

The maximum entropy principle applied to a dynamical system proposed by Lorenz

W.T.M. Verkley^a and C.A. Severijns

Royal Netherlands Meteorological Institute, PO Box 201, 3730 AE De Bilt, The Netherlands

Received 17 July 2013 / Received in final form 18 November 2013

Published online 13 January 2014

© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract. Lorenz has proposed a dynamical system in two versions (I and II) that have both proved very useful as benchmark systems in geophysical fluid dynamics. In version I of the system, used in predictability and data-assimilation studies, the system's state vector is a periodic array of large-scale variables that represents an atmospheric field on a latitude circle. The system is driven by a constant forcing, is linearly damped and has a simple form of advection that causes the system to behave chaotically if the forcing is large enough. The present paper sets out to obtain the statistical properties of version I of Lorenz' system by applying the principle of maximum entropy. The principle of maximum entropy asserts that the system's probability density function should have maximal information entropy, constrained by information on the system's dynamics such as its average energy. Assuming that the system is in a statistically stationary state, the entropy is maximized using the system's average energy and zero averages of the first and higher order time-derivatives of the energy as constraints. It will be shown that the combination of the energy and its first order time-derivative leads to a rather accurate description of the marginal probability density function of individual variables. If the average second order time-derivative of the energy is used as well, also the correlations between the variables are reproduced. By leaving out the constraint on the average energy – so that no information is used other than statistical stationarity – it is shown that the principle of maximum entropy still yields acceptable results for moderate values of the forcing.

1 Introduction

The principle of maximum entropy guarantees that the probability density function of a system reproduces all the information given on the system but is maximally noncommittal otherwise. The principle has been put forward by Jaynes in a long series of publications, beginning with two studies on statistical mechanics ([1,2], reprinted in [3]). Although Jaynes has emphasized that the principle of maximum entropy extends to systems out of statistical equilibrium, the number of studies that have explored this possibility is relatively small. It is the purpose of the present paper to investigate the principle of maximum entropy in the context of a forced-dissipative dynamical system, in particular version I of a system proposed by Lorenz [4]. This system has proved to be very useful in predictability and data-assimilation studies and is generally considered as a benchmark system in geophysical fluid dynamics. As in statistical mechanics, we wish to predict the statistics of the system's variables (the 'microscopic' variables) on the basis of a few global averages that characterize the overall state of the system (the corresponding 'macroscopic' variables).

It will be assumed that the system has been left to itself long enough that the statistics has become stationary.

Ideally, the statistics is then given by a probability density function that is a stationary solution of the Liouville equation, i.e., an invariant measure. For systems of which the flow in phase space is nondivergent, such as the unforced and undamped systems of Hamiltonian mechanics¹, any nonnegative normalized and sufficiently smooth function of the system's conserved quantities is an invariant measure. As shown in reference [1], if the measure is required to reproduce a given value of the system's average energy, the principle of maximum entropy selects a unique member out of this infinity of possibilities: Gibbs' canonical ensemble. For systems that are forced and damped the flow in phase space is generally not nondivergent so that invariant measures are much more difficult to obtain. Closest to the ideal invariant measures are the Sinai-Ruelle-Bowen (SRB) measures discussed by Ruelle [5]. However, constructing these measures requires sophisticated mathematical techniques and may not be possible for every dynamical system. In the present paper we propose to deal with the invariance of the probability density function in a limited but pragmatic way: by using zero average values of the first and possibly higher-order time-derivatives of the energy as additional constraints in the maximization of entropy. The probability density functions thus obtained are

¹ In Hamiltonian mechanics nondivergence of the flow in phase space is implied by Liouville's theorem.

^a e-mail: verkley@knmi.nl

at best approximately invariant but, as will be shown, give quite satisfactory descriptions of the system's statistics.

An approach that can be regarded as an early application of the principle of maximum entropy – in the pragmatic spirit outlined above – has been pioneered by Burgers [6]. In a series of publications, reprinted in Nieuwstadt and Steketee [7], Burgers endeavoured to apply the techniques of statistical mechanics to the problem of turbulent flow. He expanded the flow field in terms of Fourier modes, introduced a phase space spanned by the corresponding Fourier components, divided the phase space in small compartments and searched for a statistical configuration which can be realized microscopically in the largest number of ways. Instead of prescribing the average energy – which would have led to the canonical ensemble of equilibrium statistical mechanics – he required that the forcing and dissipation should balance on average. Assuming that the damping is linear, so that the dissipation of the energy is quadratic in the variables, he obtained as a remarkable result that instead of the energy the dissipation of the energy is equipartitioned among the variables.

An important problem with Burgers' method is that the phase space of a fluid dynamical system is infinite dimensional. Depending on whether energy or a balance between forcing and dissipation is used as a constraint, as a result of equipartition either the energy or the dissipation diverges. Burgers proposed several solutions to this problem but none of these completely satisfied him and at some stage he abandoned the subject. Later, also Onsager [8] took up the challenge of applying statistical mechanics to turbulence; a detailed review of the scientific context in which this took place is given by Eyink and Sreenivasan [9]. Concerning two-dimensional turbulence, Onsager decided to approximate the vorticity field by a finite set of point vortices. This leads to a finite-dimensional Hamiltonian system of ordinary differential equations to which the formalism of equilibrium statistical mechanics can be applied without qualifications. It has led to many interesting results such as the formation of large vortices in two-dimensional flows, see the review by Kraichnan and Montgomery [10]. A more recent review of the statistical mechanics of two-dimensional flows, including many new developments since 1980, is given by Bouchet and Venaille [11]. In the latter review due attention is given to the work of Miller [12], Miller et al. [13] and Robert and Sommeria ([14,15]) who developed statistical mechanical methods to deal with the infinite dimensionality of fluid systems.

All these approaches assume, however, that forcing and dissipation can be ignored as a first approximation. The question whether this is a valid procedure in studying the long-time behaviour of even slightly dissipative fluid systems is still a fundamental one (see Frisch [16], p. 245). The advantage of Burgers' approach is that forcing and dissipation are incorporated from the start. In fact, they are essential ingredients of the formulation, a property that it shares with the inertial range theory of turbulence developed by Kraichnan [17] and Batchelor [18]. The study of Verkley and Lynch [19], in which Jaynes' principle of

maximum entropy is applied to a two-dimensional forced-dissipative model of the atmosphere, can be seen as a revisit of Burgers' approach. Representing the flow fields in terms of basis functions (spherical harmonics) and then truncating the representation, a finite dimensional model was formulated. By staying within this finite dimensional model the problem of infinite dissipation was avoided. As constraints in the maximization of entropy the energy and enstrophy as well as their decay rates (assumed zero in the statistically stationary state) were used. Spectra and mean flows were obtained from the resulting probability density function. When the two types of constraints were combined, the agreement with spectra and mean flows from numerical simulations was reasonably good.

The principle of maximum entropy was also used by Verkley [20] to formulate a parametrization of the average effect of the small-scale variables on the large-scale variables in version II of the model devised by Lorenz [4]. The large-scale variables in this version represent a meteorological field at equidistant points on a latitude circle. They are subject to a simple constant forcing, a linear damping and nonlinear advection. Coupled to each large-scale variable is a group of small-scale variables that obey similar dynamics and represent the effects of motions beyond the resolution of the large-scale variables. The coupling goes in both directions; the small-scale variables are being forced by the large-scale variables and the large-scale variables are being forced by the small-scale variables. To parametrize the latter effect, the maximum entropy principle was applied to the statistics of the small-scale variables, assuming that their statistics is stationary on the time-scale of the large-scale variables. The small-scale entropy was maximized under the constraint that the average time derivative of the small-scale variables' energy is zero. The resulting probability density function was then used to calculate the average effect of the small-scale variables on the large-scale variables, yielding a simple linear damping not unlike the damping that would have resulted from an empirical approach.

In all cases mentioned, the probability density function is a product of independent normal distributions in the different variables. For an individual variable this result compares reasonably well with numerically obtained probability density functions, both in terms of averages and in terms of variances. The predicted independence of the variables is, however, usually not in accord with numerical simulations. Now, it can be understood easily that the principle of maximum entropy will not lead to covariances between the variables if the mathematical expressions of the constraints do not involve cross-terms between the variables. Cross-terms are usually present in higher order time-derivatives of the energy and covariances will therefore result from the maximization of entropy if these are used as constraints. In the context of version I of the model devised by Lorenz [4], i.e., the version without small-scale variables, we will study the consequences of using the second-order time-derivative of the energy as an additional constraint in the maximization of

entropy. The present work aims at showing that this leads to quite realistic covariances between the variables.

In the next section we give a short summary of the maximum entropy principle, in a setting that is valid for any finite dimensional dynamical system. We then, in Section 3, introduce version I of the system proposed by Lorenz [4]. As an introduction to the dynamics of this system, we discuss the stability properties of the steady state in which all variables are equal to the forcing. We then show the results of a numerical simulation in a case for which the system behaves chaotically. This case will be used as a numerical reference in Section 4. In this section we apply the maximum entropy principle, first using as a constraint a given value of the average energy, then adding a zero average first-order time-derivative of the energy and, finally, adding a zero average second-order time-derivative of the energy. In the latter two cases it is also studied what the consequences are of deleting the energy constraint. The analytical and numerical results are placed in the perspective of smaller and higher values of the forcing in Section 5. The conclusions and an outlook are given in Section 6. Some of the mathematical detail is deferred to Appendices.

2 The maximum entropy principle

We consider a dynamical system described by a point $\mathbf{X} = (X_1, X_2, \dots, X_K)$ in a K -dimensional phase space. If we are uncertain as to the exact state of the system, then it is appropriate to represent the system by means of a probability density function $\mathcal{P}(\mathbf{X})$. The amount of uncertainty is quantified by the information entropy S_I , defined as:

$$S_I = - \int \mathcal{P}(\mathbf{X}) \log \frac{\mathcal{P}(\mathbf{X})}{\mathcal{M}(\mathbf{X})} dX.$$

Here the integral is over the whole phase space, dX denoting a K -dimensional volume element $dX_1 dX_2 \dots dX_K$. The measure $\mathcal{M}(\mathbf{X})$ embodies any a priori information that we have on the system. Its presence implies that the argument of the natural logarithm is dimensionless and ensures furthermore, as emphasized by Jaynes [21], that the information entropy S_I is independent of a coordinate transformation of \mathbf{X} . Such a coordinate transformation could result from a deterministic time-evolution of the dynamical system, implying that the information entropy S_I is independent of time in that case.

The principle of maximum entropy – see Jaynes [22] and Pressé et al. [23] for general introductions – states that the information entropy should be maximal under the constraints of normalization and any other information that is available. The normalization constraint reads:

$$\int \mathcal{P}(\mathbf{X}) dX = 1.$$

The other information will be assumed to have the form of given averages

$$\langle \mathcal{K}_l \rangle \equiv \int \mathcal{P}(\mathbf{X}) \mathcal{K}_l(\mathbf{X}) dX = \mathcal{K}_l^0, \quad (1)$$

where $\mathcal{K}_l(\mathbf{X})$ is a function of the state vector \mathbf{X} , $l = 1, 2, \dots, L$, the number of averages being L . The most important example is the energy $E(\mathbf{X})$, but in this paper we will pay particular attention to $dE(\mathbf{X})/dt$ and $d^2E(\mathbf{X})/dt^2$. We note that the constraints generally leave so much room for the probability density function that maximization of the information entropy S_I selects a member from a set that is usually infinitely large. The member that is selected, however, is special in that it reproduces the information on the system's dynamics and, at the same time, is as noncommittal as possible, i.e., has not more structure than the information justifies.

Using a set of Lagrange multipliers $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_L)$ to take into account the constraints above, the maximization of S_I leads to (see Verkley and Lynch [19], their Eq. (12))

$$\mathcal{P}(\mathbf{X}) = \frac{\mathcal{M}(\mathbf{X})}{\mathcal{Z}(\boldsymbol{\lambda})} e^{-\sum_{l=1}^L \lambda_l \mathcal{K}_l(\mathbf{X})}, \quad (2)$$

where the partition function $\mathcal{Z}(\boldsymbol{\lambda})$ is the normalization factor of the probability density function, given by:

$$\mathcal{Z}(\boldsymbol{\lambda}) = \int \mathcal{M}(\mathbf{X}) e^{-\sum_{l=1}^L \lambda_l \mathcal{K}_l(\mathbf{X})} dX. \quad (3)$$

It can be seen from the expression above that the maximum entropy probability density function reduces to the a priori measure if there are no additional constraints of the form (1); in this case all Lagrange multipliers are zero².

The Lagrange multipliers must be obtained from the values of the averages, i.e., from \mathcal{K}_l^0 , $l = 1, 2, \dots, L$. An important relation that can be used to relate the average values of \mathcal{K}_l to the Lagrange multipliers λ_l is the following:

$$\langle \mathcal{K}_l \rangle = - \frac{\partial \log \mathcal{Z}}{\partial \lambda_l}. \quad (4)$$

This identity can be derived straightforwardly from (3).

3 The system

The system to be studied is version I of a model that Lorenz [4] devised to investigate problems in atmospheric predictability, data-assimilation and parametrization. E.g., Lorenz and Emanuel [24] used this model to study observation strategies in data-assimilation. In the present paper we will use the same model to study the formalism of maximum entropy as a means of obtaining the system's statistics without full recourse to numerical integrations.

