Model error and predictability over different time scales in the Lorenz '96 systems

D. Orrell

Centre for Nonlinear Dynamics, University College London, UK

Present address: Institute for Systems Biology,

1441 N 34th Street, Seattle, USA

Abstract

Prediction problems have been described by Lorenz as falling into two categories. Problems which depend on the initial condition, such as short to medium range weather forecasting, are described as 'predictions of the first kind', while problems that depend on boundary rather than initial conditions, such as in many cases the longer term climatology, are referred to as predictions of the second kind. Both kinds of prediction will be affected by error in the model equations used to approximate the true system. In this paper, we examine predictability over different time scales for the medium-dimensional Lorenz '96 systems. Models are constructed for the purposes of optimising both short-range prediction and climatological behavior, and studied over a range of forcings for which they show periodic, quasi-periodic, or chaotic behavior. It is shown that, for the models discussed here, there is a link between short and long-range predictability, which holds independently of the effects of chaos. The role of stochastic terms is considered, and the possible implications for atmospheric or oceanographic modelling are discussed.

1 Introduction

Forecasters of the Earth's atmosphere or oceans are confronted by two types of questions: what will happen over the next few hours, days, or months, as a result of the current configuration of the system; and what will happen to the climate in the longer term as a result of changed conditions such as increased carbon dioxide emissions. The former depends strongly on the initial condition, while the latter is primarily a boundary-value problem. Lorenz (1975) classified these as predictions of the first and second kind. For either problem the tool of choice is usually a physics-based model (Hunt, 1999), which will be subject to errors both in the initial condition and the model (Collins, 2002).

Given the fact that the model is imperfect, the question then arises whether the goals of long and short-term predictability, or predictions of the first and second kind, coincide. Is a model which is constructed to optimise prediction over the next few hours necessarily the best also for prediction of climatology, or are these completely different, even irreconciable, aims? Are there cases where long-term accuracy is best served by loss of short-term prediction, so the ends, it could be said, justify the means?

At first glance, it might seem that it is easier to model the climatology than the short-term weather. For example, there are many models which succeed in

3

reproducing the variability or some other aspect of financial markets; rather fewer which can predict the next recession. However, there are also models which can predict the short-term, but fail utterly to reproduce the climatology; for instance, a weather model with an overly large damping term, which over many years would eventually bring the weather to a complete halt. At the same time, it intuitively seems reasonable that a model which can predict well in the near-term is also a good choice for a climate model, so long as the model is not so large and complex as to prohibit long term calculations.

This paper addresses the above questions of model error and predictability for variants of the Lorenz '96 system (Lorenz, 1996; Hansen and Smith, 2000). Using a 40 dimension system as 'truth', we will attempt to model its large-scale behaviour with two 8 dimensional models. Section 2 introduces the system and models. Section 3 examines the effect of model error on predictability over the short to medium-term. By exploiting certain properties of the models in Section 4, we compare the short-term predictability with the long-term climatology. The effect of adding stochastic terms is also considered. Finally, in Section 5, we discuss links between the different types of predictability, and possible implications for climate modelling.

4

2   The system and models

The Lorenz '96 systems consist of two types, which we refer to as System I and System II. The System I equations are:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F, \tag{1}$$

for $i = 1$ to n. The index i is cyclic so that $x_{i-n} = x_{i+n} = x_i$. The equations simulate advection, damping and forcing of some atmospheric variable x_i that can be viewed as existing around a circle, and therefore comprise a kind of minimalist weather model. (The vagueness of this comparison is deliberate; these are toy models, and their purpose is only to help suggest approaches to the corresponding atmospheric problems.) The equations of System II are

$$\begin{aligned}
\frac{d\tilde{x}_i}{dt} &= \tilde{x}_{i-1}(\tilde{x}_{i+1} - \tilde{x}_{i-2}) - \tilde{x}_i + F - \frac{hc}{b}\sum_{j=1}^{m}\tilde{y}_{i,j} \\
\frac{d\tilde{y}_{i,j}}{dt} &= cb\tilde{y}_{i,j+1}(\tilde{y}_{i,j-1} - \tilde{y}_{i,j+2}) - c\tilde{y}_{i,j} + \frac{hc}{b}\tilde{x}_i
\end{aligned} \tag{2}$$

for $i = 1$ to n, and $j = 1$ to m. Again the variables are cyclic so that $\tilde{y}_{i+n,j} = \tilde{y}_{i,j}$ and $\tilde{y}_{i,j-m} = \tilde{y}_{i-1,j}$. The variables can again be viewed as existing around a circle, with the $\tilde{y}$'s representing fine-scale processes.

5

We here set $n = 8$, $m = 4$, for which System I has dimension 8 and System II has dimension 40. This choice was not made to reflect any property of the real atmosphere, but primarily for computational convenience and the fact that this version has particularly interesting bifurcation properties. The coefficients used are $h = 1$ and $b = c = 10$, for which the $\tilde{y}$'s tend to fluctuate ten times more rapidly but with ten times smaller magnitude than the $\tilde{x}$'s. The effect of changing the coupling coefficient h, as well as the effect of the dimension, is discussed in (Orrell and Smith, 2003).

It is possible to scale versions of these systems, including time t, so that they approximate some aspects of weather models, such as doubling time (Orrell, 2002). One could also argue that $n = 40$ is a better match of atmospheric properties such as propagation of disturbances (Lorenz and Emmanuel, 1998). However, these refinements only work if forcing is kept fixed, since model properties such as the model's internal doubling time will vary strongly with forcing. The approach here is instead to test a single version over a wide range of forcings (though we look more closely at a particular case in Section 3). Note that the analytical results in this paper hold for any choice of the dimensions n, m, or the coefficients h, b, c.

Figure 1 shows a typical orbit in the $\tilde{x}_1$ and $\tilde{y}_1$ variables for this version of

6

System II, evaluated using a Runge-Kutta scheme with a timestep of 0.005, for $F = 10$. It can be seen that a large local value of $\tilde{x}$ tends to excite the $\tilde{y}$ variables (and vice versa), due to the feedback between the two.

An alternative way of viewing the equations of the large-scale $\tilde{x}$ variables is as a variant of System I, so

$$\frac{d\tilde{x}_i}{dt} = \tilde{x}_{i-1}(\tilde{x}_{i+1} - \tilde{x}_{i-2}) - \tilde{x}_i + \tilde{F}_i(t) \tag{3}$$

for i=1 to n, where

$$\tilde{F}_i(t) = F - \frac{hc}{b} \sum_{j=1}^{m} \tilde{y}_{i,j}(t) \tag{4}$$

is treated as a forcing which varies in a complicated manner with time. If we then view the $\tilde{x}$ trajectory as the target orbit, we can attempt to approximate it using models of the form

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i(t) \tag{5}$$

where the n-dimensional vector $\mathbf{P}(t)$ with components $P_i(t)$ is some parameterization of $\tilde{\mathbf{F}}(t)$. The situation is therefore analogous to that encountered with real weather models, which incorporate parameterizations of complicated physi-

7

cal processes such as the formation and dissipation of clouds.

