A maximum entropy approach to the problem of parametrization

W. T. M. Verkley*

Royal Netherlands Meteorological Institute, The Netherlands

*Correspondence to: W. T. M. Verkley, KNMI, PO Box 201, 3730 AE De Bilt, The Netherlands.
E-mail: verkley@knmi.nl

It is shown that the principle of maximum entropy, as formulated by Jaynes, can be applied to parametrize the effect of unresolved variables on resolved variables in a dynamical system proposed by Lorenz. The starting point is the assumption that the unresolved variables are in a state of statistical equilibrium on the time-scale of the resolved variables. The probability density function that describes the statistics of the unresolved variables is then determined by requiring that its information entropy is maximal under the constraint that the average time rate of change of the unresolved variables' energy is zero. By using this probability density function to determine the average effect of the unresolved variables on the resolved variables, a linear damping of the resolved variables is found. After incorporating the linear damping in the equations of the resolved system, the principle of maximum entropy is applied a second time to shed light on the statistics of the resolved system. The consequences are studied of using *a priori* probability density functions that go beyond Laplace's principle of indifference. Copyright © 2011 Royal Meteorological Society

1. Introduction

Fluid dynamical systems evolve on high-dimensional manifolds in a state space that is dimensionally infinite. If finite difference or spectral approximations are used to construct finite-dimensional models of these systems, then generally not all variables can be resolved explicitly. The unresolved variables have to be taken into account in some approximate sense, i.e. their influence on the resolved variables has to be parametrized. This can be done by describing the unresolved variables in terms of a probability density function (p.d.f.) that is conditioned on the resolved variables. The mean impact of the unresolved variables on the resolved variables can then be calculated by averaging over this p.d.f. The p.d.f. can also be used to produce random perturbations of the resolved variables' time rate of change. These two approaches lead to deterministic and stochastic parametrizations, respectively.

In both approaches, the central issue is the choice of the p.d.f. In the present study we propose to address this issue by applying the principle of maximum entropy (Jaynes, 1957a, 1957b; Rosenkrantz, 1989). First, the unresolved variables are assumed to be in a statistically stationary state on the time-scale of the resolved variables. More specifically, it will be assumed that the p.d.f. gives a zero averaged time derivative of the unresolved variables' energy. This assumption by itself is not sufficient to fix the p.d.f., but when used as a constraint in the maximization of the information entropy, a unique p.d.f. does result. The consequences of this approach are worked out in the context of a dynamical system devised by Lorenz (1996), analyzed following the outlines of Wilks (2005) and Crommelin and Vanden-Eijnden (2008). The approach will be shown to yield a simple linear damping of the resolved variables by the unresolved variables, the strength of which depends on the coupling between resolved and unresolved variables.
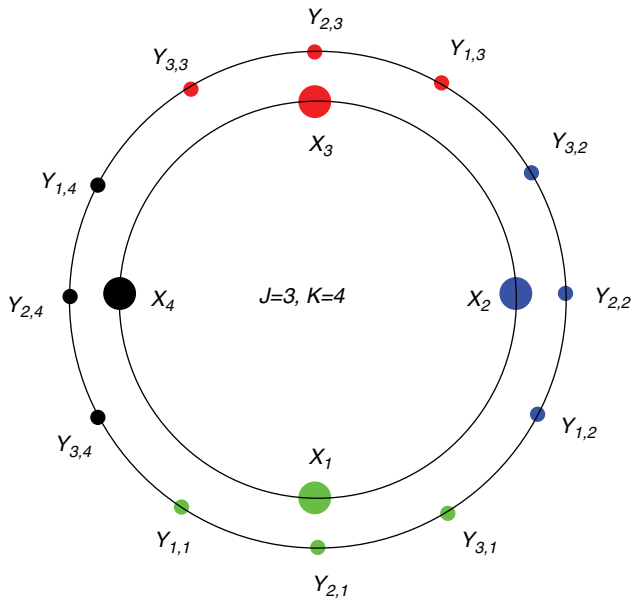
Figure 1. A schematic of the system variables for $J = 3$ and $K = 4$. The large dots represent the large-scale variables, the small dots represent the small-scale variables. Large-scale variables with a given colour (i.e. with a given index k) interact with small-scale variables of the same colour (i.e. with the index j, k).

2. The model

The system that we will study was proposed and applied by Lorenz (1996) in the context of atmospheric predictability. The model has K large-scale variables, called X_k, and JK small-scale variables, called $Y_{j,k}$, with $j = 1, ..., J$ and $k = 1, ..., K$. It is assumed that the system is cyclic in the sense that X_{k-K} and X_{k+K} are equal to X_k for all k. Also, $Y_{j,k-K}$ and $Y_{j,k+K}$ are equal to $Y_{j,k}$, whereas $Y_{j-J,k}$ equals $Y_{j,k-1}$ and $Y_{j+J,k}$ equals $Y_{j,k+1}$ for all k and j. A schematic of the system, for $J = 3$ and $K = 4$, is given in Figure 1.

The variables X_k and $Y_{j,k}$ are supposed to satisfy the following set of first-order differential equations:

$$\dot{X}_k = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F - B_k, \qquad (1)$$

$$\dot{Y}_{j,k} = c\{b\, Y_{j+1,k}(Y_{j-1,k} - Y_{j+2,k}) - Y_{j,k} + C_k\}, \qquad (2)$$

where we have defined

$$B_k = \frac{hc}{b} \sum_{j=1}^{J} Y_{j,k}, \quad C_k = \frac{h}{b} X_k. \qquad (3)$$

Both large-scale and small-scale variables have quadratic nonlinear terms that represent fluid dynamical advection. The large-scale variables X_k are forced by a forcing F, which is equal for all variables, and are damped by a linear damping $-X_k$. As the coefficient of the damping is equal to 1, the time-scale of the large-scale system is the time-scale of the damping, which Lorenz (1996) sets to 5 days.

The effect of the small-scale variables $Y_{j,k}$ on the large-scale variables X_k is given by $-B_k$; depending on the state of the small-scale variables, this term acts as a forcing or as a damping. The effect of the large-scale variables X_k on the small-scale variables $Y_{j,k}$ is given by the term C_k; it is a constant forcing for all variables $Y_{j,k}$ with a given k. The damping $-Y_{j,k}$ of the small-scale variables is linear, like the damping of the large-scale variables. The overall

time-scale of the dynamics of the small-scale variables is set by the parameter c, assumed to be positive and larger than 1. As a result, the small-scale variables evolve on shorter time-scales than the large-scale variables. According to Lorenz (1996), one could think of the variables $Y_{j,k}$ as representing convective-scale quantities and of the variables X_k as representing quantities that favour convective activity, such as the degree of static instability.

As in the original study of Lorenz (1996), we take $(K, J, F, h, c, b) = (36, 10, 10, 1, 10, 10)$, although the values of h, c and b will be varied. Integrating the system by means of a fourth-order Runge–Kutta time-stepping scheme with a time step of 0.001[*], we performed six integrations of the full system for a total period of 200 time units, i.e. 200 000 time steps. We started with a simple initial condition in which all variables are set to 0 except for the large-scale variable X_1 which is set to 0.1. Following Wilks (2005) and Crommelin and Vanden-Eijnden (2008), the results of these integrations are displayed in terms of scatter plots between B_1 and X_1. In our case, these values are plotted every 0.1 time units for the whole period of integration, the results being given in Figure 2. Figure 2(a) shows the scatter diagram in the case $(h, c, b) = (1, 10, 10)$, as in Lorenz (1996), (b) shows the scatter diagram for the case $(h, c, b) = (0.5, 10, 10)$, and (c) shows the case $(h, c, b) = (2, 10, 10)$. Figures 2(d, e, f), show the cases $(h, c, b) = (2, 10, 20)$, $(1, 5, 10)$ and $(1, 20, 10)$, respectively. We note that our choice of $k = 1$ is immaterial, i.e. the plots look the same for all k, as the system is symmetric with respect to k. In the first four columns of Table I, the system parameters are given for reference.

The different panels of Figure 2 show that for every value of the large-scale variable X_k there is a range of values of B_k. The influence of the small-scale variables on the large-scale variables, effected through the term $-B_k$ in (1), may thus depend rather sensitively on the state of the small-scale variables $Y_{j,k}$. However, we see that this sensitivity varies with the parameters of the model: in Figures 2(b and e) this sensitivity is significantly lower than in the other panels. The graphs illustrate the general idea that, to describe the effect of the small-scale (unresolved) variables on the large-scale (resolved) variables, one could use a p.d.f. $\mathcal{P}(Y_{1,1}, ..., Y_{J,K}|X_1, ..., X_K)$ of the small-scale variables $Y_{1,1}, ..., Y_{J,K}$ conditioned on the large-scale variables $X_1, ..., X_K$. In a deterministic parametrization, we then replace B_k in (1) by its expectation value over this p.d.f. As the p.d.f. depends parametrically on the large-scale variables, the result is a closed dynamical system in the large-scale variables $X_1, ..., X_K$. From the cloud of points in the panels of Figure 2, it is to be expected that, in these cases, the average effect will assume the form of a damping.

3. Small-scale variables

In this section we will discuss a procedure that will lead to a p.d.f. $\mathcal{P}(Y_{1,1}, ..., Y_{J,K}|X_1, ..., X_K)$. To this end, we consider

---

[*]Repeating the integrations with a time-step of 0.0001 did not reveal any noticeable differences in the statistics.

Table I. Values of (h, c, b) and the theoretical and numerical results of the different runs that are discussed in this article. For (K, J, F) we have $(36, 10, 10)$ in all cases. The names a, b, c, d, e and f of the runs correspond to the labels in Figures 2, 3, 4, 5 and 6. The values of κ^{num}, $\mu_{\mathrm{L}}^{\mathrm{num}}$ and $\sigma_{\mathrm{L}}^{\mathrm{num}}$ are obtained numerically by a linear least-squares fit and by calculating averages and standard deviations on the basis of the integrations that underly Figures 4, 5 and 6.

