy decreases linearly with time. Sooner or later it will hit the abscissa and become negative. This is a well-known fact for differential entropy, to which we limit our study. A simple example can illustrate this: Let  $x$  obey a uniform distribution on  $(0, \frac{1}{2})$ , then  $H = -\log 2 < 0$ .

**Figure 1.** The solution of Equations (24) and (25) with the parameters as specified in the text: **(a)** the means ( $b = 1$ ); **(b)** the covariances ( $b = 1$ ); **(c)** the absolute entropy  $H$  ( $b = 1$ ). **(d)**  $H$  versus time for the corresponding deterministic system ( $b = 0$ ); here  $H$  evolves as a straight line with a slope of  $-13.667$ .

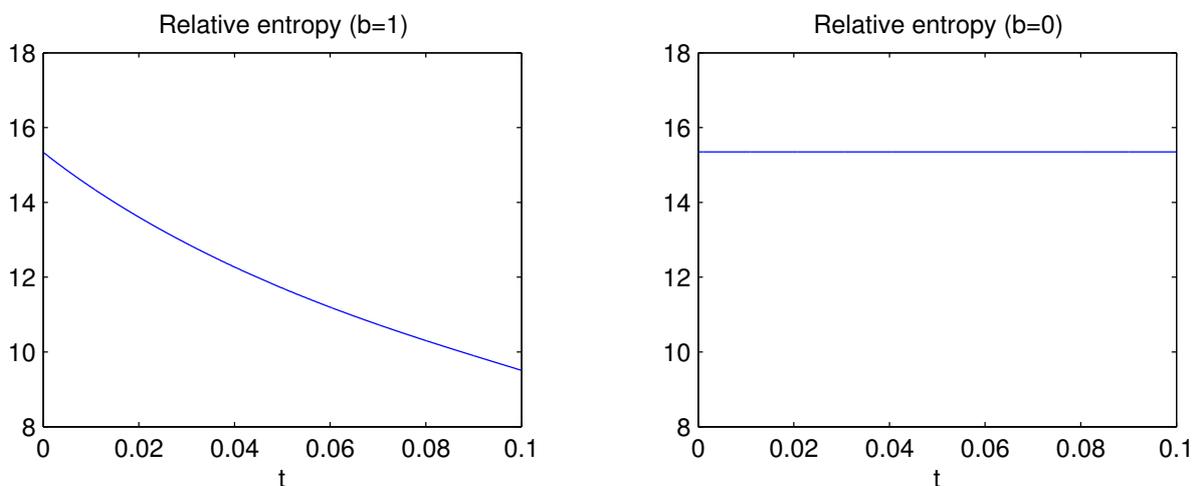


To see the relative entropy evolution, choose a mean vector and covariance matrix for the initial reference density:

$$\boldsymbol{\mu}_q(0) = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix}, \quad \mathbf{C}_q(0) = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then the relative entropy  $D$  evolution with time is shown as in Figure 2. Particularly, when the stochasticity is turned off, then  $D = 15.35$  is a constant. This is precisely what the law on relative evolution, viz., Equation (8), would state for any deterministic systems.

**Figure 2.** Relative entropy evolution with a stochastic system (**left**) and its corresponding deterministic system (**right**).



### 5. Lorenz System

We now turn to the Lorenz system,

$$\frac{dx_1}{dt} = \sigma(x_2 - x_1), \tag{30}$$

$$\frac{dx_2}{dt} = x_1(\rho - x_3) - x_2, \tag{31}$$

$$\frac{dx_3}{dt} = x_1x_2 - \beta x_3, \tag{32}$$

which has a vector field

$$\mathbf{F}(\mathbf{x}) = [\sigma(x_2 - x_1), x_1(\rho - x_3) - x_2, x_1x_2 - \beta x_3]^T.$$

We need to find its absolute entropy evolution. As normal, look at its density evolution first.

Different from the above linearized counterpart, the system is nonlinear and hence does not carry a probability density distribution known in form *a priori* as time moves on. To compute the density  $\rho$ , one naturally thinks of solving the Liouville Equation (4). This could be realized using a basis set or modes over its support, or numerical methods. An alternative but much more convenient and less expensive way is through *ensemble prediction*. As schematized in Figure 3, in order to predict  $\rho(t + \Delta t)$ , given that all information at  $t$  is known, an ensemble prediction does not seek the solution from the Liouville Equation (4); rather, it makes a detour through the steps shown in dashed box. Both these techniques namely the Liouville equation method and ensemble prediction method are equivalent, but the former is way more expensive in terms of computation.

Figure 4 is the trajectory of Equations (30)–(32) starting at  $(6, 5, 30)$ , with a time interval  $\Delta t = 0.01$ . The butterfly-like strange attractor demarcates the bounds for the 3D vector  $\mathbf{x}$ :  $-17 < x_1 < 22$ ,  $-22 < x_2 < 30$ ,  $0 < x_3 < 55$ . Thus it suffices to consider a sample space:

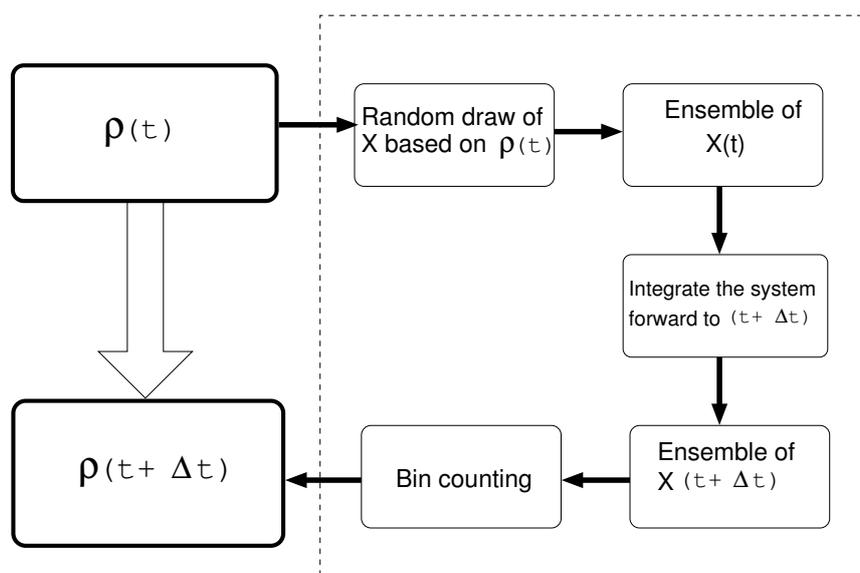
$$\Omega = [-25, 35] \times [-25, 35] \times [-2, 58]$$

for all the trajectories, at least after enough long time of integration. We now discretize the sample space equally, with 101 grid points in each dimension, forming  $100^3$  or one million cubic bins. Make random draws according to a normal distribution

$$N \left( \begin{bmatrix} 0 \\ 1 \\ 28 \end{bmatrix}, \begin{bmatrix} 4 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 16 \end{bmatrix} \right)$$

