ELSEVIER

Physics Letters A 9909 (2000) xxx

PHYSICS LETTERS A

www.elsevier.nl/locate/pla

Localized Lyapunov exponents and the prediction of predictability

Christine Ziehmann [a,*], Leonard A. Smith [a,b,c], Jürgen Kurths [a]

[a] *Universität Potsdam, Institut für Physik, Nichtlineare Dynamik, Postfach 60 15 53, D-14469, Potsdam, Germany*
[b] *Mathematical Institute,University of Oxford, Oxford, OX1 3LB, UK*
[c] *London School of Economics, London WC2A 2AE, UK*

Received 9 December 1999; received in revised form 19 April 2000; accepted 3 May 2000
Communicated by C.R. Doering

Abstract

Every forecast should include an estimate of its likely accuracy, a current measure of predictability. Two distinct types of localized Lyapunov exponents based on infinitesimal uncertainty dynamics are investigated to reflect this predictability. Regions of high predictability within which *any* initial uncertainty will decrease are proven to exist in two common chaotic systems; potential implications of these regions are considered. The relevance of these results for finite size uncertainties is discussed and illustrated numerically.

PACS: 05.45.-a; 05.10.-a; 05.45.Tp
Keywords: Lyapunov exponents; Predictability; Chaos; Weather forecasting; Ensemble prediction; Nonlinear dynamics

1. Introduction

Prediction of predictability refers to the quantitative attempt to assess the likely error in a particular forecast a priori. There are (at least) three sources of difficulty in quantifying predictability: (1) the dependence of measures of predictability upon the particular metric adopted (2) the dependence of uncertainty dynamics upon the magnitude of the uncertainty in the initial condition, and (3) the fact that errors in the model are often unknown until after predictions are observed to fail. Lyapunov exponents [1–3] quantify predictability through globally averaged effective growth rates of uncertainty in the limits of large time and small uncertainty; thus by construction they are of limited use. To obtain a quantitative estimate of the accuracy of a particular forecast, the local dynamics of uncertainties about that initial condition

* Corresponding author. Tel.: +49-331-977-1302, fax: +49-331-977-1142.
E-mail address: chriss@agnld.uni-potsdam.de (C. Ziehmann).

0375-9601/00/$ - see front matter
PII: S0375-9601(00)00336-4

are more relevant [4–9]. By allowing better risk assessment, a prediction of predictability is of value in *any* field from physics to economics; weather forecasting provides a particular example relevant to both fields.

Effective growth rates defined over a fixed duration are also used to quantify predictability; they are employed daily in the operational weather forecast centers of Europe and North America [10,11]. In Section 2, the distinction between what will be called finite time exponents and finite sample exponents is shown to lie in the particular initial *orientation of the perturbation* each considers for a given initial condition; this can result in dramatically different *effective* growth rates. Both types of exponent are called 'Local Lyapunov exponents', and recognizing the distinction between them resolves some confusion in the literature. In terms of predicting the forecast accuracy, the finite time exponents are shown to be the more relevant quantities in Section 3, where it is also proven that the mean of the largest finite time exponent does not provide an unbiased estimate of the corresponding global exponent, and similarly for the mean of the smallest finite time exponent. In addition, in regions of state space where the largest finite time exponent is less than zero *all* perturbations will shrink independent of their orientation; this is investigated in Section 4 where such regions are proven to exist in two common chaotic maps.

Each class of Lyapunov exponent discussed in this paper assumes the observational uncertainty is infinitesimal; of course as long as it remains infinitesimal it cannot limit predictability, and once it is finite its growth is no longer quantified by Lyapunov exponents. Therefore, the rigorous results restricted to infinitesimal uncertainties are contrasted with numerical demonstrations for finite uncertainties in Section 4. Part of the popularity of global Lyapunov exponents stems from the fact that their value does not depend upon the metric (or coordinate system) used; this is not the case for the exponents based upon a finite length of trajectory, yet in practice only the latter are available. We return to this issue in Section 5. Finally, there is the question of model error in nonlinear forecasting, either parametric or structural [12]. Arguably, model error may be more responsible for poor predictions of real nonlinear systems than 'chaos.' Model imperfections are not considered in this paper as there is no systematic manner to include system/model mismatches, thus it is assumed throughout the paper that the perfect model is known.

2. Localized Lyapunov exponents

The dynamics of infinitesimal uncertainties about a point $\boldsymbol{x}_0$ in an m-dimensional state space are governed by the linear propagator, $\mathcal{M}(\boldsymbol{x}_0, \Delta t)$, which evolves any infinitesimal initial uncertainty $\boldsymbol{\epsilon}_0 \in \boldsymbol{R}^m$ about $\boldsymbol{x}_0$ forward for a time Δt along the system's trajectory to $\boldsymbol{x}_{\Delta t}$:

$$\boldsymbol{\epsilon}_{\Delta t} = \mathcal{M}(\boldsymbol{x}, \Delta t)\,\boldsymbol{\epsilon}_0. \tag{1}$$

In discrete time maps, the linear propagator over k iterations is simply the product of Jacobians along the trajectory, that is $\mathcal{M}(\boldsymbol{x}_0, k) = \mathcal{J}(\boldsymbol{x}_{k-1}) \ldots \mathcal{J}(\boldsymbol{x}_1)\,\mathcal{J}(\boldsymbol{x}_0)$.

For high dimensional systems, interest tends to be focused on subspaces which are likely to contain the fastest growing perturbations [4,10,11]. Two orientations of particular interest are (i) that which will have grown the most under the linearized dynamics after k steps, $\boldsymbol{v}_1^{(k)}(\boldsymbol{x})$, and (ii) the local orientation of the globally fastest growing direction, $\boldsymbol{l}_1(\boldsymbol{x})$, which is sometimes called the Lyapunov vector [13]. The first of these orientations is defined by the singular value decomposition [14] of the propagator: the $\boldsymbol{v}_i^{(k)}(\boldsymbol{x})$ are simply the right singular vectors of $\mathcal{M}(\boldsymbol{x}, k)$. Each is associated with a singular value, $\sigma_i^{(k)}(\boldsymbol{x})$; by convention, $\sigma_i \geq \sigma_{i+1}$. The *finite time Lyapunov exponents* [15] are

$$\lambda_i^{(k)}(\boldsymbol{x}) = \frac{1}{k}\log_2\left(\|\mathcal{M}(\boldsymbol{x}, k)\,\boldsymbol{v}_i^{(k)}(\boldsymbol{x})\|\right) = \frac{1}{k}\log_2\left(\sigma_i^{(k)}(\boldsymbol{x})\right). \tag{2}$$

Properties of these exponents have been noted by Lorenz [4], Grassberger et al. [16], Abarbanel [5] and references thereof. By Oseledec's Theorem [1], in the limit $k \to \infty$ the $\lambda_i^{(k)}(\boldsymbol{x})$ converge to a unique set of values, the Lyapunov exponents Λ_i, which are the

same for almost all $\boldsymbol{x}$ with respect to an ergodic measure.

$$\Lambda_i = \lambda_i^{(\infty)} \equiv \lim_{k\to\infty} \frac{1}{k}\log_2\left(\sigma_i^{(k)}\right), \quad i = 1,2,\ldots,m \tag{3}$$

If the sum of the Λ_i is negative, a volume element in state space will shrink, on average, as it evolves along a trajectory, and most initial conditions will evolve towards an attractor of dimension less than m [17]. At each point $\boldsymbol{x}$ on such an attractor, the orientation $\boldsymbol{l}_1(\boldsymbol{x})$ denotes the orientation corresponding to Λ_1, that is the orientation towards which almost every uncertainty $\boldsymbol{\epsilon}$ in the sufficiently distant past would have evolved, when the trajectory reaches $\boldsymbol{x}$. Similarly, define $\boldsymbol{l}_m(\boldsymbol{x})$ as the orientation corresponding to Λ_m (for details, see [17]). Numerically, $\boldsymbol{l}_1(\boldsymbol{x}_0)$ and $\boldsymbol{l}_m(\boldsymbol{x}_0)$ can be approximated by evolving the singular vectors of $\mathcal{M}(\boldsymbol{x}_{(-j)}, 2j)$, (that is, the $2j$ step propagator about the j^{th} pre-image of $\boldsymbol{x}_0$), forward j steps until the trajectory reaches $\boldsymbol{x}_0$. Thus

$$\boldsymbol{l}_i^{(j)}(\boldsymbol{x}_0) = \mathcal{M}\left(\boldsymbol{x}_{-j}, j\right)\boldsymbol{v}_i^{(2j)}(\boldsymbol{x}_{-j}), \quad i = 1,m. \tag{4}$$

As $j\to\infty$, we expect $\boldsymbol{l}_i^{(j)}(\boldsymbol{x}_0)$ to approach the orientations of $\boldsymbol{l}_i(\boldsymbol{x}_0)$ for $i = 1$ and $i = m$, leading to the definition[1] of the *finite sample Lyapunov exponents*

$$\rho_1^{(k)}(\boldsymbol{x}) = \frac{1}{k}\log_2\left(\|\mathcal{M}(\boldsymbol{x},k)\boldsymbol{l}_1(\boldsymbol{x})\|\right), \tag{5}$$

and $\rho_m^{(k)}(\boldsymbol{x})$ is similarly defined using $\boldsymbol{l}_m(\boldsymbol{x})$.