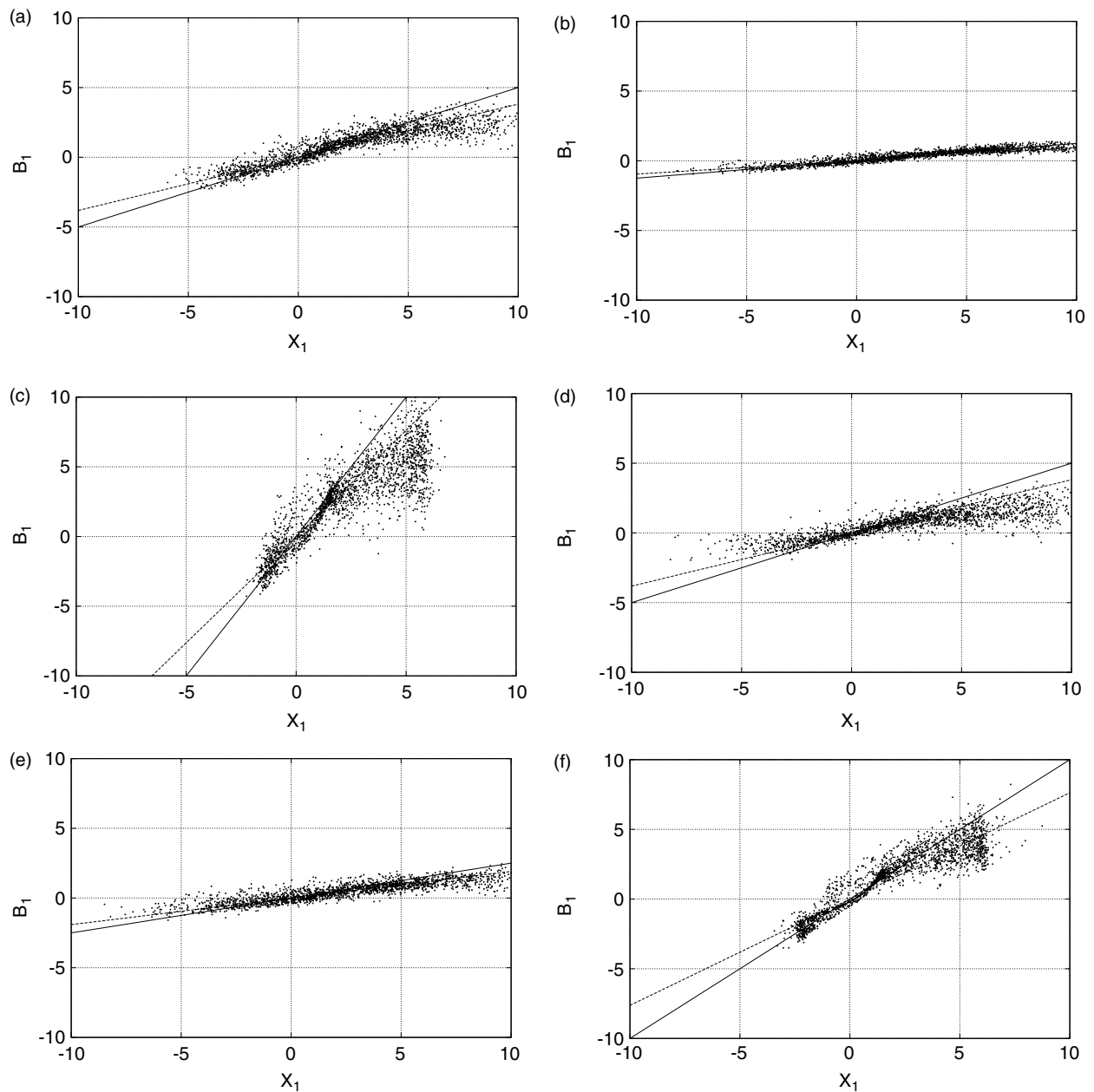| Run	h	c	b	κ	μ_{L}	σ_{L}	κ'	μ_{L}'	σ_{L}'	κ^{num}	$\mu_{\mathrm{L}}^{\mathrm{num}}$	$\sigma_{\mathrm{L}}^{\mathrm{num}}$
a	1	10	10	1.500	3.333	3.333	1.381	2.723	3.507	1.343	2.449	3.509
b	0.5	10	10	1.125	4.444	4.444	1.095	3.434	4.423	1.117	2.727	4.048
c	2	10	10	3.000	1.666	1.666	2.526	1.489	1.918	2.117	2.286	2.351
d	2	10	20	1.500	3.333	3.333	1.381	2.723	3.507	1.223	2.751	3.837
e	1	5	10	1.250	4.000	4.000	1.190	3.159	4.068	1.164	2.632	3.953
f	1	20	10	2.000	2.500	2.500	1.763	2.134	2.748	1.778	2.038	2.706

Figure 2. Scatter plots of B_1 against X_1, based on a numerical simulation of the full model equations (1), (2) and (3). The values of (h, c, b) are, from (a) to (f), $(1, 10, 10)$, $(0.5, 10, 10)$, $(2, 10, 10)$, $(2, 10, 20)$, $(1, 5, 10)$ and $(1, 20, 10)$, respectively, as summarized in the first four columns of Table I. The values of (K, J, F) are $(36, 10, 10)$ in all cases. The solid curves are the theoretical averages of B_1 as a function of X_1, given by (26). The dashed curves are the theoretical averages (72), based on the alternative *a priori* probability density function (59). Each scatter plot is based on a total number of 2001 data points.

the following global quantities:

$$E_L = \sum_{k=1}^{K} \frac{1}{2} X_k^2, \tag{4}$$

$$E_S = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{1}{2} Y_{j,k}^2. \tag{5}$$

These quantities can be regarded as the energy in the large-scale and small-scale variables, respectively. The nonlinear terms in the time-dependent equations (1) and (2) have a form that is such that, if combined with the periodicity conditions on the variables, their contribution to the time derivative of E_L and E_S vanish. We thus have

$$\dot{E}_L = \sum_{k=1}^{K} (-X_k^2 + FX_k - B_k X_k), \tag{6}$$

$$\dot{E}_S = c \sum_{j=1}^{J} \sum_{k=1}^{K} (-Y_{j,k}^2 + C_k Y_{j,k}). \tag{7}$$

Both $\dot{E}_L$ and $\dot{E}_S$ are quadratic forms, in which sense they resemble E_L and E_S themselves[†].

The p.d.f. for the small-scale variables is required to maximize the information entropy $\mathcal{S}_{IS}$, given by

$$\mathcal{S}_{IS} = -\int \cdots \int dY_{1,1} \cdots dY_{J,K}$$
$$\times \mathcal{P}(Y_{1,1}, \ldots, Y_{J,K}) \log \frac{\mathcal{P}(Y_{1,1}, \ldots, Y_{J,K})}{\mathcal{M}(Y_{1,1}, \ldots, Y_{J,K})}, \tag{8}$$

where the measure $\mathcal{M}$ is an *a priori* p.d.f. For reasons of notational convenience, we deleted the variables $X_1, \ldots, X_K$ from the argument list of $\mathcal{P}$, although it should be kept in mind that $\mathcal{P}$ depends parametrically on these large-scale variables. The same is true, generally, for $\mathcal{M}$. The constraints in the maximization of the information entropy (8) are that the average time derivative of the small-scale energy is zero,

$$\langle \dot{E}_S \rangle_S = \int \cdots \int dY_{1,1} \cdots dY_{J,K}$$
$$\times \mathcal{P}(Y_{1,1}, \ldots, Y_{J,K}) \dot{E}_S(Y_{1,1}, \ldots, Y_{J,K}) = 0, \tag{9}$$

and that the p.d.f. satisfies the normalization condition

$$\int \cdots \int dY_{1,1} \cdots dY_{J,K} \, \mathcal{P}(Y_{1,1}, \ldots, Y_{J,K}) = 1. \tag{10}$$

This is an application of Jaynes' principle of maximum entropy (Jaynes, 1957a, 1957b; Rosenkrantz, 1989)[‡]. The maximum entropy principle states that, if a system is to be described probabilistically, then its p.d.f. should be as broad as possible. The broadness is measured by the information entropy, which is to be maximized constrained by the relevant information. In our case the relevant information is

---

[†]Note that, if $\dot{E}_L$ and $\dot{E}_S$ are added, the small-scale/large-scale energy transfer terms cancel. In the absence of forcing and damping, the model of Lorenz (1996) thus conserves total energy.

[‡]The cited references to the papers by Jaynes, as well as all other manuscripts written by Jaynes, both published and unpublished, can be downloaded from http://bayes.wustl.edu/etj/node1.html

the condition that the average time derivative of E_S is zero. This condition is implied by the assumption of statistical stationarity and is evidently an important manifestation of it. It should be noted, however, that the average of higher-order time derivatives of E_S are also implied to be zero by statistical stationarity, so it is only a limited aspect of it. The issue will come up again in section 6.

There is a principal difference between the constraint of a zero time derivative of the energy and the value of the energy itself, which is the main constraint in equilibrium statistical mechanics. If the energy were available then that information would be of much value, but the energy in the small-scale variables cannot, in general, be found in terms of the large-scale variables. Indeed, this is only possible if the whole system conserves energy and this is not the case in the presence of forcing and damping. It is a fortunate state of affairs, however, that the time derivative of the energy and the energy itself are both quadratic in the variables. It implies that the mathematical analysis is similar to the analysis used in equilibrium statistical mechanics.

If the normalization condition is the only constraint in the maximization of entropy, the entropy is maximal if the p.d.f. is equal to the measure $\mathcal{M}$. This is the p.d.f. in case we only have *a priori* information. Finding the measure $\mathcal{M}$ for a given system is a non-trivial problem that in some cases can be solved by consideration of the basic symmetries of the system (Jaynes, 1968, 1973). For the moment, we note that taking $\mathcal{M}$ into account is necessary on dimensional grounds and makes the information entropy invariant to a coordinate transformation of the variables $Y_{j,k}$ once a choice of $\mathcal{M}$ has been made.

In the following we will make for $\mathcal{M}$ the simplest possible choice by taking it to be a product of constants,

$$\mathcal{M}(Y_{1,1}, \ldots, Y_{J,K}) = \prod_{j=1}^{J} \prod_{k=1}^{K} c_{j,k}^{-1}. \tag{11}$$

As we will see in due course, the actual values the constants $c_{j,k}$ (not to be confused with the constant c in the dynamical equation (2)) do not influence the resulting p.d.f. and we could just as well take them to be identical. In that case, the choice is known as Laplace's principle of indifference (Jaynes, 2003) which effectively assumes no *a priori* knowledge at all. An alternative *a priori* p.d.f. will be discussed in section 5.

In the maximization of entropy, the condition that the expectation value of $\dot{E}_S$ is zero is incorporated by using the technique of Lagrange multipliers to the constraint $\langle -(1/c)\dot{E}_S \rangle_S = 0$, the corresponding Lagrange multiplier being α. Using Eq. (12) of Verkley and Lynch (2009), this gives

$$\mathcal{P}(Y_{1,1}, \ldots, Y_{J,K}) = \frac{1}{\mathcal{Z}} \mathcal{M}(Y_{1,1}, \ldots, Y_{J,K})$$
$$\times \exp\left\{ -\alpha \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{j,k}^2 - C_k Y_{j,k}) \right\}, \tag{12}$$

where $\mathcal{Z}$, the partition function, is determined from the normalization condition. If this condition is taken into account, the above expression can be written as (Eqs. (14),

(17) and (22) of Verkley and Lynch, 2009)

$$\mathcal{P}(Y_{1,1}\ldots,Y_{J,K}) = \prod_{j=1}^{J}\prod_{k=1}^{K}\mathcal{N}(\mu_{Sk},\sigma_S,Y_{j,k}),\qquad(13)$$

where

$$\mu_{Sk} = \frac{C_k}{2}\quad\text{and}\quad\sigma_S^2 = \frac{1}{2\alpha},\qquad(14)$$

and where the functions $\mathcal{N}$,

$$\mathcal{N}(\mu,\sigma,x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\qquad(15)$$

are normal distributions with mean μ and variance σ^2. The resulting p.d.f. $\mathcal{P}$ is thus a product of normal distributions. We note that the coefficients of the *a priori* p.d.f. have been absorbed in the partition function $\mathcal{Z}$ and do not play a role in the final expression of $\mathcal{P}$.