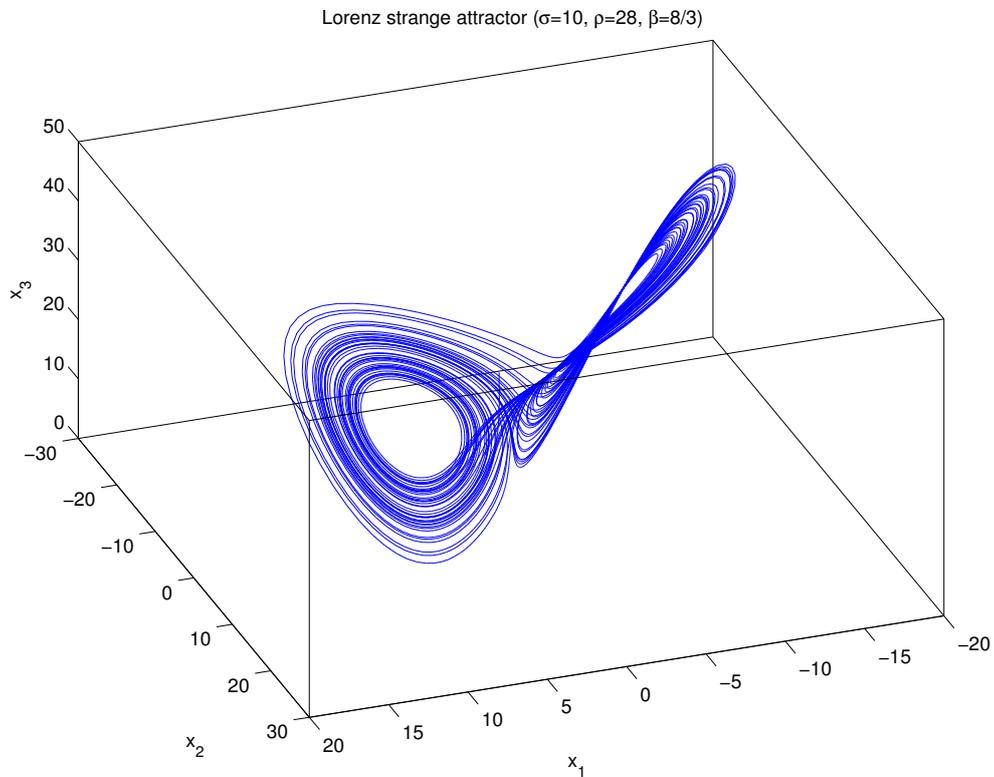
so as to ensure that there are 161 draws for each dimension (which totals more than four million draws). Just as in the preceding section, the means are not essential; we choose  $\mu = (0, 1, 28)^T$  for the convenience of computation, as it lies near the center of the attractor. We then integrate the system forward with a time step  $\Delta t = 0.01$ , and at each step count the bins with a population of  $161^3$  ensemble members. The resulting density distribution is then used to estimate the absolute entropy  $H$ .

**Figure 3.** Schematic of ensemble prediction. Given the density at  $t$ ,  $\rho(t)$ , one makes a detour as shown in the dashed line box to get the density at  $t + \Delta t$ , *i.e.*,  $\rho(t + \Delta t)$ , rather than obtains the latter directly from the Liouville equation.

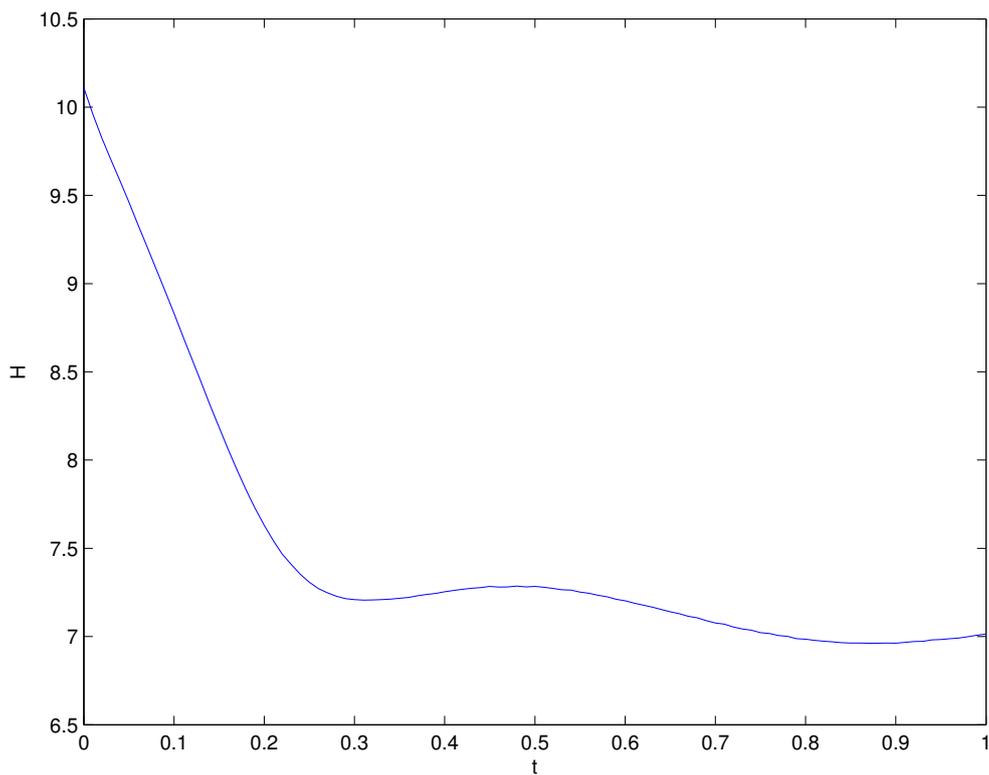


The computed  $H$  versus time is plotted in Figure 5 (thick solid line). Clearly, in the early stage of evolution, say, as  $t < 0.15$ ,  $H$  declines linearly, with a slope of approximately  $-14$ , the same as that in the linear model (see Figure 1d). But beyond that, the trend is gradually reversed and then oscillates afterwards. To see how resolution of the sample space may affect the results, we try a low resolution, with only 51 grid points in each dimension (keep the same number of draws). The result is plotted in the figure as thin dashed lines. Apparently, the linear trend is also seen, but broken earlier. In another experiment, we increase the resolution so that there are 141 grid points in each dimension and, correspondingly, increase the number of random draws to 181 for each dimension, which totals almost 6 million draws. The result, however, is almost the same as that in the standard experiment. It seems that the absolute entropy shown as the solid line in Figure 5 is what we can get through numerical computation.

**Figure 4.** Solution of the Lorenz system (30)–(32) with  $\Delta t = 0.01$ . It is initialized at  $\mathbf{x}_0 = (6, 5, 30)^T$ .



**Figure 5.** Evolution of absolute entropy obtained through ensemble prediction and pdf estimation.



The above result, however, is at odds with what Equation (5) predicts. In the preceding section the Lorenz equations are linearized and the corresponding joint absolute entropy  $H$  is found to decrease with time at a constant rate (cf. Figure 1d). In fact, this is a property of the Lorenz system:

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) = -(\sigma + 1 + \beta) = \text{const.} \quad (33)$$

Since the parameters  $\sigma$ ,  $\rho$ ,  $\beta$  are normally assumed to be positive (e.g.,  $\sigma = 10$ ,  $\rho = 28$ ,  $\beta = 8/3$ , the values originally chosen by Lorenz),  $H$  will be on the decline forever. So, what is wrong here with the computation?

As discussed in the previous section, if  $H$  keeps decreasing at the same rate, sooner or later it will cross the abscissa and become negative. A problem arises at this junction. While the differential absolute entropy  $H$  admits negative values (see the example of uniform distribution in the preceding section), its discrete counterpart does not. If  $P_i$  is the probability for state  $i$ , one has

$$-\sum_i P_i \log P_i \geq 0$$

(see Cover and Thomas, 1991). In discretizing the sample space for density estimation, one actually changes the problem of finding differential entropy into finding discrete entropy through bin counting. This is doomed to fail as the  $H$  approaches zero. As we have seen above, increasing the resolution of the sample space does help, either, since there is no way for discretized entropy to cross the abscissa! In this case, caution should be used when entropy is estimated using the technique of ensemble prediction and the subsequent bin-counting, particularly with real atmosphere-ocean problems, where this has been a widely accepted and efficient technique. This example tells us that, if absolute entropy  $H$  is computed, the concise formula (5) should be preferred.