The model state $\mathbf{X} = (X_1, X_2, \dots, X_K)$ of the system represents a certain atmospheric field at K equidistant points on a latitude circle. The system is assumed to be periodic so that $X_{k+K} = X_{k-K} = X_k$ for all k , see the schematic in Figure 1.

² From now on, we skip the normalization condition in counting constraints as normalization is always applied and taken care of by the partition function.

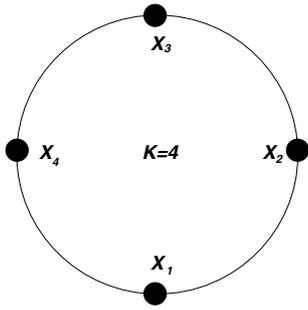


Fig. 1. A schematic of the system variables for $K = 4$, the minimum value to allow for chaotic behaviour of the system (5).

The equations that govern the system are:

$$\frac{dX_k}{dt} = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F, \quad (5)$$

where $k = 1, \dots, K$ and $K \geq 4$. The quadratic term in this equation represents advection, the linear term represents damping and the term F , assumed to be independent of k , is the forcing. In order to allow for chaotic behaviour, the dimension of the state space should be at least 4. The variables have been nondimensionalized in such a way that the coefficients in front of the advection and linear terms are 1; a nondimensional unit of time corresponds to 5 days.

The most important global quantity that characterizes the state of the system is the energy E , defined by:

$$E = \sum_{k=1}^K \frac{1}{2} X_k^2. \quad (6)$$

It can be checked that, due to the periodicity conditions, the quadratic term in the equations drop out if we calculate the first-order time-derivative of E . We thus have:

$$\begin{aligned} \frac{dE}{dt} &= \sum_{k=1}^K X_k \frac{dX_k}{dt} \\ &= \sum_{k=1}^K (-X_k^2 + X_k F) \\ &= -2E + F \sum_{k=1}^K X_k. \end{aligned} \quad (7)$$

For the second-order time-derivative of the energy we find, in a similar fashion,

$$\begin{aligned} \frac{d^2 E}{dt^2} &= -2 \frac{dE}{dt} + F \sum_{k=1}^K \frac{dX_k}{dt} \\ &= -2 \frac{dE}{dt} + F \sum_{k=1}^K (F - X_k + X_{k-1}(X_{k+1} - X_{k-2})) \\ &= -2 \frac{dE}{dt} + F \sum_{k=1}^K (F + X_k(X_{k+2} - X_{k+1} - 1)). \end{aligned} \quad (8)$$

Here we used the periodicity conditions to advance from the second to the third line in the expression above.

At this point it might be useful to note that both Lorenz [4] and Lorenz and Emanuel [24] do not attribute a specific physical meaning to the variables X_k . Instead, they remark in the latter reference (p. 400) that the model is formulated as “one of the simplest possible systems that treats all variables alike and shares certain properties with many atmospheric models, namely, (1) the nonlinear terms, intended to simulate advection, are quadratic and together conserve the total energy [...]; (2) the linear terms, representing mechanical or thermal dissipation, decrease the total energy; (3) the constant terms, representing external forcing, prevent the total energy from decaying to zero”. Be that as it may, it is to the credit of the system that, without forcing and damping, the system conserves energy, is time-reversal invariant and characterized by a nondivergent flow in phase space. As a result, the statistics of the unforced undamped version of the system is well described by equilibrium statistical mechanics, as shown at the end of Section 4.1.

A rather detailed analysis of the system’s dynamics in dependence of F , for $K = 4$, $K = 8$ and $K = 40$, is presented by Orrell and Smith [25] using a special type of bifurcation diagram. We will discuss in some detail the starting point of any analysis of such kind, namely the state \mathbf{X}_s of which all components are equal to F , i.e., $X_{sk} = F, k = 1, \dots, K$. This is a stationary solution (steady state) of the system (5). If we denote a small perturbation of X_{sk} by x_k and linearize the system (5) around this state, we see that the dynamics of these perturbations is governed by (see Lorenz and Emanuel [24], their Eq. (4))

$$\frac{dx_k}{dt} = F(x_{k+1} - x_{k-2}) - x_k.$$

Introducing a matrix \mathbf{A} with matrix elements A_{kl} ,

$$A_{kl} = -\delta_{kl} + F(\delta_{k+1,l} - \delta_{k-2,l}) \quad (9)$$

this can be written as:

$$\frac{dx_k}{dt} = \sum_{l=1}^K A_{kl} x_l,$$

where it is to be noted that the system is cyclic so that the index $k + K$ and $k - K$ denote the same variable as the index k . If all eigenvalues of the matrix \mathbf{A} have a real part that is negative, then all solutions of the linearized system decay exponentially in time and the steady state is linearly stable. If one or more eigenvalues of the matrix \mathbf{A} have a positive real part, then the steady state is linearly unstable. The special case that all eigenvalues have a negative real part, except for one or more of which the real part is zero, is characterized as neutrally stable.

If we write out the matrix \mathbf{A} , we see that each row has the same elements as the row above it but shifted one position to the right, elements disappearing on the right reappearing on the left due to the periodicity conditions.

Such matrices are called circulant; the theory of these matrices can be found in Davis [26], pp. 72-73. The matrix W of normalized eigenvectors, the columns of which denote the different eigenvectors, has elements given by:

$$W_{lm} = \frac{1}{\sqrt{K}} e^{-i\omega(l-1)(m-1)}, \quad (10)$$

where

$$\omega = \frac{2\pi}{K}. \quad (11)$$

Note that the structure of these eigenvectors is independent of the details of the matrix A but only depends on the number of variables K . The eigenvalues χ_m do depend on the details of the matrix A and are given by:

$$\chi_m = \sum_{n=1}^K A_{1n} e^{-i\omega(n-1)(m-1)}. \quad (12)$$

Using (9) and the periodicity conditions we find:

$$\chi_m = -(1 + Fz_m) - iFq_m,$$

with

$$z_m = \cos(2\omega(m-1)) - \cos(\omega(m-1)), \quad (13a)$$

$$q_m = \sin(2\omega(m-1)) + \sin(\omega(m-1)). \quad (13b)$$

From these expressions it may be concluded that the steady state is stable if, for all values of m , the quantity $1 + Fz_m$ is positive. Otherwise, the steady state is either unstable or neutrally stable. It can be checked straightforwardly that $W_{l,K+2-m} = W_{lm}^*$ and $\chi_{K+2-m} = \chi_m^*$. This implies that real-valued normal modes can be constructed by linearly combining eigenvectors with indices m and $K + 2 - m$.

Plots of z_m and q_m , for $K = 4$ (squares) and $K = 36$ (circles), are shown in the upper and lower panels of Figure 2, respectively. The solid curves are the functions $z(x) = \cos(2x) - \cos(x)$ (upper panel) and $q(x) = \sin(2x) + \sin(x)$ (lower panel), with x expressed in degrees. The function $z(x)$ has a minimum of $-9/8$ and a maximum of 2, as can be shown easily by differentiation. The minimum value of z_m is thus close to $-9/8$, while the maximum of z_m is close to 2; depending on the number of nodes K , the minimum is closer to $-9/8$ and the maximum closer to 2.

Denoting the minimum (negative) value by z_{min} and the maximum (positive) value by z_{max} it follows that the steady state is stable as long as the forcing F satisfies

$$\frac{1}{-z_{max}} < F < \frac{1}{-z_{min}}. \quad (14)$$

When the forcing F is outside this range, the steady state is either unstable or neutrally stable. For values of F slightly exceeding the maximum allowed value the flow is time-dependent and dominated by a westward traveling wave of which the structure closely resembles the first growing mode, i.e., the eigenvector of which the

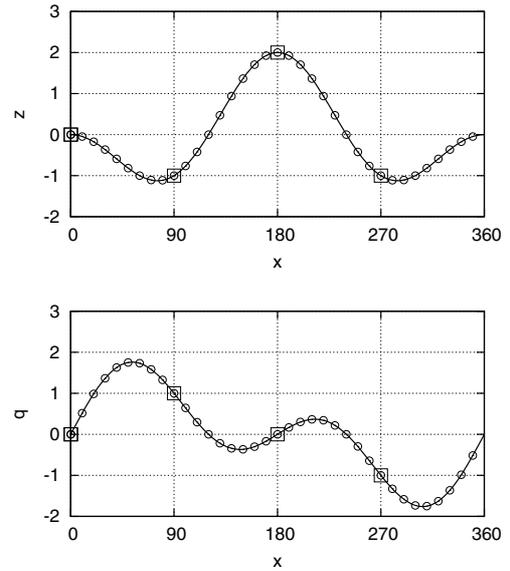


Fig. 2. Graph of $z(x) = \cos(2x) - \cos(x)$ (upper panel) and $q(x) = \sin(2x) + \sin(x)$ (lower panel) with x expressed in degrees. In the upper panel the squares and circles denote z_m for $K = 4$ and $K = 36$, respectively; in the lower panel the squares and circles denote q_m for $K = 4$ and $K = 36$, respectively.

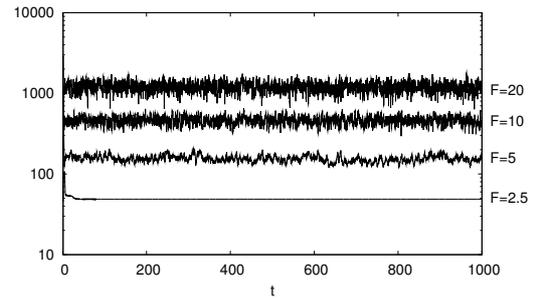


Fig. 3. The time evolution of the energy in the numerical simulation with $K = 36$ nodes and $F = 2.5, 5, 10$ and 20 .

wavenumber is given by $m - 1$ with m the value for which $z_m = z_{min}$. If F increases further the amplitude of the wave becomes larger and the dominant wavenumber decreases. However, with increasing F the dominance by waves gives way to increasingly chaotic behaviour, see also Orrell and Smith [25]. Analogous statements are valid for negative values of F .

In order to explore the nature of the system's dynamics outside the range of forcing values for which the steady state is stable, we integrated the system (5) numerically for the values $F = 2.5, 5, 10$ and 20 for a period of 20 000 time units. We used a fourth order Runge-Kutta time integration scheme with a time step of 0.001 and sampled the model's state every 0.1 time unit. We took $K = 36$, as in Lorenz [4], and started the integration with all variables X_k equal to 0, except for X_1 which was set to 0.1. The evolution of the energy (6) of the system over the first 1000 time units is shown in Figure 3. The graphs show that in all four cases the energy quickly reaches a well-defined equilibrium value over this period of time.

Table 1. Overview of the results of the numerical integrations (first set of four rows and collectively denoted by ‘Num. sim.’) and the results of the maximum entropy analysis (next five sets of four rows and collectively denoted by ‘Maxent’ and the equation numbers referring to the constraints used). The different columns give the values of the forcing F , the average energy $E_0 = \langle E \rangle$, the average $\mu_k = \gamma$ and standard deviation $C_{kk}^{1/2}$ of an individual variable X_k , as well as the parameters σ and φ . Values that are underlined have been used as input of the maximum entropy analysis.

Case	F	$\langle E \rangle$	γ	$C_{kk}^{1/2}$	σ	φ
Num.	2.5	48.78	1.084	1.239	–	–
sim.	5	154.3	1.715	2.374	–	–
	10	465.2	2.586	4.385	–	–
	20	1199.	3.331	7.458	–	–
Maxent	2.5	<u>48.78</u>	0.000	1.646	1.646	–
(16)	5	<u>154.3</u>	0.000	2.928	2.928	–
	10	<u>465.2</u>	0.000	5.084	5.084	–
	20	<u>1199.</u>	0.000	8.162	8.162	–
Maxent	2.5	<u>48.78</u>	1.084	1.239	1.239	–
(16)	5	<u>154.3</u>	1.714	2.373	2.373	–
(17)	10	<u>465.2</u>	2.584	4.378	4.378	–
	20	<u>1199.</u>	3.331	7.451	7.451	–
Maxent	2.5	56.25	1.250	1.250	1.250	–
	5	225.0	2.500	2.500	2.500	–
(17)	10	900.0	5.000	5.000	5.000	–
	20	3600.	10.000	10.000	10.000	–
Maxent	2.5	<u>48.78</u>	1.084	1.239	0.557	0.869
(16)	5	<u>154.3</u>	1.714	2.373	1.843	0.680
(17)	10	<u>465.2</u>	2.584	4.378	3.965	0.465
(18)	20	<u>1199.</u>	3.331	7.451	7.043	0.355
Maxent	2.5	72.32	1.607	1.198	0.893	0.714
	5	244.8	2.720	2.490	2.280	0.441
(17)	10	920.0	5.111	4.999	4.889	0.222
(18)	20	3619.	10.053	10.000	9.947	0.107

We are interested in the statistics of an individual variable X_k as well as in correlations C_{kl} between the variable X_k with any of the other variables X_l . The statistics of an individual variable can be described by giving its full marginal probability density function or, more compactly but less complete, by giving its mean μ_k and variance C_{kk} . We note that, due to the system’s symmetry, the mean μ_k is independent of the index k and the covariance C_{kl} is only dependent on the difference between k and l . All numerical averages are calculated straightforwardly by adding the sampled values of the relevant quantity and dividing by the total number of samples, i.e., by 200 001. Marginal probability density functions are obtained from histograms of X_k . An appropriate interval of X_k is divided into 100 equal-sized bins, the number of samples within each bin is counted and the result is divided by the total number of samples and the bin size. The numerical results corresponding to the four values of F are given in the first set of rows of Table 1.