The condition that $\mathcal{P}$ also satisfies constraint (9) fixes the value of the Lagrange multiplier α. To obtain an expression for α we need the following properties of the normal distribution:

$$\int\mathcal{N}(\mu,\sigma,x)\,\mathrm{d}x = 1,\qquad(16)$$

$$\int\mathcal{N}(\mu,\sigma,x)x\,\mathrm{d}x = \mu,\qquad(17)$$

$$\int\mathcal{N}(\mu,\sigma,x)x^2\,\mathrm{d}x = \mu^2 + \sigma^2.\qquad(18)$$

From these properties it can be deduced that

$$\langle Y_{j,k}\rangle_S = \mu_{Sk},\qquad(19)$$

$$\langle Y_{j,k}\,Y_{l,m}\rangle_S = \delta_{jl}\,\delta_{km}\,\sigma_S^2 + \mu_{Sk}\,\mu_{Sm}.\qquad(20)$$

It thus follows that the constraint (9) gives, using the first expression in (14),

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\{-\langle Y_{j,k}^2\rangle_S + C_k\langle Y_{j,k}\rangle_S\} = 0$$

$$\Rightarrow\sum_{j=1}^{J}\sum_{k=1}^{K}\{-\sigma_S^2 - \mu_{Sk}^2 + 2\mu_{Sk}^2\} = 0$$

$$\Rightarrow\sum_{k=1}^{K}\{-\sigma_S^2 + \mu_{Sk}^2\} = 0,\qquad(21)$$

from which it may be concluded that, again using the first expression in (14),

$$\sigma_S^2 = \frac{1}{K}\sum_{k=1}^{K}\mu_{Sk}^2 = \frac{1}{K}\sum_{k=1}^{K}\left(\frac{C_k}{2}\right)^2.\qquad(22)$$

The value of the Lagrange multiplier α can now be obtained from the second expression in (14). This concludes the calculation of the p.d.f. $\mathcal{P}$. We note that the means μ_{Sk} and the variance σ_S^2 both depend on the state of the large-scale system, as expressed by the variables $X_1,\ldots,X_K$. The variance σ_S^2 is actually proportional to the energy E_L, as can be seen

by comparing (22) with (4), using the second expression in (3).

From the p.d.f. of the small-scale variables $Y_{j,k}$, it is possible to obtain the p.d.f. of the terms B_k. From the first expression in (3) it can be seen that B_k is $(hc)/b$ times a sum of independent normally distributed variables $Y_{j,k}$, so that the p.d.f. of B_k is given by a similar expression,

$$\mathcal{P}(B_1,\ldots,B_K) = \prod_{k=1}^{K}\mathcal{N}(\mu_{Bk},\sigma_B,B_k),\qquad(23)$$

where the means μ_{Bk} and the variance σ_B^2 are given by (items 3.6.11 and 4.2.1 of Zwillinger and Kokoska, 2000)

$$\mu_{Bk} = \frac{hc}{b}\sum_{j=1}^{J}\mu_{Sk}\quad = Jc\left(\frac{h}{b}\right)\mu_{Sk},\qquad(24)$$

$$\sigma_B^2 = \left(\frac{hc}{b}\right)^2\sum_{j=1}^{J}\sigma_S^2\quad = Jc^2\left(\frac{h}{b}\right)^2\sigma_S^2,\qquad(25)$$

i.e. a product of K normal distributions with means μ_{Bk} and variance σ_B^2. The expressions for μ_{Bk} and σ_B^2 can be checked to be identical to $\langle B_k\rangle_S$ and $\langle(B_k - \langle B_k\rangle_S)^2\rangle_S = \langle B_k^2\rangle_S - \langle B_k\rangle_S^2$, respectively, by applying the identities (19) and (20) to the first expression of (3). Substituting (14)(first expression) and (22) for μ_{Sk} and σ_S^2, respectively, and using the second expression of (3), we get

$$\mu_{Bk} = J\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2 X_k,\qquad(26)$$

$$\sigma_B^2 = J\left(\frac{c}{2}\right)^2\left(\frac{h}{b}\right)^4\frac{1}{K}\sum_{k=1}^{K}X_k^2,\qquad(27)$$

which give μ_{Bk} and σ_B^2 directly in terms of the large-scale variables X_k.

The deterministic parametrization is now obtained by replacing the term B_k in (1) by its expectation value $\langle B_k\rangle_S = \mu_{Bk}$, as given by (26). The equation for the large-scale variables then becomes

$$\dot{X}_k = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F - J\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2 X_k.\qquad(28)$$

By defining an effective damping coefficient κ that includes the contribution of the parametrized small-scale variables,

$$\kappa \equiv 1 + J\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2,\qquad(29)$$

the system may be written

$$\dot{X}_k = X_{k-1}(X_{k+1} - X_{k-2}) - \kappa X_k + F.\qquad(30)$$

As c was assumed positive, our approach is seen to yield an additional linear damping of the large-scale variables. We note that the extra damping is proportional to J and c and varies quadratically with h/b. A stronger coupling between resolved and unresolved variables thus implies a stronger damping or, equivalently, a damping time-scale that is shorter. Note that the quadratic dependence of the damping on h/b results from the fact that both B_k and C_k

are proportional to h/b. Values of the resulting effective damping coefficient κ, given by (29), are given in the fifth column of Table I. The theoretical expression of $\langle B_k \rangle_S = \mu_{Bk}$ is represented by the solid curves in Figure 2. We see that it gives a reasonably good representation of the average values of the numerically obtained distributions, however the slopes are somewhat too large, in particular for large values of $|X_k|$.

4. Large-scale variables

We now address the question whether the maximum entropy approach enables us to predict more details of the scatter plots of Figure 2. Information on the mean and variance of X_k and thereby information on the range of X_k and B_k will allow us to predict the horizontal and the vertical extent of the cloud of points in the scatter plots of Figure 2. We recall that the mean μ_{Bk} depends on X_k only, but that the variance σ_B^2 depends on the sum over $k = 1, ..., K$ of X_k^2. In the following we attempt to shed light on the statistics of the large-scale variables $X_1, ..., X_K$ by applying the principle of maximum entropy to the large-scale system, parametrizing the effect of the small-scale variables in the way described above.

Replacing B_k by its expectation value $\langle B_k \rangle_S = \mu_{Bk}$, as given by (26), we have for the time derivative of the large-scale energy, given by (6),

$$\dot{E}_L = \sum_{k=1}^{K} \left\{ -X_k^2 + FX_k - J\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2 X_k^2 \right\} \quad (31)$$

$$= \sum_{k=1}^{K} (-\kappa X_k^2 + FX_k), \quad (32)$$

where κ is defined by (29). We will now assume that the large-scale system is in a statistically stationary state and describe the large-scale variables X_k in terms of a p.d.f. $\mathcal{P}(X_1, ..., X_K)$. In analogy to the procedure described in the previous section, we will require that the information entropy $\mathcal{S}_{IL}$ is maximal under the constraint that the average time derivative of E_L is zero and that the p.d.f. is normalized. The information entropy $\mathcal{S}_{IL}$ is defined as

$$\mathcal{S}_{IL} = - \int \cdots \int dX_1 \cdots dX_K$$
$$\times \mathcal{P}(X_1, \ldots, X_K) \log \frac{\mathcal{P}(X_1, \ldots, X_K)}{\mathcal{M}(X_1, \ldots, X_K)}. \quad (33)$$

The definition is identical to (8), the large-scale variables X_k replacing the small-scale variables $Y_{j,k}$.

Again, the *a priori* p.d.f. is assumed to be a product of constants

$$\mathcal{M}(X_1, ..., X_K) = \prod_{k=1}^{K} c_k^{-1}. \quad (34)$$

Using as a constraint that $\langle (-1/\kappa)\dot{E}_L \rangle_L = 0$, with Lagrange multiplier β, we find, using Eq. (12) of Verkley and Lynch (2009) in much the same way as before,

$$\mathcal{P}(X_1, \ldots, X_K) = \frac{1}{\mathcal{Z}} \mathcal{M}(X_1, \ldots, X_K)$$
$$\times \exp\left\{ -\beta \sum_{k=1}^{K} \left(X_k^2 - \frac{F}{\kappa} X_k \right) \right\}, \quad (35)$$

where $\mathcal{Z}$ is determined from the normalization condition. The resulting normalized p.d.f. is seen to be a product of K independent normal distributions (Eqs (14), (17) and (22) of Verkley and Lynch, 2009),

$$\mathcal{P}(X_1, ..., X_K) = \prod_{k=1}^{K} \mathcal{N}(\mu_L, \sigma_L, X_k), \quad (36)$$

with

$$\mu_L = \frac{F}{2\kappa} \quad \text{and} \quad \sigma_L^2 = \frac{1}{2\beta}. \quad (37)$$

The value of the Lagrange multiplier β can be found by using the following identities, analogous to (19) and (20),

$$\langle X_k \rangle_L = \mu_L, \quad (38)$$
$$\langle X_k X_m \rangle_L = \delta_{km} \sigma_L^2 + \mu_L^2. \quad (39)$$

Writing out the constraint $\langle \dot{E}_L \rangle_L = 0$ and using these identities in addition to the first expression in (37), one finds

$$\sigma_L^2 = \mu_L^2 = \left(\frac{F}{2\kappa}\right)^2, \quad (40)$$

from which β follows by using the second expression in (37). We see that the normal distributions in (36) all have the same average μ_L and variance σ_L^2. The values of μ_L and σ_L, corresponding to the parameters of the different runs, are listed in the sixth and seventh columns of Table I.