## 6. Application to a Large-Dimensional Problem

We now consider a large-dimensional system, a fluid flow system. This kind of problems, which are described by partial differential equations (PDEs), are essentially of infinite size. But if we instead look at their numerical models, the dimensionality becomes finite, though very large in general. This section supplies a demonstration how the above formulas, namely, Equations (5), (8), (12), (14) and (15), may be applied to such a system. The purpose here is two-fold: The first is to provide a concrete example with technical details; the second is, in demonstrating the application, to look at how the uncertainty of the originally deterministic system is influenced after some stochasticity is added. For the former, we specifically need to

- Form the dynamical system;
- Find a way to initialize the system so as to get a chaotic attractor (only some particular pattern can destabilize the system);
- Drive the system toward a statistical equilibrium;
- Perform an EOF analysis to reduce the model to a low-dimensional system;
- Make random draws with the reduced model;
- Perform ensemble forecasts;

- Compute the entropy evolution using the above laws.

These will be shown in sequel in the following.

On the physics side, the motivation comes from the recent surge of interest in stochastic modeling of geophysical fluid flows—atmospheric and oceanic circulations. Realizing that model inaccuracy is inevitable (for example, it is impossible to accurately specify the boundary conditions in a weather forecast model), people have been turning their attention to stochastic PDEs, or SPDEs, with a stochastic term replacing the poorly represented physics in the governing equations (e.g., [33]). On the other hand, fluid flows are in general multiscale interactive. The atmospheric system, for example, contains millions of scales, from millimeters to tens of thousands kilometers, which cannot be all resolved in a computer model; the effect of the unresolved or subgrid processes must be parameterized for a reliable modeling. A promising approach to the parameterization problem is stochastic closure, which has been of great interest recently (e.g., [34,35]). All these concern of adding stochasticity to the original deterministic equations; it would be of interest to see how uncertainty may change when a system is changed from deterministic to stochastic.

Consider a flow with a model domain as schematized in Figure 6. For the sake of simplicity, the study is limited to a horizontal flow, with independent variables  $\xi$  and  $\eta$ . (We write this way, instead of the conventional  $(x, y)$ , to avoid confusion in notation.) Further, assume that the flow is quasi-geostrophic, so that only one state variable, namely, streamfunction, is needed (cf. [36,37]). Consider the perturbation from a zonal basic flow  $(U, V) = (U(\eta), 0)$ , and define a perturbation streamfunction  $\psi$  such that  $u = -\frac{\partial\psi}{\partial\eta}$ ,  $v = \frac{\partial\psi}{\partial\xi}$ , where  $(u, v)$  is the perturbation velocity. The SPDE of concern is

$$\frac{\partial\mathcal{L}\psi}{\partial t} = -\alpha \left( \frac{\partial\psi}{\partial\xi} \frac{\partial\mathcal{L}\psi}{\partial\eta} - \frac{\partial\psi}{\partial\eta} \frac{\partial\mathcal{L}\psi}{\partial\xi} \right) - U \frac{\partial\mathcal{L}\psi}{\partial\xi} - (\beta - U_{\eta\eta}) \frac{\partial\psi}{\partial\xi} + \mathcal{K}h(\xi, \eta)\dot{w}. \tag{34}$$

Here  $\mathcal{L}$  is the Laplacian operator:  $\mathcal{L} = \partial^2/\partial\xi^2 + \partial^2/\partial\eta^2$ ,  $\mathcal{K}$  a preset linear operator,  $h$  a given function of  $(\xi, \eta)$ , and  $\dot{w}$  white noise with respect to some given filtration  $\{\mathcal{F}_t\}_{t \geq 0}$ . The boundary conditions are such that at  $\eta = \pm 1$ ,  $\psi = 0$  (slip boundaries), and  $\xi = 0$  and  $\xi = 10$  are periodic. For simplicity,  $h$  is taken to be a constant. In this section two types of stochasticity are considered according to the operator  $\mathcal{K}$ : (1)  $\mathcal{K} = \mathcal{L}^{-1}$ ; (2)  $\mathcal{K} = 1$ . The first means that stochasticity applies to the streamfunction at each location, while in the second case stochasticity comes in with the vorticity equation at each location. In forming the discretized version large-dimensional system, the former results in a diagonal matrix  $\mathbf{B} = h\mathbf{E}$  in Equation (9); the latter leaves a  $\mathbf{B} = h\mathbf{L}^{-1}$ , where  $\mathbf{L}$  is the discretized version of the Laplacian  $\mathcal{L}$ .

The deterministic counterpart of the above SPDE is the barotropic quasi-geostrophic perturbation evolution equation in the absence of external forcing and friction. It and can be found in, say [37]. Its discretization is referred to [38], Appendix B.

We choose the shear instability model for our demonstration purpose. This extensively studied model has a velocity profile with a constant shear in the neighborhood of the axis, and zero elsewhere. Specifically, let

$$U(\eta) = \begin{cases} 1, & \eta > 0.15 \\ -1 + \frac{20}{3}(\eta + 0.15), & -0.15 \leq \eta \leq 0.15 \\ -1, & \eta < -0.15 \end{cases}$$

With this background flow we investigate the entropy evolution for the system (34). The forming of the large-dimensional system follows what we did in [38]. First, the domain as shown in Figure 6 is discretized into a grid with  $80 \times 40$  cells; next Equation (34), together with the boundary conditions, is discretized on the grid using the central difference scheme. Invert the Laplacian and we then obtain the system. Specifically, Equation (34) is now replaced by its difference form:

$$\begin{aligned}
 L \frac{d\psi_{i,j}}{dt} &= -\alpha \left( \frac{\psi_{i+1,j} - \psi_{i-1,j}}{2\Delta\xi} \times \frac{(\mathcal{L}\psi)_{i,j+1} - (\mathcal{L}\psi)_{i,j-1}}{2\Delta\eta} - \frac{\psi_{i,j+1} - \psi_{i,j-1}}{2\Delta\eta} \times \frac{(\mathcal{L}\psi)_{i+1,j} - (\mathcal{L}\psi)_{i-1,j}}{2\Delta\xi} \right) \\
 &\quad - U_j \times \frac{(\mathcal{L}\psi)_{i+1,j} - (\mathcal{L}\psi)_{i-1,j}}{2\Delta\xi} - (\beta - U_{\eta,j}) \times \frac{\psi_{i+1,j} - \psi_{i-1,j}}{2\Delta\xi} + Kh_{i,j}\dot{w}_{i,j} \\
 &\equiv N_{i,j} + Kh_{i,j}\dot{w}_{i,j},
 \end{aligned}
 \tag{35}$$