We will use two parameterization schemes. The first, and simplest, is to set $P_i(t)$ equal to a constant, as in System I. A sensible choice of constant would appear to be the mean forcing (we justify its use below). We therefore define the constant parameterization as

$$P^c = \langle \tilde{F_i(t)} \rangle \tag{6}$$

where the mean is calculated over a long orbit on the System II attractor (by symmetry it is independent of the index i used). In general P^c is smaller than F by a small amount (Orrell, 2001), so for example if $F = 10$ the corresponding value of P^c is found to be 9.62. The *constant model* is then:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P^c. \tag{7}$$

The second parameterization incorporates a linear fit to the forcing term. It was noted above that the magnitude of the $\tilde{y}_{i,j}$ variables in Figure 1 tends to be greater in regions where the local $\tilde{x}_i$ variable is also larger. It is therefore reasonable to look for a connection between the local $\tilde{x}_i$ component and the forcing $\tilde{F}_i$ experienced by that component, which depends on the local $\tilde{y}_{i,j}$'s. This is shown more clearly by Figure 2, which is a plot of pairs $(\tilde{x}_i, \tilde{F}_i)$ where $\tilde{x}_i$ is the value of a

8

particular $\tilde{x}$ component and $\tilde{F}_i$ from Eqn. 4 is the local forcing for $F = 10$. There

is a definite linear tendency to the data, which can be fit using linear regression,

leading to a formula

$$P_i^l(x_i) = \alpha_0 + \alpha_1 x_i, \tag{8}$$

where

$$\alpha_0 = \frac{\langle \tilde{x}_i^2 \rangle \langle \tilde{F}_i \rangle - \langle \tilde{x}_i \rangle \langle \tilde{x}_i \tilde{F}_i \rangle}{\langle \tilde{x}_i^2 \rangle - \langle \tilde{x}_i \rangle^2} \tag{9}$$

and

$$\alpha_1 = \frac{\langle \tilde{x}_i \tilde{F}_i \rangle - \langle \tilde{x}_i \rangle \langle \tilde{F}_i \rangle}{\langle \tilde{x}_i^2 \rangle - \langle \tilde{x}_i \rangle^2}. \tag{10}$$

We define the *linear model* to be

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P_i^l(x_i). \tag{11}$$

The constant α_0 and slope α_1 must be calculated for each value of F. At $F = 10$,

$\alpha_0 = 9.95$ and $\alpha_1 = -0.122$.

3 Optimising short-term prediction

The choice of constant and linear parameterizations P^c and P^l was justified by the fact that they minimise the expected error in the forcing compared to the System II forcing. It turns out, not surprisingly, that this is exactly what is required to optimise short-term predictability. Suppose the System II target orbit which we wish to approximate is $\tilde{\mathbf{x}}(t)$. We write the model equations in the form

$$\frac{d\mathbf{x}}{dt} = \mathbf{G}(\mathbf{x}(t)), \tag{12}$$

and the target system equations in the form

$$\frac{d\tilde{\mathbf{x}}}{dt} = \tilde{\mathbf{G}}(\tilde{\mathbf{x}}(t)). \tag{13}$$

Define the error vector to be

$$\mathbf{e}(t) = \mathbf{x}(t) - \tilde{\mathbf{x}}(t). \tag{14}$$

Then a first order Taylor series expansion gives

$$\mathbf{e}(t) \approx \mathbf{e}(0) + \frac{d\mathbf{e}(0)}{dt}t. \tag{15}$$

10

If the initial condition error is zero, so $\mathbf{x}(0) = \tilde{\mathbf{x}}(0)$, then the error at small times t is approximately given by

$$\mathbf{e}(t) \approx \frac{d\mathbf{e}(0)}{dt} t. \tag{16}$$

It therefore follows that, in the limit as prediction time goes to zero, the short-term root-mean-square prediction error $e(t)$ can be optimised by minimising the root-mean-square value of the initial *tendency error* (Klinker and Sardeshmukh, 1992; Schubert and Schang, 1996). The tendency error is given by

$$\begin{aligned} \frac{d\mathbf{e}(0)}{dt} &= \frac{d\mathbf{x}(0)}{dt} - \frac{d\tilde{\mathbf{x}}(0)}{dt} \\ &= \mathbf{G}_e(\tilde{\mathbf{x}}(0)) \end{aligned} \tag{17}$$

where

$$\mathbf{G}_e(\tilde{\mathbf{x}}(t)) = \mathbf{G}(\tilde{\mathbf{x}}(t)) - \tilde{\mathbf{G}}(\tilde{\mathbf{x}}(t)). \tag{18}$$

For the Lorenz systems studied here, this term is just the forcing error

$$\mathbf{G}_e(\tilde{\mathbf{x}}(0)) = \mathbf{P}(0) - \tilde{\mathbf{F}}(0). \tag{19}$$

Suppose now that there is an initial error of root-mean-square magnitude S,

11

that is uncorrelated with the tendency error. Then the expected error is given by

$$e(t) \approx \sqrt{S^2 + (G_e t)^2} \tag{20}$$

where G_e is the root-mean-square forcing error. Thus the relative importance of the model error depends on the magnitude of initial condition error; and to reduce short-term errors, one must minimize both initial error and tendency error. Given a fixed S, however, the optimal model is that with the smallest tendency error G_e. Since the choice of the coefficients in the constant and linear models minimise this term, it follows that these choices are optimal for short-term prediction within their model class (i.e. constant or linear).

Better models can of course be built, but these models are interesting to study in that they contain a significant degree of model error, and also correspond to the kind of simple first-order parameterizations that one might adopt in real physical models. Compared to the constant model, the linear model reduces the initial tendency error by nearly a half. One can therefore loosely view the constant model as a 'high-error' case, and the linear model as a 'low-error' case, though these terms are relative to the size of the initial condition error.

For either model, error arises due to inadequacies in the parameterization of

12

the $\tilde{y}$ variables. As discussed in (Lorenz, 1996), these variables can be interpreted as representing fine-scale atmospheric processes, and are important in the early stages of initial error growth within System II when small perturbations are made to the initial condition. Therefore it might come as no surprise that parameterizing them introduces model error. However, it is important to distinguish between sensitivity to initial condition in System II, and model error in the constant or linear model. The two are separate phenomena, and need not be directly related; for example, even if System II was relatively insensitive to changes in the $\tilde{y}$ variables, a poor parameterization could still introduce significant model error.