Both the $\rho^{(k)}$ and the $\lambda^{(k)}$ are often called 'local Lyapunov exponents' [5,19–23]. To avoid the confusion of this polysemy, we will call the $\lambda^{(k)}$ 'finite-time' (since they are completely defined by a finite segment of trajectory) and the $\rho^{(k)}$ 'finite-sample' (since they sample the growth of an orientation defined by the global dynamics). Both involve the same linear forward propagator, but each reflects the growth of a different orientation: the $\boldsymbol{l}_1(\boldsymbol{x})$ for the $\rho_1^{(k)}$ and the $\boldsymbol{v}_1(\boldsymbol{x})$ for the $\lambda_1^{(k)}$. As $k\to\infty$ both $\rho_1^{(k)}$ and $\lambda_1^{(k)}$ approach Λ_1, yet for the relatively small k over which forecasts are typically made their properties are quite different and neither is constrained by the value of Λ_1.

3. Properties of localized Lyapunov exponents

General constraints on the relative magnitudes of the largest and the smallest finite time and finite sample exponents are now derived from the definitions above, and then illustrated below. By construction, maximum growth corresponds to $\boldsymbol{v}_1$, thus $\rho_1^{(k)}(\boldsymbol{x}) \le \lambda_1^{(k)}(\boldsymbol{x})$ for each $\boldsymbol{x}$, and therefore this inequality also holds for the mean values $\langle \rho_1^{(k)}\rangle \le \langle \lambda_1^{(k)}\rangle$, where $\langle\cdot\rangle$ denotes an arithmetic average taken with respect to the natural measure[2], and $\langle\cdot\rangle_N$ a numerical approximation with sample size N. Examples from two chaotic systems are given in Fig. 1. The Henon map [24] is

$$x_{i+1} = 1 - ax_i^2 + y_i, \quad y_{i+1} = bx_i, \tag{6}$$

where $a = 1.4$ and $b = 0.3$; the Jacobian is independent of y_i and has constant determinant equal to $-b$. The Ikeda map [25]

$$x_{i+1} = 1 + \mu\left[x_i\cos(t) - y_i\sin(t)\right],$$
$$y_{i+1} = \mu\left[x_i\sin(t) + y_i\cos(t)\right], \tag{7}$$

with

$$t = 0.4 - \frac{6}{x^2 + y^2 + 1}$$

and $\mu = 0.9$ provides a rather more complex Jacobian, still with constant determinant, equal to μ^2 in this case. Note the non-Gaussian shape of all the distributions in Fig. 1, particularly those for small k. Contrasting the shapes of the distributions for the Henon and Ikeda systems suggests that such distributions will be strongly system dependent. Also note how the distributions sharpen with increasing k and that for given k the distributions of $\lambda_1^{(k)}(\boldsymbol{x})$ and $\rho_1^{(k)}(\boldsymbol{x})$ differ.

[1] If a specific $\boldsymbol{x}$ is not of interest, approximation of $\boldsymbol{l}_1(\boldsymbol{x}(t))$ at $\boldsymbol{x}(t)$ along a numerical trajectory can be simply approximated by a very long integration of the system and tangent equations for an arbitrary initial uncertainty. The quality of this approximation may be unknown, however, see [18] and the discussion in Section 5.

[2] We assume throughout that there exists a unique natural measure which is well approximated by the numerical iteration of the system.

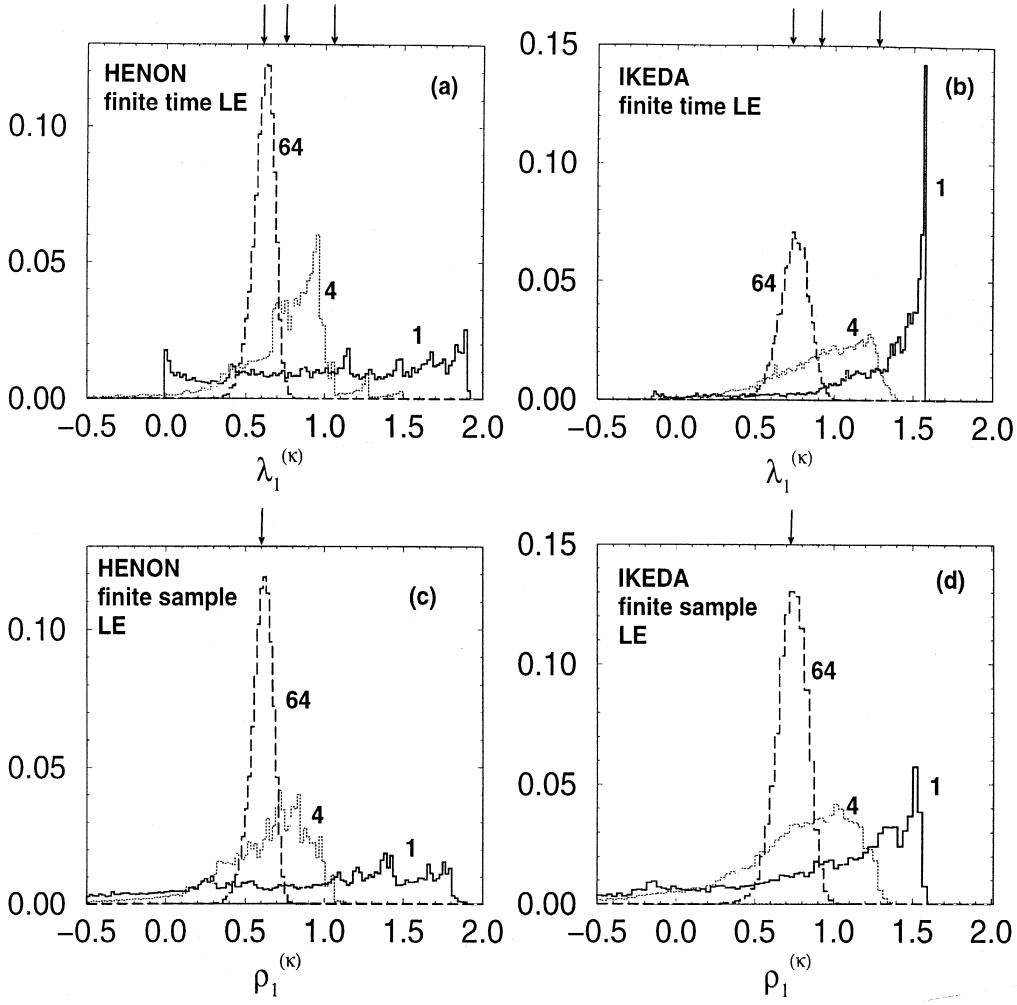

Fig. 1. Distributions of $\lambda_1^{(k)}$ [(a),(b)] and $\rho_1^{(k)}$ [(c),(d)] in the Henon and the Ikeda systems for $k = 1$ (thick), $k = 4$ (thin), and $k = 64$ (dashed), each with $N = 4096$. The arrows at the top axis indicate $\langle\rangle_N$. Larger bin widths have been used in the lower panels.

The mean values $\langle \lambda_1^{(k)} \rangle_N$ do not increase with increasing k. In fact $\langle \lambda_1^{(2k)} \rangle \leq \langle \lambda_1^{(k)} \rangle$ for any k, as can be seen by considering the matrix $\mathscr{C}$, the product of Jacobians $\mathscr{J}(\boldsymbol{x}_i), i = 1,2,\ldots,2k$. Divide $\mathscr{C}$ into two sub-products $\mathscr{A}$ and $\mathscr{B}$, each of length k:

$$\underbrace{\underbrace{\mathscr{J}(\boldsymbol{x}_{2k}) \cdots \mathscr{J}(\boldsymbol{x}_{k+1})}_{\mathscr{B}} \underbrace{\mathscr{J}(\boldsymbol{x}_k) \cdots \mathscr{J}(\boldsymbol{x}_1)}_{\mathscr{A}}}_{\mathscr{C}}$$

The first singular value $\sigma_1^{\mathscr{A}}$ of $\mathscr{A}$ reflects the maximum possible growth over the first k steps; the first singular value of matrix $\mathscr{C}$ must be less than or equal to the product of the first singular values of the matrices $\mathscr{A}$ and $\mathscr{B}$, thus[3] $\sigma_1^{\mathscr{C}} \leq \sigma_1^{\mathscr{B}} \sigma_1^{\mathscr{A}}$. The equality holds only if the first left singular vector $\boldsymbol{u}_1^{\mathscr{A}}$ of $\mathscr{A}$ is aligned with the first right singular vector $\boldsymbol{v}_1^{\mathscr{B}}$ of $\mathscr{B}$ (i.e. $\boldsymbol{u}_1^{\mathscr{A}} \cdot \boldsymbol{v}_1^{\mathscr{B}} = 1$), as in the uniform Baker's map [17] and Baker's Apprentice Maps [7]). From Eq. (2), the largest finite time Lyapunov exponent