The theoretical equivalent of a scatter diagram, such as shown in Figure 2, is the p.d.f. $\mathcal{P}(X_1, B_1)$, given by

$$\mathcal{P}(X_1, B_1) = \int \cdots \int dX_2 \cdots dX_K \, dB_2 \cdots dB_K$$
$$\times \mathcal{P}(X_1, ..., X_K, B_1, ..., B_K), \quad (41)$$

in which $\mathcal{P}(X_1, ..., X_K, B_1, ..., B_K)$ is the joint p.d.f. of finding the system in the state $(X_1, ..., X_K, B_1, ..., B_K)$. The latter p.d.f. is given by the following product, an expression of the product rule of probabilities,

$$\mathcal{P}(X_1, ..., X_K, B_1, ..., B_K) =$$
$$\mathcal{P}(X_1, ..., X_K)\mathcal{P}(B_1, ..., B_K|X_1, ..., X_K). \quad (42)$$

Here $\mathcal{P}(B_1, ..., B_K|X_1, ..., X_K)$ is the p.d.f. of the variables B_k, conditioned on the large-scale variables X_k, for which an expression was derived in the previous section and $\mathcal{P}(X_1, ..., X_K)$ is the p.d.f. of the large-scale variables X_k, discussed above. Upon substituting the p.d.f. as given by (23) and (36) into (41) and (42), we obtain

$$\mathcal{P}(X_1, B_1) = \int \cdots \int dX_2 \cdots dX_K \, dB_2 \cdots dB_K$$
$$\times \prod_{k=1}^{K} \mathcal{N}(\mu_L, \sigma_L, X_k) \prod_{m=1}^{K} \mathcal{N}(\mu_{Bm}, \sigma_B, B_m). \quad (43)$$

The integrals over $B_2, ..., B_K$ can be carried out immediately and the normal distribution in the variable X_1 can be placed in front of the integral. This gives

$$\mathcal{P}(X_1, B_1) = \mathcal{N}(\mu_L, \sigma_L, X_1)$$
$$\times \int \cdots \int dX_2 \cdots dX_K \prod_{k=2}^{K} \mathcal{N}(\mu_L, \sigma_L, X_k)\mathcal{N}(\mu_{B1}, \sigma_B, B_1). \quad (44)$$
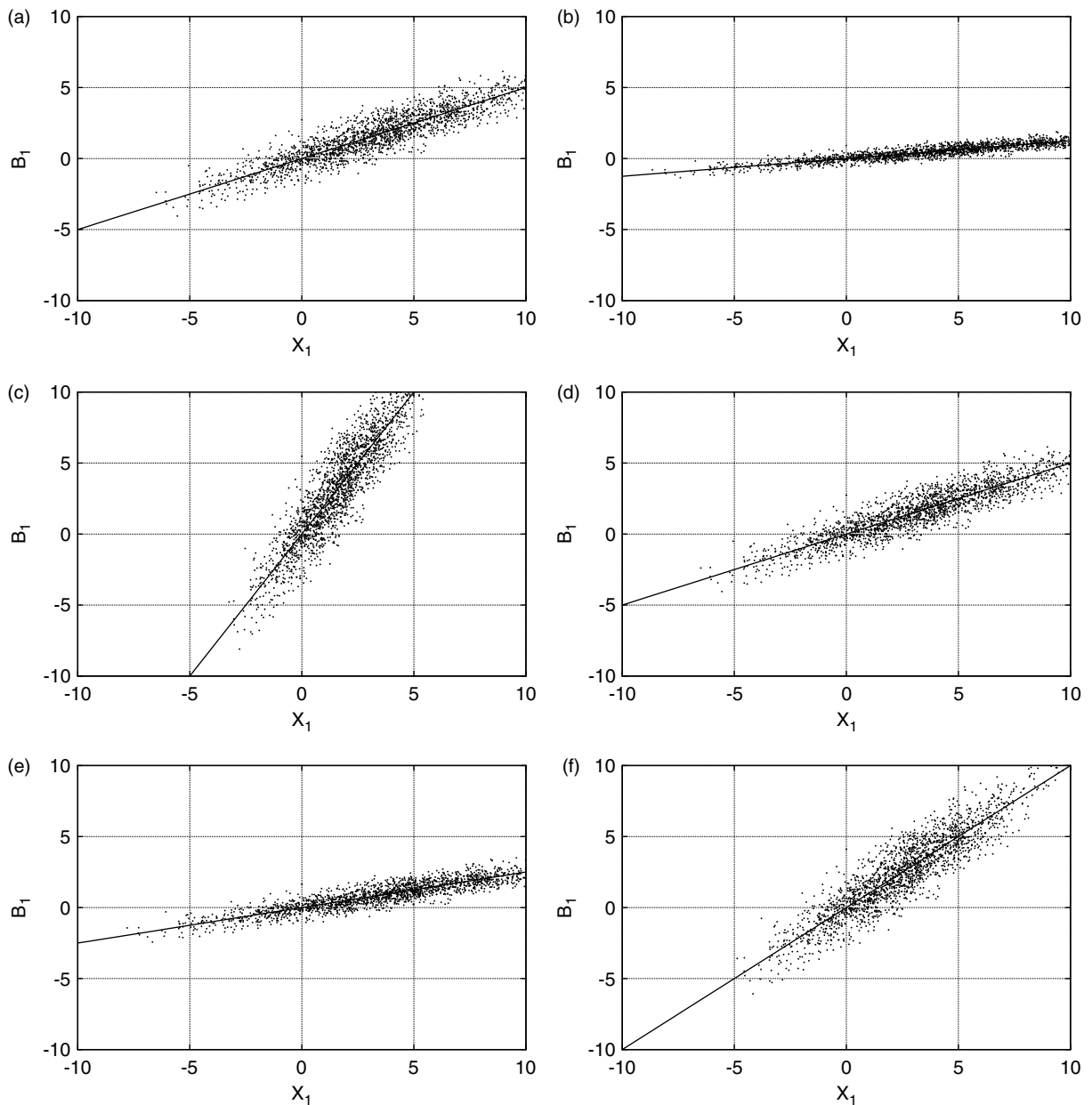
Figure 3. Scatter plots of B_1 against X_1, calculated by randomly drawing from the theoretical probability density function $\mathcal{P}(X_1, B_1)$, given by (51). The panels of this figure and of the following figures are ordered as in Figure 2, with the parameters of the different panels given in the first four columns of Table I. The solid lines show the theoretical average μ_{B1}, given by (26). The plots are based on 2001 data points.

We see that the normal distribution in B_1 depends on the variables $X_2, ..., X_K$ via σ_B for which we have ((27)),

$$\sigma_B = \sigma_B(X_1, Z)$$
$$= \left(\frac{J}{K}\right)^{1/2}\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2 \sigma_L \left\{\left(\frac{X_1}{\sigma_L}\right)^2 + Z\right\}^{1/2}, \quad (45)$$

where

$$Z = \sum_{k=2}^{K}\left(\frac{X_k}{\sigma_L}\right)^2. \quad (46)$$

The variables X_k are normally distributed with non-zero mean μ_L. The integral over $X_2, .., X_K$ can therefore be replaced by an integral over Z, from 0 to ∞, where Z is distributed according to the non-central χ^2 distribution.

We thus have

$$\mathcal{P}(X_1, B_1) = \mathcal{N}(\mu_L, \sigma_L, X_1)$$
$$\times \int dZ\, p(Z)\mathcal{N}\{\mu_{B1}(X_1), \sigma_B(X_1, Z), B_1\}, \quad (47)$$

where the non-central χ^2 distribution $p(Z)$ is given by (Johnson *et al.*, 1995, Eq. (29.4))

$$p(Z) = \frac{1}{2}\exp\left(-\frac{Z+\Lambda}{2}\right)\left(\frac{Z}{\Lambda}\right)^{\nu/4-1/2}I_{\nu/2-1}[(\Lambda Z)^{1/2}]. \quad (48)$$

Here I is a modified Bessel function of the first kind, ν is the number of degrees of freedom (the number of terms in Z) and Λ is the non-centrality parameter (the value of Z if the means of X_k are substituted for X_k). In our case these

parameters are given by, using (40) in the second expression below,

$$\nu = K - 1 \quad \text{and} \quad \Lambda = \sum_{k=2}^{K} \left(\frac{\mu_L}{\sigma_L}\right)^2 = K - 1. \quad (49)$$

For the mean value of Z, averaged over the distribution (48), we have the following simple expression (Johnson *et al.*, 1995, Eq. (29.28))

$$\langle Z \rangle = \nu + \Lambda = 2(K - 1), \quad (50)$$

an expression that can also be derived from (39) and (46).

We numerically evaluated the integral in (47) by using the routine `bessik` from Press *et al.* (1996) for the modified Bessel function I and integrating by means of the trapezoidal rule with 200 values of Z, evenly distributed over the interval $[0, 200]$. In the course of experimenting we noted that a satisfactory approximation[§] is obtained by replacing $p(Z)$ by a delta function at the average value $\langle Z \rangle$. This gives

$$\mathcal{P}(X_1, B_1) \approx \mathcal{N}(\mu_L, \sigma_L, X_1) \\ \times \mathcal{N}\{\mu_{B1}(X_1), \sigma_B(X_1, \langle Z \rangle)), B_1\}, \quad (51)$$

where we recall that $\mu_{B1}(X_1)$ is given by (26) and $\sigma_B(X_1, Z)$ by (45).

In Figure 3, we show scatter plots generated by randomly drawing from the p.d.f. (51). Using the routine `gasdev` from Press *et al.* (1996), we calculated 2001 pairs of values X_1 and B_1 using the same parameters as used for the scatter plots in Figure 2, i.e. for (a), (b), (c), (d), (e) and (f) we have $(h, c, b) = (1, 10, 10), (0.5, 10, 10), (2, 10, 10), (2, 10, 20), (1, 5, 10)$ and $(1, 20, 10)$, respectively. The other parameters are the same as before, i.e. $(K, J, F) = (36, 10, 10)$. The solid lines are the theoretical averages μ_{B1}, given by (26), as in Figure 2. Note that the theoretical scatter diagrams of Figures 3(a) and (d) are identical, as only the fraction h/b enters into the theoretical expressions. The figures compare reasonably well with the scatter plots in Figure 2, given that no time integration had to be carried out to produce them. In accordance with remarks made earlier on the theoretical average μ_{Bk}, we see that the theoretical scatter plots of Figure 3 are somewhat steeper than those of Figure 2. We also notice that, generally, the theoretical scatter plots have a larger spread in values of B_1 than the numerical scatter plots.

5. *A priori* probability

In choosing the *a priori* p.d.f.s (11) and (34), we have been guided by Laplace's principle of indifference. We will now consider the consequences of an alternative possibility, anticipating that this might influence the theoretical average μ_{Bk}. A consideration that we did not take into account is that the presence of forcing and damping causes the energy of the small-scale and large-scale systems to be bounded from above. This effectively limits the volume in phase space in which these systems must reside if they are in a statistically stationary state. The derivation that follows is consistent with Eq. (7) of Lorenz (1963), an equation from which, for a wide variety of dynamical systems, it can be concluded that the attractor occupies a finite region of phase space.