As seen in the next section, the Lorenz '96 systems display a variety of behaviour at different values of the forcing, including periodic, quasi-periodic, and chaotic states. Interestingly, though, the expected tendency error for either the constant or linear models varies quite smoothly, and goes approximately with the square-root of forcing, as shown in Figure 3. It is shown later that climatological quantities such as mean and variance also vary in a fairly smooth manner.

While the optimality of the chosen models only applies for times near zero, it is interesting to ask whether it extends also into longer times. In the remainder of this section, we examine error growth over the short to medium-term, and attempt to generalise the concept of tendency error for these time periods.

13

In (Orrell, 2001; Orrell et al., 2001) it was proposed that the forecast error over the short range is approximately equal to the *drift*, which is defined as the integral of the tendency error over the target orbit:

$$d(\tau) = \| \int_0^\tau \mathbf{G}_e(\tilde{\mathbf{x}}(t))dt \|. \tag{21}$$

The drift is derived in a manner similar to Eq. 15, by considering a Taylor series expansion of the error, and the accuracy of the approximation depends on the model and the system. The approximation is improved, i.e. holds for longer times, if instead the *propagated drift*

$$d_p(\tau) = \| \int_0^\tau \mathbf{M}(\tau, t)\mathbf{G}_e(\tilde{\mathbf{x}}(t))dt \| \tag{22}$$

is used (Orrell, 2002). Here $\mathbf{M}(\tau, t) = e^{\int_t^\tau \mathbf{J}(t)dt}$ is the linear propagator (Strang, 1986) evaluated along the target orbit $\tilde{\mathbf{x}}$ from time t to time τ, and $\mathbf{J}(t)$ is the Jacobian of $\mathbf{G}(\tilde{\mathbf{x}}(t))$. The propagated drift accounts for the amplification of model error by the nonlinear dynamics. Again, the reliability of the approximation depends on the particular model and system. Because in chaotic models the effect of the nonlinear dynamics is usually to amplify the model errors, the drift will generally be smaller than the propagated drift.

14

Figure 4 compares the drift and propagated drift with the total forecast error for both the constant and linear models, where the true system again has $F = 10$. In the upper panels, the true system is perfectly observed, so there is no initial error. The drift, which we write $d^m(t)$, is therefore due to the model equations alone. The drift underestimates the total error, while the propagated drift is almost indistinguishable from the total error up to $t = 1$. For comparison, the model's internal doubling time at this value of forcing F is around $\tau_d = 0.23$ time units.

In the middle panels, Gaussian noise with standard deviation 0.1 is added to the true system, to simulate the observation error which is encountered when working with real physical systems. Because of their different natures, it is hard to directly compare the magnitudes of observational error and model error, but one useful measure is to compare the drift $d^m(\tau_d)$ due to model error at the doubling time τ_d with the initial condition error. For the constant model, the drift $d^m(\tau_d)$ at $\tau_d = 0.23$ is about 0.33, while for the linear model it is about 0.15. Since the initial error of 0.1 will roughly double in this time, model error will slightly dominate for the constant model at this time, while initial error, magnified by the model dynamics, will slightly dominate for the linear model. Note that the effect of this amount of observation error on the drift calculation at times greater than τ_d is relatively small.

15

The lower panels show a zoomed view of the error growth for times up to $t = 0.4$. As a rough guide to comparison with weather models, based on doubling times, this could correspond to error growth over the first few days. Of course the assimilation procedure used with real weather models will affect the shape of the growth curve; the plots here would correspond to errors relative to raw or minimally treated observations.

It can be seen that the effect on the drift of the observational error is, to first order, to increase it to

$$d(t) = \sqrt{d^m(t)^2 + 2S^2},\qquad(23)$$

which compares with Eq. 20 for the tendency error. [1] The resulting estimate is indicated in the figure by the circle symbols. It therefore follows that, to minimise the total drift $d(t)$, one must minimise the drift due to model error $d^m(t)$. In cases where the dynamical model errors are similar to white noise, $d^m(t)$ increases in a square-root fashion at a rate determined by the initial tendency error (Orrell, 2002). An example of a model with nearly white-noise error is the linear model, as shown in Figure 5 which compares the power spectrum of the forcing error with that of the constant model. For such models, the drift is essentially determined by

---

[1]This topic, and the subject of medium-term predictability in general, will be explored in greater detail in a future paper.

the initial tendency error; and since the effect of the propagator is generally to amplify the model errors, the propagated drift is similarly affected. Therefore the influence of the initial tendency error on predictability extends well beyond the short-term. In the next section, we consider the relationship between errors at very short times, and those in the limit as time goes to infinity.

4 Climatologies

While optimising short-term prediction error is a clearly defined problem, the question of climatology is somewhat more nebulous, and depends on what properties of the target climatology we wish the model to preserve. One possibility would be to attempt to match the power spectrum of System II. Unfortunately, this is an impossible task for the models considered here, simply because their bifurcation diagrams are different: the models may be periodic or quasi-periodic when System II is chaotic. Figure 6 is a spectral bifurcation diagram for System II and the constant and linear models (see also (Orrell and Smith, 2003) for a discussion of these figures). The horizontal axis is the System II forcing, vertical axis is frequency. The shading indicates the logarithm of the power at each frequency. In a periodic orbit, all the power is at a base frequency, corresponding to

17

the period of the orbit, or higher multiples, so these regions appear as a sequence of equispaced lines (for example, around $F = 3$). Chaotic regions are characterised by power at all frequencies, such as around $F = 7$. Quasi-periodic orbits, meanwhile, are indicated by lines which are at incommensurate frequencies, such as around $F = 5.3$ for the constant model. (For clarity, only power above a certain cut-off is shown, so what appears to be a periodic region may sometimes also include a small amount of power at other frequencies.) Note the high degree of structure in the quasi-periodic regions - the main advantage of spectral bifurcation diagrams over conventional bifurcation diagrams is that they reveal such features. Because the models have periodic or quasi-periodic orbits in places where System II is chaotic, it is not possible to perfectly reconcile the power spectra between model and target.

An alternative and somewhat simpler measure of climatology is given by the mean and the variance of the large-scale variables. These quantities also have atmospheric relevance: a typical climate question might be what effect additional carbon dioxide would have on the average temperature and its variability. For the Lorenz systems, it is evident from the spectral bifurcation diagrams that the model and system undergo complex changes in behaviour as forcing is increased. Nevertheless, we will see that the mean and variance behave in a fairly smooth

18

and predictable manner. We begin by establishing some analytic results for the model and system, which prove useful in determining when and how the models can match the system climatology.