[3] This follows immediately from the singular value decomposition (SVD) of a square matrix, $\mathscr{A} = \mathscr{U}\Sigma\mathscr{V}^T$, where the superscript ^{T} denotes the transpose of a matrix. Σ is a diagonal matrix whose largest entry is σ_1 and $\mathscr{U}$ and $\mathscr{V}$ are orthonormal rotation matrices. Noting that rotation matrices cannot enhance growth yields the desired result. A brief proof is given in the Appendix

defined by $\mathscr{C}$ must be less than or equal to the *average* of those defined by $\mathscr{A}$ and $\mathscr{B}$: $\lambda_1^{(2k)}(\boldsymbol{x}_1) = \frac{1}{2k}\log\sigma_1^{\mathscr{C}} \le \frac{1}{2k}\log(\sigma_1^{\mathscr{B}}\sigma_1^{\mathscr{A}}) = \frac{1}{2}[\lambda_1^{(k)}(\boldsymbol{x}_{k+1}) + \lambda_1^{(k)}(\boldsymbol{x}_1)]$. Similarly, the smallest finite time exponent defined by $\mathscr{C}$ must satisfy $\lambda_m^{(2k)}(\boldsymbol{x}_1) \ge -\frac{1}{2}[\lambda_m^{(k)}(\boldsymbol{x}_{k+1}) + \lambda_m^{(k)}(\boldsymbol{x}_1)]$. As this is true for each individual $\lambda_1^{(2k)}(\boldsymbol{x}_i)$ (alternatively $\lambda_m^{(2k)}(\boldsymbol{x}_i)$), the *mean* of the largest (smallest) finite time Lyapunov exponent will not increase (not decrease) as k increases by a factor of two:

$$\langle \lambda_1^{(2k)}\rangle \le \langle \lambda_1^{(k)}\rangle \quad \text{and} \quad \langle \lambda_m^{(2k)}\rangle \ge \langle \lambda_m^{(k)}\rangle. \tag{8}$$

When $\mathscr{A}$ and $\mathscr{B}$ are of different lengths, k_1 and k_2, then

$$(k_1+k_2)\langle \lambda_1^{(k_1+k_2)}\rangle \le k_2\langle \lambda_1^{(k_2)}\rangle + k_1\langle \lambda_1^{(k_1)}\rangle \tag{9}$$

and a similar relation is obtained for $\langle \lambda_m^{(k_1+k_2)}\rangle$, i.e., the $k\lambda_1^{(k)}$ ($k\lambda_m^{(k)}$) are sub-additive (super-additive) sequences of functions [26,27]. While this does not guarantee a monotonic decrease of $\langle \lambda_1^{(k)}\rangle$ (or increase of $\langle \lambda_m^{(k)}\rangle$) with increasing k, it does imply

$$\langle \lambda_1^{(k)}\rangle \ge \Lambda_1 \quad \text{and} \quad \langle \lambda_m^{(k)}\rangle \le \Lambda_m \quad \text{for all } k, \tag{10}$$

proving that the mean of the $\lambda_1^{(k)}$ distribution is *not* an unbiased estimate of the global Lyapunov exponent Λ_1 for any finite k [5,28].

The mean of the distribution of finite sample exponents is equal to Λ_1 by definition, independent of k. While in Fig. 1 each $\langle \lambda_1^{(k)}\rangle_N$ decreases with increasing k (as indicated by the arrows), the $\langle \rho_1^{(k)}\rangle_N$ coincide, providing a consistency check as to whether N might be large enough so that $\langle \rho_1^{(k)}\rangle_N$ approximates the limiting value $\langle \rho_1^{(k)}\rangle$.

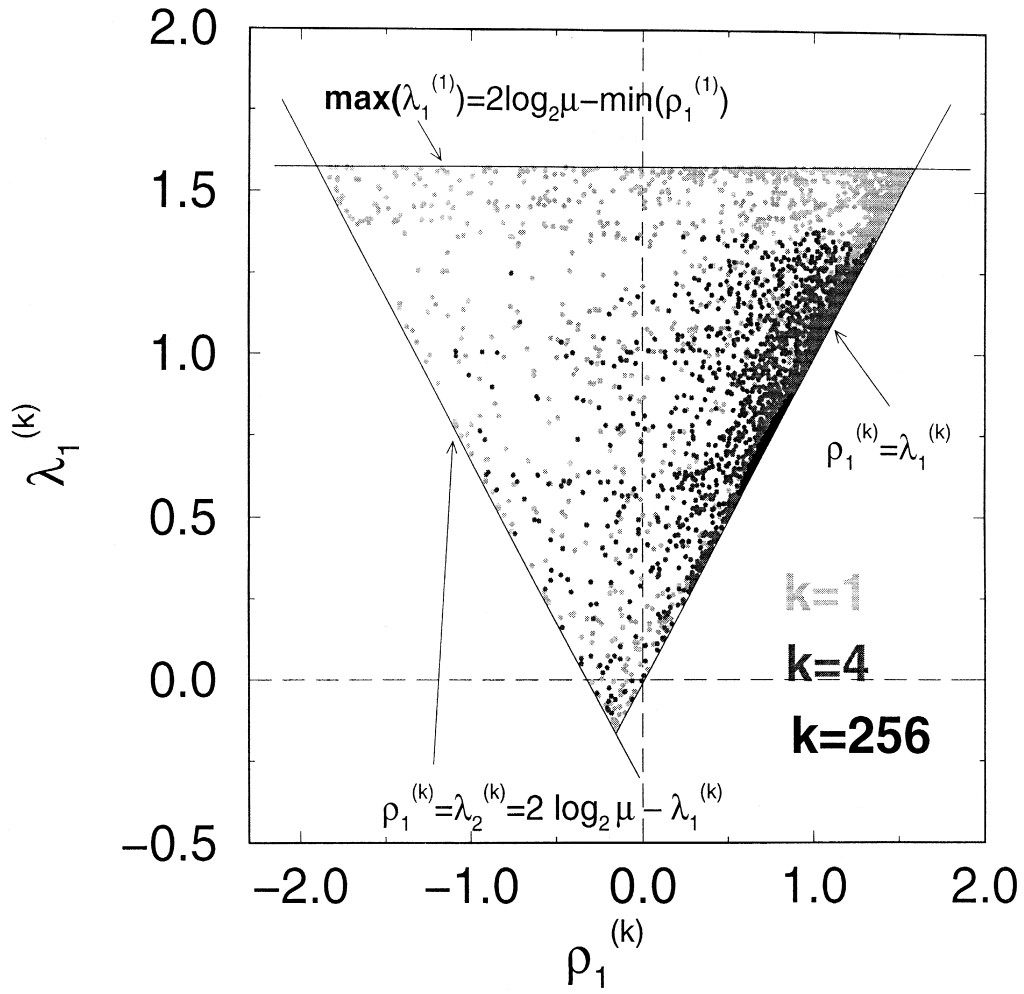

Fig. 2. Contrasting finite sample Lyapunov exponents (abscissa) with the corresponding finite time Lyapunov exponents (ordinate) in the Ikeda system for $k = 1$ (light grey), $k = 4$ (grey), and $k = 256$ (black) iterations. Note, that in the Ikeda system $\mathrm{Det}(J) = \mu^2$.

If the Jacobian determinant is constant, then for all $\boldsymbol{x}$

$$\sum_{i=1}^{m} \lambda_i^{(k)}(\boldsymbol{x}) = \log_2(\alpha), \qquad (11)$$

where $\alpha = |\det \mathcal{J}|$. Recalling that $\rho_1^{(k)}(\boldsymbol{x}) \le \lambda_1^{(k)}(\boldsymbol{x})$ and $\rho_m^{(k)}(\boldsymbol{x}) \ge \lambda_m^{(k)}(\boldsymbol{x})$, it follows that for both the Henon system and the Ikeda system $\rho_1^{(k)} \le \lambda_1^{(k)}$, $\rho_1^{(k)} \ge \lambda_2^{(k)} = \log_2(\alpha) - \lambda_1^{(k)}$, and

$$\lambda_1^{(k)} \le \max_{\boldsymbol{x}}\left(\lambda_1^{(1)}\right) = \log_2(\alpha) - \min_{\boldsymbol{x}}\left(\rho_1^{(1)}\right), \qquad (12)$$

defining a triangle which bounds the distribution of $(\rho_1^{(k)}(\boldsymbol{x}), \lambda_1^{(k)}(\boldsymbol{x}))$, as illustrated in Fig. 2 for the Ikeda map. With increasing k the distributions of points approach the line $\rho_1^{(k)} = \lambda_1^{(k)}$. For $k = 256$, the largest observed value of $|\lambda_1^{(k)}(\boldsymbol{x}) - \rho_1^{(k)}(\boldsymbol{x})|$ was 0.03. Yet the widths of each of these distributions exceeds 0.3, indicating that this variation stems from the different initial conditions on the attractor, not the initial orientation.