5.1. *Small-scale variables*

We begin with the small-scale variables and apply Schwarz's inequality[¶]

$$\sum_{j=1}^{J} \sum_{k=1}^{K} C_k Y_{j,k} \leq \left(\sum_{j=1}^{J} \sum_{k=1}^{K} C_k^2\right)^{1/2} \left(\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2\right)^{1/2}, \quad (52)$$

from which it follows that

$$\dot{E}_S \leq c \left(\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2\right)^{1/2} \\ \times \left\{-\left(\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2\right)^{1/2} + \left(\sum_{j=1}^{J} \sum_{k=1}^{K} C_k^2\right)^{1/2}\right\}. \quad (53)$$

We therefore have the implication

$$\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2 > \sum_{j=1}^{J} \sum_{k=1}^{K} C_k^2 \Rightarrow \dot{E}_S < 0, \quad (54)$$

or

$$E_S > \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{1}{2} C_k^2 \Rightarrow \dot{E}_S < 0. \quad (55)$$

This means that, in a statistically stationary state in which we cannot have a steady decrease of energy, we should have that

$$\sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2 \leq \sum_{j=1}^{J} \sum_{k=1}^{K} C_k^2, \quad (56)$$

from which it may be concluded that the unresolved variables' energy is bounded from above by the values of the resolved variables. We may translate this into an *a priori* p.d.f. $\mathcal{M}^*$ of the form

$$\mathcal{M}^*(Y_{1,1}, ..., Y_{J,K}) = A\mathcal{H}\left(\sum_{j=1}^{J} \sum_{k=1}^{K} C_k^2 - \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2\right), \quad (57)$$

where $\mathcal{H}(x)$ is the Heaviside step function and A is a normalization constant.

In order to check the consequences of using $\mathcal{M}^*$ instead of $\mathcal{M}$ without getting into too much calculational trouble, we approximate $\mathcal{M}^*$ by $\mathcal{M}'$:

$$\mathcal{M}'(Y_{1,1}, ..., Y_{J,K}) \\ = B \exp\left\{\rho\left(\sum_{j=1}^{J} \sum_{k=1}^{K} C_k^2 - \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2\right)\right\}, \quad (58)$$

[§]The differences between (51) and (47) are between 1 and 2% of the maximum value of (47) for the cases displayed in Figure 2.

[¶]For the reasoning that leads to the implication (55), the author is indebted to his colleague Mr Roel Stappers.

where B is a normalization constant and the parameter ρ is determined by requiring that $\mathcal{M}^*$ and $\mathcal{M}'$ give the same expectation value of E_S. As shown in the appendix, this gives

$$\mathcal{M}'(Y_{1,1}, ..., Y_{J,K})$$
$$= \left(\frac{1}{\sigma_{MS}\sqrt{2\pi}}\right)^{JK} \exp\left(-\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{Y_{j,k}^2}{2\sigma_{MS}^2}\right), \quad (59)$$

where

$$\sigma_{MS}^2 = \frac{1}{JK+2}\sum_{j=1}^{J}\sum_{k=1}^{K}C_k^2 = \frac{\sigma_S^2}{\gamma_S^2}. \quad (60)$$

Here we have defined a parameter γ_S^2,

$$\gamma_S^2 \equiv \frac{1}{4}\frac{JK+2}{JK}, \quad (61)$$

and σ_S^2 is given by (22).

If we now, in the maximization of entropy, go back to (12) and substitute $\mathcal{M}'$ for $\mathcal{M}$, we obtain

$$\mathcal{P}'(Y_{1,1}, ..., Y_{J,K}) = \frac{1}{\mathcal{Z}}\left(\frac{1}{\sigma_{MS}\sqrt{2\pi}}\right)^{JK}$$
$$\times \exp\left\{-\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\frac{Y_{j,k}^2}{2\sigma_{MS}^2} + \alpha Y_{j,k}^2 - \alpha C_k Y_{j,k}\right)\right\}. \quad (62)$$

Using the definition of the normal distribution and expressing C_k in terms of μ_{Sk} by using (14), this can be rewritten as (13)

$$\mathcal{P}'(Y_{1,1}..., Y_{J,K}) = \prod_{j=1}^{J}\prod_{k=1}^{K}\mathcal{N}(\mu'_{Sk}, \sigma'_S, Y_{j,k}), \quad (63)$$

where

$$\mu'_{Sk} = \frac{\sigma_{MS}^2}{\sigma_{MS}^2 + 1/(2\alpha)}\mu_{Sk}, \quad (64)$$

$$\sigma_S'^2 = \frac{\sigma_{MS}^2}{\sigma_{MS}^2 + 1/(2\alpha)}\frac{1}{2\alpha}. \quad (65)$$

The Lagrange multiplier α can be found from the condition that $\langle \dot{E}_S\rangle'_S = 0$, the prime denoting that the average is taken over the primed p.d.f. Using (7), (14), (19), (20) and (22), we find

$$\frac{1}{2\alpha} = \frac{\sigma_{MS}^2 + 1/\alpha}{\sigma_{MS}^2 + 1/(2\alpha)}\sigma_S^2. \quad (66)$$

From this expression, $1/(2\alpha)$ may be obtained by solving a quadratic equation. Keeping only the root that leads to a postive value of $1/(2\alpha)$, this gives

$$\frac{1}{2\alpha} = \sigma_{MS}^2\left[\frac{\sigma_S^2}{\sigma_{MS}^2} - \frac{1}{2} + \left\{\left(\frac{\sigma_S^2}{\sigma_{MS}^2}\right)^2 + \frac{1}{4}\right\}^{1/2}\right]. \quad (67)$$

Using (60), this can be written as

$$\frac{1}{2\alpha} = \sigma_{MS}^2\left[\gamma_S^2 - \frac{1}{2} + \left\{\gamma_S^4 + \frac{1}{4}\right\}^{1/2}\right]. \quad (68)$$

If substituted in (64) and (65) we obtain

$$\mu'_{Sk} = f_S\,\mu_{Sk} \quad \text{and} \quad \sigma_S'^2 = g_S^2\,\sigma_S^2, \quad (69)$$

where

$$f_S = \frac{1}{\gamma_S^2 + \frac{1}{2} + (\gamma_S^4 + \frac{1}{4})^{1/2}}, \quad (70)$$

$$g_S^2 = \frac{\gamma_S^2 - \frac{1}{2} + (\gamma_S^4 + \frac{1}{4})^{1/2}}{\gamma_S^2 + \frac{1}{2} + (\gamma_S^4 + \frac{1}{4})^{1/2}}\frac{1}{\gamma_S^2}. \quad (71)$$

If we take the values $J = 10$ and $K = 36$, we get $\gamma_S^2 = 0.251$ and thus $f_S = 0.763$ and $g_S = 0.971$.

If μ'_{Sk} is substituted into (24) and $\sigma_S'^2$ is substituted into (25), we obtain

$$\mu'_{Bk} = f_S J\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2 X_k, \quad (72)$$

$$\sigma_B'^2 = g_S^2 J\left(\frac{c}{2}\right)^2\left(\frac{h}{b}\right)^4\frac{1}{K}\sum_{k=1}^{K}X_k^2, \quad (73)$$

whereas for κ' we get

$$\kappa' \equiv 1 + f_S J\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2. \quad (74)$$

Thus, if we take (59) as the *a priori* p.d.f. for the small-scale variables, the theoretical lines in the scatter diagrams of Figure 2 acquire a slope that is 0.763 times the original value. In Figure 2 these theoretical lines are dashed. The resulting values of κ', listed in the eighth column of Table I, become smaller as well. Consequently, the effective damping of the large-scale variables by the small-scale variables becomes less. From the fact that g_S is not much smaller than 1, we expect to see not much difference in terms of the vertical spread of the scatter plots.

5.2. Large-scale variables

We next consider an alternative to the *a priori* p.d.f. (34) and investigate the implications for the statistics of the large-scale variables. From (32) we may deduce, using Schwarz's inequality in much the same way as before[||]

$$E_L > \sum_{k=1}^{K}\frac{1}{2}\left(\frac{F}{\kappa}\right)^2 \Rightarrow \dot{E}_L < 0. \quad (75)$$

In a statistically stationary state, we should thus have that

$$\sum_{k=1}^{K}X_k^2 \leq \sum_{k=1}^{K}\left(\frac{F}{\kappa}\right)^2, \quad (76)$$

[||]For reasons of notational simplicity, we use an unaccented variable κ, but this should be replaced by an accented κ if the modified *a priori* p.d.f. for the small-scale variables is used.

from which it may be concluded that the large-scale variables' energy is bounded from above, in this case by the forcing F and the effective damping coefficient κ. A p.d.f. that takes this into account could be

$$\mathcal{M}^*(X_1, ..., X_K) = A\mathcal{H}\left\{\sum_{k=1}^{K}\left(\frac{F}{\kappa}\right)^2 - \sum_{k=1}^{K} X_k^2\right\}. \quad (77)$$

As before, for reasons of computational convenience, $\mathcal{M}^*$ is approximated by $\mathcal{M}'$:

$$\mathcal{M}'(X_1, ..., X_K) = B\exp\left[\rho\left\{\sum_{k=1}^{K}\left(\frac{F}{\kappa}\right)^2 - \sum_{k=1}^{K} X_k^2\right\}\right], \quad (78)$$

where the parameter ρ is determined by requiring that $\mathcal{M}^*$ and $\mathcal{M}'$ give the same expectation value of E_L. This leads to

$$\mathcal{M}'(X_1, ..., X_K) = \left(\frac{1}{\sigma_{ML}\sqrt{2\pi}}\right)^K \exp\left(-\sum_{k=1}^{K}\frac{X_k^2}{2\sigma_{ML}^2}\right), \quad (79)$$

where

$$\sigma_{ML}^2 = \frac{1}{K+2}\sum_{k=1}^{K}\left(\frac{F}{\kappa}\right)^2 = \frac{\sigma_L^2}{\gamma_L^2}. \quad (80)$$

Here we have defined a parameter γ_L^2,

$$\gamma_L^2 \equiv \frac{1}{4}\frac{K+2}{K}, \quad (81)$$

and σ_L^2 is given by (40).

If we go back to (35) and substitute $\mathcal{M}'$ for $\mathcal{M}$, we obtain

$$\mathcal{P}'(X_1, ..., X_K) = \frac{1}{\mathcal{Z}}\left(\frac{1}{\sigma_{ML}\sqrt{2\pi}}\right)^K$$
$$\times \exp\left[-\sum_{k=1}^{K}\left\{\frac{X_k^2}{2\sigma_{ML}^2} + \beta X_k^2 - \beta\left(\frac{F}{\kappa}\right) X_k\right\}\right]. \quad (82)$$

This can be rewritten as (36)

$$\mathcal{P}'(X_1 ..., X_K) = \prod_{k=1}^{K}\mathcal{N}(\mu_L', \sigma_L', X_k), \quad (83)$$

where

$$\mu_L' = \frac{\sigma_{ML}^2}{\sigma_{ML}^2 + 1/(2\beta)}\mu_L, \quad (84)$$

$$\sigma_L'^2 = \frac{\sigma_{ML}^2}{\sigma_{ML}^2 + 1/(2\beta)}\frac{1}{2\beta}. \quad (85)$$

The Lagrange multiplier β can be found from the condition that $\langle \dot{E}_L \rangle_L' = 0$ and gives, in the same manner as before,

$$\frac{1}{2\beta} = \sigma_{ML}^2\left[\frac{\sigma_L^2}{\sigma_{ML}^2} - \frac{1}{2} + \left\{\left(\frac{\sigma_L^2}{\sigma_{ML}^2}\right)^2 + \frac{1}{4}\right\}^{1/2}\right]. \quad (86)$$

Using (80), this can be written as

$$\frac{1}{2\beta} = \sigma_{ML}^2\left[\gamma_L^2 - \frac{1}{2} + \left\{\gamma_L^4 + \frac{1}{4}\right\}^{1/2}\right]. \quad (87)$$

If substituted in (84) and (85), we obtain

$$\mu_L' = f_L\,\mu_L \quad \text{and} \quad \sigma_L'^2 = g_L^2\,\sigma_L^2, \quad (88)$$

where

$$f_L = \frac{1}{\gamma_L^2 + \frac{1}{2} + (\gamma_L^4 + \frac{1}{4})^{1/2}}, \quad (89)$$

$$g_L^2 = \frac{\gamma_L^2 - \frac{1}{2} + (\gamma_L^4 + \frac{1}{4})^{1/2}}{\gamma_L^2 + \frac{1}{2} + (\gamma_L^4 + \frac{1}{4})^{1/2}}\frac{1}{\gamma_L^2}. \quad (90)$$

For $K = 36$, we get $\gamma_L^2 = 0.264$ and this gives $f_L = 0.752$ and $g_L = 0.969$.