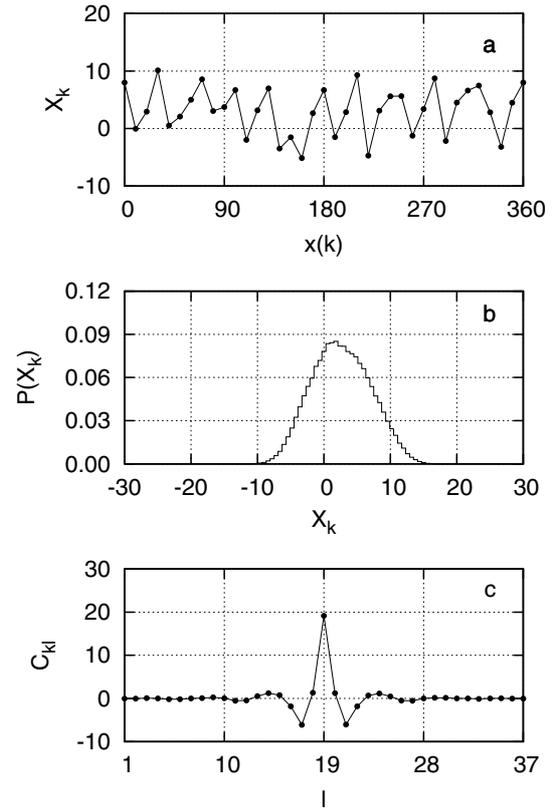


Fig. 4. A few results of a numerical time integration of the Lorenz system (5) for $F = 10$ and $K = 36$. A snapshot at $t = 1000$ is shown in the first graph (panel a), where the variables X_k are plotted as a function of $x(k) = (360/36)(k - 1)$. The second panel (b) shows the marginal probability density function for a single variable X_k , as calculated from the 200 001 sampled values of the numerical integration. The last panel shows the correlation function C_{kl} as a function of l , for $k = 19$, determined numerically from the same sampled values of X_k .

We will discuss in more detail the case $F = 10$ for which the numerical average energy E_0 is 465.2, or 12.92 for E_0/K , i.e., for the energy per degree of freedom. The outcomes of this numerical integration will be used in the next section as illustration of the results to be discussed there. In the first graph (panel a) of Figure 4 we show a snapshot of the system at $t = 1000$, i.e., at the end of the integration window of Figure 3. Due to the fact that the forcing F is positive, the average value of an individual variable X_k is positive as well. We see from this snapshot that a wave pattern is present, with a wave number around 10. This wave behaves very erratically, although a moderate westward motion can be discerned. The second panel of Figure 4 shows the numerical marginal probability density function, based on a histogram over the interval $[-30, 30]$, of the values of X_k , $k = 1$. We see that, for $F = 10$, the form is not unlike a normal distribution and is characterized by an average of 2.586 and a standard deviation of 4.385, see Table 1.

As the presence of a moving wave pattern already indicates, the variables X_k are correlated. This is confirmed

by the last graph (panel c) of Figure 4 in which we show the correlation coefficient C_{kl} for $k = 19$ as a function of l . The choice of $k = 19$ is motivated by the wish to have the maximum (equal to the variance or squared standard deviation) in the middle of the figure but is otherwise arbitrary due to the system's symmetry. The covariance can be described as a damped oscillation away from its reference point $k = 19$. In the next paragraph we will investigate to what extent the statistics, as revealed by the marginal probability density function of X_k and the correlation function C_{kl} , can be reproduced from the maximum entropy principle.

4 Applying the maximum entropy principle

In this paper we are interested in the statistical properties of (5). We will assume that the system has been integrated numerically for a sufficiently long time so that the statistics can be considered stationary. The results discussed in the previous section support this assumption, at least for the cases considered.

The principle of maximum entropy, as outlined in Section 2, will be applied to calculate the system's probability density function $\mathcal{P}(\mathbf{X})$. We will make no a priori assumptions about the probability density function and thus use:

$$\mathcal{M}(\mathbf{X}) = 1.$$

In Section 4.1 we will derive an expression for $\mathcal{P}(\mathbf{X})$, using as constraints the normalization condition, written as:

$$\langle 1 \rangle = 1, \tag{15}$$

and the condition that the average energy has a given value:

$$\langle E \rangle = E_0. \tag{16}$$

This is essentially a recapitulation of the approach of equilibrium statistical mechanics and leads to Gibbs' canonical ensemble. It will be investigated to what extent the statistics of the system is reproduced if the energy E_0 is taken from the corresponding numerical simulation.

In Section 4.2, we will discuss the consequences of adding the constraint

$$\left\langle \frac{dE}{dt} \right\rangle = 0. \tag{17}$$

This is based on the assumption that the statistics of the system is stationary. Requiring the average time rate of change of the energy to be zero is an attempt to bring the probability density function closer to a stationary solution of the Liouville equation. To what extent this is the case is not investigated but we will see that, by adding the latter constraint, a marked improvement is achieved in the description of the system's statistics.

In Section 4.3 we will study the consequences of adding the constraint

$$\left\langle \frac{d^2E}{dt^2} \right\rangle = 0. \tag{18}$$

This may also be justified on the basis of the system's assumed statistical stationarity and can be seen as a step further in the direction of a stationary solution of the Liouville equation. The use of this constraint will lead to additional, quite realistic, structure in the statistics in the form of correlations between the variables. In all cases considered we compare the analytical results with the numerical results for $F = 10$ given in the previous section. The results for the other values of the forcing are summarized in Table 1, discussed in Section 4.4.

4.1 The case $\langle E \rangle = E_0$

Computing the maximum entropy probability density function $\mathcal{P}(\mathbf{X})$ for the system (5) using only the normalization condition (15) and condition (16) as constraints is straightforward. From (2) and (16) it follows immediately that:

$$\mathcal{P}(\mathbf{X}) = \frac{1}{\mathcal{Z}(\lambda_0)} e^{-\lambda_0 E}, \tag{19}$$

where λ_0 is the Lagrange multiplier associated with the constraint $\langle E \rangle = E_0$ and the partition function $\mathcal{Z}(\lambda_0)$ is given by:

$$\mathcal{Z}(\lambda_0) = \int e^{-\lambda_0 E} dX.$$

By writing:

$$\lambda_0 E = \frac{1}{2\sigma^2} \sum_{k=1}^K X_k^2, \tag{20}$$

in which we have defined (see (A.1) and (A.2) of Appendix A)

$$\sigma^2 = \frac{1}{\lambda_0}, \tag{21}$$

the probability density function assumes the form:

$$\mathcal{P}(\mathbf{X}) = \frac{1}{\mathcal{Z}(\lambda_0)} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^K X_k^2}.$$

By introducing the matrix B with matrix elements

$$B_{kl} = \delta_{kl}$$

the probability density function can be written as:

$$\mathcal{P}(\mathbf{X}) = \frac{1}{\mathcal{Z}(\lambda_0)} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^K \sum_{l=1}^K X_k B_{kl} X_l}. \tag{22}$$

Now, a general multivariate normal distribution is defined by (see [27], their Eq. (6.19.1))

$$\mathcal{N}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{X}) = \frac{1}{(2\pi)^{K/2} |\mathbf{C}|^{1/2}} \times e^{-\frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K (X_k - \mu_k) C_{kl}^{-1} (X_l - \mu_l)}, \tag{23}$$

where $\boldsymbol{\mu}$ is a vector of means (μ_1, \dots, μ_K) and \mathbf{C} is the covariance matrix with matrix elements C_{kl} and determinant $|\mathbf{C}|$. By definition, a multivariate normal distribution has the following properties:

$$\int \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{X}) d\mathbf{X} = 1, \quad (24a)$$

$$\int \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{X}) X_k d\mathbf{X} = \mu_k, \quad (24b)$$

$$\int \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{X}) X_k X_l d\mathbf{X} = \mu_k \mu_l + C_{kl}, \quad (24c)$$

and its entropy is given by (see [27], their Eq. (17.2.2))

$$S_{\mathcal{N}} = \frac{1}{2} \log [(2\pi e)^K |\mathbf{C}|].$$

It can thus be concluded that the sought-for probability density function is simply given by:

$$\mathcal{P}(\mathbf{X}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{X}), \quad (25)$$

with

$$\mu_k = 0, \quad (26a)$$

$$C_{kl} = \sigma^2 B_{kl}^{-1} = \sigma^2 \delta_{kl}, \quad (26b)$$

and

$$|\mathbf{C}| = \sigma^{2K}.$$

For the maximum entropy, denoted by S_M , we thus have:

$$S_M = \frac{1}{2} \log [(2\pi e \sigma^2)^K], \quad (27)$$

whereas for the partition function we obtain:

$$\mathcal{Z}(\lambda_0) = (2\pi)^{K/2} |\mathbf{C}|^{1/2} = (2\pi \sigma^2)^{K/2}. \quad (28)$$

The latter result follows by equating expressions (22) and (23) of the probability density function.

In the resulting probability density function (25) the value of σ , related to the Lagrange multiplier λ_0 , is still unknown. A relation for σ^2 can be obtained by applying (4) to (28), but we prefer to follow a different route. Using (24b), (24c) and (26), one obtains:

$$\begin{aligned} \langle X_k \rangle &= \mu_k = 0, \\ \langle X_k^2 \rangle &= \mu_k^2 + C_{kk} = \sigma^2. \end{aligned}$$

Combining the constraint (16) with (6) gives:

$$\langle E \rangle = \sum_{k=1}^K \frac{1}{2} \langle X_k^2 \rangle = \frac{K}{2} \sigma^2 = E_0.$$

It thus follows that we have:

$$\sigma^2 = \frac{2E_0}{K}, \quad (29)$$

so that

$$\lambda_0 = \frac{K}{2E_0}.$$

This completes the determination of the probability density function.

If we substitute (29) into expression (27) for the entropy S_M we may calculate the entropy per degree of freedom (S_M/K) as a function of the energy per degree of freedom (E_0/K). A grey shading plot of S_M/K as a function of F and E_0/K is given in the upper panel of Figure 5. The inclusion of F is in anticipation of future plots of a similar nature, but at the moment the plot's independence of F may remind us that in the analytic treatment the forcing has not entered the stage yet. The numerical integration discussed in the previous section has $F = 10$ and $E_0/K = 12.92$ and is represented by the dot in the graph, the corresponding value of S_M/K being 3.04.

When we use the numerical average value 12.92 of E_0/K , we may calculate σ^2 and thus the probability density function $\mathcal{P}(\mathbf{X})$. This probability density function involves all variables X_k . The marginal probability density function of any subset of variables may be found by retaining only the means and covariances of this subset. If there is only a single variable X_k in this subset, we have:

$$\mathcal{P}(X_k) = \mathcal{N}(\mu_k, C_{kk}, X_k), \quad (30)$$

i.e., a univariate normal distribution with mean $\mu_k = 0$ and variance

$$C_{kk} = \sigma^2 = \frac{2E_0}{K}. \quad (31)$$

A plot of this marginal probability density function, together with the numerically obtained probability density function of the previous section, is shown in the second panel of Figure 5. The analytical marginal probability density function has its maximum at the origin and not at a finite positive value, but the width is only mildly overestimated, see Table 1. From (26b) we conclude that the analytical covariance matrix is proportional to the identity matrix. The correlations between the variables, as exemplified by the last panel of Figure 4, are thus not reproduced.

As remarked above, the probability density function (25) is Gibbs' canonical ensemble. This probability density function is expected to give a proper description of the statistics of the unforced and undamped system. To demonstrate that this is indeed the case, we continued the time integration discussed in the previous section for another 20 000 time units but with the forcing and damping put to zero and sampling the results over the latter period. When comparing the flow patterns of this integration with the flow patterns of the forced and damped integration, it was noticed that waves remain present and continue to behave erratically but without overall wave propagation. This suggests that correlations between the different variables will be small and, indeed, they are: not larger in amplitude than 0.2, in accord with the absence of correlations in (25). The marginal probability density function for a single variable is shown in the third panel of Figure 5, together with the analytical distribution of panel b. The match is nearly perfect. Similar results apply when this is repeated for the other integrations.

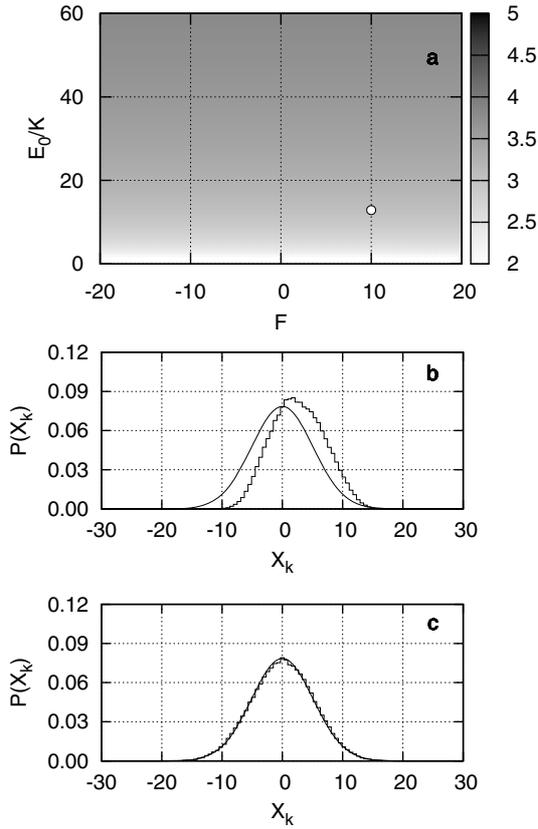


Fig. 5. Upper panel (a): a grey shading plot of the maximum entropy per degree of freedom (S_M/K) as a function of the forcing F and the energy per degree of freedom (E_0/K) in the case that the average energy is used as the single constraint in the maximization of entropy. The dependence on F has been included in anticipation of plots of a similar nature to be discussed in the following subsections. The dot has coordinates $F = 10$ and $E_0/K = 12.92$ and corresponds to the time integration discussed in the previous section. The value of S_M/K at this point is 3.04. Middle panel (b): the corresponding marginal probability density function (30) together with the numerically obtained distribution of Figure 4, panel b. Lower panel (c): the same marginal probability density function (30) but now combined with the numerical distribution, obtained by setting the forcing and damping to zero and continuing the numerical integration for another 20 000 time units with sampling over the latter period.