Looking first at the constant model, the equation for x_i is:

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P^c. \tag{24}$$

Multiplying each side of this equation by x_i gives

$$x_i\frac{dx_i}{dt} = \frac{1}{2}\frac{d(x_i^2)}{dt} = x_1 x_{i-1}(x_{i+1} - x_{i-2}) - x_i^2 + P^c x_i. \tag{25}$$

Summing over all i, the advection terms cancel out, leaving

$$\frac{1}{2}\sum_{i=1}^{n}\frac{d(x_i^2)}{dt} = -\sum_{i=1}^{n}x_i^2 + P^c\sum_{i=1}^{n}x_i. \tag{26}$$

This summation can be viewed as a global integration over all the variables. Now, let L be a real number. Then

$$\frac{1}{L}\int_0^L \frac{1}{2}\sum_{i=1}^{n}\frac{d(x_i^2)}{dt}dt = -\frac{1}{L}\int_0^L \sum_{i=1}^{n}x_i^2 dt + \frac{1}{L}P^c\int_0^L\sum_{i=1}^{n}x_i dt. \tag{27}$$

19

Taking the limit as L goes to infinity, the left hand side is just

$$\lim_{L \to \infty} \frac{1}{L} \frac{1}{2} \sum_{i=1}^{n} x_i^2. \tag{28}$$

It is easily seen, for example by the Trapping Region Lemma (Alligood et al., 1996), that x_i is bounded, and so the above term goes to zero in the limit. The first term on the right hand side, meanwhile, converges to $n\langle x_i^2 \rangle$, where the average is over the attractor, and the second term is $n\langle x_i \rangle$. Therefore we obtain the result that the mean of x_i^2 is equal to the forcing times the mean of x_i:

$$\langle x_i^2 \rangle = P^c \langle x_i \rangle. \tag{29}$$

A similiar technique can be applied to System II. The equation for the large scale variables is

$$\frac{d\tilde{x}_i}{dt} = \tilde{x}_{i-1}(\tilde{x}_{i+1} - \tilde{x}_{i-2}) - \tilde{x}_i + F - \frac{hc}{b} \sum_{j=1}^{m} \tilde{y}_{i,j}. \tag{30}$$

The same procedure as that followed above gives a similiar result, but now there is

20

an additional term which depends on the correlation between $\tilde{x}_i$ and $\tilde{y}_{i,j}$ variables:

$$\langle \tilde{x}_i^2 \rangle = F \langle \tilde{x}_i \rangle - \frac{mhc}{b} \langle \tilde{x}_i \tilde{y}_{i,j} \rangle. \tag{31}$$

The system equation for the $\tilde{y}_{i,j}$ variables is

$$\frac{d\tilde{y}_{i,j}}{dt} = cb\tilde{y}_{i,j+1}(\tilde{y}_{i,j-1} - \tilde{y}_{i,j+2}) - c\tilde{y}_{i,j} + \frac{hc}{b}\tilde{x}_i. \tag{32}$$

Multiplying now by $\tilde{y}_{i,j}$ and proceeding as above gives

$$\langle \tilde{y}_{i,j}^2 \rangle = \frac{mh}{b} \langle \tilde{x}_i \tilde{y}_{i,j} \rangle. \tag{33}$$

Combining these equations yields

$$\langle \tilde{y}_{i,j}^2 \rangle = \frac{1}{c}(F \langle \tilde{x}_i \rangle - \langle \tilde{x}_i^2 \rangle). \tag{34}$$

This result means that information about the fine-scale $\tilde{y}_{i,j}$ variables can be deduced by observing only the large-scale $\tilde{x}$ variables.

Suppose now that we wish to model the macroscopic behaviour of System II

21

using the constant model with forcing P^c. For the model, we have

$$\frac{\langle x_i^2 \rangle}{\langle x_i \rangle} = P^c \tag{35}$$

while for the system

$$\frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle} = F - \frac{c\langle \tilde{y}_{i,j}^2 \rangle}{\langle \tilde{x}_i \rangle}. \tag{36}$$

If the aim is a perfect match to climatology, then the ratio of the first and second moments should agree, so

$$\frac{\langle x_i^2 \rangle}{\langle x_i \rangle} = \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle}. \tag{37}$$

In this case, the optimal choice for the constant model forcing would be

$$P^c = \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle}. \tag{38}$$

The value of P^c arrived at is not the same as the value used above to minimise tendency error. At $F = 10$, for example, the optimal forcing is 8.87 as opposed to 9.63. Also, the chosen value of P^c gives the correct ratio of $\langle x_i^2 \rangle$ to $\langle x_i \rangle$, but never the correct value of either term. Therefore it is impossible to exactly match both climatological quantities.

Another approach is to keep P^c at the value used for minimising tendency

error, and test how this fits System II in terms of climatology. Figure 7 shows $\langle \tilde{x}_i \rangle$ and $\langle \tilde{x}_i^2 \rangle$ respectively, along with the corresponding values of $\langle x_i \rangle$ and $\langle x_i^2 \rangle$ for the constant model with forcing P^c. Except for certain regions such as around $F = 5$, the model mean is always slightly too low, and the variance too high. It therefore follows for example that if the model forcing is adjusted to match System II, then the variance will be too high. However the match in general is quite good, and it appears that the constant model offers a reasonable compromise between the two quantities.

Note that the quantities vary in a regular manner with forcing. The mean $\langle \tilde{x}_i \rangle$ goes approximately with the square root, and $\langle \tilde{x}_i^2 \rangle$ with the square of forcing. Since the variance is just the sum of the power spectra, a conclusion is that the total power in the system, summed over all frequencies, varies in a smooth manner with the square root of forcing. This simple result would not be evident from the spectral bifurcation diagram, Figure 6. The spectrum varies between periodic, quasi-periodic and chaotic behaviour in a discontinuous fashion, but the total power is a relatively smooth function of forcing.

One might expect to do an even better job of modelling System II with the linear model, which has two parameters to adjust. The relationship between $\langle x_i^2 \rangle$ and $\langle x_i \rangle$ can be computed for the linear model just as for the constant model.

23

Recall that the linear model has a forcing term $\mathbf{P^l}$ with components given by

$$P_i^l(x_i) = \alpha_0 + \alpha_1 x_i. \tag{39}$$

Following the same procedure above gives

$$\frac{\langle x_i^2 \rangle}{\langle x_i \rangle} = \frac{\alpha_0}{1 - \alpha_1} \tag{40}$$

with the additional linear term in the parameterization introducing a factor $1 - \alpha_1$

in the denominator. Thus, to preserve the ratio $\frac{\langle x_i^2 \rangle}{\langle x_i \rangle}$ of the true system, we require

$$\frac{\alpha_0}{1 - \alpha_1} = \frac{\langle \tilde{x}_i^2 \rangle}{\langle \tilde{x}_i \rangle} \tag{41}$$

which solved for α_1 gives

$$\alpha_1 = 1 - \alpha_0 \frac{\langle \tilde{x}_i \rangle}{\langle \tilde{x}_i^2 \rangle}. \tag{42}$$

Therefore, given a value of α_0, the corresponding value of α_1 can be found.