4. Regions of high predictability in chaotic maps

Lyapunov exponents are often said to reflect predictability, and a positive global Lyapunov exponent is often said to destroy any hope of 'long-term' predictability. But since they are defined via the linear propagator Lyapunov exponents need only quantify the growth of infinitesimal uncertainties in the initial condition, this is a high price to pay for invariance under a smooth change of coordinate. Both the Ikeda system (Eq. (6)) and the Henon system (Eq. (7)) are considered chaotic for the parameters considered above, since in each case it is believed that $\Lambda_1 > 0$; yet this does not imply $\lambda_1^{(k)}(\boldsymbol{x}) > 0$ for any finite k. Indeed it is clear from Fig. 2 that there are many points on the attractor (about 1.5 %) of the Ikeda system for which the leading $k = 1$ finite time exponent is negative, i.e. there are states about which *every* infinitesimal uncertainty will shrink regardless of its orientation.

We now proceed to locate the corresponding regions in state space with negative largest finite time Lyapunov exponents, $\lambda_1^{(k)} < 0$, which we interpret as likely regions of relatively high predictability: *all* infinitesimal initial uncertainties will *decrease* in these regions. Recently [9,29], similar regions have been determined analytically in the Lorenz [30] system (see also [20,31]); we now present new results for the Ikeda system and the Henon system. This result is then demonstrated (numerically) to hold in several cases for finite uncertainties, but the exact results below are subject to the caveat of infinitesimal uncertainties, as are all general arguments regarding the prediction of deterministic chaotic systems.

4.1. Infinitesimal uncertainties

Naturally, exact results are most easily obtained for small k. Therefore we consider only $k = 1$ and $k = 2$ analytically; numerical results are given for larger values. In a map $\lambda_1^{(1)}(\boldsymbol{x}) < 0$ implies that the largest singular value of the Jacobian $\mathcal{J}(\boldsymbol{x})$ is less than one. For the Ikeda system with μ in the range $(\sqrt{13} - 3)/2 < \mu < 1$, the one-step finite time Lyapunov exponent $\lambda_1^{(1)}(\boldsymbol{x})$ passes through zero at two circles about the origin with radii $r_{o,i} = \sqrt{6c - 1 \pm \sqrt{(6c-1)^2 - 1}}$ where $c = |\mu/(\mu^2 - 1)|$. In this case $\lambda_1^{(1)}(\boldsymbol{x}) < 0$ for all points either within the inner circle (i.e. those with $\sqrt{x^2 + y^2} < r_i$) or outside the outer circle (i.e. $\sqrt{x^2 + y^2} > r_o$). Fig. 3 shows points on the Ikeda attractor where the sign of $\lambda_1^{(1)}(\boldsymbol{x})$ is indicated by the grey scale. For $\mu = 0.9$, the radii are $r_i = 0.135$ and $r_o = 7.404$ (thus this attractor lies well within r_o, and the outer circle is not visible in Fig. 3). As $|\mu|$ approaches one, the radius of the inner circle goes to zero.

In the Henon system there are no regions in which every uncertainty will shrink after one iteration, that is, $\lambda_1^{(1)}(\boldsymbol{x}) \ge 0$ for *all* $\boldsymbol{x}$. As shown below, this is not the case for $\lambda_1^{(2)}$, however. The smallest value of $\lambda_1^{(1)}(\boldsymbol{x})$ is found for points $\boldsymbol{x}$ on the y-axis. Here $\lambda_1^{(1)}(\boldsymbol{x}) = 0$, and thus all uncertainties shrink except those aligned with $\boldsymbol{v}_1$, which remain unchanged in magnitude. Note that for points on the y-axis, $\boldsymbol{v}_1$ is parallel to the y-axis. We now prove that there exist a finite region within which all points have $\lambda_1^{(2)}(\boldsymbol{x}) < 0$; the region includes a portion of the first preimage of the y-axis.

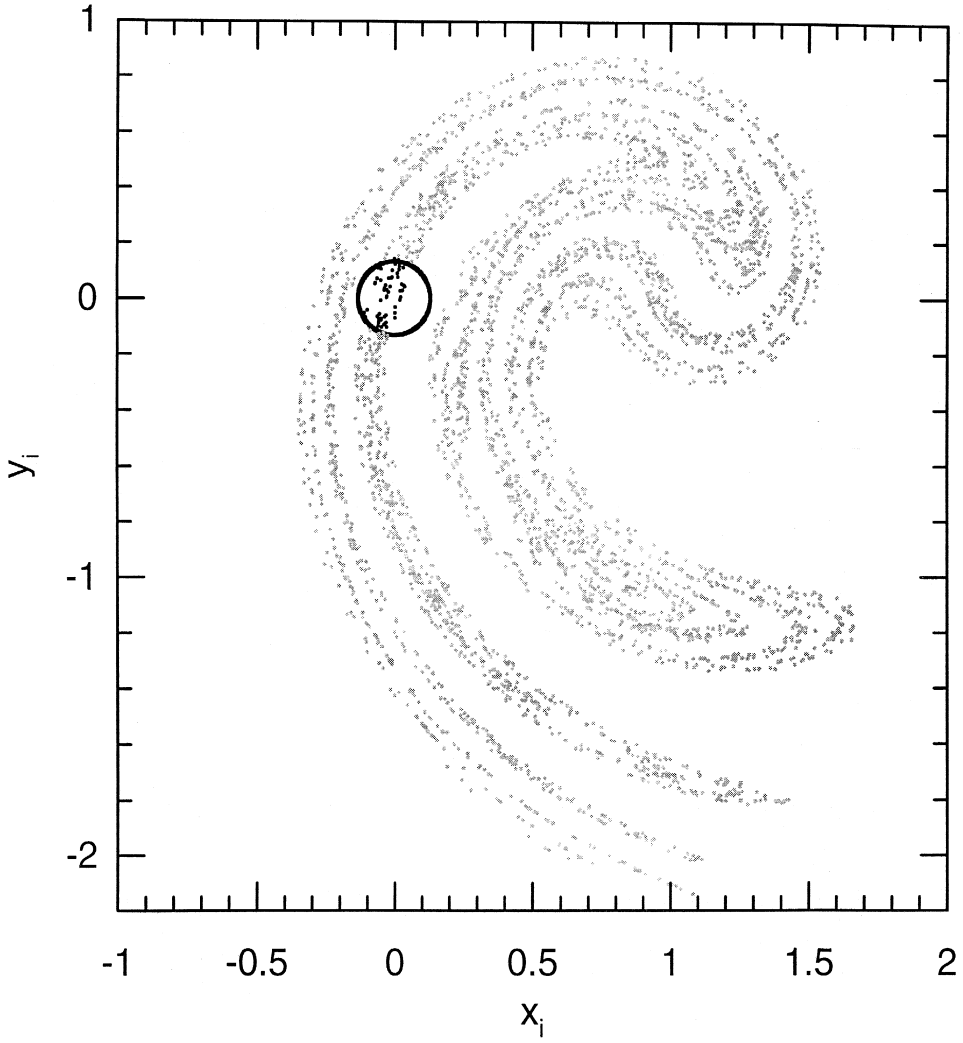

Fig. 3. Regions of decreasing uncertainty in the Ikeda system. Points on the attractor are colored grey if $\lambda_1^{(1)}(\boldsymbol{x}) > 0$, black otherwise. Within the circular region near the origin, $\lambda_1^{(1)}(\boldsymbol{x}) < 0$ for all $\boldsymbol{x}$.

The preimage of the y-axis is the parabola, $y = ax^2 - 1$, and the two step propagator for points $(\xi, a\xi^2 - 1)$ is

$$\begin{aligned}\mathcal{M}\big((\xi, a\xi^2 - 1), k = 2\big) &= \mathcal{J}_1(x = 0)\,\mathcal{J}_0(x = \xi) \\ &= \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix}\begin{pmatrix} -2a\xi & 1 \\ b & 0 \end{pmatrix} \\ &= \begin{pmatrix} b & 0 \\ -2ab\xi & b \end{pmatrix},\end{aligned}$$

where we have used the fact that in the Henon system $\mathcal{J}(\boldsymbol{x})$ is only a function of x. To locate those $\boldsymbol{x}$ with $\lambda_1^{(k=2)} < 0$, we determine the singular values σ_i of $\mathcal{M}((\xi, a\xi^2 - 1), k = 2)$, noting that $\sigma_i^2 = \nu_i$, where ν_i are the roots of the characteristic polynomial of $\mathcal{M}^T\mathcal{M}$. Thus

$$\nu^2 - 2b^2\big(2a^2\xi^2 + 1\big)\nu + b^4 = 0. \qquad (13)$$

We can now test whether $\sigma^2 < 1$ for any ξ; alternatively, we can solve for $\sigma = 1$ to find

$$\xi(\sigma = 1) = \pm\frac{b^2 - 1}{2ab} = \mp\frac{0.91}{0.84} \approx \pm 1.08\overline{3} \qquad (14)$$

for $a = 1.4$, $b = 0.3$. Note that the parabola and the x-axis intersect at the point (0,-1). At this intersection $x = \xi = 0$ and it follows from the equation above that $\sigma_1^{(k=2)} = \sigma_2^{(k=2)} = b < 1$ or, equivalently, that $\lambda_1^{(k=2)} = \frac{1}{2}\log_2(b) = \lambda_2^{(k=2)} < 0$. Showing that a point is negative in the range given by Eq. 14 proves

that all points on the parabola within these limits have $\lambda_1^{(k=2)} < 0$. By continuity, there exists a finite region in the neighborhood of the parabola for

$$-\frac{1-b^2}{2ab} < x < \frac{1-b^2}{2ab},$$

within which the largest two-step finite-time Lyapunov exponent is negative. A numerical estimate of this region is shown in Fig. 5 a.