4.2 The case $\langle \mathbf{E} \rangle = \mathbf{E}_0$, $\langle d\mathbf{E}/dt \rangle = \mathbf{0}$

If we maximize the entropy under the normalization condition (15) and the constraints (16) and (17) we obtain, according to (2),

$$\mathcal{P}(\mathbf{X}) = \frac{1}{\mathcal{Z}(\lambda_0, \lambda_1)} e^{-\lambda_0 E - \lambda_1 \frac{dE}{dt}}, \quad (32)$$

where λ_0 and λ_1 are the Lagrange multipliers associated with the constraints (16) and (17), respectively, and the partition function $\mathcal{Z}(\lambda_0, \lambda_1)$ is given by:

$$\mathcal{Z}(\lambda_0, \lambda_1) = \int e^{-\lambda_0 E - \lambda_1 \frac{dE}{dt}} dX.$$

The argument of the exponential can be written as:

$$\lambda_0 E + \lambda_1 \frac{dE}{dt} = \frac{1}{2\sigma^2} \sum_{k=1}^K (X_k - \gamma)^2 - \gamma^2, \quad (33)$$

where σ^2 and γ are given by (see (A.1) and (A.2) of Appendix A)

$$\sigma^2 = \frac{1}{\lambda_0 - 2\lambda_1}, \quad (34a)$$

$$\gamma = -\frac{\lambda_1 F}{\lambda_0 - 2\lambda_1}. \quad (34b)$$

Using the matrix \mathbf{B} introduced in the previous subsection (with matrix elements $B_{kl} = \delta_{kl}$) the probability density function can be written

$$\mathcal{P}(\mathbf{X}) = \frac{e^{-\frac{K\gamma^2}{2\sigma^2}}}{\mathcal{Z}(\lambda_0, \lambda_1)} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^K \sum_{l=1}^K (X_k - \gamma) B_{kl} (X_l - \gamma)}.$$

If normalized, this distribution is seen to be a multivariate normal distribution of the form (23) with

$$\mu_k = \gamma, \quad (35a)$$

$$C_{kl} = \sigma^2 B_{kl}^{-1} = \sigma^2 \delta_{kl} \quad (35b)$$

and

$$|\mathbf{C}| = \sigma^{2K}.$$

The entropy is given by (27) and for the partition function we have:

$$\mathcal{Z}(\lambda_0, \lambda_1) = e^{\frac{K\gamma^2}{2\sigma^2}} (2\pi)^{K/2} |\mathbf{C}|^{1/2} = \left(e^{\frac{\gamma^2}{\sigma^2}} 2\pi\sigma^2 \right)^{K/2}.$$

The resulting probability density function is thus the multivariate normal distribution (25) with a diagonal covariance matrix given by (35b) and nonzero means that are given by (35a).

In obtaining the values of σ^2 and γ we might use (4) but we prefer to reason as follows. We have:

$$\begin{aligned} \langle X_k \rangle &= \mu_k = \gamma, \\ \langle X_k^2 \rangle &= \mu_k^2 + C_{kk} = \gamma^2 + \sigma^2, \end{aligned}$$

so that the constraints (16) and (17) assume the form

$$\langle E \rangle = \sum_{k=1}^K \frac{1}{2} \langle X_k^2 \rangle = \frac{K}{2} (\gamma^2 + \sigma^2) = E_0,$$

$$\left\langle \frac{dE}{dt} \right\rangle = \sum_{k=1}^K (-\langle X_k^2 \rangle + \langle X_k \rangle F) = K(-\gamma^2 - \sigma^2 + \gamma F) = 0.$$

From the first condition we obtain:

$$\gamma^2 + \sigma^2 = \frac{2E_0}{K}$$

and if we substitute this in this second condition we get an expression of γ in terms of the energy E_0 and the forcing F :

$$\gamma = \frac{2E_0}{FK}. \quad (36)$$

The value of γ now being known, the same condition can be used to obtain σ^2 :

$$\sigma^2 = \gamma(F - \gamma). \quad (37)$$

From σ^2 and γ^2 thus obtained we may deduce the values of the Lagrange multipliers λ_0 and λ_1

$$\lambda_0 = \frac{1}{F\sigma^2} (F - 2\gamma), \quad (38a)$$

$$\lambda_1 = \frac{1}{F\sigma^2} (-\gamma). \quad (38b)$$

These equations follow from (34) by dividing the expression of γ by the expression of σ^2 , giving λ_1 , and then using the expression of σ^2 to obtain λ_0 . This completes the determination of the probability density function.

Substituting expression (36) of γ into (37) we obtain:

$$\sigma^2 = \frac{2E_0}{K} \left(1 - \frac{2}{F^2} \frac{E_0}{K} \right). \quad (39)$$

This result is quite remarkable as it shows that not all values of the energy E_0 are allowed as only values of E_0 for which

$$\frac{E_0}{K} \leq \frac{F^2}{2} \quad (40)$$

lead to an acceptable solution, i.e., a solution for which σ^2 is nonnegative³. If the expression for σ^2 is substituted in equation (27) of the entropy, we see that the entropy depends on the energy as well as on the forcing and that the entropy is well-defined as long as (40) is satisfied.

In panel a of Figure 6 the entropy is plotted as a function of the forcing and the energy. The upper solid parabola, at which the left-hand side and the right-hand side of (40) are equal, bounds from above the region of allowed values of the energy. The lower dot in this panel, at which $S_M/K = 2.90$, denotes the values $F = 10$ and $E_0/K = 12.92$ and refers to the numerical integration discussed in the previous section. The marginal probability density function is (30) with $\mu_k = \gamma$ and C_{kk} given by:

$$C_{kk} = \sigma^2 = \gamma(F - \gamma). \quad (41)$$

This probability density function is plotted as the smooth solid curve in the second panel of Figure 6, together with the numerically obtained probability density function discussed in the previous section. The agreement between the analytical and the numerical curves is now much better, especially because the analytical mean of 2.584 (based on the numerically given value of the energy) is very close to the numerical value of 2.586, see Table 1. Another improvement is that the analytical probability density function has become somewhat slimmer and has acquired a somewhat higher maximum value. The covariance matrix is diagonal, see (35b), so that the covariances as seen in panel c of Figure 4 are not reproduced.

³ We note that the existence of an upper limit to the energy of a forced-dissipative system that has converged to its attractor is also implied by equation (7) of Lorenz [28].

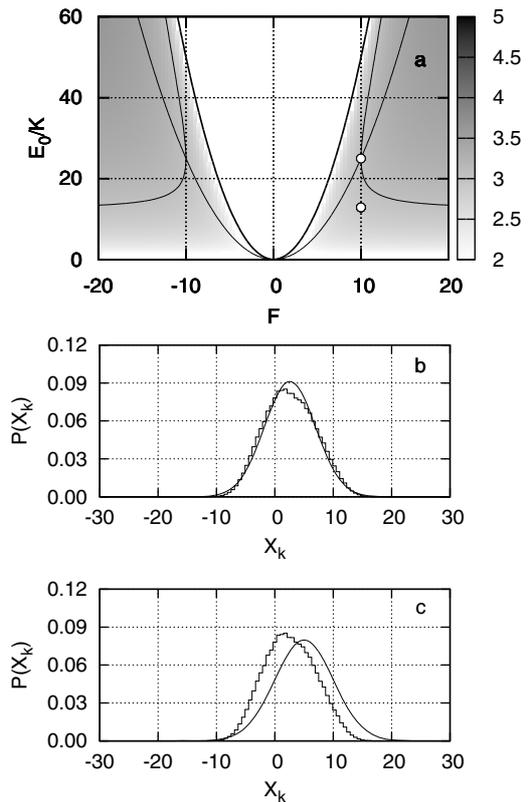


Fig. 6. Upper panel (a): a grey shading plot of the maximum entropy per degree of freedom (S_M/K) as a function of the forcing F and the energy per degree of freedom (E_0/K) in the case that both (16) and (17) are used as constraints in the maximization of entropy. The lower dot has the same significance as in Figure 5 and marks the coordinates $F = 10$ and $E_0/K = 12.92$; the value of S_M/K at this point is 2.90. The upper dot denotes the coordinates $F = 10$ and $E_0/K = 25.0$ and lies on the moderately thin parabola that represents the values of E_0/K that result from a maximum entropy analysis without the energy constraint (16). The two thin solid curves are isolines of the entropy passing through the coordinates of the upper dot; the value of S_M/K at this point is 3.03. The thick solid parabola marks the upper boundary of the allowed values of E_0/K . Middle panel (b): the analytical marginal probability density function corresponding to the lower dot together with the numerically obtained distribution of Figure 4, panel b. Lower panel (c): the analytical marginal probability density function corresponding to the upper dot in combination with the same numerical distribution as in panel b.

For each value of the forcing F there is a value of the energy E_0 for which the entropy is larger than for other values of E_0 . This energy value would be the result of our analysis, had the energy not been used to constrain the maximization of entropy. For any value of the forcing F , we may obtain this maximum by repeating the analysis above but leaving out the constraint (16). The result can be obtained straightforwardly by putting the Lagrange multiplier λ_0 equal to zero in (38a). It follows that:

$$\gamma = \frac{F}{2}, \quad (42)$$

so that, from (36) and (37), we have:

$$\frac{E_0}{K} = \frac{F^2}{4}, \quad (43a)$$

$$\sigma^2 = \frac{F^2}{4}. \quad (43b)$$

The points in panel a of Figure 6 for which the first of these relations is valid are represented by the intermediately thin parabola. It is a curve halfway between zero and the maximum allowed value of E_0/K and connects the entropy maxima that pertain to given values of the forcing F .

For $F = 10$ the value of E_0/K on this curve is 25.00 and the corresponding point is denoted by the upper dot in panel a of Figure 6. The isoline of S_M/K that passes through this point, with a value of 3.03, is displayed in the form of a thin solid curve. The marginal probability density function that corresponds to these values is shown in the third panel of Figure 6, along with the same numerical distribution as in panel b. In accordance with the fact that the energy is higher than the corresponding numerical simulation the mean is higher than the numerical mean. Also the width is somewhat larger. Although the latter result is less accurate than the former is it about as accurate as the result in the previous subsection. To the credit of the result shown in panel c of Figure 6 is that no information from the numerical simulation is used; it is exclusively based on the assumption that the system has reached a statistically stationary state.

4.3 The case $\langle E \rangle = E_0$, $\langle dE/dt \rangle = 0$, $\langle d^2E/dt^2 \rangle = 0$

In this subsection we consider the case in which, apart from the normalization condition, the conditions $\langle E \rangle = E_0$, $\langle dE/dt \rangle = 0$ and $\langle d^2E/dt^2 \rangle = 0$ are used as constraints in the maximization of entropy. According to (2) the probability density function can be written

$$\mathcal{P}(\mathbf{X}) = \frac{1}{\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2)} e^{-\lambda_0 E - \lambda_1 \frac{dE}{dt} - \lambda_2 \frac{d^2E}{dt^2}}, \quad (44)$$

where λ_0 , λ_1 and λ_2 are the Lagrange multipliers associated with the constraints and the partition function $\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2)$ is given by:

$$\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2) = \int e^{-\lambda_0 E - \lambda_1 \frac{dE}{dt} - \lambda_2 \frac{d^2E}{dt^2}} dX.$$

The argument of the exponential can be written as:

$$\lambda_0 E + \lambda_1 \frac{dE}{dt} + \lambda_2 \frac{d^2E}{dt^2} = \frac{1}{2\sigma^2} \sum_{k=1}^K ((X_k - \gamma)^2 + \varphi X_k (X_{k+2} - X_{k+1}) + \varphi F - \gamma^2), \quad (45)$$

with σ^2 , γ and φ given by (see (A.1) and (A.2) of Appendix A)

$$\sigma^2 = \frac{1}{\lambda_0 - 2\lambda_1 + 4\lambda_2}, \quad (46a)$$

$$\gamma = -\frac{(\lambda_1 - 3\lambda_2)F}{\lambda_0 - 2\lambda_1 + 4\lambda_2}, \quad (46b)$$

$$\varphi = \frac{2\lambda_2 F}{\lambda_0 - 2\lambda_1 + 4\lambda_2}. \quad (46c)$$

By introducing the matrix D with the matrix elements

$$D_{kl} = \delta_{kl} + \frac{\varphi}{2} (\delta_{k-2,l} - \delta_{k-1,l} - \delta_{k+1,l} + \delta_{k+2,l}),$$

we obtain for the probability density function (44)

$$\mathcal{P}(\mathbf{X}) = \frac{e^{-\frac{K(\gamma^2 - \varphi F)}{2\sigma^2}}}{\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2)} e^{-\frac{1}{2\sigma^2} \sum_{k=1}^K \sum_{l=1}^K (X_k - \gamma) D_{kl} (X_l - \gamma)}.$$

After normalization, this distribution becomes a multivariate normal distribution (23) with

$$\mu_k = \gamma, \quad (47a)$$

$$C_{kl} = \sigma^2 D_{kl}^{-1}, \quad (47b)$$

and

$$|C| = \sigma^{2K} |D|^{-1}.$$

The entropy is given by:

$$S_M = \frac{1}{2} \log [(2\pi e \sigma^2)^K |D|^{-1}], \quad (48)$$

and for the partition function we have:

$$\begin{aligned} \mathcal{Z}(\lambda_0, \lambda_1, \lambda_2) &= e^{-\frac{K(\gamma^2 - \varphi F)}{2\sigma^2}} (2\pi)^{K/2} |C|^{1/2} \\ &= (e^{-\frac{(\gamma^2 - \varphi F)}{\sigma^2}} 2\pi \sigma^2)^{K/2} |D|^{-1/2}. \end{aligned}$$

This is the same general form as obtained in the cases with one and two constraints, but with a covariance matrix C that has more structure as it is proportional to the inverse of a matrix D that is not diagonal.