Using this relationship, the parameters can be optimised to match the climato-

logical quantities, and compared with the short-term linear model. Figure 8 shows

how the ratio of $\langle x_i \rangle$ to $\langle \tilde{x}_i \rangle$ changes with the forcing offset $\alpha_0 - F$. A graph of

the ratio of $\langle x_i^2 \rangle$ to $\langle \tilde{x}_i^2 \rangle$ is indistinguishable. The ratio is approximately 1.0 when

24

the offset is approximately zero, or $\alpha_0 = F$. The corresponding value of α_1 is then

$$\alpha_1 = 1 - F\frac{\langle \tilde{x}_i \rangle}{\langle \tilde{x}_i^2 \rangle}. \tag{43}$$

For $F = 10$, the resulting slope is $\alpha_1 = -0.127$. To the margin of error, these coefficients are indistinguishable from the values $\alpha_0 = 9.95$ and $\alpha_1 = -0.122$ used in the linear model. In other words, the same model is arrived at whether the aim is to optimise long-term or short-term performance (so long as initial condition error has an assumed magnitude and is uncorrelated with tendency error).

The reason for this becomes more clear if we consider the choice of α_0 and α_1 for short-term predictability. From Eqs. 9 and 10, it is easily checked that these values have the property

$$\alpha_0 \langle \tilde{x}_i \rangle + \alpha_1 \langle \tilde{x}_i^2 \rangle = \langle \tilde{x}_i \tilde{F}_i \rangle. \tag{44}$$

But from Eqs. 4 and 31,

$$\langle \tilde{x}_i \tilde{F}_i \rangle = F\langle \tilde{x}_i \rangle - \frac{mhc}{b}\langle \tilde{x}_i \tilde{y}_{i,j} \rangle \tag{45}$$

$$= \langle \tilde{x}_i^2 \rangle. \tag{46}$$

25

Therefore

$$\alpha_0 \langle \tilde{x}_i \rangle + \alpha_1 \langle \tilde{x}_i^2 \rangle = \langle \tilde{x}_i^2 \rangle, \tag{47}$$

so α_0 and α_1 satisfy Eq. 41. It follows that if α_0 and α_1 are chosen to minimize the short-term error, they also preserve the climatological ratio of variance to mean.

When checked over a range of forcings, the standard linear model does a fine job of reproducing System II's macroscopic behaviour. The upper panel of Figure 9 shows $\langle x_i \rangle$ for the linear model compared to $\langle \tilde{x}_i \rangle$ for the system as a function of system forcings, while the lower panel shows $\langle x_i^2 \rangle$ compared to $\langle \tilde{x}_i^2 \rangle$. Agreement is excellent except in the regions near $F = 1.3$ (where the fine-scale variables become non-zero in the true system) and $F = 6.5$. Apart from these areas, correspondence is almost exact. The variance/mean ratio (not shown) for either the model or system are identical, as expected from the choice of α_0 and α_1.

Since the model forcing differs from that of the system by an amount that appears to vary in a somewhat random manner, one might ask whether the models could be improved by adding a stochastic term to the model equations, with a variance that is compatible with the forcing error. The constant model, for example, could become

$$\frac{dx_i}{dt} = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + P^c + N(\epsilon) \tag{48}$$

26

where $N(\epsilon)$ is a random variable with variance ϵ, which is updated each time step (here 0.005 units). Following the procedure used above, Eqn. 29 then becomes

$$\langle x_i^2 \rangle = P^c \langle x_i \rangle + \langle N(\epsilon) x_i \rangle. \tag{49}$$

Since $N(\epsilon)$ is a random variable, the last term should be zero. Therefore the relationship between model mean and variance is unchanged.

The addition of a random term has little noticeable effect on either the variance, or, because the two are linked, the mean. This is the case even if the random term is significant, and only breaks down when the term is so large that it dominates the forcing, at which point the integration scheme becomes affected. Therefore, although it might seem counter-intuitive, adding a random term to the model equations generally has negligible effect on the climatology.

The effect on short term predictability, however, is both large, and in the wrong direction. The square of the root-mean-square tendency error is given by

$$\langle (P^c + N(\epsilon) - \tilde{F}_i(t))^2 \rangle = \langle (P^c - \tilde{F}_i(t))^2 \rangle + \epsilon \tag{50}$$

so the effect of the stochastic term is to increase the expected initial tendency error. The drift over a time τ will similarly increase, with $N(\epsilon)$ adding a random

27

walk component to the integral.

It therefore appears that adding a stochastic term to the model has little or no effect on long term climatology, but reduces short term accuracy. A corollary is that, if the model differs from the true system by only a stochastic forcing term, then the two climatologies should match. Since the errors for the linear model are closer to being stochastic than the constant model, as shown in Figure 5, it follows that the climatological fit should also be better.

5 Conclusions

This paper has investigated links between predictability over different time scales for the Lorenz '96 systems. Clearly, model error has an effect not just on short-term prediction, but also long-term climatology. For the models considered here, there appears to be a connection between the two. The constant model, designed to optimise short-term accuracy, also gives a good match to climatological mean and variance. A linear model does an even better job of fitting these quantities; furthermore, the same model is arrived at whether the aim is long-term or short-term prediction.

Simple relationships between the climatological variables were established for

the Lorenz systems, by summing the equations for each variable, and integrating the result over a long orbit. These were used to prove that the constant model cannot exactly match both the System II mean and variance simultaneously, and to determine the correct parameters for fitting the linear model to climatological data. Of course, these results do not apply to general systems, and such simple relationships will not exist for actual weather or ocean models. Nevertheless, the structural similarity between the Lorenz systems and atmospheric models may imply that similar relationships hold in an approximate sense. For example, it was found that, when the system equations were summed over the index, the advection terms canceled. This operation is the equivalent of integrating over space in a weather model, and it should be the case that the integrated advection terms will be small relative to other terms. It may therefore be possible to generalise the results here to broader classes of models.

The performance of the linear model relative to the constant model shows that even a simple improvement in parameterization can lead to significantly enhanced accuracy in both the short and long range. An implication for atmospheric or ocean models is that improvements to the parameterization of complex processes, such as cloud formation and dissipation, may have significant benefits, not just to the forecast of tomorrow's weather, but also to climate predictions. The addition

of stochastic terms to the model, however, might have a small effect on long term predictions but significantly reduce short range accuracy.