Quite analogous to the small-scale variables, the mean of the distributions is shifted to values that are a factor 0.752 smaller, whereas the variance does not change very much. Values of μ_L' and σ_L', using κ' in the definition of μ_L and σ_L, are given in the ninth and tenth columns of Table I.

To further investigate the effects of the alternative *a priori* p.d.f.s, we now discuss the theoretical scatter plots. Equation (51) becomes, after replacing the different parameters by their primed equivalents,

$$\mathcal{P}'(X_1, B_1) \approx \mathcal{N}(\mu_L', \sigma_L', X_1)$$
$$\times \mathcal{N}\{\mu_{B1}'(X_1), \sigma_B'(X_1, \langle Z'\rangle'), B_1\}. \quad (91)$$

Here μ_L' and σ_L' are given by (88) above, μ_{Bk}' is given by (72) and for σ_B' we have from (73)

$$\sigma_B'(X_1, Z') = g_S\left(\frac{J}{K}\right)^{1/2}\left(\frac{c}{2}\right)\left(\frac{h}{b}\right)^2\sigma_L'\left\{\left(\frac{X_1}{\sigma_L'}\right)^2 + Z'\right\}^{1/2}, \quad (92)$$

where Z' is defined by

$$Z' = \sum_{k=2}^{K}\left(\frac{X_k}{\sigma_L'}\right)^2. \quad (93)$$

In the same way as before, we have that Z' is distributed according to the non-central χ^2 distribution with $\nu = K - 1$ degrees of freedom and the following expression of the non-centrality parameter:

$$\Lambda' = \sum_{k=2}^{K}\left(\frac{\mu_L'}{\sigma_L'}\right)^2 = \left(\frac{f_L}{g_L}\right)^2(K - 1). \quad (94)$$

For $\langle Z'\rangle'$ we thus obtain

$$\langle Z'\rangle' = \nu + \Lambda' = \left\{1 + \left(\frac{f_L}{g_L}\right)^2\right\}(K - 1). \quad (95)$$

If we substitute expression (74) of κ' in the expressions for μ_L and σ_L, we may use (91) to calculate a theoretical scatter diagram such as Figure 3. In accordance with the dashed lines in Figure 2, the cloud of points is expected to have a somewhat smaller slope. We calculated these scatter diagrams (not shown) and noticed that the cloud of points is slightly thinner than in Figure 3. Taking an overall view, these features bring the results somewhat closer to the numerically obtained scatter plots, displayed in Figure 2.
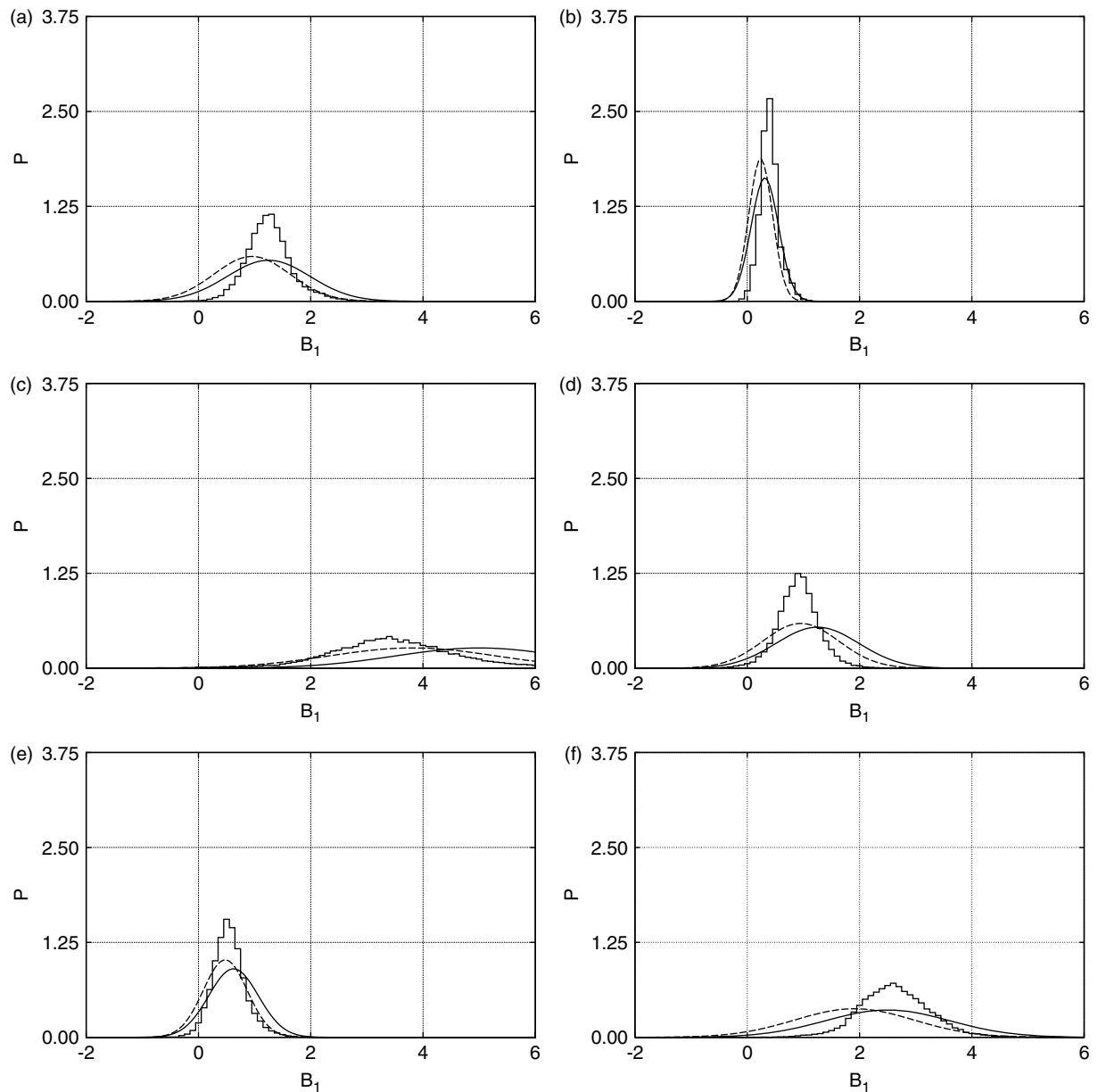
Figure 4. Numerically obtained probability density functions calculated by binning the values of B_1 from a numerical time integration of 20 000 time units, selecting from data points in which X_1 lies within an interval of width 1 around $X_1^a = 2.50$. Out of the total number of 200 001 data points, an average of 18 500 satisfy the selection criterion on X_1. The results are displayed in the form of step functions (histograms). The smooth solid curves are the theoretical probability density functions as given by (96), based on the *a priori* probability density functions (11) and (34). The smooth dashed curves are given by (97), based on the *a priori* probability density functions (59) and (79).

6. Discussion

In line with the work of Wilks (2005) and Crommelin and Vanden-Eijnden (2008), we will discuss in somewhat more detail the p.d.f.s of X_1 and B_1, given by (51) and (91). Interpreting the right-hand sides of these expressions in terms of the product rule of probabilities, we may consider the second factor as the p.d.f. of B_1 conditioned on X_1. An impression of the width of the theoretical scatter diagrams is obtained by evaluating these expressions at a fixed value of X_1, for which we take $X_1 = X_1^a = 2.50$. For the unprimed and primed cases we then have

$$\mathcal{P}(B_1 | X_1^a) \approx \mathcal{N}\{\mu_{B1}(X_1^a), \sigma_B(X_1^a, \langle Z \rangle), B_1\}, \qquad (96)$$

$$\mathcal{P}'(B_1 | X_1^a) \approx \mathcal{N}\{\mu'_{B1}(X_1^a), \sigma'_B(X_1^a, \langle Z' \rangle'), B_1\}. \qquad (97)$$

These p.d.f.s can be evaluated by substituting X_1^a into the expressions (26) and (72) of μ_{Bk} and μ'_{Bk}, and into (45) and (92) for σ_B and σ'_B, respectively. We also need (50) and (95) for $\langle Z \rangle$ and $\langle Z' \rangle'$, and $\sigma_L^2 = (F/2\kappa)^2$ and $\sigma_L'^2 = g_L^2(F/2\kappa')^2$, where κ and κ' are given by (29) and (74), respectively.

To check these results numerically, we performed another six integrations, with the same parameters as before, but for a longer period of 20 000 time units. Evaluating the results at every 0.1 time unit, this gives us a set of 200 001 values of X_1 and B_1. By applying a linear least-squares fit to these 200 001 data points, we obtained the numerical equivalent of the effective damping parameter, denoted by κ^{num}. The values are given in the eleventh column of Table I and demonstrate, quantitatively, that the alternative *a priori* p.d.f.s improve the theoretical values of the effective damping parameter. The only exception concerns run b in which κ' differs more from κ^{num} than does κ, although the difference is not very

large. We also used the 200 001 values of X_1 to calculate, using standard formulae, their means and variances. The resulting μ_L^{num} and σ_L^{num} are displayed in the last two columns of Table I. Note that the value 2.50 for X_1^a is quite representative of the values of μ_L^{num}.

We then selected values of X_1 in an interval of width 1 centred around X_1^a and binned the corresponding values of B_1 in 251 intervals of width 0.1, ranging from -12.55 to $+12.55$. In this way, we obtained numerical p.d.f.s of B_1 conditioned on $X_1 = X_1^a$ which can be compared with the theoretical expressions (96) and (97). The average number of points on which the numerical p.d.f.s are based, is 18 500, about 1/10 of the total number of data points.