What about larger values of k? Trajectories passing through this region of $\lambda_1^{(2)} < 0$ are often found to have $\lambda_1^{(k)} < 0$ for $k > 2$ as well; negative values of $\lambda_1^{(4)}$ are clearly visible in Fig. 1(a). In the following, we will illustrate the relation between the trajectories of points on the attractor with $\lambda_1^{(k)} < 0$ for $k > 2$ and the y-axis; namely that such points tend to lie near preimages of the y-axis. The first three preimages of the x-axis can be obtained analytically. The y-axis is the preimage of the x-axis, while the parabola noted above is the first preimage of the y-axis. The first preimage of the parabola is

$$x = \frac{a\xi^2 - 1}{b}, \quad y = \xi - 1 + a\left(\frac{(a\xi^2 - 1)}{b}\right)^2,$$

$$-\frac{1-b^2}{2ab} < \xi < \frac{1-b^2}{2ab}.$$

The first 4 preimages of the x-axis are shown together with the attractor in each panel of Fig. 4, while points on the attractor with $\lambda_1^{(k)} < 0$ are shown for the specific values $k = 2,3,4,5$ in the four panels.

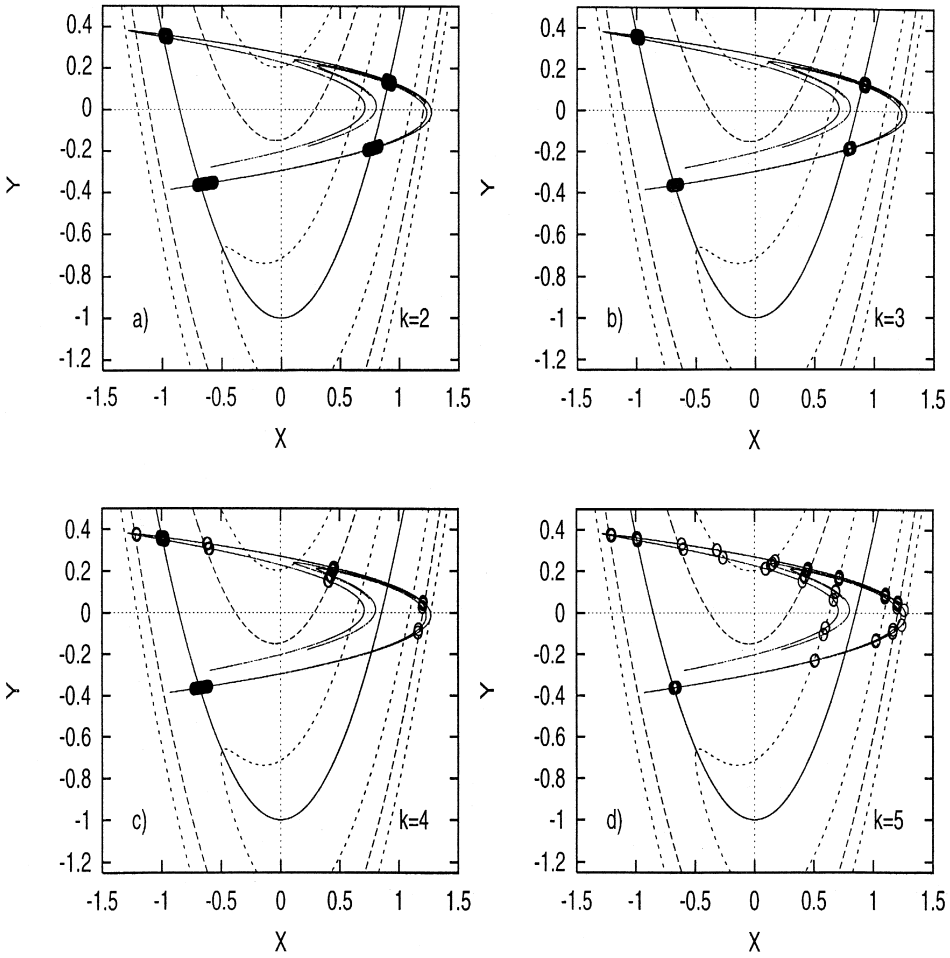

Fig. 4. All panels show the Henon attractor, the x-axis, the y-axis and its last three preimages: the parabola (solid line), its second preimage (long dashed) and its third preimage (short dashed). The different panels show points on the attractor with $\lambda_1^{(k)} < 0$ for $k = 2$ (a), $k = 3$ (b), $k = 4$ (c), and $k = 5$ (d).

It is clear that the regions with $\lambda_1^{(k)} < 0$ are related to the intersections of the attractor and preimages of the y-axis. We next show that this is even more evident for initial conditions in the general vicinity of the attractor. As in Fig. 4, Fig. 5 shows both the preimages of the x-axis and the attractor; in addition, test points for which $\lambda_1^{(k)} < 0$ are plotted as well, where the test points were drawn at random from the region shown in this figure. Points with decreasing uncertainties for $k \leq 5$ are found in the vicinity of preimages of the y-axis, i.e. they often have trajectories which include a near approach to the y-axis. Typically, this occurs towards the end of that trajectory: points with $\lambda_1^{(2)} < 0$ are close to the first preimage of the y-axis, points with $\lambda_1^{(3)} < 0$ are close to its first and second preimage and so forth.

Next, we investigate the behavior for even larger k in the Henon system. Grassberger et al. [16] showed that one should expect $\lambda_1^{(k)} < 0$ for arbitrarily large k, assuming a behavior essentially like averages of random variables correlated only over short times. This general picture is correct although the details are sometimes important, as we have argued elsewhere [15]. Here, we conjecture that, due to the deterministic nature of the Henon system, this fraction of $\lambda_1^{(k)}$ decreases more quickly than the random variable argument would suggest. As shown in Fig. 6, the fraction of initial conditions on the attractor with $\lambda_1^{(k)} < 0$ is observed to decrease exponentially with k, as are the corresponding fraction when linear propagators of the map are combined at random. The random case for $k = 1,2,4$ and 8 are shown; for each,

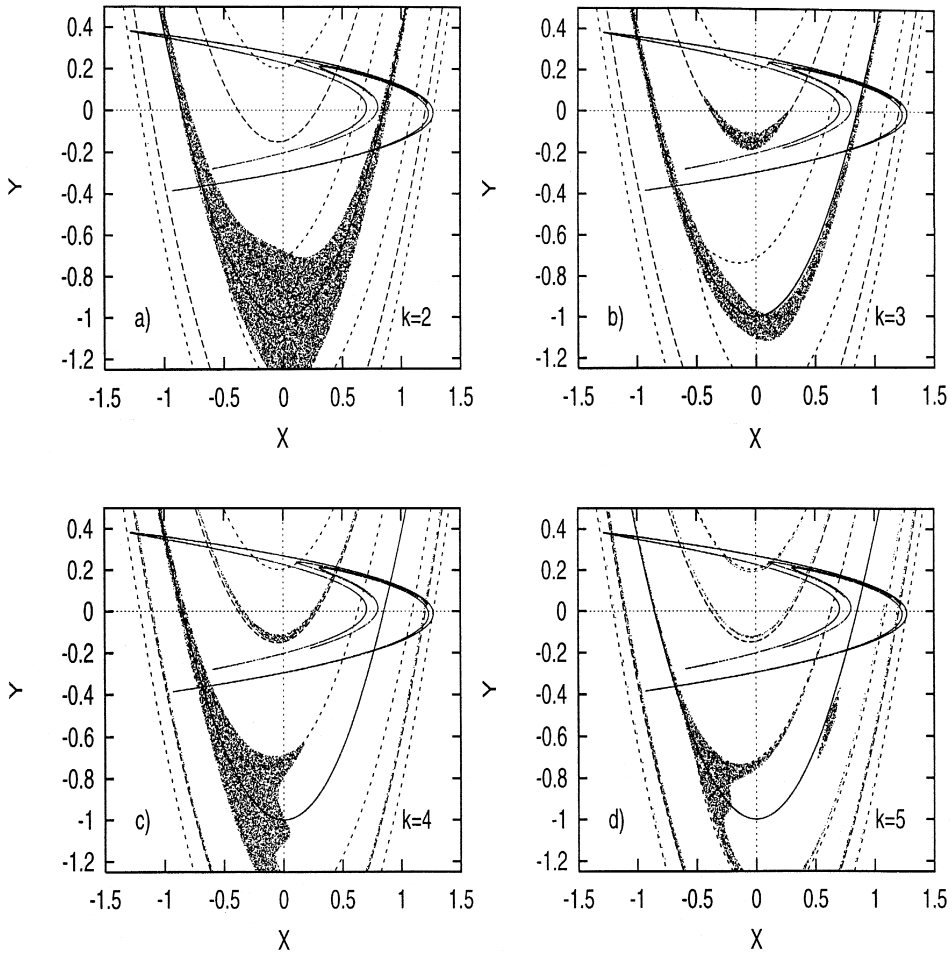

Fig. 5. All panels show the Henon attractor, the x-axis, the y-axis and its last three preimages: the parabola (solid line), its second preimage (long dashed) and its third preimage (short dashed). The different panels show points in the vicinity of the attractor with $\lambda_1^{(k)} < 0$ for $k = 2$ (a), $k = 3$ (b), $k = 4$ (c), and $k = 5$ (d).