where

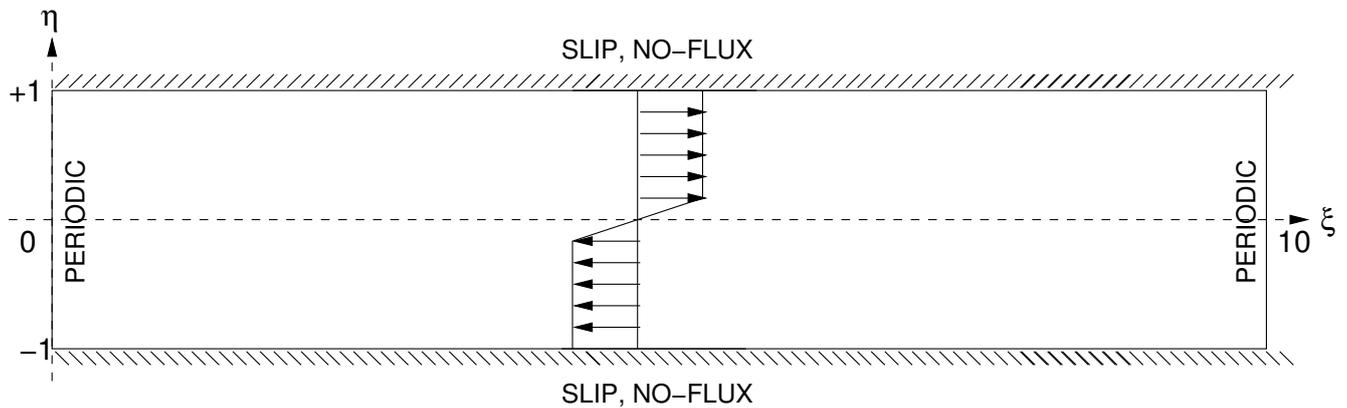
$$(\mathcal{L}\psi)_{i,j} = \frac{\psi_{i+1,j} - 2\psi_{i,j} + \psi_{i-1,j}}{\Delta\xi^2} + \frac{\psi_{i,j+1} - 2\psi_{i,j} + \psi_{i,j-1}}{\Delta\eta^2},$$

$N_{i,j}$  is a nonlinear function of  $\psi$  and  $\mathcal{L}\psi$  at the five grid points  $(i + 1, j)$ ,  $(i, j)$ ,  $(i - 1, j)$ ,  $(i, j + 1)$ ,  $(i, j - 1)$ , and  $L$  and  $K$  are, respectively, the discretized forms of  $\mathcal{L}$  and  $\mathcal{K}$ . Left multiplication by  $L^{-1}$ , together with the boundary conditions, yields a stochastic dynamical system in the desired form:

$$\frac{d\psi_{i,j}}{dt} = L^{-1}N + L^{-1}Kh_{i,j}\dot{w}_{i,j}.
 \tag{36}$$

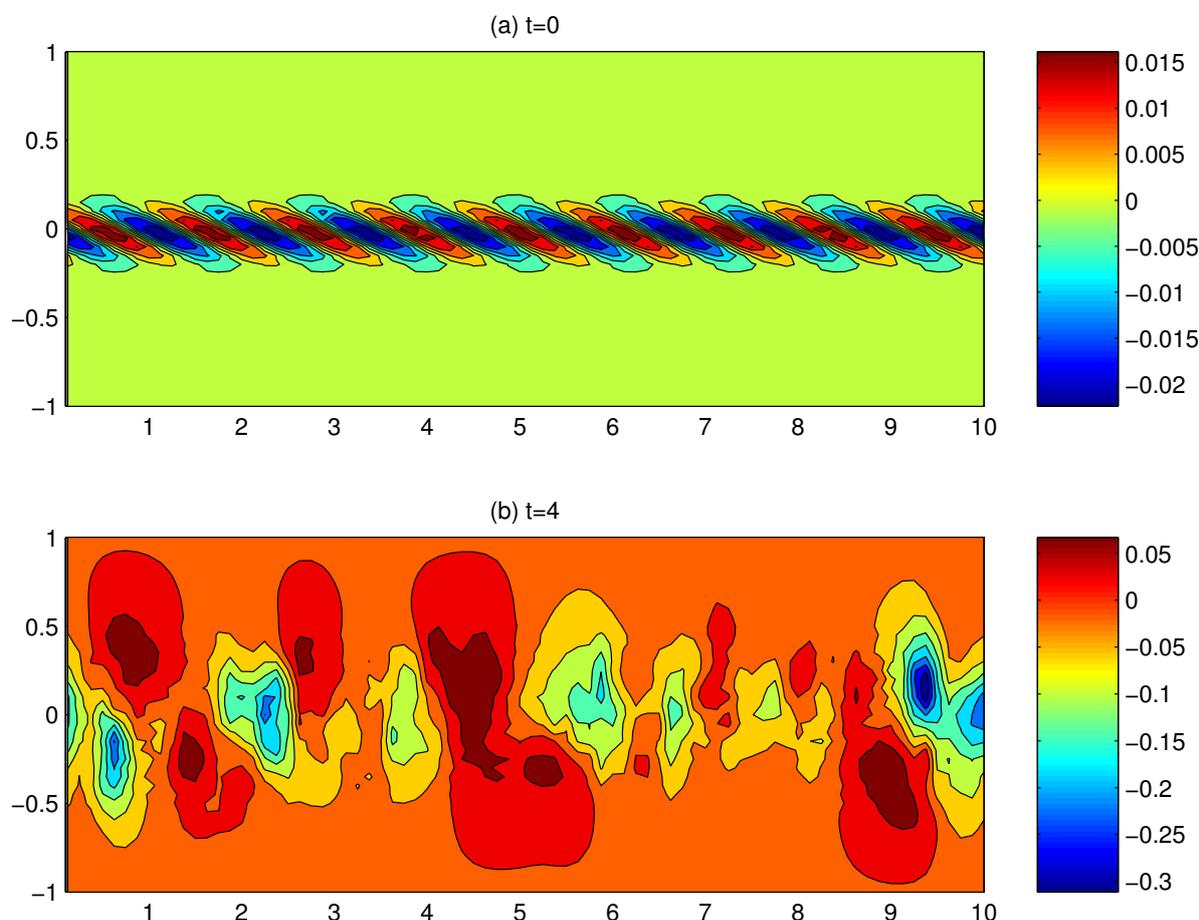
In this study, the two parameters  $\alpha = 2$  and  $\beta = 1$  are used.

**Figure 6.** Configuration of the shear instability model.



We look at how entropy evolves after the flow loses its stability and becomes chaotic. The first issue is how the integration should be initialized—different initializations will generally give different predictions, and not just any initialization may destabilize the flow. Here we use the optimal perturbation technique of [39] to find the optimal excitation mode (for details, see [38]). This mode is plotted in Figure 7a, where the perturbation amplitude is chosen to be 0.02. Once initialized, the deterministic model is integrated forward until it reaches a quasi-equilibrium, as shown in the chaotic state of Figure 7b.

**Figure 7.** Snapshots of  $\psi$ , the dimensionless perturbation streamfunction: (a) initial condition; (b) distribution at  $t = 4$ .



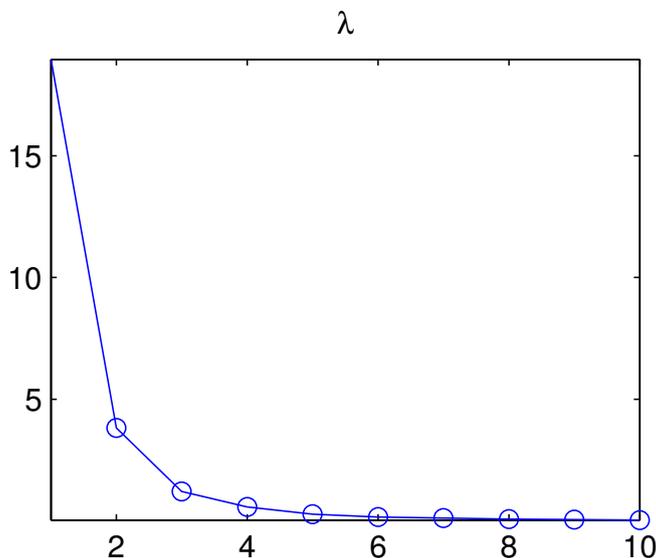
The second issue is how the ensemble may be generated. Different from the Lorenz system as shown before, the dimensionality is huge here, and the ensemble generation method cannot be applied directly. To solve this problem, we use the empirical orthogonal function (EOF) technique [40] to reduce the system size. Starting from the chaotic state shown in Figure 7b, the model is further integrated for 5000 steps (or a duration  $t = 1.25$  with time step  $\Delta t = 2.5 \times 10^{-4}$ ), to form a time sequence of the  $\psi$  distribution.