The results, in the form of step functions (histograms), are displayed in Figure 4. The theoretical distributions (96) and (97) are shown in terms of the smooth solid and dashed curves, respectively. Generally, the difference between the primed and unprimed p.d.f.s is that the latter are shifted towards smaller values and are slightly less broad. In some cases this leads to an improved resemblance with the numerical results (Figure 4(c, d, e)) but in other cases the resemblance becomes less (Figure 4(a, b, f)). For $X_1^a = 5.0$ (not shown) the shift to smaller average values improves the resemblance in the majority of cases. The graphs all confirm that the theoretical scatter diagrams have a larger spread in values of B_1 than the numerical diagrams. The spread is only slightly smaller in the primed cases so that in that respect there is not much difference between the primed and unprimed cases.

It is also of interest to consider the p.d.f. of an individual variable X_1 of the large-scale system. Theoretically, we have for $\mathcal{P}(X_1)$ and $\mathcal{P}'(X_1)$, as follows from (36) and (83) by integrating over $X_2, ..., X_K$ or from (51) and (91) by integrating over B_1,

$$\mathcal{P}(X_1) = \mathcal{N}(\mu_L, \sigma_L, X_1), \tag{98}$$

$$\mathcal{P}'(X_1) = \mathcal{N}(\mu_L', \sigma_L', X_1), \tag{99}$$

where we recall that $\mu_L = F/(2\kappa)$ and $\mu_L' = f_L F/(2\kappa')$. To check these results in somewhat more detail than is possible on the basis of scatter plots, we calculated numerical p.d.f.s for the variable X_1. These are obtained by binning the values of X_1 at every 0.1 time unit in 101 intervals with width 0.50, ranging from -25.25 to $+25.25$. As now all data are used, the numerical p.d.f.s are based on 200 001 data points. The results for the different cases are displayed in Figure 5, again in the form of solid step functions (histograms). Also shown are the theoretical results based on (98) and (99), the smooth solid curves for the unprimed values of μ_L and σ_L, and smooth dashed curves for the primed values of these parameters. We recall that the underlying difference is the choice of *a priori* p.d.f., i.e. (11) and (34) in the unprimed case and (59) and (79) in the primed case. We see that the differences are not too large; the primed p.d.f.s are shifted to smaller values and are slightly broader than their unprimed counterparts. The shift to smaller values is generally of the right sign.

It is remarkable that four out of the six numerically obtained p.d.f.s are quite close to Gaussian, in accordance with theory. The two exceptions are the p.d.f.s displayed in Figures 5(c) and (f): they are trimodal instead of unimodal and have sharp peaks. They show that, if the effective damping becomes too large, the system enters a less chaotic state. In fact, as inspection of the variables X_k has shown,

the system in Figures 5(c) and (f) is dominated by a westward travelling wave of wave number 7. The theoretical p.d.f.s, being Gaussian, do not capture these features, but they do yield averages and variances of the right order of magnitude.

Figure 6 displays the numerical covariances $\langle (X_1 - \langle X_1 \rangle)(X_k - \langle X_k \rangle) \rangle$ for $k = 1, ..., K + 1$, the index $K + 1$ referring to the same variable as the index 1.

Figures 6(c) and (f) show very clearly that the system in these cases is dominated by a wave of wavenumber 7. As in the two previous figures, the results are based on a numerical simulation of 20 000 time units of the full model equations (1), (2) and (3), using the output every 0.1 time unit. Although the cases (c) and (f) are quite exceptional, the figure generally illustrates that the large-scale variables $X_1, ..., X_K$ are not statistically independent, in contrast to either (36) or (83). The numerical covariance between the variables depends on the effective damping and becomes smaller if this damping becomes smaller, as the other panels of the figure indicate. Although the numerical covariances ($k > 1$) are not reproduced, the numerical variances ($k = 1$) are in reasonable accord with the theory, as can be judged from Table I. Covariances will be predicted if higher-order time derivatives of the energy are used as additional constraints in the maximization of entropy. This becomes evident if one studies e.g. the second-order time derivatives of (4) and (5). The mathematical analysis of this problem is beyond the scope of the present study.

7. Conclusions

By applying the principle of maximum entropy, we obtained a p.d.f. of the small-scale variables $Y_{j,k}$ of a model devised by Lorenz (1996). The information entropy of the p.d.f. is maximized under the constraint that the average time derivative of the energy of the small-scale variables is zero. The p.d.f. of the small-scale variables entails a p.d.f. of B_k which, like the p.d.f. of the small-scale variables $Y_{j,k}$, is a product of independent normal distributions. The mean values of these distributions are proportional to the large-scale variables X_k whereas the variance is proportional to the energy E_L of the large-scale variables.

Without first performing numerical time integrations, the maximum entropy approach leads to an expression of the average interaction of the small-scale variables with the large-scale variables. The corresponding parametrization (26) assumes the form of a linear damping of the large-scale variables, the strength of which depends on the parameters of the interaction. The theory predicts how the damping, and thus the average slope in scatter plots such as displayed in Figure 2, vary with the relevant parameters. These predictions are borne out by numerical integrations of the full system, as the solid curves in Figure 2, representing (26), clearly show.

We also applied the maximum entropy approach to the large-scale system, parametrizing the influence of the small-scale variables in the manner described above. In analogy to the small-scale system, it was assumed that the large-scale variables are in a statistically stationary state which can be described by a maximum entropy p.d.f. The entropy is maximized under the constraint that the averaged time derivative of the large-scale energy is zero. The procedure is analogous to the procedure applied to the small-scale variables and leads to similar results. These results allow
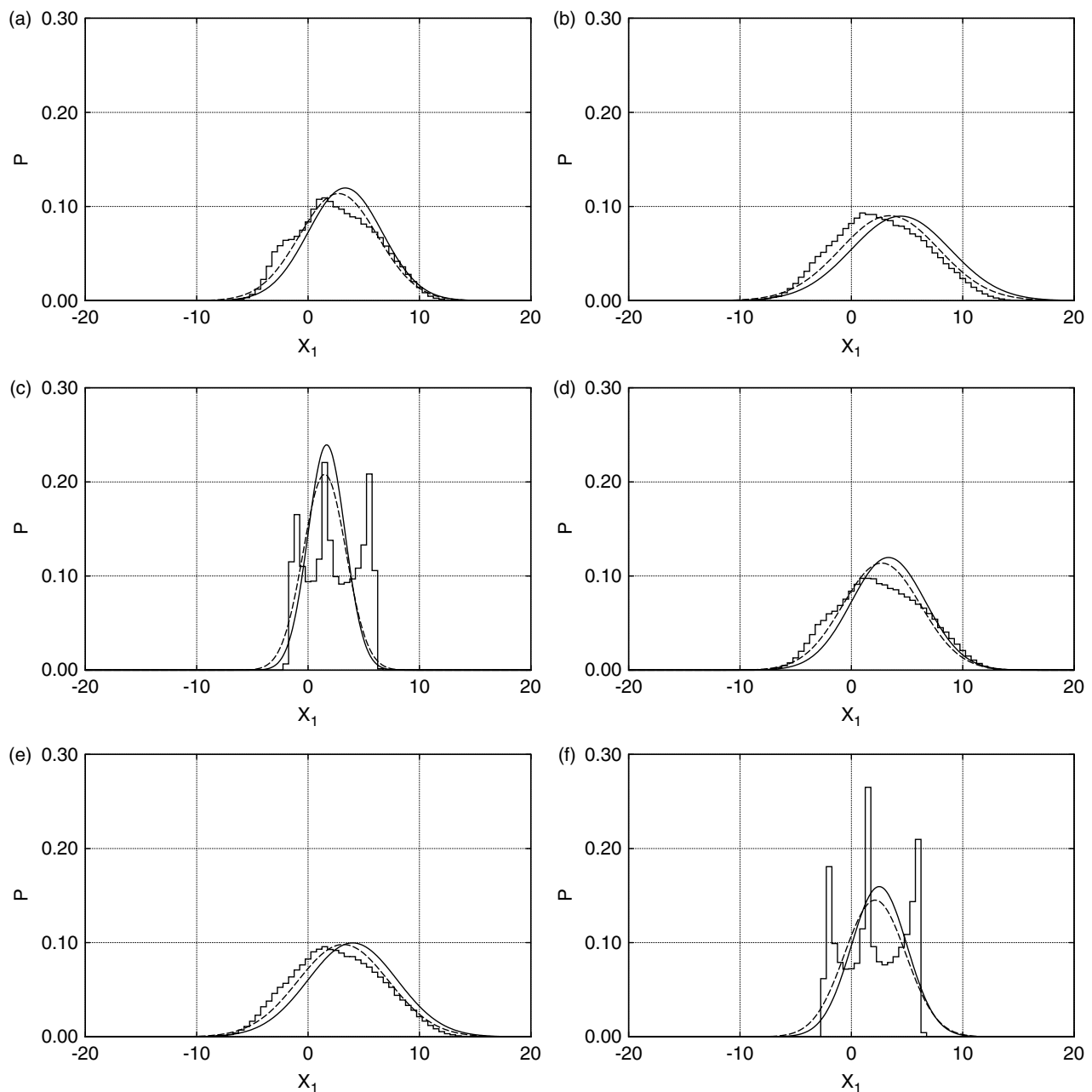
Figure 5. Numerically obtained probability density functions calculated by binning the values of X_1 from a numerical time integration of 20 000 time units. The means and standard deviations of the numerically obtained probability density functions are given in the last two columns of Table I. The results are displayed in the form of step functions (histograms). The smooth solid curves represent the theoretical probability density functions given by (98), with unprimed μ_L and σ_L, i.e. based on the *a priori* probability density functions (11) and (34). The smooth dashed curves represent (99), with μ_L' and σ_L', i.e. based on the *a priori* probability density functions (59) and (79). The numerical probability density functions in (c) and (f) differ markedly from the others because here the large-scale system is dominated by a westward traveling wave with wave number 7.

us to predict more details of the scatter plots given in Figure 2. The theoretical scatter plots of Figure 3 show that these predictions are in reasonable agreement with results obtained numerically.

We have also discussed the influence of alternative *a priori* p.d.f.s. These alternatives are based on the observation that, in a statistically stationary state, the presence of forcing and dissipation implies that the system's energy is bounded from above. By approximating the alternative *a priori* p.d.f.s by a suitable multivariate Gaussian, it was possible to study their influence on the statistics of both the small-scale and the large-scale variables. For the small-scale variables, it leads to a smaller effective friction, as can be seen from the solid and dashed lines in Figure 2 and the solid and dashed curves

in Figure 4. For the large-scale variables, it leads to p.d.f.s that are shifted to somewhat smaller values, as indicated by the solid and dashed curves in Figure 5. It was illustrated by Figure 6 that, depending on the amount of effective damping, the numerical results show covariances between the different variables. These covariances are not reproduced by the present theory and will arise only if zero averages of higher-order time derivatives of the energy are used in the maximization of entropy.