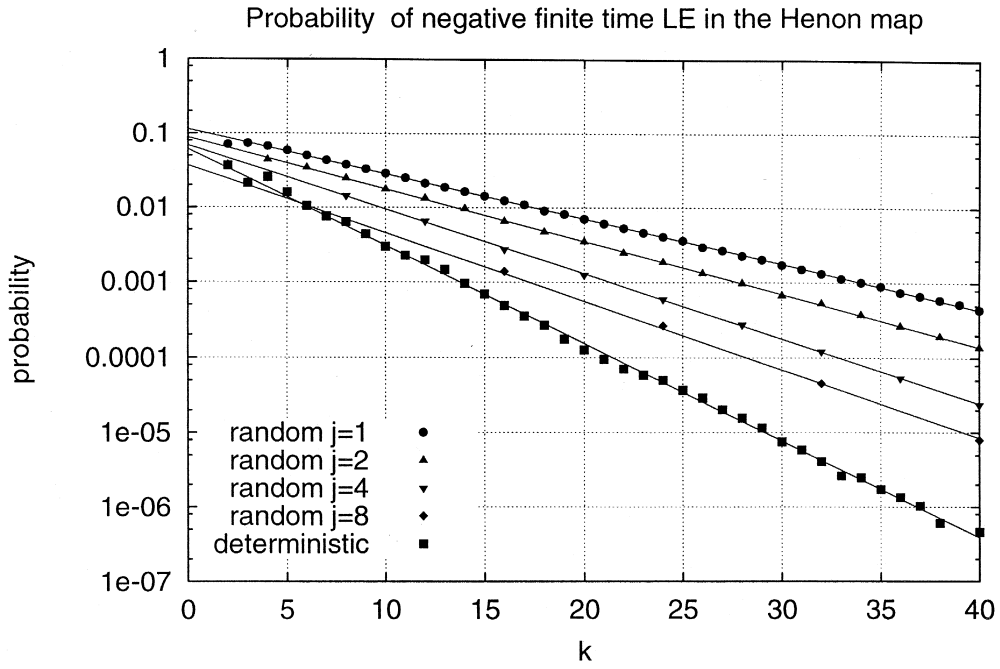

Fig. 6. For the Henon system $(a = 1.4, b = 0.3)$, the fraction of points with $\lambda_1^{(k)} < 0$ as a function of k in the deterministic case (squares). Also shown are the results for random matrices, where the matrices are drawn from the distribution of the linear propagators of the Henon map, that is $\mathscr{M}^{(j)}$, for $j = 1,2,4$ and 8. The solid lines reflect the best fit to an exponential decay over the range $8 \le k \le 40$. For large k, about 2^{30} iterations of the map where considered in the deterministic case. Note, that determinism is a strong constraint reducing the likelihood of finding negative finite time exponents.

the fraction decreases less quickly than in the deterministic case. Disregarding the determinism in the series of Jacobians leads to frequencies of negative $\lambda_1^{(k)}$ which for the larger k exceed those of the deterministic case by orders of magnitudes.

4.2. Finite uncertainties

From the practical point of view of estimating predictability, the knowledge of such points would be of limited utility for large k, since the shrinking region around each point may be very small. Further, recall that all estimates of predictability based upon Lyapunov exponents assume an infinitesimal initial error. Therefore, we next consider finite uncertainties explicitly, first exploring Gaussian distributed uncertainties in the Ikeda map, and then uncertainties of uniform magnitude in the Henon map. In each case we allow the expected magnitude of the error to vary and discuss the relation between regions of enhanced predictability and the regions where $\lambda_1^{(k)} < 0$.

In the Ikeda map, we consider normally distributed uncertainties in each coordinate of the initial condition with zero mean and the same standard deviation ϵ. Taking a point on the attractor at random, 128 'observations' were generated and predicted forward k steps; if the distance from truth at final time was less than the initial perturbation applied in more than 50% of the 128 cases, then the original point on the attractor was considered to be within a region of high predictability (for that value of ϵ). For $k = 1$ there is a region (not shown) of high predictability centered on the circle derived above and shown in Fig. 3. Points of high predictability have been observed for $\epsilon = 2^{-9}$ and persist for $\epsilon = \frac{1}{8}$. While all are near the origin, many fall outside the circle, but this is not surprising as the definition of high predictability in this numerical experiment is much less restrictive than requiring a negative finite time Lyapunov exponent, which guarantees 100% of the uncertainties to shrink if they have infinitesimal magnitudes. The same tests for $k = 4$ are shown in Fig. 7. The results based upon infinitesimal uncertainties shown in the upper left panel are seen to reflect the region of high predictability very well up until $\epsilon = 1/8$, which is a fair

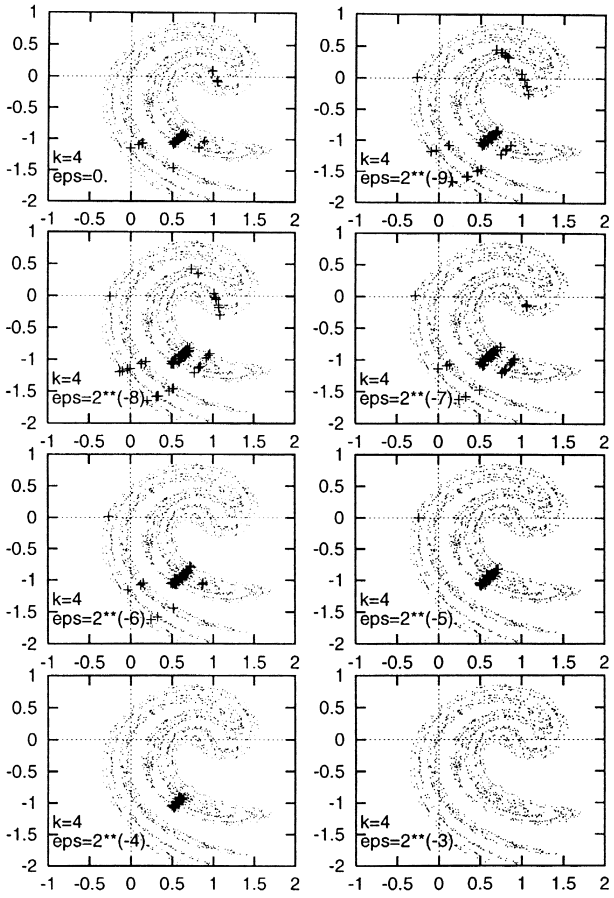

Fig. 7. All panels show the Ikeda attractor; the dots represent the attractor. Initial conditions on the attractor with enhanced predictability are marked with a '+' when more than 50% of 128 initial uncertainties show decreased magnitudes at final time $k = 4$. The different panels reflecting different initial magnitudes *eps* are contrasted with the linear dynamics in the upper left panel.

fraction of the diameter of the attractor. We stress that results of this kind will be extremely system specific.

For the Henon map the portrait of negative $\lambda_1^{(4)}$ revealed in Fig. 5 is contrasted with regions of high predictability for finite uncertainties in initial condition with a much sharper test than for the Ikeda map. In this case each observation is placed at random on a circle of radius ϵ about the true state; 1000 such 'observations' were considered for each true state and only if the final time distance of *every one* of them was less than ϵ was the point recorded as 'high predictability.' This is shown for 3 different initial magnitudes in Fig. 8. For small magnitudes, $\epsilon = 0.001$, the cartography is quite similar to that of infinitesimals; of course, structures smaller than ϵ cannot be detected. It becomes harder to identify regions with differing properties in predictability with increasing prediction time; sensitive dependence on initial condition will limit the prediction of predictability as well as prediction itself.

Thus far, we have only considered the value of an exponent *at* a particular value of $k = K$; alternatively one might consider regions in which the exponent is negative for all $k \leq K$, with corresponding uncertainties decreasing monotonically for the total duration of K iterations. While such subsets are of interest, they are not investigated here since the existence of small positive values at intermediate k are consistent with regions of high predictability; indeed the points omitted from the set of points for a particular k are those that are said to display 'return of skill' in meteorology [32]. Although beyond the scope of this paper, it would be interesting to examine the spatial distribution and fraction of initial condition in these subsets, both as a function of k and the magnitude of the initial uncertainty.