The systems have been studied over a range of forcings, and a mix of periodic, quasi-periodic, and chaotic states. While the system behavior can change radically with a small change in forcing, however, the degree of predictability is much less sensitive. Indeed, both the tendency error at short times and the climatological error over long times varies in a surprisingly smooth fashion with forcing. It appears that, if model error is present, chaos is sometimes of secondary importance in determining predictability. Again, these results do not all generalise to other nonlinear systems; for example, (Lorenz, 1964) provides an example in which the climate mean is unstable when the system is in its chaotic phase.

To sum up, for the Lorenz '96 systems studied here, the long and the short of it is that the long and the short are the same: if observation error is uncorrelated with tendency error, then optimising for short range accuracy also gives the best model for long range climatology, and vice versa. It is a topic of future work to investigate whether, and to what extent, these conclusions can be applied also to full weather or ocean models. The structural similarities between the Lorenz '96 systems and such models imply that there may at least be useful parallels. If this is the case, then improvements to the detailed near-term forecasts may also pay

30

significant dividends in longer-term climatology predictions; and studies of short-term model error may yield information on the effect of model error for climate studies.

Acknowledgements

References

Alligood, K., Sauer, T., and Yorke, J., 1996: *Chaos: An Introduction to Nonlinear Dynamical Systems*. Springer-Verlag, New York.

Collins, M., 2002: Climate predictability on interannual to decadal time scales: The initial value problem. *Climate Dynamics*, **19**, 671-692.

Gilmour, I., 1998: Nonlinear model evaluation: ι-shadowing, probabilistic prediction and weather forecasting. *D. Phil. Thesis, Oxford University*.

Gilmour, I., Smith, L., and Buizza, R., 2001: On the duration of the linear regime: Is 24 hours a long time in weather forecasting? *J. Atmos. Sci.*, **58**, 3525-3539.

Hansen, J. and Smith, L., 2000: The Role of Operational Constraints in Selecting Supplementary Observations. *J. Atmos. Sci.*, **57**, 2859-2871.

Hunt, J., 1999: Environmental forecasting and turbulence modeling. *Physica D*, **133**, 270-295.

Klinker, E. and Sardeshmukh, P., 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *J. Atmos. Sci.*, **49**, 608-627.

Lorenz, E., 1996: Predictability - a problem partly solved. *Predictability*,

edited by T. Palmer, European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading UK.

Lorenz, E., 1964: The problem of deducing the climate from the governing equations. *Tellus*, **16**,1-11.

Lorenz, E., 1975: Climate predictability: the physical basis of climate modeling. *WMO, GARP Publication Series*, **16**, 132-136.

Lorenz, E. and Emmanuel, K., 1998: Optimal Sites for Supplementary Weather Observations. *J. Atmos. Sci.*, **55**,399-414.

Orrell, D., 2002: Role of the metric in forecast error growth: how chaotic is the weather? *Tellus*, **54A**, 350-362

Orrell, D., 2001: Modelling nonlinear dynamical systems: chaos, error and uncertainty. *D.Phil. thesis, Oxford University*.

Orrell, D., Smith, L., Barkmeijer, J., and Palmer, T., 2001: Model error in weather forecasting. *Nonlin. Proc. Geo.*, **8**, 357-371.

Orrell, D. and Smith, L., 2003: Visualizing bifurcations in high dimensional systems: The spectral bifurcation diagram. *Int. J. Bifurcations and Chaos*, in press.

Schubert, S. and Schang, Y., 1996: An objective method for inferring sources of model error. *Monthly Weather Review*, **124**, 325-340.

Smith, L., 1996: Accountability in ensemble prediction. *Predictability*, edited by T. Palmer, European Centre for Medium-Range Weather Forecasting, Shinfield Park, Reading UK.

Strang, G., 1986: *Introduction to applied mathematics*. Wellesley-Cambridge Press.
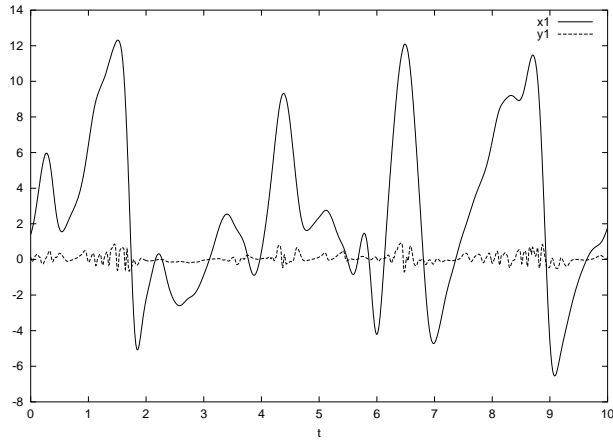
Figure 1: Time series of the System II variables, evaluated using a Runge-Kutta scheme with a timestep of 0.005. The $\tilde{y}$'s tend to fluctuate ten times more rapidly but with ten times smaller magnitude than the $\tilde{x}$'s.
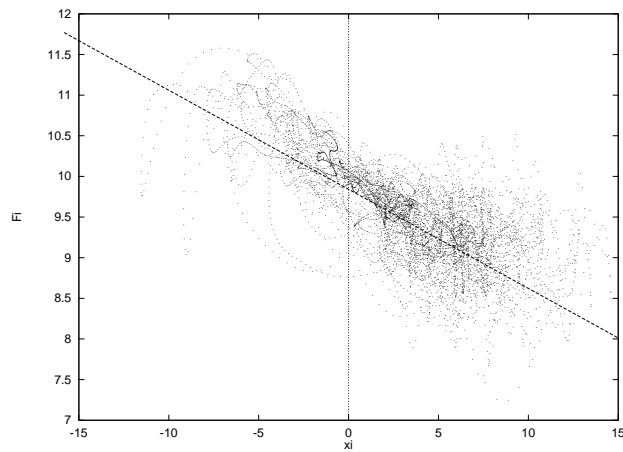

Figure 2: Scatter plot of local forcing versus local $\tilde{x}$ at $F = 10$. Also shown is the linear interpolation used in the corresponding linear model.