Acknowledgements
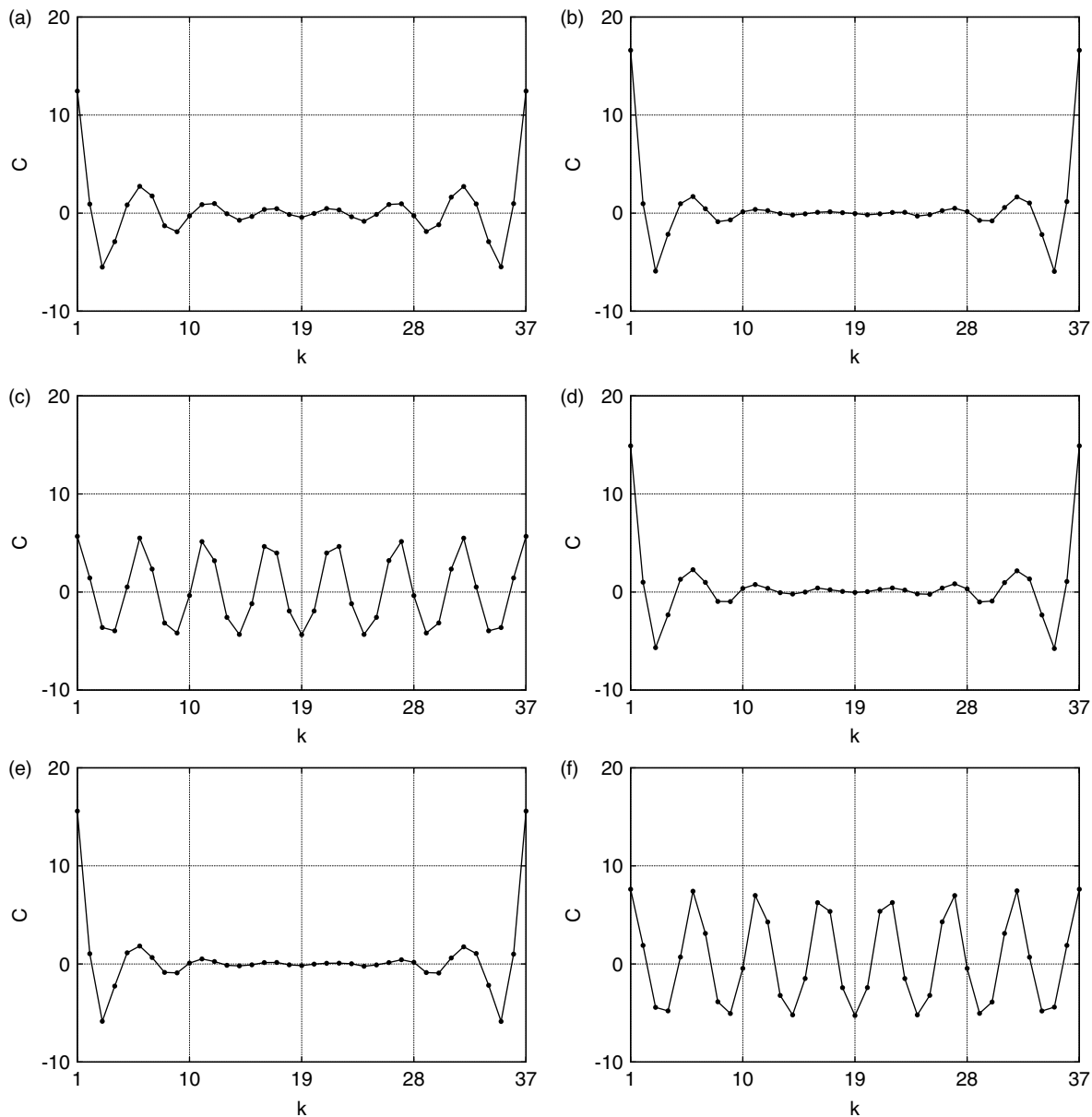
Figure 6. Plots of the covariances $\langle (X_1 - \langle X_1 \rangle)(X_k - \langle X_k \rangle) \rangle$ for $k = 1, \ldots, K + 1$, the index $K + 1$ signifying the same variable as the index 1, based on a numerical simulation of 20 000 time units of the full model equations (1), (2) and (3).

Dr Daan Crommelin in May 2009. He wishes to express his gratitude to Drs Daan Crommelin, Rubén Pasmanter and Frank Selten for supplying him with critical comments and useful suggestions during the preparation of this article. Discussions with Mr Roel Stappers on *a priori* p.d.f.s and with Dr Camiel Severijns on the treatment of higher-order time derivatives are also gratefully acknowledged. The manuscript has benefited substantially from the remarks made by the referees.

Appendix

Mathematical details

In this appendix we will indicate how we obtain values of ρ in the two expressions (58) and (78) for $\mathcal{M}'$ that give the same expectation value of the energy as $\mathcal{M}^*$ in (57) and (77). Concerning the *a priori* p.d.f. $\mathcal{M}^*$ for the small-scale

variables, we first define

$$r^2 = \sum_{J=1}^{J} \sum_{k=1}^{K} Y_{j,k}^2, \quad R^2 = \sum_{J=1}^{J} \sum_{k=1}^{K} C_k^2. \quad \text{(A.1)}$$

The normalization condition can then be written

$$A \int \cdots \int_{r^2 \leq R^2} \mathrm{d}Y_{1,1} \cdots \mathrm{d}Y_{J,K} = 1. \quad \text{(A.2)}$$

Here the integral is over a spherical volume in JK-dimensional space defined by $r^2 \leq R^2$. This volume is given by the following integral over r:

$$\int \cdots \int_{r^2 \leq R^2} \mathrm{d}Y_{1,1} \cdots \mathrm{d}Y_{J,K} = \int_0^R \mathrm{d}r\, \mathcal{A}_{JK}(r), \quad \text{(A.3)}$$

where $\mathcal{A}_n(r)$ is the area of a sphere with radius r in an n-dimensional Euclidean space. For $\mathcal{A}_n(r)$ we have (Zwillinger,

1996, formula 4.18.2), and using $\Gamma(x + 1) = x\Gamma(x)$,

$$\mathcal{A}_n(r) = n \frac{\pi^{n/2}}{\Gamma(n/2 + 1)} r^{n-1}, \qquad (A.4)$$

where Γ is the gamma function. We thus find for the normalization constant A

$$A^{-1} = \frac{(\pi R^2)^{JK/2}}{\Gamma(JK/2 + 1)}. \qquad (A.5)$$

The normalization condition for $\mathcal{M}'$ reads

$$B \int \cdots \int dY_{1,1} \cdots dY_{J,K}$$
$$\times \exp\left[\rho\left(\sum_{j=1}^{J}\sum_{k=1}^{K} C_k^2 - \sum_{j=1}^{J}\sum_{k=1}^{K} Y_{j,k}^2\right)\right] = 1, \qquad (A.6)$$

or, using the expressions given above,

$$B JK \frac{\pi^{JK/2}}{\Gamma(JK/2 + 1)} \exp(\rho R^2) \int_0^\infty dr\, r^{JK-1} \exp(-\rho r^2)$$
$$= 1. \qquad (A.7)$$

According to Zwillinger (1996, definite integral 642), we have

$$\int_0^\infty dr\, r^{n-1} \exp(-\rho r^2) = \frac{1}{2}\rho^{-n/2}\,\Gamma(n/2), \qquad (A.8)$$

from which we find, again making use of $\Gamma(x + 1) = x\Gamma(x)$,

$$B^{-1} = \left(\frac{\pi}{\rho}\right)^{JK/2} \exp(\rho R^2). \qquad (A.9)$$

For the normalized *a priori* p.d.f.s $\mathcal{M}^*$ and $\mathcal{M}'$ for the small-scale variables, we thus have

$$\mathcal{M}^* = \Gamma(JK/2 + 1)\,(\pi R^2)^{-JK/2}\,\mathcal{H}(R^2 - r^2), \qquad (A.10)$$

$$\mathcal{M}' = \left(\frac{\rho}{\pi}\right)^{JK/2} \exp(-\rho r^2). \qquad (A.11)$$

If we express the small-scale energy E_S in terms of r,

$$E_S = \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{1}{2} Y_{j,k}^2 = \frac{1}{2} r^2, \qquad (A.12)$$

and denote the expectation values of E_S with respect to $\mathcal{M}^*$ and $\mathcal{M}'$ by $\langle E_S\rangle^*$ and $\langle E_S\rangle'$, respectively, we find, using analogous integral identities as before,

$$\langle E_S\rangle^* = \frac{1}{2}\frac{JK}{JK + 2} R^2, \qquad (A.13)$$

$$\langle E_S\rangle' = \frac{JK}{4\rho}. \qquad (A.14)$$

If ρ is fixed by requiring that $\langle E_S\rangle^* = \langle E_S\rangle'$, it follows that

$$\frac{1}{2\rho} = \frac{1}{JK + 2} R^2 = \frac{1}{JK + 2}\sum_{j=1}^{J}\sum_{k=1}^{K} C_k^2. \qquad (A.15)$$

We arrive at (59) if we identify $1/(2\rho)$ with σ_{MS}^2. In an analogous way, we may obtain (79) for the large-scale variables.

References

Publications by Jaynes (1957a, 1957b, 1968, 1973) are reprinted in Rosenkrantz 1989.

Crommelin D, Vanden-Eijnden E. 2008. Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.* **65**: 2661–2675.

Jaynes ET. 1957a. Information theory and statistical mechanics. *Phys. Rev.* **106**: 620–630.

Jaynes ET. 1957b. Information theory and statistical mechanics II. *Phys. Rev.* **108**: 171–190.

Jaynes ET. 1968. Prior probabilities. *IEEE Trans. Sys. Sci. Cyber.* **4**: 227–241.

Jaynes ET. 1973. The well-posed problem. *Found. Phys.* **3**: 477–493.

Jaynes ET. 2003. *Probability Theory – The Logic of Science*. Cambridge University Press: Cambridge, UK.

Johnson NL, Kotz S, Balakrishnan N. 1995. *Continuous Univariate Distributions* – Vol. 2. (2nd ed.) John Wiley & Sons: New York.

Lorenz EN. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**: 130–141.

Lorenz EN. 1996. 'Predictability – A problem partly solved'. In *Proceedings of Seminar on Predictability, Vol. 1*. Reading, UK, 4–8 Sept 1995. ECMWF: Reading, UK. 1–18.

Press WH, Teukolsky SA, Vetterling WT, Flannery BP. 1996. *Numerical Recipes in Fortran 77*. (2nd ed.) Cambridge University Press: Cambridge, UK.

Rosenkrantz RD. (ed.) 1989. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*. Kluwer Academic Publishers: Dordrecht, Netherlands.

Verkley WTM, Lynch P. 2009. Energy and enstrophy spectra of geostrophic turbulent flows derived from a maximum entropy principle. *J. Atmos. Sci.* **66**: 2216–2236.

Wilks DS. 2005. Effects of stochastic parametrizations in the Lorenz '96 system. *Q. J. R. Meteorol. Soc.* **131**: 389–407.

Zwillinger D. 1996. *Standard Mathematical Tables and Formulae*. (30th ed.) CRC Press: Boca Raton, FL.

Zwillinger D, Kokoska S. 2000. *Standard Probability and Statistics Tables and Formulae*. CRC Press: Boca Raton, FL.