4.3. Coexistence of chaos and regions of high predictability

Another interesting aspect concerning the regions with $\lambda_1^{(k)} < 0$ is the interplay between these regions and the location of unstable periodic orbits, which are believed to form the skeleton of the attractor in many chaotic systems [33]. Clearly, an unstable period-k orbit cannot contain a point within a region for which $\lambda_1^{(k)} < 0$, since that point would then be stable. In short, each point on each unstable period k orbit must avoid all regions of the state space in which $\lambda_1^{(k)} < 0$: it is not easy to see how this comes about *if* the orbits are dense on the attractor *and* the area of the regions does not vanish. If the regions do not vanish, then this observation suggests a new angle from which to view the extreme sensitivity of the structure of the attractor to small changes in parameter value.

It is also interesting to consider the implications positive $\lambda_1^{(k)}$ might have on numerical results when the true attractor is a stable attracting periodic orbit; no point in a periodic orbit need lie in a region for which $\lambda_1^{(k)} < 0$. For $a \approx 1.37511006867, b = 0.3$, the

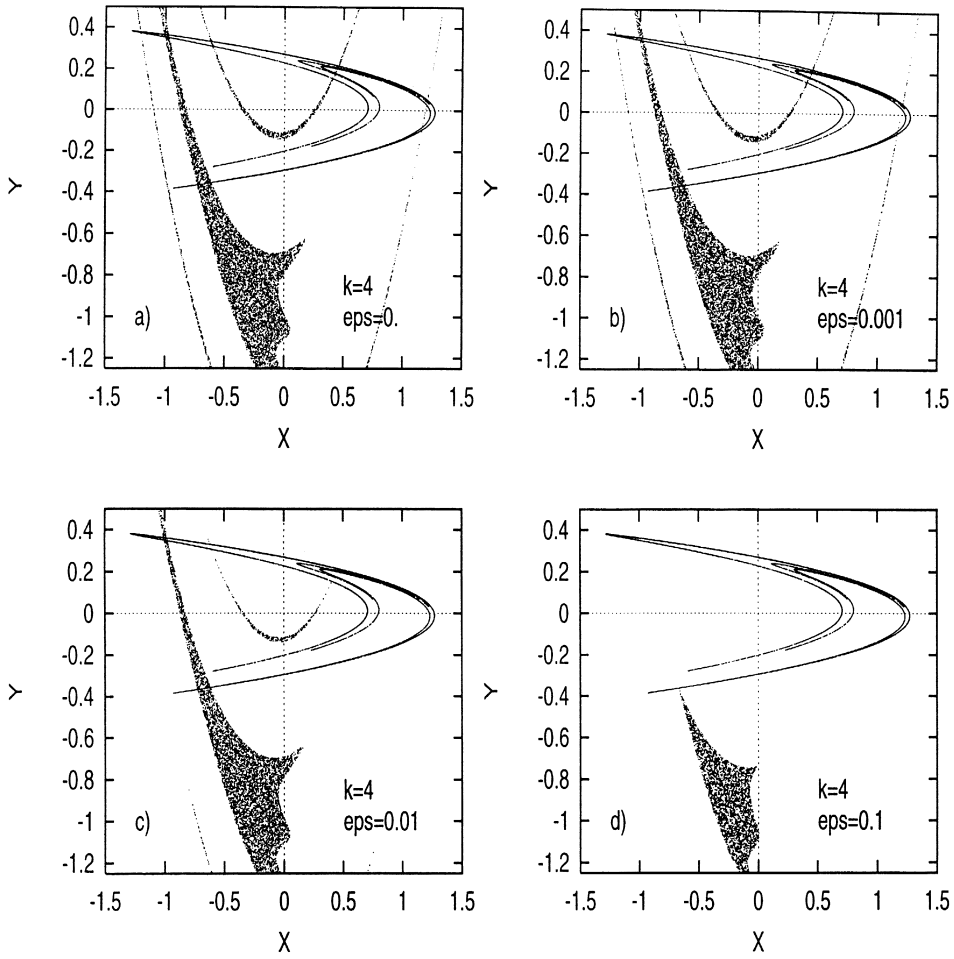

Fig. 8. All panels show the Henon attractor; the dots represent initial conditions with enhanced predictability for finite uncertainties. The different panels belong to different initial magnitudes of initial uncertainty: (a) *eps* = 0 (thus coinciding with Fig. 5c). In panels (b), (c), and (d), a dot at $\boldsymbol{x}$ indicates that the distance between the image of the true state and *each one* of 1000 inexact 'observations' decreased at $k = 4$. The observations were initially distributed on a circle of radius *eps* centered on $\boldsymbol{x}$.

Henon system has a stable period 24 orbit. The majority of points on this orbit have $\lambda_1^{(24)} < 0$, but for several, $\lambda_1^{(24)} > 0$; the largest observed value is $\lambda_1^{(24)} \approx 0.3$, implying a magnification factor of more than a hundred within one cycle. When $\mathcal{M}$ is non-normal, the magnitude of the leading singular value may be quite large regardless of whether or not the orbit is asymptotically stable. With a slight increase in a the system appears chaotic; the dynamics still resemble those of the stable orbit, but the attractor now consists of 24 small, clearly separated chaotic 'regions,' each visited in turn. It would be interesting to examine the distribution of leading singular values about stable periodic points on the same orbit, as a function of parameter; is there a positive lower bound on the angle between these eigenvectors?

This observation suggests an interesting possible parallel between these simple two dimensional maps and the onset of turbulence in laminar fluid flows. It has long been known that shear flows can become turbulent at Reynolds numbers well below the critical value as defined by the classical linear stability theory based on eigenvalues ([34] and references thereof; for a recent overview see [35]). Perturbations in the directions of the singular vectors may grow rapidly for a (finite) time, exciting nonlinear terms and thereby dominating the onset of turbulence; the long term behavior described by the eigenvalues becomes irrelevant. Non-normality might hold similar consequences for the numerical iteration of nonlinear systems. The smallest nonzero *numerical* perturbation is finite, being defined by the numerical

grid; and it could be quite difficult to identify a stable period k orbit with $\lambda_1^{(k)} > 0$ by numerically iterating the map. The smallest nonzero numerical perturbation might well grow sufficiently to bring the nonlinear terms into play, resulting in sustained, complicated dynamics up to the time-scales at which the numerical orbit closed exactly upon itself, as all trajectories on digital computers will [36]. We do not claim that this is the case in the maps considered above, but merely note the dynamics might appear similar and thus stress the value of performing a bifurcation analysis in addition to numerical iteration.

5. Discussion and conclusions

The results presented in this article hold implications for two questions of general interest: the approximation of largest Lyapunov exponent Λ_1 and the estimation of likely forecast accuracy. Noting that the finite time Lyapunov exponents $\langle \lambda_i^{(k)} \rangle$ can be computed accurately by standard methods, a lower bound on the error in assuming $\Lambda_1 \approx \langle \lambda_1^{(k)} \rangle_N$ is given by $\langle \lambda_1^{(k)} \rangle_N - \langle \lambda_1^{(2k)} \rangle_N \equiv \delta$, while $\langle \lambda_1^{(k)} \rangle$ provides an upper bound on Λ_1. When a good approximation of the 'Lyapunov vector' $\boldsymbol{l}_1$ is available, one can also require for the difference between finite sample and finite time exponents $|\langle \rho_1^{(k)} \rangle - \langle \lambda_1^{(k)} \rangle| < \delta$. Yet since both $\boldsymbol{l}_1$ and Λ_1 are multiplicative ergodic statistics [1], uncertainty in numerical estimates of $\boldsymbol{l}_1(\boldsymbol{x})$ remains largely unquantified. Similarly, inasmuch as matrix multiplication does not commute, attempts to estimate the uncertainty in Λ_1 via the standard bootstrap approach must be treated with care for deterministic systems [15]. The goal of a sufficient condition for the convergence of Λ_1 estimates remains allusive. Quantitative necessary conditions, along with explicit tests for convergence in aleatoric systems (stochastic systems with positive Λ_1) are discussed elsewhere [28,37].

In terms of identifying the 'worst forecast bust', the λ_1 are more important than the ρ_1 simply because the λ_1 are larger. While it is sometimes argued that the corresponding singular vectors may point 'off the attractor', the λ_1 remain relevant, as possible uncertainties about the true initial state will also lie 'off the attractor', almost certainly. Infinitesimal uncertainties along $\boldsymbol{l}_1(\boldsymbol{x})$ have the advantage to be free of transients, but if of finite magnitude they also may lie 'off the attractor'. And even for finite uncertainties 'on the attractor', finite time growth is not bound by ρ_1. These facts imply that the 'super-Lyapunov' growth found by Nicolis et al. [8] is to be expected: after time t an uncertainty may be magnified by more than the larger of $2^{\Lambda_1 t}$ and $2^{\rho_1 t}$, even if the initial uncertainty is infinitesimal. Over what duration can realistic (i.e. operational) uncertainties be treated as infinitesimal? Or equivalently, what is the extent of the *linear regime*? This is an interesting and open question, even in numerical weather forecasting [38].