35
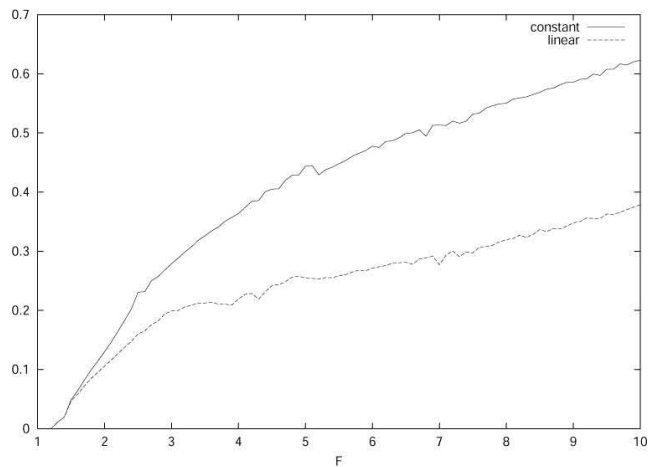
Figure 3: Plot of root-mean-square tendency error for the constant (upper line) and linear (lower line) models as a function of F. Errors have been divided by a factor $\sqrt{n}$, where $n = 8$ is the model dimension, so that they represent the error per component and can be compared with F. The tendency error varies in a fairly smooth manner with F.
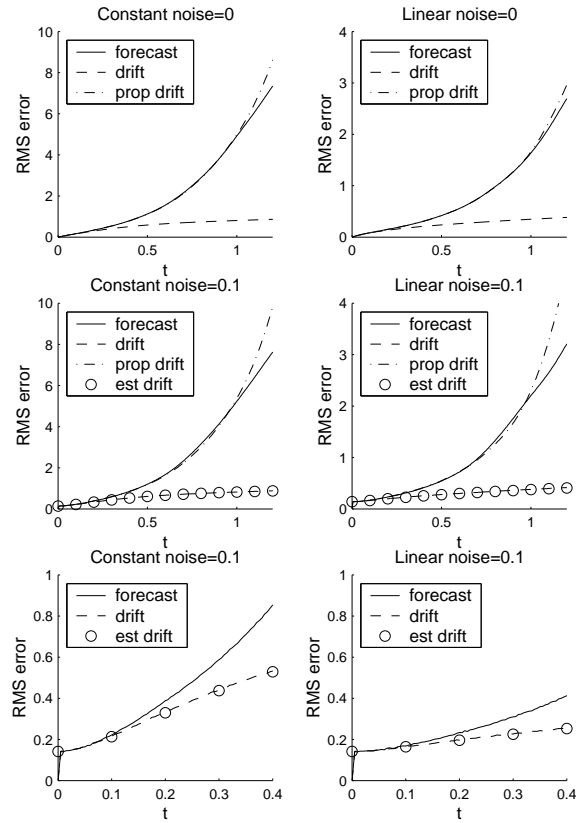
Figure 4: Comparison of drift and propagated drift with the total forecast error for constant (left) and linear (right) models. True system has $F = 10$. Results are root-mean-square over 200 starting points. Upper panels show case where observation error is zero. For either model, the drift underestimates the total error, while the propagated drift is almost indistinguishable from the total error up to about $t = 1.0$. Middle panels show the effect of adding Gaussian noise with standard deviation 0.1 to observations of the true system. Circle symbols indicate drift estimated using Eq. 23. Lower panels show zoomed view to $t = 0.4$. For comparison, the model's internal doubling time is about $t = 0.23$.

37
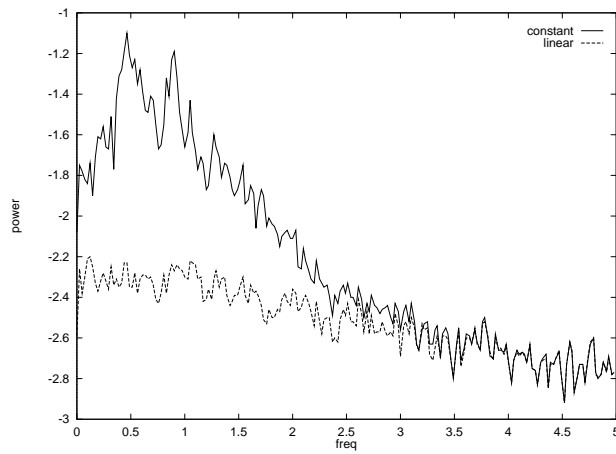
Figure 5: Power spectra of forcing error for constant (upper line) and linear (lower line) models at $F = 10$. The linear model error spectrum is relatively flat, while the constant model error contains more power at low frequency.
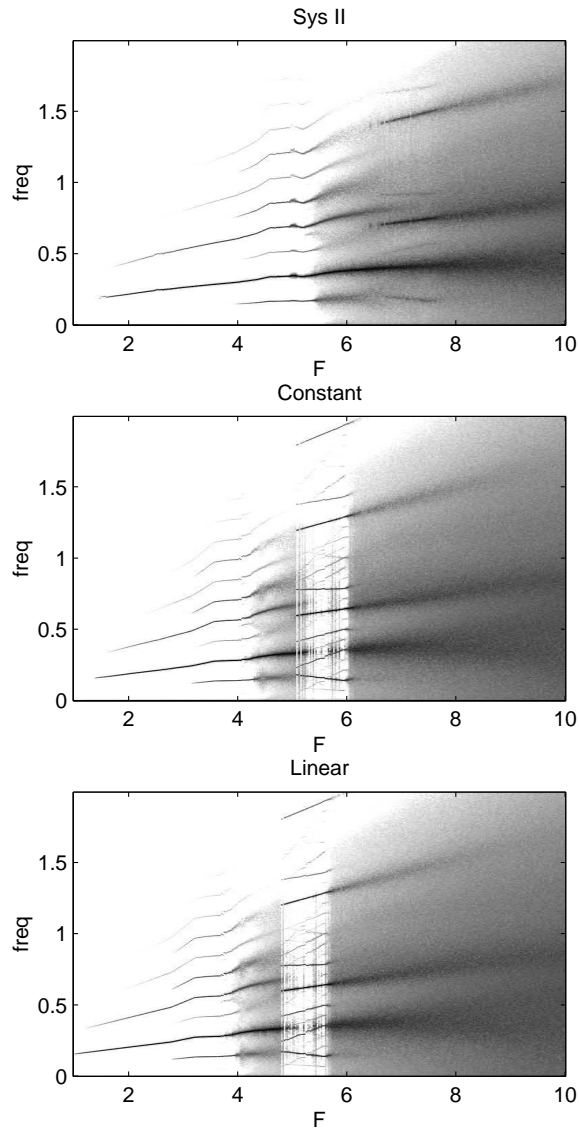
Figure 6: Spectral bifurcation diagrams of System II (upper panel), constant model (middle panel) and linear model (lower panel). Horizontal axis is the System II forcing F (as opposed to the model forcing), vertical axis is frequency. Shading indicates the logarithm of the power at each frequency. Periodic orbits appear as equispaced bands, chaotic regions show power at all frequencies. System II is periodic at low forcing and becomes fully chaotic around $F = 5.5$. The models are similar, but become chaotic earlier, and there are bands of periodic and quasi-periodic behaviour inside the chaotic regions, for example around $F = 5.3$.
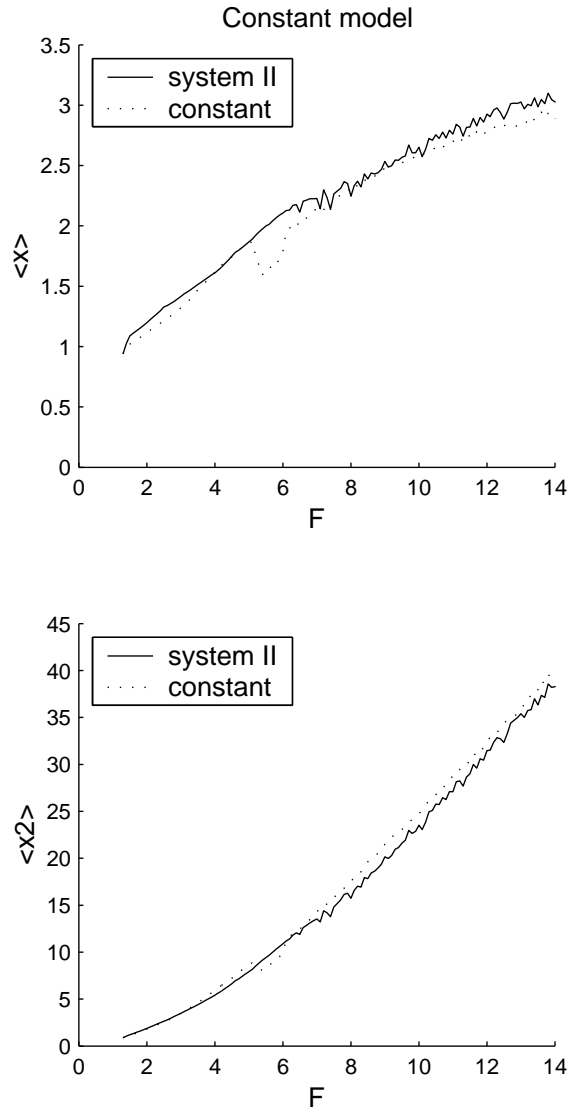
39