The matrix D is circulant like the matrix A of the linear stability problem of the steady state. Therefore, the matrix of eigenvectors – of which the elements are denoted by W_{lm} – is given by (10) and (11). For the eigenvalues, denoted here by ψ_m , the expression is (12) with matrix elements D_{1n} instead of A_{1n} . This yields

$$\psi_m = 1 + \varphi z_m,$$

where z_m is given by (13a). As the determinant of a matrix is the product of its eigenvalues, we have:

$$|D| = \prod_{m=1}^K \psi_m = \prod_{m=1}^K (1 + \varphi z_m), \quad (49)$$

which implies that

$$|D|^{-1} = \prod_{m=1}^K \frac{1}{1 + \varphi z_m}.$$

The entropy can thus be written

$$S_M = \frac{1}{2} \log \left[\prod_{m=1}^K \frac{2\pi e \sigma^2}{1 + \varphi z_m} \right]. \quad (50)$$

It also follows from (49) that

$$|D|^{-1/2} = \prod_{m=1}^K \left[\frac{1}{1 + \varphi z_m} \right]^{1/2},$$

which gives

$$\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2) = \prod_{m=1}^K \left[\frac{e^{\frac{(\gamma^2 - \varphi F)}{\sigma^2} 2\pi \sigma^2}}{1 + \varphi z_m} \right]^{1/2}$$

for the partition function.

The partition function contains square roots of $\psi_m = 1 + \varphi z_m$. To ensure that the partition function is real and positive (as it should be as it is the normalization factor of the probability density function) we require that φ be such that $1 + \varphi z_m$ is positive for all values of m . We have seen in Section 3 that the minimum z_{min} of z_m is negative (close to $-9/8$) and that the maximum z_{max} of z_m is positive (close to 2) so that all values of $1 + \varphi z_m$ are positive as long as φ satisfies

$$\frac{1}{-z_{max}} < \varphi < \frac{1}{-z_{min}}. \quad (51)$$

This limits the range of φ . The values of σ^2 , φ and γ – and therewith the Lagrange multipliers λ_0 , λ_1 and λ_2 – are to be determined from the constraints (16), (17) and (18). Again, we might make use of (4) to obtain the necessary relations, but we will follow a route which, we believe, is more clarifying.

An important ingredient in the evaluation of the constraints is the inverse of the matrix D . In Appendix B it is shown that the matrix elements of D^{-1} , denoted by D_{kl}^{-1} , are given by:

$$D_{kl}^{-1} = \frac{1}{K} \sum_{m=1}^K \frac{\cos(\omega(m-1)(l-k))}{1 + \varphi z_m}. \quad (52)$$

The elements C_{kl} of the covariance matrix C thus become

$$C_{kl} = \frac{\sigma^2}{K} \sum_{m=1}^K \frac{\cos(\omega(m-1)(l-k))}{1 + \varphi z_m}. \quad (53)$$

We recall that it follows from the expressions (6) and (7) that:

$$\begin{aligned} \langle E \rangle &= \sum_{k=1}^K \frac{1}{2} \langle X_k^2 \rangle \\ \left\langle \frac{dE}{dt} \right\rangle &= \sum_{k=1}^K \left(-\langle X_k^2 \rangle + \langle X_k \rangle F \right). \end{aligned}$$

Furthermore, it follows from the properties (24) of a multivariate normal distribution that:

$$\begin{aligned} \langle X_k \rangle &= \gamma, \\ \langle X_k^2 \rangle &= \gamma^2 + \sigma^2 D_{kk}^{-1} = \gamma^2 + \frac{\sigma^2}{K} \sum_{m=1}^K \frac{1}{1 + \varphi z_m}, \end{aligned}$$

where we also used (47) and (52). We thus have:

$$\begin{aligned} \langle E \rangle &= \frac{1}{2} \left(K \gamma^2 + \sigma^2 \sum_{m=1}^K \frac{1}{1 + \varphi z_m} \right) \\ \left\langle \frac{dE}{dt} \right\rangle &= - \left(K \gamma (\gamma - F) + \sigma^2 \sum_{m=1}^K \frac{1}{1 + \varphi z_m} \right). \end{aligned}$$

From (8) we deduce that:

$$\begin{aligned} \left\langle \frac{d^2 E}{dt^2} \right\rangle + 2 \left\langle \frac{dE}{dt} \right\rangle &= \\ F \sum_{k=1}^K \left(F + \langle X_k X_{k+2} \rangle - \langle X_k X_{k+1} \rangle - \langle X_k \rangle \right). \end{aligned}$$

Equations (24), (47) and (52) imply that

$$\begin{aligned} \langle X_k X_{k+1} \rangle &= \gamma^2 + \sigma^2 D_{k,k+1}^{-1} \\ &= \gamma^2 + \frac{\sigma^2}{K} \sum_{m=1}^K \frac{\cos(\omega(m-1))}{1 + \varphi z_m}, \\ \langle X_k X_{k+2} \rangle &= \gamma^2 + \sigma^2 D_{k,k+2}^{-1} \\ &= \gamma^2 + \frac{\sigma^2}{K} \sum_{m=1}^K \frac{\cos(2\omega(m-1))}{1 + \varphi z_m}. \end{aligned}$$

This gives

$$\left\langle \frac{d^2 E}{dt^2} \right\rangle + 2 \left\langle \frac{dE}{dt} \right\rangle = F \left(K(F - \gamma) + \sigma^2 \sum_{m=1}^K \frac{z_m}{1 + \varphi z_m} \right).$$

To advance from the previous two expressions to the last we used the definition (13a) of z_m . The constraints on the energy and on the first- and second-order time derivatives of the energy are thus seen to lead to the following three conditions on σ^2 , γ and φ

$$\gamma^2 + \frac{\sigma^2}{K} \sum_{m=1}^K \frac{1}{1 + \varphi z_m} = \frac{2E_0}{K}, \quad (54a)$$

$$\gamma(\gamma - F) + \frac{\sigma^2}{K} \sum_{m=1}^K \frac{1}{1 + \varphi z_m} = 0, \quad (54b)$$

$$(F - \gamma) + \frac{\sigma^2}{K} \sum_{m=1}^K \frac{z_m}{1 + \varphi z_m} = 0. \quad (54c)$$

Once we have obtained these parameters, the Lagrange multipliers λ_0 , λ_1 and λ_2 can be derived from (46). Indeed,

we see that:

$$\lambda_0 = \frac{1}{2F\sigma^2}(2F - 4\gamma + 2\varphi), \quad (55a)$$

$$\lambda_1 = \frac{1}{2F\sigma^2}(3\varphi - 2\gamma), \quad (55b)$$

$$\lambda_2 = \frac{1}{2F\sigma^2} \varphi. \quad (55c)$$

The expression for λ_2 follows by dividing (46c) by (46a). By dividing (46b) by (46a) and using the result for λ_2 , one finds λ_1 . Expression (46a) can be used to obtain λ_0 .

By subtracting (54b) from (54a) we find:

$$\gamma = \frac{2E_0}{FK}, \quad (56)$$

which is the same expression for γ as we obtained in the previous subsection. Substituting this value in (54a) we see that

$$\frac{\sigma^2}{K} \sum_{m=1}^K \frac{1}{1 + \varphi z_m} = \frac{2E_0}{K} \left(1 - \frac{2}{F^2} \frac{E_0}{K} \right).$$

This expression resembles (39) and, as $1 + \varphi z_m$ should be positive for all m , implies again that

$$\frac{E_0}{K} \leq \frac{F^2}{2} \quad (57)$$

in order for σ^2 to be nonnegative. By multiplying (54c) by γ and adding the result to (54b) we obtain, excluding $\sigma^2 = 0$,

$$\sum_{m=1}^K \frac{1 + \gamma z_m}{1 + \varphi z_m} = 0. \quad (58)$$

This equation is generally nonlinear in φ and has to be solved numerically. (For $K = 4$, however, the equation is quadratic and can be solved analytically.) After solving (58), any of equations (54) can be used to obtain σ^2 from φ , by which the determination of the probability density function is completed.

We noted above that, in order for $1 + \varphi z_m$ to be positive for all m , the parameter φ should lie between $-1/z_{max}$ and $-1/z_{min}$, see (51). Now, for a sum of fractions (58) with positive denominators to add up to zero, at least one of the nominators has to be negative. This implies that, if φ has to have a value within the limits (51), then γ needs to have a value outside these limits. The immediate consequence is that

$$\gamma \geq \frac{1}{-z_{min}} \quad \text{or} \quad \gamma \leq \frac{1}{-z_{max}}.$$

The first case is relevant for positive F for which it implies that

$$\frac{E_0}{K} \geq \frac{F}{2} \frac{1}{-z_{min}}. \quad (59)$$

The second case is relevant for negative F in which case it implies that

$$\frac{E_0}{K} \geq \frac{F}{2} \frac{1}{-z_{max}}. \quad (60)$$

Given a value of F and a value of E_0/K such that (57) and either (59) or (60) are satisfied we may solve (58) for φ and thus acquire all necessary parameters to determine the probability density function.

However, when the forcing F lies between the limits $-1/z_{max}$ and $-1/z_{min}$ values of E_0/K such that (57) and either (59) or (60) are satisfied do not exist because then the conditions (59) or (60) are incompatible with (57). We have seen in the previous section that for these values of the forcing the steady state solution of the Lorenz system is stable. Equation (58) cannot be solved in this case although the system (54) admits the special solution (for which the value of φ is not relevant)

$$\sigma^2 = 0, \quad \gamma = F,$$

implying that there is only one acceptable value for E_0/K , namely

$$\frac{E_0}{K} = \frac{F^2}{2}.$$

This is the situation in which the probability density function is infinitely sharply peaked at the stable stationary state $X_k = X_{sk} = F$. It is quite remarkable that the added constraint $(d^2E/dt^2) = 0$ identifies this special case.

As in the previous subsections, the marginal probability density function may be found by retaining only the means and covariances of the subset of variables in which we are interested. For a single variable X_k , this gives (30) in which $\mu_k = \gamma$ and in which C_{kk} follows from (53):

$$C_{kk} = \frac{\sigma^2}{K} \sum_{m=1}^K \frac{1}{1 + \varphi z_m} = \gamma(F - \gamma). \quad (61)$$

In the second equality we used (54b). The covariances between the variables X_k are given by (53). Note that the resulting marginal probability density function is identical to the one we obtained in the previous subsection. An important difference, however, is that the covariance matrix (53) is no longer diagonal.

In the upper panel of Figure 7, in the form of a grey shading plot, we show the entropy per degree of freedom (S_M/K) as a function of the forcing (F) and the energy per degree of freedom (E_0/K). The plot is similar to the plot shown in Figure 6, except for the lower limits of the allowed energy. These limits, in which the left-hand sides and the right-hand sides of (59) and (60) are equal, are displayed by the thick solid lines. The parameters of the numerical simulation, $F = 10$ and $E_0/K = 12.92$, are represented by the lower dot; here the value of S_M/K is 2.85. The corresponding marginal probability density function is identical to the one shown in panel b of Figure 6 and is not shown again. Instead, we give in panel b the correlation matrix C_{kl} as a function of l , fixing k at 19, in the form of open dots connected by thin straight lines. The corresponding numerical result, shown earlier in panel c of Figure 4, is given by the filled dots. The resemblance between the analytical and numerical correlation matrix is quite good. With increasing $|l - k|$, the elements of the analytical correlation matrix decrease in amplitude somewhat faster than their numerical counterparts.