With the time sequence we perform the EOF analysis. Shown in Figure 8 are the variances associated with the modes. By calculation the first three account for 95% of the total variance; that is to say, the system essentially has a dimensionality of 3, rather than  $80 \times 40 = 3200$ . For safety, we choose 1 more mode to generate the ensemble. Shown in Figure 9 are these four modes.

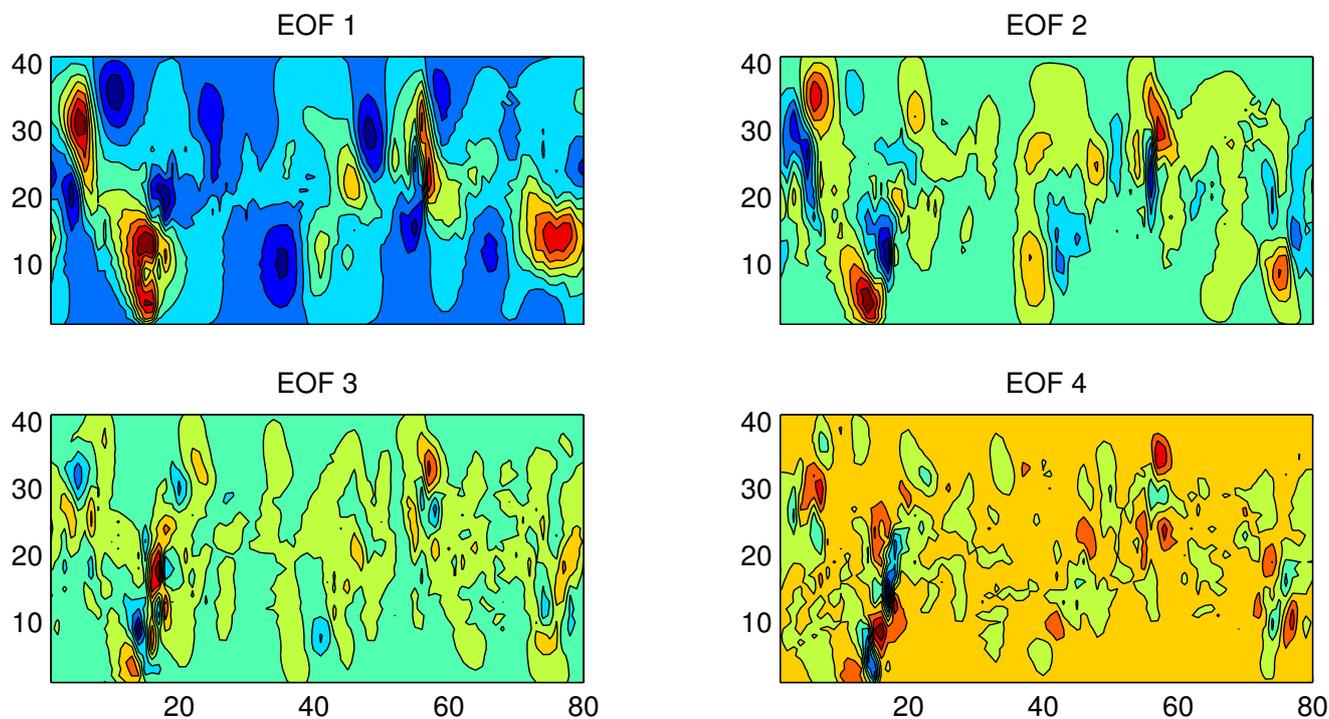
We make random draws according to a joint Gaussian distribution of the above four modes. We use the Gaussian distribution not just because of its simplicity, but also because of the observation that, in a large system after sufficiently large number of iterates, the resulting distribution tends to be a Gaussian, thanks to the central limit theorem. Besides, given mean and variance, a distribution that maximizes absolute entropy must be a Gaussian. For convenience, let initially the four modes be uncorrelated so that we may consider each dimension independently (for a Gaussian, uncorrelatedness is equivalent to independence). From the time-varying coefficients in the above EOF analysis, the means and variances  $(\mu_i, \sigma_i)$  for the

four modes are estimated as:  $(0, 3.8 \times 10^{-2})$ ,  $(0, 7.6 \times 10^{-3})$ ,  $(0, 2.4 \times 10^{-3})$ ,  $(0, 1.1 \times 10^{-3})$ , and the distributions are formed accordingly. We make five draws in each dimension; this yields an ensemble with a total of  $5^4 = 625$  members.

**Figure 8.** Variance versus EOF modal number.



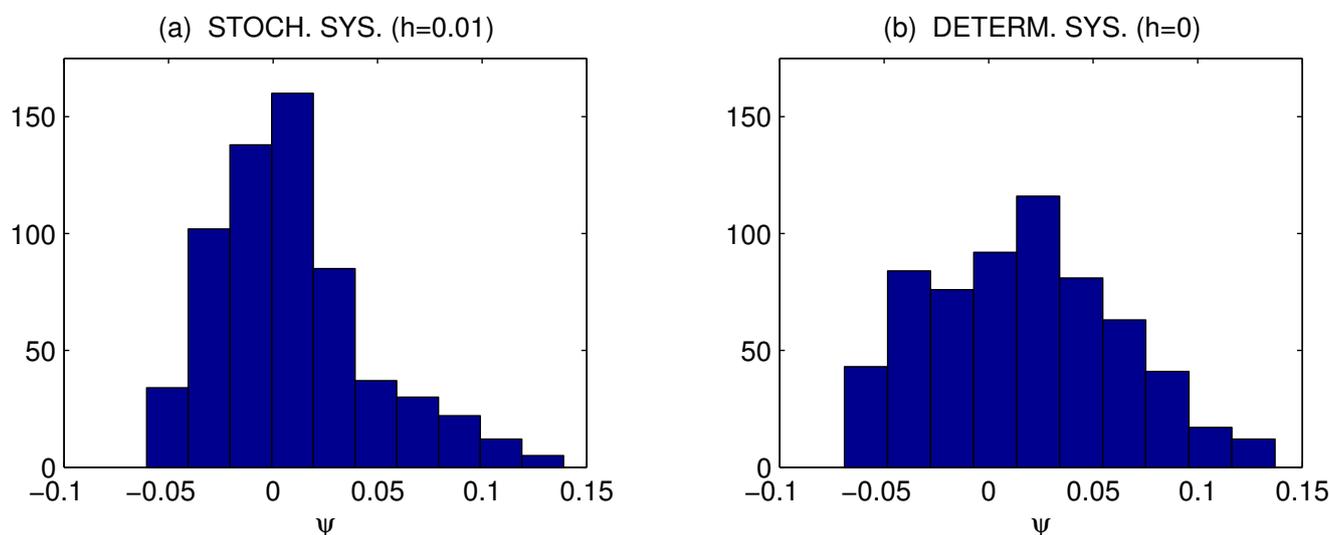
**Figure 9.** Structures of the first four EOF modes. These modes are normalized.