Note that computing exponents for finite time is somewhat gratuitous in that *any* increase will yield a positive *effective* exponent; a positive exponent implies effectively exponential growth then only in the limit of infinite time. For finite time, a positive exponent implies growth, but not exponential growth; it only reflects the time dependence of the uncertainty under the *additional* assumption that the growth was exponential. The width of the distributions in Figs. 1 and 2 does not indicate uniform exponential growth on these time scales. An alternative approach to quantify predictability by computing the time required to reach an uncertainty threshold is contrasted with the use of effective rates in [7,9], where examples with *both* large Λ_1 *and* large uncertainty doubling times are discussed. As proven in Section 4, there are initial conditions for which *no* perturbations grow for two paradigm attractors; it would be interesting to investigate the relative location of regions within which $\lambda_1^{(k)} < 0$ and unstable period k orbits for large k, as a function of parameter in a variety of low dimensional maps; the numerics near stable period k points with $\lambda_1^{(k)} \gg 0$ may also prove of interest.

In this paper each system has been considered in its natural state space, its original, physically relevant co-ordinate system. It should be noted, that neither the typical measures of forecast error[4] nor the finite time Lyapunov exponents nor the finite sample Lyapunov exponents are invariant under co-

[4] See [39] for an atypical approach.

ordinate changes (or even changes in a Riemannian metric). In this study, the forecast error is simply the Euclidean distance between two points in state space, specifically, $|\boldsymbol{\epsilon}| = \sqrt{\boldsymbol{\epsilon}^T \mathcal{I} \boldsymbol{\epsilon}}$ where $\mathcal{I}$ is the identity matrix. The singular vector corresponding to the largest finite time exponent maximizes this distance at prediction time k. There may be physically more relevant definitions of distance between two forecasts. Then the identity might be replaced by another Riemannian metric $\mathcal{P}$, for example the inverse of the covariance matrix may serve as a natural choice when the different directions in state space display different variances. Alternatively $\mathcal{P}$ may be used to account for different levels of noise on different state space variables, to target the variables whose prediction is of particular concern, or even to decide which variables to observe in order to minimize the prediction error (see [40] and references therein). The application determines the choice of metric.

In conclusion, we again stress that the relevance of *all three* types of exponent is restricted to cases where the uncertainties are sufficiently small that their growth is well approximated by the linear propagator [38]: $\mathcal{M}$ is exact only for infinitesimal uncertainties. Behavior of larger (finite) uncertainties requires the use of ensembles of initial conditions, each consistent with the observation; the relative performance of ensembles in the subspaces defined by $\boldsymbol{v}_1$ are contrasted with those defined in the subspace defined by $\boldsymbol{l}_1$ for several chaotic flows in [9]. The construction of ensembles for forecast evaluation in imperfect models remains an important issue for all nonlinear systems.

Appendix A.

Here we establish that for a product of matrices $\mathcal{C} = \mathcal{B}\mathcal{A}$ we have $\sigma_1^{\mathcal{C}} \le \sigma_1^{\mathcal{B}} \sigma_1^{\mathcal{A}}$.

The spectral norm of a matrix [14] is defined by $|\mathcal{C}| = \max_{\boldsymbol{x} \neq 0} |\mathcal{C}\boldsymbol{x}| / |\boldsymbol{x}|$ where the maximum is taken over all nonzero vectors $\boldsymbol{x}$. It bounds the amplifying power of a matrix, i.e., $|\mathcal{C}\boldsymbol{x}| \le |\mathcal{C}||\boldsymbol{x}|$. The spectral norm of a rotation matrix is 1, while that of a diagonal matrix corresponds to the maximum element. The singular value decomposition decomposes any square matrix into the product of a rotation matrix, a diagonal matrix, and another rotation matrix, hence the spectral norm of a matrix is identical to its first singular value, $|\mathcal{C}| = \sigma_1^{\mathcal{C}}$. Given $\mathcal{C} = \mathcal{B}\mathcal{A}$, then for any nonzero $\boldsymbol{x}$ we have

$$|\mathcal{C}||\boldsymbol{x}| \le |\mathcal{C}\boldsymbol{x}| = |\mathcal{B}(\mathcal{A}\boldsymbol{x})| \le |\mathcal{B}||\mathcal{A}\boldsymbol{x}| \le |\mathcal{B}||\mathcal{A}||\boldsymbol{x}|.$$

Dividing by $|\boldsymbol{x}|$, substitution yields $\sigma_1^{\mathcal{C}} = |\mathcal{C}| = |\mathcal{B}\mathcal{A}| \le |\mathcal{B}||\mathcal{A}| = \sigma_1^{\mathcal{B}} \sigma_1^{\mathcal{A}}$ as desired.

References

[1] V.I. Oseledec, Transactions of the Moscow Mathematical Society 19 (1968) 197.
[2] J.-P. Eckmann, D. Ruelle, Rev. Mod. Phys. 57 (1985) 617.
[3] L. Arnold, Random Dynamical Systems, Springer, Berlin, 1998.
[4] E.N. Lorenz, Tellus 17 (1965) 321.
[5] H.D. I Abarbanel, R. Brown, M.B. Kennel, Int. J. Mod. Phys. B 5 (1991) 1347.
[6] R. Doerner, B. Hübinger, W. Martiensen, S. Grossmann, S. Thomae, Chaos Solitons and Fractals 1 (1991) 553.
[7] L. Smith, Phil. Trans. R Soc. Lond. A (1994).
[8] C. Nicolis, S. Vannitsem, J.-F. Royer, Q.J.R. Meteorol. Soc. 121 (1995) 705.
[9] L. Smith, C. Ziehmann, K. Fraedrich, Q.J.R. Meteorol. Soc. 125 (1999) 2855.
[10] Z. Toth, E. Kalnay, Bull. Am. Meteorol. Soc. 74 (1993) 2317.
[11] T.N. et al., Palmer, Phil. Trans. R. Soc. Lond., 1994.
[12] L.A. Smith, Nonlinear Dynamics and Statistics chapter Disentangling Uncertainty and Error: On the Predictability of Nonlinear Systems, Birkhäuser, Boston 2000.
[13] S. Vannitsem, C. Nicolis, J. Atmos. Sci. 54 (1997) 347.
[14] G. Strang, Linear algebra and its application, Hartcourt Brace Jovanovich, San Diego, 1988.
[15] C. Ziehmann, L.A. Smith, J. Kurths, Physica D 126 (1999) 49.
[16] P. Grassberger, R. Badii, A. Politi, J. Statistical Physics 51 (1988) 135.
[17] E. Ott, Chaos in dynamical systems, Cambridge University Press, Cambridge, New York, Melbourne, 1993.
[18] G. Froyland, K. Judd, A. Mees, Phys. Rev. E 51 (1995) 2844.
[19] J.S. Nicolis, G. Mayer-Kress, G. Haubs, Z. Naturforsch. 38a, 1983, p. 1157–1169.
[20] J.M. Nese, Physica D 35 (1989) 237.
[21] A.S. Pikovsky, Chaos 3 (1993) 225.
[22] B. Eckhardt, D. Yao, Physica D 65 (1993) 100.
[23] U. Feudel, J. Kurths, A.S. Pikovsky, Physica D 88 (1995) 176.
[24] M. Henon, Commun. Math. Phys. 50 (1976) 69.
[25] K. Ikeda, Opt. Commun. 30 (1979) 257.
[26] D. Ruelle, Publications Mathematiques de l'Institut Hautes Etudes Scientifiques 50 (1979) 27.

[27] A. Katok, B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, 1995.
[28] S. Ellner, R. Gallant, D. McGaffrey, D. Nychka, Phys. Lett. A 153 (1991) 357.
[29] C. Ziehmann-Schlumbohm, Vorhersagestudien in chaotischen Systemen und in der Praxis, PhD thesis Freie Universität Berlin. Meteorologische Abhandlungen. Neue Folge Serie A. Band 8 Heft 3 1994.
[30] E.N. Lorenz, J. Atmos. Sci. 20 (1963) 130.
[31] H. Mukougawa, M. Kimoto, S. Yoden, J. the Atmospheric Sciences 48 (1991) 1231.
[32] J.L. Anderson, H.M. van den Dool 122 (1994) 507.
[33] P. Cvitanovic, Phys. Rev. Letters 61 (1988) 2729.
[34] L. Boberg, U. Brosa, Z. Naturforsch. 43a, 1988.
[35] L.N. Trefethen, A.E. Trefethen, S.C. Reddy, T.A. Driscoll, Science, 1993, p. 578–584.
[36] L.A. Smith, Lacunarity and Chaos in Nature, PhD thesis Columbia University, New York, NY. 1987, See Appendix 1.
[37] L. Smith, C. Ziehmann, J. Kurths, 2000.
[38] I. Gilmour, Nonlinear model evaluation: ι-shadowing, probabilistic forecasts, and weather forecasting, PhD thesis Oxford University 1998.
[39] P. McSharry, L.A. Smith, Phys. Rev. Lett. 83 (1999) 4285.
[40] J.A. Hansen, L.A. Smith, J. Atmos. Sci., 1999, in press.