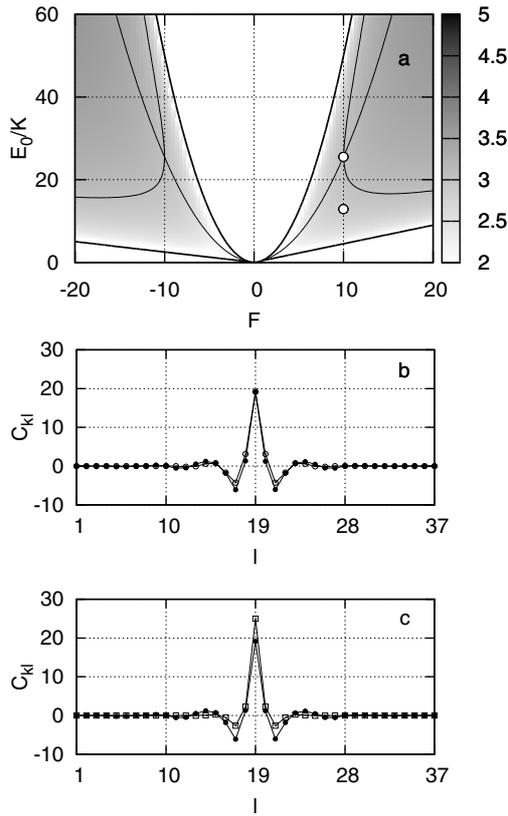


Fig. 7. Upper panel (a): a grey shading plot of the maximum entropy per degree of freedom (S_M/K) as a function of the forcing F and the energy per degree of freedom (E_0/K). All three constraints (16), (17) and (18) are used to constrain the maximization of entropy. The lower dot has the same significance as in Figures 5 and 6 and marks the coordinates $F = 10$ and $E_0/K = 12.92$, corresponding to the numerical simulation discussed in Section 3; the value of S_M/K at this point is 2.85. The upper dot denotes the coordinates $F = 10$ and $E_0/K = 25.55$ and lies on the curve of intermediate thickness that represents the values of E_0/K that result from a maximum entropy analysis without the energy constraint (16). The two thin solid curves are isolines of the entropy passing through the coordinates of the upper dot; the value of S_M/K is 3.02 at this point. The thick solid parabola and the two solid lines mark the upper and lower boundaries of the allowed values of E_0/K . Middle panel (b): the analytical correlation function (open dots) corresponding to the lower dot in panel a, together with the numerically obtained correlation (filled dots) of Figure 4, panel c. Lower panel (c): the analytical correlation function (open squares) corresponding to the upper dot in combination with the same numerical correlation as in panel b.

Also here we note that for each value of the forcing F , there is a value of E_0 at which the entropy S_M is largest. Plotted as a function of F and represented by a curve of intermediate thickness, these values are included in panel a of Figure 7. The value of E_0/K at $F = 10$ for which the entropy is largest is 25.55 and is marked by the upper dot. The entropy S_M/K at this point is 3.02; the corresponding isoline is shown in the form of a thin curve

passing through this point. The value of E_0/K at this point is somewhat higher than the energy of the corresponding point in Figure 6 – there it is 25.00 – so that its associated marginal probability density function is slightly shifted towards higher values. The difference between this graph and the graph shown in panel c of Figure 6 is too small to warrant a separate plot. Instead, we show the analytical correlation function in panel c. In accord with the higher value of the energy, its peak is higher than the corresponding plot in the second panel of Figure 7.

The curve which marks the value of the energy at which the entropy is maximal is obtained by repeating the maximum entropy calculation without the constraint on the energy, i.e., without (16). The results can be obtained rather simply by setting the Lagrange multiplier λ_0 equal to 0. If we do this in the first equation of (55), we see that the average γ is given directly in terms of the forcing F and the parameter γ by:

$$\gamma = \frac{F + \varphi}{2}. \quad (62)$$

The calculations proceed as before, except that (54a) is not to be considered as a constraint but as one of the expressions from which E_0 might be calculated once γ , σ^2 and φ are known – an alternative is (56).

The value of φ is obtained from (58) by substituting (62) for γ and solving directly for φ in terms of F . After substitution of (62) into (58), the equation becomes:

$$\sum_{m=1}^K \frac{1 + Fz_m}{1 + \varphi z_m} + K = 0. \quad (63)$$

This equation will not yield acceptable values of φ for all values of the forcing F ; the forcing has to lie outside the range from $-1/z_{max}$ to $-1/z_{min}$, i.e., outside the range for which the steady state $X_k = X_{sk} = F$ is stable. This can be seen as follows. In order to have a real and positive partition function, all denominators in (63) should be positive. But if F would lie between $-1/z_{max}$ and $-1/z_{min}$, then also all nominators are positive so that the first term on the left-hand side of (63) cannot cancel the second term on the left-hand side. The only option to satisfy (54) is to have $\sigma^2 = 0$ and $\gamma = F$ so that $E_0/K = F^2/2$, in accord with what we have seen above. If F lies outside the range from $-1/z_{max}$ to $-1/z_{min}$, then an acceptable value of φ can be found from (63), either analytically (if $K = 4$) or numerically by a root finding procedure. From the resulting value one may calculate γ using (62) and σ^2 from either the second or the third equation in (54). The energy per degree of freedom E_0/K can then be calculated by using e.g. (56). To give some insight into the corresponding graphs for small values of F , we reproduce an enlarged portion of panel a of Figure 7 in Figure 8.

Of central importance in the use of $\langle d^2 E/dt^2 \rangle = 0$ as a constraint in the maximization of entropy is (58) or, if we leave out $\langle E \rangle = E_0$ as a constraint, its counterpart (63). If γ or F lies outside the interval in which the steady state is stable, then both (58) and (63) can be solved to give acceptable values of φ . As remarked above, these relations

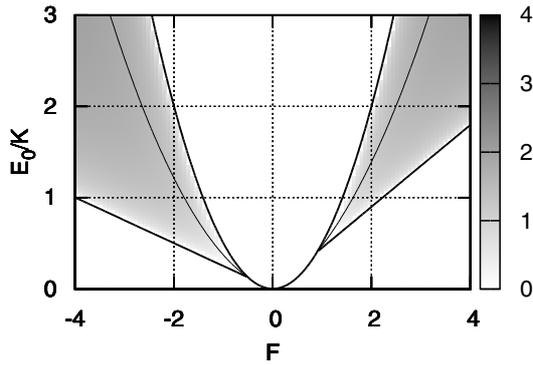


Fig. 8. Enlarged part of Figure 7 showing the entropy in the region of allowed values of the energy E_0 for small values of the forcing F . If the forcing is between $-1/z_{max}$ and $-1/z_{min}$ the only allowed values of the energy are $E_0/K = F^2/2$, the energy of the stable steady state.

are generally nonlinear in φ and need to be solved numerically. However, experience has learned us that the equations are very well behaved, the functions involved being monotonic with a single zero if either γ or F lies in the region of allowed values (if this is not so, the function has singularities and multiple roots). If $K = 4$, then $z_1 = 0$, $z_2 = -1$, $z_2 = 2$ and $z_3 = -1$, see Figure 2, and both functions are quadratic. Relation (58) is then equivalent with:

$$\varphi^2 + (3\gamma - 2)\varphi - 2 = 0,$$

and its solutions are:

$$\begin{aligned} \gamma \geq 1: \quad \varphi &= 1 - \frac{3}{2}\gamma + \left(2 + \left(1 - \frac{3}{2}\gamma\right)^2\right)^{1/2}, \\ \gamma \leq -\frac{1}{2}: \quad \varphi &= 1 - \frac{3}{2}\gamma - \left(2 + \left(1 - \frac{3}{2}\gamma\right)^2\right)^{1/2}. \end{aligned}$$

Relation (63) becomes in the case $K = 4$

$$\varphi^2 + \left(\frac{3}{5}F - \frac{4}{5}\right)\varphi - \frac{4}{5} = 0,$$

and has solutions

$$\begin{aligned} F \geq 1: \quad \varphi &= \frac{2}{5} - \frac{3}{10}F + \left(\frac{4}{5} + \left(\frac{2}{5} - \frac{3}{10}F\right)^2\right)^{1/2}, \\ F \leq -\frac{1}{2}: \quad \varphi &= \frac{2}{5} - \frac{3}{10}F - \left(\frac{4}{5} + \left(\frac{2}{5} - \frac{3}{10}F\right)^2\right)^{1/2}. \end{aligned}$$

With these exact results, the maximum entropy analysis for the Lorenz system with $K = 4$ is fully analytical; this may be useful in further studies. In panel a of Figure 9 we show the solutions of (58), i.e., φ as a function of γ , for $K = 4$ in the form of dashed curves and for $K = 36$ in the form of solid curves. In panel b of Figure 9 the solutions of (63) are shown, i.e., φ as a function of F , in the form of dashed ($K = 4$) and solid ($K = 36$) curves. The graphs for $K = 4$ are based on the equations above; the graphs for $K = 36$ are obtained numerically.

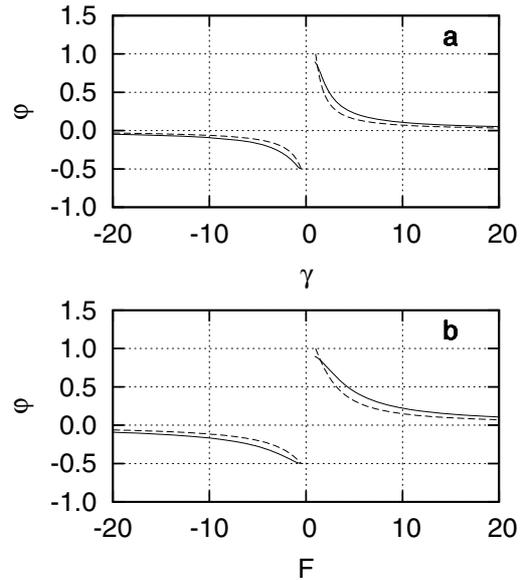


Fig. 9. Graphs of the solutions of (58), i.e., φ as a function of γ (upper panel) and the solutions of (63), i.e., φ as a function of F (lower panel). The dashed graphs refer to $K = 4$ and are based on the analytical expressions given in the text, the solid graphs refer to $K = 36$ and are obtained numerically.

4.4 Summary

For both $F = 10$, the case which was used as a reference in the previous subsections, as well as for $F = 2.5$, $F = 5$ and $F = 20$ the results of the numerical simulations and the maximum entropy analysis are given in Table 1. We recall that the time evolution of the energy in all four cases is shown in Figure 3. The first four rows of the table (Num. sim.) contain the results of the numerical integration for the different values of the forcing F ; the values of F are given in the first column of the table. The average energy $E_0 = \langle E \rangle$ is given in the second column, the mean value $\mu_k = \gamma$ of an individual variable X_k is given in the third column and the standard deviation $C_{kk}^{1/2}$ is given in the fourth column. The fifth and sixth column are reserved for the analytical variables σ and φ .

The second set of rows shows the analytical results obtained in Section 4.1 in which the entropy is maximized with (16) as a constraint, i.e., under the condition that $\langle E \rangle = E_0$. The values of E_0 are taken from the numerical simulation and are repeated in the second column. The values are underlined to emphasize that they are input of the maximum entropy (maxent) analysis. The third set of rows refers to Section 4.2 in which the entropy is maximized with (16) and (17) as constraints, i.e., with $\langle E \rangle = E_0$ and $\langle dE/dt \rangle = 0$. Also here the average energy is input of the analysis. This is different in the fourth set of rows in which we give the analytical results in case only (17), or $\langle dE/dt \rangle = 0$, is used as a constraint. Here the average energy is output of the analysis. The last two sets of rows contain the results of Section 4.3 in which (18), i.e. $\langle d^2E/dt^2 \rangle = 0$ is added as a constraint.

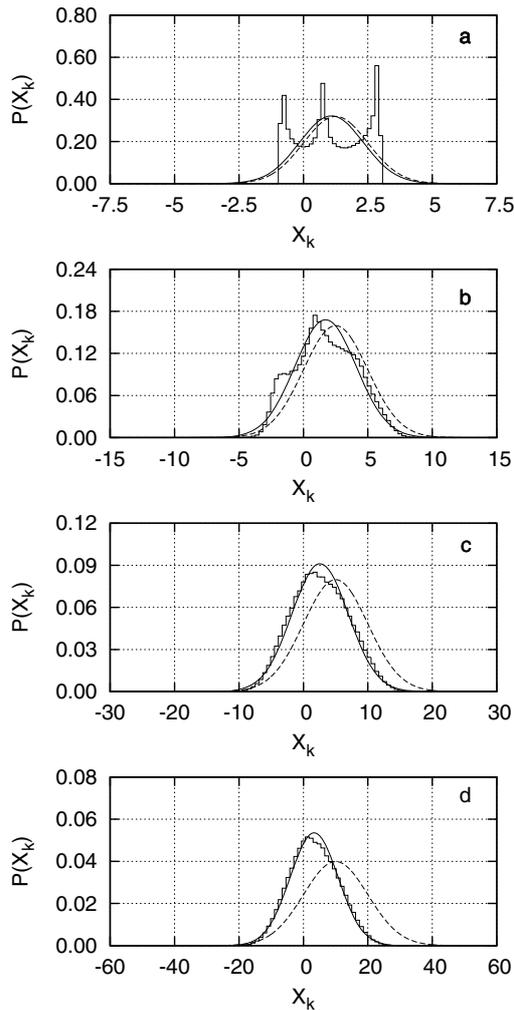


Fig. 10. The marginal probability density function $\mathcal{P}(X_k)$ for Lorenz' system with $K = 36$ nodes and forcing values $F = 2.5$ (panel a), 5 (panel b), 10 (panel c) and 20 (panel d). As in earlier plots of a similar nature, the numerical results are displayed by means of stepped curves. When the maximum entropy principle is applied with (16) and (17) as constraints, i.e., under the conditions $\langle E \rangle = E_0$ and $\langle dE/dt \rangle = 0$, one obtains the smooth solid curves. If only (17) is used as a constraint, i.e., if only $\langle dE/dt \rangle = 0$ is used to constrain the maximization of entropy, one obtains the smooth dashed curves. The analytical graphs are normal distributions with mean γ and standard deviation $C_{kk}^{1/2}$, of which the values are given in the third and fourth column of Table 1. The results displayed in panel c are also shown in the panels b and c of Figure 6.