Figure 7: Comparison of System II and constant model climatologies. Upper panel $\langle \tilde{x}_1 \rangle$ for System II (solid line), and $\langle x_1 \rangle$ for the constant model (dotted line), as a function of F. Lower panel shows $\langle \tilde{x}_1^2 \rangle$ and $\langle x_1^2 \rangle$. The model mean is in most places slightly too low, and the variance too high, but the overall match is good.
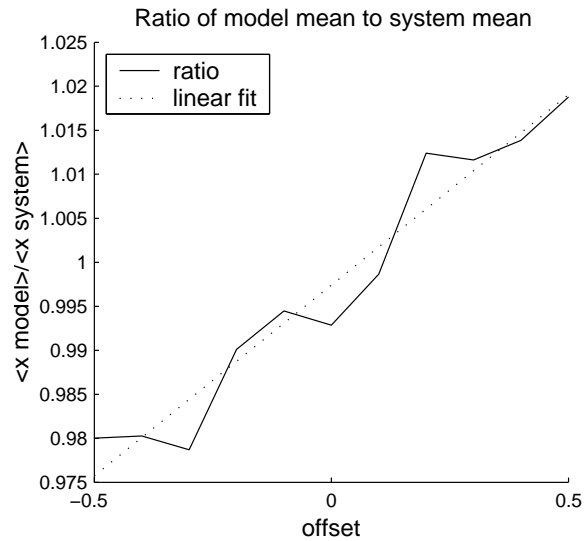
Figure 8: Ratio of $\langle x_i \rangle$ to $\langle \tilde{x}_i \rangle$ for values of the linear model offset $\alpha_0 - F$ where $F = 10$. Dotted line shows a straight line interpolation. The linear model gives $\langle x_i \rangle \approx \langle \tilde{x}_i \rangle$ for offsets near zero. The corresponding slope agrees with that of the linear model used for shadowing purposes. The variation may be due to numerical effects.
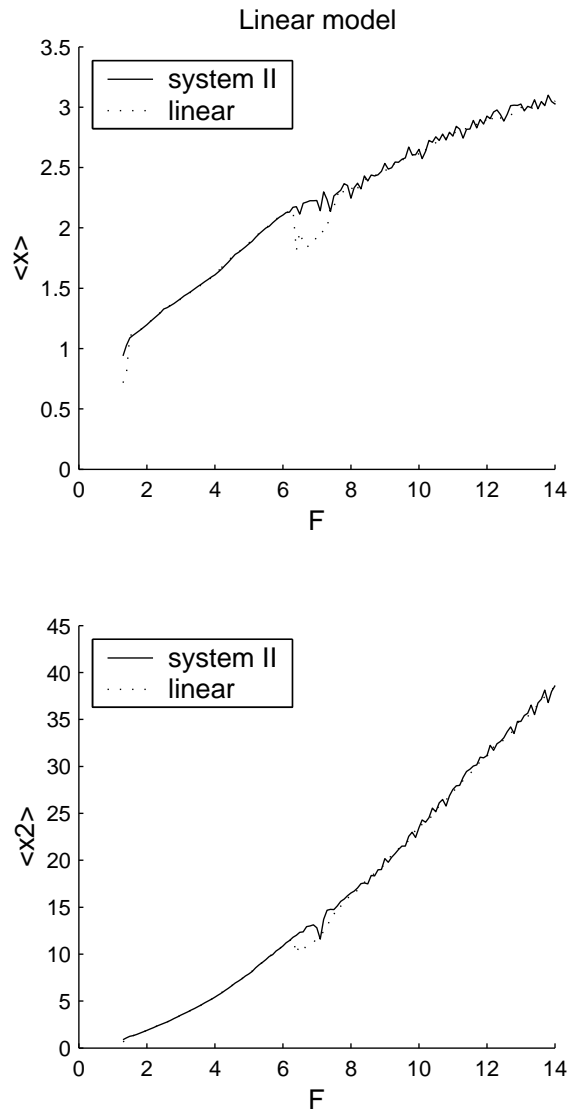
Figure 9: Comparison of System II and linear model climatologies. Upper panel shows $\langle \tilde{x}_1 \rangle$ for System II (solid line), and $\langle x_1 \rangle$ for the linear model (dotted line), as a function of F. Lower panel shows $\langle \tilde{x}_1^2 \rangle$ and $\langle x_1^2 \rangle$. Agreement is excellent except in the region of $F = 1.3$ and $F = 7$.