For each member of the ensemble, the state is steered forth by the stochastic system; for simplicity, the stochastic perturbation amplitude is set to be a constant  $h = 0.01$ . This results in an ensemble, and hence a distribution, at each time step. As an example, shown in Figure 10a is a snapshot of the

computed histogram of  $\psi$  at point (70, 25). An observation about the distributions: the spread growth is inhomogeneous in space; actually at some points it may not grow at all, though it is generally believed that, for a weather system, any prediction tends to deteriorate rapidly. In order to examine the impact of stochasticity perturbation, a model run with  $h = 0$  is launched. For comparison, we draw the histograms of  $\psi$  at some points; shown in Figure 10b is an example at (70, 25). The interesting thing here is, though stochasticity literally means uncertainty, at some points the spread of  $\psi$  with the stochastic system is slower than that with the corresponding deterministic system. Spread alone, of course, is not enough to characterize the uncertainty. We next estimate the entropy evolution. Note that the initial condition  $H(0) = \text{const}$  does not affect our study of the evolution of  $H(t)$ , this constant can be anything.

**Figure 10.** A snapshot of the histograms of the streamfunction at point (70, 25) predicted with the stochastic and deterministic systems: (a) stochastic system ( $h = 0.01$ ); (b) deterministic system ( $h = 0$ ).

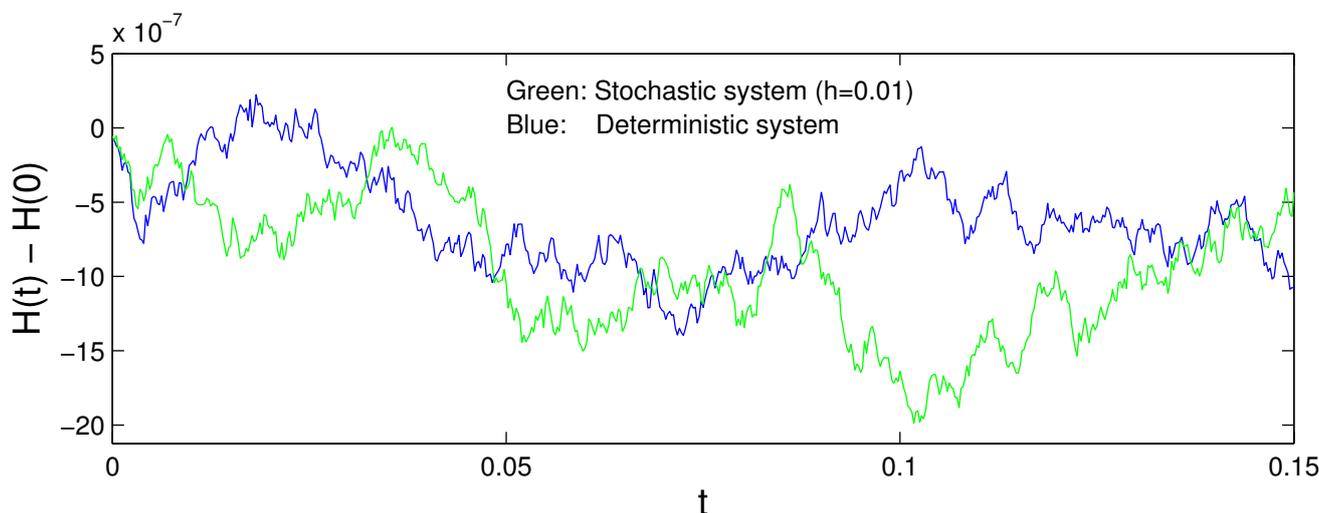


The computed entropy evolutions with the deterministic and stochastic systems are shown in Figure 11, where the blue curve stands for the case  $h = 0$ , and the green curve for  $h = 0.01$ . Different from its relative entropy counterpart, the absolute entropy  $H$  does not need to grow as stochasticity comes in. In the figure this is particularly clear around  $t = 0.1$ , where the entropy of the deterministic case attains a local maximum but that of the stochastic system is a minimum. Obviously, stochasticity does not always leave a system more uncertain.

The phenomenon of uncertainty reduction, albeit seemingly counterintuitive, actually can be understood from the formulas. By Equations (8) and (15), a deterministic system conserves its relative entropy  $D$ , and stochasticity reduces it. That is to say, stochasticity always causes a system to lose predictability. This is the natural result that one would expect. For absolute entropy  $H$ , one may also argue, based on a comparison of Equation (12) to (5), that stochasticity increases  $H$  and hence uncertainty. This, unfortunately, need not always be true. The issue here is, the part  $E(\nabla \cdot \mathbf{F})$  in the two formulas are not the same thing; when stochasticity is introduced,  $E(\nabla \cdot \mathbf{F})$  may vary accordingly—it may rise or fall. If it falls, the decrease could surpass the increase due to the stochastic process, resulting

in a deficit in absolute entropy, and hence a reduced uncertainty. This is what is shown in the above model result.

**Figure 11.** Evolution of absolute entropy  $H$  with the system (34) for  $h = 0.01$  (green) and  $h = 0$  (blue).



### 7. Discussion

We have presented a comprehensive introduction and a systematic derivation of the evolutionary laws for absolute entropy  $H$  and relative entropy  $D$  with respect to dynamical systems, both deterministic and stochastic, and explored applications to the renowned Lorenz system and a large-dimensional system namely a stochastic fluid flow problem. For easy reference, the resulting formulas are wrapped up in the following.

If a deterministic system

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) \tag{37}$$

is considered, its absolute entropy  $H$  and relative entropy  $D$  evolve as:

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}), \tag{38}$$

$$\frac{dD}{dt} = 0. \tag{39}$$

If the system has stochasticity included:

$$d\mathbf{x} = \mathbf{F}(\mathbf{x}, t)dt + \mathbf{B}(\mathbf{x}, t)d\mathbf{w}, \tag{40}$$

then

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) - \frac{1}{2}E(\mathbf{G} : \nabla \nabla \ell_\rho), \tag{41}$$

$$\frac{dD}{dt} = -\frac{1}{2}E[\nabla \ell_{\rho/q} \cdot \mathbf{G} \cdot \nabla \ell_{\rho/q}], \tag{42}$$

where  $\ell_\rho = \log \rho$ ,  $\ell_{\rho/q} = \log \rho/q$ , and  $\mathbf{G} = \mathbf{B}\mathbf{B}^T$ . In Equation (40), if  $\mathbf{B}$  has no dependence on  $\mathbf{x}$ , in other words, if the noise is additive, then

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) + \frac{1}{2}E(\nabla \ell_\rho \cdot \mathbf{G} \cdot \nabla \ell_\rho). \tag{43}$$

Moreover, if the parameters, say, the means  $\mu_i$ , of the pdf are bound to the state variables in a form of translation (as that in a Gaussian), then the above formula is further reduced to

$$\frac{dH}{dt} = E(\nabla \cdot \mathbf{F}) + \frac{1}{2}\mathbf{G} : \mathbf{I}, \tag{44}$$

where  $\mathbf{I} = (I_{ij})$  is the Fisher information matrix.

Among the above Equations (38), (39) and (41)–(44), Equations (41), (43) and (44) are new. Equation (42) is a generalization of that in [30] with their positivity assumption relaxed. Since a mathematical expectation may be well estimated by an ensemble mean, and there are many ways to estimate  $\mathbf{I}^{-1}$  the Cramér–Rao bound, Equation (44) enables us to avoid the expensive yet inaccurate pdf estimation in investigating the uncertainty evolution in large-dimensional systems.