5 Discussion

In this section we place the results for $F = 10$, on which we focused in the previous section, in the perspective of higher and lower values of F . In Figure 10 we show the numerically obtained marginal probability density functions for $F = 2.5$ (panel a), $F = 5$ (panel b), $F = 10$ (panel c) and $F = 20$ (panel d). The probability density functions were calculated numerically by dividing the ranges $[-7.5, 7.5]$, $[-15, 15]$, $[-30, 30]$ and $[-60, 60]$ of X_k

in 100 equal-sized intervals and counting the number of samples of the variable X_k ($k = 1$) within each interval. We also show, using solid lines, the analytical marginal probability density functions that result from the maximum entropy analysis with (16) and (17) as constraints, i.e., using the conditions $\langle E \rangle = E_0$ and $\langle dE/dt \rangle = 0$. The parameters γ and $C_{kk}^{1/2}$ that characterize these distributions are given in the third set of rows of Table 1. By means of the dashed curves we show the marginal probability density functions in the case that only (17), i.e., $\langle dE/dt \rangle = 0$, is used as a constraint. The corresponding parameters are given in the fourth set of rows in Table 1. The two different set of theoretical curves illustrate the differences between the cases with and without the constraint $\langle E \rangle = E_0$. Note that the results shown in panel c have been shown before in panels b and c of Figure 6. They are included here to place them in the context of similar graphs for lower and higher values of the forcing F . We recall that all plots refer to $K = 36$.

Looking at the numerical probability density functions, it becomes evident that their resemblance with a normal distribution becomes less if the forcing decreases. At $F = 2.5$ this is quite outspoken. Of the three distinct maxima that can be discerned in the marginal probability density function, the outer two are related to the fact, mentioned in Section 3, that the behaviour of the system at small values of the forcing is dominated by a (westward) propagating wave. This implies that the behaviour of an individual variable X_k is dominated by an oscillation, meaning that a relatively large amount of time is spent at the upper and lower extremes of X_k . Generally speaking, the solid curves demonstrate that the accurate description of the marginal probability density functions in the case that both $\langle E \rangle = E_0$ and $\langle dE/dt \rangle = 0$ are taken into account in the maximization of entropy is valid in the whole range of forcing values shown, although, as we have just seen, the form of the numerical marginal probability density function can be quite unlike a normal distribution for smaller values of F .

If the constraint $\langle E \rangle = E_0$ is left out, the average energy $\langle E \rangle$ is overestimated. This was already noted in the previous section for $F = 10$, but Table 1 (fourth and sixth set of rows) shows this to be also true for the other values of F . It reflects itself in analytical marginal probability density functions (dashed curves) that have mean values that are higher than the numerically obtained ones. We also see, and this is confirmed by a closer look at Table 1, that the overestimation, if expressed in relative terms, is larger for larger values of the forcing F . For moderate values of F , say for values not larger than 10, the difference between the results with and without the constraint $\langle E \rangle = E_0$ is small enough for the results to be useful as an estimate of the marginal probability density function of a variable X_k . And this is quite remarkable as no information from a numerical integration is needed to produce this estimate. This is different from the case in which $\langle E \rangle = E_0$ is used as the single constraint in the maximization of entropy. For values of F smaller than about 10 this case yields marginal probability density functions that are

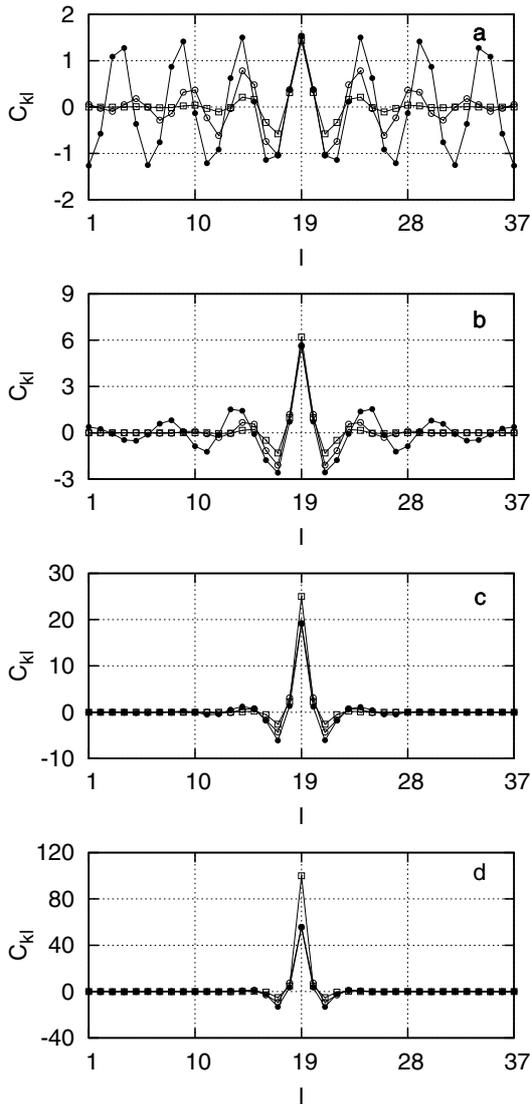


Fig. 11. The covariance C_{kl} with $k = 19$ as a function of l . As in similar plots shown in previous sections, the numerical covariances are represented by filled dots connected by straight lines. The analytical results from the maximum entropy procedure, using (16), (17) and (18) as constraints ($\langle E \rangle = E_0$, $\langle dE/dt \rangle = 0$ and $\langle d^2E/dt^2 \rangle = 0$) are represented by the open dots. The results from the entropy maximization procedure with (17) and (18), i.e. with only $\langle dE/dt \rangle = 0$ and $\langle d^2E/dt^2 \rangle = 0$ as constraints, are given by the open squares. The results displayed in panel c are also shown in the panels b and c of Figure 7.

about as accurate as those obtained by using $\langle dE/dt \rangle = 0$, but rely on information from a numerical integration in the form of a value of E_0 .

For the same series of forcing values, i.e., for $F = 2.5, 5, 10$ and 20 , the covariance matrix C_{kl} as a function of l with $k = 19$ is shown in Figure 11. The numerically obtained covariance matrix is displayed, as in the panels b and c of Figure 7, by means of solid dots connected by thin solid lines. The analytical results, from the analysis

in which both $\langle E \rangle = E_0$, $\langle dE/dt \rangle = 0$ and $\langle d^2E/dt^2 \rangle = 0$ are used as constraints in the maximization of entropy, are displayed by means of open dots connected by thin solid lines. The analytical results from maximizing the entropy with $\langle dE/dt \rangle = 0$ and $\langle d^2E/dt^2 \rangle = 0$ are displayed by means of open squares. The graphs in panel c of this figure were shown earlier in panels b and c of Figure 7. The figure illustrates that the resemblance between numerical and analytical results is generally quite good but decreases with decreasing values of F . Indeed, the numerical results display correlations that extend over larger differences between l and k than the analytical results, the more so for smaller values of F . Note that the discrepancy is least serious in case the constraint $\langle E \rangle = E_0$ is taken into account.

6 Conclusions

We have used the principle of maximum entropy to obtain the probability density function of a dynamical system proposed by Lorenz [4]. The state vector \mathbf{X} of this system has K components X_k that represent an atmospheric field on a latitude circle and satisfy an equation that incorporates forcing, damping and advection. In the four numerical integrations that we have studied, in which the forcing F was varied, the system quickly reaches a statistically stationary state in which the energy E fluctuates around a well-defined and constant average E_0 , see Figure 3. An important constraint in the maximization of the entropy is thus $\langle E \rangle = E_0$, where the brackets denote the average defined in terms of the probability density function of the system. The fact that statistical stationarity quickly sets in also justifies the constraint $\langle dE/dt \rangle = 0$, a condition used previously by Burgers [6] in the context of turbulence. Statistical stationarity furthermore justifies constraints of the type $\langle d^n E/dt^n \rangle = 0$, with $n \geq 2$. The constraints on the average time-derivatives of the energy enforce approximate time-invariance of the probability density function and generate additional structure in the statistics. In the present paper we concentrate on $\langle E \rangle = E_0$, $\langle dE/dt \rangle = 0$ and $\langle d^2E/dt^2 \rangle = 0$.

Using the single constraint $\langle E \rangle = E_0$, the maximum entropy principle yields a probability density function in the form of a multivariate normal distribution $\mathcal{P}(\mathbf{X}) = \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}, \mathbf{X})$, in which the means μ_k are zero and the covariance matrix C_{kl} is proportional to the identity matrix, see (25) and (26). The proportionality factor σ^2 is given in terms of the energy E_0 and the number of variables K by (29). The variables are therefore uncorrelated and each variable satisfies a marginal probability density function in the form of a univariate normal distribution with mean $\mu_k = 0$ and variance $C_{kk} = \sigma^2$. The forcing has not been taken into account yet so that the entropy diagram of Figure 5a, where the entropy per degree of freedom (S_M/K) is plotted as a function of the forcing F and the energy per degree of freedom (E_0/K), is independent of F . For the case $F = 10$, where E_0/K is 12.92, the analytically obtained marginal probability density function is somewhat too broad as compared with the numerically

obtained distribution and has a zero instead of a nonzero mean, see Figure 5b and Table 1.

Augmenting the constraint $\langle E \rangle = E_0$ with $\langle dE/dt \rangle = 0$ leads to the same multivariate normal distribution (25), but with nonzero means $\mu_k = \gamma$ that are given by (36) and a diagonal covariance matrix C_{kl} of which the factor σ^2 is given in terms of F and γ by (37). From (36) and (37) it follows that the energy E_0 cannot assume all values but should satisfy $E_0/K \leq F^2/2$. This has rather important consequences for the entropy diagram in Figure 6a, which is now bounded from above by the parabola $E_0/K = F^2/2$. The analytical marginal probability density function of an individual variable X_k , for $F = 10$ and its corresponding numerical value of E_0/K , is shown in Figure 6b, together with the numerical distribution. The match between the two distributions is now much better than in the previous case. In Figure 6a the curve of intermediate thickness is the parabola $E_0/K = F^2/4$, which gives the energy that would result if $\langle E \rangle = E_0$ is skipped as a constraint. For $F = 10$ and the corresponding value of the energy the analytical marginal probability density function is shown in Figure 6c. In accordance with an energy that is overestimated, the probability distribution has a mean that is too large compared with the numerical distribution.

The use of $\langle d^2E/dt^2 \rangle = 0$ as a constraint in the maximization of entropy leads to a further limitation of the allowed values of E_0/K . For positive F , the value of E_0/K should be larger than or equal to $(F/2)(-1/z_{min})$ and for negative F , the value of E_0/K should be larger than or equal to $(F/2)(-1/z_{max})$. Here z_{min} and z_{max} are the minimum and maximum values of the quantity z_m , defined in (13a). For F -values between $-1/z_{max}$ and $-1/z_{min}$ these conditions are incompatible with the condition that E_0/K be smaller than or equal to $F^2/2$. This is exactly the range of F -values for which the steady state $X_k = X_{sk} = F$ is stable. In this range the probability density function collapses into a delta function at the stable steady state. The entropy diagram for this case is shown in Figure 7a, with an enlargement of the region around the steady state in Figure 8.

On the condition that F lies outside the region for which the steady state is stable, the constraint $\langle d^2E/dt^2 \rangle = 0$ turns out not to influence the marginal probability density function of an individual variable X_k . It results, however, in a covariance matrix C_{kl} that has nonzero off-diagonal elements. For $F = 10$ and $E_0/K = 12.92$ the function C_{kl} is plotted as a function of l for $k = 19$ in Figure 7b (open circles) together with the numerically obtained covariance matrix (solid dots). The resemblance between the two is quite good, the analytical covariances decreasing in amplitude somewhat faster with increasing $|l - k|$ than the numerical covariances. Leaving out the constraint $\langle E \rangle = E_0$ leads to energy values that are determined directly in terms of the forcing. These values are represented by the curve of intermediate thickness in Figure 7a. For $F = 10$ and the corresponding value of E_0/K , the associated covariance matrix is plotted in Figure 7c, using open squares. Also here the energy is

overestimated, reflecting itself in a maximum of the analytical C_{kl} that is larger than its numerical counterpart.

Figures 10 and 11 place the results of the maximum entropy analysis in the perspective of smaller and larger values of the forcing F . Both figures show that the marginal probability density functions and the covariance matrices show the best match between numerical and analytical results if the constraint $\langle E \rangle = E_0$ is included in the analysis. Skipping this constraint leads to an overestimation of the energy. As a consequence, the marginal probability density functions have too large mean values and the covariance matrices too high maxima. We note, however, that if the forcing F is moderate, say less than about 10, leaving out the constraint $\langle E \rangle = E_0$ gives results that are still useful as first estimates of the corresponding numerical integration. They deserve credit in relying only on the assumption that the system is in a statistically stationary state.