For a linear system

$$d\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{B}dw, \tag{45}$$

the above formulas can be explicitly written out in terms of means and covariances. Here  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , and  $\mathbf{A}$  and  $\mathbf{B}$  are constant matrices. For such as system, if initially  $\mathbf{x}$  is Gaussian-distributed, it will keep being a Gaussian forever. Let the mean vector be  $\boldsymbol{\mu}$  and covariance matrix be  $\mathbf{C}$ . Then

$$H = \log \left[ (2\pi)^{n/2} (\det \mathbf{C})^{1/2} \right] + \frac{n}{2}. \tag{46}$$

$$D = \frac{1}{2} \log \left( \frac{\det \mathbf{C}_q}{\det \mathbf{C}} \right) - \frac{n}{2} + \frac{1}{2}\mathbf{C} : \mathbf{C}_q^{-1} + \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_q), \tag{47}$$

where  $\boldsymbol{\mu}_q$  and  $\mathbf{C}_q$  are the means and covariances of the reference distribution. Subsequently we have the time rate of change of  $H$  and  $D$

$$\frac{dH}{dt} = Tr(\mathbf{A}) + \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : \mathbf{C}^{-1}, \tag{48}$$

$$\frac{dD}{dt} = (\mathbf{B}\mathbf{B}^T) : \mathbf{C}_q^{-1} - \frac{1}{2}(\mathbf{B}\mathbf{B}^T) : (\mathbf{C}^{-1} + \mathbf{C}_q^{-1}\mathbf{C}\mathbf{C}_q^{-1}) - \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{\mu}_q)^T \mathbf{C}_q^{-1} \mathbf{B}\mathbf{B}^T \mathbf{C}_q^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_q). \tag{49}$$

Particularly, in the absence of stochasticity,  $\frac{dH}{dt}$  can be obtained without solving the governing equation or the equation for  $\boldsymbol{\mu}$  and  $\mathbf{C}$ :  $dH/dt = Tr(\mathbf{A})$ . In this case, the time rate of absolute entropy change is precisely equal to the trace of  $\mathbf{A}$ . For relative entropy, we recover the conservation law, *i.e.*,  $dD/dt = 0$ . These results have been tested with a specific linear system.

One may ask about how the result may comply to the second law of thermodynamics. As discussed extensively before [21,25], relative entropy is always dissipated and non-increasing, in accordance to the law; Equation (42) particularly shows that its dissipative representation in the equation has a form reminiscent of the Fisher information. The absolute entropy, in contrast, need not be increasing all the

time (see below for a specific example). It is a proved fact in probability theory that a distribution that maximizes absolute entropy unconditionally must be uniform. One may argue that, if initially a system has a uniform distribution, then it cannot change any more, otherwise the entropy will fall as time goes on, violating the above second law. This is, of course, not true in reality, and hence absolute entropy need not increase with a dynamical system. This, however, does not mean that the second law of thermodynamics is violated. What the second law states applies to isolated systems, excluding the possibility of phase space expansion or contraction. When the conformity between the law and Equation (41) is discussed, it would be more appropriate to limit the discussion to divergence-free systems like the Hamiltonian systems. (Recently, the second law is generalized for open systems; see the fluctuation theorem in [41].)

We have put the obtained evolutionary laws to application to the renowned Lorenz system, and a large-dimensional system. The Lorenz system, though highly nonlinear, actually has a constant vector field divergence, and the constant is negative. By Equation (38), this implies a constantly decreasing absolute entropy  $H$ . That is to say, on a plot of  $H$  versus time, sooner or later it will hit the abscissa and become negative, no matter how it is initialized. Though absolute entropy in the discrete sense is always above or equal to zero, a negative differential entropy is not uncommon; a simple example is a uniformly distributed random variable  $x$  on  $(0, \frac{1}{2})$ , which by calculation has an  $H = -\log 2 < 0$ . This makes a big issue in computing uncertainties in ensemble prediction problems (such as weather forecasting), where sample space coarse-graining is a common practice. In these computations, statistical properties are estimated through counting the bins on the discretized sample space—The widely used histogram plot drawing is just such an exercise. The resulting estimate of  $H$  is, therefore, a discrete one, and hence can never go below zero. In this study we showed that, the  $H$  estimated through bin counting is relevant only on the short interval before it approaches zero. Here the moral is: when dealing with highly dissipative systems such as the Lorenz system, caution should be used in estimating its absolute entropy.

Heuristically, if we imagine a dynamical system with a point attractor then, when the system is deterministic, its ensemble density will shrink to a delta function over the point attractor. During this process, absolute entropy will decrease at a constant rate. If we now add stochastic fluctuations, at some point the contraction of phase space to the point attractor and the dispersive effect of fluctuations will balance and the absolute entropy will asymptote to a (usually negative) constant value (at the steady state solution of the Fokker–Planck equation). On the other hand, if we consider the relative entropy, for deterministic systems this will be conserved as time progresses. The addition of random fluctuations will, in contrast, render the initial distributions indistinguishable, making the relative entropy fall to its lower bound of zero as the ensemble converges on the point attractor.

The application of the formulas requires ensemble prediction, and this is computationally tractable only for low-dimensional systems (as the above Lorenz system). To see how this may work for large-dimensional systems, we examined a stochastic quasi-geostrophic flow model. Stochastic modeling is hot these days in oceanography and atmospheric science, partially reflecting the effort to recover the unresolved small-scale processes and to remedy the model inaccuracies. In this study we particularly investigated through a shear-instability problem how the addition of stochasticity may change the uncertainty of the system which would otherwise be deterministic. Initialized with an optimal excitation perturbation, the pattern appears chaotic. To see how its uncertainty evolves, first we

reduced the model to a computationally tractable 3D problem, using the EOF analysis, then generated an ensemble with respect to the resulting EOF modes. The above formulas, after being applied to this ensemble and the vector field derived from the discretized system, yield an absolute entropy oscillating with time. An interesting observation is that, though stochasticity has almost been taken as a synonym of uncertainty, it does not necessarily leave a system more uncertain. In fact, sometimes introduction of noise may reduce the uncertainty of the system!

Uncertainty reduction due to noise addition occurs in a system when the reduction in phase volume expansion exceeds the uncertainty that the noise brings in. In Equation (41), that is to say, the reduction of  $E(\nabla \cdot \mathbf{F})$  exceeds the reduction of  $\frac{1}{2}E(\mathbf{G} : \nabla \nabla \log \ell_\rho)$ . For this reason, uncertainty reduction does not happen just under any circumstances. First, it is not permissible in linear systems. By Equation (48), a linear system always has a constant  $\nabla \cdot \mathbf{F}$ . For example, the most common mean reverting stochastic process, Ornstein–Uhlenbeck process,

$$dx = \lambda(\mu - x)dt + \sigma dw \quad (50)$$

( $\lambda > 0$ ,  $\sigma > 0$ , and  $\mu$  are parameters), has a constant divergence  $-\lambda$ . There is no way to have it decreased with the addition of  $\sigma dw$  in the equation. Clearly, uncertainty reduction can only occur in nonlinear systems.

Nonlinearity, however, is not a sufficient condition. The Lorenz system shown above is an example—albeit highly nonlinear,  $E(\nabla \cdot \mathbf{F})$  is always a constant. Another example is the Schwartz type 1 process, a popular stochastic process in finance:

$$dx = \kappa(\mu - \log x)xdt + \sigma xdw. \quad (51)$$

Here  $E(\nabla \cdot \mathbf{F}) = -\kappa E(\log x) - \kappa$ . This is not a constant, but depending on  $E(\log x)$ . However, the process is actually a log price Ornstein–Uhlenbeck process:

$$d \log x = \kappa(\mu - \log x)dt + \sigma dw,$$

and hence

$$\frac{dE(\log x)}{dt} = \kappa(\mu - \log x),$$

which gives an evolution of  $E(\log x)$  independent of  $\sigma$ . That is to say, stochasticity exerts no effect on  $E(\nabla \cdot \mathbf{F})$ ; in this case uncertainty always rises when stochasticity is included.

The Lorenz system and the quasi-geostrophic flow system are examples of self-organization, a process with which a system spontaneously moves from more probable states towards states of low probability, and through which the system reduces its absolute entropy, via exchanging information with the outside world (e.g., [42]). Both systems show that self-organization is manifested in phase space contraction. The flow instability problem particularly shows that the white noise from without may facilitate this process, leading to a reduction of uncertainty in comparison to the corresponding deterministic case. Obviously, for this to happen, nonlinearity is a necessary, albeit not sufficient, condition. But how nonlinearity and stochasticity work together to fulfill a reorganization of otherwise random states is not clear. This is, of course, beyond the scope of the present study, which is only for a demonstrating purpose. We will follow that up in future investigations.

## Acknowledgments

The comments from an anonymous referee are appreciated. This study was supported by Jiangsu Government through the “Jiangsu Specially-Appointed Professor Program” (Jiangsu Chair Professorship), and by the National Science Foundation of China (NSFC) under Grant No. 41276032.

## Conflicts of Interest

The author declares no conflict of interest.

## References

1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423.
2. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
3. Georgescu-Roegen, N. *The Entropy and the Economic Process*; Harvard University Press: Cambridge, MA, USA, 1971.
4. Lin, S.-K. Diversity and entropy. *Entropy* **1999**, *1*, 1–3.
5. Maes, C.; Redig, F.; Moffaert, J. On the definition of entropy production via examples. *Math. Phys.* **2000**, *41*, 1528.
6. Ruelle, D. How should one define entropy production for nonequilibrium quantum spin systems? *Math. Phys.* **2002**, *14*, 701–707.
7. Jakšić, V.; Pillet, C.-A. A note on the entropy production formula. *Contemp. Math.* **2003**, *327*, 175.
8. Gallavotti, G. Entropy production in nonequilibrium stationary states: A point of view. *Chaos* **2004**, *14*, 680–690.
9. Goldstein, S.; Lebowitz, J. On the (Boltzmann) entropy of nonequilibrium systems. *Physica D* **2004**, *224*, 53–66.
10. Tomé, T. Entropy production in nonequilibrium systems described by a Fokker–Planck equation. *Braz. J. Phys.* **2006**, *36*, 1285–1289.
11. Pavon, M.; Ticozzi, F. On entropy production for controlled Markovian evolution. *J. Math. Phys.* **2006**, *47*, 1–11.
12. Poletini, M. Fact-checking Ziegler’s maximum entropy production principle beyond the linear regimes and towards steady states. *Entropy* **2013**, *15*, 2570–2584.
13. Martyushev, L.M. Entropy and entropy production: Old misconceptions and new breakthroughs. *Entropy* **2013**, *15*, 1152–1170.
14. Lorenz, E. A study of the predictability of a 28-variable atmospheric model. *Tellus* **1965**, *17*, 321–333.
15. Epstein, E.S. Stochastic dynamic prediction. *Tellus* **1969**, *21*, 739–759.
16. Leith, C.E. Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.* **1974**, *102*, 409–418.
17. Ehrendorfer, E.; Tribbia, J. Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.* **1997**, *54*, 286–313.
18. Moore, A.M. The dynamics of error growth and predictability in a model of the Gulf Stream. II. Ensemble prediction. *J. Phys. Oceanogr.* **1999**, *29*, 762–778.

19. Schneider, T.; Griffies, S.M. A conceptual framework for predictability studies. *J. Clim.* **1999**, *12*, 3133–3155.
20. Palmer, T.N. Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.* **2000**, *63*, 71–116.
21. Kleeman, R. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.* **2002**, *59*, 2057–2072.
22. Kirwan, A.D.; Toner, M.; Kantha, L., Jr. Predictability, uncertainty, and hyperbolicity in the ocean. *Int. J. Eng. Sci.* **2003**, *41*, 249–258.
23. Lermusiaux, P.F.J. Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *J. Comput. Phys.* **2006**, *217*, 176–199.
24. Evangelinos, C.; Lermusiaux, P.F.J.; Xu, J.; Haley, P.J.; Hill, C.N. Many task computing for real-time uncertainty prediction and data assimilation in the ocean. *IEEE Trans. Parallel. Distr. Syst.* **2011**, *22*, 1012–1024.
25. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; Wiley: New York, NY, USA, 1991.
26. Andrey, L. The rate of entropy change in non-Hamiltonian systems. *Phys. Lett.* **1985**, *111A*, 45–46.
27. Liang, X.S.; Kleeman, R. Information transfer between dynamical system components. *Phys. Rev. Lett.* **2005**, *95*, 244101.
28. Plastino, A.R.; Daffertshofer, A. Liouville Dynamics and the conservation of information. *Phys. Rev. Lett.* **2004**, *93*, 138701.
29. Liang, X.S. Local predictability and information flow in complex dynamical systems. *Physica D* **2013**, *248*, 1–15.
30. Plastino, A.R.; Miller, H.G.; Plastino, A. Minimum Kullback entropy approach to the Fokker–Planck equation. *Phys. Rev. E* **1997**, *56*, 3927.
31. Liang, X.S.; Kleeman, R. A rigorous formalism of information transfer between dynamical system components. I. Discrete mapping. *Physica D* **2007**, *231*, 1–9.
32. Friston, K.; Ao, P. Free energy, value, and attractors. *Comput. Math. Methods Med.* **2012**, doi:10.1155/2012/937860.
33. Duan, J.; Gao, H.; Schmalfu, B. Stochastic dynamics of a coupled atmosphere-ocean model. *Stoch. Dyn.* **2002**, doi:10.1142/S0219493702000467.
34. Majda, A.J.; Timofeyev, I.; Vanden-Eijnden, E. A mathematical framework for stochastic climate models. *Commun. Pure Appl. Math.* **2001**, *54*, 891–974.
35. Mana, P.G.L.; Zanna, L. Stochastic parameterization of ocean mesoscale eddies. *Ocean Model.* **2014**, *79*, 1–20.
36. Pedlosky, J. *Geophysical Fluid Dynamics*, 2nd ed., Springer: New York, NY, USA, 1979; p. 710.
37. Vallis, G.K. *Atmospheric and Oceanic Fluid Dynamics*; Cambridge University Press: Cambridge, MA, USA, 2006; p. 745.
38. Liang, X.S. Uncertainty generation in deterministic flows: Theory and application with an atmospheric jet stream model. *Dyn. Atmos. Ocean.* **2011**, *52*, 51–79.
39. Farrell, B.F.; Ioannou, P.J. Generalized stability theory. Part I. Autonomous operators. *J. Atmos. Sci.* **1996**, *53*, 2025–2040.

40. Preisendorfer, R.W. *Principal Component Analysis in Meteorology and Oceanography*. Elsevier: Amsterdam, The Netherlands, 1988.
41. Evans, D.J. A non-equilibrium free energy theorem for deterministic systems. *Mol. Phys.* **2003**, *101*, 1551–1554.
42. Bar-Yam, Y. *Dynamics of Complex Systems*; Perseus Books: Reading, MA, USA, 1997; p. 848.

© 2014 by the author; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).