We recall from Lorenz and Emanuel [24] that the variables of the system proposed by Lorenz are not to be associated with a particular atmospheric field. Instead, the system is meant to be representative of general atmospheric models by incorporating the most basic atmospheric processes in an elementary way. Due to its simplicity, the system is well suited to illustrate the main message of this article: that the fundamental principles of statistical mechanics can be extended to systems out of equilibrium by using the principle of maximum entropy in combination with constraints on the average time-derivatives of the energy (or of other relevant global quantities). Future research on Lorenz' system might concern the use of time-derivatives higher than second-order, as these are expected to introduce additional detail into the probability density function. Mathematically this could pose a challenge as it will lead to non-normal (non-Gaussian) statistics. Of much interest is the application of the method to concrete physical problems, such as atmospheric convection, cloud physics or the climate itself, see Marston [29]. Research on these subjects might yield parametrizations of sub-grid scale processes that are less dependent on ad-hoc assumptions and tuning. In any case, we believe that the method holds sufficient promise to warrant testing on a variety of physical problems.

The authors would like to thank the referees for constructively criticizing their manuscript.

Appendix A: The argument of the exponential

The expression for the probability density function if all three constraints (16), (17) and (18) are applied in the maximization of entropy is given by (44), which we repeat here for convenience

$$\mathcal{P}(\mathbf{X}) = \frac{1}{\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2)} e^{-\lambda_0 E - \lambda_1 \frac{dE}{dt} - \lambda_2 \frac{d^2E}{dt^2}}.$$

In this expression λ_0 , λ_1 and λ_2 are the corresponding Lagrange multipliers and $\mathcal{Z}(\lambda_0, \lambda_1, \lambda_2)$ is the partition function. If $\lambda_0 E + \lambda_1 dE/dt + \lambda_2 d^2E/dt^2$ is written out by substituting (6), (7) and (8), we obtain:

$$\lambda_0 E + \lambda_1 \frac{dE}{dt} + \lambda_2 \frac{d^2E}{dt^2} = \sum_{k=1}^K \left((\lambda_0/2 - \lambda_1 + 2\lambda_2) X_k^2 + (\lambda_1 - 2\lambda_2 - \lambda_2) F X_k + \lambda_2 F (X_k X_{k+2} - X_k X_{k+1}) + \lambda_2 F^2 \right).$$

By placing the factor $\lambda_0/2 - \lambda_1 + 2\lambda_2$ in front of the summation sign, one obtains

$$\begin{aligned} \lambda_0 E + \lambda_1 \frac{dE}{dt} + \lambda_2 \frac{d^2E}{dt^2} &= (\lambda_0/2 - \lambda_1 + 2\lambda_2) \\ &\times \sum_{k=1}^K \left(X_k^2 + \frac{(\lambda_1 - 2\lambda_2 - \lambda_2)F}{\lambda_0/2 - \lambda_1 + 2\lambda_2} X_k + \frac{\lambda_2 F}{\lambda_0/2 - \lambda_1 + 2\lambda_2} \right. \\ &\quad \left. \times (X_k X_{k+2} - X_k X_{k+1}) + \frac{\lambda_2 F^2}{\lambda_0/2 - \lambda_1 + 2\lambda_2} \right) = \\ &(\lambda_0/2 - \lambda_1 + 2\lambda_2) \sum_{k=1}^K \left(\left(X_k + \frac{(\lambda_1 - 2\lambda_2 - \lambda_2)F}{2(\lambda_0/2 - \lambda_1 + 2\lambda_2)} \right)^2 \right. \\ &\quad \left. + \frac{\lambda_2 F}{\lambda_0/2 - \lambda_1 + 2\lambda_2} (X_k X_{k+2} - X_k X_{k+1}) \right. \\ &\quad \left. + \frac{\lambda_2 F^2}{\lambda_0/2 - \lambda_1 + 2\lambda_2} - \left(\frac{(\lambda_1 - 2\lambda_2 - \lambda_2)F}{2(\lambda_0/2 - \lambda_1 + 2\lambda_2)} \right)^2 \right). \end{aligned}$$

By defining the parameters

$$\sigma^2 = \frac{1}{2(\lambda_0/2 - \lambda_1 + 2\lambda_2)}, \quad (\text{A.1a})$$

$$\gamma = -\frac{(\lambda_1 - 2\lambda_2 - \lambda_2)F}{2(\lambda_0/2 - \lambda_1 + 2\lambda_2)}, \quad (\text{A.1b})$$

$$\varphi = \frac{\lambda_2 F}{\lambda_0/2 - \lambda_1 + 2\lambda_2}, \quad (\text{A.1c})$$

we may write

$$\begin{aligned} \lambda_0 E + \lambda_1 \frac{dE}{dt} + \lambda_2 \frac{d^2E}{dt^2} &= \frac{1}{2\sigma^2} \\ &\times \sum_{k=1}^K \left((X_k - \gamma)^2 + \varphi X_k (X_{k+2} - X_{k+1}) + \varphi F - \gamma^2 \right). \end{aligned} \quad (\text{A.2})$$

These expressions are equivalent with (45) and (46). We recover (20), (21) by keeping in these expressions only the Lagrange multiplier λ_0 , and (33), (34) by keeping only the Lagrange multipliers λ_0 and λ_1 .

Appendix B: The covariance matrix

In the case the constraint $\langle d^2E/dt^2 \rangle = 0$ is incorporated into the analysis, the elements C_{kl} of the covariance matrix

C are related to the elements D_{kl}^{-1} of the matrix D^{-1} by (47b). In this Appendix we will derive expression (52) of D_{kl}^{-1} . Obtaining the matrix D^{-1} means that we have to solve the following system of equations for Y_1, \dots, Y_K

$$\sum_{l=1}^K D_{kl} Y_l = Z_k, \quad (\text{B.1})$$

from Z_1, \dots, Z_K for $k = 1, \dots, K$. To accomplish this, we use the fact that any Y_k or Z_k can be written in terms of the eigenvectors W_{km} , i.e.,

$$Y_l = \sum_{m=1}^K U_m W_{lm}, \quad Z_k = \sum_{m=1}^K V_m W_{km},$$

and, conversely,

$$U_m = \sum_{p=1}^K Y_p W_{mp}^*, \quad V_m = \sum_{p=1}^K Z_p W_{mp}^*.$$

The latter identities can be seen to follow from

$$\sum_{p=1}^K W_{qp} W_{pm}^* = \delta_{qm},$$

which, in turn, is a consequence of the identity

$$\sum_{p=1}^K e^{i\omega(p-1)(m-q)} = K \delta_{qm},$$

and the definition (10) of W_{lm} . By expressing Y_l in terms of U_m and Z_k in terms of V_m we may write for (B.1)

$$\begin{aligned} \sum_{l=1}^K D_{kl} \left(\sum_{m=1}^K U_m W_{lm} \right) &= \sum_{m=1}^K V_m W_{km} \implies \\ \sum_{m=1}^K U_m \left(\sum_{l=1}^K D_{kl} W_{lm} \right) &= \sum_{m=1}^K V_m W_{km} \implies \\ \sum_{m=1}^K U_m \psi_m W_{km} &= \sum_{m=1}^K V_m W_{km}. \end{aligned}$$

Here we used the fact that W_{lm} is the matrix of eigenvalues of D and that ψ_m are the corresponding eigenvalues. The solution of the latter equation is simply

$$U_m = \frac{V_m}{\psi_m}.$$

For Y_l we thus find

$$\begin{aligned} Y_l &= \sum_{m=1}^K \frac{V_m}{\psi_m} W_{lm} = \sum_{m=1}^K \frac{W_{lm}}{\psi_m} \sum_{p=1}^K Z_p W_{mp}^* \\ &= \sum_{p=1}^K \left(\sum_{m=1}^K \frac{W_{lm} W_{mp}^*}{\psi_m} \right) Z_p = \sum_{k=1}^K \left(\sum_{m=1}^K \frac{W_{lm} W_{mk}^*}{\psi_m} \right) Z_k, \end{aligned}$$

or, interchanging k and l ,

$$Y_k = \sum_{l=1}^K \left(\sum_{m=1}^K \frac{W_{km}W_{ml}^*}{\psi_m} \right) Z_l.$$

It thus follows that

$$D_{kl}^{-1} = \sum_{m=1}^K \frac{W_{km}W_{ml}^*}{\psi_m} = \frac{1}{K} \sum_{m=1}^K \frac{e^{i\omega(m-1)(l-k)}}{\psi_m},$$

where we have substituted (10) in the last equality. Now, writing

$$m = K + 2 - m',$$

where m' runs from 2 to $K + 1$ (we recall here that the index $K+1$ refers to the same variable as does the index 1), we have:

$$\begin{aligned} D_{kl}^{-1} &= \frac{1}{K} \sum_{m=1}^K \frac{1}{2} \frac{e^{i\omega(m-1)(l-k)}}{\psi_m} + \frac{1}{K} \sum_{m'=2}^{K+1} \frac{1}{2} \frac{e^{i\omega(K+1-m')(l-k)}}{\psi_{K+2-m'}} \\ &= \frac{1}{K} \sum_{m=1}^K \frac{1}{2} \frac{e^{i\omega(m-1)(l-k)}}{\psi_m} + \frac{1}{K} \sum_{m'=2}^{K+1} \frac{1}{2} \frac{e^{-i\omega(m'-1)(l-k)}}{\psi_{m'}} \\ &= \frac{1}{K} \sum_{m=1}^K \frac{1}{2} \frac{e^{i\omega(m-1)(l-k)}}{\psi_m} + \frac{1}{K} \sum_{m'=1}^K \frac{1}{2} \frac{e^{-i\omega(m'-1)(l-k)}}{\psi_{m'}} \\ &= \frac{1}{K} \sum_{m=1}^K \frac{\cos(\omega(m-1)(l-k))}{\psi_m}. \end{aligned}$$

Here we used that, for the matrix D , $\psi_{K+2-m'}$ is identical to $\psi_{m'}$. Writing $\psi_m = 1 + \varphi z_m$ we thus have

$$D_{kl}^{-1} = \frac{1}{K} \sum_{m=1}^K \frac{\cos(\omega(m-1)(l-k))}{1 + \varphi z_m}.$$

This is the result (52).

References

1. E.T. Jaynes, Phys. Rev. **106**, 620 (1957)
2. E.T. Jaynes, Phys. Rev. **108**, 171 (1957)
3. E.T. Jaynes: *Papers on Probability, Statistics and Statistical Physics*, edited by R.D. Rosenkrantz (Kluwer Academic, Dordrecht, 1989)
4. E.N. Lorenz, Predictability – A problem partly solved, in *Proc. Seminar on Predictability* (ECMWF, Reading, Berkshire, UK, 1996), Vol. 1, pp. 1–18
5. D. Ruelle, *Chaotic Evolution and Strange Attractors, Lezioni Lincee* (Cambridge University Press, Cambridge, 1989)
6. J.M. Burgers, Proc. Roy. Neth. Acad. Sci. Amsterdam **17**, 1 (1939)
7. *Selected papers of J.M. Burgers*, edited by F.T.M. Nieuwstadt, J.A. Steketee (Kluwer Academic, Dordrecht, 1995)
8. L. Onsager, Nuovo Cimento **6**, 279 (1949)
9. G.L. Eyink, K.R. Sreenivasan, Rev. Mod. Phys. **78**, 87 (2006)
10. R.H. Kraichnan, D. Montgomery, Rep. Prog. Phys. **43**, 547 (1980)
11. F. Bouchet, A. Venaille, Phys. Rep. **515**, 227 (2012)
12. J. Miller, Phys. Rev. Lett. **65**, 2137 (1990)
13. J. Miller, P.B. Weichman, M.C. Cross, Phys. Rev. A **45**, 2328 (1992)
14. R. Robert, J. Sommeria, J. Fluid Mech. **229**, 291 (1991)
15. R. Robert, J. Sommeria, Phys. Rev. Lett. **69**, 2776 (1992)
16. U. Frisch, *Turbulence – The Legacy of A.N. Kolmogorov* (Cambridge University Press, Cambridge, 1995)
17. R.H. Kraichnan, Phys. Fluids **10**, 1417 (1967)
18. G.K. Batchelor, Phys. Fluids (Suppl. II) **12**, 233 (1969)
19. W.T.M. Verkley, P. Lynch, J. Atmos. Sci. **66**, 2216 (2009)
20. W.T.M. Verkley, Q.J.R. Meteorol. Soc. **137**, 1872 (2011)
21. E.T. Jaynes, *Information theory and statistical mechanics (1962)*, *Brandeis Summer Institute in Theoretical Physics*, edited by K. Ford (Benjamin, 1963)
22. E.T. Jaynes, *Probability Theory – The Logic of Science* (Cambridge University Press, Cambridge, 2003)
23. S. Pressé, K. Ghosh, J. Lee, K.A. Dill, Rev. Mod. Phys. **85**, 1115 (2013)
24. E.N. Lorenz, K.A. Emanuel, J. Atmos. Sci. **55**, 399 (1998)
25. D. Orrell, L.A. Smith, Int. J. Bifurc. Chaos **13**, 3015 (2003)
26. P.J. Davis, *Circulant Matrices* (Chelsea Publishing, New York, 1979)
27. D. Zwillinger, S. Kokoska, *Standard Probability and Statistics Tables and Formulae* (CRC Press, Boca Raton, 2000)
28. E.N. Lorenz, J. Atmos. Sci. **20**, 130 (1963)
29. B. Marston, Phys. Trends **4**, 20 (2011)

Open Access This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.